JMIR Biomedical Engineering

Engineering for health technologies, medical devices, and innovative medical treatments and procedures
Volume 5 (2020), Issue 1 ISSN 2561-3278 Editor in Chief: Syed A. A. Rizvi, MD, PhD, MBA, MPH, BSN

Contents

Original Papers

Review

Viewpoints

Original Paper

Telerehabilitation for Patients With Knee Osteoarthritis: A Focused Review of Technologies and Teleservices

MReza Naeemabadi[1*], MSc, BME; Hesam Fazlali[2*], MSc, BME; Samira Najafi[3*], MSc, BME; Birthe Dinesen[1], PhD; John Hansen[4], PhD

[1]Laboratory of Welfare Technologies - Telehealth and Telerehabilitation, Integrative Neuroscience Research Group, Center for Sensory-Motor Interaction, Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

[2]Kinesiology Research Center, Kharazmi University, Tehran, Iran

[3]Department of Treatment Affairs, Mashhad University of Medical Sciences, Mashhad, Iran

[4]Laboratory for Cardio-Technology, Medical Informatics Group, Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

[*]these authors contributed equally

Corresponding Author:
MReza Naeemabadi, MSc, BME
Laboratory of Welfare Technologies - Telehealth and Telerehabilitation, Integrative Neuroscience Research Group, Center for Sensory-Motor Interaction
Department of Health Science and Technology
Aalborg University
Fredrik Bejars Vej 7D
Aalborg, 9220
Denmark
Phone: 45 71348333
Email: mr.naeemabadi@gmail.com

Abstract

Background: Telerehabilitation programs are designed with the aim of improving the quality of services as well as overcoming existing limitations in terms of resource management and accessibility of services. This review will collect recent studies investigating telerehabilitation programs for patients with knee osteoarthritis while focusing on the technologies and services provided in the programs.

Objective: The main objective of this review is to identify and discuss the modes of service delivery and technologies in telerehabilitation programs for patients with knee osteoarthritis. The gaps, strengths, and weaknesses of programs will be discussed individually.

Methods: Studies published in English since 2000 were retrieved from the EMBASE, Scopus, Web of Science, Cumulative Index to Nursing and Allied Health Literature (CINAHL), PubMed, Physiotherapy Evidence Database (PEDro), and PsycINFO databases. The search words “telerehabilitation,” “telehealth,” “telemedicine,” “teletherapy,” and “ehealth” were combined with “knee” and “rehabilitation” to generate a data set of studies for screening and review. The final group of studies reviewed here includes those that implemented teletreatment for patients for at least 2 weeks of rehabilitation.

Results: In total, 1198 studies were screened, and the full text of 154 studies was reviewed. Of these, 38 studies were included, and data were extracted accordingly. Four modes of telerehabilitation service delivery were identified: phone-based, video-based, sensor-based, and expert system–based telerehabilitation. The intervention services provided in the studies included information, training, communication, monitoring, and tracking. Video-based telerehabilitation programs were frequently used. Among the identified services, information and educational material were introduced in only one-quarter of the studies.

Conclusions: Video-based telerehabilitation programs can be considered the best alternative solution to conventional treatment. This study shows that, in recent years, sensor-based solutions have also become more popular due to rapid developments in sensor technology. Nevertheless, communication and human-generated feedback remain as important as monitoring and intervention services.

(JMIR Biomed Eng 2020;5(1):e16991) doi:10.2196/16991

KEYWORDS

telemedicine; telerehabilitation; communication technologies; knee osteoarthritis; total knee replacement

Introduction

Osteoarthritis and chronic musculoskeletal disorders are considered the second most frequent medical condition and are the primary causes of physical disability and pain [1-4]. Knee osteoarthritis (KOA) among seniors is estimated at 10% and 13% in men and women, respectively [5]. Moreover, it has been reported that symptomatic KOA has doubled among women and tripled among men in the past 20 years [6]. Previous studies indicate that postsurgical physical rehabilitation is a crucial component of the recovery process [7] and that exercises and education are frequently recommended for patients with KOA [8].

Telecommunication technologies have been used to provide health care, monitoring, and rehabilitation services for patients who experience stroke [9], pulmonary disorders [10], COPD [11,12], dermatological disorders [13], oral diseases [14], and musculoskeletal conditions [15,16]. Such technologies have also been used for remote consultations [17]. Using internet access, early telehomecare programs were implemented as a substitute for home care visits, and user perception and satisfaction were assessed and reported to be high [18,19]. Russell [20] introduced telerehabilitation as a means of augmenting traditional rehabilitation by employing telecommunication technologies that would provide services such as assessment, education, intervention, and interview. Previous studies indicated that telerehabilitation programs can provide better clinical services in rural and remote communities compared with conventional therapy [21,22], as well as improve cost efficiency and resource management of the services [23-25] with high validity and reliability [26].

Previous studies have noted that a telerehabilitation program for KOA not only improves patients' quality of life [27] but also introduces a better functional recovery after arthroplasty in comparison to conventional therapy [28]. Sharareh et al [29] showed that providing a postoperative telerehabilitation program reduced the frequency of postoperative visits and increased patient satisfaction. Tousignant et al [30] and Chalupka et al [31] also indicated that home telerehabilitation enhanced accessibility to health care services and was as effective as conventional therapy.

Russell divided the technologies used for telerehabilitation into image-based, sensor-based, virtual environment, and virtual reality telerehabilitation [32]. Real-time video conferences were extensively used in physical telerehabilitation [33-36]. Giantomassi et al [37] indicated that current video game technologies could be used in physical rehabilitation. Several telerehabilitation programs have been developed using the Microsoft Kinect sensor [38-41] and the Nintendo Wii board [42-45]. Wearable sensors were also proposed as a means to facilitate a telerehabilitation program and monitor patient performance [46]. Moreover, Strecher [47] remarked that health care services could be delivered using decision-making algorithms and expert systems [48] over the internet. Rini et al [49], using an expert system approach, adapted the face-to-face therapeutic intervention into an internet-based intervention. However, there is no clear vision of the strengths and weaknesses of each solution as well as the existing gaps and limitations of the presented solutions. In addition, there is a lack of focused reviews investigating the presented services as part of telerehabilitation programs.

Therefore, the main objective of this review was to identify and discuss the modes of service delivery and technologies used as a telerehabilitation program for patients with KOA in recent studies. The gaps, strengths, and weaknesses of programs were discussed individually.

Methods

Search Strategy

The search strategy was designed to identify relevant literature regarding telerehabilitation solutions for KOA while focusing on the technologies and services of the programs [15,26,28,50]. The literature search used here investigated studies that had implemented and evaluated a telerehabilitation program for patients with KOA using an experimental study design. The EMBASE, Scopus, Web of Science, CINAHL, PubMed, Physiotherapy Evidence Database (PEDro), and PsycINFO databases were searched. Searches were undertaken in March 2020, and comprised medical subject heading (MeSH) [51] terms and keyword search terms. The MeSH terms "telemedicine," "rehabilitation," and "knee" and keyword terms "telehealth," "ehealth," "teletherapy," "telecare," and "knee" were used. Moreover, both *Telemedicine and e-Health* and the *Journal of Telemedicine and Telecare* were searched independently using knee rehabilitation key search terms.

Research Question

The research questions of this review are the following: (1) Which modes of service delivery were used to establish a telerehabilitation program for patients with KOA? (2) What services were introduced by these programs? (3) What are the strengths and weaknesses of each solution?

Inclusion Criteria

Original English-language studies published from January 2000 to January 2020 were included if they fit the eligibility criteria, which were based on the PICOS framework [52].

Participants

Studies with adult participants (aged 18 years and above) with KOA were included. The studies in which participants' primary medical condition was not related to KOA (eg, stroke, upper limb disability, pulmonary disorders) were excluded.

Intervention

Only studies where telecommunication technology was employed as an interventional rehabilitation method in an experimental or observational study were included. The study intervention had to focus on knee pain management or knee

rehabilitation for a period of at least two weeks via synchronous or asynchronous telerehabilitation (eg, phone, email, website report, videoconference, multimedia messages). Studies with insufficient technical explanations were excluded.

Comparison

All trials were included, whether they did or did not employ a control group.

Study Design

Any randomized controlled trial (RCT), quasi-RCT, non-RCT, controlled clinical trial, and pilot study designs, regardless of the blinding of the assessor, were included. Protocol manuscripts, review studies, abstracts, and guidelines were excluded.

Data Collection

According to the inclusion and exclusion criteria, a two-step study identification and data extraction process was used. Two authors (MRN and HF) independently screened the electronic search results. First, the retrieved studies were screened for eligibility based on their title and abstract. The full text of studies selected in the first stage was then reviewed and analyzed as a candidate for final inclusion. Any disagreement between the two authors was resolved through discussion between the authors; if necessary, a third author (JH) was referred to for arbitration. Two authors (MRN and SN) were responsible for data extraction from the included articles. The extracted items were study design, study population, medicinal condition (population), outcomes, modes of telerehabilitation program (intervention) delivery, and rehabilitation duration.

Results

Study Identification

Figure 1 shows an overview of the relevant study identification process using a four-step PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [53] flow diagram. A total of 1198 studies were identified through the literature search: 119 in EMBASE, 807 in Scopus, 99 in Web of Knowledge, 165 in CINAHL, 97 in PubMed, 8 in PsycINFO, and 11 in the PEDro database. In total, 210 duplicated studies were found in the identified documents, and 909 articles were excluded by screening the titles and abstracts of the identified articles based on the defined inclusion criteria. The full text of 154 papers was reviewed by the authors in the eligibility stage, and 38 of the studies were included in this review. The eligible studies were reviewed, and studies using the same experimental setup and population were grouped together. Eventually, 24 group studies were chosen (Table 1).

Figure 1. Flowchart of the results from the literature search. CINAHL: Cumulative Index to Nursing and Allied Health Literature; PEDro: Physiotherapy Evidence Database.

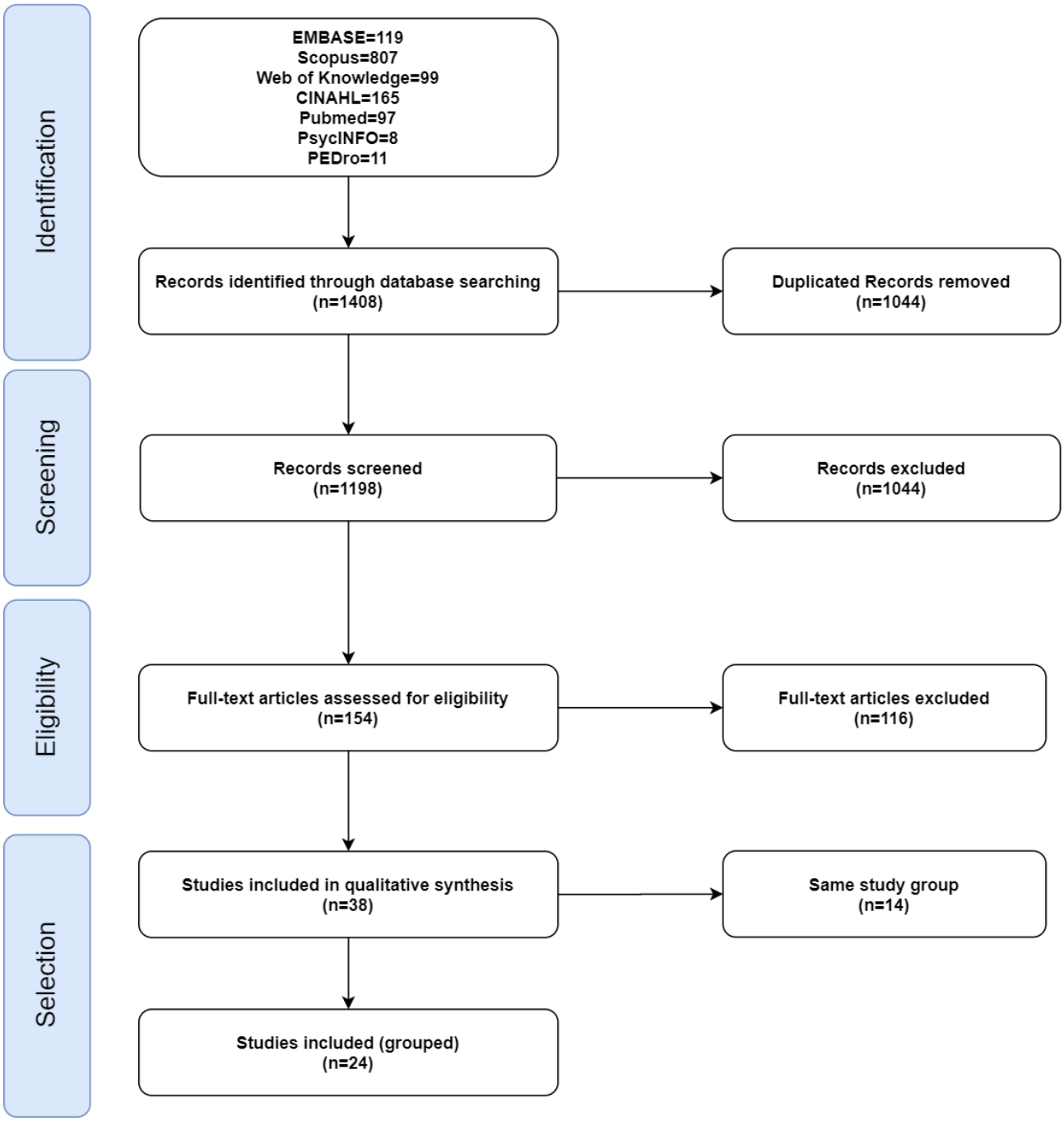


Table 1. Study characteristics.

Group	Author(s), year	Study design[a]	Population[b]	Patients (control, target)	Duration (weeks)	Intervention type[c] and detail	Finding
1	Kramer et al, 2003 [54]	RCT	TKA	80, 80	12	P; Patients got the exercise plan in a booklet and received at least two phone calls within the rehabilitation period	No statistically difference between the groups. A phone call by a physiotherapist can effectively treat patients after TKA.
2	Han et al, 2015 [55]	RCT	TKA	196, 194	6	P; Patients got a daily exercise program and were asked to perform exercises 3 times per day. They received a phone call every week.	The intervention group achieved noninferior outcomes (pain and function) compared with usual care physiotherapy.
3	Chen et al, 2016 [56]	RCT	TKA	101, 101	12	P; The patients received the standard rehabilitation program and were asked to perform exercises for 1 hour per day. They also received 3 phone calls.	A structured telephone follow-up may improve patient adherence as well as enhance patient mental health and range of motion.
4	Azma et al, 2017 [57]	RCT	KOA	27, 27	6	P; Home training with a weekly phone call and logbook	The telerehabilitation was as effective as regular rehabilitation, and no significant difference was observed between the study groups.
5	Wong et al, 2005 [58]	PiS	KnP	—[d], 20	12	V; Video call at the secondary center with a group of patients and unsupervised home training.	Videoconferencing was accepted as a mode of health care service delivery among the users. Significant reductions were observed in pain level and stiffness, and there was an improvement in physical function and the Berg Balance Scale score.
6	Tousignant et al, 2009 [59]	PiS	TKA	—[d], 5	8	V; Over 8 weeks, there were 16 video calls, which included the prescribing of an individualized training program.	A high level of participant satisfaction was achieved, and positive patient-therapist relationships were established.
6[e]	Tousignant et al, 2011 [30]	RCT	TKA	24, 24	8	V; Over 8 weeks, there were 16 video calls, which included the prescribing of an individualized training program.	The home telerehabilitation was as effective as conventional home visits in terms of reducing disability and improving function in the short term.
6	Tousignant et al, 2011 [35]	RCT	TKA	20, 22	8	V; Over 8 weeks, there were 16 video calls, which included the prescribing of an individualized training program.	Patient and therapist satisfaction were high and comparable to that of conventional therapy.
7[e]	Moffet et al, 2015 [60]	RCT	TKA	101, 104	8	V; In-home training using 16 video calls over 8 weeks. Treatment, assessment, and recommendations were considered in the video sessions.	The telerehabilitation was as effective as conventional face-to-face rehabilitation in terms of functional recovery and quality of life.
7	Tousignant et al, 2015 [23]	RCT	TKA	100, 97	8	V; In-home training using 16 video calls over 8 weeks. Treatment, assessment, and recommendations were considered in the video sessions.	The telerehabilitation program was less expensive compared to conventional home visits when the distance between the health care center and patients was more than 30 km.
7	Moffet et al, 2017 [36]	RCT	TKA	98, 84	8	V; In-home training using 16 video calls over 8 weeks. Treatment, assessment, and recommendations were considered in the video sessions.	Patient satisfaction was reported high for both control and intervention groups. No strong correlation was found between reported satisfaction and measurements.
7	Boissy et al, 2015 [61]	Mix	TKA	—[d], 97	8	V; In-home training using 16 video calls over 8 weeks. Treatment, assessment, and recommendations were considered in the video sessions.	The telerehabilitation program was reliable; however, the program required technical maintenance, support, and initial installation.

Group	Author(s), year	Study design[a]	Population[b]	Patients (control, target)	Duration (weeks)	Intervention type[c] and detail	Finding
8	Bini et al, 2017 [62]	RCT	TKA	15, 13	12	V; Asynchronous video communication using a mobile app.	No significant difference was observed between study groups in any of the outcomes. A high level of patient acceptance and satisfaction was achieved. Asynchronous communication may overcome the limitations of real-time telerehabilitation.
9	Doiron-Cadrin et al, 2018 [63]	RCT	HKA	12, 22	12	V; Video telerehabilitation sessions were performed twice per week, and patients were asked to repeat the exercises unsupervised.	High level of patient satisfaction and adherence to the intervention program were achieved. No significant differences were observed between groups for all measurements. Telerehabilitation seems feasible and safe.
10	Eisermann et al, 2004 [64]	RCT	HKA	142, 154	3-4	S; Home training using a computer (software) 3-5 times per week	The provided solution was feasible and as effective as conventional therapy.
11	Piqueras et al, 2013 [65]	RCT	TKA	70, 72	2	S: Included 1 week of on-site rehabilitation and 1 week of home training using a PC and sensors.	The proposed telerehabilitation program was at least as effective as conventional therapy. Significant improvement was observed in active extension range and quadriceps muscle strength.
12	Ayoade and Baillie, 2014 [66]	RCT	TKA	7, 8	6	S; Home training using a computer (software) and a video call (Week 3)	The system was acceptable among the seniors. However; it did not improve their adherence.
13[e]	Correia et al, 2018 [67]	RCT	TKA	30, 29	8	S; Training at home using 3 wearable sensors and a tablet 5-7 days per week.	Performance tests, range of motion, and patient-reported outcomes were significantly higher in the intervention group.
13	Correia et al, 2019 [68]	RCT	TKA	30, 29	8	S; Training at home using 3 wearable sensors and a tablet 5-7 days per week.	High levels of adherence to the program and satisfaction were reported. All the outcomes were significantly higher in the intervention group at 3 months and primary outcomes at 6 months.
14	Argent et al, 2019 [69]	Mix	TKA	—[d], 15	2	S; Training at home using a wearable sensor and customized Android application.	High level of adherence to the intervention was reported, and overall positive user-experience achieved. Technical issues caused some negative experiences.
15	Ramkumar et al, 2019 [70]	FiS	TKA	—[d], 25	12	S; Training at home using a leg sleeve equipped with two wearable sensors communicating with an iPhone. Daily activities were measured based on the internal pedometer of the phone.	Patients found the telerehabilitation program engaging, motivating, and easy to use.
16	Eichler et al, 2019 [71]	RCT	HKA	55, 56	12	S; Microsoft Kinect was used and therapist could actively modify the exercises. Real-time video communication was used to establish the communication.	The effect of the telerehabilitation therapy was equivalent to the usual aftercare in terms of functional testing, quality of life, and pain.
17	Bettger et al, 2019 [72]	RCT	TKA	144, 143	12	S; An interactive training program was presented, and Microsoft Kinect was used to track the exercises. The therapist could track patient performance and modify the training program. The system provided real-time video communication as well.	The program was as effective as traditional treatment (function and disability) and as safe as traditional treatment (pain and rehospitalization).

Group	Author(s), year	Study design[a]	Population[b]	Patients (control, target)	Duration (weeks)	Intervention type[c] and detail	Finding
18	Kuether et al, 2019 [73]	PiS	HKA	—[d], 40	≈ 8	S; An interactive training program was presented and Microsoft Kinect was used to track the exercises. The therapist could track patient performance and modify the training program. The system provided real-time video communication as well.	High satisfaction rate was observed among the patients, and comparable improvement in the patient-reported outcomes were achieved.
19	Chughtai et al, 2019 [74]	FiS	KA	—[d], 157	≈ 4	S; An interactive training program was presented and Microsoft Kinect was used to track the exercises. The therapist could track patient performance and modify the training program. The system provided real-time video communication as well.	The intervention was cost-effective, convenient, and improved patient adherence and overall satisfaction.
20	Bossen et al, 2013 [75]	PiS	HKO	—[d], 20	9	E; Web-based exercise management gradually increases the activity level over 9 weeks without physiotherapist involvement.	The intervention was feasible and accepted among the patients. Patient satisfaction was reported high.
20	Bossen et al, 2013 [76]	Mix	HKO	—[d], 100	9	E; Web-based exercise management gradually increases the activity level over 9 weeks without physiotherapist involvement.	Higher age, presence of comorbidity, lack of self-discipline, and physical activity baseline have a negative impact on the intervention adherence rate.
20[e]	Bossen et al, 2013 [77]	RCT	HKO	99, 100	9	E; Web-based exercise management gradually increases the activity level over 9 weeks without physiotherapist involvement.	In the intervention group, physical function status improved in the short term; in the long term, higher levels of subjective and objective physical activity were observed.
21	Rini et al, 2015 [78]	RCT	HKO	55, 58	8-10	E; Automatically generated training program and management based on feedback without therapist participation.	The intervention group reported significantly lower pain compared with the control group and a very high level of adherence to the intervention program was observed. The acceptability of the program was demonstrated based on the strong evidence.
22	Kim et al, 2016 [79]	RCT	KnP	20, 50	6	E; Training program generated and updated automatically with therapist involvement.	No statistical difference was observed between the control and intervention groups. Only 8% of the participants completed the program.
23	Hinman et al, 2017 [33]	Mix	KnP/KOA	—[d], 12	12	E/V; Home training (3 times per week) and 7 video calls during the treatment period.	Patients were satisfied with the provided solution due to the time efficiency, flexibility of service, and ease of access. The solution also improved users' confidence and sense of self-efficacy.
23[e]	Bennell et al, 2017 [80]	RCT	KnP	74, 74	12	E/V; Home training (3 times per week) and 7 video calls during the treatment period.	Short-term and long-term effectiveness of the intervention was achieved. A significant short-term improvement was observed in the primary outcomes.
23	Lawford et al, 2018 [81]	RCT	KnP	74, 74	12	E/V; Home training (3 times per week) and 7 video calls during the treatment period.	Employed participants with a higher self-efficacy had better improvement in their health condition compared to the rest of the study group.

Group	Author(s), year	Study design[a]	Population[b]	Patients (control, target)	Duration (weeks)	Intervention type[c] and detail	Finding
24	Bossen et al, 2016 [82]	Mix	HKO	—[d], 8	9	E; Web-based exercise management with gradual increases in the activity level over 12 weeks; the therapist could update the program, which included 5 face-to-face visits.	The proposed treatment seems to be feasible, and a good level of usability was reported. Lack of possibility to monitor patients between sessions was considered a limitation.
24	Vries et al, 2017 [83]	Mix	HKO	—[d], 90	12	E; Web-based exercise management with gradual increases in the activity level over 12 weeks; the therapist could update the program, which included 5 face-to-face visits.	Internet skills, self-discipline, usability of the intervention, added value, time required, flexibility, and execution of the exercise plan and participating in research identified as determinants of patients' adherence.
24	Kloek et al, 2018 [84]	RCT	HKO	99, 109	12	E; Web-based exercise management with gradual increases in the activity level over 12 weeks; the therapist could update the program, which included 5 face-to-face visits.	The intervention cost was significantly lower in the intervention group. However; the total societal and healthcare costs were not statistically significant between groups.
24	Kloek et al, 2018 [85]	Mix	HKO	—[f], —[f]	12	E; Web-based exercise management with gradual increases in the activity level over 12 weeks; the therapist could update the program, which included 5 face-to-face visits.	Appropriateness, required time, workload, added value, environmental factors, professional autonomy, and financial consequences were identified as determinants for physiotherapists' usage of the proposed intervention
24[e]	Kloek et al, 2018 [86]	RCT	HKO	99, 109	12	E; Web-based exercise management with gradual increases in the activity level over 12 weeks; the therapist could update the program, which included 5 face-to-face visits.	No statistical differences were found between control and intervention groups in physical functioning and free-living physical activity.

[a]The study designs are randomized controlled trial (RCT), mixed method study (Mix), quasi-experimental study (QeS), pilot study (PiS), feasibility study (FiS), and cross-sectional study (CrS).

[b]The population groups are total hip/knee arthroplasty (HKA), knee osteoarthritis (KOA), total knee arthroplasty (TKA), knee pain (KnP), and hip/knee osteoarthritis (HKO).

[c]For the intervention type, the designations P, V, S, and E stand for phone-based, video-based, sensor-based, and expert system–based telerehabilitation, respectively.

[d]Not available.

[e]These studies are considered the reference study of their group.

[f]There were 123 physiotherapists that participated in the study.

Technologies and Mode of Rehabilitation Service Delivery

We identified four modes of telerehabilitation programs: phone-based, video-based, sensor-based, and expert system–based; we focused on the given training solution.

Phone-Based Telerehabilitation

Azma et al [57] and Han et al [55] used weekly phone communication initiated by a health care professional who would instruct patients and track their progress over the rehabilitation period. Kramer et al [54] and Chen et al [56] performed less frequent phone calls during the rehabilitation period. In the studies, training instructions were provided using a guidebook of the exercises. Information about adherence to the exercise programs was collected by filling out a logbook of activities [57] or during the phone calls [55].

Video-Based Telerehabilitation

In previous studies, real-time video streaming communication was frequently used to deliver rehabilitation services to patients. Wong et al [58] established a weekly supervised training session with a group of patients using video conference communication at a secondary center. Patients were asked to perform the prescribed exercises 3 times per week for 12 weeks using a booklet of exercise instructions; patient activity/adherence was reported using a logbook. Tousignant et al [30] provided a home telerehabilitation service by setting up a video conferencing system in the patient's home, and rehabilitation sessions were carried out twice per week for 8 weeks. Moffet et al [60] developed the telerehabilitation solution by employing a hardware/software video communication platform (TelAge), which was the treatment program for the target group. In the TelAge system, custom computer software (TeRa) was developed to control the streaming video with a user-friendly

graphical interface [61]. Doiron-Cadrin et al [63] employed a medical teleconsultation application (REACTS Lite, Innovative Imaging Technologies) to establish real-time video communication.

Bini et al [62] introduced asynchronous video communication by employing smart devices (iPod Touch, Apple Inc) and media file–sharing applications (CaptureProof). The system established a two-way asynchronous communication between the physiotherapist and patient. It also enabled the physiotherapist to instruct the patient using prerecorded exercise introductions and provide supplementary media.

Sensor-Based Telerehabilitation

Eisermann et al [64] provided a telerehabilitation program using custom computer software and several sensors to track the patient's performance. Patients were asked to perform individualized exercises based on the training program prescribed by the therapist; the program could be modified based on the patient's feedback. Accelerometers, webcams, chest sensors, and wristbands were employed in the study to monitor training performance and generate relevant reports and online feedback.

Ayoade et al [66] and Piqueras et al [65] developed a telerehabilitation program by using two wireless sensors equipped with 9 degrees of freedom (9DOF) inertial measurement units to track the knee angle. Participants were asked to wear the sensors on their operated leg (shin and thigh) using elastic bands while performing the recommended exercises. Argent et al [69] used a classic Bluetooth 9DOF sensor (Shimmer3, Shimmer Sensing) fixed on the patient's shin, and Correia et al [67] increased the number of Bluetooth Low Energy 9DOF sensors to three; these were placed on the chest, thigh, and shin. Ramkumar et al [70,87] employed Focus Motion (Focus Ventures) sleeves to track the operated knee's range of motion. The sleeve was designed for the lower limb and equipped with two classic Bluetooth 9DOF sensors. In addition, the user's cellphone was used to track daily activities based on the internal pedometer.

Eichler et al [71,88] used Microsoft Kinect (Version 2, Microsoft Corp) in the training program to track the patient's performance. The VERA (Reflexion Health) system also used Microsoft Kinect to track the exercises and it has been used in several clinical studies [72-74]. The Microsoft Kinect software development kit (Version 2.0) [89] can provide an estimation of 25 joins (including the knee) in space. Therefore, the telerehabilitation program could produce an avatar of the user performing the exercises.

All the introduced telerehabilitation programs were able to track the number of performed exercises and to provide real-time visual feedback on user performance.

Expert System–Based Telerehabilitation

Bossen et al [77] provided a web-based training program (Join2Move). The training program was automatically generated based on reported baseline measurements. The intensity of the exercises was increased over time, based on the behavioral graded activity concept [90]. The expert system collected weekly patient adherence reports and provided autogenerated messages and reports without any intervention by the physiotherapist. The telerehabilitation program was improved and developed using a participatory design method [82,91]. In the improved program (E-exercise), online information and 5 face-to-face visits were included in the internet-based intervention [86]. Moreover, in E-exercise, it was observed that the therapist could deviate from the suggested training program.

Kim et al [79] used a decision-making system that introduced an adaptive training program based on the patient's adherence, pain level, and difficulty reports. However, the physiotherapists were not involved in adjusting the training program; they were able to monitor the patient's reports and respond to the patient's questions via a text messaging service embedded in the program.

Rini et al [78] employed an expert system to provide internet-based pain coping skills training (PainCoach). An individualized training program was automatically generated based on the patient's baseline without physiotherapist participation. In addition, the program enabled the patients to access the appropriate instructions and a history of their performance; they could also ask other patients about their experiences and share experiences. Bennell et al [80] extended the PainCoach program by including 7 videoconference sessions with a physiotherapist over 12 weeks.

Intervention and Services

In total, five different services were identified in the included papers. Table 2 shows the introduced services as part of the telerehabilitation program. The details regarding services were provided as follows.

Information

The information service provided relevant educational material for the target group and was accessible on a 24/7 basis without any interruption. The information provided could cover a wide variety of instructions and answer questions; in addition, it could introduce critical challenges that the patient might encounter. Only four of the studies introduced education materials via an online service as a part of the rehabilitation program.

Communication

The two-way communication between the patient and health care professionals can be used for consultation, recommendation, and interview purposes. The majority of the studies provided this service (19 studies). Real-time communication (using the phone or a video call) was used more than asynchronized communication (asynchronous SMS text messaging or video messaging). In the phone-based and video-based solutions, the communication platform was also used to deliver the training services. It should be mentioned that two of the studies [67,86] chose to include regular in-person visits. Eichler et al [71] used both real-time videoconferencing and asynchronous messaging approaches in the program.

Training

The training service includes exercise instructions, daily/weekly rehabilitation plans (number of repetitions and sets for each exercise) and relevant interactive materials for each exercise. All the studies provided training services using an interactive

training program with visual feedback, video rehabilitation sessions, or a printed booklet of instructions.

Intervention

Intervention in a telerehabilitation program can be carried out based on the patient's reports and is done individually by either a physiotherapist or a decision-making algorithm (as part of an expert system). Making adjustments and modifications to the training program (ie, repetitions, intensity, number of exercises) and providing relevant feedback are considered as interventions in the treatment.

Monitoring and Tracking

Monitoring and tracking services enable physiotherapists or an expert system to perform a predefined assessment or diagnosis remotely. In addition, this service may provide a history of the patient's performance. The data can be recorded manually by a physiotherapist, self-reported by the patient (such as adherence or pain level), or collected automatically using motion tracking sensors.

Table 2. Details of the services provided in the included telerehabilitation programs.

Group	Main study	Information	Communication	Training	Intervention	Tracking
1	Kramer et al, 2003 [54]	Booklet	Phone call	Instruction booklet of the exercises	Phone call	N/A[a]
2	Han et al, 2015 [55]	N/A	Phone call	A hard copy of the instructions	Phone call	Adherence (by phone call)
3	Chen et al, 2016 [56]	N/A	Phone call	A hard copy of the instructions	Phone call	Unclear
4	Azma et al, 2017 [57]	N/A	Phone call	Instruction booklet of the exercises	Phone call	Adherence logbook
5	Wong et al, 2005 [58]	N/A	Real-time video communication	Instruction booklet of the exercises	Unclear	Real-time video communication
6	Tousignant et al, 2011 [30]	N/A	Real-time video communication	Real-time video communication	A therapist can modify the training program	Real-time video communication
7	Moffet et al, 2015 [60]	N/A	Real-time video communication	Real-time video communication	A therapist can modify the training program	Real-time video communication
8	Bini et al, 2017 [62]	N/A	Asynchronous video communication	Video instruction of the training program	A therapist can modify the training program	Video reports
9	Doiron-Cadrin et al, 2018 [63]	N/A	Real-time video communication	Real-time video communication	Real-time video communication	Adherence using logbook
10	Eisermann et al, 2004 [64]	N/A	Asynchronous text messaging	An interactive training program with real-time feedback using motion sensors	A therapist can modify the training program	Patient performance collected by sensors and reported
11	Piqueras et al, 2013 [65]	N/A	Unclear	An interactive training program with real-time feedback using motion sensors	A therapist can modify the training program	Patient performance collected by sensors and reported
12	Ayoade and Baillie, 2014 [66]	N/A	Real-time video communication	An interactive training program with real-time feedback using motion sensors	Real-time video communication	Unclear
13	Correia et al, 2018 [67]	N/A	Face-to-face, phone call	Visual real-time feedback and audio instructions using motion sensors	A therapist can modify the training program	System generated performance
14	Argent et al, 2019 [69]	Provided by the application	Unclear	An interactive training program with real-time feedback using motion sensors	Unclear	Adherence (performance) collected by the system and patient reports (pain, difficulty)
15	Ramkumar et al, 2019 [70]	Unclear	Unclear	An interactive training program with real-time feedback using motion sensors	Unclear	Adherence and daily steps were collected automatically, and patient-reported outcome data were reported every week electronically.
16	Eichler et al, 2019 [71]	N/A	Asynchronous SMS text and voice messaging as well as real-time video communication	An interactive training program with real-time feedback using Microsoft Kinect	A therapist can modify the training program.	Patient adherence and performance collected and reported
17	Bettger et al, 2019 [72]	N/A	Real-time video communication	An interactive training program with real-time feedback using Microsoft Kinect	A therapist can modify the training program.	Patient adherence and performance collected and reported

Group	Main study	Information	Communication	Training	Intervention	Tracking
18	Kuether et al, 2019 [73]	N/A	Real-time video communication	An interactive training program with real-time feedback using Microsoft Kinect	A therapist can modify the training program.	Adherence to the program reported automatically.
19	Chughtai et al, 2019 [74]	N/A	Real-time video communication	An interactive training program with real-time feedback using Microsoft Kinect	A therapist can modify the training program.	Adherence to the program reported automatically.
20	Bossen et al, 2013 [77]	Online education materials	N/A	Interactive training program	The expert system modifies the training program	Adherence report
21	Rini et al, 2015 [78]	Online education materials	N/A	Interactive training program	The expert system modifies the training program	Patient reports
22	Kim et al, 2016 [79]	Unclear	Asynchronous text messaging	Interactive training program	The expert system modifies the training program	Patient reports
23	Bennell et al, 2017 [80]	Online education materials	Real-time video communication	Interactive training program	The expert system modifies the training program	Unclear
24	Kloek et al, 2018 [86]	Online education materials	Face-to-face	Interactive training program	The expert system and therapist modify the training program	Adherence report
Overall	N/A	6	19	24	21	20

[a]N/A: not applicable.

Discussion

This focused review had two purposes: to investigate the technologies used in telerehabilitation programs for patients with KOA and to identify the services that were introduced for the target group. The review identified 24 group studies. The majority of studies (87.5%) were conducted in the second decade of the investigation period (2010-2020) and half of the studies published in the last 4 years. Four different modes of service delivery and five groups of services were identified. The findings showed that video-based communication was the most well-established mode of service delivery, and the studies primarily emphasized establishing training and intervention services rather than providing online education materials.

It is believed that limited services can be provided using a phone-based telerehabilitation program such as phone consultations, recommendations, and interviews [57,92,93], while real-time video communication can be seen as an alternative implementation of an in-person physiotherapy session. Real-time video can be employed not only for consultation purposes but also for training, intervention, and assessment services [30,94,95]. Video conferencing enables the physiotherapist to provide individualized instructions, feedback, and training programs for each patient in real time [59]. Cottrell et al [15] also concluded that real-time video telerehabilitation might be as effective as conventional therapy. Furthermore, the clinical assessment can be carried out by a physiotherapist via the visual observation of a patient's performance while performing a clinical test [59]. Capturing high-quality still images to assess the range of motion of a patient's knee is also recommended [96,97]. In addition, it has been shown that video conferencing can be carried out using a low-bandwidth internet connection [94,98].

Bini et al [62] remarked that real-time video communication might involve several limitations when compared to asynchronous video communication, such as time restrictions and limited or no access to the previous records. The video conference session is usually conducted according to predefined schedules (for example, twice per week in [30,60]), and patients were asked to repeat the exercises without any supervision [63]. Consequently, the patient cannot initiate on-demand communication. Moreover, storing the real-time video stream for later use requires more complicated infrastructure; therefore, neither the physiotherapist nor patients would have access to the previous sessions to track treatment progress. Russell et al [99] recommended a store and forward method to provide video instructions with higher quality. Perez-Manchon et al [100] stated that asynchronous telemedicine could be an efficient method for providing health care services at a distance. However, this method still requires the active engagement of the therapists to review the recorded video session.

Sensor-based telerehabilitation programs are a more independent service than video communication as they use one or more sensors to record the patient's physical activity and collect movement information using software running on a computer device (PC, smartphone, tablet). Ayoade and Baillie [66] showed that a sensor-based telerehabilitation program can offer increased time flexibility and independence for patients by using semi supervised training sessions. Interactive telerehabilitation can be provided by presenting real-time graphical feedback of the

patient's performance using the software. In addition, interactive telerehabilitation enables users to review their performance over time. Moreover, several services can also be provided by software such as asynchronous or synchronous communication, educational materials, and patient reports. The program can be divided into wearable sensor–based and Microsoft Kinect–based programs. Earlier studies remarked that wearable sensors can provide accurate and precise details of movement [101-103]. However, several limitations were observed in the studies. Ayoade and Baillie [66] evaluated the system on only five patients, and Piqueras et al [65] reported that patients used the program for five days. Argent et al [69] reported that patients had a negative experience due to inconsistencies in the automatic measuring system. Ramkumar et al [70] remarked that the users did not appreciate the frequent charging of sensors. In addition, Naeemabadi et al [104] showed that Bluetooth Low Energy–based motion sensors might have a less accurate estimation of the sensor orientation due to a low sampling rate. Moreover, the sampling rate will also decline when the number of sensors is increased. Therefore, further investigation might be required to assess the real-time responsiveness of telerehabilitation programs like those introduced by Correia et al [67,68]. In general, wearable sensors can only represent the orientation of the limb to which they are attached. Therefore, the depicted avatar cannot represent the movement of the whole body. However, Microsoft Kinect–based solutions can track the whole body without the need to wear sensors. In addition, no calibration process is required, and users can immediately start the exercises. The portability of these solutions are debatable due to the computational requirement for Kinect. Eichler et al [71] used a small form factor PC attached to the user's TV; for the VERA solution, the PC, display, and Kinect sensor were placed in a case. Conversely, cellphones and tablets were used in the wearable sensor–based solutions. It was also shown that Microsoft Kinect might impose practical limitations on particular exercises [105]. Microsoft Kinect requires a large space to track the body and Eichler et al [71] consider this in the inclusion criteria. Hence, using Microsoft Kinect to track the exercises might be controversial. It is believed that we still lack a robust solution for the sensor-based telerehabilitation program. Hence, further studies are needed to provide a better understanding of these challenges.

The recent investigations showed that the expert system could partly or entirely interact with the patient and take the therapist's responsibilities. In only 2 out of the 5 identified studies, the expert system was entirely responsible for managing and supervising the treatment procedure, and the patient-to-therapist connection was disrupted [77,78]. Later on, both studies saw improvements, with physiotherapists being more involved in the treatment process via face-to-face visits [86] and video conference communication [80]. Kim et al [79] also used an asynchronous communication with a physiotherapist in the expert system that they introduced. We conclude that the expert system can be effectively used as an assistant system that allows the physiotherapist to have responsibility and maintain physiotherapist-patient telecommunication.

Russell [32] also recognized virtual reality–based telerehabilitation programs as a mode of telerehabilitation. However, we were not able to identify any studies that employed this mode based on the taxonomy of virtual reality displays [106].

In summary, video conference–based programs can be considered the well-established alternative solution to the conventional rehabilitation program for the target group; however, there remain several limitations, such as flexibility and resource management. The recent studies justified the effectiveness of this approach. Although sensor-based solutions might offer higher flexibility and better resource management. The investigations indicated, more studies are being conducted utilizing the sensor technology as a telerehabilitation in the last two years thanks to the existing demand for a more flexible and portable telerehabilitation with better human resource management.

Acknowledgments

This study was supported by the Aage and Johanne Louis-Hansen Foundation, Aalborg University, as well as the Orthopedic Surgery Research Unit, Science and Innovation Center, Aalborg University Hospital, Aalborg, Denmark.

Conflicts of Interest

None declared.

References

1. Guccione AA, Felson DT, Anderson JJ, Anthony JM, Zhang Y, Wilson PW, et al. The effects of specific medical conditions on the functional limitations of elders in the Framingham Study. Am J Public Health 1994 Mar;84(3):351-358. [doi: 10.2105/ajph.84.3.351] [Medline: 8129049]
2. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. In: The Lancet. Amsterdam, Netherlands: The Lancet; Dec 15, 2012:2163-2196.
3. Woolf AD. Global burden of osteoarthritis and musculoskeletal diseases. BMC Musculoskelet Disord 2015 Dec 1;16(S1). [doi: 10.1186/1471-2474-16-s1-s3]

4. Smith E, Hoy DG, Cross M, Vos T, Naghavi M, Buchbinder R, et al. The global burden of other musculoskeletal disorders: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014 Aug 03;73(8):1462-1469. [doi: 10.1136/annrheumdis-2013-204680] [Medline: 24590181]
5. Zhang Y, Jordan JM. Epidemiology of osteoarthritis. Clin Geriatr Med 2010 Aug;26(3):355-369 [FREE Full text] [doi: 10.1016/j.cger.2010.03.001] [Medline: 20699159]
6. Nguyen U, Zhang Y, Zhu Y, Niu J, Zhang B, Felson DT. Increasing prevalence of knee pain and symptomatic knee osteoarthritis: survey and cohort data. Ann Intern Med 2011 Dec 06;155(11):725-732 [FREE Full text] [doi: 10.7326/0003-4819-155-11-201112060-00004] [Medline: 22147711]
7. Artz N, Elvers KT, Lowe CM, Sackley C, Jepson P, Beswick AD. Effectiveness of physiotherapy exercise following total knee replacement: systematic review and meta-analysis. BMC Musculoskelet Disord 2015 Feb 07;16(1):15 [FREE Full text] [doi: 10.1186/s12891-015-0469-6] [Medline: 25886975]
8. da Costa BR, Vieira ER, Gadotti IC, Colosi C, Rylak J, Wylie T, et al. How Do Physical Therapists Treat People with Knee Osteoarthritis, and What Drives Their Clinical Decisions? A Population-Based Cross-Sectional Survey. Physiother Can 2017 Feb;69(1):30-37 [FREE Full text] [doi: 10.3138/ptc.2015-83] [Medline: 28154442]
9. Lai JC, Woo J, Hui E, Chan W. Telerehabilitation - a new model for community-based stroke rehabilitation. J Telemed Telecare 2004 Jun 24;10(4):199-205. [doi: 10.1258/1357633041424340] [Medline: 15273029]
10. Inskip JA, Lauscher HN, Li LC, Dumont GA, Garde A, Ho K, et al. Patient and health care professional perspectives on using telehealth to deliver pulmonary rehabilitation. Chron Respir Dis 2018 Feb 18;15(1):71-80 [FREE Full text] [doi: 10.1177/1479972317709643] [Medline: 28569116]
11. Pinnock H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicentre, randomised controlled trial. BMJ 2013 Oct 17;347(oct17 3):f6070-f6070 [FREE Full text] [doi: 10.1136/bmj.f6070] [Medline: 24136634]
12. Zanaboni P, Dinesen B, Hjalmarsen A, Hoaas H, Holland AE, Oliveira CC, et al. Long-term integrated telerehabilitation of COPD Patients: a multicentre randomised controlled trial (iTrain). BMC Pulm Med 2016 Aug 22;16(1):126 [FREE Full text] [doi: 10.1186/s12890-016-0288-z] [Medline: 27549782]
13. Piette E, Nougairède M, Vuong V, Crickx B, Tran V. Impact of a store-and-forward teledermatology intervention versus usual care on delay before beginning treatment: A pragmatic cluster-randomized trial in ambulatory care. J Telemed Telecare 2016 Aug 05;23(8):725-732. [doi: 10.1177/1357633x16663328]
14. Estai M, Kanagasingam Y, Xiao D, Vignarajan J, Huang B, Kruger E, et al. A proof-of-concept evaluation of a cloud-based store-and-forward telemedicine app for screening for oral diseases. J Telemed Telecare 2016 Sep 08;22(6):319-325. [doi: 10.1177/1357633X15604554] [Medline: 26377126]
15. Cottrell MA, Galea OA, O'Leary SP, Hill AJ, Russell TG. Real-time telerehabilitation for the treatment of musculoskeletal conditions is effective and comparable to standard practice: a systematic review and meta-analysis. Clin Rehabil 2017 May 02;31(5):625-638. [doi: 10.1177/0269215516645148] [Medline: 27141087]
16. Vaish A, Ahmed S, Shetty A. Remote physiotherapy monitoring using the novel D+R Therapy iPhone application. J Clin Orthop Trauma 2017 Jan;8(1):21-24 [FREE Full text] [doi: 10.1016/j.jcot.2016.08.008] [Medline: 28360491]
17. Deldar K, Bahaadinbeigy K, Tara A. Teleconsultation and Clinical Decision Making: a Systematic Review. Acta Inform Med 2016 Jul 16;24(4):286-292 [FREE Full text] [doi: 10.5455/aim.2016.24.286-292] [Medline: 27708494]
18. Finkelstein SM, Speedie SM, Demiris G, Veen M, Lundgren JM, Potthoff S. Telehomecare: quality, perception, satisfaction. Telemed J E Health 2004 Jun;10(2):122-128. [doi: 10.1089/tmj.2004.10.122] [Medline: 15319041]
19. Demiris G, Speedie SM, Finkelstein S. Change of patients' perceptions of TeleHomeCare. Telemed J E Health 2001 Sep;7(3):241-248. [doi: 10.1089/153056201316970948] [Medline: 11564360]
20. Russell TG. Telerehabilitation: a coming of age. Australian Journal of Physiotherapy 2009;55(1):5-6. [doi: 10.1016/s0004-9514(09)70054-6]
21. Cary MP, Spencer M, Carroll A, Hand DH, Amis K, Karan E, et al. Benefits and Challenges of Delivering Tele-rehabilitation Services to Rural Veterans. Home Healthcare Now 2016;34(8):440-446. [doi: 10.1097/nhh.0000000000000441]
22. Wiborg A, Widder B. Teleneurology to Improve Stroke Care in Rural Areas. Stroke 2003 Dec;34(12):2951-2956. [doi: 10.1161/01.str.0000099125.30731.97]
23. Tousignant M, Moffet H, Nadeau S, Mérette C, Boissy P, Corriveau H, et al. Cost analysis of in-home telerehabilitation for post-knee arthroplasty. J Med Internet Res 2015 Mar 31;17(3):e83 [FREE Full text] [doi: 10.2196/jmir.3844] [Medline: 25840501]
24. Kairy D, Lehoux P, Vincent C, Visintin M. A systematic review of clinical outcomes, clinical process, healthcare utilization and costs associated with telerehabilitation. Disabil Rehabil 2009 Jul 07;31(6):427-447. [doi: 10.1080/09638280802062553] [Medline: 18720118]
25. Fusco F, Turchetti G. Telerehabilitation after total knee replacement in Italy: cost-effectiveness and cost-utility analysis of a mixed telerehabilitation-standard rehabilitation programme compared with usual care. BMJ Open 2016 May 17;6(5):e009964 [FREE Full text] [doi: 10.1136/bmjopen-2015-009964] [Medline: 27188803]

26. Mani S, Sharma S, Omar B, Paungmali A, Joseph L. Validity and reliability of Internet-based physiotherapy assessment for musculoskeletal disorders: a systematic review. J Telemed Telecare 2016 Mar 31;23(3):379-391. [doi: 10.1177/1357633x16642369]
27. Odole AC, Ojo OD. Is telephysiotherapy an option for improved quality of life in patients with osteoarthritis of the knee? Int J Telemed Appl 2014;2014:903816-903819 [FREE Full text] [doi: 10.1155/2014/903816] [Medline: 24778645]
28. Jiang S, Xiang J, Gao X, Guo K, Liu B. The comparison of telerehabilitation and face-to-face rehabilitation after total knee arthroplasty: A systematic review and meta-analysis. J Telemed Telecare 2016 Dec 27;24(4):257-262. [doi: 10.1177/1357633x16686748]
29. Sharareh B, Schwarzkopf R. Effectiveness of telemedical applications in postoperative follow-up after total joint arthroplasty. J Arthroplasty 2014 May;29(5):918-922.e1. [doi: 10.1016/j.arth.2013.09.019] [Medline: 24342278]
30. Tousignant M, Moffet H, Boissy P, Corriveau H, Cabana F, Marquis F. A randomized controlled trial of home telerehabilitation for post-knee arthroplasty. J Telemed Telecare 2011 Mar 11;17(4):195-198. [doi: 10.1258/jtt.2010.100602] [Medline: 21398389]
31. Chalupka S. Internet-based outpatient telerehabilitation following total knee arthroplasty. AAOHN J 2011 Mar;59(3):144-144. [doi: 10.3928/08910162-20110223-05] [Medline: 21366204]
32. Russell TG. Physical rehabilitation using telemedicine. J Telemed Telecare 2007 Jun 23;13(5):217-220. [doi: 10.1258/135763307781458886] [Medline: 17697506]
33. Hinman RS, Nelligan RK, Bennell KL, Delany C. "Sounds a Bit Crazy, But It Was Almost More Personal:" A Qualitative Study of Patient and Clinician Experiences of Physical Therapist-Prescribed Exercise For Knee Osteoarthritis Via Skype. Arthritis Care Res (Hoboken) 2017 Dec 02;69(12):1834-1844 [FREE Full text] [doi: 10.1002/acr.23218] [Medline: 28217864]
34. Chumbler NR, Quigley P, Li X, Morey M, Rose D, Sanford J, et al. Effects of Telerehabilitation on Physical Function and Disability for Stroke Patients. Stroke 2012 Aug;43(8):2168-2174. [doi: 10.1161/strokeaha.111.646943]
35. Tousignant M, Boissy P, Moffet H, Corriveau H, Cabana F, Marquis F, et al. Patients' satisfaction of healthcare services and perception with in-home telerehabilitation and physiotherapists' satisfaction toward technology for post-knee arthroplasty: an embedded study in a randomized trial. Telemed J E Health 2011 Jun;17(5):376-382. [doi: 10.1089/tmj.2010.0198] [Medline: 21492030]
36. Moffet H, Tousignant M, Nadeau S, Mérette C, Boissy P, Corriveau H, et al. Patient Satisfaction with In-Home Telerehabilitation After Total Knee Arthroplasty: Results from a Randomized Controlled Trial. Telemed J E Health 2017 Feb;23(2):80-87. [doi: 10.1089/tmj.2016.0060] [Medline: 27529575]
37. Giantomassi A, Capecci M, Benettazzo F, Iarlori S, Ferracuti F, Freddi A, et al. Training and retraining motor functions at home with the help of current technology for video games: Basis for the project. In: Ambient Assisted Living. Switzerland: Springer Nature; 2015:439-447.
38. Antón D, Nelson M, Russell T, Goñi A, Illarramendi A. Validation of a Kinect-based telerehabilitation system with total hip replacement patients. J Telemed Telecare 2015 Jun 30;22(3):192-197. [doi: 10.1177/1357633x15590019]
39. Anton D, Berges I, Bermúdez J, Goñi A, Illarramendi A. A Telerehabilitation System for the Selection, Evaluation and Remote Management of Therapies. Sensors (Basel) 2018 May 08;18(5):1459 [FREE Full text] [doi: 10.3390/s18051459] [Medline: 29738442]
40. Capecci M, Ceravolo MG, Ferracuti F, Grugnetti M, Iarlori S, Longhi S, et al. An instrumental approach for monitoring physical exercises in a visual markerless scenario: A proof of concept. J Biomech 2018 Mar 01;69:70-80. [doi: 10.1016/j.jbiomech.2018.01.008] [Medline: 29398000]
41. Da Cunha Neto JS, Reboucas Filho P, Da Silva GPF, Da Cunha Olegario NB, Duarte JBF, De Albuquerque VHC. Dynamic Evaluation and Treatment of the Movement Amplitude Using Kinect Sensor. IEEE Access 2018;6:17292-17305. [doi: 10.1109/access.2018.2811720]
42. Carregosa AA, Aguiar Dos Santos LR, Masruha MR, Coêlho MLDS, Machado TC, Souza DCB, et al. Virtual Rehabilitation through Nintendo Wii in Poststroke Patients: Follow-Up. J Stroke Cerebrovasc Dis 2018 Feb;27(2):494-498. [doi: 10.1016/j.jstrokecerebrovasdis.2017.09.029] [Medline: 29100855]
43. Karakoc Z, Colak TK, Yurdalan S. The results of adding virtual rehabilitation to standard rehabilitation program on balance and lower extremity functionality after ACL reconstruction: pilot study. Physiotherapy 2016 Nov;102:e247. [doi: 10.1016/j.physio.2016.10.309]
44. Pua Y, Clark RA, Ong P. Evaluation of the Wii Balance Board for walking aids prediction: proof-of-concept study in total knee arthroplasty. PLoS One 2015 Jan 23;10(1):e0117124 [FREE Full text] [doi: 10.1371/journal.pone.0117124] [Medline: 25615952]
45. Baltaci G, Harput G, Haksever B, Ulusoy B, Ozer H. Comparison between Nintendo Wii Fit and conventional rehabilitation on functional performance outcomes after hamstring anterior cruciate ligament reconstruction: prospective, randomized, controlled, double-blind clinical trial. Knee Surg Sports Traumatol Arthrosc 2013 Apr 29;21(4):880-887. [doi: 10.1007/s00167-012-2034-2] [Medline: 22543515]
46. Iosa M, Picerno P, Paolucci S, Morone G. Wearable inertial sensors for human movement analysis. Expert Rev Med Devices 2016 Jul 17;13(7):641-659. [doi: 10.1080/17434440.2016.1198694] [Medline: 27309490]

47. Strecher V. Internet methods for delivering behavioral and health-related interventions (eHealth). Annu Rev Clin Psychol 2007 Apr;3(1):53-76. [doi: 10.1146/annurev.clinpsy.3.022806.091428] [Medline: 17716048]
48. Jackson P. Introduction to expert systems. United States: Addison-Wesley; 1998.
49. Rini C, Porter LS, Somers TJ, McKee DC, Keefe FJ. Retaining critical therapeutic elements of behavioral interventions translated for delivery via the Internet: recommendations and an example using pain coping skills training. J Med Internet Res 2014 Dec 19;16(12):e245 [FREE Full text] [doi: 10.2196/jmir.3374] [Medline: 25532216]
50. Shukla H, Nair S, Thakker D. Role of telerehabilitation in patients following total knee arthroplasty: Evidence from a systematic literature review and meta-analysis. J Telemed Telecare 2016 Jul 09;23(2):339-346. [doi: 10.1177/1357633x16628996]
51. Fatehi F, Bird D, Gray LC. Pubmed Searching Using Mesh Terms to Identify Randomized Controlled Trials on Telemedicine for Diabetes. J Telemed Telecare 2013 Apr 23;19(3):175-176. [doi: 10.1177/1357633x13479708]
52. Tacconelli E. Systematic reviews: CRD's guidance for undertaking reviews in health care. The Lancet Infectious Diseases 2010 Apr;10(4):226. [doi: 10.1016/s1473-3099(10)70065-7]
53. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
54. Kramer JF, Speechley M, Bourne R, Rorabeck C, Vaz M. Comparison of clinic- and home-based rehabilitation programs after total knee arthroplasty. Clin Orthop Relat Res 2003 May;410(410):225-234. [doi: 10.1097/01.blo.0000063600.67412.11] [Medline: 12771834]
55. Han ASY, Nairn L, Harmer AR, Crosbie J, March L, Parker D, et al. Early rehabilitation after total knee replacement surgery: a multicenter, noninferiority, randomized clinical trial comparing a home exercise program with usual outpatient care. Arthritis Care Res (Hoboken) 2015 Feb 27;67(2):196-202 [FREE Full text] [doi: 10.1002/acr.22457] [Medline: 25220488]
56. Lin F, Chen M, Li P. Influence of structured telephone follow-up on patient compliance with rehabilitation after total knee arthroplasty. PPA 2016 Mar:257. [doi: 10.2147/ppa.s102156]
57. Azma K, RezaSoltani Z, Rezaeimoghaddam F, Dadarkhah A, Mohsenolhosseini S. Efficacy of tele-rehabilitation compared with office-based physical therapy in patients with knee osteoarthritis: A randomized clinical trial. J Telemed Telecare 2017 Aug 03;24(8):560-565. [doi: 10.1177/1357633x17723368]
58. Wong YK, Hui E, Woo J. A community-based exercise programme for older persons with knee pain using telemedicine. J Telemed Telecare 2005 Sep 01;11(6):310-315. [doi: 10.1258/1357633054893346] [Medline: 16168168]
59. Tousignant M, Boissy P, Corriveau H, Moffet H, Cabana F. In-home telerehabilitation for post-knee arthroplasty: a pilot study. Int J Telerehabil 2009 Sep 04;1(1):9-16 [FREE Full text] [doi: 10.5195/ijt.2009.5997] [Medline: 25945158]
60. Moffet H, Tousignant M, Nadeau S, Mérette C, Boissy P, Corriveau H, et al. In-Home Telerehabilitation Compared with Face-to-Face Rehabilitation After Total Knee Arthroplasty. The Journal of Bone and Joint Surgery-American Volume 2015;97(14):1129-1141. [doi: 10.2106/jbjs.n.01066]
61. Boissy P, Tousignant M, Moffet H, Nadeau S, Brière S, Mérette C, et al. Conditions of Use, Reliability, and Quality of Audio/Video-Mediated Communications During In-Home Rehabilitation Teletreatment for Postknee Arthroplasty. Telemed J E Health 2016 Aug;22(8):637-649. [doi: 10.1089/tmj.2015.0157] [Medline: 26958932]
62. Bini S, Mahajan J. Clinical outcomes of remote asynchronous telerehabilitation are equivalent to traditional therapy following total knee arthroplasty: A randomized control study. J Telemed Telecare 2016 Jul 09;23(2):239-247. [doi: 10.1177/1357633x16634518]
63. Doiron-Cadrin P, Kairy D, Vendittoli P, Lowry V, Poitras S, Desmeules F. Feasibility and preliminary effects of a tele-prehabilitation program and an in-person prehablitation program compared to usual care for total hip or knee arthroplasty candidates: a pilot randomized controlled trial. Disabil Rehabil 2020 Apr 13;42(7):989-998. [doi: 10.1080/09638288.2018.1515992] [Medline: 30638076]
64. Eisermann U, Haase I, Kladny B. Computer-aided multimedia training in orthopedic rehabilitation. Am J Phys Med Rehabil 2004 Sep;83(9):670-680. [doi: 10.1097/01.phm.0000137307.44173.5d] [Medline: 15314531]
65. Piqueras M, Marco E, Coll M, Escalada F, Ballester A, Cinca C, et al. Effectiveness of an interactive virtual telerehabilitation system in patients after total knee arthoplasty: a randomized controlled trial. J Rehabil Med 2013 Apr;45(4):392-396 [FREE Full text] [doi: 10.2340/16501977-1119] [Medline: 23474735]
66. Ayoade M, Baillie L. A novel knee rehabilitation system for the home. New YorkNYUnited States: Association for Computing Machinery; 2014 Presented at: Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14; 26 Apr - 1 May; Toronto Ontario Canada p. 2521-2530. [doi: 10.1145/2556288.2557353]
67. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Home-based Rehabilitation With A Novel Digital Biofeedback System versus Conventional In-person Rehabilitation after Total Knee Replacement: a feasibility study. Sci Rep 2018 Jul 26;8(1):11299 [FREE Full text] [doi: 10.1038/s41598-018-29668-0] [Medline: 30050087]
68. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Medium-Term Outcomes of Digital Versus Conventional Home-Based Rehabilitation After Total Knee Arthroplasty: Prospective, Parallel-Group Feasibility Study. JMIR Rehabil Assist Technol 2019 Feb 28;6(1):e13111 [FREE Full text] [doi: 10.2196/13111] [Medline: 30816849]

69. Argent R, Slevin P, Bevilacqua A, Neligan M, Daly A, Caulfield B. Wearable Sensor-Based Exercise Biofeedback for Orthopaedic Rehabilitation: A Mixed Methods User Evaluation of a Prototype System. Sensors (Basel) 2019 Jan 21;19(2):432 [FREE Full text] [doi: 10.3390/s19020432] [Medline: 30669657]
70. Ramkumar PN, Haeberle HS, Ramanathan D, Cantrell WA, Navarro SM, Mont MA, et al. Remote Patient Monitoring Using Mobile Health for Total Knee Arthroplasty: Validation of a Wearable and Machine Learning-Based Surveillance Platform. J Arthroplasty 2019 Oct;34(10):2253-2259. [doi: 10.1016/j.arth.2019.05.021] [Medline: 31128890]
71. Eichler S, Salzwedel A, Rabe S, Mueller S, Mayer F, Wochatz M, et al. The Effectiveness of Telerehabilitation as a Supplement to Rehabilitation in Patients After Total Knee or Hip Replacement: Randomized Controlled Trial. JMIR Rehabil Assist Technol 2019 Nov 07;6(2):e14236 [FREE Full text] [doi: 10.2196/14236] [Medline: 31697239]
72. Prvu Bettger J, Green C, Holmes D, Chokshi A, Mather RC, Hoch BT, et al. Effects of Virtual Exercise Rehabilitation In-Home Therapy Compared with Traditional Care After Total Knee Arthroplasty: VERITAS, a Randomized Controlled Trial. J Bone Joint Surg Am 2020 Jan 15;102(2):101-109. [doi: 10.2106/JBJS.19.00695] [Medline: 31743238]
73. Kuether J, Moore A, Kahan J, Martucci J, Messina T, Perreault R, et al. Telerehabilitation for Total Hip and Knee Arthroplasty Patients: A Pilot Series with High Patient Satisfaction. HSS J 2019 Oct 21;15(3):221-225 [FREE Full text] [doi: 10.1007/s11420-019-09715-w] [Medline: 31624476]
74. Chughtai M, Newman J, Sultan A, Khlopas A, Navarro SM, Bhave A, et al. The Role of Virtual Rehabilitation in Total Knee and Hip Arthroplasty. Surg Technol Int 2018 Jun 01;32:299-305. [Medline: 29566421]
75. Bossen D, Veenhof C, Dekker J, de Bakker D. The usability and preliminary effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis. BMC Med Inform Decis Mak 2013 May 28;13(1):61 [FREE Full text] [doi: 10.1186/1472-6947-13-61] [Medline: 23714120]
76. Bossen D, Buskermolen M, Veenhof C, de Bakker D, Dekker J. Adherence to a web-based physical activity intervention for patients with knee and/or hip osteoarthritis: a mixed method study. J Med Internet Res 2013 Oct 16;15(10):e223 [FREE Full text] [doi: 10.2196/jmir.2742] [Medline: 24132044]
77. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
78. Rini C, Porter LS, Somers TJ, McKee DC, DeVellis RF, Smith M, et al. Automated Internet-based pain coping skills training to manage osteoarthritis pain: a randomized controlled trial. Pain 2015 May;156(5):837-848 [FREE Full text] [doi: 10.1097/j.pain.0000000000000121] [Medline: 25734997]
79. Kim TWB, Gay N, Khemka A, Garino J. Internet-Based Exercise Therapy Using Algorithms for Conservative Treatment of Anterior Knee Pain: A Pragmatic Randomized Controlled Trial. JMIR Rehabil Assist Technol 2016 Dec 14;3(2):e12 [FREE Full text] [doi: 10.2196/rehab.5148] [Medline: 28582256]
80. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an Internet-Delivered Exercise and Pain-Coping Skills Training Intervention for Persons With Chronic Knee Pain. Ann Intern Med 2017 Feb 21;166(7):453. [doi: 10.7326/m16-1714]
81. Lawford BJ, Hinman RS, Kasza J, Nelligan R, Keefe F, Rini C, et al. Moderators of Effects of Internet-Delivered Exercise and Pain Coping Skills Training for People With Knee Osteoarthritis: Exploratory Analysis of the IMPACT Randomized Controlled Trial. J Med Internet Res 2018 May 09;20(5):e10021. [doi: 10.2196/10021] [Medline: 29743149]
82. Bossen D, Kloek C, Snippe HW, Dekker J, de Bakker D, Veenhof C. A Blended Intervention for Patients With Knee and Hip Osteoarthritis in the Physical Therapy Practice: Development and a Pilot Study. JMIR Res Protoc 2016 Feb 24;5(1):e32 [FREE Full text] [doi: 10.2196/resprot.5049] [Medline: 26912378]
83. de Vries HJ, Kloek CJ, de Bakker DH, Dekker J, Bossen D, Veenhof C. Determinants of Adherence to the Online Component of a Blended Intervention for Patients with Hip and/or Knee Osteoarthritis: A Mixed Methods Study Embedded in the e-Exercise Trial. Telemed J E Health 2017 Dec;23(12):1002-1010. [doi: 10.1089/tmj.2016.0264] [Medline: 28525310]
84. Kloek CJJ, van Dongen JM, de Bakker DH, Bossen D, Dekker J, Veenhof C. Cost-effectiveness of a blended physiotherapy intervention compared to usual physiotherapy in patients with hip and/or knee osteoarthritis: a cluster randomized controlled trial. BMC Public Health 2018 Aug 31;18(1):1082 [FREE Full text] [doi: 10.1186/s12889-018-5975-7] [Medline: 30170586]
85. Kloek CJ, Bossen D, de Vries HJ, de Bakker DH, Veenhof C, Dekker J. Physiotherapists' experiences with a blended osteoarthritis intervention: a mixed methods study. Physiother Theory Pract 2020 May 28;36(5):572-579. [doi: 10.1080/09593985.2018.1489926] [Medline: 29952687]
86. Kloek C, Bossen D, Spreeuwenberg P, Dekker J, de Bakker DH, Veenhof C. Effectiveness of a Blended Physical Therapist Intervention in People With Hip Osteoarthritis, Knee Osteoarthritis, or Both: A Cluster-Randomized Controlled Trial. Phys Ther 2018 Jul 01;98(7):560-570 [FREE Full text] [doi: 10.1093/ptj/pzy045] [Medline: 29788253]
87. Ramkumar PN, Haeberle HS, Navarro SM, Sultan AA, Mont MA, Ricchetti ET, et al. Mobile technology and telemedicine for shoulder range of motion: validation of a motion-based machine-learning software development kit. J Shoulder Elbow Surg 2018 Jul;27(7):1198-1204. [doi: 10.1016/j.jse.2018.01.013] [Medline: 29525490]
88. Eichler S, Rabe S, Salzwedel A, Müller S, Stoll J, Tilgner N, ReMove-It study group. Effectiveness of an interactive telerehabilitation system with home-based exercise training in patients after total hip or knee replacement: study protocol

for a multicenter, superiority, no-blinded randomized controlled trial. Trials 2017 Sep 21;18(1):438 [FREE Full text] [doi: 10.1186/s13063-017-2173-3] [Medline: 28934966]
89. Microsoft. Kinect for Windows SDK 2.0. 2014. URL: https://www.microsoft.com/en-us/download/details.aspx?id=44561 [accessed 2020-06-30]
90. Veenhof C, Köke AJA, Dekker J, Oostendorp RA, Bijlsma JWJ, van Tulder MW, et al. Effectiveness of behavioral graded activity in patients with osteoarthritis of the hip and/or knee: A randomized clinical trial. Arthritis Rheum 2006 Dec 15;55(6):925-934 [FREE Full text] [doi: 10.1002/art.22341] [Medline: 17139639]
91. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2016 Dec 27;23(9):780-785. [doi: 10.1177/1357633x16686747]
92. Li L, Gan Y, Zhang L, Wang Y, Zhang F, Qi J. The effect of post-discharge telephone intervention on rehabilitation following total hip replacement surgery. International Journal of Nursing Sciences 2014 Jun;1(2):207-211. [doi: 10.1016/j.ijnss.2014.05.005]
93. Hørdam B, Sabroe S, Pedersen P, Mejdahl S, Søballe K. Nursing intervention by telephone interviews of patients aged over 65 years after total hip replacement improves health status: a randomised clinical trial. Scand J Caring Sci 2010 Mar;24(1):94-100. [doi: 10.1111/j.1471-6712.2009.00691.x] [Medline: 19422632]
94. Russell TG, Buttrum P, Wootton R, Jull GA. Rehabilitation after total knee replacement via low-bandwidth telemedicine: the patient and therapist experience. J Telemed Telecare 2004 Dec 02;10 Suppl 1(1_suppl):85-87. [doi: 10.1258/1357633042614384] [Medline: 15603622]
95. Lemaire E, Boudrias Y, Greene G. Low-bandwidth, Internet-based videoconferencing for physical rehabilitation consultations. J Telemed Telecare 2001;7(2):82-89. [doi: 10.1258/1357633011936200] [Medline: 11331045]
96. Cabana F, Boissy P, Tousignant M, Moffet H, Corriveau H, Dumais R. Interrater agreement between telerehabilitation and face-to-face clinical outcome measurements for total knee arthroplasty. Telemed J E Health 2010 Apr;16(3):293-298. [doi: 10.1089/tmj.2009.0106] [Medline: 20406116]
97. Russell T, Jull G, Wootton R. Can the Internet be used as a medium to evaluate knee angle? Manual Therapy 2003 Nov;8(4):242-246. [doi: 10.1016/s1356-689x(03)00016-x]
98. Russell TG, Wootton R, Jull GA. Physical outcome measurements via the Internet: reliability at two Internet speeds. J Telemed Telecare 2002 Dec 02;8 Suppl 3(6):50-52. [doi: 10.1258/13576330260440853] [Medline: 12537905]
99. Russell TG, Buttrum P, Wootton R, Jull GA. Internet-Based Outpatient Telerehabilitation for Patients Following Total Knee Arthroplasty. The Journal of Bone and Joint Surgery-American Volume 2011;93(2):113-120. [doi: 10.2106/jbjs.i.01375]
100. Pérez-Manchón D, Caramés-Sánchez C, Pfang B. An asynchronous telemedicine program: Three years' experience with African patients treated in Spain. J Telemed Telecare 2016 Jul 03;23(5):558-560. [doi: 10.1177/1357633x16658159]
101. Chen K, Chen P, Liu K, Chan C. Wearable sensor-based rehabilitation exercise assessment for knee osteoarthritis. Sensors (Basel) 2015 Feb 12;15(2):4193-4211 [FREE Full text] [doi: 10.3390/s150204193] [Medline: 25686308]
102. Han S, Xie M, Chien C, Cheng Y, Tsao C. Using MEMS-based inertial sensor with ankle foot orthosis for telerehabilitation and its clinical evaluation in brain injuries and total knee replacement patients. Microsyst Technol 2015 Feb 3;22(3):625-634. [doi: 10.1007/s00542-015-2439-1]
103. Kobsar D, Osis S, Boyd J, Hettinga B, Ferber R. Wearable sensors to predict response to a hip strengthening exercise intervention in patients with knee osteoarthritis. Osteoarthritis and Cartilage 2017 Apr;25:S23-S24. [doi: 10.1016/j.joca.2017.02.052]
104. Naeemabadi M, Dinesen B, Najafi S, Thogersen M, Hansen J. Feasibility of employing AHRS algorithms in the real-time estimation of sensor orientation using low-cost and low sampling rate wearable sensors in IoT application. 2018 Presented at: 2018 IEEE 8th International Conference on Consumer Electronics - Berlin (ICCE-Berlin); 2 - 5 Sep 2018; Berlin, Germany p. 1-6. [doi: 10.1109/icce-berlin.2018.8576239]
105. Naeemabadi M, Dinesen B, Andersen O, Najafi S, Hansen J. Evaluating Accuracy and Usability of Microsoft Kinect Sensors and Wearable Sensor for Tele Knee Rehabilitation after Knee Operation. 2018 Presented at: Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies; 19-21 Jan; Funchal, Portegual p. 128-135. [doi: 10.5220/0006578201280135]
106. Milgram P, Kishino F. A taxonomy of mixed reality visual displays. IEICE TRANSACTIONS on Information and Systems 1994 Dec 25;77:1321-1329.

Abbreviations

9DOF: 9 degrees of freedom
CINAHL: Cumulative Index to Nursing and Allied Health Literature
PEDro: Physiotherapy Evidence Database
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
MeSH: medical subject heading
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 09.11.19; peer-reviewed by D Kairy, C Snoswell; comments to author 09.12.19; revised version received 16.05.20; accepted 14.06.20; published 21.07.20.

Please cite as:
Naeemabadi M, Fazlali H, Najafi S, Dinesen B, Hansen J
Telerehabilitation for Patients With Knee Osteoarthritis: A Focused Review of Technologies and Teleservices
JMIR Biomed Eng 2020;5(1):e16991
URL: http://biomedeng.jmir.org/2020/1/e16991/
doi:10.2196/16991
PMID:

Review

Current Status and Future Challenges of Sleep Monitoring Systems: Systematic Review

Qiang Pan[1*], MSc; Damien Brulin[1*], PhD; Eric Campo[1*], PhD

LAAS-CNRS, University of Toulouse, Toulouse, France

[*]all authors contributed equally

Corresponding Author:
Eric Campo, PhD
LAAS-CNRS
University of Toulouse
7, avenue du Colonel Roche
Toulouse, 31400
France
Phone: 33 561 337 961
Email: eric.campo@univ-tlse2.fr

Abstract

Background: Sleep is essential for human health. Considerable effort has been put into academic and industrial research and in the development of wireless body area networks for sleep monitoring in terms of nonintrusiveness, portability, and autonomy. With the help of rapid advances in smart sensing and communication technologies, various sleep monitoring systems (hereafter, sleep monitoring systems) have been developed with advantages such as being low cost, accessible, discreet, contactless, unmanned, and suitable for long-term monitoring.

Objective: This paper aims to review current research in sleep monitoring to serve as a reference for researchers and to provide insights for future work. Specific selection criteria were chosen to include articles in which sleep monitoring systems or devices are covered.

Methods: This review investigates the use of various common sensors in the hardware implementation of current sleep monitoring systems as well as the types of parameters collected, their position in the body, the possible description of sleep phases, and the advantages and drawbacks. In addition, the data processing algorithms and software used in different studies on sleep monitoring systems and their results are presented. This review was not only limited to the study of laboratory research but also investigated the various popular commercial products available for sleep monitoring, presenting their characteristics, advantages, and disadvantages. In particular, we categorized existing research on sleep monitoring systems based on how the sensor is used, including the number and type of sensors, and the preferred position in the body. In addition to focusing on a specific system, issues concerning sleep monitoring systems such as privacy, economic, and social impact are also included. Finally, we presented an original sleep monitoring system solution developed in our laboratory.

Results: By retrieving a large number of articles and abstracts, we found that hotspot techniques such as big data, machine learning, artificial intelligence, and data mining have not been widely applied to the sleep monitoring research area. Accelerometers are the most commonly used sensor in sleep monitoring systems. Most commercial sleep monitoring products cannot provide performance evaluation based on gold standard polysomnography.

Conclusions: Combining hotspot techniques such as big data, machine learning, artificial intelligence, and data mining with sleep monitoring may be a promising research approach and will attract more researchers in the future. Balancing user acceptance and monitoring performance is the biggest challenge in sleep monitoring system research.

(JMIR Biomed Eng 2020;5(1):e20921) doi:10.2196/20921

KEYWORDS

EEG; ECG; classification; mobile phone

Introduction

Background

Sleep is crucial for human health and quality of life. Poor sleep and sleep disorders have been increasingly prevalent among the world's older population [1,2]. Polysomnography (PSG) has long been the gold standard to quantify sleep time, to differentiate sleep stages, and to assess sleep fragmentation. PSG provides comprehensive physiological information during sleep including electroencephalograms (EEGs), electrocardiograms (ECGs), electromyography (EMG), electrooculography (EOG), oral-nasal airflow, body position, thoracic and abdominal movements, pulse oximetry, and limb movements. The serious impact of sleep on health and well-being is the dominant motivation for sleep monitoring.

- *On personal health*, sleep is a foundation for good health as important as diet and exercise, according to the National Sleep Foundation. Poor sleep can lead to adverse health consequences, including obesity [3], cardiovascular disease [4], and depressive disorders [5]. Sleep is also associated with creativity [6], memory consolidation [7], and cognitive function [8]. Patients affected by sudden cardiac death with sleep disorders such as obstructive sleep apnea (OSA) have been reported to have peak mortality during sleep hours [9].
- *On society*, the incidence of sleep disorders appears to be a concern worldwide. Among the global population, 16.6% of people in Africa and Asia [10], 18% of people in Europe [11], and more than 20% of people in North America [12,13] are affected by nocturnal sleep disorders. Such prevalence has led to a range of societal problems, such as high rates of chronic diseases, road traffic accidents, and workplace accidents. Approximately 13% of work injuries are due to sleep problems [14]. In the United States, the expenditure for the treatment of moderate-to-severe sleep disorders and related disorders amounts to US $165 billion per year, far more than the costs of treating diseases such as heart failure, stroke, hypertension, and asthma, which range from US $20 to US $80 per year [15]. In 5 OECD (Organisation for Economic Co-operation and Development) countries, the economic costs of sleep deprivation are 1.35% (Canada), 1.56% (Germany), 1.88% (United Kingdom), 2.28% (United States), and 2.92% (Japan) of their respective gross domestic product (GDP) [16].

Sleep Stages Scoring Rules

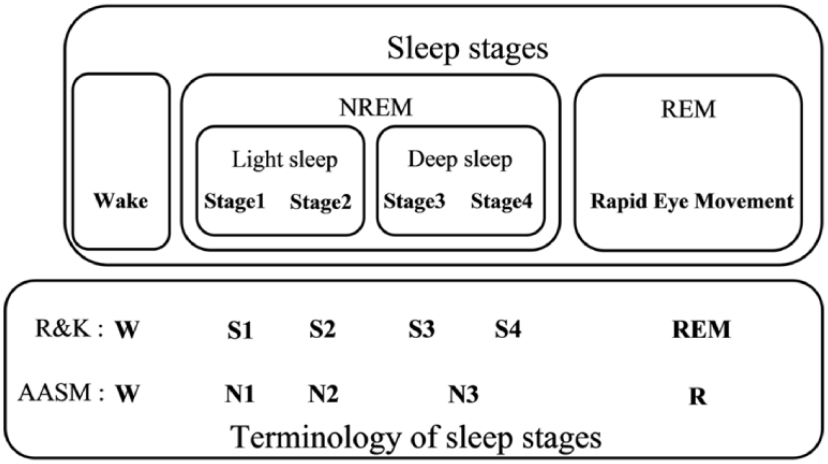
Clinicians can obtain reliable sleep monitoring results, such as sleep stages, by analyzing the PSG recording during the night. For sleep stage guidelines and scoring rules, the R&K rules proposed by Rechtschaffen and Kales in 1968 [17] were used until 2007, when the American Academy of Sleep Medicine (AASM) updated the scoring manual commonly referred to as the AASM scoring manual [18]. The R&K rules and the AASM rules differ in the terminology used. The R&K rules divide sleep into 6 distinct stages: W (wake); non–rapid eye movement (non-REM [NREM]) stages S1, S2, S3, and S4; and REM sleep stage. The AASM rules recognize 5 sleep stages: W (wake) stage N1 (formerly stage 1 sleep), stage N2 (formerly stage 2 sleep), stage N3 (formerly stages 3 and 4 sleep), and stage R sleep (formerly stage REM sleep), as illustrated in Figure 1.

For the same sleep, the scoring results obtained from R&K rules and the AASM rules will be slightly different. One study [19] adopted both rules to score PSG sleep recordings of healthy subjects and patients (38 women and 34 men) aged 21 to 86 years. The results showed that sleep latency, REM latency, total sleep time, and sleep efficiency were not affected by the classification standard. In contrast, the time (in minutes and as a percentage of total sleep time) spent in stage 1 (S1/N1), stage 2 (S2/N2), and slow wave sleep (S3+S4/N3) differed significantly between the R&K and AASM classifications. Although light and deep sleep increased (S1 vs N1 [+10.6 min, (+2.8%)]: $P<.01$; S3+S4 vs N3 [+9.1 min (+2.4%)]: $P<.01$), stage 2 sleep decreased significantly according to the AASM rules (S2 vs N2 [−20.5 min, (–4.9%)]: $P<.01$).

The differences between the results of the 2 sleep standards can be attributed to the different rules used [20].

The reader is reminded that sleep stages should not be considered as distinct entities but rather as a gradual transition of a waveform. Sleep usually follows a predictable pattern, moving cyclically between the light sleep stage, the deep sleep stage, and REM. Each sleep cycle typically lasts about 90 min and is repeated 4 to 6 times during the night. In each sleep cycle, people usually experience a transition from light to deep sleep first and then switch to REM. However, some stages can be skipped during sleep. For example, one can switch to REM or return directly to deep sleep from REM sleep [21]. Sleep quality is analyzed using standard parameters such as sleep efficiency, total sleep time, sleep latency, sleep stages 1 and 2, slow-wave sleep (sleep stages 3 and 4), rapid eye movement sleep, wake time after sleep onset, and nocturnal wake time [22].

Figure 1. Terminology used by R&K and AASM for sleep stages scoring.



Sleep disorders, the most recently released third edition of the *International Classification of Sleep Disorders* (*ICSD-3*) has identified 7 major categories of sleep disorders that include insomnia, sleep-related breathing disorders, central hypersomnolence disorders, circadian rhythm sleep-wakefulness disorders, sleep-related movement disorders, parasomnia, and other sleep disorders [23]. Most sleep disorders can be monitored by sleep monitoring systems, and some of them are detailed below:

- *Insomnia* refers to impairment in the quality and quantity of sleep. According to Ohayon [24], 10% to 30% of the adult population is affected by insomnia. The ICSD-3 criteria for this diagnosis include (1) a report of sleep initiation or maintenance problems, (2) adequate opportunity and circumstances for sleep, and (3) consequences during the day. The ICSD-3 duration criterion for chronic insomnia disorder is 3 months, and a frequency criterion (at least 3 times per week) was added.
- *Sleep apnea* is characterized by pauses in breathing or instances of shallow breathing during sleep [25]. Due to sleep apnea, the patient wakes up regularly throughout the night to retrieve breathing. Frequent awakening results in very poor quality of sleep and excessive daytime fatigue. Usually, sleep apnea may be accompanied by loud snoring, which can be easily monitored by a microphone (many researchers have studied Snoring signal processing–based methods to achieve a supplementary diagnosis way of sleep apnea) [26-28].
- *Restless legs syndrome (RLS)* is based on an urge to move the legs, sometimes accompanied by an uncomfortable sensation that (1) occurs primarily with rest or inactivity, (2) is partially or totally relieved by the movement, for as long as the movement occurs, and (3) occurs primarily in the evening or night [29]. Up to 30% of cases are caused by iron deficiency. These abnormal leg movements can be easily monitored with an accelerometer [30-32].
- *Periodic limb movement disorder (PLMD)* is characterized by abnormal limb movements and is responsible for deterioration in sleep quality [33]. For young people, it will be considered as pathologic when the index of periodic limb movement during sleep (PLMS; number of PLMS per hour) is greater than 5. For older people, an index of PLMS greater than 15 is usually adopted as the pathological threshold. This disorder can be detected by using EMG [34] or actigraphy [30].
- *Disorders of arousal from NREM (DAN)* include confusion arousal, sleepwalking, sleep terrors, and sleep-related eating disorders [23,35]. The general criteria for disorders of arousal include (1) recurrent episodes of incomplete awakening, (2) absent or inappropriate responsiveness, (3) limited or no cognition or dream report, and (4) partial or complete amnesia for the episode. Detection of repeated wakes during the NREM stage can be a sign of DAN. This disorder can be detected using EEG.
- *REM sleep behavior disorder (RBD)* is characterized by the intermittent loss of REM sleep atonia and the appearance of elaborate motor activity associated with the situation in dreams, such as repeated episodes of behavior or vocalization arising from REM [36]. When specific movements and sounds are detected during the REM stage, a suspicion of RBD should be considered. This disorder could be detected by using a microphone, actigraphy, and EEG.

Therefore, there appears to be a growing interest in researching new sleep monitoring system solutions to provide rapid, reliable, and long-term monitoring results to users and clinicians. Innovative home-used sleep monitoring systems offer users access to sleep phases and quality by themselves and can be a reference for the diagnosis of sleep disorders by clinicians.

For sleep monitoring, a sleep monitoring system can include a wide range of wearable or noncontact devices, including sensors, actuators, smart fabrics, power supplies, wireless communication networks, processing units, multimedia devices, user interfaces, software, and algorithms for data capture, processing, and decision support. These systems are able to measure vital signs, such as body and limb movements, body and skin temperature, heart rate, ECGs, EEGs, EMG, and respiration rate. The measurements are transmitted via a sensor network either to a central connection node, such as a personal digital assistant, or directly to a medical center for storage, data processing, and decision making.

To discuss this potential, this paper aimed to review the current state of research and development in the field of sleep

monitoring system, highlighting the main features of the most promising projects being developed and future challenges.

Methods

Overview

With the growing appearance of sleep problems, the sleep monitoring system study has been one of the hotspots in the field of smart human monitoring. As a result, advances in sleep monitoring system development technology are continuously accelerating. Simple, lightweight, and small-size sensor systems are being adopted to acquire sleep-related physiological information. These systems are designed to be adaptable to the gold standard PSG, including sleep stages, sleep or wake, sleep apnea, sleep positions, and so on. Furthermore, the advantages of such a system over the traditional PSG method are that it is affordable, requires little or no technician intervention, is installed in the home, and can be used over the long term. The systems have developed rapidly with the progress made in the miniaturization of sensors, the reduction of energy consumption, and various communication possibilities (Bluetooth, Wi-Fi, Sigfox, LoRa, and NB-IoT). These technologies enable current sleep monitoring systems to be less intrusive and effective, with remote and continuous monitoring. In this review, specific selection criteria are chosen as reference articles on sleep monitoring systems.

Inclusion Criteria for Sleep Monitoring System Search

Most sleep monitoring system research projects have focused on smart, portable, and nonintrusive devices that encompass wireless communication, moving the monitoring site from the hospital to the home, in patch or noncontact form. Systems that have the following features are included:

- Wearable, portable, nonintrusive, wireless, and noncontact.
- Patch, body sensor system, and sensor network.
- Band, watch, textile, bedsheet, and belt.
- Mobile, stationary, ambulatory, at home, and remote.

Automatic collection and transmission of acquired data and processing results can help physicians and caregivers easily follow sleep conditions over time. In addition, it can make it easier to find trends in the data, providing insight into individualized sleep patterns.

Search Methods and Strategies

This literature review focuses on the presentation of the hardware and software adopted in the current sleep monitoring system. We have included journal publications, conference publications, and information on related websites. The keywords for material collection are shown in Textbox 1. We conducted a keyword search in Web of Science, PubMed, and PubMed Central.

Textbox 1. Keywords used for the literature search.

- Sleep
- Sleep quality
- Sleep monitor
- Sleep monitor system
- Sleep monitor and sensor
- Sleep monitor and smart patch
- Sleep monitor and commercial products
- Sleep monitor at home
- Polysomnography
- Electroencephalography
- Rapid eye movement or light or deep sleep or wake
- Long-term sleep monitor
- Sleep phase classification
- Sleep stage classification
- Noncontact sleep monitor
- Nonintrusive sleep monitor
- Noninvasive sleep monitor
- Sleep big data
- Sleep data mining
- Sleep deep learning
- Sleep machine learning
- Sleep artificial intelligence

Results

Owing to a large number of articles and abstracts retrieved, it was decided to include only articles published for the period 2013 to 2018 in Web of Science, PubMed, and PubMed Central. By counting the number of hits in the bibliographic database for each keyword, we can find search hotspots in this field and aspects that are still rarely covered. In this first search, we tried to find articles and abstracts, and websites with the keywords listed in Textbox 1. Keywords are used alone or combined using and, or operators. The article should report a clear description of the systems, the recipients or users requiring these systems, and issues related to sleep monitoring systems, including the parameters measured, wireless sensor network (WSN), user needs, and user acceptance. As this review does not constitute an exhaustive presentation of the scientific literature in the field of sleep monitoring systems, only a few representative sleep monitoring system research and development projects or products from academia or industry are presented.

The number of hits in the sleep monitoring system research field between 2013 and 2018 is shown in Table 1.

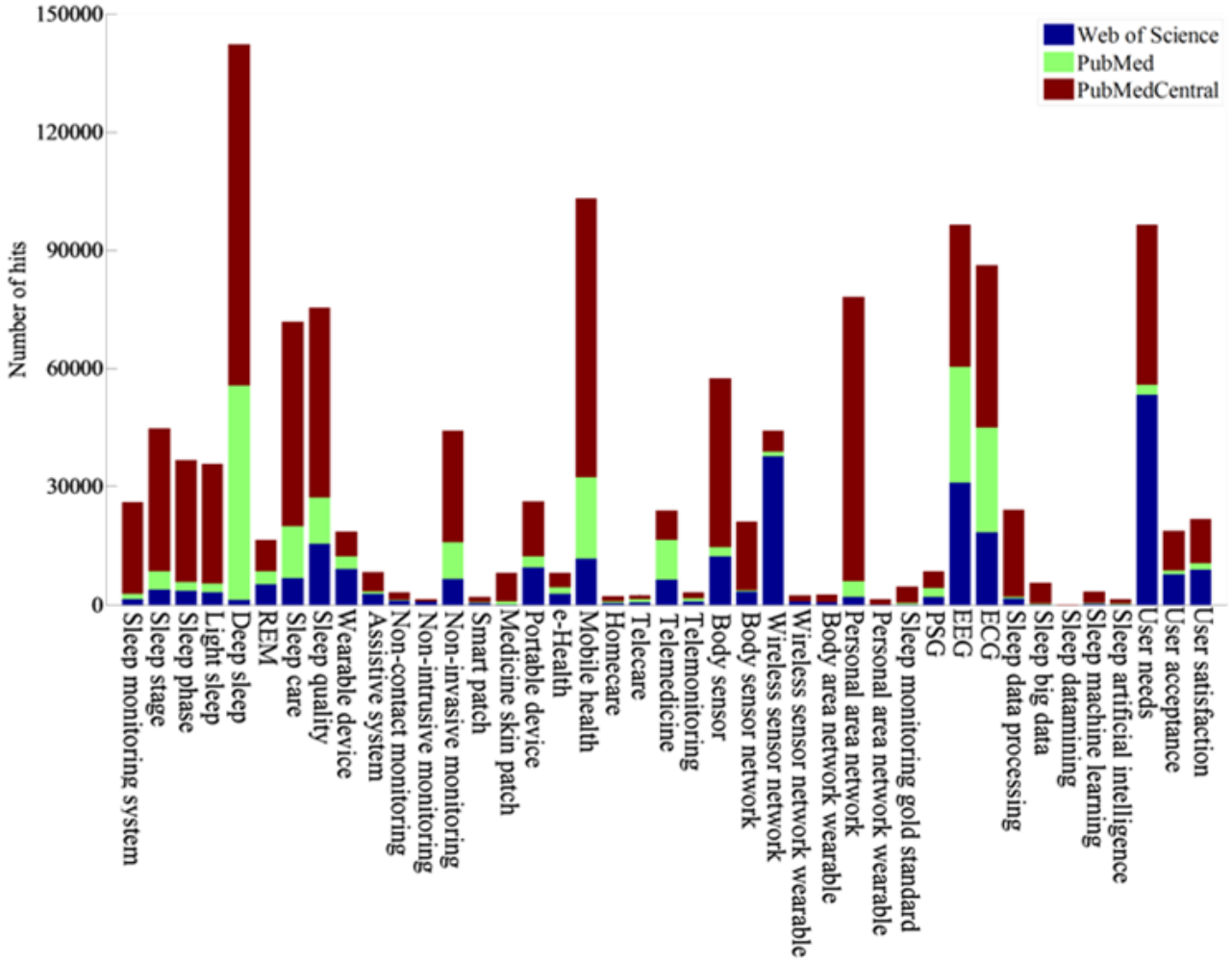
Table 1 shows that the number of hits for sleep big data, sleep machine learning, sleep artificial intelligence, and sleep data mining is lower than the other keywords. It would appear that techniques such as big data, machine learning, artificial intelligence, and data mining have not been widely applied to the sleep monitoring research area, although they are now focused in other research areas. Therefore, the combination of these hotspot techniques and sleep monitoring may be a promising research direction and will attract more researchers in the future. To facilitate reading of the data in Table 1, a histogram is drawn, as shown in Figure 2.

Table 1. Number of occurrences in the field of sleep monitoring system research over 5 years.

Keywords	Web of Science, n	PubMed, n	PubMed Central, n
Sleep monitoring system	1416	1345	23,212
Sleep stage	3955	4561	36,146
Sleep phase	3542	2224	31,022
Light sleep	3174	2204	30,367
Deep sleep	1172	54,438	86,653
REM[a]	5224	3220	8044
Sleep care	6760	13,154	51,882
Sleep quality	15,517	11,616	48,324
Wearable device	9001	3379	6232
Assistive system	2735	663	4912
Noncontact monitoring	1017	205	1929
Nonintrusive monitoring	854	73	537
Noninvasive monitoring	6507	9355	28,298
Smart patch	561	56	1462
Medicine skin patch	82	834	7232
Portable device	9501	2760	13,910
eHealth	2808	1633	3613
mHealth[b]	11,712	20,583	70,836
Homecare	508	404	1226
Telecare	613	761	934
Telemedicine	6378	10,211	7343
Telemonitoring	921	617	1575
Body sensor	12,356	2220	42,906
Body sensor network	3368	271	17,487
Wireless sensor network	37,771	1083	5239
Wireless sensor network wearable	790	56	1528
Body area network wearable	592	30	1982
Personal area network	1975	4013	72,117
Personal area network wearable	67	5	1389
Sleep monitoring gold standard	131	277	4294
PSG[c]	2036	2240	4219
EEG[d]	31,027	29,396	36,069
ECG[e]	18,358	26,588	41,139
Sleep, data processing	1534	437	22,206
Sleep, big data	131	69	5377
Sleep data mining	0	0	4
Sleep machine learning	277	187	2807
Sleep artificial intelligence	27	202	1194
User needs	53,375	2362	40,766
User acceptance	7711	902	10,274
User satisfaction	8892	1618	11,250

[a]REM: rapid eye movement.
[b]mHealth: mobile health.
[c]PSG: polysomnography.
[d]EEG: electroencephalography.
[e]ECG: electrocardiography.

Figure 2. Number of results in the field of research on sleep monitoring systems over a 5-year period.



Discussion

Through the observation and analysis of search results, we found that current and rapidly developing frontier hotspot techniques, such as large data sets, data mining, artificial intelligence, and machine learning, are not yet widely used in the field of sleep monitoring. However, due to the powerful performance and wealth of applications of these techniques, their application in the field of sleep monitoring is expected to be the future development trend and certainly presents great research value and market opportunities.

Big data can be defined by 3 key concepts: volume, velocity, and variety. *Volume* refers to the amount of data generated and stored. In general, the larger the amount of data, the greater the statistical power for descriptive and predictive analysis. Applied to sleep monitoring systems, it could better describe sleep behavior and predict sleep-related disorders and health status. *Velocity* refers to the speed of data generation and processing. Big data are often available in real time. This makes it easier for people to get their sleep monitoring results in a timely manner, while helping subjects or medical staff to respond quickly to abnormalities and emergencies discovered during sleep monitoring. Finally, the term *variety* refers to different sources, types, and formats of data. Nowadays, more data types are being collected via sleep monitoring systems, including text, audio, image, and video data. Big data allow missing data to be completed through data fusion. This enables comprehensive sleep information to be obtained efficiently. In addition, big data can provide targeted information through the comprehensive and detailed collection of various relevant information, such as age, gender, BMI, place of residence, occupation, and so on. For example, certain age groups, a certain gender, people living in a certain location, people working in a certain profession, and people of a certain body type have a higher rate of poor sleep quality. On the basis of this, it is possible to organize more medical resources in certain areas, and at the same time, more attention in terms of sleep health

can be given to some people who have a higher rate of poor sleep quality. This will improve the efficiency of the medical system.

Artificial intelligence refers to technology that presents human intelligence through computer programs. Machine learning uses data or previous experience to automatically improve the performance of specific algorithms. Data mining is a computational process that uses artificial intelligence, machine learning, statistics, and databases to discover patterns in relatively large data sets. In general, machine learning is considered a subset of artificial intelligence and consists mainly of data mining. As a tedious and repetitive task, sleep monitoring is well suited to the adoption of these promising and powerful approaches. When applying these approaches, it is more convenient and timely to obtain a large amount and variety of sleep-related data through the continued development of big data technology. This allows these approaches to be used to train more powerful models and to progressively extract higher-level features from the raw data to create a smarter, more efficient, and more convenient sleep monitoring systems.

Current Issues With Sleep Monitoring Systems

User Needs, Perception, and Acceptance

A good sleep monitoring system should take into account user needs, perception, and acceptance before it is designed. User needs for sleep monitoring systems are diverse. These could include obtaining accurate and complete information about sleep. These needs could be met by professional medical instruments such as PSG and EEG. In addition, the user needs could also obtain auxiliary reference information, that is, only a small amount of key information such as sleep duration, number of awakenings, proportion of different sleep stages, and even only a summary of the sleep score. These needs are typically met by various apps in consumer electronics and smartphones with sleep monitoring functions. Compared with professional medical devices, this type of commercial product takes better account of the user's perception and is usually noninvasive, nonintrusive, or even contact free. The user's perception of the sleep monitoring system is closely related to the number of electrodes or sensors attached to the body, the position, and the method of attachment on the body.

For the number of electrodes or sensors, the fewer the number, the better the user's perception. For the position of the attachment, it is preferable to attach it to the distal limbs such as the wrist, fingers, ankle, instep, and toes rather than to the main body, face, and head. The method of attachment to the body may be by using adhesive tape or a belt, such as a chest belt. In general, adhesive bonding may give a better user perception than a belt because there is less contact area with the body and less restraint on the body. User acceptance of sleep monitoring systems depends on the satisfaction of the user's needs and perception. Usually the satisfaction of user needs and the satisfaction of user perception are contradictory. To meet user needs as much as possible, more complete and accurate human physiological information needs to be collected, which often means that more sensors need to be attached to more body positions, often worsening user perception. Therefore, the design of a good sleep monitoring system has to find a compromise between users' needs and their perception to achieve a good user perception, which is usually related to ease of use while trying to meet user needs as much as possible.

Effectiveness

Although PSG is the gold standard for sleep monitoring, it is expensive, highly invasive, and complicated to perform. PSG monitoring is not easily accessible, especially in developing countries [37]. Owing to the many limitations of PSG, most people are only subject to PSG monitoring for one night. However, monitoring at night is not sufficient to determine the actual sleep status. To improve effectiveness and obtain appropriate follow-up, long-term, at-home monitoring is necessary.

Guettari et al [38] proposed a system based on one thermal sensor is used over a long period of time to supervise changes in sleep quality and can be used at home and consulted remotely by sleep medicine experts. Changes in sleep quality derived from long-term monitoring are very useful for assessing sleep health. Using existing equipment daily for sleep monitoring is found to be an effective approach, such as sleep monitoring using our smartphone router. Liu et al [39] proposed monitoring vital signs of breathing and heart rate during sleep using a single Wi-Fi access point (such as a router) and a single Wi-Fi device (PC or smartphone) without any wearable or dedicated devices. Thus, the system has the potential to be widely deployed and to perform continuous long-term monitoring. Smartphone apps are considered a good choice for large-scale, low-cost, and long-term sleep monitoring, which will improve effectiveness and accuracy [21,40]. Sleep Hunter [41] is a mobile service that uses smartphones' built-in sensors. It is implemented on the Android platform and can detect the transitions between sleep stages for monitoring sleep quality and the intelligent wake-up call, which wakes users in light sleep. The ability to perform long-term monitoring is important for the effectiveness of a sleep monitoring system. Long-term monitoring is essential for reliable results and early detection of abnormal sleep changes. To do this, sleep monitoring systems should be as inexpensive, easy to use, and easily accessible as possible.

Interoperability

Sleep monitoring devices are useful in health care. The value of these devices will increase if sleep monitoring system software apps can seamlessly collect medical data and upload the data to a database. The ISO/IEEE11073 (X73) family of interoperability standards was originally designed for point-of-care clinical environments. The latest branch of X73, X73 for personal health devices (X73PHD), enables the development of interoperable personal health ecosystems and brings benefits to both technology producers (design cost reduction, experience sharing, and marketing facilities) and users (plug and play, accessibility, ease of integration, and prices) [42]. OpenICE is an open-source software project of the Medical Device Plug-and-Play Interoperability Program at Massachusetts General Hospital, leveraging much of the program's research performed since 2004 to support 4 distinct user groups: use case demonstrations, clinical adoption, regulatory science, and commercial adoption [43]. Data sharing and interoperability are positive for users, researchers,

physicians, and businesses. With the development and popularization of big data technology, improving interoperability is a hotspot in sleep monitoring system research and will be the trend for future development.

Hardware and Software Considerations

The main considerations for sleep monitoring system hardware focus on 3 aspects: cost, comfort, and convenience, which could be the determining factors for the acceptance of implementation. In terms of cost, the equipment should be affordable for most people. In addition, for devices that require frequent maintenance, such as the need for frequent replacement of specific components and the consumption of specific reagents or materials, their costs should be considered. For devices designed to use disposable batteries instead of rechargeable batteries, the energy consumption should be taken into account, as frequent battery replacement will significantly increase the cost of use. In terms of comfort, contactless systems have the greatest advantage, but for contact systems, the emphasis is on wireless, miniaturization, and weight reduction. The comfort aspect includes ease of implementation and maintenance convenience. In terms of the convenience of implementation, daily implementation does not have to be complex and time consuming. The main objective is to allow users to carry out the application themselves without the intervention of professional technicians. Generally, in terms of maintenance convenience, the longer the maintenance interval, the easier the operations are.

The main considerations for sleep monitoring system software are two-fold: effectiveness and efficiency. First, the sleep monitoring software must be able to effectively process the data collected to obtain the most accurate monitoring results. In terms of efficiency, this includes temporal efficiency and energy efficiency. It is very important for real-time sleep monitoring system processing to consider temporal efficiency. The cost of the execution time must be short enough to meet the real-time requirement.

For non–real-time sleep monitoring systems that process data after the end of monitoring, time efficiency is also of great importance. After sleep, users tend to be concerned about the results obtained. Waiting time will have an impact on the user experience, so the shorter the processing time, the better. Energy efficiency depends on 2 aspects: optimization of the algorithm and sleep or wake programming for the hardware. If the algorithm can be optimized well, it will significantly reduce the energy consumption for the execution of the algorithm. Furthermore, with reasonable sleep or wake programming of the hardware, unnecessary energy consumption can be avoided.

Medical, Wellness, and Quality-of-Life Benefits

Sleep quality is a crucial factor for human health and quality of life. There is growing recognition of the harmful effects of poor sleep quality and sleep disorders. Patients with sleep disorders are prone to chronic diseases such as obesity, diabetes, and hypertension. The use of sleep monitoring systems could reduce the incidence of sleep-related illnesses or illnesses could be predicted by sleep through long-term monitoring and trend analysis. McHill et al [44] demonstrated the relationship between obesity and sleep time. Lee et al [45] examined the impact of sleep quantity and sleep quality on blood glucose control in type 2 diabetes. Fuchs et al [46] showed that OSA is a clear risk factor for resistant hypertension. The application of sleep monitoring system can overcome infrequent clinical visits that may not detect transient events that predict dangerous future events. Early diagnosis through long-term trend analysis could prevent the potential severity of a disease. These analyses could provide an instant diagnosis of acute events, issue alerts to health care professionals, and reduce the time of intervention through telediagnosis and teletherapy. Some typical sleep disorders, such as sleep apnea, restless leg syndrome, and periodic limb movement disorders can be detected in time through sleep monitoring. Unfortunately, people with sleep disorders such as OSA tend to go undiagnosed [47] because they are usually not even aware that sleep apnea events have occurred. This lack of awareness of symptoms during sleep is a serious health problem for modern life [48]. The early signs of these disorders could be monitored and treated with mild medications [49].

Cost, Psychological, and Socioeconomic Barriers

Wireless patches, wristbands, chest belts, headbands, or other wearable devices that connect a sleeper to formal or informal caregivers, a data center or call center, who can then notify medical services in the event of abnormal sleep, are affordable and reliable. This technology has been available for more than 15 years, but despite its affordability [50], its adoption is minimal in almost all countries. Wearing permanent health care mobile devices and systems has psychological effects on patients. Significant barriers limit the widespread use of these systems due to the lack of studies on testing smart wearable systems by end users who provide feedback and preferences [51]. The high cost of current sleep monitoring system services limits their expansion. Wireless networks are another barrier to sleep monitoring system deployment. Until the end of 2019, the global internet penetration rate was only 58.7% [52]. Consequently, access to services via internet is not always available. People affected by sleep disorders may have difficulties in finding adequate sleep monitoring devices and services to support them in monitoring the quality of their sleep. Economic and social issues also need to be addressed to ensure that the market for sleep monitoring systems is opened up. A sound analysis of the costs and benefits of sleep monitoring systems has not been conducted. Some studies focus solely on system technology and performance [40,53]. Sociotechnical design science needs to be taken into consideration to ensure that sleep health care meets the needs of society. Ultimately, Coiera [54] argued that it is the beliefs and values of our culture that shape what we will create and what we will dream about. A total of 4 rules govern the design of health services: (1) technical systems have strong social consequences, (2) social systems have technical consequences, (3) we do not design technology; designing sociotechnical systems does not just mean designing technology, and (4) the design of sociotechnical systems must consider the way in which people and technologies interact.

Privacy, Ethics, and Legal Barriers

With the continuous development of sleep monitoring technology, the collection of user information via sleep monitoring system has become increasingly detailed and diverse. At the same time, it has gradually evolved from traditional overnight monitoring to long-term monitoring. This series of developments has improved the accuracy and reliability of sleep monitoring but has significantly increased the risk of leakage of user privacy information. To protect the privacy of users, the traditional method is to provide informed consent before receiving sleep monitoring, and data collection can only take place once the user has signed the informed consent. Consent is normally used to authorize a single study, and there are no specific regulations for data sharing in the research community. Given the high value and gradually popular trend of big data apps, privacy issues related to data sharing need to be addressed urgently by legislators.

Impact of Sleep Monitoring Systems on Society

Sleep disorders affect a significant part of the population [10-12]. The socioeconomic consequences can be dramatic. It includes drowsiness while driving, drowsiness in the workplace, and cardiovascular diseases [55]. Surantha et al [56] argued that sleep quality monitoring is one of the solutions to maintaining sleep quality and preventing chronic diseases, mental problems, or accidents caused by sleep disorders.

On the basis of these considerations and issues, many types of sleep monitoring systems have been developed. The features of these are detailed in the following section.

Sleep Monitoring System Features

Conventional Sleep Monitoring Systems

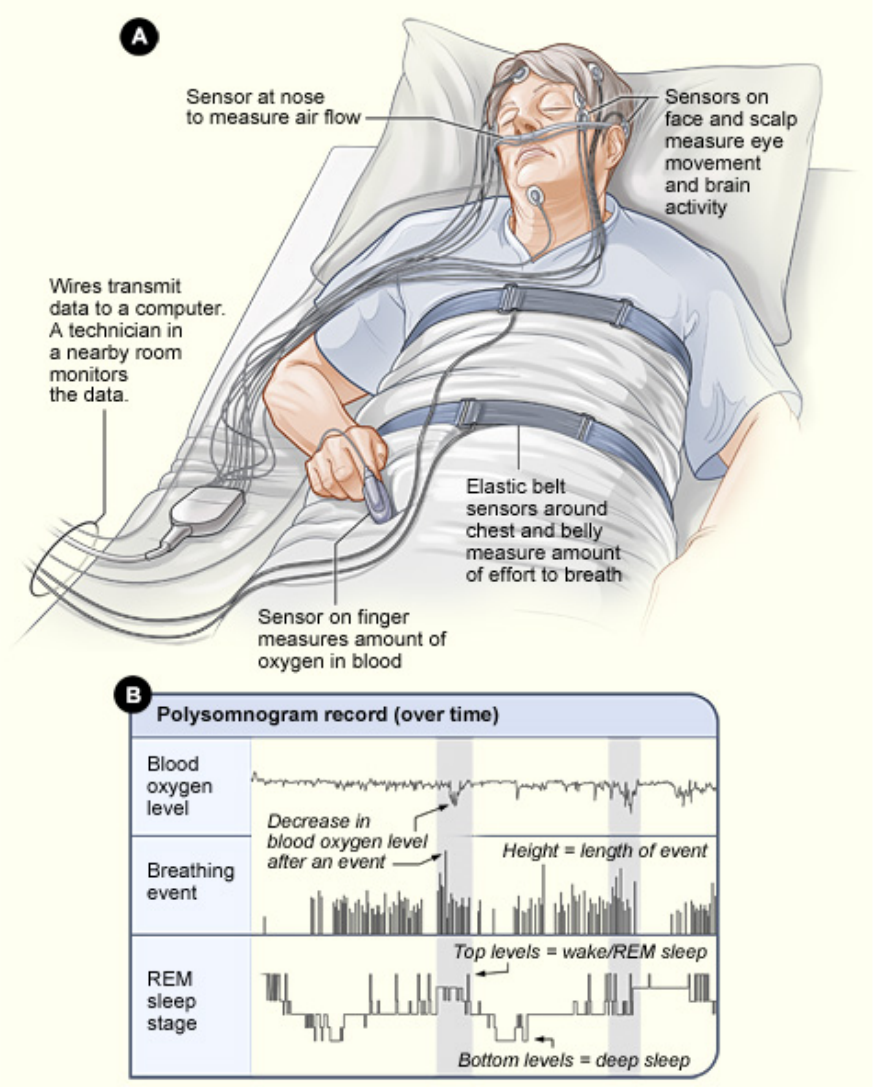
Polysomnography

Polysomnography (PSG) is the gold standard in sleep assessment introduced in the 1960s as a tool for assessing sleep disorders. The subject equipped with a PSG is illustrated in Figure 3. A PSG records a minimum of 12 channels requiring a minimum of 22 wires attached to the patient. These channels vary in each laboratory and can be adapted to meet the physician's requirements. There are a minimum of 3 channels for the EEG, 1 or 2 measure airflow, 1 or 2 are for chin muscle tone, 1 or more for leg movements, 2 for eye movements (EOG), 1 or 2 for heart rate and rhythm, 1 for oxygen saturation, and 1 for each waist belt, which measures movements of the chest wall and upper abdominal wall. Belt motion is usually measured using piezoelectric sensors or respiratory inductance plethysmography. Breathing amplitude is often measured with the temperature changes that occur with breathing, as measured by a thermistor or thermocouple placed in the path of the airflow (nose and mouth). Body movement was measured by using EMG. Oximetry is adopted to measure oxygen saturation levels in the blood by passing infrared light through the finger and measuring absorption patterns (made by the oxygen-carrying pigment, hemoglobin, in the blood). The body position sensor is used to distinguish between lying, standing, and lateral positions during sleep.

Although PSG provides the most accurate and objective measurement of sleep, specialized equipment, an elaborate facility, and dedicated and experienced PSG technologists are required to perform and analyze recordings, which are costly and labor intensive. This technique is not practical for large-scale and long-term sleep monitoring [57].

Figure 3 shows the standard configuration of a polysomnogram. In Figure 3, the patient lies in a bed with sensors attached to the body. In Figure 3, the polysomnogram recording shows the blood oxygen level, the respiratory event, and the REM sleep stage over time.

Figure 3. The PSG sleep monitoring - National Heart Lung and Blood Institute (NIH), November 2013.



Electroencephalography

EEG is an electrophysiological monitoring method that records electrical activity of the brain using electrodes placed along the scalp to measure the voltage fluctuations resulting from ionic current in the neurons of the brain [58]. The EEG signal is the most important signal in the classification of sleep stages [59].

Different sleep stages are characterized by different brain activities that can be detected by EEG recordings. The EEG patterns of the different sleep stages are shown in Figure 4. Stage 1 is the transition stage between wakefulness and sleep. It usually lasts between 1 and 5 min. This stage consists of a low voltage EEG trace with well-defined alpha (Figure 5) and theta (Figure 5) activity, occasional vertex peaks, and slow eye movements. This stage, on average, represents 4% to 5% of total sleep and is free of sleep spindles (Figure 5) and K-complexes (Figure 5). Stage 2 is the *baseline* of sleep and is characterized by the occurrence of sleep spindles and K-complexes and a relatively low voltage, mixed frequency EEG background. In addition, high voltage delta waves may account for up to 20% of stage 2 epochs. Stage 3 is a period in which at least 20% and no more than 50% of sleep consists of EEG signals with a frequency less than or equal to 2 Hz and an amplitude greater than 75 μV (delta waves; Figure 5). Stage 4 is quite similar to stage 3, except that delta waves cover 50% or more of the recording. Stage 4 typically represents 12% to 15% of total sleep time.

Stages 3 and 4 together are also called *deep sleep* or slow wave sleep (SWS), and it is the most restorative part of sleep. REM is the sleep stage in which dreams occur and makes up 20% to 25% of a normal night's sleep. It is well known that the incidence of rapid eye movements under closed eyelids, motor atonia, and low voltage EEG patterns. During REM sleep, the brain activity is reversed from stage 4 to a pattern similar to that in stage 1. The characteristics of each sleep stage are summarized in Table 2. Although the EEG is accurate for determining sleep stages, the complexity and intrusiveness of the user make it difficult to achieve large-scale, long-term, and home sleep monitoring.

Figure 4. Typical EEG pattern for different stages of sleep [60].

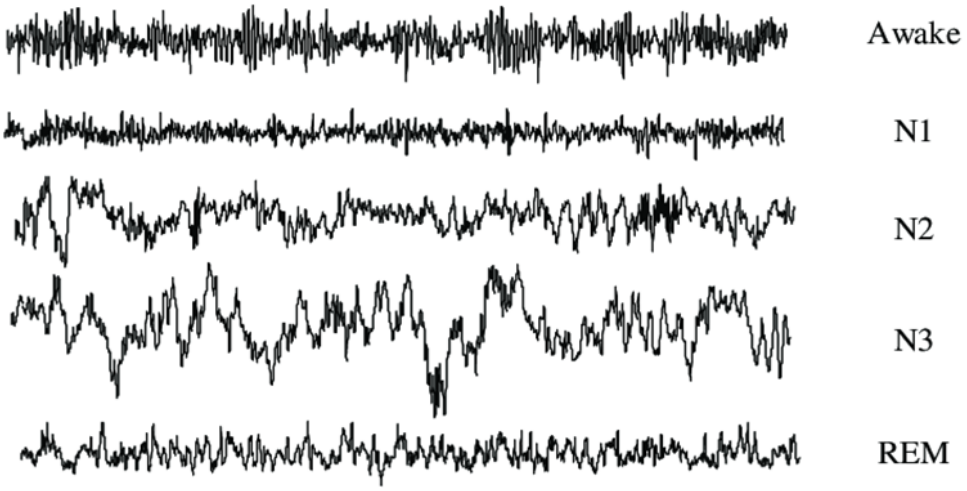


Figure 5. Typical EEG wave types.

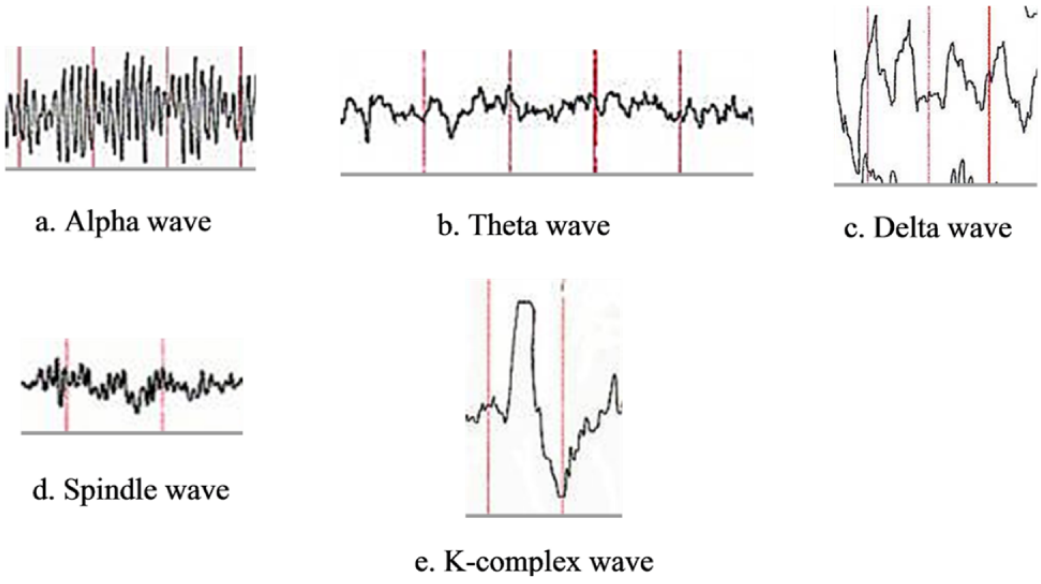


Table 2. Characteristics of each stage of sleep.

Sleep stages	Proportion of sleep, %	EEG[a] frequency (Hz)	EEG amplitude (mv)	EEG percentage (one screen)
Awake	<5	15-50	<50	α>50%
N1	2-5	4-8	50-100	Theta wave >50% or alpha wave <50%
N2	45-55	4-15	50-150	Delta wave <20%; K-complex >1.7%
N3	3-8	2-4	100-150	Delta wave 20% to 50%
N4	10-15	0.5-2	100-200	Delta wave >50%
REM[b]	20-25	15-30	<50	EEG with mixed wave

[a]EEG: electroencephalography.

[b]REM: rapid eye movement.

Electrocardiography

Electrocardiography (ECG) is the process of recording the electrical activity of the heart over a period of time using electrodes placed on the skin. These electrodes detect tiny electrical changes on the skin that result from the electrophysiological pattern of depolarization and repolarization of the heart muscle with each heartbeat. In a conventional 12-lead ECG, 10 electrodes are placed on the patient's limbs and on the surface of the chest. The correlation behavior in the heartbeat rate significantly differed for light sleep, deep sleep, and REM sleep. During deep sleep, the heartbeat rate is reduced, whereas a relative increase is observed in REM sleep [61]. Furthermore, spontaneous movements during sleep are preceded

by an increase in heart rate [62]. These observations indicate a functional link between cardiac activities and sleep stages. As with PSG and EEG, the complexity, high equipment, and expertise requirements of the standard ECG are barriers to its use as a large-scale, long-term, home sleep monitoring method.

Wireless Body Area Network

Introduction

The wireless body area network (WBAN) is a wireless sensor network that aims to monitor the vital signs and physiological information of the user by deploying sensors on or next to the human body. Parameters acquired from the WBAN can include brain waves, heart rate, body movements, body temperature, blood oxygen saturation level, sound, and so on, or environmental conditions such as temperature, brightness, noise level, and humidity. Owing to advances in sensor and communication technology, WBAN enables the exchange of information or commands over short distances between sensor components. Moreover, remote data transmission or control between the sensor components and the database or control center is also available based on WBAN. Such features make the WBAN a very suitable tool for performing continuous monitoring tasks without requiring too much manual intervention, which meets the requirements of sleep monitoring. As a result, many sleep monitoring systems have been developed by researchers and technicians based on the WBAN.

WBAN technology is highly valued in the fields of medical sciences and human health care [63]. In the health care field, WBAN has established itself as a leading technology capable of providing real-time patient health monitoring in hospitals, asylums, and even at home [64]. WBAN allows the removal of cables and the delocalization of instrumentation and intelligence to the sensor nodes themselves, which is useful for establishing a nonintrusive, portable, continuous home sleep monitoring system [55]. Currently, WBAN-based sleep monitoring system has attracted increasing attention from researchers worldwide [65-67]. Similar to the evolution of the WBAN, the trends in sleep monitoring system are miniaturization, intelligence, and long-term monitoring capability. In this section, we have selected a few representative works on WBAN-based sleep monitoring systems in recent years, which we briefly present from a hardware and software perspective.

Hardware Implementation

A WBAN-based sleep monitoring system is a sensor network application in which the sensor is an essential piece of hardware. The choice of sensor determines the type of body parameters that will be acquired, and the position of the sensors directly influences the efficiency, quality of data acquisition, and user acceptance. For these reasons, we list the type and position of sensors used in several works, specify the possible description of sleep phases, and briefly analyze the advantages and drawbacks of each type of sensor, as presented in Multimedia Appendix 1 [68-99]. Table 3 lists the sensors used in each study.

As shown in Multimedia Appendix 1 [68-99], the accelerometer is the most commonly used sensor in these works, usually placed on the wrist or chest or close to both positions. The microphone was adopted only once among these works. However, the microphone is a widely used sensor in sleep apnea monitoring [63,104,105]. As a sound recording sensor, the microphone is useful for detecting snoring or even abnormal breathing [106], which are also important physiological parameters related to the sleep state. Both the ECG sensor and the pulse sensor are used for heart rate monitoring, but due to different detection principles, their positions are different. Multimedia Appendix 1 [68-99] shows that in most cases, the ECG sensor is placed near the chest, whereas the pulse sensor is placed near the wrist. Thus, for user acceptance, the pulse sensor is better than the ECG sensor.

Both the accelerometer and the thermopile sensor can be used for motion detection, but they have their own advantages. In terms of user acceptance, the accelerometer should generally be attached to the user's body, but the thermopile sensor is a noncontact sensor, so the thermopile sensor is preferable. However, with regard to measurement accuracy, thermopile sensors are easily disturbed by the user's coatings, such as duvets, which affect the measurements. In addition, thermopile sensors can only monitor effectively in a limited and fixed area. It is difficult for thermopile sensors to specifically measure the movement of certain parts of the body, such as measuring only leg movement to detect periodic leg movements during sleep. As a result, the accelerometer outperforms the thermopile sensor. In short, the type of sensor to be chosen depends on the application scenario and specific requirements.

Table 3. Choice of sensor for different works related to sleep monitoring.

Sources	Accelerometer	Pressure sensor	Temperature sensor	Thermopile sensor (Infrared)	Microphone	ECG[a] sensor	Pulse sensor
Kalkbrenner et al [70]	✓[b]	—[c]	—	—	✓	—	—
Guettari et al [38]	—	—	—	✓	—	—	—
Seba et al [100]	✓	—	✓	✓	—	—	—
Velicu et al [68]	✓	—	—	—	—	✓	
Suzuki et al [96]	✓	—	✓	—	—	✓	✓
Suzuki et al [76]	✓	—	—	—	—	—	✓
Saad et al [92]	✓	—	✓	—	—	—	✓
Sadek et al [98]	—	✓	—	—	—	—	—
Sadek et al [101]	—	✓	—	—	—	—	—
Lee et al [102]	✓	—	—	—	—	✓	—
Chan et al [97]	✓	—	—	—	—	✓	—
Samy et al [103]	—	✓	—	—	—	—	—

[a]ECG: electrocardiography.

[b]The sensor is included.

[c]The sensor is not included.

Software and Algorithm Processing

Software or algorithms are used to process the data collected by the hardware. Table 4 presents the algorithms, software, and system results illustrated in several books or articles.

Data or signal processing algorithms usually include spectral analysis, wavelet transformation, empirical mode decomposition (EMD), and various varieties of filters. Many sleep-related physiological signals, such as EEG and ECG, are nonstationary. A wavelet analysis is very useful for processing nonstationary signals, which is why it has been adopted by many researchers specializing in sleep monitoring. EMD, proposed by Huang et al [109], is usually used to extract breathing and heartbeat signals from measured data. Unlike wavelet-based decomposition methods, this method is data-driven and does not require a parent wavelet to be defined beforehand. With this technique, any complicated signal can be decomposed into a defined number of high- and low-frequency components, called intrinsic mode functions. This technique is suitable for the analysis of nonlinear and nonstationary biosignals [110] and can extract local temporal structures such as heartbeats superimposed on respiration signals [111]. In sleep monitoring, several types of biosignals of different frequencies are acquired simultaneously. Therefore, filters are effective and simple tools for signal discrimination that are widely adopted in this field.

The classification algorithm is usually used for the classification of sleep stages. Sleep stage classification is an important and common output of sleep monitoring system. Although sleep stages include stages 1, 2, 3, and 4 and stage REM according to the AASM [18], most research classifies sleep stages in a simpler way as *wake*, *lightsleep* (stages 1 and 2), *deep sleep* (stages 3 and 4), *REM* [108] or *wake*, *NREM* (stages 1, 2, 3, and 4), REM [102], or some other similar way. This simplification of sleep stages involves balancing the difficulty of the task with the application requirements. Commonly used classifiers include random forest (RF), support vector machine (SVM), multilayer, feedforward neural network (NN), linear discriminant analysis (LDA), decision tree (DT), and Bayes. Some papers compare the performance of several classifiers in their work to find the best one [101,108].

Table 4. Implementation of software or algorithm in different works.

Sources	Involved algorithms or software	Outputs
Kalkbrenner et al [73]	1. An FIR[a] bandpass filter with boundaries between 200 and 2000 Hz was used on the initial raw tracheal body sound signal acquired by microphone to obtain a pure breathing sound signal 2. A bandpass filter with the boundaries between 5 and 30 Hz was applied on the initial raw tracheal body sound signal acquired by microphone to suppress breathing and most of the artifacts to get heart beat sound 3. A LD[b] classifier was used on cardiorespiratory features and movement features for automated sleep staging	1. Sleep and wake classification 2. Wake, REM[c], and NREM[d] classification 3. Wake, REM, light sleep, and deep sleep classification
Sadek et al [98]	1. Wavelet decomposition was used on microbend fiber optic sensor data to achieve the measuring of heart rate 2. Third-order polynomial fit and Savitzky-Golay smoothing was used on microbend fiber optic sensor data to achieve the measuring of respiratory rate 3. Adaptive thresholding method was used on SD of the respiratory signal for apnea or nonapnea classification 4. Chebyshev type-I bandpass filter was used on microbend fiber optic sensor data to extract BCG[e] and respiratory signals 5. The MODWT[f] with the multiresolution analysis was used on microbend fiber optic sensor data to estimate heart rate	Heart rate, respiratory, and apnea
Guettari et al [38]	1. SAX[g] method was used on thermal sensor data for segmentation processing of the thermal signal 2. SOM[h] algorithm—Kohonen maps is used on features of thermal signal segmentation level, thermal signal segmentation duration and the variance of each thermal signal segmentation for achieving classification	Classification of signal segments as 3 phases of sleep: 1. Deep and paradoxical sleep (*REM*, N3) 2. Agitated and light sleep (N1, N2) 3. Awake phase (*Wake*)
Seba et al [100]	K-means algorithm was used on IButton skin temperature sensor data to achieve data clustering and classification	Classification of the activities into 3 classes: *awakening*, *calm sleep*, and *agitated sleep*
Saad et al [92]	This sleep monitoring system involves Arduino IDE[i] software and Visual Studio 2015. 1. Arduino IDE is programmed that consist of sensor algorithms to enable those sensors and to read the value that has been captured from room ambience and body condition. 2. A window application was programed by Visual Studio to display the value of those parameters	The relationship between the room ambience and quality of sleep
Sadek et al [99]	1. Multiresolution analysis of the maximal overlap discrete wavelet transform was used on piezoelectric sensor data to compute heart rate 2. Bandpass Butterworth filter is used on microbend sensor data to retrieve BCG signal	Heart rate of the person sitting in the massage chair
Velicu et al [68]	Kushida algorithm–derived equation was used on accelerometer data as the discriminator for wake or sleep, wake or REM, and light or deep by applying 3 different thresholds	Discrimination for wake or sleep, wake or REM, and light or deep
Sadek et al [101]	1. For comparison, 5 classifiers are employed, that is, RF[j], SVM[k], multilayer, feed-forward NN[l], LDA[m], and DT[n] 2. Butterworth bandpass filter with frequency limits of 1 Hz and 12 Hz was used on microbend fiber optic sensor data to extract BCG component 3. MATLAB based software was developed as a data labeling tool	Classification of informative and noninformative signal for further heart rate detection work
Kalkbrenner et al [70]	1. The developed software for visualizing and storing received data 2. Heart sound was extracted by applying a bandpass filter ranging from 15 Hz up to 80 Hz on microphone data 3. Breath sound was extracted by applying a bandpass filter ranging from 100 Hz up to 1.5 kHz on microphone data 4. Stable results of accelerometer are provided by using the Madgwick-Filter	Heartbeats, breathing, snoring, sleeping positions, and movements of the volunteer
Sadek et al [107]	1. The BCG signal is decomposed using CEEMDAN (complete ensemble empirical mode decomposition with adaptive noise) 2. Sensor data fusion method: time domain average 3. The BCG signal is extracted using a Butterworth high-pass filter (fifth-order with a cutoff frequency of 0.2 Hz) followed by a Butterworth low-pass filter (10th order with a cutoff frequency of 30 Hz) on microbend fiber optic sensor data	Heart rate

Sources	Involved algorithms or software	Outputs
Suzuki et al [96]	1. Silmee framework provides basic functionality of Silmee system by locating Silmee sensor node, smartphone (or tablet or wearable terminal) and cloud server 2. Silmee firmware provides vital signal processing capabilities such as noise reduction, important information extraction, or data compression 3. Silmee API[o]: This API provides basic information to realize wide-variety of smart healthcare MW[p] and apps 4. Silmee MWs are located in smartphone (or tablet or wearable terminal) or in health care cloud server. The MWs provide less medical expert API than Silmee API. For example, determination of REM and non-REM sleep, which is a popular term, is one of Silmee MW API, which is calculated by R-R intervals information included in the Silmee API	ECG wave, pulse wave, body temperature and body movements were measured by the set and send to a smartphone using a Bluetooth wireless connection
Suzuki et al [76]	1. The Cole algorithm for wake and sleep identification from the amount of activity data 2. Fast Fourier transformation (FFT) is executed for the even-interval pulse-to-pulse intervals to get the frequency spectrum 3. The k-means clustering method is adopted to classify sleep stages	Wristwatch-shaped physiological sensor that monitors user's wrist motion and pulse wave interval
Lee et al [102]	1. To capture the respiratory signal, first-order derivation is used to compensate for the drifting phenomenon of pressure sensor 2. A low-pass filter is applied to eliminate short-term fluctuations in respiration signals 3. Sum all the pixels in the lower half of the pressure image and mark a leg movement when a significant drop or increase in pressure is detected 4. A simple thresholding technique for movement reporting	Classification of sleep stages: 1. Wake 2. NREM 3. REM
Beattie et al [108]	The Scikit library used to explore different types of classifiers: LD classifiers, quadratic discriminant classifiers, RF, and SVM approaches, and the LD classifier achieved the best performance	Classification of sleep stages: 1. Wake 2. Light (N1 or N2) sleep 3. Deep (N3) sleep 4. REM

[a]FIR: Finite impulse response.

[b]LD: linear discriminant.

[c]REM: rapid eye movement.

[d]NREM: nonrapid eye movement.

[e]BCG: ballistocardiography.

[f]MODWT: maximal overlap discrete wavelet transform.

[g]SAX: symbolic aggregate approximation.

[h]SOM: self-organizing map.

[i]IDE: integrated development environment.

[j]RF: random forest.

[k]SVM: support vector machine.

[l]NN: neural network.

[m]LDA: linear discriminant analysis.

[n]DT: decision tree.

[o]API: application programming interface.

[p]MW: middleware.

Research Prototypes

Noncontact Methods

Seba et al [100] discussed the development of a new approach to sleep analysis. This system, based on temperature monitoring (patient and ambient), aims to be integrated into the telemedicine platform developed in the framework of the Smart-EEG project by the SYEL—SYstèmes ELectroniques team. The proposed method is based on the thermal signature to classify the activity into 3 classes: *awakening*, *calm sleep*, and *agitated sleep* by k-means clustering. A thermopile sensor (TMP007) was placed above the bed at a distance close to 2 m to measure the upper *Bed+Patient* temperature. A thermal camera giving images in medical format but also information on the target temperature according to a spatial distribution is used to label the different events related to changes in the patient's posture in the bed by visual analysis of an expert. An inertial unit is used to obtain the wrist acceleration in 3 axes to compare the responses of the thermopile sensor. The system measured the wrist, distal, and proximal skin temperatures using IButtons [112]. The day and night alternation corresponds, on the one hand, to the alternation between wakefulness and sleep and, on the other hand, to the alternation between high and low temperature. During sleep,

the body temperature decreases whereas it increases during the day. Skin temperature, unlike the body temperature, increases during sleeping and decreases after awakening. The work in [113] examined possible mechanisleep monitoring system linking rhythms in sleep and core body and skin temperature, focusing on the causal effects of changes in core and skin temperature on sleep regulation. Several studies refer to the links between body temperature and sleep [112,114]. This work gave a figure of an example of classification for sleep base on thermal data.

Guettari et al [38] presented the design and first evaluation of a new monitoring system based on contactless sensors to estimate sleep quality. A passive thermopile sensor fixed on the wall produces thermal signals to detect human presence in the bed and then to estimate the quality of sleep. The Symbolic Aggregate Approximation (SAX) method has been implemented [115], which uses Gaussian window in the thermal signal segmentation processing. Each segment is generated by the SAX method based on a segmentation of the midvariance and then by identifying its sleep phase. The system extracted 3 features: the duration of the thermal data segment, the variance of the thermal segment of each segment, and the level of each segment. The Kohonen self-organized map (SOM) [116] was used to classify the signal segments into 3 sleep phases: deep or paradoxal sleep (R, N3), agitated or light sleep (N1, N2), and awake phase (W). It synchronized the thermal signal with the sleep stage labels based on the physiological parameters measured by the PSG with a hypnogram being established manually by the physicians. This study involved 13 patients, 11 people for learning the SOM model, and 2 other patients for evaluation of the learned SOM model. In total, 87% (40/46) of evaluation results showed good classifications.

Gu et al [41] presented Sleep Hunter, a mobile service that detects the transition between sleep stages for monitoring sleep quality and intelligent wakefulness. The smartphone was placed next to the participant's pillow. Using sensors integrated in smartphones, Sleep Hunter integrates body movements, acoustic events, environmental lighting conditions, sleep duration, and personal factors using a statistical model: linear-chain conditional random field (CRF) [117] for sleep stage detection. It argued that, compared with the hidden Markov model [118], CRFs are more relevant for sequences that have long interdependencies and may therefore perform better in this application. On the basis of the duration of each sleep stage, Sleep Hunter also provides a report on sleep quality and a smart call service for users. In this work [41], commercial product *Zeo* [119] was adopted as the reference device. One study [120] indicated that the quality of sleep is actually determined by the distribution of the different stages of sleep rather than the length of sleep during the night. This work distinguished sleep stages between wakefulness, light sleep, deep sleep, and REM. The sleep quality score is then calculated based on the duration of each sleep stage. The detection accuracy of the Sleep Hunter proposed in this work [41] was 64.55%.

Krishna et al [121] proposed SleepSensei, an automated sleep quality monitor that estimates the sleep duration for the user. It uses (1) the built-in web camera and microphone of a personal computer connected to a power source, and custom software to collect environmental features and (2) the accelerometer sensor of a smartphone to detect body movements. Smartphones are placed close to the user (next to the pillow). In this system, the user can be in 1 of 2 sleep states: *deep sleep* or *light sleep*. The user's sleep state (sound or light) is determined solely on the basis of the variance of the user's body movements during sleep. Environmental features such as light intensity, ambient noise, temperature, and humidity have been entered by using custom software.

Temperature, ambient sound (noise and music), and light conditions have been found to be strong indicators of the user's environment that clearly affect sleep [122]. The system proposed a regression model consisting of linear regression and SVM regression. The regression model estimates the share of each time slot (30-min window) that contributes to the completion of a user's sleep quota (the total duration of sleep a user needs to obtain *satisfactory* sleep). The ground truth of this system comes from the data provided by users on sleep quality by answering the question: was the sleep *fulfilling*? It uses SVM and naive Bayes models as classifiers. By comparing the results of each classifier with 2 and 4 times cross-validation, the SVM model with 2 times cross validations has the best results and has an average accuracy of 79.84%.

Contact Methods

Distributed Sensor System on the Body

Velicu et al [68] proposed a system based on an accelerometer and an ECG sensor for the classification of sleep phases (wakefulness, light sleep, deep sleep, and REM). The accelerometer was embedded into a wristband, but the position of the ECG sensor was not mentioned. It described the classification logic: (1) body movements become less intense and less frequent as we enter the deeper phases of sleep and (2) HR becomes more stable as sleep deepens. The Kushida algorithm–derived equation [69] was adopted in this system as a discriminator between wake and sleep using accelerometer data collected every minute, with a 9-min sliding time window, showing 69% agreement with the EEG sensor result. This work shows a part of the classification results for an experiment lasting 3 hour and 43 min. However, the results have not been validated against PSG or any other reliable standard.

Kalkbrenner et al [70] presented the first step in the development of a sleep monitoring system. It includes the capabilities of capturing heartbeats, breathing, snoring, sleeping positions, and movements of 2 volunteers. In this system, a microphone was set up at the suprasternal notch to record breathing sounds and heart sounds. The heart signal is extracted by applying a bandpass filter from 15 Hz to 80 Hz. Nakano [71] and Yadollahi [72] showed that placing a stethoscope such as a microphone in the suprasternal notch at night can detect sleep apnea. At the same time, an MPU6000 inertial measurement unit embedded in an abdominal belt worn by the patient determines the sleep position and movements. The data are transmitted wirelessly to the laptop via Bluetooth and processed, visualized, and stored using developed software. The validation of the proposed system by comparison with the gold standard was published in [73]. A total of 60 adult subjects were subjected to overnight diagnosis, and a PSG screening was included for validation of the proposed

system. A total of 30 dimensions of features were extracted from data on breath, heartbeat, and movement. A linear discriminant (LD) classifier was used for automated sleep staging. The classifier achieved 86.9% accuracy and a kappa of 0.69 for sleep or wake classification, 76.3% accuracy and a kappa of 0.42 for Wake or REM or NREM classification, and 56.5% accuracy and a kappa of 0.36 for wake or REM or light sleep or deep sleep classification.

Lee et al [74] proposed smart patches and wearable bands (W-band) for recording biosignals during sleep. The system consists of 15 smart patches attached to the user's face to monitor multiple biosignals (EEG, ECG, EMG, and EOG). A total of 14 biosignal sensing (SN) patches to monitor biosignals, a network controller (NC) patch placed behind the ear to manage the whole system and used as a reference electrode for ECG, EEG, and EOG signals. All electrodes are implemented on a multilayered fabric patch based on the Planar Fashionable Circuit Board technology. The biosignals recorded by the SN patches were collected in the NC patch with an internal 20 kb SRAM via the W-band. When the memory is full, the recorded data are transmitted to an external device via an inductively coupled interface. The program displaying the data runs on an external PC so that the user can check the monitoring result after waking up. The performance of biosignal recording by this system has not been compared with a gold standard.

Stand-Alone System With Several Sensors

Shambroom et al [75] evaluated a wireless system for automatic collection and scoring of human sleep. This system uses dry silver-coated fabric sensors in the headband to collect electrophysiological signals from the forehead, which include contributions from the EEG and eye and frontal muscle movements. The resulting signal is transmitted to a base station using an ultra–low-power wireless protocol at 2.4 GHz. The system was compared with the PSG data scored by 2 technicians according to the R&K criteria. A reduced set of sleep stage classifications was adopted, including awake, REM, light sleep (combined stages N1 and N2), and deep sleep (combined Stages N3 and N4) [17]. A total of 26 healthy adults were subjected to simultaneous sleep measurements using this system and the PSG. The agreement was 62% and 56%, respectively, with regard to PSG1 (PSG recording scored by technician 1) and PSG2 (PSG recording scored by technician 2). The mean (SD) entire night sleep stage agreement for the 26 subjects was 75.9% (7.0%) for this system versus PSG1 (PSG recording scored by technician 1), 74.7% (8.5%) for this system versus PSG2 (PSG recording scored by technician 2), and 81.2% (7.4%) for this system versus PSGC (PSG recording scored consistently by 2 technicians).

Suzuki et al [76] described a wristwatch-shaped wearable sleep monitoring system for home use. The sensor incorporates a photoelectric pulse wave sensor and a 3-axis accelerometer to measure pulse waves and accelerations on a user’s wrist and stores the computed pulse intervals (PPIs) and amount of activity in a flash memory (4 MB). It uses the Cole algorithm to identify wake or sleep from the amount of activity data [77]. The system compared the estimation result with the PSG results. Fast Fourier transform (FFT) is performed to obtain the heart rate spectrum. In the frequency domain, the integral value of the power from 0.04 Hz to 0.15 Hz is called LF (low frequency), which shows both sympathetic and parasympathetic nervous activities. The integral value of the power from 0.15 Hz to 0.4 Hz is called HF (high frequency), which shows parasympathetic nervous activity. The balance between sympathetic and parasympathetic nervous activity is related to sleep stages. According to the study by Baharav et al [78], there is a decrease in LF during sleep, with minimal values during non-REM slow-wave sleep, that is, deep sleep, and high levels similar to those of wakefulness during REM.

The HF increased with the onset of sleep, reaching maximal values during slow-wave sleep, and behaved as a mirror image of LF, as shown in Figure 6. The correlation between HF, LF of PPI, and sleep stages is summarized in Table 5.

The sympathetic predominance that characterizes wakefulness decreases during non-REM sleep, is minimal during slow wave sleep, and approaches average levels of wakefulness during REM sleep. Autonomic balance shifts to parasympathetic predominance during slow-wave sleep. To classify sleep stages from the LF and HF data sets, the k-means clustering method is adopted. It defines the coincidence ratio as a moving average sleep stage correlation coefficient (20-min window) between the stages estimated by this method and those estimated by the PSG. A mean coincidence ratio of 0.735 (SD 0.052) was obtained for the classification of the SWS, REM, non-REM, and wake stages.

Beattie et al [108] estimated sleep stages using a wrist-worn device that measured movements using a 3D accelerometer and an optical pulse photoplethysmograph, which provided data on movement, breathing, and heart rate variability. Overnight recordings were obtained from 60 adult participants wearing these devices on their left and right wrists, simultaneously with a type III home sleep testing device (Embletta MPR) that included EEG channels for sleep staging. The reference Embletta recordings were scored for sleep stages using the AASM guidelines [79], which labeled sleep stages as awake, light (N1 or N2), deep (N3), and REM over 30 second epoch level. Motion-based features include the number of activities over 30 second, the magnitude of rotation (using the 3D accelerometer to combine the maximum−minimum of each axis), the time from the last significant movement, and the time to the next significant movement. It extracted heart rate features such as HF power (0.15-0.4 Hz), LF power (0.04-0.15 Hz), very low frequency (VLF) power (0.01-0.04 Hz), root mean square of the successive differences, pNN50 (proportion of the number of pairs of successive RR intervals (the interval between R waves of ECG, that is, the time between heart beats) that differ by more than 50 msec divided by the total number of RR intervals), delta RR (intervals between beats), and mean heart rate.

The spectral features of the estimated breathing rate on a 1 s basis such as HF power (0.15-0.4 Hz), LF power (0.04-0.15 Hz), and VLF power (0.015-0.04 Hz) were formed. This system used the Scikit library to explore different types of classifiers: LD classifiers, quadratic discriminant classifiers, RFs, and SVM approaches. The LD mode seems to work slightly better than

the others, so it was chosen as the final model. On the basis of a single validation, the overall accuracy per epoch of the automated algorithm was 69%, with a Cohen kappa of 0.52 (SD 0.14).

Figure 6. Illustration of HF and LF from PPI [97].

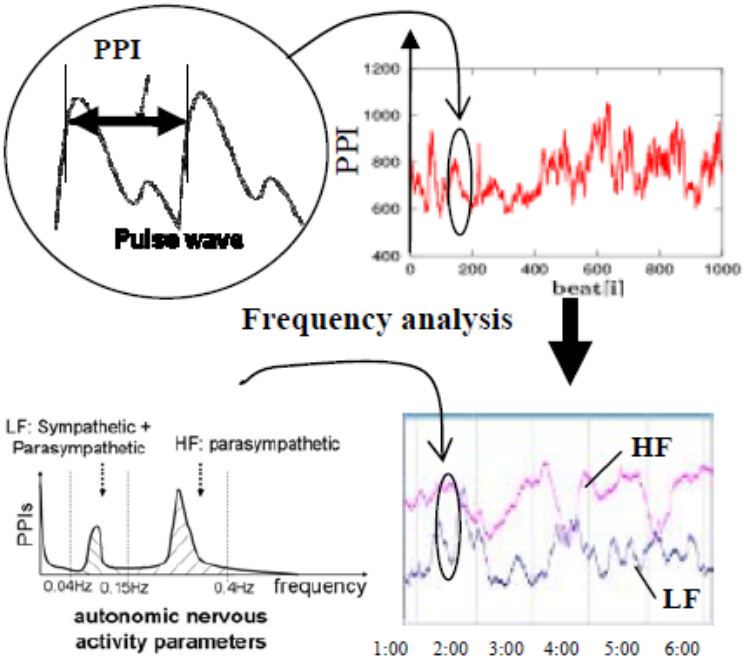


Table 5. Correlation between high frequency and low frequency from pulse-to-pulse intervals and sleep stages.

Frequency bands	Sleep onset	Slow wave sleep	Wakefulness
HF[a]	Increase	Maximal	Low level
LF[b]	Decrease	Minimal	high level

[a]HF: high frequency.

[b]LF: low frequency.

Only One Sensor Attached to One Site of Body

Tataraidze et al [80] presented an algorithm for detecting wakefulness, REM, and non-REM sleep based on a set of 33 features extracted from the respiratory inductive plethysmography signal captured by the PSG thoracic belt. The features extracted include the logarithm of power in different frequency ranges, time and frequency domain features, motion, breathing, and volume-based features. A bagging classifier was used in the experiments and a heuristic algorithm was applied to increase the performance of the classification. Compared with the PSG gold standard, an accuracy of 80.38 (SD 8.32%) and a Cohen kappa of 0.65 (SD 0.13) were obtained with the classifier.

Commercial Products

Given the various shortcomings of PSG, such as its invasiveness, high cost, and one-night monitoring, the industry has shown great enthusiasm for the development of commercial sleep monitoring products with the advantages of being portable, noninvasive, and suitable for long-term monitoring. Commercial products used for home sleep monitoring are currently available for direct purchases on the market. Some of the most popular and representative products are briefly introduced below.

Zeo [119] is a headband based on a true lightweight EEG brainwave *pod* monitor. It can provide a classification of sleep stages into *awake*, *light* (stages 1 and 2 combined), *deep* (stages 3 and 4 combined), and REM sleep. The *Companion for Zeo* smartphone app was developed for data collection. A validation study has been published [75]. Compared with the PSG, the epoch-to-epoch concordance of *light*, *deep*, and REM sleep is greater than 74%.

Up (Jawbone) [81] is a soft rubber wristband. In terms of sleep monitoring, it provides total sleep duration, time to fall asleep, and the number of nighttime awakenings. It also interacts with smartphone apps. To date, there have been no validation studies.

Fitbit [82] is also a wristband product. Its sleep monitoring algorithm classifies night sleep into *awake*, *light sleep*, *deep sleep*, and *REM* based on wrist movements and heart rate data. It also provides total sleep duration, sleep starting time, and end time. The publication [83] evaluates the performance of the Fitbit against the PSG. It showed a sensitivity of 0.96 (sleep detection accuracy), a specificity of 0.61 (wakefulness detection accuracy), an accuracy of 0.81 for the detection of N1+N2 sleep (*light sleep*), an accuracy of 0.49 for the detection of N3 sleep (*deep sleep*), and an accuracy of 0.74 for the detection of REM sleep.

RestOn [84] is a thin belt. It uses a single click of the magnetic cover to fix the device on the bedsheet; the position corresponds to the user's chest. RestOn can measure heart rate and respiratory rate in real time. Medical-grade sensors that are 2-foot long are embedded into a thin belt less than a length of 2 mm. The device can provide sleep time, actual sleep time, and sleep stages including *awake*, *light*, *medium*, and *deep* sleep. Its smart alarm can wake the user during the lightest sleep time.

The Sleep Dot [85] measures sleep cycles and body movements by simply attaching it to the upper corner of the pillow. It can play music to help the user fall asleep. Soothing sounds and music are adopted as alarm tones to wake the user more naturally during the lightest sleep. This product works with a smartphone app and generates a sleep report that can be shared with family and friends.

Withings Aura [86] is a bedside device with a white cloth sleep sensor placed under the mattress, aligned in position with the user's chest. It is recommended that the 11-inch high bedside device be placed at least 1 m from the bed. The bedside device measures environmental parameters such as temperature, light, and noise. The white cloth sleep sensor indicates the time to sleep, number of awakenings during the night, duration of light sleep or deep sleep or REM sleep, and percentage of sleep goal achieved.

Our Proposition of Sleep Monitoring System

The LAAS-CNRS (Laboratory for Analysis and Architecture of Systems-French National Centre for Scientific Research) has been developing systems for monitoring people since 1990 [87,88]. This mainly includes research on health monitoring for older adults and frail people at home based on a wireless local body area network (WLBAN) [89-91]. On the basis of the techniques and experiences of studying the WLBAN monitoring system for many years, our laboratory has started research on sleep monitoring. The objective is to perform continuous and long-term sleep monitoring at home, focusing on the variance of sleep conditions night by night and monitoring indicator deviations after each night. The first week's sleep is considered the reference nights and is used for comparison of the following nights. When an abnormal condition is detected, the user is warned via a specific visual interface (graph or other) on a computer, tablet, or smartphone.

Many indicators may be relevant for monitoring sleep conditions, such as the following:

1. Distribution of sleep stages during the night.
2. Number of wrist activities.
3. Amount of leg activity during sleep (this is also the main indicator of Restless Legs Syndrome [RLS]).
4. Body turnover time.
5. Onset of snoring. If so, the duration and number of apneas.

The contribution here is to propose the following indexes:

- Change in apnea behavior.
- Sleep phase deviation.
- Leg movement.
- Heart rate and blood oxygen saturation.
- Finger, big toe, and chest temperature.
- Temperature and brightness of the environment.

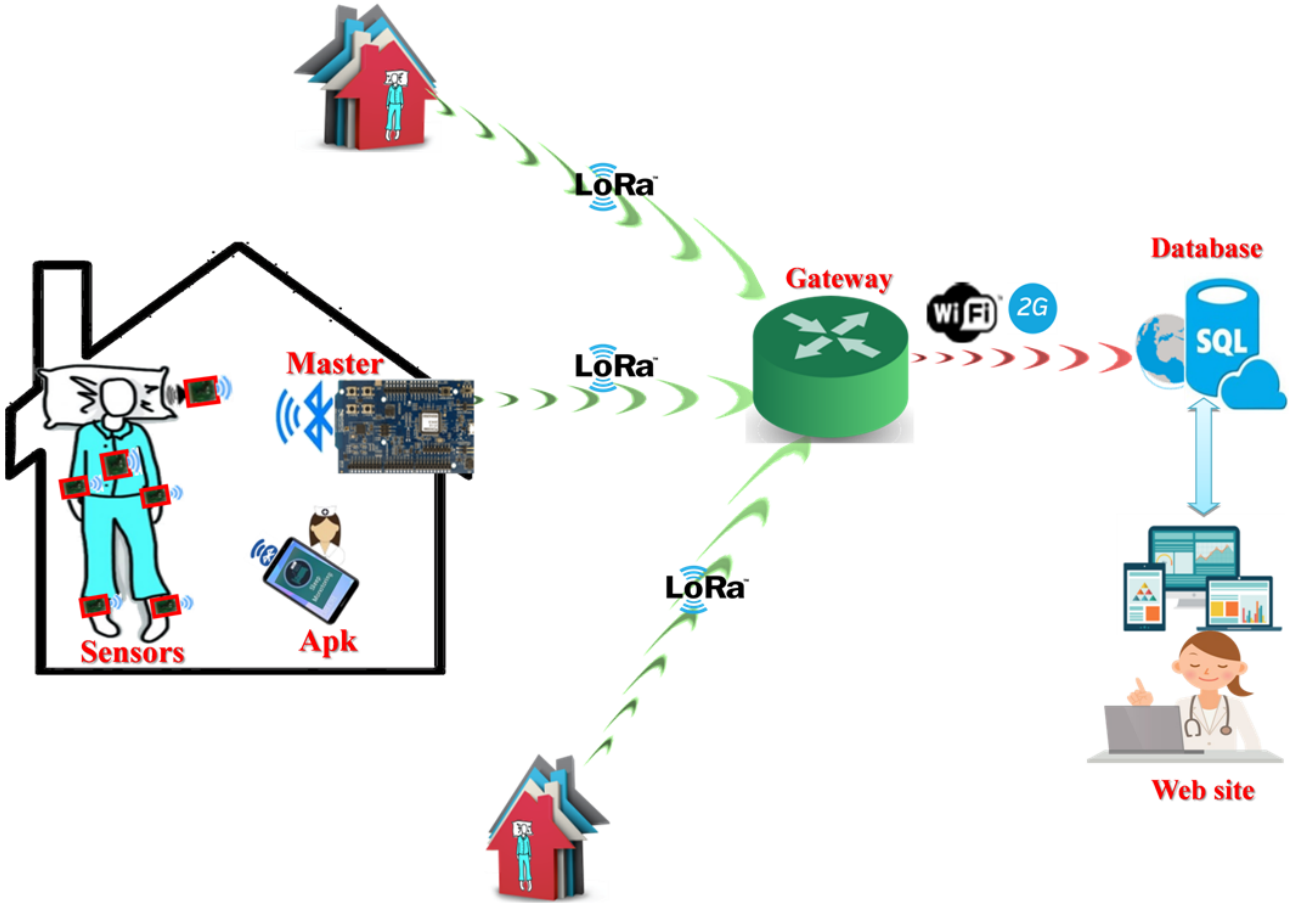
Thresholds on these indexes could trigger alerts to the doctor. It is also possible to correlate these indices to provide an overall index of sleep quality.

The current research at the LAAS-CNRS on sleep monitoring is shown in Figure 7. The developed system consists of a local network integrating sensors embedded on the person, and these sensors communicate using Bluetooth to a data concentrator (Master). The master communicates via an Android app with the person itself. The master transmits the data via the LoRa technology to a gateway that takes care of sending the data to a database that can be remotely consulted by the physician.

Using this system, we collect a hypnogram based on wrist movement data using the clustering algorithm and to monitor the restless leg syndrome using leg movement data and to detect symptoms such as snoring using sound data.

We proposed threshold-based and k-means clustering based methods to process acceleration data from the nondominant wrist. The threshold-based method uses 3 thresholds to achieve falling asleep or waking up detection and 4-sleep stages classification (*awake*, *light sleep*, *deep sleep*, and *REM*). The k-means clustering-based method performs 5-iteration of k-means clustering for epochs between falling asleep and waking up to also achieve a 4-sleep stage classification (*awake*, *light sleep*, *deep sleep* and *REM*). The epochs between falling asleep and waking up are determined by falling asleep or waking up detection from the threshold-based method. Furthermore, a method for calculating sleep scores for a night's sleep is also proposed in our system. The calculation of the sleep score was based on the sleep duration and the duration of each sleep stage. With tests carried out on 5 volunteers, all the methods we propose give promising results. The related article will be submitted soon.

Figure 7. Sleep monitoring system ongoing research at LAAS (Laboratory for Analysis and Architecture of Systems).



Conclusions

The objective of this paper is to provide an overview of the current state and future prospects of research and development of sleep monitoring system s. Observation and sleep monitoring is a very important medical issue for the possible consequences on life behavior. The gold standard used is the PSG technique, which is an intrusive method that can only be used in a clinical setting. In addition, several studies have focused on the development of methods and strategies for lighter and longitudinal monitoring. sleep monitoring systems have been proposed but they raise questions about the acceptance of the wearing of these devices by users, socioeconomic aspects, privacy, and impact on society, but also about the performance of the proposed algorithmic processing. This document deals with these issues and the different solutions reported in the literature and available on the market. The sleep monitoring system features a broad and heterogeneous range of devices, WSN standards, apps, and involve the efforts of numerous researchers, developers, and users. Owing to its interdisciplinary nature, a number of apps related to sleep monitoring integrate biomedical engineering and medical informatics. Other knowledge in the fields of medicine, social sciences, psychology, economics, ethics, and law must be taken into account and integrated into the development and deployment of wearable health care systems. Most systems are still in their prototype stages, and developers have not yet faced deployment issues. Information technology and electronics are mature fields and can provide viable, disposable, and affordable wearable systems. Systematic evaluations of the effectiveness and efficiency of sleep monitoring system are considered crucial to ensure potential user acceptance. Sleep monitoring is important for individuals and clinicians. Beyond the interest in healthy lifestyle and clinical diagnosis, sleep monitoring may also be important in reducing fatigue-related workplace injuries, particularly for shift workers. However, this type of monitoring will only be practical if systems with proven reliability and validity are in place. Consumers and patients will have the opportunity to take part in the revolution in personal health data. Increasingly powerful and convenient wearable technologies will be able to provide rich health information, but it is not clear that this will translate into workable health decisions. The democratization of devices previously *reserved for doctors* should improve access to health data and overall awareness of personal health. It is important that such information is properly communicated and understood by consumers. More complex integrated sensor technologies, detection, and analytical algorithms are likely to be developed in the coming years. Other wearable diagnostic tools for consumers, or even implantable devices and nanotechnologies, are currently under development. Products exist that can integrate sensors into clothing. Ideally, these technologies will empower consumers and patients and promote preventive medicine. The most important challenges are the development of nonintrusive hardware implementation, smart signal processing, data analysis and interpretation, communication standards interoperability, electronic component efficiency, energy self-sufficiency, and long-term monitoring.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Hardware implementation in different works.
[DOCX File , 24 KB - biomedeng_v5i1e20921_app1.docx]

References

1. van de Straat V, Buffel V, Bracke P. Medicalization of sleep problems in an aging population: a longitudinal cross-national study of medication use for sleep problems in older European adults. J Aging Health 2018 Jun;30(5):816-838. [doi: 10.1177/0898264317696775] [Medline: 28553808]
2. Bao YP, Han Y, Ma J, Wang RJ, Shi L, Wang TY, et al. Cooccurrence and bidirectional prediction of sleep disturbances and depression in older adults: meta-analysis and systematic review. Neurosci Biobehav Rev 2017 Apr;75:257-273. [doi: 10.1016/j.neubiorev.2017.01.032] [Medline: 28179129]
3. Reutrakul S, van Cauter E. Sleep influences on obesity, insulin resistance, and risk of type 2 diabetes. Metabolism 2018 Jul;84:56-66. [doi: 10.1016/j.metabol.2018.02.010] [Medline: 29510179]
4. Tobaldini E, Costantino G, Solbiati M, Cogliati C, Kara T, Nobili L, et al. Sleep, sleep deprivation, autonomic nervous system and cardiovascular diseases. Neurosci Biobehav Rev 2017 Mar;74(Pt B):321-329. [doi: 10.1016/j.neubiorev.2016.07.004] [Medline: 27397854]
5. Suzuki H, Savitz J, Kent Teague T, Gandhapudi SK, Tan C, Misaki M, et al. Altered populations of natural killer cells, cytotoxic T lymphocytes, and regulatory T cells in major depressive disorder: association with sleep disturbance. Brain Behav Immun 2017 Nov;66:193-200 [FREE Full text] [doi: 10.1016/j.bbi.2017.06.011] [Medline: 28645775]
6. Lewis PA, Knoblich G, Poe G. How memory replay in sleep boosts creative problem-solving. Trends Cogn Sci 2018 Jun;22(6):491-503. [doi: 10.1016/j.tics.2018.03.009] [Medline: 29776467]
7. Huber R, Born J. Sleep, synaptic connectivity, and hippocampal memory during early development. Trends Cogn Sci 2014 Mar;18(3):141-152. [doi: 10.1016/j.tics.2013.12.005] [Medline: 24462334]
8. Scullin MK, Bliwise DL. Sleep, cognition, and normal aging: integrating a half century of multidisciplinary research. Perspect Psychol Sci 2015 Jan;10(1):97-137 [FREE Full text] [doi: 10.1177/1745691614556680] [Medline: 25620997]
9. Gami AS, Howard DE, Olson EJ, Somers VK. Day-night pattern of sudden death in obstructive sleep apnea. N Engl J Med 2005 Mar 24;352(12):1206-1214. [doi: 10.1056/NEJMoa041832] [Medline: 15788497]
10. Stranges S, Tigbe W, Gómez-Olivé FX, Thorogood M, Kandala NB. Sleep problems: an emerging global epidemic? Findings from the INDEPTH WHO-SAGE study among more than 40,000 older adults from 8 countries across Africa and Asia. Sleep 2012 Aug 1;35(8):1173-1181 [FREE Full text] [doi: 10.5665/sleep.2012] [Medline: 22851813]
11. Dregan A, Armstrong D. Cross-country variation in sleep disturbance among working and older age groups: an analysis based on the European social survey. Int Psychogeriatr 2011 Nov;23(9):1413-1420. [doi: 10.1017/S1041610211000664] [Medline: 21554795]
12. Sutton DA, Moldofsky H, Badley EM. Insomnia and health problems in Canadians. Sleep 2001 Sep 15;24(6):665-670. [doi: 10.1093/sleep/24.6.665] [Medline: 11560179]
13. Kessler RC, Berglund PA, Coulouvrat C, Hajak G, Roth T, Shahly V, et al. Insomnia and the performance of US workers: results from the America insomnia survey. Sleep 2011 Sep 1;34(9):1161-1171 [FREE Full text] [doi: 10.5665/SLEEP.1230] [Medline: 21886353]
14. Uehli K, Mehta AJ, Miedinger D, Hug K, Schindler C, Holsboer-Trachsler E, et al. Sleep problems and work injuries: a systematic review and meta-analysis. Sleep Med Rev 2014 Feb;18(1):61-73 [FREE Full text] [doi: 10.1016/j.smrv.2013.01.004] [Medline: 23702220]
15. Filip I, Tidman M, Saheba N, Bennett H, Wick B, Rouse N, et al. Public health burden of sleep disorders: underreported problem. J Public Health 2016 Dec 6;25(3):243-248 [FREE Full text] [doi: 10.1007/s10389-016-0781-0]
16. Hafner M, Stepanek M, Taylor J, Troxel WM, van Stolk C. Why sleep matters-the economic costs of insufficient sleep: a cross-country comparative analysis. Rand Health Q 2017 Jan;6(4):11 [FREE Full text] [Medline: 28983434]
17. Allan Hobson J. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Clin Neurophysiol 1969 Jun;26(6):644. [doi: 10.1016/0013-4694(69)90021-2]
18. Iber C, Ancoli-Israel S, Chesson A, Quan S. The AASM Manual for the Scoring of Sleep and Associated Events. American Academy of Sleep Medicine. 2007. URL: https://aasm.org/clinical-resources/scoring-manual/ [accessed 2020-08-13]
19. Moser D, Anderer P, Gruber G, Parapatics S, Loretz E, Boeck M, et al. Sleep classification according to AASM and Rechtschaffen & Kales: effects on sleep scoring parameters. Sleep 2009 Feb;32(2):139-149 [FREE Full text] [doi: 10.1093/sleep/32.2.139] [Medline: 19238800]

20. Himanen S, Hasan J. Limitations of Rechtschaffen and Kales. Sleep Med Rev 2000 Apr;4(2):149-167. [doi: 10.1053/smrv.1999.0086] [Medline: 12531164]
21. Fino E, Mazzetti M. Monitoring healthy and disturbed sleep through smartphone applications: a review of experimental evidence. Sleep Breath 2019 Mar;23(1):13-24. [doi: 10.1007/s11325-018-1661-3] [Medline: 29687190]
22. Westerlund A, Lagerros YT, Kecklund G, Axelsson J, Åkerstedt T. Relationships between questionnaire ratings of sleep quality and polysomnography in healthy adults. Behav Sleep Med 2016;14(2):185-199. [doi: 10.1080/15402002.2014.974181] [Medline: 25384098]
23. Sateia MJ. International classification of sleep disorders-third edition: highlights and modifications. Chest 2014 Nov;146(5):1387-1394. [doi: 10.1378/chest.14-0970] [Medline: 25367475]
24. Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. Sleep Med Rev 2002 Apr;6(2):97-111. [doi: 10.1053/smrv.2002.0186] [Medline: 12531146]
25. Avcı C, Akbaş A. Sleep apnea classification based on respiration signals by using ensemble methods. Bio-Med Mater Eng 2015 Aug 17;26(s1):S1703-S1710. [doi: 10.3233/bme-151470]
26. Hou L, Xie S, Kai S. Detection of OSAHS Using Only Time-Domain Property of Snoring Signal. In: International Conference on Multimedia Technology. 2011 Presented at: ICMT'11; Jul 26-28, 2011; Hangzhou, China. [doi: 10.1109/icmt.2011.6002183]
27. Karunajeewa AS, Abeyratne UR, Hukins C. Multi-feature snore sound analysis in obstructive sleep apnea-hypopnea syndrome. Physiol Meas 2011 Jan;32(1):83-97. [doi: 10.1088/0967-3334/32/1/006] [Medline: 21119221]
28. Solà-Soler J, Fiz JA, Morera J, Jané R. Multiclass classification of subjects with sleep apnoea-hypopnoea syndrome through snoring analysis. Med Eng Phys 2012 Nov;34(9):1213-1220. [doi: 10.1016/j.medengphy.2011.12.008] [Medline: 22226588]
29. Walters AS, LeBrocq C, Dhar A, Hening W, Rosen R, Allen TP, International Restless Legs Syndrome Study Group. Validation of the international restless legs syndrome study group rating scale for restless legs syndrome. Sleep Med 2003 Mar;4(2):121-132. [doi: 10.1016/s1389-9457(02)00258-7] [Medline: 14592342]
30. Plante DT. Leg actigraphy to quantify periodic limb movements of sleep: a systematic review and meta-analysis. Sleep Med Rev 2014 Oct;18(5):425-434 [FREE Full text] [doi: 10.1016/j.smrv.2014.02.004] [Medline: 24726711]
31. Terrill PI, Leong M, Barton K. Measuring Leg Movements During Sleep Using Accelerometry: Comparison With EMG and Piezo-Electric Scored Events. In: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2013 Presented at: EMBC'13; July 3-7, 2013; Osaka, Japan p. -6865. [doi: 10.1109/embc.2013.6611134]
32. Madhushri P, Ahmed B, Penzel T. Periodic Leg Movement (PLM) Monitoring Using a Distributed Body Sensor Network. In: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2015 Presented at: EMBC'15; August 25-29, 2015; Milan, Italy. [doi: 10.1109/embc.2015.7318738]
33. Ohayon MM, Roth T. Prevalence of restless legs syndrome and periodic limb movement disorder in the general population. J Psychosom Res 2002 Jul;53(1):547-554. [doi: 10.1016/s0022-3999(02)00443-9] [Medline: 12127170]
34. Montplaisir J, Allen RP, Arthur W. Restless Legs Syndrome and Periodic Limb Movements During Sleep. Italian Ministry of Health. 2010. URL: https://moh-it.pure.elsevier.com/en/publications/restless-legs-syndrome-and-periodic-limb-movements-during-sleep-2 [accessed 2020-08-13]
35. Zucconi M, Ferini-Strambi L. NREM parasomnias: arousal disorders and differentiation from nocturnal frontal lobe epilepsy. Clin Neurophysiol 2000 Sep;111(Suppl 2):S129-S135. [doi: 10.1016/s1388-2457(00)00413-2] [Medline: 10996566]
36. Ferini-Strambi L, Zucconi M. REM sleep behavior disorder. Clin Neurophysiol 2000 Sep;111:S136-S140. [doi: 10.1016/s1388-2457(00)00414-4]
37. Gitanjali B. Establishing a polysomnography laboratory in India: problems and pitfalls. Sleep 1998 Jun 15;21(4):331-332. [doi: 10.1093/sleep/21.4.331] [Medline: 9646376]
38. Guettari T, Istrate D, Boudy J, Benkelfat B, Fumel B, Daviet J. Design and first evaluation of a sleep characterization monitoring system using a remote contactless sensor. IEEE J Biomed Health Inform 2017 Nov;21(6):1511-1523. [doi: 10.1109/jbhi.2016.2639823]
39. Liu J, Chen Y, Wang Y, Chen X, Cheng J, Yang J. Monitoring vital signs and postures during sleep using WiFi signals. IEEE Internet Things J 2018 Jun;5(3):2071-2084. [doi: 10.1109/jiot.2018.2822818]
40. Roomkham S, Lovell D, Cheung J, Perrin D. Promises and challenges in the use of consumer-grade devices for sleep monitoring. IEEE Rev Biomed Eng 2018;11:53-67. [doi: 10.1109/RBME.2018.2811735] [Medline: 29993607]
41. Gu W, Shangguan L, Yang Z, Liu Y. Sleep hunter: towards fine grained sleep stage tracking with smartphones. IEEE Trans on Mobile Comput 2016 Jun 1;15(6):1514-1527. [doi: 10.1109/tmc.2015.2462812]
42. Barrón-González HG, Martínez-Espronceda M, Trigo JD, Led S, Serrano L. Proposal of a novel remote command and control configuration extension for interoperable personal health devices (PHD) based on ISO/IEEE11073 standard. Conf Proc IEEE Eng Med Biol Soc 2014;2014:6312-6315. [doi: 10.1109/EMBC.2014.6945072] [Medline: 25571440]
43. Arney D, Plourde J, Goldman JM. OpenICE medical device interoperability platform overview and requirement analysis. Biomed Tech (Berl) 2018 Feb 23;63(1):39-47. [doi: 10.1515/bmt-2017-0040] [Medline: 28734113]
44. McHill AW, Wright KP. Role of sleep and circadian disruption on energy expenditure and in metabolic predisposition to human obesity and metabolic disease. Obes Rev 2017 Feb;18(Suppl 1):15-24. [doi: 10.1111/obr.12503] [Medline: 28164449]

45. Lee SW, Ng KY, Chin WK. The impact of sleep amount and sleep quality on glycemic control in type 2 diabetes: a systematic review and meta-analysis. Sleep Med Rev 2017 Feb;31:91-101. [doi: 10.1016/j.smrv.2016.02.001] [Medline: 26944909]
46. Fuchs FD, Fuchs SC, Martinez D. Obstructive sleep apnea-Hypertension link: almost there? J Thorac Dis 2017 Oct;9(10):3537-3540 [FREE Full text] [doi: 10.21037/jtd.2017.08.162] [Medline: 29268335]
47. Hossain JL, Shapiro CM. The prevalence, cost implications, and management of sleep disorders: an overview. Sleep Breath 2002 Jun;6(2):85-102. [doi: 10.1007/s11325-002-0085-1] [Medline: 12075483]
48. Deng F, Dong J, Wang X, Fang Y, Liu Y, Yu Z, et al. Design and implementation of a noncontact sleep monitoring system using infrared cameras and motion sensor. IEEE Trans Instrum Meas 2018 Jul;67(7):1555-1563. [doi: 10.1109/tim.2017.2779358]
49. Okubo M, Imai Y, Ishikawa T. Development of Automatic Respiration Monitoring for Home-care Patients of Respiratory Diseases With Therapeutic AIDS. In: 4th European Conference of the International Federation for Medical and Biological Engineering. 2008 Presented at: ECIFMBE'08; November 23-27, 2008; Antwerp, Belgium. [doi: 10.1007/978-3-540-89208-3_267]
50. Coughlin JF. Aging & Family Caregiving: Why Should Financial Services Care? Massachusetts Institute of Technology. 2006. URL: http://web.mit.edu/coughlin/Public/Publications/Coughlin%20Caregiving%20&%20Financial%20Services.pdf [accessed 2020-08-13]
51. Barlow J, Bayer S, Curry R. Implementing complex innovations in fluid multi-stakeholder environments: experiences of 'telecare'. Technovation 2006 Mar;26(3):396-406. [doi: 10.1016/j.technovation.2005.06.010]
52. World Internet Usage and Population Statistics. Internet World Stats. 2020. URL: https://www.internetworldstats.com/stats.htm [accessed 2020-03-03]
53. Kelly JM, Strecker RE, Bianchi MT. Recent developments in home sleep-monitoring devices. ISRN Neurol 2012;2012:768794 [FREE Full text] [doi: 10.5402/2012/768794] [Medline: 23097718]
54. Coiera E. Four rules for the reinvention of health care. Br Med J 2004 May 15;328(7449):1197-1199 [FREE Full text] [doi: 10.1136/bmj.328.7449.1197] [Medline: 15142933]
55. de VN, Robert F, Penders J. Wireless Body Area Network for Sleep Staging. In: IEEE Biomedical Circuits and Systems Conference. 2007 Presented at: BIOCAS'07; November 27-30, 2007; Montreal, Quebec. [doi: 10.1109/biocas.2007.4463334]
56. Surantha N, Kusuma GP, Isa SM. Internet of Things for Sleep Quality Monitoring System: A Survey. In: 11th International Conference on Knowledge, Information and Creativity Support Systems. 2016 Presented at: KICSS'16; November 10-12, 2016; Yogyakarta, Indonesia. [doi: 10.1109/kicss.2016.7951426]
57. van de Water AT, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography--a systematic review. J Sleep Res 2011 Mar;20(1 Pt 2):183-200 [FREE Full text] [doi: 10.1111/j.1365-2869.2009.00814.x] [Medline: 20374444]
58. Schomer DL, de Silva HL. Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. New York, USA: Lippincott Williams & Wilkins; 2005.
59. Chokroverty S. Atlas of Sleep Medicine. New York, USA: CRC Press; 2010.
60. Khalighi S, Sousa T, Pires G, Nunes U. Automatic sleep staging: a computer assisted approach for optimal combination of features and polysomnographic channels. Expert Systems with Applications 2013 Dec;40(17):7046-7059. [doi: 10.1016/j.eswa.2013.06.023]
61. Vaughn B, Quint S, Messenheimer J, Robertson K. Heart period variability in sleep. Clin Neurophysiol 1995 Mar;94(3):155-162. [doi: 10.1016/0013-4694(94)00270-u]
62. Townsend RE, Johnson LC, Naitoh P, Muzet AG. Heart rate preceding motility in sleep. Psychophysiology 1975 Mar;12(2):217-219. [doi: 10.1111/j.1469-8986.1975.tb01280.x] [Medline: 1135356]
63. Wu T, Wu F, Redoute J, Yuce MR. An autonomous wireless body area network implementation towards IoT connected healthcare applications. IEEE Access 2017;5:11413-11422. [doi: 10.1109/access.2017.2716344]
64. Rathee D, Rangi S, Chakarvarti SK, Singh VR. Recent trends in wireless body area network (WBAN) research and cognition based adaptive WBAN architecture for healthcare. Health Technol 2014 May 24;4(3):239-244. [doi: 10.1007/s12553-014-0083-x]
65. Milici S, Lazaro A, Villarino R, Girbau D, Magnarosa M. Wireless wearable magnetometer-based sensor for sleep quality monitoring. IEEE Sensors J 2018 Mar 1;18(5):2145-2152. [doi: 10.1109/jsen.2018.2791400]
66. Younes M, Soiferman M, Thompson W, Giannouli E. Performance of a new portable wireless sleep monitor. J Clin Sleep Med 2017 Feb 15;13(2):245-258 [FREE Full text] [doi: 10.5664/jcsm.6456] [Medline: 27784419]
67. Finan PH, Richards JM, Gamaldo CE, Han D, Leoutsakos JM, Salas R, et al. Validation of a wireless, self-application, ambulatory electroencephalographic sleep monitoring device in healthy volunteers. J Clin Sleep Med 2016 Nov 15;12(11):1443-1451 [FREE Full text] [doi: 10.5664/jcsm.6262] [Medline: 27707438]
68. Velicu OR, Madrid NM, Seepold R. Experimental Sleep Phases Monitoring. In: International Conference on Biomedical and Health Informatics. 2016 Presented at: BHI'16; February 24-27, 2016; Las Vegas, NV, USA. [doi: 10.1109/bhi.2016.7455976]

69. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. Sleep Med 2001 Sep;2(5):389-396. [doi: 10.1016/s1389-9457(00)00098-8] [Medline: 14592388]
70. Kalkbrenner C, Eichenlaub M, Brucher R. Development of a new homecare sleep monitor using body sounds and motion tracking. Curr Dir Biomed Eng 2015;1(1):30-33. [doi: 10.1515/cdbme-2015-0008]
71. Nakano H, Hayashi M, Ohshima E, Nishikata N, Shinohara T. Validation of a new system of tracheal sound analysis for the diagnosis of sleep apnea-hypopnea syndrome. Sleep 2004 Aug 1;27(5):951-957. [doi: 10.1093/sleep/27.5.951] [Medline: 15453554]
72. Yadollahi A, Giannouli E, Moussavi Z. Sleep apnea monitoring and diagnosis based on pulse oximetry and tracheal sound signals. Med Biol Eng Comput 2010 Nov;48(11):1087-1097. [doi: 10.1007/s11517-010-0674-2] [Medline: 20734154]
73. Kalkbrenner C, Brucher R, Kesztyüs T, Eichenlaub M, Rottbauer W, Scharnbeck D. Automated sleep stage classification based on tracheal body sound and actigraphy. Ger Med Sci 2019;17:Doc02 [FREE Full text] [doi: 10.3205/000268] [Medline: 30996721]
74. Lee S, Yan L, Roh T. The Smart Patches and Wearable Band (W-Band) for Comfortable Sleep Monitoring System. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011 Presented at: EMBC'11; August 30- September 3, 2011; Boston, USA. [doi: 10.1109/iembs.2011.6091741]
75. Shambroom JR, Fábregas SE, Johnstone J. Validation of an automated wireless system to monitor sleep in healthy adults. J Sleep Res 2012 Apr;21(2):221-230 [FREE Full text] [doi: 10.1111/j.1365-2869.2011.00944.x] [Medline: 21859438]
76. Suzuki T, Ouchi K, Kameyama K. Development of a Sleep Monitoring System with Wearable Vital Sensor for Home Use. Science and Technology Publications. 2009. URL: https://www.scitepress.org/papers/2009/17842/pdf/index.html [accessed 2020-08-13]
77. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992 Oct;15(5):461-469. [doi: 10.1093/sleep/15.5.461] [Medline: 1455130]
78. Baharav A, Kotagal S, Gibbons V, Rubin BK, Pratt G, Karin J, et al. Fluctuations in autonomic nervous activity during sleep displayed by power spectrum analysis of heart rate variability. Neurology 1995 Jun;45(6):1183-1187. [doi: 10.1212/wnl.45.6.1183] [Medline: 7783886]
79. Berry RB, Brooks R, Gamaldo CE, Harding SM, Lloyd RM. The AASM Manual for the Scoring of Sleep and Associated Events. American Academy of Sleep Medicine. 2015. URL: https://aasm.org/clinical-resources/scoring-manual/ [accessed 2020-08-13]
80. Tataraidze A, Anishchenko L, Korostovtseva L, Kooij BJ, Bochkarev M, Sviryaev Y. Sleep stage classification based on respiratory signal. Conf Proc IEEE Eng Med Biol Soc 2015;2015:358-361. [doi: 10.1109/EMBC.2015.7318373] [Medline: 26736273]
81. Jawbone Up: Fitness Tracker Review. Live Science. 2013. URL: https://www.livescience.com/40107-jawbone-up-review.html [accessed 2013-10-02]
82. FitBit. 2020 Aug 11. URL: https://www.fitbit.com/fr/store?utm_source=&utm_medium=paidsearch&gclid=Cj0KCQjwv-DaBRCcARIsAI9sba9Dh12KY1zqUWJ1WLOZbS5BDheFt-ITkH3s-tE-652MtxsDMvoP3LgaAkgWEALw_wcB&dclid=CLevqriyutwCFcriGwodGHQL0Q [accessed 2020-08-11]
83. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit charge 2 compared with polysomnography in adults. Chronobiol Int 2018 Apr;35(4):465-476. [doi: 10.1080/07420528.2017.1413578] [Medline: 29235907]
84. Sleepace. 2020. URL: http://www.sleepace.com/en/reston.html?category=reston [accessed 2020-08-11]
85. Sleep Dot B501. 2020. URL: http://www.sleepace.com/en/dot.html?category=dot [accessed 2020-08-11]
86. Withings Aura Sleep System. Sleep Trackers. 2020. URL: https://sleeptrackers.io/withings-aura/ [accessed 2020-08-11]
87. Chan M, Campo E, Estève D. Assessment of activity of elderly people using a home monitoring system. Int J Rehabil Res 2005 Mar;28(1):69-76. [doi: 10.1097/00004356-200503000-00010] [Medline: 15729100]
88. Campo E, Bonhomme S, Chan M. Remote Tracking Patients in Retirement Home Using Wireless Multisensor System. In: 12th IEEE International Conference on e-Health Networking, Applications and Services. 2010 Presented at: HEALTH'10; July 1-3, 2010; Lyon, France. [doi: 10.1109/health.2010.5556567]
89. Charlon Y, Bourennane W, Bettahar F, Campo E. Activity monitoring system for elderly in a context of smart home. IRBM 2013 Feb;34(1):60-63. [doi: 10.1016/j.irbm.2012.12.014]
90. Charlon Y, Campo E, Brulin D. Design and evaluation of a smart insole: application for continuous monitoring of frail people at home. Expert Syst Appl 2018 Apr;95:57-71. [doi: 10.1016/j.eswa.2017.11.024]
91. Charlon Y, Fourty N, Campo E. A telemetry system embedded in clothes for indoor localization and elderly health monitoring. Sensors (Basel) 2013 Sep 4;13(9):11728-11749 [FREE Full text] [doi: 10.3390/s130911728] [Medline: 24008286]
92. Saad WH, Khoo CW, Ab Rahman SI, Ibrahim MM, Saad NH. Development of sleep monitoring system for observing the effect of the room ambient toward the quality of sleep. In: IOP Conference Series: Materials Science and Engineering. 2017 Presented at: IOP'17; June 12-16, 2017; Prague, Czech Republic. [doi: 10.1088/1757-899x/210/1/012050]
93. Webb WB, Agnew HW. Sleep stage characteristics of long and short sleepers. Science 1970 Apr 3;168(3927):146-147. [doi: 10.1126/science.168.3927.146] [Medline: 4313684]

94. Purves D, Augustine G, Fitzpatrick D. Physiological Changes in Sleep States. Second Edition. Washington, DC, USA: Neuroscience; 2001.
95. 2015 Alaska Sleep Education Center. Improve Your Sleep. Improve Your Life. 2020 Aug 11. URL: http://www.alaskasleep.com/ [accessed 2020-08-11]
96. Suzuki T, Tanaka H, Minami S. Wearable Wireless Vital Monitoring Technology for Smart Health Care. In: 7th International Symposium on Medical Information and Communication Technology. 2013 Presented at: ISMICT'13; March 6-8, 2013; Tokyo, Japan. [doi: 10.1109/ismict.2013.6521687]
97. Chan AM, Selvaraj N, Ferdosi N. Wireless Patch Sensor for Remote Monitoring of Heart Rate, Respiration, Activity, and Falls. In: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2013 Presented at: EMBC'13; July 3-7, 2013; Osaka, Japan. [doi: 10.1109/embc.2013.6610948]
98. Sadek I, Seet E, Biswas J, Abdulrazak B, Mokhtari M. Nonintrusive vital signs monitoring for sleep apnea patients: a preliminary study. IEEE Access 2018;6:2506-2514. [doi: 10.1109/access.2017.2783939]
99. Sadek I, Biswas J, Abdulrazak B. Continuous and Unconstrained Vital Signs Monitoring With Ballistocardiogram Sensors in Headrest Position. In: International Conference on Biomedical & Health Informatics. 2017 Presented at: BHI'17; February 16-19, 2017; Orlando, Florida, USA. [doi: 10.1109/bhi.2017.7897262]
100. Seba A, Istrate D, Guettari T, Ugon A, Pinna A, Garda P. Thermal-signature-based sleep analysis sensor. Informatics 2017 Oct 28;4(4):37. [doi: 10.3390/informatics4040037]
101. Sadek I, Biswas J, Yongwei Z. Sensor Data Quality Processing for Vital Signs With Opportunistic Ambient Sensing. In: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2016 Presented at: EMBC'16; August 16-20, 2016; Orlando, FL, USA. [doi: 10.1109/embc.2016.7591234]
102. Lee WK, Yoon H, Park KS. Smart ECG monitoring patch with built-in R-peak detection for long-term HRV analysis. Ann Biomed Eng 2016 Jul;44(7):2292-2301. [doi: 10.1007/s10439-015-1502-5] [Medline: 26558395]
103. Samy L, Huang M, Liu JJ, Xu W, Sarrafzadeh M. Unobtrusive sleep stage identification using a pressure-sensitive bed sheet. IEEE Sensors J 2014 Jul;14(7):2092-2101. [doi: 10.1109/jsen.2013.2293917]
104. Lee L, Lo Y, Yu J, Lee G, Ni Y, Chen N, et al. Snoring sounds predict obstruction sites and surgical response in patients with obstructive sleep apnea hypopnea syndrome. Sci Rep 2016 Jul 29;6:30629 [FREE Full text] [doi: 10.1038/srep30629] [Medline: 27471038]
105. Hou L, Wang C, Zhang C. Calculating AHI Combine Oximetery and Snore Sound. In: China Summit and International Conference on Signal and Information Processing. 2015 Presented at: ChinaSIP'15; July 12-15, 2015; Chengdu, China. [doi: 10.1109/chinasip.2015.7230364]
106. Emoto T, Abeyratne UR, Kawano K, Okada T, Jinnouchi O, Kawata I. Detection of sleep breathing sound based on artificial neural network analysis. Biomed Signal Proces 2018 Mar;41:81-89. [doi: 10.1016/j.bspc.2017.11.005]
107. Sadek I, Biswas J, Fook VF. Automatic Heart Rate Detection From FBG Sensors Using Sensor Fusion and Enhanced Empirical Mode Decomposition. In: International Symposium on Signal Processing and Information Technology. 2015 Presented at: ISSPIT'15; December 7-19, 2015; Abu Dhabi, United Arab Emirates. [doi: 10.1109/isspit.2015.7394358]
108. Beattie Z, Oyang Y, Statan A, Ghoreyshi A, Pantelopoulos A, Russell A, et al. Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals. Physiol Meas 2017 Oct 31;38(11):1968-1979. [doi: 10.1088/1361-6579/aa9047] [Medline: 29087960]
109. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond A 1998 Mar 8;454(1971):903-995. [doi: 10.1098/rspa.1998.0193]
110. Charleston-Villalobos S, González-Camarena R, Chi-Lem G, Aljama-Corrales T. Crackle sounds analysis by empirical mode decomposition. Nonlinear and nonstationary signal analysis for distinction of crackles in lung sounds. IEEE Eng Med Biol Mag 2007;26(1):40-47. [doi: 10.1109/memb.2007.289120] [Medline: 17278771]
111. Balocchi R, Menicucci D, Santarcangelo E, Sebastiani L, Gemignani A, Ghelarducci B, et al. Deriving the respiratory sinus arrhythmia from the heartbeat time series using empirical mode decomposition. Chaos Soliton Fract 2004 Apr;20(1):171-177. [doi: 10.1016/s0960-0779(03)00441-7]
112. van Marken Lichtenbelt WD, Daanen HA, Wouters L, Fronczek R, Raymann RJ, Severens NM, et al. Evaluation of wireless determination of skin temperature using iButtons. Physiol Behav 2006 Jul 30;88(4-5):489-497. [doi: 10.1016/j.physbeh.2006.04.026] [Medline: 16797616]
113. van Someren EJ. Mechanisms and functions of coupling between sleep and temperature rhythms. Prog Brain Res 2006;153:309-324. [doi: 10.1016/S0079-6123(06)53018-3] [Medline: 16876583]
114. Kräuchi K, Deboer T. Body temperatures, sleep, and hibernation. Principl Pract Sleep Med 2011:323-334. [doi: 10.1016/b978-1-4160-6645-3.00028-1]
115. Keogh E, Lin J, Lee SH, Herle H. Finding the most unusual time series subsequence: algorithms and applications. Knowl Inf Syst 2006 Nov 23;11(1):1-27. [doi: 10.1007/s10115-006-0034-6]
116. van Hulle MM. Self-organizing maps. In: Handbook of Natural Computing. Berlin, Heidelberg: Springer; 2012:585-622.

117. Lafferty J, McCallum A, Pereira FC. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. University of Pennsylvania ScholarlyCommons. 2001. URL: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1162&context=cis_papers [accessed 2020-08-13]
118. Schuster-Böckler B, Bateman A. An introduction to hidden Markov models. Curr Protoc Bioinformatics 2007 Jun;Appendix 3:Appendix 3A. [doi: 10.1002/0471250953.bia03as18] [Medline: 18428778]
119. Zeo Sleep Manager Pro: Usage Tips. Gibson Research Corporation. 2016. URL: https://www.grc.com/zeo.htm [accessed 2016-05-10]
120. Carpenter JS, Andrykowski MA. Psychometric evaluation of the Pittsburgh sleep quality index. J Psychosom Res 1998 Jul;45(1):5-13. [doi: 10.1016/s0022-3999(97)00298-5] [Medline: 9720850]
121. Krishna A, Mallick M, Mitra B. SleepSensei: An Automated Sleep Quality Monitor and Sleep Duration Estimator. In: Proceedings of the First Workshop on IoT-enabled Healthcare and Wellness Technologies and Systems. 2016 Presented at: IoT of Health'16; June 25-30, 2016; Singapore,. [doi: 10.1145/2933566.2933570]
122. Kay M, Choe EK, Shepherd J. Lullaby: a Capture & Access System for Understanding the Sleep Environment. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. 2012 Presented at: UbiComp'12; September 5-8, 2012; Pittsburgh, USA. [doi: 10.1145/2370216.2370253]

Abbreviations

AASM: American Academy of Sleep Medicine
BCG: ballistocardiography
CRF: conditional random field
DAN: disorders of arousal from NREM
ECG: electrocardiography
EEG: electroencephalography
EMD: empirical mode decomposition
EMG: electromyography
EOG: electrooculography
FFT: fast Fourier transformation
GDP: gross domestic product
HF: high frequency
ICS: intercostal space
ICSD-3: International Classification of Sleep Disorders
LAAS-CNRS: Laboratory for Analysis and Architecture of Systems-French National Centre for Scientific Research
LD: linear discriminant
LDA: linear discriminant analysis
LF: low frequency
NC: network controller
NREM: nonrapid eye movement
OSA: obstructive sleep apnea
PLMS: periodic limb movement during sleep
PPI: pulse to pulse interval
PSG: polysomnography
RBD: REM sleep behavior disorder
REM: rapid eye movement
RF: random forest
RLS: restless legs syndrome
SAX: symbolic aggregate approximation
SOM: self-organized map
SVM: support vector machine
SWS: slow wave sleep
VLF: very low frequency
WBAN: wireless body area network
W-band: wearable bands
WLBAN: wireless local body area network
WSN: wireless sensor network

Edited by G Eysenbach; submitted 01.06.20; peer-reviewed by LA Lee, E Kutafina; comments to author 25.06.20; revised version received 05.07.20; accepted 22.07.20; published 26.08.20.

Please cite as:
Pan Q, Brulin D, Campo E
Current Status and Future Challenges of Sleep Monitoring Systems: Systematic Review
JMIR Biomed Eng 2020;5(1):e20921
URL: http://biomedeng.jmir.org/2020/1/e20921/
doi:10.2196/20921
PMID:

Original Paper

Usability and Practicality of a Novel Mobile Attachment for Aural Endoscopy (endoscope-i): Formative Usability Study

Rowena Williams[1*], BSc, MSc; Jonathan Daw Ern Lee[1*], MBChB; Jameel Muzaffar[1], MSc, FRCS; Tom Clutton-Brock[2], MSc, FRCS; Chris Coulson[1], PhD, FRCS

[1]Queen Elizabeth Hospital Birmingham, Birmingham, United Kingdom

[2]Medical Devices Testing and Evaluation Centre, Birmingham, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Chris Coulson, PhD, FRCS
Queen Elizabeth Hospital Birmingham
Mindelsohn Way, Birmingham B15 2TH
Birmingham
United Kingdom
Phone: 44 1213712000 ext 14512
Email: Christopher.Coulson@uhb.nhs.uk

Abstract

Background: Our aims were to determine the usability and practicality of the endoscope-i system, a novel mobile attachment for aural endoscopy. This incorporated assessing the ease of use of the endoscope-i for different professionals, and ultimately improving the system by receiving constructive feedback.

Objective: Our objectives were to assess the ease of the endoscope-i system in conducting an aural examination and to assess its feasibility for integrating its use into clinical practice. We looked to assess its ease, effectiveness, and efficiency; to compare this to current practices with otoscopes; and to determine whether participants perceived the system to be able to produce an image of sufficient quality to make a clinical assessment. Finally, we wanted to assess the usefulness of the current training given for using the system, and we sought to gain feedback for the product from the differing specialists.

Methods: A formative usability study of the endoscope-i system was conducted with 5 health care professionals. Each session lasted 40 minutes and involved audio/video consent, a hands-on session, a private semistructured interview, and an option to discuss the device with a company representative.

Results: All participants found the endoscope-i system easy to use. The image quality was perceived to be greater than that achieved by current otoscopes. The ability to record images and view them retrospectively was also seen as a positive.

Conclusions: This study has not identified any significant issues relating to the design, functionality, or application of the endoscope-i. Participants perceived the system as superior to current options with a directly positive impact on their clinical practice.

(JMIR Biomed Eng 2020;5(1):e18850) doi:10.2196/18850

KEYWORDS

otology; endoscopy; smartphone; telemedicine; usability studies

Introduction

Background

Otological diseases make up a large proportion of global disease. In 2015, the Global Burden of Disease Project ranked otitis media as the third-most common short-term disease with an incidence of 471 million worldwide [1]. Furthermore, hearing loss was ranked the fourth-most common chronic disease, affecting over 1 billion people. More recently, in 2017, otitis media had a reported incidence of 318 million cases [2].

The World Health Organization has noted a discrepancy between burden of disease and current resources for many specialties including otology, encouraging advances in technology and telemedicine to help bridge this gap around the world [3]. Recently, with the continual rise of availability of smartphones, mobile apps and attachments have become a way to achieve this, with some apps showing promise for streamlining referrals

to ear, nose, and throat (ENT) specialists [4,5]. For example, Biagio showed substantial agreement in otological diagnoses between face-to-face assessments with an otoscope and remote viewing of videos of the findings, termed video-otoscopy [6]. This may help alleviate the aforementioned imbalance of resources by facilitating specialist expertise in remote areas. However, this is dependent on the speed of transfer of the video and the image quality [7]. The key aim of this study was to assess the usability and practicality of a novel otological technology, the endoscope-i, by gaining professional opinions on the effectiveness and comprehensiveness of the system when compared to the current standard, the otoscope (full aims are mentioned in the Test Objectives section). This study explores its potential for improving patient care [8].

The endoscope-i

Overview

During an interview conducted on June 14, 2019, the endoscope-i representative corroborated the following information. The endoscope-i system comprises an adapter designed for a smartphone to be attached to an endoscope, in order to look into a patient's ear with a wide-angle lens, using a short rigid scope. The system is already commercially available and 3500 have been sold internationally online. The system comes packaged with instructions for use, including an explanatory diagram. Previous trials have been undertaken in Staffordshire, UK, with primary care clinicians using a slightly modified version of the current app. A key group of targeted users of the endoscope-i are audiology staff, nurses, audiologists, general practitioners, health care assistants, and community nurses. Once trained in using the device, it is intended that these users will be able to send the image obtained to an ENT surgeon for assessment of the ear. Therefore, the operator only needs to be skilled in using the device, with appropriate anatomy and physiology knowledge, and not in aural clinical assessment. This would then negate the need for all aural patients to see an ENT surgeon in person and, in turn, could potentially reduce waiting lists.

Differences and Innovations When Compared to the Traditional Otoscope

Firstly, the endoscope-i is attached to a smartphone camera that can be used to view the external auditory canal and tympanic membrane, once it is attached to the scope via the adapter, thereby producing a better image than an otoscope. An app is then used to optimize the image. These high-quality images will also only improve further with the advance in technology and camera quality of mobile phones. An app is then used to optimize the image. Secondly, in current practice, the eye needs to be close to the ear in order to see through the otoscope. With the endoscope-i system, an eyepiece sits on the adapter to view the image, allowing greater distance between the patient and clinician for easier working practice and improved view. This shared image may also be beneficial in the training of medical professionals. In addition, the wide-angle lens of an endoscope produces a broader perspective of the tympanic membrane not afforded by conventional otoscopes.

Another strength of the endoscope-i is its ability to capture high-quality photos and videos. These can be shown to the patient for reassurance and explanation of pathology, or they can be used for a remote referral, such as a junior doctor sending it to a senior colleague or a general practitioner sending it to an ENT surgeon in a distant location. This may enable a specialist management plan to be given for a patient, which could in turn eliminate the need to see the ENT specialist in person, further reducing waiting lists.

Finally, the introduction of the endoscope-i introduces the possibility of simple interventions alongside vision in the future, similar to the mainstream use of endoscopes in major operations, such as in endoscopic tympanoplasty [9].

Cons of the endoscope-i

With a new instrument or device comes a learning curve and a need for learning a new technique. Training is required to utilize the app and adapter but is thought to be minimal, with practice allowing for familiarity with its use. This assumption is tested as one of the aims of this study. The larger part of the training is anticipated to be for the skill of oto-endoscopy itself, which was also assessed. Staff who are not specialized in ENT, audiology, or endoscopy will require initial training in endoscopy, which is more time-consuming than training in only the endoscope-i device.

Further drawbacks are the risks associated with using the device; these include a perforated eardrum at the severe end of the scale and pain on insertion of the endoscope. Significant complications are thought to be rare, with none reported thus far.

Test Objectives

The aims of this formative usability study were to assess the extent to which the endoscope-i could demonstrate effectiveness, efficiency, novel advances, and satisfaction in aural examination on patients attending an outpatient clinic as a part of the clinical assessment process, as per the International Organization for Standardization (ISO) standard, ISO 9241-11 [10]. This study is part of the iterative product development process, and it is different from a summative usability study, which would be conducted for validation purposes [11].

The specific objectives were to gain professional opinions on using the endoscope-i. Specifically, these objectives were as follows:

1. To assess its ability to conduct an aural examination: to be able to assemble the equipment, insert the endoscope into the ear, view the ear drum, and remove the endoscope.
2. To assess its feasibility for integrating its use into clinical practice.
3. To assess its ease and practicality over current practices with otoscopes.
4. To confirm or deny the ability of the device to produce an image of sufficient quality to make a clinical assessment.
5. To assess the usefulness of the current training given for using the device.
6. To gain insights into future recommendations for the product from the targeted specialists.

Methods

Ethics Approval

The project received institutional ethical approval. All participants consented to taking part in the usability study and did so voluntarily.

Participants

The study was advertised via emails and a poster circulated through the audiology and research and development departments within the University of Birmingham and the University Hospitals Birmingham National Health Service (NHS) Foundation Trust as well as through correspondence with professional contacts within the ENT department. All recruits were required to be trained in the skill of endoscopy or otoscopy in order for them to be able to compare it to their current practice. A total of 5 volunteers participated in this study, among which were a lead research nurse, a consultant ENT surgeon, an audiology clinical scientist, an endoscopy nurse, and the head of audiology; these volunteers represented a broad range of experience levels. All participants currently work for the University Hospitals Birmingham NHS Foundation Trust; their roles are listed in Table 1. Information on the endoscope-i was not revealed prior to the study, although 2 of the volunteers had previously seen and used the device through their own professional contact with the developer. The advertising email and flyer mentioned the scope of the study, including the task involved (eg, using an adapter for a smartphone to be attached to an endoscope in order to perform an aural examination).

Table 1. Participants' roles.

Participant number	Job title
EI1	Lead research nurse
EI2	Consultant ENT[a] surgeon
EI3	Endoscopy nurse
EI4	Audiology clinical scientist
EI5	Head of audiology

[a]ENT: ear, nose, and throat.

Tasks and Test Design

A formative usability study took place on June 14, 2019, in the Medical Device Testing and Evaluation Centre (MD-TEC) simulation suite at the Institute of Translational Medicine, Birmingham, on commission of the company. Results were collected, analyzed, and presented independent of the company. The project was fully funded by the European Regional Development Fund. A team of three planned and moderated the study. The study procedure was built around three core elements: users, user environment, and user experience of the device interface [12].

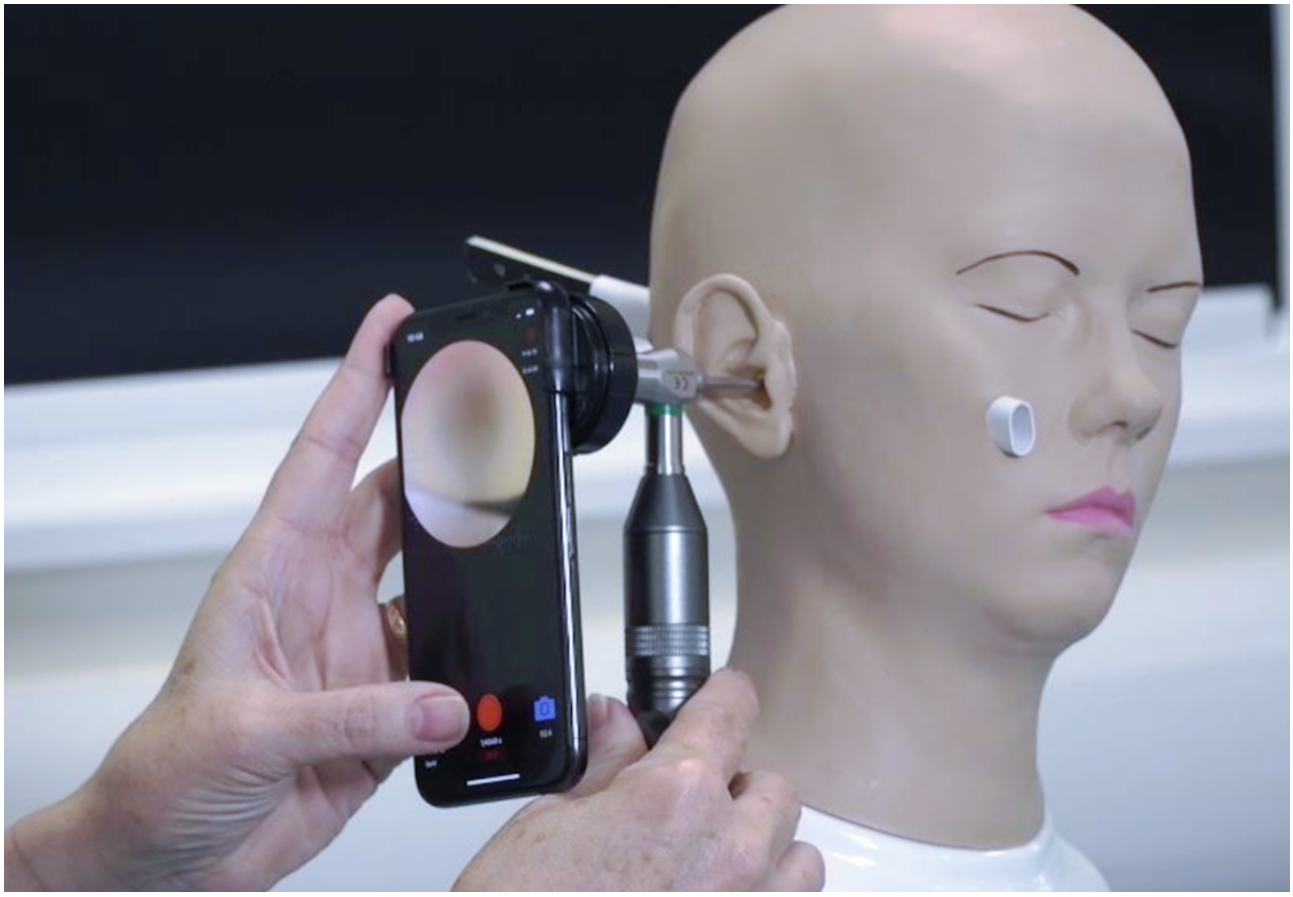
The conducted usability study was structured around a task-based scenario: mannequin heads with accessible ear canals designed for audiology training were used in a mock audiology clinic setting; this can be seen in Figure 1. An endoscope-i representative assisted on the day, giving a training session and a demonstration of the device. The participants were then presented with the device in its case and asked to use it as if they would in a clinical assessment of an ear canal.

The tasks involved in the hands-on session included the following:

1. Take the device out of the box and assemble it according to the instructions.
2. Set the device up to record.
3. Insert the device into the ear.
4. Confirm visual image of the ear canal on the phone.
5. Remove the device from the ear canal.
6. Utilize the app to send the image to a clinician.
7. Dissemble the device and replace it into the box.

For the scenario, the participant acted as the assessing clinician. The company representative was available for issues and queries during the hands-on session, guiding participants if they required, considering it being a formative study [12]. There was also a member of the MD-TEC team present to facilitate the study. The hands-on session was recorded by a high-definition camera on a tripod—with exception of participant 4 who wished not to be recorded—operated by a member of the NHS Foundation Trust's A/V (audio/video) team. Immediately after the session, a voice recorder was used for the interviews.

Figure 1. The endoscope-i in use on a mannequin head. Having a simple design, the product aims to be easy to assemble and use in various settings, including with outpatients, in general practitioner surgeries, and in the community.



Results

Usability Metrics

Task Completion Success Rate

All participants were able to complete the tasks of assembling the device, carrying out an aural endoscopy, obtaining an image, and dissembling the device; therefore, there was a completion rate of 100% (5/5) for the representative end users. Participants expressed that this was an excellent instrument to obtain a high-quality image that is superior to the current image quality from otoscopes. From the feedback given in this usability study, it is deemed that the relevant clinicians would easily and gratefully adopt this device.

There were no significant errors made during the task-based session of the usability study. On two occasions, it was necessary for the company representative to step in and assist the participant with an aspect of using the device (EI1 and EI4); however, following this, the procedure was then completed completely independently by the participants, suggesting a high error tolerance. Bearing in mind, one of these participants had no prior experience in aural assessment; this indicates that operating the device would be easy to learn for those already assessing ears in practice. Further usability testing would be required to assess the use of the device by lesser-trained health care professionals. All participants (5/5, 100%) were able to complete all seven tasks within 5 minutes, with an average of 3.5 minutes, suggesting it was efficient as well as easy to use.

Usability Overview

All the participants were able to successfully use the endoscope-i and gave positive responses, with comments such as “The instrument itself is excellent” and “Perfect for a busy clinic” (EI1) being made. It was described as “small, compact, clear” (EI1) and was recognized as a “different way of examining a patient,” as there is “no need to get as close [to the patient]” (EI2), but it was also thought of as not difficult to get used to (EI2 and EI4). All participants (5/5, 100%) showed a good awareness of the endoscope-i’s use and purpose, regardless of their previous experience levels of aural examinations, and all seemed keen to integrate it into practice. Their feedback is highly relevant, considering their previous use of alternatives and experience levels in clinical assessments, communication between departments, and transmission systems within the primary and secondary care infrastructure. Figure 1 shows the endoscope-i in use at the point of contact with the patient.

Ease of Use of the Device

Ease of use was specifically named as one of the positive aspects of the device by 2 out of 5 (40%) participants (EI3 and EI5), with Participant EI1 also complimenting the simplicity of putting the parts together. Additionally, all 5 (100%) participants found the endoscope-i easy to use, with phrases such as “easier than anticipated” (EI1) being used to describe it by a participant with no previous experience of the device or of aural examinations and “pretty straightforward” said by a participant who had used a version of it in the past. It is also important to mention that

Participant EI3 claimed it was "easy to do even with arthritic fingers."

Image Quality

One of the advantages of this device over the current practice of using an otoscope was reported as the high-quality image, which was highlighted by all (5/5, 100%) participants. One participant (EI2) described it as "vastly superior" and went on to explain that, in addition to the better quality, this was because the point of vision is next to the ear drum, and the device provides the ability to look around corners to see the full ear drum. They also pointed out that there is also the opportunity to show the image to the patient and the relatives. Storing the actual image was seen as a notable advantage by 4 out of 5 (80%) participants (EI1, EI2, EI4, and EI5) (ie, "the recording element is a good idea").

The important aspect of image quality was whether it would be of sufficient quality to make a clinical assessment from. The 4 participants out of 5 (80%) (EI2, EI3, EI4, and EI5) with experience in otoscopy or endoscopy positively confirmed that it would be of sufficient quality; the 1 (20%) participant (EI1) without experience said she guessed it would be, taking into account the vast experience she does have in clinical care. It was even described by one participant (EI4) as "far exceed[ing] current image quality." The image quality can be seen in Figure 1.

Comparison to Current Practice

In comparison to current practice, it was seen as different because, as Participant EI4 explained, there are "clear guidelines on safety and bracing against a patient's head, which is not able to be done with the endoscope-i." In addition, Participant EI5 stated that there is "a knack to using it which is different to an otoscope," which would mean that frequent use would be required to maintain an adequate skill level.

Training Required

The opinions on training needs for the endoscope-i varied between an ongoing competency document for the trainee to complete while being assessed by a competent user in a clinic (EI1), face-to-face group sessions (EI2 and EI4), and video teaching (EI3 and EI5). By mentioning that it would need to be used regularly to gain the specific technique, one participant's answer was in correlation to Participant EI1 suggesting a competency document. After the initial training, a degree of experience would be required, and the user would need to use the device regularly in order to become fully competent and comfortable with it, according to Participant EI5. Another participant (EI4) also posited that there could be a divide in the acceptance and ease of adaptation of learning between older staff and younger staff due to the difference in familiarity with technology.

For those who suggested group face-to-face training, the emphasis was on having the device in front of the trainees so they could be talked through the process, try it themselves, and have a trainer available for assistance (EI2 and EI4). Those staff who were already trained in endoscopy would require less training than those who were not, since, for those who were inexperienced, "training would be around endoscopy of the ear as a new skill" (EI2) (eg, passing an endoscope through the ear canal without damaging it), in addition to using the adapter and app. A review of anatomy would also be useful, as the user will be getting a totally different view of the ear drum (EI2). Table 2 summarizes the methods of training that were recommended.

Table 2. Summary of recommendations of training methods for the endoscope-i system.

Participant	Video	Face-to-face	Other
EI1	Nice to have, but not essential	N/A[a]	Competency document in practice
EI2	Videos to refer back to	Group session with demonstration	Review of anatomy Practical training in the skill of endoscopy
EI3	Video taught	N/A	N/A
EI4	e-learning video	Hands-on practical	N/A
EI5	Short videos	N/A	Ongoing for practice of endoscope-i

[a]N/A: not applicable; the participant did not have recommendations for this category.

iPhone Versus Android

With regard to the app only being compatible with iPhones, it was explored whether this would present an issue for some users. Out of the 5 participants, 2 (40%) (EI1 and EI3) did consider that this may be difficult and would cause some confusion to navigate initially, but Participant EI4 reported that no difficulties were encountered as an Android user.

Design Feedback

The participants were asked to comment on the positive and negative aspects of the device. These responses are expanded upon below and summarized in Table 3.

In terms of assembling the device, it was recorded to be "relatively straightforward" and very easy to put the light source and endoscope together and screw in the battery pack onto the endoscope. However, 2 out of 5 (40%) participants (EI1 and EI2) reported having difficulty with assembly, both assembling the device and the phone in a different order. Participant EI1 found it challenging to clip the adapter, with the endoscope already in place, onto the phone, and Participant EI2 had trouble attaching the endoscope onto the adapter that clips onto the phone. Participant EI2 explained "the weight of the battery makes it very awkward." Both made suggestions on how to resolve this: Participant EI1 laid the phone flat on the table to then screw it in place, and Participant EI2 attempted changing

the order of assembly to attach the heavy battery pack last. It was also stated that the screw-fix to attach the endoscope is fiddly, and Participant EI2 suggested that if the endoscope could clip in to the adapter that would be easier; however, it was recognized that perhaps the screw-fix is the most secure fixation to hold the weight of the phone and battery.

Table 3. Overview of likes and dislikes.

Participant	Likes	Dislikes
EI1	Clear image Simplicity of putting parts together	If the user is not used to the function of an iPhone, initially this may cause some confusion to navigate. An Android user may struggle initially. An iPhone user will not have any trouble at all.
EI2	Much better picture quality, as the point of vision is next to the ear drum Can look around corners so can see full ear drum. "Vastly superior image." Can show the patient the image. Relatives can view image in real time. Otoscope relies on a description or a picture drawn—the endoscope-i can store actual image of ear. Data storage is straightforward, once one is used to where images are stored. Images are anonymous, so can share.	Device: order of assembly and weight of battery pack Use of program: change recording and saving buttons to be accessible to thumb. Switch controls for ISO (International Organization for Standardization) and focus, so the focus is easier to access, irrespective of which side is holding the phone, whichever hand.
EI3	Ease of use Amazing picture	Concern would be information on a phone—solution would be to leave it at the hospital.
EI4	Clarity of picture Theory of being able to share photos	Use of the app is not intuitive. Having to flick between the screens and twist it upside down to activate. User position—[add a] handhold [to] it. Long sharp pointy end—lacking safety element which should be fine for experienced users, but not juniors. Cumbersome to hold compared to what she is used to.
EI5	Picture quality Ease of use	Practice and experience to find the best way to hold it, as it is more weighted than an otoscope.

For the software app itself, suggestions were made by Participant EI2 to change the recording and saving buttons so they are both accessible to the thumb, and to switch around the controls such as the focus button, in order for the focus to be easier to access, irrespective of which side the phone is being held on. The app was not found to be intuitive by 1 out of 5 (20%) participants (EI4), because the user is required to flick between the screens and twist it upside down to activate. This participant also found the user position or handhold challenging, as it was "cumbersome to hold" compared to current practice; this view was shared by Participant EI5, positing that it may require practice and experience to find the best way to hold it. Out of the 5 participants, 1 (20%) (EI4) had safety concerns about the distal end entering a patient's ear when used by junior or inexperienced staff. This could be addressed through the ongoing practical competency element of the training. A final query by a separate participant (EI3) was around storing confidential data in the form of images on the phone. However, the ability to view the images repeatedly after the clinical assessment and share them was also seen to be of benefit.

As a guide for further development with the user in mind, the participants were also encouraged to make usability and design recommendations, resulting in some useful design and handling ideas being proposed. These have been separated into *two separate categories* of *assembly and handling* and *interface recommendations* for ease of assessment; these are all listed in Textbox 1.

Textbox 1. Endoscope assembly and handling and interface recommendations.

Attachment of the endoscope:

- Reassess the connection system on the adapter to make it easier to put together (EI2).
- Use a lighter battery pack (EI2).
- Add a rubber or plastic part on the end of the endoscope to make it softer for entering the patient's ear (EI3).
- Enable the use of a speculum on the end for safety (EI4).
- Allow for the device to be held horizontally to more closely resemble the handling of an otoscope, meaning a smoother transition step (EI4).
- Have it come in a case where it can remain assembled, rather than assembling it before every use (eg, a box where it can stay in one piece) (EI5).
- Advise users of the importance of sterility and cleaning between patients—using antibacterial wipes and storing in the box (EI2).

Interface:

- Reconsider the location of the buttons on the screen, switching the position of focus (EI2); put the film button at the bottom of the screen and have the buttons all on one screen (EI4).
- Incorporate the shutter button on both sides of screen, or have it flip over to the other side of the screen, so it can be used more accessibly with both hands (EI2).
- Label the focus and light exposure buttons (EI5).
- Set up a function to see a printed-off image or see it on the electronic patient record (EI1).

The recommendation of adding the rubber bungs to the part entering the patient was offered by 2 out of 5 (40%) participants (EI3 and EI4), due to the consideration that there may be some anxiety attached to the use of an endoscope on a live patient for the first few times by junior staff. Using a speculum on the end of the endoscope was also suggested for the same reason (EI4).

Discussion

Principal Findings

This formative usability study conducted by MD-TEC revealed a high acceptance rate of the endoscope-i, including its concept and method of use. Based on the study, as well as participant feedback, the current design is suitable for the intended purpose and its targeted end users, although some adjustments to the user interface and assembly order of the device may be considered. Furthermore, all participants perceived the device as having superior image quality, inducing a direct positive impact on their usual clinical practice, and having potential to reduce waiting lists and implicitly improve patient safety.

Despite the number of volunteers to the study (N=5), the wide range of experience of the participants was deemed a strength. In particular, one participant was not familiar with aural assessment and had not seen the endoscope-i, two participants were familiar in otoscopy but had not seen the endoscope-i, and two participants were familiar with otoscopy and had previously used the endoscope-i. Those who were familiar with otoscopy could compare their own experience directly against the use of this device; the one that was not trained in otoscopy could be used to see how easy it was for a complete novice to learn to use the device. Regardless of prior experience, a 100% completion rate within 5 minutes (averaged at 3.5 minutes) suggests that the device is easy to learn and use. Another favorable aspect was in the setup of the testing scenario: the use of a mannequin ensured patient safety, the room was designed to mimic clinical practice, and the setting remained uniform across each testing session to reduce variables between participants. The use of different types of feedback (ie, verbal, written, and video) gave a greater level of insight into the device's usability, the suggestions of which are mentioned in the concluding remarks. The videos, for example, revealed a learning curve where participants initially fumbled to assemble the device, yet soon figured out how to complete the task. In this study, these participants had live training preceding their test; however, other users apart from this study will likely not have this luxury. Therefore, as 80% (4/5) of the participants agreed, a written, annotated, and/or video guide would be extremely helpful for overcoming this learning curve, to use as a reference when using the device; note that this aforementioned reference was not present during the live test for participants in this study, which could have contributed to the fumbling.

Nevertheless, in spite of realistically replicating the use scenario, there are some well-acknowledged limitations of usability studies, such as "testing is always an artificial situation"; personal preferences are only those of the participants and are not representative of the entire user population [13]. In addition, the use of retrospective think-aloud interviews were a good asset to reveal further insights into the usability of the device.

Conclusions

The endoscope-i is a novel device and system that allows the user to take high-quality videos of the eardrum and canal with an improved angle compared to the current otoscope. This study shows that it would enable an aural assessment to be carried out from the image it produces, with the feature to save and export the image allowing for remote examination of patients. This may reduce patient waiting lists to see a specialist in person. On usability testing, the study further found the device to be efficient, effective, error tolerant, and easy to learn. This study did find that further exploration in the following areas might help with the user's engagement:

1. Explore different options of assembling the device parts and the possibility of a lighter battery pack to assist with this.
2. Consider locations of buttons on the screen in the app interface.
3. Evaluate the possibility of a feature to upload the images directly to the patient notes on an electronic patient record or to print for paper notes.

Acknowledgments

This usability study was fully funded by the European Regional Development Fund.

Conflicts of Interest

CC is Managing Director of endoscope-i Ltd and was not involved in the data collection or analysis. The remaining authors declare no conflict of interest.

References

1. Saunders JE, Rankin Z, Noonan KY. Otolaryngology and the global burden of disease. Otolaryngol Clin North Am 2018 Jun;51(3):515-534. [doi: 10.1016/j.otc.2018.01.016] [Medline: 29773124]
2. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
3. Telemedicine: Opportunities and Developments in Member States. Report on the Second Global Survey on eHealth. Geneva, Switzerland: World Health Organization; 2010. URL: https://www.who.int/goe/publications/goe_telemedicine_2010.pdf [accessed 2019-12-05]
4. Casale M, Costantino A, Rinaldi V, Forte A, Grimaldi M, Sabatino L, et al. Mobile applications in otolaryngology for patients: An update. Laryngoscope Investig Otolaryngol 2018 Dec;3(6):434-438 [FREE Full text] [doi: 10.1002/lio2.201] [Medline: 30599026]
5. Erkkola-Anttinen N, Irjala H, Laine MK, Tähtinen PA, Löyttyniemi E, Ruohola A. Smartphone otoscopy performed by parents. Telemed J E Health 2019 Jun;25(6):477-484. [doi: 10.1089/tmj.2018.0062] [Medline: 30040525]
6. Biagio L, Swanepoel DW, Laurent C, Lundberg T. Video-otoscopy recordings for diagnosis of childhood ear disease using telehealth at primary health care level. J Telemed Telecare 2014 Jun 23;20(6):300-306. [doi: 10.1177/1357633x14541038]
7. Chowdhury A, Hafeez-Baig A, Gururajan R, McCubbin A, Akmal Sharif M. Image quality in telehealth: Challenges in developing countries. Research Square (preprint) 2019 Dec 18:1-19 [FREE Full text] [doi: 10.21203/rs.2.19114/v1]
8. Mistry N, Coulson C, George A. endoscope-i: An innovation in mobile endoscopic technology transforming the delivery of patient care in otolaryngology. Expert Rev Med Devices 2017 Nov;14(11):913-918. [doi: 10.1080/17434440.2017.1386548] [Medline: 28972409]
9. Tseng C, Lai M, Wu C, Yuan S, Ding Y. Comparison of the efficacy of endoscopic tympanoplasty and microscopic tympanoplasty: A systematic review and meta-analysis. Laryngoscope 2017 Aug;127(8):1890-1896. [doi: 10.1002/lary.26379] [Medline: 27861950]
10. ISO 9241-11:1998(en) Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability. International Organization for Standardization (ISO). 1998. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en [accessed 2020-09-07]
11. US Food and Drug Administration. Applying Human Factors and Usability Engineering to Medical Devices: Guidance for Industry and Food and Drug Administration Staff. Rockville, MD: US Food and Drug Administration; 2016 Feb 03. URL: https://www.fda.gov/media/80481/download [accessed 2019-10-01]
12. North B. The Growing Role of Human Factors and Usability Engineering for Medical Devices: What's Required in the New Regulatory Landscape. London, UK: BSI Standards Ltd; 2015. URL: https://www.bsigroup.com/LocalFiles/de-de/Medizinprodukte/Growing-role-of-human-factors.pdf [accessed 2019-10-01]
13. Dicks RS. Mis-usability: On the uses and misuses of usability testing. In: Proceedings of the 20th Annual International Conference on Computer Documentation (SIGDOC '02). 2002 Presented at: 20th Annual International Conference on Computer Documentation (SIGDOC '02); October 20-23, 2002; Toronto, Canada p. 26-30. [doi: 10.1145/584955.584960]

Abbreviations

A/V: audio/video
ENT: ear, nose, and throat
ISO: International Organization for Standardization
MD-TEC: Medical Device Testing and Evaluation Centre

NHS: National Health Service

Edited by G Eysenbach; submitted 23.03.20; peer-reviewed by C Falls, M Ghozali, JR Bautista; comments to author 12.06.20; revised version received 28.06.20; accepted 16.08.20; published 18.09.20.

Please cite as:
Williams R, Lee JDE, Muzaffar J, Clutton-Brock T, Coulson C
Usability and Practicality of a Novel Mobile Attachment for Aural Endoscopy (endoscope-i): Formative Usability Study
JMIR Biomed Eng 2020;5(1):e18850
URL: http://biomedeng.jmir.org/2020/1/e18850/
doi:10.2196/18850
PMID:

Viewpoint

Dementia-Related Products on an e-Commerce Platform

Benjamin K P Woo[1], MD

University of California, Los Angeles, Sylmar, CA, United States

Corresponding Author:
Benjamin K P Woo, MD
University of California, Los Angeles
14445 Olive View Drive
Sylmar, CA, 91342
United States
Phone: 1 747 210 3830
Email: bkpwoo@gmail.com

Abstract

Dementia is a neurocognitive disorder, which affects older adults. There are currently no medication treatments available to cure dementia, but a number of biomedical technologies could be useful in assisting patients with dementia. With the continued growth of electronic commerce (e-commerce), online shopping for aging and health-related products will only continue to increase. Using the Tmall marketplace as an example, the purpose of this viewpoint is to describe the current trends of dementia-related products and devices available on an e-commerce platform. Feedback and critiques in the form of consumer reviews should be used to improve the design of dementia-related products. Online medical product consumers, however, must be vigilant about the effectiveness and risks of these biomedical devices.

(JMIR Biomed Eng 2020;5(1):e17514) doi:10.2196/17514

KEYWORDS

consumer review comments; dementia; e-commerce; biomed; technology; devices; biomedical engineering

Electronic commerce (e-commerce) has allowed patients and their caregivers to directly buy health products and medical devices from a seller over the internet. The largest of these online corporations include Alibaba, Amazon, and eBay. Tmall.com, a Chinese-language business-to-consumer (B2C) website operated by Alibaba in China, has approximately 260 million monthly visits from potential consumers and ranks number 9 among all websites in the e-commerce and shopping category [1]. A 2015 report from the Chinese government highlighted that while 59% of items sold online were genuine or of good quality, at least 40% of online goods sold were either counterfeits or of bad quality [2]. Tmall.com may eventually offer a reliable marketplace for consumers to purchase authentic medical products and devices; however, the sheer number of sellers on the B2C platform makes eliminating fake or counterfeit health products or biomedical devices from the Tmall marketplace difficult.

Dementia—also known as neurocognitive disorder—involves a constellation of symptoms including apathy, behavioral changes, confusion, and the impairment of executive functions [3-5]. A systematic review found that the estimated number of people with dementia in China was 9.2 million in 2010 [6]. There are currently no medication treatments available to cure dementia or to alter its progressive course [7-9]. However, there are a number of biomedical engineering technologies that can assist patients with dementia. For example, GPS locator devices could alert family members when someone with dementia wanders away from home [10]. With the continued growth of e-commerce and the worldwide popularity of online shopping for aging and health-related products, all biomedical engineers and health professionals should be prepared to further understand such trends. Therefore, the keyword dementia ( ) was searched on Tmall.com to provide a viewpoint on these trends of online shopping for dementia-related products.

On Tmall, an online marketplace for Chinese-language consumers at home and abroad to make purchases, there were a total of 1038 dementia-related products identified on November 29, 2019. Among these dementia-related items, 788 (75.92%) were books, 137 (13.20%) were medications or supplements, 65 (6.26%) were bracelets or wristbands, 42 (4.05%) were dementia toys, and 6 (0.57%) were biomedical devices. Of the 6 biomedical technologies, 4 were GPS-tracking Smart Watches for older adults, and 2 were transcranial magnetic stimulators.

There were no reviewer comments for the 2 transcranial magnetic stimulators; however, the consumer ratings for the 4 Smart Watches were relatively positive (4.8 out of 5.0). Some

reviewer comments (translated from Chinese to English) were included to understand how older Chinese adults may feel about technology for dementia. A reviewer wrote, "Very suitable for elderly with visual impairments and falls… Voice control is necessary [for elderly]." Another reviewer commented, "Easy to use, can make phone calls… but GPS can miss by 5 meters indoor." An older adult reviewer stated, "My children will know my whereabouts, and I feel more emotionally secured because I know my children are on the other side of the watch."

On the other hand, using online shopping for buying biomedical devices is not without risks. A recent small study found that transcranial magnetic stimulation (TMS) may temporarily improve memory; however, the study did not examine which areas of the brain should be targeted with stimulation nor how effective the treatment could be [11]. In addition, the Food and Drug Administration (FDA) of the United States has recently rejected the first TMS device to be used for treating symptoms of dementia, as the clinical trial for this device failed to demonstrate efficacy [12]. This particular TMS device is registered and classified as a Class II medical device by China's National Medical Products Administration (NMPA); however, the Tmall marketplace failed to point out that no assessment concerning the safety and efficacy of this device has been conducted for patients with dementia. Online shopping from the Tmall marketplace may become commonplace, but consumers and patients from China and abroad must be vigilant about the safety and effectiveness of these non-FDA or non-NMPA approved biomedical devices. The Tmall marketplace must ensure sellers are not overstating the medical device benefits and, more importantly, use algorithms to get rid of fake or unregulated health products targeting vulnerable older adults and their caregivers.

Future work should examine how to incorporate online consumer review comments on dementia-related products, especially feedback and critiques from older adult users, to improve the design of biomedical engineering technologies. Excellence in design may only be achieved when biomedical engineers incorporate the needs of all stakeholders, that is dementia patients, their caregivers, and their families. Reviewer comments may serve as a proxy for quality improvement. However, in addition to the likelihood of purchasing online biomedical engineering products that could be fraudulent or counterfeit, this viewpoint highlights the issues of safety and effectiveness of dementia-related products purchased from the Tmall marketplace. Online medical product consumers need to look beyond the cost savings and focus on the effectiveness and risks of such medical treatments. This viewpoint demonstrates how widespread these dementia products and medical devices are in a Chinese-language e-commerce marketplace, but a limitation to keep in mind is the possibility of fake reviews, fake comments, or even fake transactions on the Tmall marketplace. Nevertheless, Alibaba has used machine learning to lower search rankings for stores engaged in false transactions. In addition, in 2017, the Chinese government sentenced a merchant associated with generating fake reviews and transactions to approximately 6 years in prison [13].

Conflicts of Interest

None declared.

References

1. SimilarWeb. Top sites ranking for e-commerce And shopping > marketplace in the world URL: https://www.similarweb.com/top-websites/category/e-commerce-and-shopping/marketplace
2. Jourdan A. Reuters. 2015 Nov 02. Over 40 percent of China's online sales counterfeit, shoddy: Xinhua URL: https://tinyurl.com/rvc4waa
3. Woo BKP, Harwood DG, Melrose RJ, Mandelkern MA, Campa OM, Walston A, et al. Executive deficits and regional brain metabolism in Alzheimer's disease. Int J Geriatr Psychiatry 2010 Nov;25(11):1150-1158. [doi: 10.1002/gps.2452] [Medline: 20069587]
4. Woo BK. Dementia health promotion for Chinese Americans. Cureus 2017 Jun 29;9(6):e1411 [FREE Full text] [doi: 10.7759/cureus.1411] [Medline: 28856076]
5. Woo BK. Family history and its relationship with dementia stigma beliefs among Chinese Americans. Geriatr Gerontol Int 2017 Jan;17(1):122-125. [doi: 10.1111/ggi.12686] [Medline: 26694867]
6. Chan KY, Wang W, Wu JJ, Liu L, Theodoratou E, Car J, Global Health Epidemiology Reference Group (GHERG). Epidemiology of Alzheimer's disease and other forms of dementia in China, 1990-2010: a systematic review and analysis. Lancet 2013 Jun 08;381(9882):2016-2023 [FREE Full text] [doi: 10.1016/S0140-6736(13)60221-4] [Medline: 23746902]
7. Lam NHT, Woo BKP. YouTube as a new medium for dementia education among Chinese Americans. Community Ment Health J 2019 Oct 22. [doi: 10.1007/s10597-019-00493-7] [Medline: 31641910]
8. Lam NH, Woo BK. Digital media recruitment for fall prevention among older Chinese-American individuals: observational, cross-sectional study. JMIR Aging 2018 Nov 01;1(2):e11772 [FREE Full text] [doi: 10.2196/11772] [Medline: 31518249]
9. Cheng TY, Liu L, Woo BK. Analyzing Twitter as a platform for Alzheimer-related dementia awareness: thematic analyses of tweets. JMIR Aging 2018 Dec 10;1(2):e11542 [FREE Full text] [doi: 10.2196/11542] [Medline: 31518232]
10. Topfer L. GPS locator devices for people with dementia. CADTH Issues in Emerging Health Technologies 2016 Aug 31. [Medline: 27809428]

11. Nilakantan AS, Mesulam M, Weintraub S, Karp EL, VanHaerents S, Voss JL. Network-targeted stimulation engages neurobehavioral hallmarks of age-related memory decline. Neurology 2019 May 14;92(20):e2349-e2354. [doi: 10.1212/WNL.0000000000007502] [Medline: 30996057]
12. Strickland E. IEEE Spectrum. 2019 Mar 19. First device to treat Alzheimer's is up for approval by the FDA URL: https://spectrum.ieee.org/the-human-os/biomedical/devices/the-first-device-to-treat-alzheimers-is-up-for-approval-by-the-fda
13. Davis K. Sixth Tone. 2017 Jun 22. In judicial first, man imprisoned for fake Taobao reviews URL: https://www.sixthtone.com/news/1000374/in-judicial-first%2C-man-imprisoned-for-fake-taobao-reviews

Abbreviations

B2C: business-to-consumer
e-commerce: electronic commerce
FDA: Food and Drug Administration
NMPA: National Medical Products Administration
TMS: transcranial magnetic stimulation

Edited by G Eysenbach; submitted 17.12.19; peer-reviewed by O Liran, X Zheng, P Dunn, T Kawashita; comments to author 05.01.20; revised version received 05.01.20; accepted 07.02.20; published 30.03.20.

Please cite as:
Woo BKP
Dementia-Related Products on an e-Commerce Platform
JMIR Biomed Eng 2020;5(1):e17514
URL: http://biomedeng.jmir.org/2020/1/e17514/
doi:10.2196/17514
PMID:

Viewpoint

Innovation in Pediatric Medical Devices: Proceedings From The West Coast Consortium for Technology & Innovation in Pediatrics 2019 Annual Stakeholder Summit

Juan Espinoza[1,2], MD; Kathryne Cooper[1], MBA; Nadine Afari[1,3], MS; Payal Shah[1,2], MS; Sriharinarayana Batchu[1], BS; Yaniv Bar-Cohen[1,4], MD

[1]The West Coast Consortium for Technology & Innovation in Pediatrics, Los Angeles, CA, United States

[2]Children's Hospital Los Angeles, Division of General Pediatrics, Los Angeles, CA, United States

[3]Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States

[4]Children's Hospital Los Angeles, Division of Cardiology, Los Angeles, CA, United States

Corresponding Author:
Juan Espinoza, MD
Children's Hospital Los Angeles
Division of General Pediatrics
4650 Sunset Blvd MS#76
Los Angeles, CA
United States
Phone: 1 323 361 2721
Email: jespinoza@chla.usc.edu

Abstract

Pediatric medical devices cover a broad array of indications and risk profiles, and have helped to reduce disease burden and improve quality of life for numerous children. However, many of the devices used in pediatrics are not intended for or tested on children. Several barriers have been identified that pose difficulties in bringing pediatric medical devices to the market. These include a small market and small sample size; unique design considerations; regulatory complexities; lack of infrastructure for research, development, and evaluation; and low return on investment. In 2007, the Food and Drug Administration (FDA) created the Pediatric Device Consortia (PDC) Grants Program under the administration of the Office of Orphan Products Development. In 2018, the FDA awarded over US $30 million to five new PDCs. The West Coast Consortium for Technology & Innovation in Pediatrics (CTIP) is one of these PDCs and is centered at the Children's Hospital Los Angeles. In February 2019, CTIP convened its primary stakeholders to discuss its priorities and activities for the new grant cycle. In this paper, we have presented a report of the summit proceedings to raise awareness and advocate for patients and pediatric medical device innovators as well as to inform the activities and priorities of other organizations and agencies engaged in pediatric medical device development.

(JMIR Biomed Eng 2020;5(1):e17467) doi:10.2196/17467

KEYWORDS

medical device development; pediatrics; innovation; proceedings; United States Food and Drug Administration

Introduction

Pediatric medical devices cover a broad array of indications and risk profiles and have helped to reduce disease burden and improve quality of life for numerous children. However, many of the devices used in pediatrics are not intended for or tested on children. Although some devices such as infant incubators were designed specifically for children, many others are adult devices adapted for pediatric use. Although children and adults can suffer from similar diseases, their device needs are different because of anatomical and physiological differences, physical activity, body structure and functions, and the challenges of growth [1]. Owing to the potential for much longer device use as compared with adults, device longevity and adverse effects of long-term implanted materials can be more significant issues in children. According to a national survey of government-associated clinicians conducted by the Food and Drug Administration (FDA) and the National Center for Advancing Translational Sciences, despite cutting-edge research and improved technologies to advance pediatric device development, the percentage of novel pediatric devices designed, evaluated, and approved for pediatrics is only about a quarter

of those approved for adults [2]. In the same survey, 91% of clinicians reported that a new or improved device is needed, and 64% were dissatisfied with existing devices [3].

Several factors have been identified that pose difficulties in bringing pediatric medical devices to the market. These include a small market and small sample size; unique design considerations; regulatory complexities; lack of infrastructure for research, development, and evaluation; and low return on investment [4]. The end result of these barriers is that very few new devices end up receiving specific pediatric regulatory approval from the FDA. In 2017, 66 new devices were approved through the premarket approval and humanitarian device extension pathways; only 18 of those were indicated for use in the pediatric population and even fewer for children younger than 18 years [2]. Of the remaining 48 devices approved for adults, 88% (n=42) were determined by pediatric experts to have potential applicability in pediatric diseases [2]. Due to the paucity of approved devices, children are potentially exposed to greater risk because providers are forced to use either off-label devices or less-effective therapies.

The FDA has committed to advancing policies to encourage the development of safe and effective medical devices designed specifically for pediatric patients. In 2007, with the enactment of Section 305 of the Pediatric Medical Device Safety and Improvement Act, the Pediatric Device Consortia (PDC) Grants Program was created and administered by the FDA's Office of Orphan Products Development [5,6]. The primary aim of this program is to facilitate the development, production, and distribution of medical devices. PDC serves providers with scientific and regulatory advice, physical and design resources, identifying funding sources and business assistance for the development of medical devices. The consortia have assisted the development of more than 1000 proposed medical devices in various stages of the total product life cycle. There are 19 unique pediatric medical devices available to patients as a result of this program, including a needle-free blood collection device, a surgical vessel sealing system, and a rapid vascular infusion device [7]. The PDC program has awarded US $37 million to various consortia since 2009 for pediatric device research and development (Figure 1) [6,7]. The FDA is also collaborating with industry stakeholders to build the National Evaluation System for health Technology [8] to generate better evidence for medical device evaluation and regulatory decision making, and incorporating real-world evidence generation strategies for more efficient and balanced approaches toward pre- and postmarket data collection [7,9].

Figure 1. Past and current Pediatric Device Consortia 2009-2018.

2009	2011	2013	2018
The Michigan Pediatric Device Consortium (**A**)	University of Michigan MPED & PMDI Pediatric Medical Device Consortium (**A**)	University of Michigan Pediatric Device Consortium (**A**)	UCSF-Stanford Pediatric Device Consortium (**B+C**)
University of California, San Francisco Pediatric Device Consortium (**B**)	University of California, San Francisco Pediatric Device Consortium (**B**)	Boston Pediatric Device Consortium (**D**)	National Capital Consortium for Pediatric Device Innovation 2.0 (**F**)
The MISTRAL Device Consortium (**C**)	Atlanta Pediatric Device Consortium (**E**)	Atlantic Pediatric Device Consortium (**E**)	The West Coast Consortium for Technology & Innovation in Pediatrics (**H**)
The Pediatric Cardiovascular Device Consortium (**D**)		National Capital Consortium for Pediatric Device Innovation (**F**)	Pennsylvania Pediatric Medical Device Consortium (**I**)
		New England Pediatric Device Consortium (**G**)	Southwest National Pediatric Device Innovation Consortium (**J**)
		Southern California Center for Technology and Innovation in Pediatrics (**H**)	
		Philadelphia Regional Pediatric Medical Device Consortium (**I**)	

About The West Coast Consortium for Technology & Innovation in Pediatrics

The West Coast Consortium for Technology & Innovation in Pediatrics (CTIP) is a PDC centered at the Children's Hospital Los Angeles (CHLA) and the University of Southern California (USC). Established in 2011 and first funded by the US FDA in 2013, CTIP promotes the commercialization and clinical use of pediatric medical device technology. In August 2018, CTIP was awarded a new US $6.6 million P50 grant from the FDA to continue its efforts to advance the research and development of medical devices for children.

CTIP attempts to address an important component often missing from pediatric device innovation, that is, simultaneously engaging clinicians, engineers, regulators, hospital administrators, patients, and the business community in the process of assessment and development of technology. For portfolio companies, CTIP fosters networking opportunities, direct and indirect financial support, and guidance on issues related to, but not limited to, intellectual property (IP), prototyping, engineering, testing, grant writing, and clinical trial design. CTIP has a network of children's hospitals, academic institutions, accelerators, and incubators across The West Coast to support the commercialization of pediatric medical devices. CTIP network members include the University of California, Los Angeles; Oregon Health & Science University; University of Southern California; University of California, San Diego; University of California, Berkeley; Seattle Children's Hospital; Cedars-Sinai Accelerator; The Lundquist Institute (formerly La BioMed); and Project Zygote (Figure 2).

Figure 2. List of The West Coast Consortium for Technology & Innovation in Pediatrics (CTIP) member institutions.

No.	Institution	Role
1	Children's Hospital Los Angeles	Principal Site
2	Oregon Health & Science University	Funded Academic Partner
3	University of California Los Angeles	Funded Academic Partner
4	University of California San Diego	Unfunded Partners and Collaborators
5	University of Southern California	Unfunded Partners and Collaborators
6	Seattle Children's Hospital	Unfunded Partners and Collaborators
7	University of California Berkeley	Unfunded Partners and Collaborators
8	Cedars-Sinai Accelerator	Unfunded Partners and Collaborators
9	The Lundquist Institute	Unfunded Partners and Collaborators
10	Project Zygote	Unfunded Partners and Collaborators

CTIP aims to do the following:

- Build upon our national network of multidisciplinary stakeholders to identify and foster promising pediatric medical device projects.
- Increase awareness around the need for novel pediatric medical device development.
- Overcome current barriers to commercialization with a particular focus on establishing academia's role in alleviating these barriers.
- Develop and implement strategies that will sustain a productive needs-driven pipeline of new pediatric medical devices.

The 2019 West Coast Consortium for Technology & Innovation in Pediatrics Summit

On February 1, 2019, the CTIP hosted a consortium summit at CHLA to discuss pediatric medical devices, regulatory challenges, reimbursement strategies, and prototype, design, and business development. The CTIP's advisory board and steering committee members shared current industry knowledge and research in the field of Medtech and explored different pathways to commercialization success. During the discussion sessions, participants were encouraged to focus on ways to enhance collaboration between their institutions and CTIP. The information summarized herein reflects the knowledge and opinions of the 2019 summit participants (Table 1).

Table 1. List of 2019 summit attendees.

Name	Title	Organization	The West Coast Consortium for Technology & Innovation in Pediatrics affiliation
Katz, Samantha	Venture General Manager and Health Director	BCG Digital Ventures	Advisory board
Rousset, Jessica	Chief Operating Officer	Cure Pharmaceuticals	Advisory board
Edgerton, Kathryn	Partner	Nelson Hardiman Healthcare Lawyers	Advisory board
Cramer, Charlotte	Innovation Strategist, Innovation Studio	CHLA[a]	CHLA staff
Crown, Kelly	Innovation Strategist, Innovation Studio	CHLA	CHLA staff
Eric, Meyer	Innovation Strategist, Innovation Studio	CHLA	CHLA staff
Kulkarni, Omkar	Chief Innovation Officer	CHLA	CHLA staff
Martine, Broome	Portfolio and Alliance Manager, Office of Technology Commercialization	CHLA	CHLA staff
Ng, Victoria	Innovation Consultant, Innovation Studio	CHLA	CHLA staff
Afari, Nadine	Program Manager	CTIP[b], CHLA	CTIP staff
Bar-Cohen, Yaniv	Co-Director	CTIP, CHLA	CTIP staff
Batchu, Srihari-narayana	Program Administrator	CTIP, CHLA	CTIP staff
Cooper, Kathryne	Co-Director	CTIP, CHLA	CTIP staff
Espinoza, Juan	Director	CTIP, CHLA	CTIP staff
Shah, Payal	Research Associate	CTIP, CHLA	CTIP staff
Senn, Sean	Managing Partner	Accelerator AIX	Entrepreneur in residence
Bernstein, Christina	President	BB Medical Surgical, Inc	Entrepreneur in residence
Furth, Neelima	Regulatory Consultant	Best Laid Plans, Inc	Entrepreneur in residence
Patnaik, Meeta	SVP Clinical Development and Medical Affairs	Bionaut Labs	Entrepreneur in residence
Rushi, Amit	Vice President, Product, Marketing, and Business Development	GraftWorx	Entrepreneur in residence
Hager, Earle	Managing Partner	The Neutrino Donut, LLC	Entrepreneur in residence
Ulmer, Kwame	Principal	Ulmer Ventures, LLC	Entrepreneur in residence
Plush, Robert	VP of Pediatrics	Surgical Theater	Entrepreneur in residence
Sheridan, David	Codirector, Emergency Medicine Clinical Innovation Program	Oregon Health & Science University	Steering committee
Watson, Andrew	Senior Director of Technology Transfer	Oregon Health & Science University	Steering committee
Sue Swanson, Wendy	Chief of Digital Innovation	Seattle Children's Hospital	Steering committee
Horse-Grant, Desert	Senior Director of Research and Innovation	University of California Los Angeles	Steering committee
Levi, Dan	Interventional Pediatric Cardiologist	University of California Los Angeles	Steering committee
Flores, Ruben	Director, Office of Innovation and Commercialization	University of California San Diego	Steering committee
Blanco, Cesar	Senior Director of Research and Development Programs, Regulatory and Quality	USC Alfred E. Mann Institute for Biomedical Engineering	Steering committee
Grossman, Elissa	Associate Professor of Clinical Entrepreneurship	USC Marshall School of Business	Steering committee
Richmond, Frances	Chair, Department of Regulatory and Quality Sciences	USC School of Pharmacy	Steering committee
Khoo, Michael CK	Professor of Biomedical Engineering and Pediatrics	USC Viterbi School of Engineering	Steering committee

Name	Title	Organization	The West Coast Consortium for Technology & Innovation in Pediatrics affiliation
Loeb, Gerald	Professor of Biomedical Engineering and Neurology	USC Viterbi School of Engineering	Steering committee
Tolomiczenko, George	Admininistrative Director, Health, Technology, and Engineering Program	USC Viterbi School of Engineering	Steering committee

[a]CHLA: Children's Hospital Los Angeles.

[b]CTIP: The West Coast Consortium for Technology & Innovation in Pediatrics.

Summit Proceedings

Session 1: Pediatric Medical Devices—Where Should Consortium for Technology & Innovation in Pediatrics Focus Resources and Efforts?

Session 1 was led by Dr. Juan Espinoza and Dr. Yaniv Bar-Cohen, the principal investigator (PI) and co-PI of CTIP, respectively. They began by leading a discussion about the importance of defining "success" for CTIP. For the FDA, "success" refers to device approvals and commercialization. These are important milestones that validate the need for programs such as the Pediatric Device Consortia. However, these are rare events on longer timelines; therefore, the focus should be on process measures that show progress toward those key milestones.

One of CTIP's goals is to increase awareness in the Medtech community around the unique issues faced by pediatric medical devices, pediatric indications, and labeling. Although the legal and regulatory framework used by the FDA for devices is complex, the economic and market barriers to medical and surgical device development for children are significant. In the future, as early as 2020, CTIP could explore current medical devices on the market that could potentially pursue pediatric classification. CTIP's leadership team discussed that the portfolio should have a range of Class 2 and Class 3 devices. Whether CTIP should focus in certain areas, such as orthopedics or implantable devices, was also discussed, with the consensus being that CTIP should continue to support projects that address unmet needs regardless of the focus area.

The majority of medical device companies continue to shy away from developing pediatric medical devices because the process is costly and complicated. Many believe that the market is small with limited opportunities for return on investment (ROI). However, it was pointed out that the United States is the largest market for pediatric medical devices whose growth is attributed to an increase in the incidence of chronic diseases such as heart disease, diabetes, and obesity, increasing awareness about adolescent health and behavioral health and increasing the demand for medical devices in pediatric hospitals and clinics. Pediatric Medtech remains to be a dynamic sector, with new technologies and products emerging every year. Many investors often overlook investment opportunities within pediatric medical devices; CTIP firmly believes that the pediatric Medtech sector can generate excellent investment opportunities.

CTIP understands that one of the best ways to assess whether a given company is focused on innovation and success is to look at the company's product pipeline, team structure, and research and development (R&D) efforts. Innovation in medicine is primarily driven by both need and the size of a market, and the demand for pediatric devices does not always cover the necessary R&D investment required to bring these devices to the market. Investors will also look for these trends and patterns. Entrepreneurs in this space need to maintain an unconventional way of thinking when presenting their devices to investors and appeal to a variety of motivations, including long-term ROI, social impact, and launching additional, often adult-targeted, products. Investors should also be educated on evaluating pediatric Medtech differently, taking into consideration many of the previously mentioned factors.

There are several key points in a Medtech company's life cycle, and each stage has certain ramifications for an investor. Startups often face years of losses and cash outflows, as the management tries to lead new products through clinical and device trials, regulatory processes, and eventually into the market. CTIP and our consultants working with the portfolio members understand that very few Medtech companies mature into large, independent players. For many companies, the goal is licensing or acquisition of their technology or IP. Despite the unique considerations for pediatrics, the total product life cycle of pediatric medical devices will be familiar to Medtech investors; companies should leverage that familiarity in their investor pitches.

CTIP's leadership team shared with the group that as pediatrics devices are a unique subset of the medical device industry, PDC program activities need to cater to that market. Before a medical device can be marketed in the United States, the FDA has to evaluate and approve the device. Although not all product approvals require expensive device trials, many products that drive revenue growth for the pediatric sector require data on efficacy and safety before the FDA permits their sale. It is also worth noting that approval is not the end of the regulatory process; the FDA requires ongoing monitoring and reporting and can order devices off the market if hidden dangers reveal themselves in subsequent years.

Risk analysis is an essential component of a quality management system. CTIP should focus on derisking devices early on. CTIP consultants work closely with the startup's risk management team to create a plan for applying risk management procedures to the design and manufacture of their pediatric medical device. It is crucial to document how risks are identified, evaluated, and traced. The pediatric device innovator's plan should define the entire scope of the risk management process, including the purpose of the device, its life cycle, responsible parties and authorities, and data collection and analysis all the way through postproduction.

Emphasis was put on CTIP's initial evaluation of the company and asking clear go/no-go questions for every company: "How good or bad is this idea? What are the risks? What will the market bear? How will payers receive this?"

Other key questions that companies should be able to answer include:

1. How many units do you think you can sell?
2. What price could you get?
3. Walk us through how is this going to work?
4. How is your business sustainable?

This is meant to push companies to think through their business and sustainability plans.

Suggestions from the participants included the following:

1. CTIP should create risk-analysis worksheets for companies to complete during the initial evaluation.
2. During the evaluation call, CTIP should ask companies to share the top three markets for their device and consider how they might pursue market expansion. During the evaluation call, companies should also be asked, "What your sustainability plan is for the long term?" "What the other commercialization possibilities are?"
3. CTIP should keep in mind that a mature technology company is valued partly by traditional methods, including profit, revenue growth, and overall sales. Derisking these elements, particularly in the early stages, can add significant value. Of course, because technology is an ever-changing space, even what we consider a successful Medtech company can rise or fall based on an unproven product or even an announcement of a new development.

One of the other considerations that CTIP and companies need to be aware of is the issue of reimbursement. The reimbursement strategy is just as important as the regulatory strategy. Reimbursement assessment is used largely to help make informed decisions about the coverage of health care services by private and public payers. Reimbursement influences product development strategy from an early stage and drives decision making about potential asset acquisitions. It also combines health economics, outcomes research, pricing and reimbursement, external affairs, key opinion leader engagement, advocacy, and government lobbying. The suggestion was that CTIP should provide portfolio members with key reimbursement information to ensure startup success.

Some of the key reimbursement questions CTIP should address with portfolio members are as follows:

1. What are the priorities in the hospital?
2. How do pediatric hospitals assess value and who is involved?
3. How do hospital decision makers decide what they want?
4. What are the barriers for adoption?
5. What does the reimbursement approval process look like?
6. What are the relevant Centers for Medicare & Medicaid Services and insurance policies involved?

The discussion also focused on creating an innovation roadmap to connect entrepreneurs with the decision makers in the hospitals and/or create a roadmap for important decisions. This innovation roadmap should consider both the hospital business development process and entrepreneurs. This roadmap would remind innovators of whom they need to connect with during this journey. Having regular opportunities to network and build relationships between stakeholders is important to more effectively work together. Clear communication that centers on sharing resources, information, and best practices among stakeholders can also help to achieve shared goals.

As a part of CTIP Activities for year 1, the group recommended developing a business relationship with the Los Angeles Mayor's office with the possibility of a pipeline between the Mayor's Office, the Los Angeles Bioscience Investment Fund, CHLA, and others to create more awareness about opportunities in pediatric Medtech. In 2015, the Los Angeles County Board of Supervisors made economic development a priority to stimulate regional job growth and lift residents out of poverty. Seven industries were targeted based on their proven ability to create jobs and wealth. Bioscience was the first of the industry sectors selected by Los Angeles county for focused support. The Los Angeles region generates cutting-edge bioscience R&D and a trained workforce capable of launching and supporting enterprises emerging from local research institutions and incubators. Over past economic cycles, including the Great Recession, bioscience jobs have proven unaffected by economic downturn. Increased awareness in the business and investor community would facilitate follow-on investments in CTIP portfolio members.

Nearing the end of the first session, the participants highlighted the need to understand that in most hospitals, if the projects do not benefit the home institution, then there is limited incentive to move forward. Seattle's Children's Hospital encourages a dual incentive for both the hospital and the entrepreneur, which also encourages quality control and safety. One suggestion was brought up to host office hours at CHLA and document what has been done/shared between entrepreneurs and hospitals. This could reduce communication barriers between entrepreneurs, clinicians, and administrators. CTIP could incentivize this process by funding participants' time as consultants. This could be a bridge building activity that makes the process more official. The CTIP's leadership will explore connecting with human resources on challenges or issues regarding this suggestion.

The first session ended energetically with other programing solutions as follows:

1. Collecting success stories and sharing the pathways/stories from both the entrepreneur side and the hospital side.
2. Sharing stories in a video format or blog featured on the CTIP site could be beneficial to Medtech commercialization.

Session 2: Business Models for Pediatric Medical Devices—How Do We Advise Companies on Their Path Toward Commercialization?

Session 2 was led by Elissa Grossman, MBA, PhD-CTIP steering committee member; Kathryne Cooper, MBA, CTIP Co-Director; and Christopher Del Vecchio, MS, JD-CTIP, steering committee member. The session began with a discussion

of the current CTIP portfolio. Many CTIP companies (roughly 50) are at an early stage. The pediatric market as a subdivision of health care accounts for approximately 25% of the US population (individuals younger than 21 years). This smaller market results in many businesses having limited resources and needing to pursue nontraditional funding opportunities. CTIP should consider the nuances of the matchmaking process and match investor money with the right entrepreneur. It is important to ensure that interests are aligned, connect with investors to strategize where they can invest their money, and work with business development consultants to create different kinds of business models and commercialization strategies.

Historically, the requirements of the American market have set standards for the design and development of products, functions, and processes. Although the United States will continue to be the most important market for the next decade, global markets will become increasingly relevant for pediatric Medtech startups. These markets will provide opportunities for new customers and expand operational activities, such as manufacturing, shared services, and R&D.

For therapeutics and diagnostics markets where Medtech holds significant new potential, R&D and business development investments should lead to the creation of new technologies that have both clinical efficacy and cost-effectiveness. This will require companies to leverage existing global commercial, development, and operating capabilities. Companies may consider partnerships with other organizations that have existing capabilities. However, all partnerships should clearly delineate any IP considerations. When an innovator engages with a contractor or consultant, and the innovator expects to own the IP arising from the engagement, the innovator should expressly provide clarification for this in the engagement agreement. Lack of clarity around the important issues such as who owns the company assets, documents, and IP can pose a significant risk.

Given the rapidly evolving nature of Medtech, it is important for entrepreneurs to reflect on how successful teams choose which ideas to pursue. Their business model should include a clearly articulated plan for how they make these decisions. These decisions are informed by a clear understanding of health care technology purchasing and the relevant processes and people involved in those decisions. The Medtech industry as a whole faces challenges to the legacy business models and strategic choices that companies have used to excel in the past.

Success in Medtech requires a clear and consistent focus on delivering differentiated value and performance to customers, shareholders, and users. A new “value bar” forces Medtech startups to rethink how they can effectively create a product portfolio that will meet this ever-increasing set of expectations. It is no longer sufficient to demonstrate the marginal product benefits for new product launches. A more diverse set of stakeholders are now involved in the adoption and decision-making process for Medtech products. A broader set of stakeholders to support product adoption is needed.

Suggestions from the participants included the following:

1. CTIP companies should try to work with hospitals to evaluate outcome measures and costs.
2. Business models that CTIP companies should consider include value-based models that focus on cost containment and help the hospital be more profitable.
3. One needs to be careful about the technology sector, which can be boom or bust. The same is true for individual companies and market segments within the Medtech space.
4. It will be helpful to read the 2013 book by Michael Raynor and Mumtaz Ahmed (The Three Rules: How Exceptional Companies Think) [10], which brings discipline to the field by identifying rigorous, research-backed principles that guide exceptional companies in two ways: (1) better before cheaper: rather than competing solely on price, companies achieve sustainable success by focusing on delivering differentiated value; and (2) revenue before cost: the advantages of higher revenue are more valuable and durable than the advantages of lower cost.

Medtech companies could have a significant customer base and still not show consistent profits. In many cases, they lose money, sometimes a significant amount, as these companies build out capacity and develop a market for their products. To build differentiated products, many companies need to consider transformational innovation: innovation that creates and delivers customer value through novel products, solutions, and business models that address these unmet market needs.

To improve their business model, companies should ask themselves these key questions:

1. How will your customers hear about your medical device?
2. What is the value proposition of your device?
3. Who do entrepreneurs need to speak with to educate the customer?

CTIP can assist in finding a systematic approach to facilitate entrepreneurs’ connection with those making purchasing decisions. Entrepreneurs need to understand who the decision makers are for their potential customers. CTIP is exploring creating a repeatable operational process around this topic and understanding how it can incentivize decision makers to attend, present, and engage. Encouraging innovation and using analytics to share success metrics are proactive ways to connect with hospital administrators. With the help of CTIP, companies can complete a derisk tech worksheet to decide if they want to design for manufacturing, acquisition, or licensing.

There are a couple of key points to consider when hospitals partner with an entrepreneur:

1. How is the collaboration structured and defined?
2. How are entrepreneurs evaluating their technology? Solely based on profit?

Hospitals evaluate new technology based on value and cost savings. Companies and CTIP can work with hospitals to collectively define success and quantify the positive clinical impact of devices.

For existing companies with adult devices, there are opportunities to expand into pediatrics. Options include developing the product themselves or licensing their IP to a second company to develop the pediatric version. This could be a novel approach to bring more pediatric medical devices to

the market. CTIP can match companies with Entrepreneurs-In-Residence who can help portfolio members explore and negotiate these types of arrangements.

The group also discussed other nonprofit–driven business models such as B-Corps and 501(c)3 nonprofits. CTIP can help companies find a way to blend the nonprofit model with how to operate as a business that supports the device. Companies that form a nonprofit organization do not have to make decisions based on a profit motive; rather, they can operate on mission instead. If the values of new investors are not aligned, companies do not have to take their offers, giving entrepreneurs more latitude. A major consideration is that entrepreneurs need to demonstrate how their product development and mission meet the legal requirements for being structured under these mechanisms.

The discussion then moved to small business innovation research and small business technology transfer (SBIR/STTR) program issues:

1. Some institutes have started evaluating SBIR/STTR proposals using a similar rubric to National Institutes of Health research grants such as R01. Most small companies do not typically have the skills or background to write these grants. CTIP can assist companies by connecting them to grant writing support.
2. The question raised was if CTIP could assemble a team of PhDs and other experts to help with grant writing.
3. It would be very helpful if CTIP can host workshops for portfolio companies at different stages of development but focus on similar class devices and projects. These workshops can teach reimbursement pathways, SBIR grant basics and different business model strategies, perhaps as a 4-week course.
4. CTIP should curate case studies from its portfolio that exemplify key processes and decisions and share them with the rest of the portfolio to learn from.

Session 3: Food for Thought—Derisking Strategies for the Commercialization of Medical Devices

Session 3 was led by Cesar E. Blanco, PhD - CTIP Steering Committee, on derisking strategies for the commercialization of medical devices. The “valley of death” [11] in startup formation is often difficult to traverse for pediatric medical device companies in particular. This is because investors are reluctant to support startups if their products have not yet been derisked. In addition, the pediatric device market is seen by some as niche and considered risky from the beginning. CTIP can assist with derisking in two main ways: (1) by providing funding for prototype and design so entrepreneurs can take their first iteration product to their stakeholders for feedback and (2) by pairing the entrepreneur with an Entrepreneur-in-Residence or a consultant via CTIP’s Consultant-Company Match Program. CTIP can serve as a device derisking incubator, with the aim of helping entrepreneurs with business planning, regulatory support, clinical trials, prototyping, and preclinical work.

Medtech startups transform and change in many ways over time and entrepreneurs should know that as they develop a Medtech device. The team and individuals involved in the process change as the company needs change. This will impact how the company matures. In the ideation phase, a company will seek technical expertise, creativity, and team members who can push the boundaries in terms of development and innovation. In the next stage of the R&D process, companies consider and filter design input/requirements. This is where innovators start focusing on meeting design requirements; the need for out of the box thinkers may be less. The team is likely to change again as R&D expands, and their funding increases. Companies should be as intentional about the people on their team as they are about the product. Teams and companies can create a list of competencies, skills, and values. As companies continue to grow and transform, founders and teams should continually re-evaluate their product, their market, and their team composition.

The cost of medical technology is not declining, and many medical professionals equate progress in medicine to the increasing use of sophisticated technology that is often expensive. There is an urgent need to address high-cost small markets through the development and use of appropriate technology in accordance with the needs and priorities of pediatric patients. A number of simple and inexpensive quality measures that have the potential to improve outcomes substantially without the need for expensive equipment should be instituted before embracing high-end technology. Innovations to reduce health care costs are another key component of commercialization success. Medtech entrepreneurs need to learn how to decide how much to charge the product, which requires more thought than simply calculating costs and adding a markup. Entrepreneurs, hospital administrators, and clinicians need to assess and consider how much they value the product or service they buy. Figuring out how much the customer (hospital or patient) values the product or service and pricing is paramount to success. Another important consideration for pricing is whether a product should have multiple price points based on features, use case, or customer.

There was also a discussion about reviewing prototyping, which can be funded by CTIP and is an essential part of medical device development and manufacturing. Not only is it critical for refining the design of a product and testing safety and performance, prototyping is a must for startups, so that they can demonstrate how the product looks and functions. Without prototypes, it would be nearly impossible to secure funding for production. Some nimble startups use rapid prototyping and low-volume manufacturing to compete with established companies and gain market share to save time and money.

Prototypes also play a critical role in obtaining quality feedback from end users, clinicians, and hospital administrators; having a high-fidelity prototype can lead to better design optimization within accelerated development windows. Prototyping is about speed and fast turnaround time. The goal is to find flaws, imperfections, and other opportunities for improvement and then address them quickly. As design specifications advance, the device may be able to enter more than one market. Design should be simplified, and entrepreneurs and engineers may need to let go of the bells and whistles. This can be a critical issue for software and digital health and can lead to issues they did not foresee. Clinicians with solutions often have too many

features and need to refine their designs. Finally, innovators need to be aware of different types of prototypes, such as prototypes for clinical studies, as opposed to prototypes for the evaluation of commercial product fit.

When thinking about derisking their commercialization strategy, entrepreneurs should consider the following questions:

1. What is your market and how big is it?
2. How will you capture and protect IP?
3. When and how can you add value?
4. When is this exciting to another strategic partner?
5. Can this be leveraged in some other market or application?

While discussing team dynamics with innovators, it is important to focus on identifying if the team that founded the company wants to be in full control; how much control is willing to let go of? It is important to know this upfront; as the company matures and acquires investors, they will have to release some control. Companies need to understand their market and development costs. CTIP can host a workshop to help companies identify opportunities to expand, fully explore their value proposition, and make sure they define their customers correctly (the end user is often not the customer in health care). Understanding the dynamic changes and shareholders who both create and hinder change is critical to Medtech success. Marketing leaders push for quicker cycle times, engineering leaders want cutting-edge technology, supply chain leaders seek low-cost initiatives, clinicians want to provide quality care, investors want an ROI, and hospitals want to maintain high quality and low costs. Companies need to be able to deliver a compelling proposition to each of these stakeholders.

With respect to the regulatory process, there are two important parts to a medical device: (1) the device itself and (2) the documentation that supports the design and indication for use, including preclinical studies, efficacy studies, quality management systems, patents, etc. There is a significant value in how well this documentation is organized and updated. It is important for entrepreneurs to develop these assets early. CTIP will need to remind portfolio members about the need to appropriately document quality control and reproducibility. Documentation is critical in the early stages; in some ways, the paperwork becomes more important than the device itself, early on.

Finally, the group addressed the timing of investor outreach. In general, this should be in line with device and company maturity. An example discussed was a device startup company that ran into difficulties with a thermally triggered adhesive. This device uses an adhesive to secure and hold in place pads that connect to a string device that helps close sutures. It was intended to minimize scarring and promote faster healing. The device did this by reducing the strain across incisions, thereby decreasing the size of scars. Currently, this device is challenged by temperature and adhesive strength. Current testing could not meet the high adhesive strength requirements. This company is working on a new formula but has limited funding for this phase of R&D. In this scenario, although the suture closure device is very mature, they still have not solved significant product issues and are having trouble attracting investors. In such cases, CTIP can help companies identify exits, strategize multiple exit points, and articulate concepts to investors.

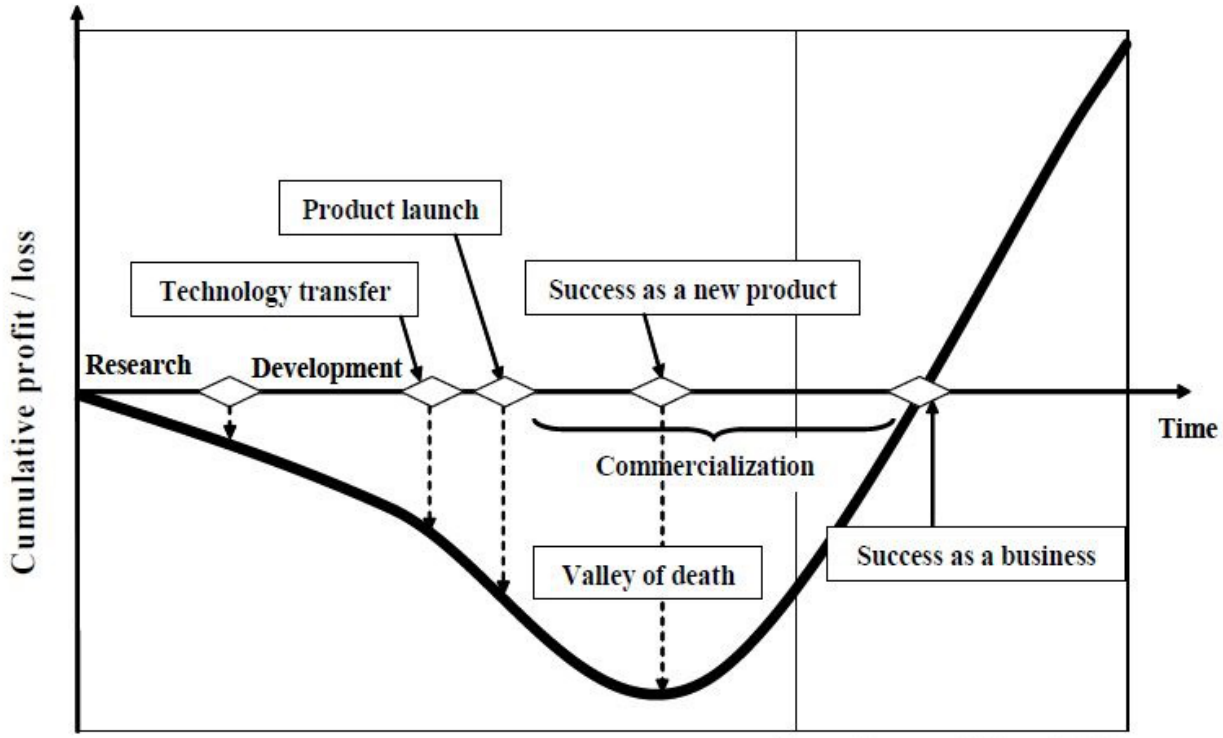
Session 4: The Business of Pediatric Medical Devices—Reimbursement Strategy and Food and Drug Association Regulation

Session 4 was led by Frances Richmond, PhD - CTIP Steering Committee, in which she discussed the role of reimbursement and regulatory strategies. The innovation “valley of death” (Figure 3) [11] prevents the progress of science from the laboratory bench to the point where it provides the basis of a commercially successful business or product in a couple key ways. Medtech startups often need help to access knowledge, capacity, resources, and people outside of their company. Organizations such as CTIP could develop innovation processes and infrastructure to reduce risk and uncertainty for companies, allowing them to invest with confidence to advance a given technology beyond the validation stage all the way through to commercialization. During the CTIP evaluation phase, CTIP can help companies navigate the “valley of death” by teaching companies how to assess the ideal strategic fit; matching their device or pivot project to a new device that will have high value to a strategic partner, leading to commercial success.

The group discussed how CTIP can build bridges between the regulatory and reimbursement processes that reside on either side of the valley. For many products, the reimbursement budget is larger than the regulatory budget. The National Science Foundation (NSF) I-Corp program hosted at USC can assist CTIP companies through their entrepreneurship boot camp. Once completed, companies are eligible for NSF development grants. One way CTIP can support the portfolio members is to have a program where we embed a regulatory program graduate student with a portfolio member at a reasonable cost; together, the student and company can create the required documentation. CTIP can also create sample templates for quality management systems, documents/work orders, and templates for regulatory submissions. The Southern California Clinical and Translational Science Institute (USC’s Clinical and Translational Science Award) has developed web-based modules for training in clinical and translational science. CTIP could develop a similar set of resources for its portfolio companies.

CTIP and PDCs could provide another type of value to early stage companies: validation. The PDCs might consider developing a joint certification or validation, a “stamp of approval” for early stage innovators, which in some cases can be as valuable as financial support. This can also be of value to more mature companies who need that type of endorsement to pursue investors and customers. It would create some branding and coordination across the PDCs but may be worth pursuing as a value add-on.

Figure 3. Valley of death: overall process from research to commercialization.



The discussion was extended on the list of templates and checklists that CTIP can create for portfolio members, including the following: (1) tech maturity checklist; (2) quality control checklist; (3) regulatory checklist; (4) reimbursement value analysis; (5) 5-10 elements essential for the company's progression; (6) company self-evaluation; and (7) design control checklist.

CTIP should pursue partnerships with other organizations such as the USC collaborative fund, to sponsor educational events. These events can feature industry experts to discuss a variety of topics, including the development of social capital. Another idea is to host a "shark tank"–style pitch competition with key stakeholders at CHLA and other hospitals. There would be an idea submission process that would include the checklist, and ideas processed are given feedback by the team. There are some retired individuals with regulatory and reimbursement expertise available and are often eager to assist in the process. Many of these experts are looking to find ways to stay involved in the field and give back to the Medtech community. CTIP could play a role in developing a list of individuals who want to provide pro bono services to pediatric medical device innovators.

Conclusions

As FDA-funded pediatric medical device accelerators, the PDCs are uniquely positioned to not only support pediatric medical device innovators but also raise awareness about the barriers to pediatric medical device development and advocate for both patients and innovators to streamline and accelerate the pathway to the market. Follow-up discussions with Summit attendees also raised challenges that CTIP hopes to address in year 2. These include the dynamic nature of software regulatory requirements, considerations regarding artificial intelligence and machine learning, and the limitations of other resources based on different devices with different levels of complexity. The ability of artificial intelligence to transform health care by analyzing vast amounts of clinical data presents unique challenges. This includes unstructured datasets, data ownership, and privacy issues to store global data while adhering to evolving regulatory definitions and requirements. These challenges demand engagement from our experienced partners and network to provide novel insights and increase product innovation. Our goal in sharing the proceedings of our annual stakeholder meeting is to provide insights into our process and priorities for individuals, institutions, and agencies with a vested interest in pediatric medical device development who wish to engage in similar work. By sharing these insights, we hope to begin to align efforts and incentives across sectors to achieve our common goal of improving child health outcomes through technology.

Acknowledgments

The authors would like to express their gratitude to all the participants of the 2019 summit. They would especially like to thank Nandinee Rajyagor for taking notes, Steven Du for helping with the event, Omkar Kulkarni and the Innovation Studio team for their ongoing support and camaraderie, and the staff at the FDA Office of Orphan Products Development for their leadership and support of the PDC program.

This work was supported by the FDA under award number P50FD006425 (PI: Espinoza) for The West Coast Consortium for Technology & Innovation in Pediatrics. The funding source had no involvement in the development of this manuscript or in the decision to submit the paper for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the FDA.

Conflicts of Interest

None declared.

References

1. Safe Medical Devices for Children. Washington, DC: The National Academies Press; 2006.
2. Gottlieb S. Public Meeting - Pediatric Medical Device Development. US Food and Drug Administration. 2018. URL: https://www.fda.gov/medical-devices/workshops-conferences-medical-devices/public-meeting-pediatric-medical-device-development-august-13-14-2018-fda-white-oak-campus-08132018 [accessed 2020-06-01]
3. 2018 FDA/NCATS Report on Unmet Medical Device Needs for Patients with Rare Diseases. US Food and Drug Administration. 2019. URL: http://www.fda.gov/industry/designating-humanitarian-use-device-hud/2018-fdancats-report-unmet-medical-device-needs-patients-rare-diseases [accessed 2020-05-26]
4. Bergsland J, Elle OJ, Fosse E. Barriers to medical device innovation. Med Devices (Auckl) 2014;7:205-209 [FREE Full text] [doi: 10.2147/MDER.S43369] [Medline: 24966699]
5. Pediatric Device Consortia Grants Program. US Food and Drug Administration. 2019. URL: http://www.fda.gov/industry/developing-products-rare-diseases-conditions/pediatric-device-consortia-grants-program [accessed 2020-05-26]
6. Ulrich LC, Joseph FD, Lewis DY, Koenig RL. FDA's pediatric device consortia: national program fosters pediatric medical device development. Pediatrics 2013 May;131(5):981-985. [doi: 10.1542/peds.2012-1534] [Medline: 23569100]
7. FDA Awards Five Grants to Advance the Development of Pediatric Medical Devices. US Food and Drug Administration. 2019. URL: http://www.fda.gov/news-events/press-announcements/fda-awards-five-grants-advance-development-pediatric-medical-devices [accessed 2020-05-26]
8. National Evaluation System for health Technology (NEST). US Food and Drug Administration. 2019. URL: http://www.fda.gov/about-fda/cdrh-reports/national-evaluation-system-health-technology-nest [accessed 2020-05-26]
9. Real-World Evidence. US Food and Drug Administration. 2019. URL: http://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence [accessed 2020-05-26]
10. Michael R, Mumtaz A. The Three Rules: How Exceptional Companies Think. London, UK: The Penguine Group; 2013.
11. Osawa Y, Miyazaki K. An empirical analysis of the valley of death: large - scale R&D project performance in a Japanese diversified company. Asian J Technol Innov 2006 Jan;14(2):93-116. [doi: 10.1080/19761597.2006.9668620]

Abbreviations

CHLA: Children's Hospital Los Angeles
CTIP: The West Coast Consortium for Technology & Innovation in Pediatrics
FDA: Food and Drug Administration
IP: intellectual property
NSF: National Science Foundation
PDC: Pediatric Device Consortia
PI: principal investigator
R&D: research and development
ROI: return on investment
SBIR: small business innovation research
STTR: small business technology transfer
USC: University of Southern California

Edited by G Eysenbach; submitted 17.12.19; peer-reviewed by F Meza, K Chong; comments to author 10.02.20; revised version received 03.04.20; accepted 14.05.20; published 03.07.20.

Please cite as:
Espinoza J, Cooper K, Afari N, Shah P, Batchu S, Bar-Cohen Y
Innovation in Pediatric Medical Devices: Proceedings From The West Coast Consortium for Technology & Innovation in Pediatrics 2019 Annual Stakeholder Summit
JMIR Biomed Eng 2020;5(1):e17467
URL: http://biomedeng.jmir.org/2020/1/e17467/
doi:10.2196/17467
PMID:

Viewpoint

Fingerprint Biometric System Hygiene and the Risk of COVID-19 Transmission

Kenneth Okereafor[1*], HND, PGD, BSc, MSc, PhD, PhD; Iniobong Ekong[2*], MPH, MBBS; Ini Okon Markson[3*], MBBS; Kingsley Enwere[1*], BSc

[1]Department of Information & Communications Technology, National Health Insurance Scheme, Abuja, Nigeria

[2]Department of Health Planning, Research & Statistics, Federal Capital Territory Health & Human Services Secretariat, Abuja, Nigeria

[3]Evangelical Church Winning All (ECWA) Comprehensive Medical Centre, Karu, Abuja, Nigeria

[*]all authors contributed equally

Corresponding Author:
Kenneth Okereafor, HND, PGD, BSc, MSc, PhD, PhD
Department of Information & Communications Technology
National Health Insurance Scheme
NHIS Data Centre
Abuja
Nigeria
Phone: 234 8023148494
Email: nitelken@yahoo.com

Abstract

Biometric systems use scanners to verify the identity of human beings by measuring the patterns of their behavioral or physiological characteristics. Some biometric systems are contactless and do not require direct touch to perform these measurements; others, such as fingerprint verification systems, require the user to make direct physical contact with the scanner for a specified duration for the biometric pattern of the user to be properly read and measured. This may increase the possibility of contamination with harmful microbial pathogens or of cross-contamination of food and water by subsequent users. Physical contact also increases the likelihood of inoculation of harmful microbial pathogens into the respiratory tract, thereby triggering infectious diseases. In this viewpoint, we establish the likelihood of infectious disease transmission through touch-based fingerprint biometric devices and discuss control measures to curb the spread of infectious diseases, including COVID-19.

(JMIR Biomed Eng 2020;5(1):e19623) doi:10.2196/19623

KEYWORDS

biometric; contact; contaminate; coronavirus; COVID-19; cybersecurity; disease; fingerprint; hygienic; infectious; pathogen; scanner; surface; verification

Introduction

The primary purpose of fingerprint recognition or identification systems is to provide trustworthy verification of users to control access to resources such as computers as well as to critical facilities such as offices and hospitals. These systems also provide useful data for generating summary periodic reports on usage statistics [1], including the number of participants inside a facility, frequency of entrances and exits, duration of stay, and other intelligence required for monitoring, surveillance, and security administration.

Despite their many benefits, fingerprint scanners are potential sources of disease transmission due to contamination from multiple touches by various users in a wide range of questionable hygienic conditions; as a result, these scanners pose potential transmission risks. Serial use of finger scanners in a given setting may play a more significant role in transmission, as latent prints left on the scanner surface by the deposition of finger moisture, sweat, or oils can soil the surface [2].

Unhygienic thumbs can potentially leave surviving bacteria [3], fungi, and viruses on the surface [4] of the scanner after use, thereby increasing the possibility of transmitting germs that cause illnesses [5], including COVID-19, which is predominantly spread via droplets and contaminated hands or surfaces [6].

Applications of Fingerprint Biometric Systems

Like other biometric devices, fingerprint systems provide a secure and facile means of verifying the identities of humans [7] using their unique finger attributes. These systems are used in a number of applications for various cybersecurity and access control purposes, as follows.

Employee Time and Attendance Systems

Traditional workplaces [8] use fingerprint biometric systems to identify personnel, grant authorized access to offices, and maintain accurate records of staff attendance [1,7,9].

Secure Login Access to Software and Electronic Systems

Fingerprints are used to control login access [10] to computers and mobile phones as well as to banking, accounting, and other software applications [11].

Vehicle Door and Access Control Systems

Increasing numbers of electronic automobile systems are incorporating biometric doors based on fingerprint technologies, enabling drivers to restrict, control, and track usage of their vehicles.

Access Control to Physical Facilities

Fingerprint biometrics are used to regulate access to and within secure environments [12], including airports, hospitals, stadiums, educational institutions [13], and shopping malls.

Verification for Civic Services

Biometric systems are used to control access to citizens' benefits, including population demographics and patient verification, to prevent impersonation of health care beneficiaries by scammers [14].

Automatic Teller Machine Login Security

Many automatic teller machines (ATMs) and point-of-sale vendor systems contain embedded fingerprint biometric login functionalities to improve the security of web-based financial and banking transactions.

Immigration and Embassy Verification Systems

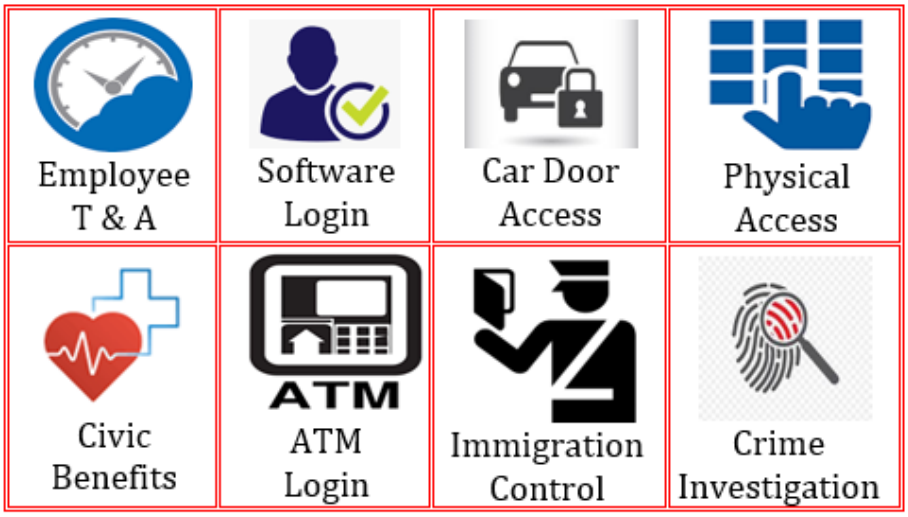
A biometric ePassport [15] contains the passport holder's fingerprint records in line with the United Nations International Civil Aviation Organization (ICAO) 9303 [16] standard for machine-readable travel documents (MRTDs). Embassies and High Commissions also use automated fingerprint systems to capture, store, and exchange profiles for visa verification at border areas.

Crime Investigation

Law enforcement agencies use automated fingerprint identification systems (AFISs) to manage fingerprint databases for crime investigation.

Figure 1 shows illustrations of the various applications of fingerprint biometric systems in commerce, industry, business, and government.

Figure 1. Applications of fingerprint biometric systems. ATM: automatic teller machine; T & A: time and attendance.



Fingerprint systems require direct touch of a finger on the scanner to extract the user's unique biometric image [10]. In each of the application areas discussed above, direct finger contact with the scanner surface is a mandatory operational requirement to obtain the user's unique fingerprint for verification, identification, and secure access.

Overview of Fingerprint Scanners as a Potential Source of Harmful Microbial Transmission

Background and Concerns

Due to their simplicity, ease of use, and cost-effectiveness, the commercial applications of fingerprint systems for identity verification have become widespread. Although the popularity

of these systems suggests a high level of industrial endorsement and user acceptance, their frequent use under various environmental conditions also raises serious concerns regarding their hygienic safety [17] and risks of disease exposure.

A biometric scanner is the component of the biometric system that retrieves the electrical equivalent of the biometric sample for unique measurement. Physical contact is part of the normal operational routine of most biometric scanners. In the case of a fingerprint system, the user is required to touch the surface of the scanner to provide the biometric sample; the thumb must be directly placed in a certain manner for a specified duration to obtain a reliable measurement.

In the process of detecting and acquiring an electrical interpretation of the finger pattern, the structure of the fingertip (composed of ridges, valleys, and furrows) is scanned [18,19]. Directly touching the scanner surface with the thumb can potentially lead to exchange of particles; this can increase the chances of either acquiring or depositing harmful microorganisms, including pathogens that can spread bacterial, fungal, and viral infectious diseases, possibly including the novel COVID-19.

These germs may accumulate as deposits after each successive use, resulting in contamination of the surface of the fingerprint scanner [4]. According to Kramer and Assadian [20], contaminated surfaces can serve as important vectors for cross-transmission after hand contact, and in the case of frequent use, contamination can occur even more rapidly. This is particularly common in medical environments, where contaminated surfaces also play an important role in the transmission of hospital-acquired infections, also known as nosocomial infections [4].

Given their ability to retain microorganisms and germs, contaminated fingerprint scanner surfaces can become secondary reservoirs of infection that can in turn cross-contaminate food or water through hands or objects that come in contact with them [21].

Possible Disease Spread via Pathogenic Deposits on Fingerprint Scanner Surfaces

Microorganisms, including pathogens [22], can potentially be transmitted from animate sources to inanimate environmental sources through disease-carrying objects (fomites) or materials. As a result, pathogens can propagate among humans from touching communal surfaces in public places [3], including fingerprint scanner surfaces, where they can trigger the spread of infectious diseases. Contracting infectious diseases through fingerprint scanners is classified as an indirect means of transmission.

Depending on the infectious microorganism in question, and provided that the media or aerosol droplets remain alive on the surface, some pathogens can survive for only few seconds or minutes before dying. However, other pathogens can live on surfaces in wait for a potential new host for hours, days, weeks, or even months [23]; therefore, they remain a continuous source of transmission depending on frequency of use and whether regular preventive surface disinfection is performed. For example, studies [6,24] indicate that the novel coronavirus that causes COVID-19, SARS-CoV-2, can remain viable and infectious in aerosols for hours and on surfaces for up to days depending on the environmental temperature, humidity, and inoculum. Particularly, lower environmental temperatures increase the persistence of the COVID-19 pathogen and further prolong its survivability on external surfaces [6].

For these reasons, long-lasting pathogens found in body fluids and carried by fingertips are sources of concern in the administration of fingerprint scanners. Table 1 shows a partial listing of body fluids and secretions that can be deposited on or acquired from fingerprint scanners, alongside some of their commonly associated pathogenic isolates.

Table 1. List of body fluids and secretions and some commonly associated microbial isolates capable of causing disease.

Body fluid/secretion	Implicated pathogenic isolates
Sweat	*Staphylococcus* species [25,26], Ebola virus [27], HBV[a] [28]
Tears	Unconfirmed pathogenic association with tears, although chronic HBV has been found in the tears of children [29]
Saliva	SARS-CoV-2 [30,31], filoviruses [32,33], HBV [34], HSV[b] [35], *Escherichia coli* [36]
Nasal mucus fluid	SARS-CoV-2 [37], filoviruses [33], *Streptococcus* [38], *Haemophilus influenzae* [39,40], parainfluenza [41,42], *Mirabella catarrhal* [43], *Actinomyces* [44,45], *Mycoplasma pneumoniae* [46]
Pus	*Staphylococcus* [26], *Klebsiella* species [47-49], *Pseudomonas* species [50], *E coli* [36], *Proteus* species [51]
Hemoglobin/blood	SARS-CoV-2 [52], filoviruses [33], HBV [26,28]
Semen	*E coli* [53], hepatitis C [54,55], HSV [56], SARS-CoV-2 [57,58], Zika virus [59-61], *Ureaplasma urealyticum* [62,63], *Enterococcus faecalis* [64], alpha-hemolytic streptococci [46], Lassa virus [65], Marburg virus [66], Ebola virus [66], Chikungunya virus [56]
Vaginal swab	*Candida albicans* [67,68], *Trichomonas vaginalis* [69], *Neisseria gonorrhoeae* [70,71], yeast cells [72], Group B streptococci [38,46,73], *E coli* [74], *Klebsiella* [48], HSV [56], warts [46]
Urine	Yeast cells [72], adenovirus [75], BK virus [76,77], *Schistosoma* [78], *E coli* [79-81], *Klebsiella* [48], Zika virus [61], Coxsackievirus [82], *Proteus* [43], *Staphylococcus saprophyticus* [26], *Candida albicans* [67]
Feces	*Salmonella* [83], paratyphoid [41], *Entamoeba histolytica* [84]
Sputum	SARS-CoV-2 [85,86], *Haemophilus influenzae* [39], parainfluenza [41], *E coli* [87], *Streptococcus pneumoniae* [38], *Actinomyces* [46]
Nasal swabs	SARS-CoV-2 [30,31], Zika virus [61], *Haemophilus influenzae* [39], parainfluenza [41]
Urinary tract swabs	*Chlamydiatrachomatis* [88,89], *Mycoplasma* [90,91], HSV [92,93], *Streptococcus saprophyticus* [38], *Neisseria gonorrhoeae* [46,94], *Pseudomonas aeruginosa* [95]
Ear wax	Usually sterile, but yeast cells may be present [96]
Eye secretion	*Neisseria gonorrhoeae* [97], *Chlamydia trachomatis* [94], HSV [98], *Pseudomonas aeruginosa* [99,100], *Staphylococcus aureus* [101,102]
Vomit	Norovirus in viral gastroenteritis [103,104]
Secretion from wounds and sores	*Pseudomonas aeruginosa[105],Enterobacteriaceae* [105], HBV [28], HSV [26,51]

[a]HBV: hepatitis B virus.

[b]HSV: herpes simplex virus.

The various pathogenic isolates from common body fluids and secretions listed in Table 1 can potentially be retained on surfaces that are frequently touched; this raises major hygienic concerns, including possible transmission of respiratory viruses [106]. However, the transmissibility of these pathogens through contaminated biometric fingerprint surfaces needs to be specifically determined.

Sources of Fingerprint Scanner Surface Contamination

Essentially, persons who are nasal carriers of *Staphylococcus aureus*, whose hands are soiled with *S aureus*, or who exudate *S aureus* from skin lesions such as pimples, boils, infected cuts, and burns may contaminate fingerprint biometric surfaces through contact. An exposed injury that leaks pus-like or clear fluid composed of serum, fibrin, and white blood cells can also contaminate surfaces through contact. Contamination has also been shown to occur from persons in incubatory stages of hepatitis A infection, persons infected with norovirus or who have come in contact with excreta of persons with salmonellosis or cholera, and persons who are intestinal carriers of *Shigella* species [4,107-109].

Other sources of contamination include persons whose hands are soiled with raw poultry, pork, beef, and other meats contaminated with *Salmonella*, *Campylobacter jejuni*, *Yersinia enterocolitica*, *Clostridium perfringens*, *S aureus*, *Escherichia coli*, etc.

It should be noted, however, that the pathogenicity of any disease-causing microbe will depend not only on its transmissibility from a host (human user of the fingerprint device) to the reservoir (fingerprint device) and to a fresh host (another user of the device) but also on its survivability [107] in the new host, its ability to breach the host's immune system (infectivity), and its capacity to harm the host (virulence) [56].

Factors That Influence Infectious Disease Transmission via Direct Contact With Fingerprint Scanner Surfaces

Frequency of Use and Location of the Device

By virtue of their function, fingerprint scanners constitute frequently touched surfaces and are used by a large number of people on a continuous basis. The greater the number of people

who use a scanner, the higher the risk of contamination. The location of the devices also significantly affects the likelihood of contamination and disease transmission; for example, surfaces in hospital settings and microbiology or biomedical laboratories are more likely to be contaminated because of their proximity to patients and their body secretions [19] or samples [20], with a high probability of nosocomial infections.

Personal Hygiene of Users

Kramer and Assadian [20] showed that a single hand contact with a contaminated surface can result in different levels of potential microbe transfer from the surface to the hand. The risks of this cross-contamination of hands were shown to be higher with *E coli*, *Salmonella* species, *S aureus* (all 100%), *C albicans* (90%), rhinovirus (61%), hepatitis A virus (22% to 33%), and rotavirus (16%).

In addition to transmission through food or water, certain microbes spread through direct inoculation into the respiratory tract from contaminated hands. For example, influenza A [110] viruses, in addition to transmission through aerosol, are believed to spread between humans through contact and large respiratory droplets; these droplets can settle on the ground or other environmental surfaces within 1 to 2 meters [111], including nearby fingerprint biometric surfaces.

Pathogenic Potential of the Microorganism and Level of Infective Inoculum

The ability to contract disease from cross-contamination of food or water from the use of a contaminated biometric fingerprint scanner will depend on several factors, including the pathogenic potential or virulence of the microorganism [108]. This potential varies between microorganisms and is directly proportional to the size of the infective inoculum irrespective of whether this inoculum is a preformed toxin, such as in botulism and staphylococcal food poisoning, or requires direct microbial inoculation, such as *Salmonella* serovar Typhi or hepatitis A. Similarly, the immunity of the individual host plays an important role in the eventual development of the disease.

Extent of Microbial Survival on Environmental Surfaces

Several complex factors influence the survivability of microorganisms on environmental surfaces [20]; the longer the organisms persist on a surface, the longer the contaminated surface can remain a reservoir and a source of transmission to individuals who come in contact with the surfaces.

While the literature is limited on this topic, available studies have shown varying results. However, many bacteria, fungi, and viruses can persist for durations of hours to days and months on environmental surfaces depending on certain environmental conditions, including but not limited to temperature, humidity, and biofilm formation.

Notably, fungi can persist for a few days; other spore-forming organisms can persist for longer periods due to the thermotolerant properties of the spores [20]. It is currently uncertain how long SARS-CoV-2 survives on surfaces; however, preliminary studies [6,23,24,112,113] suggest that this novel coronavirus may survive on surfaces for periods ranging from a few hours to several days, weeks, or longer.

Relative Humidity, Temperature, and Biofilm Formation

Although the amount of moisture on users' fingers has an impact on overall biometric performance [114], the effects of humidity on the survival of microorganisms on environmental surfaces such as fingerprint scanners vary greatly between different classes of microorganisms and are significantly influenced by the presence of a cell wall or protective membrane.

While viruses with lipid envelopes, such as influenza virus or coronaviruses, tend to survive longer in low relative humidity (20% to 30%), some viruses survive longer in moist surfaces with higher relative humidity (70% to 90%). This is also true for different types of bacteria, such as gram-positive bacteria (eg, *Staphylococcus* species), which survive better at low humidity than gram-negative bacteria such as *E coli*. However, in some cases, the results inexplicably conflict, including the variability in fingerprint performance [115] with temperature and humidity [116].

Higher temperatures affect the integrity of viral proteins and enzymes as well as viral genomes (DNA or RNA) and will reduce the survival of viruses on surfaces.

Biofilms are a collection of two or more types of microorganisms (bacteria, fungi, and protists) that can grow on many different external surfaces, including on glass or plastic fingerprint scanner materials. A biofilm is an enclosure of microbial cells in environmental surfaces composed of an extracellular polymeric substance that protects the microorganisms from environmental influences [109] and can potentially prolong their survival, enabling them to live longer as contaminants on fingerprint scanner surfaces.

Intrinsic Properties of Surfaces

The physiochemical properties of surfaces define the rate and extent of attachment and the retention of microorganisms to form biofilms. Flagellate motile organisms attach to surfaces more readily than nonmotile organisms. According to Donlan [109], microbes attach more frequently to hydrophobic, nonpolar surfaces such as plastics than to hydrophilic materials such as glass and metals. In another study [117], surfaces lacking pores, such as plastics, were found to allow microorganisms to remain on the surface, thereby facilitating cross-contamination. Microorganisms entrapped in pores are more subject to lysis, particularly in the absence of water and nutrients.

Meanwhile, the upper surface of a fingerprint scanner is composed of a very small piece of plain glass or transparent plastic material that is used to capture the electrical equivalent of the finger pattern. The act of touching leaves an impression of the person's finger on the scanner surface, which the system interprets as an electrical representation of the user's unique finger pattern. It is a theoretical possibility that in the touch-dependent capture process, the scanner surface can also retain surviving microorganisms or germs (eg, bacteria, viruses, fungi, and protozoa) that can potentially contaminate the surface

with disease-causing pathogens that are transmissible from person to person through direct contact and serial use.

Other Influential Factors

Other subtle factors that favor transfer of microorganisms across fingerprint biometric system users are the large skin-surface contact of flat fingers, type of fingerprint-capturing device, nonporous contact surface, large overlap of users' contact areas on the surface, short turnaround time between successive users, and high contact pressure on the surface [106].

Recommendations

Universal Hygiene Measures

Universal hygiene measures are highly necessary to control disease transmission through fingerprint devices. These measures should be adopted by the facilities in which these scanners are used, particularly those in high-touch areas and in areas with close proximity to infectious patients, such as hospital settings and microbiology and biomedical laboratories. The adopted hygienic measures should ensure that the fingerprint systems are safely decontaminated for use in all circumstances on a continuous basis without compromising their integrity and biometric functionality.

For instance, while the use of gloves would be ideal for preventing contamination of a fingerprint device surface or cross-contamination of food and water [19], the glove obliterates the physiological patterns of the natural fingertips, rendering them unreadable and thus completely defeating the operational purpose of unique biometric identification. The use of gloves is therefore not a suitable means of controlling disease transmission by fingerprint systems.

Furthermore, some fingerprint biometric systems are capable of liveness detection [115,118-120], which is the ability to verify that the finger presented to the scanner is a natural, live human finger possessing all the attributes of vitality, including the presence of normal human temperature, normal pulse beat rate, and natural blood flow. With these systems, the use of gloves could operationally return a "fake finger detected" result due to the inanimate nature of the glove.

In keeping with global cybersecurity standards, measurement of liveness (ie, liveness detection) is a means to verify that the finger presented by a user is not a dummy [115], fake, cloned from a molded artefact, or cut from a corpse/cadaver. In a biometric spoofing attack [119,121], an impostor masquerades as an authentic user by replicating a legitimate user's biometrics (eg, with a counterfeit finger); therefore, the system could likely interpret the wearing of gloves as a typical biometric spoofing attack.

Table 2 indicates liveness parameters that can be measured on a fingerprint biometric system and demonstrates why the use of gloves to prevent infectious diseases can have a negative impact on the cybersecurity performance and overall accuracy of the fingerprint system and is thus discouraged.

Table 2. Effects of glove use for minimizing fingerprint contamination on the pattern obtained by a fingerprint biometric system.

Measurable liveness parameter	Biometric pattern	
	Expected	With hand in glove
Texture	Natural human body tissue	Latex or rubber
Temperature	37±0.5 ºC (human range)	Inconsistent
Pulse	60 to 100 beats per minute (human rate)	None
Pattern	Human finger ridges, valleys, and furrows [19,115]	Smooth
Blood flow	Evidence of natural human blood flow	None

Table 2 shows that on the basis of biometric functionality, a finger in a glove could be misinterpreted as a fake or spoofed finger; therefore, wearing gloves is completely unsuitable because it prevents the system from accurately detecting the biometric parameters of the real finger required for unique human identity verification and identification.

Clean the Scanner Surface Regularly

The surface of a fingerprint biometric device should be regularly decontaminated with disinfectants. The scanner surface should be cleaned after each use, either by designating a human assistant to manually disinfect it on a regular basis using a recommended cleaning agent with a damp nonabrasive cloth or a paper towel or by compelling each user to clean the scanner after use with a damp wipe soaked with an approved cleaning agent such as those contained in the US Environmental Protection Agency (EPA) [122] list of products for use against emerging viral pathogens.

In environments of heavy use, an automated spray and wipe mechanism can be set up to automatically scrub and disinfect the scanner surface between successive uses. Any combination of these measures will leave a safe fingerprint surface for subsequent users. Automating the postuse sanitization of fingerprint scanner surface may require additional investment, which will increase the total cost of biometric deployment and maintenance.

Use Alcohol-Based Cleaning Fluid

Alcohol-based cleaning agents should be used to enable rapid drying of the scanner surface through evaporation, as a wet scanner surface will introduce numerous errors. These errors include false negatives, where legitimate persons are wrongly denied access, and false positives, where unauthorized users are granted access. Both false negatives and false positives have cybersecurity implications in the management of fingerprint biometric systems in addition to user apathy, and they should be avoided. Furthermore, wet surfaces can encourage the growth

of pathogens and increase their survival on the device; thus, they should be avoided.

The recommendation of alcohol-based sanitizers is occasioned by the wide range of germ-killing efficacy of these sanitizers as well as their rapid evaporation without leaving wet surfaces, which can impact the biometric performance of the scanners and lead to operational failures.

Use Noncorrosive Cleaning Fluid

When choosing a suitable cleaning fluid for scanners, harsh products containing high levels of acidic constituents should be avoided. These products can corrode the scanner surface and alter its sensitivity, negatively impacting the biometric performance of the system with inaccurate readings including high probabilities of false positives (false acceptance) and false negatives (false rejection) [14,123].

The use of a highly corrosive cleansing agent (liquid, gel, or powder) is discouraged as a cleanser for the fingerprint scanner surface because the corrosive nature of the agent could damage the surface structure of the scanner, resulting in a decrease of its ability to function as a biometric device for validating users.

Enforce Finger Hygiene and Hand Sanitization Culture Through Education

It is essential to educate users on handwashing, cleaning of fingernails, and strict hygienic behaviors as well as the dangers of using a fingerprint scanning device with exposed skin and after contact with mouth, nose, or eye infections and uncovered wounds. This can prevent contamination of the scanner and avert cross-contamination of food or water [18,21,22,117].

Users should be compelled through frequent sensitization to use an antimicrobial substance that is effective in killing microorganisms or stopping their growth, such as alcohol-based hand sanitizers applied on the palm and thumbs before and after each use of the fingerprint device.

The EPA has provided a recommended list of approved tier 1 products [122] for cleaning touch devices and open surfaces. The products are in compliance with the EPA's Emerging Viral Pathogen Guidance for Antimicrobial Pesticides. Table 3 shows a partial listing of EPA-recommended cleaners for emerging enveloped viral pathogens, including SARS-CoV-2, which can be used for each type of finger scanner surface.

Table 3. List of recommended cleaners for scanner surfaces.

Surface type [124]	Composition	Recommended cleaners [122]	Justification
Nonporous	Glass, marble, metal, and plastic	Fiberlock IAQ 2500, CA-MRSA disinfectant spray, Clorox commercial solution hydrogen peroxide cleaner disinfectant	Contains noncorrosive chlorine, which does not alter the biometric functionality of the fingerprint scanner surface.
Porous	Fabric, unfinished wood, and paper	Clorox broad spectrum quaternary disinfectant cleaner, germicidal cleaner and disinfectant, detergent disinfectant pump spray	Antibacterial quaternary ammonium compounds with no harsh phenolics. Suitable for absorbent surfaces.
Textured	Rough, hard coating	Solucide hard surface disinfectant spray, Opti-Cide3 surface wipes, peroxide multisurface cleaner and disinfectant	The hydrogen peroxide constituent is very effective against fungi, viruses, bacteria, and other germs on rough and cracked surfaces.

Adjust the Scanner Sensitivity

The sensitivity of a biometric scanner determines how easily it can extract a fingerprint image or template from a user. A highly sensitive scanner has a tendency to reject user attempts, while a less sensitive scanner can be too porous; hence, a balance is necessary. The sensitivity of the fingerprint scanner should be adjusted to a level that minimizes the duration of contact required to read the finger sample and reduces the likelihood of germ transmission.

Deploy Multifactor Authentication

A multifactor authentication system includes an additional means of verification, such as a password or token. This system is beneficial to prevent overreliance on single-factor fingerprint authentication, which can be relaxed in emergency health situations such as the COVID-19 pandemic when the risk of propagating infectious diseases is high. By adding a second or third factor of authentication, an alternative system can be proactively prepared to switch modes in accordance with the demands of the moment.

Use an Alternative Means of Authentication

In the absence of multifactor authentication, other standalone but temporary means of authentication can be used, including password verification, personal identification numbers, and tokens.

Temporarily Disable Biometric Sign-In

For environments that rely on fingerprint biometrics as the only user access verification technique, temporary suspension and deactivation is highly recommended to enable the establishment of the preventive surface disinfection actions discussed above, particularly those that relate to the hygiene of the scanner surface. Biometric use can resume after these countermeasures have been established.

Use Contactless Biometric Technologies

Although many biometric systems require physical contact with the scanner to obtain the user's biometric image, research on completely contactless fingerprint scanners is ongoing at different stages among technology vendors. Contactless identification of fingerprints has gained considerable attention [125] for a number of reasons, including limiting the transmission and spread of touch-dependent infectious diseases.

While the prospects are high for fingerprint systems that can acquire biometric samples without compelling the user to touch the scanner surface, studies [17,126,127] have shown that the achievement of a highly accurate contactless fingerprint recognition system is challenged by possible delays and recognition errors. Research in this area is advancing, and some prototypes may soon be commercially available; therefore, contactless fingerprint biometric systems will hopefully address hygiene concerns by eliminating contamination from frequent direct touching.

Shifting to contactless biometric systems such as iris recognition appears to be an alternative approach. Despite cost considerations, iris recognition is a faster and more accurate biometric modality than fingerprint recognition. Iris cameras are technologically advanced and are suitable for use in health care, laboratory and medical research facilities, and other environments where poor hygiene exposes users to significant health risks. Iris recognition alternatives are also used in airports and military bases, where fast and reliable human identification is both an operational requirement and a security imperative. Other contactless options include vascular biometrics (finger vein and palm vein), voice biometrics, and gait recognition.

Limitations and Further Research

Potential Setbacks of Incorporating Automatic Sanitization in Fingerprint Systems

Incorporating the autosanitizer framework proposed in this work into pre-existing fingerprint biometric systems may be difficult to achieve. Future research and development efforts could include investigating ways to integrate automatic sanitization directly into the design of new fingerprint recognition systems in a vendor-neutral manner to ensure interoperability without impacting performance and convenience.

Extensive Study on Transmissibility of Isolated Pathogens

The list of microbial isolates that can be found in common body fluids and secretions as considered in this work appears to be nonexhaustive because of the limited literature and research associating the identified pathogens with known infectious diseases that can be transmitted via fingerprint scanners.

This creates a window of opportunity for focused research to investigate pathogenic linkages with infectious disease, including COVID-19. Research in this area is currently scarce but is certainly expedient and in high demand.

It is likely that a study in which each pathogenic isolate is matched to a corresponding infectious disease condition will become increasingly complex, with conflicting overlaps. This could lead to a shift of focus from fingerprint scanner surface contamination to disease mapping; this could be addressed by applying a phased research methodology in which each pathogen or group of microbial isolates is studied in isolation to facilitate research data interpretation.

Elucidation of Microbial Infectivity and Virulence in Relation to the Size of the Inoculum

While other complex factors are involved in the pathogenicity of diseases in users of biometric fingerprint scanner devices, further studies are needed to isolate specific pathogens on these devices, profile the common pathogens, identify their sources, and demonstrate the extent to which they cause disease in other users of the devices.

Infectious Disease Transmission via Other Touch-Based Devices

This study primarily focused on hygiene concerns related to fingerprint devices. It has therefore suggested further research opportunities to investigate the likelihood of transmission of the novel COVID-19 pathogen via other electronic surfaces and touch-based devices, including but not limited to ATM keypads, touch screen computer monitors, computer keyboards, touch screen phone buttons, elevator control keypads, machinery buttons, as well as shared surfaces including shopping carts, stairway rails, door handles, currency notes, table tops, fabrics, papers, cardboard, and plastics.

Conclusion

Principal Findings

Physical contact has been proven to be the most common source of surface contamination, depositing harmful microbial pathogens that can be directly inoculated into the respiratory tract of another person or cross-contaminate food and water, triggering infectious disease [111]. These disease-causing pathogens vary from bacteria and fungi to viruses, including SARS-CoV-2, the novel pathogen that causes COVID-19.

Fingerprint scanners are often repeatedly touched by multiple users, thereby raising serious issues of hygiene as their use exposes the risk of contracting and transmitting harmful diseases [19,128] due to direct finger vs scanner contact while extracting the biometric image [10]. These devices are increasingly hazardous, particularly when used in health, medical, or diagnostic facilities where aerosolized droplets and body secretions of sick patients can be easily deposited on them; thus, they may play an important role in the transmission of hospital-acquired infections [4].

While the physiochemical properties of the device material, the environmental conditions, and the survival ability as well as the virulence of the disease pathogen significantly contribute to the risk of disease transmission, several hygienic measures have been shown to be effective in controlling disease transmission from these surfaces. Other approaches and strategies can also be adopted to control and prevent transmission of infectious disease through use of fingerprint systems; however, these approaches must not interfere with the primary biometric performance of the systems. Poor hand washing after defecation and urination as well as contact with infected body fluids, foods, soil, or surfaces can lead to contamination of fingerprint scanner surfaces, which then serve as vectors of disease-carrying pathogens.

Because there are indications of prolonged longevity of the novel SARS-CoV-2 pathogen on certain external surfaces [6,112,113], there is a theoretical risk of transmitting COVID-19 via contaminated or unhygienic fingerprint surfaces because the pathogen may remain alive on the surfaces. In this paper, we proposed ideas for preventing contamination of fingerprint scanners with germs without compromising their primary biometric performance.

As a way of controlling the spread of COVID-19, this paper proposes the adoption of a defense-in-depth approach to managing fingerprint scanner cleanliness to reduce the propensity of these scanners to harbor pathogens that could lead to disease transmission. This approach involves the simultaneous application of recommended measures, all of which aim to make fingerprint systems more hygienic and safer to use while retaining their optimal functions as a biometric verification technique. Beyond the present COVID-19 pandemic, a defense-in-depth approach to managing fingerprint cleanliness will improve overall biometric acceptability and minimize user apathy related hygiene.

Acknowledgments

The logistic support of the management of the NHIS, the Federal Capital Territory Administration (FCTA), and the ECWA Comprehensive Medical Centre, all in Abuja, are appreciated. The authors value the immense input from the NHIS Department of Information and Communications Technology. The conclusions and views expressed in this paper are those of the authors and do not necessarily represent the official positions of the NHIS, the FCTA, or the ECWA Comprehensive Medical Centre. Names of specific vendors, manufacturers, or products, wherever expressly mentioned or implied, are for public health educational and informational purposes only; implicit or expressed inclusion does not suggest endorsement of the vendors, manufacturers, or products by the NHIS, FCTA, or ECWA Comprehensive Medical Centre.

Conflicts of Interest

None declared.

References

1. Mir GM, Balkhi AA, Lala NA, Sofi NA, Kirmani MM, Mir IA, et al. The Benefits of Implementation of Biometric Attendance System. Orient J Comp Sci and Technol 2018 Mar 22;11(1):50-54. [doi: 10.13005/ojcst11.01.09]
2. Sano E, Maeda T, Nakamura M, Shikai K, Sakata M, Matsushita M, et al. Fingerprint Authentication Device Based on Optical Characteristics Inside a Finger. 2006 Presented at: 2006 Conference on Computer Vision and Pattern Recognition Workshop; June 17-22, 2006; New York, NY. [doi: 10.1109/cvprw.2006.83]
3. Ibfelt T, Engelund EH, Permin A, Madsen JS, Schultz AC, Andersen LP. Presence of Pathogenic Bacteria and Viruses in the Daycare Environment. J Environ Health 2015 Oct;78(3):24-29. [Medline: 26591334]
4. Otter JA, Yezli S, Salkeld JA, French GL. Evidence that contaminated surfaces contribute to the transmission of hospital pathogens and an overview of strategies to address contaminated surfaces in hospital settings. Am J Infect Control 2013 May;41(5 Suppl):S6-11. [doi: 10.1016/j.ajic.2012.12.004] [Medline: 23622751]
5. Abrampah NM, Montgomery M, Baller A, Ndivo F, Gasasira A, Cooper C, et al. Improving water, sanitation and hygiene in health-care facilities, Liberia. Bull World Health Organ 2017 Apr 25;95(7):526-530. [doi: 10.2471/blt.16.175802]
6. Kampf G, Todt D, Pfaender S, Steinmann E. Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. J Hosp Infect 2020 Mar;104(3):246-251 [FREE Full text] [doi: 10.1016/j.jhin.2020.01.022] [Medline: 32035997]
7. Raut M, Kokate S, Shinde S, Karpe S, Barahate S. Virtual Biometric Fingerprint Attendance System. IJSRD 2015;3(1):511-513.
8. Muhtahir OO, Adeyinka AO, Kayode AS. Fingerprint Biometric Authentication for Enhancing Staff Attendance System. IJAIS 2013 Feb 08;5(3):19-24. [doi: 10.5120/ijais12-450867]
9. Verma M, Khan N. A Study on Benefits of Biometrics Attendance System: A Technological based Human Resource Management Practice. TMIMT Int J 2016;Special Issue [FREE Full text]
10. Naveed G, Batool R. Biometric Authentication in Cloud Computing. J Biom Biostat 2015;06(05). [doi: 10.4172/2155-6180.1000258]
11. El-Sisi A. Design and Implementation Biometric Access Control System Using Fingerprint for Restricted Area Based on Gabor Filter. Int Arab J Inf Technol 2011;8(4):355-363.
12. Shakil M, Nandi RN. Attendance Management System for Industrial Worker using Finger Print Scanner. GJCST-F 2013;13(6):16-22.
13. Walia H, Jain N. Fingerprint Based Attendance Systems - A Review. Int Res J Eng Tech 2016;3(5):1166-1171.
14. Rajarajan S, Palanivel S, Sekar KR, Arunkumar S. Study on the diseases and deformities causing false rejections for fingerprint authentication. Int J Pure Appl Math 2018;119(15):443-454.
15. Malčík D, Drahanský M. Anatomy of biometric passports. J Biomed Biotechnol 2012;2012:490362 [FREE Full text] [doi: 10.1155/2012/490362] [Medline: 22969272]

16. Doc 9303, Machine Readable Travel Documents, 7th Edition. International Civil Aviation Organization. 2015. URL: https://www.icao.int/publications/pages/publication.aspx?docnum=9303
17. Yin X, Zhu Y, Hu J. Contactless Fingerprint Recognition Based on Global Minutia Topology and Loose Genetic Algorithm. IEEE Trans Inform Forensic Secur 2020;15:28-41. [doi: 10.1109/tifs.2019.2918083]
18. Drahansky M, Dolezel M, Urbanek J, Brezinova E, Kim T. Influence of skin diseases on fingerprint recognition. J Biomed Biotechnol 2012;2012:626148 [FREE Full text] [doi: 10.1155/2012/626148] [Medline: 22654483]
19. Dolezel M, Drahansky M, Urbanek J, Brezinova E, Kim TH. Influence of Skin Diseases on Fingerprint Quality and Recognition. In: Yang J, Xie SJ, editors. New Trends and Developments in Biometrics. London, UK: IntechOpen; 2012:275.
20. Kramer A, Assadian O. Survival of Microorganisms on Inanimate Surfaces. In: Borkow G, editor. Use of Biocidal Surfaces for Reduction of Healthcare Acquired Infections. Cham, Switzerland: Springer; 2014.
21. Foodborne Disease Outbreaks: Guidelines for investigation and Control. World Health Organization. 2008. URL: https://www.who.int/foodsafety/publications/foodborne_disease/outbreak_guidelines.pdf?ua=1 [accessed 2020-04-12]
22. Control of Communicable Diseases Manual, 17th Edition. American Public Health Association. 2000. URL: https://www.ciphi.ca/hamilton/Content/documents/ccdm.pdf
23. Kramer A, Schwebke I, Kampf G. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. BMC Infect Dis 2006 Aug 16;6:130 [FREE Full text] [doi: 10.1186/1471-2334-6-130] [Medline: 16914034]
24. van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. N Engl J Med 2020 Apr 16;382(16):1564-1567 [FREE Full text] [doi: 10.1056/nejmc2004973]
25. Kawamura Y, Hou XG, Sultana F, Hirose K, Miyake M, Shu SE, et al. Distribution of Staphylococcus species among human clinical specimens and emended description of Staphylococcus caprae. J Clin Microbiol 1998 Jul;36(7):2038-2042 [FREE Full text] [doi: 10.1128/JCM.36.7.2038-2042.1998] [Medline: 9650958]
26. Pishchany G, McCoy AL, Torres VJ, Krause JC, Crowe JE, Fabry ME, et al. Specificity for human hemoglobin enhances Staphylococcus aureus infection. Cell Host Microbe 2010 Dec 16;8(6):544-550 [FREE Full text] [doi: 10.1016/j.chom.2010.11.002] [Medline: 21147468]
27. Ebola (Ebola Virus Disease). US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/vhf/ebola/index.html [accessed 2020-04-14]
28. Hepatitis B. Missouri Department of Health & Senior Services. URL: https://health.mo.gov/living/healthcondiseases/communicable/hepatitisb/ [accessed 2020-04-14]
29. Komatsu H, Inui A, Sogo T, Tateno A, Shimokawa R, Fujisawa T. Tears from children with chronic hepatitis B virus (HBV) infection are infectious vehicles of HBV transmission: experimental transmission of HBV by tears, using mice with chimeric human livers. J Infect Dis 2012 Aug 15;206(4):478-485. [doi: 10.1093/infdis/jis293] [Medline: 22508939]
30. Khurshid Z, Asiri FYI, Al Wadaani H. Human Saliva: Non-Invasive Fluid for Detecting Novel Coronavirus (2019-nCoV). Int J Environ Res Public Health 2020 Mar 26;17(7):2225 [FREE Full text] [doi: 10.3390/ijerph17072225] [Medline: 32224986]
31. Peng X, Xu X, Li Y, Cheng L, Zhou X, Ren B. Transmission routes of 2019-nCoV and controls in dental practice. Int J Oral Sci 2020 Mar 03;12(1):9 [FREE Full text] [doi: 10.1038/s41368-020-0075-9] [Medline: 32127517]
32. Dovih P, Laing ED, Chen Y, Low DHW, Ansil BR, Yang X, et al. Filovirus-reactive antibodies in humans and bats in Northeast India imply zoonotic spillover. PLoS Negl Trop Dis 2019 Oct;13(10):e0007733-e0007710 [FREE Full text] [doi: 10.1371/journal.pntd.0007733] [Medline: 31671094]
33. Emperador DM, Mazzola LT, Wonderly Trainor B, Chua A, Kelly-Cirino C. Diagnostics for filovirus detection: impact of recent outbreaks on the diagnostic landscape. BMJ Glob Health 2019 Feb 07;4(Suppl 2):e001112 [FREE Full text] [doi: 10.1136/bmjgh-2018-001112] [Medline: 30899573]
34. Villar L, Medina H, Villela-Nogueira C, Nabuco L, Rodrigues do Ó K, da Silva E, et al. Saliva as a Source for Hpatitis B Virus Diagnosis. J Hepatol 2010 Apr;52:S248 [FREE Full text] [doi: 10.1016/s0168-8278(10)60636-8]
35. Zuo Y, Whitbeck JC, Haila GJ, Hakim AA, Rothlauf PW, Eisenberg RJ, et al. Saliva enhances infection of gingival fibroblasts by herpes simplex virus 1. PLoS One 2019;14(10):e0223299 [FREE Full text] [doi: 10.1371/journal.pone.0223299] [Medline: 31581238]
36. Lim JY, Yoon J, Hovde CJ. A Brief Overview of Escherichia coli O157:H7 and Its Plasmid O157. J Microbiol Biotechn 2010 Jan 28;20(1):5-14. [doi: 10.4014/jmb.0908.08007]
37. Chilvers M, McKean M, Rutman A, Myint B, Silverman M, O'Callaghan C. The effects of coronavirus on human nasal ciliated respiratory epithelium. Eur Respir J 2001 Dec 01;18(6):965-970 [FREE Full text] [doi: 10.1183/09031936.01.00093001] [Medline: 11829103]
38. Azike CA, Agi VN, Akere BB. The Prevalence of Group A Streptococcus as a Re-Emerging Microorganism in Port Harcourt Metropolis. IJPR 2019 Apr 27:1-6. [doi: 10.9734/ijpr/2019/v2i330072]
39. Agrawal A, Murphy TF. Haemophilus influenzae Infections in the H. influenzae Type b Conjugate Vaccine Era. J Clin Microbiol 2011 Sep 07;49(11):3728-3732. [doi: 10.1128/jcm.05476-11]
40. Slack MPE. A review of the role of in community-acquired pneumonia. Pneumonia 2015 Jun 29;6(1):26-43 [FREE Full text] [doi: 10.15172/pneu.2015.6/520] [Medline: 31641576]

41. Henrickson KJ. Parainfluenza viruses. Clin Microbiol Rev 2003 May;16(2):242-264 [FREE Full text] [doi: 10.1128/cmr.16.2.242-264.2003] [Medline: 12692097]
42. Mao N, Ji Y, Xie Z, Wang H, Wang H, An J, et al. Human parainfluenza virus-associated respiratory tract infection among children and genetic analysis of HPIV-3 strains in Beijing, China. PLoS One 2012;7(8):e43893 [FREE Full text] [doi: 10.1371/journal.pone.0043893] [Medline: 22937119]
43. Schaffer JN, Pearson MM. Proteus mirabilis and Urinary Tract Infections. Microbiol Spectr 2015 Oct;3(5) [FREE Full text] [doi: 10.1128/microbiolspec.UTI-0017-2013] [Medline: 26542036]
44. Kim S, Kim D, Choi K, Cho K. Nasal Cavity Actinomycosis Mimicking Rhinolith. J Craniofac Surg 2018 May;29(3):e255-e257. [doi: 10.1097/SCS.0000000000004304] [Medline: 29461370]
45. Ferry T, Valour F, Karsenty J, Breton P, Gleizal A, Braun E, et al. Actinomycosis: etiology, clinical features, diagnosis, treatment, and management. Infect Drug Resist 2014 Jul:183 [FREE Full text] [doi: 10.2147/idr.s39601]
46. Van TT, Mata K, Dien Bard J. Automated Detection of Pharyngitis by Use of Colorex Strep A CHROMagar and WASPLab Artificial Intelligence Chromogenic Detection Module Software. J Clin Microbiol 2019 Oct 23;57(11):1-21. [doi: 10.1128/jcm.00811-19]
47. Kumar AR. Antimicrobial sensitivity pattern of Klebsiella pneumonia isolated from pus from tertiary care hospital and issues related to the rational selection of antimicrobials. J Chem Pharm 2013;5(11):326-311 [FREE Full text]
48. Piperaki E, Syrogiannopoulos GA, Tzouvelekis LS, Daikos GL. Klebsiella pneumoniae: Virulence, Biofilm and Antimicrobial Resistance. Pediatr Infect Dis J 2017 Oct;36(10):1002-1005. [doi: 10.1097/INF.0000000000001675] [Medline: 28914748]
49. Patilaya P, Husori DI, Marhafanny L. Susceptibility of Isolated from Pus Specimens of Post-Surgery Patients in Medan, Indonesia to Selected Antibiotics. Open Access Maced J Med Sci 2019 Dec 30;7(22):3861-3864 [FREE Full text] [doi: 10.3889/oamjms.2019.520] [Medline: 32127992]
50. Yilmaz S, Saklamaz A, Maden A. Pseudomonas keratitis. Ophthalmology 2006 May;113(5):883-884. [doi: 10.1016/j.ophtha.2006.02.009] [Medline: 16650685]
51. Khanam RA, Islam MR, Sharif A, Parveen R, Sharmin I, Yusuf MA. Bacteriological Profiles of Pus with Antimicrobial Sensitivity Pattern at a Teaching Hospital in Dhaka City. Bangladesh J Infect Dis 2018 Aug 04;5(1):10-14. [doi: 10.3329/bjid.v5i1.37710]
52. Chang L, Yan Y, Wang L. Coronavirus Disease 2019: Coronaviruses and Blood Safety. Transfus Med Rev 2020 Apr;34(2):75-80 [FREE Full text] [doi: 10.1016/j.tmrv.2020.02.003] [Medline: 32107119]
53. Prabha V, Sandhu R, Kaur S, Kaur K, Sarwal A, Mavuduru RS, et al. Mechanism of sperm immobilization by Escherichia coli. Adv Urol 2010;2010:240268 [FREE Full text] [doi: 10.1155/2010/240268] [Medline: 20379358]
54. Cavalheiro NDP, Santos ACDO, Melo CE, Morimitsu SR, Barone AA. Hepatitis C virus detection in the semen of infected patients. Braz J Infect Dis 2008 Oct;12(5):358-361. [doi: 10.1590/s1413-86702008000500003] [Medline: 19219272]
55. Turner SS, Gianella S, Yip MJS, van Seggelen WO, Gillies RD, Foster AL, et al. Shedding of Hepatitis C Virus in Semen of Human Immunodeficiency Virus-Infected Men. Open Forum Infect Dis 2016 Mar;3(2):ofw057 [FREE Full text] [doi: 10.1093/ofid/ofw057] [Medline: 27186582]
56. Ala'Aldeen DAA, Wooldridge KG. Bacterial Pathogenicity. In: Greenwood D, Barer M, Slack R, Irving W, editors. Medical Microbiology, 18th Edition. Nottingham, UK: Elsevier; 2012:156-157.
57. Paoli D, Pallotti F, Colangelo S, Basilico F, Mazzuti L, Turriziani O, et al. Study of SARS-CoV-2 in semen and urine samples of a volunteer with positive naso-pharyngeal swab. J Endocrinol Invest 2020 May 23 [FREE Full text] [doi: 10.1007/s40618-020-01261-1] [Medline: 32329026]
58. Holtmann N, Edimiris P, Andree M, Doehmen C, Baston-Buest D, Adams O, et al. Assessment of SARS-CoV-2 in human semen-a cohort study. Fertil Steril 2020 Aug;114(2):233-238 [FREE Full text] [doi: 10.1016/j.fertnstert.2020.05.028] [Medline: 32650948]
59. Mead PS, Duggal NK, Hook SA, Delorey M, Fischer M, Olzenak McGuire D, et al. Zika Virus Shedding in Semen of Symptomatic Infected Men. N Engl J Med 2018 Apr 12;378(15):1377-1385. [doi: 10.1056/nejmoa1711038]
60. Medina FA, Torres G, Acevedo J, Fonseca S, Casiano L, De León-Rodríguez CM, et al. Duration of the Presence of Infectious Zika Virus in Semen and Serum. J Infect Dis 2019 Jan 01;219(1):31-40. [doi: 10.1093/infdis/jiy462] [Medline: 30059980]
61. Musso D, Gubler DJ. Zika Virus. Clin Microbiol Rev 2016 Mar 30;29(3):487-524. [doi: 10.1128/cmr.00072-15]
62. Pajovic B, Radojevic N, Vukovic M, Stjepcevic A. Semen analysis before and after antibiotic treatment of asymptomatic Chlamydia- and Ureaplasma-related pyospermia. Andrologia 2013 Aug 16;45(4):266-271. [doi: 10.1111/and.12004] [Medline: 22897222]
63. Zhou YH, Ma HX, Shi XX, Liu Y. Ureaplasma spp. in male infertility and its relationship with semen quality and seminal plasma components. J Microbiol Immunol Infect 2018 Dec;51(6):778-783 [FREE Full text] [doi: 10.1016/j.jmii.2016.09.004] [Medline: 28739435]
64. Grande G, Vincenzoni F, Mancini F, Baroni S, Luca G, Calafiore R, et al. Semen Proteomics Reveals the Impact of Enterococcus faecalis on male Fertility. Protein Pept Lett 2018 Jul 03;25(5):472-477. [doi: 10.2174/0929866525666180412161818] [Medline: 29651938]

65. Bausch DG, Rollin PE, Demby AH, Coulibaly M, Kanu J, Conteh AS, et al. Diagnosis and clinical virology of Lassa fever as evaluated by enzyme-linked immunosorbent assay, indirect fluorescent-antibody test, and virus isolation. J Clin Microbiol 2000 Jul;38(7):2670-2677 [FREE Full text] [doi: 10.1128/JCM.38.7.2670-2677.2000] [Medline: 10878062]
66. Brainard J, Pond K, Hooper L, Edmunds K, Hunter P. Presence and Persistence of Ebola or Marburg Virus in Patients and Survivors: A Rapid Systematic Review. PLoS Negl Trop Dis 2016 Mar;10(2):e0004475 [FREE Full text] [doi: 10.1371/journal.pntd.0004475] [Medline: 26927697]
67. Dadar M, Tiwari R, Karthik K, Chakraborty S, Shahali Y, Dhama K. Candida albicans - Biology, molecular characterization, pathogenicity, and advances in diagnosis and control - An update. Microb Pathog 2018 May;117:128-138. [doi: 10.1016/j.micpath.2018.02.028] [Medline: 29454824]
68. Kabir MA, Hussain MA, Ahmad Z. Candida albicans: A Model Organism for Studying Fungal Pathogens. ISRN Microbiol 2012;2012:538694 [FREE Full text] [doi: 10.5402/2012/538694] [Medline: 23762753]
69. Leitsch D. Recent Advances in the Trichomonas vaginalis Field. F1000Res 2016 Feb 11;5:162 [FREE Full text] [doi: 10.12688/f1000research.7594.1] [Medline: 26918168]
70. Piszczek J, St Jean R, Khaliq Y. Gonorrhea: Treatment update for an increasingly resistant organism. Can Pharm J (Ott) 2015 Mar 10;148(2):82-89 [FREE Full text] [doi: 10.1177/1715163515570111] [Medline: 25918540]
71. Ngonzi J, Bebell LM, Bazira J, Fajardo Y, Nyehangane D, Boum Y, et al. Risk Factors for Vaginal Colonization and Relationship between Bacterial Vaginal Colonization and In-Hospital Outcomes in Women with Obstructed Labor in a Ugandan Regional Referral Hospital. Int J Microbiol 2018;2018:6579139 [FREE Full text] [doi: 10.1155/2018/6579139] [Medline: 30327672]
72. Janssens GE, Veenhoff LM. The Natural Variation in Lifespans of Single Yeast Cells Is Related to Variation in Cell Size, Ribosomal Protein, and Division Time. PLoS One 2016;11(12):e0167394 [FREE Full text] [doi: 10.1371/journal.pone.0167394] [Medline: 27907085]
73. Krzyściak W, Pluskwa KK, Jurczak A, Kościelniak D. The pathogenicity of the Streptococcus genus. Eur J Clin Microbiol Infect Dis 2013 Dec 3;32(11):1361-1376 [FREE Full text] [doi: 10.1007/s10096-013-1914-9] [Medline: 24141975]
74. Sáez-López E, Cossa A, Benmessaoud R, Madrid L, Moraleda C, Villanueva S, et al. Characterization of Vaginal Escherichia coli Isolated from Pregnant Women in Two Different African Sites. PLoS One 2016 Jul 7;11(7):e0158695 [FREE Full text] [doi: 10.1371/journal.pone.0158695] [Medline: 27387665]
75. Lion T. Adenovirus Infections in Immunocompetent and Immunocompromised Patients. Clin Microbiol Rev 2014 Jun 30;27(3):441-462. [doi: 10.1128/cmr.00116-13]
76. Sawinski D, Trofe-Clark J. BK Virus Nephropathy. CJASN 2018 Sep 21;13(12):1893-1896. [doi: 10.2215/cjn.04080318]
77. Krajewski W, Kamińska D, Poterek A, Małkiewicz B, Kłak J, Zdrojowy R, et al. Pathogenicity of BK virus on the urinary system. Cent European J Urol 2020;73(1):94-103 [FREE Full text] [doi: 10.5173/ceju.2020.0034] [Medline: 32395331]
78. Tzanetou K, Adamis G, Andipa E, Zorzos C, Ntoumas K, Armenis K, et al. Urinary tract Schistosoma haematobium infection: a case report. J Travel Med 2007;14(5):334-337 [FREE Full text] [doi: 10.1111/j.1708-8305.2007.00137.x] [Medline: 17883465]
79. Lee DS, Lee S, Choe H. Community-Acquired Urinary Tract Infection by in the Era of Antibiotic Resistance. Biomed Res Int 2018;2018:7656752 [FREE Full text] [doi: 10.1155/2018/7656752] [Medline: 30356438]
80. Luna-Pineda VM, Ochoa SA, Cruz-Córdova A, Cázares-Domínguez V, Reyes-Grajeda JP, Flores-Oropeza MA, et al. Features of urinary Escherichia coli isolated from children with complicated and uncomplicated urinary tract infections in Mexico. PLoS ONE 2018 Oct 4;13(10):e0204934. [doi: 10.1371/journal.pone.0204934]
81. Eljamay SM. Escherichia Coli Bacteria Infection in Females Urinary Tract. Jacobs Journal of Nephrology and Urology 2019 Feb 28 [FREE Full text]
82. Wang S, Wang A, Liu P, Zhang W, Du J, Xu S, et al. Divergent Pathogenic Properties of Circulating Coxsackievirus A6 Associated with Emerging Hand, Foot, and Mouth Disease. J Virol 2018 May 14;92(11). [doi: 10.1128/jvi.00303-18]
83. Giannella RA. Salmonella. In: Baron S, editor. Medical Microbiology, 4th Edition. Galveston, TX: University of Texas Medical Branch at Galveston; 1996.
84. Nowak P. Entamoeba histolytica - Pathogenic Protozoan of the Large Intestine in Humans. J Clin Microbiol Biochem Technol 2015 Dec 30:010-017. [doi: 10.17352/jcmbt.000003]
85. Qu Y, Kang E, Cong H. Positive result of Sars-Cov-2 in sputum from a cured patient with COVID-19. Travel Med Infect Dis 2020;34:101619 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101619] [Medline: 32160971]
86. Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA 2020 Mar 11;323(18):1843-1844 [FREE Full text] [doi: 10.1001/jama.2020.3786] [Medline: 32159775]
87. Saha AK, Dhar P. Spectrum of Antimicrobial Sensitivity of Escherichia Coli in Sputum in a Tertiary Medical Centre in Kolkata, West Bengal, 7 Years' Experience. Int J Contemp Med Res 2017;4(10):2177.
88. Giffard PM, Lilliebridge RA, Wilson J, Murray G, Phillips S, Tabrizi SN, et al. Contaminated fingers: a potential cause of positive urine specimens. Sex Transm Infect 2018 Feb;94(1):32-36 [FREE Full text] [doi: 10.1136/sextrans-2016-053081] [Medline: 28600332]

89. Thielemans E, Wyndham-Thomas C, Henrard S, De Vleeschouwer A, Steensels D, Montesinos I, et al. Screening for Chlamydia trachomatis and Neisseria gonorrhoeae Infections in Men Who Have Sex With Men. Sex Transm Dis 2018;45(3):195-198. [doi: 10.1097/olq.0000000000000722]
90. Jensen J, Cusini M, Gomberg M, Moi H. Background review for the 2016 European guideline on Mycoplasma genitalium infections. J Eur Acad Dermatol Venereol 2016 Oct;30(10):1686-1693 [FREE Full text] [doi: 10.1111/jdv.13850] [Medline: 27605499]
91. Combaz-Söhnchen N, Kuhn A. A Systematic Review of Mycoplasma and Ureaplasma in Urogynaecology. Geburtshilfe Frauenheilkd 2017 Dec;77(12):1299-1303 [FREE Full text] [doi: 10.1055/s-0043-119687] [Medline: 29269957]
92. Hanaoka N, Ito S, Konagaya M, Nojiri N, Yasuda M, Fujimoto T, et al. Infectious human adenoviruses are shed in urine even after disappearance of urethral symptoms. PLoS One 2019;14(3):e0212434 [FREE Full text] [doi: 10.1371/journal.pone.0212434] [Medline: 30840641]
93. Parra-Sánchez M, Marcuello López A, García-Rey S, Zakariya-Yousef Breval I, Bernal Martínez S, Pueyo Rodríguez I, et al. Performance of the HSV OligoGen kit for the diagnosis of herpes simplex virus type 1 and 2. Diagn Microbiol Infect Dis 2016 Jul;85(3):315-317 [FREE Full text] [doi: 10.1016/j.diagmicrobio.2016.04.019] [Medline: 27185644]
94. Khattab R, Abdelfattah MM. Study of the prevalence and association of ocular chlamydial conjunctivitis in women with genital infection by Chlamydia trachomatis, Mycoplasma genitalium and Candida albicans attending outpatient clinic. Int J Ophthalmol 2016;9(8):1176-1186 [FREE Full text] [doi: 10.18240/ijo.2016.08.15] [Medline: 27588273]
95. Bashir Ahmed O. Incidence and Antibiotic Susceptibility Pattern of Pseudomonas aeruginosa Isolated from Inpatients in Two Tertiary Hospitals. Clin Microbiol 2016;05(02). [doi: 10.4172/2327-5073.1000248]
96. Megarry S, Pett A, Scarlett A, Teh W, Zeigler E, Canter R. The activity against yeasts of human cerumen. J Laryngol Otol 2007 Jun 29;102(8):671-672. [doi: 10.1017/S0022215100106115]
97. Adamson P, Judson SD, Klausner JD, Kelesidis T. Neisseria gonorrhoeae as a Rare Cause of Preseptal Cellulitis. Sex Transm Dis 2019;46(12):813-815 [FREE Full text] [doi: 10.1097/olq.0000000000001055]
98. Getahun E, Gelaw B, Assefa A, Assefa Y, Amsalu A. Bacterial pathogens associated with external ocular infections alongside eminent proportion of multidrug resistant isolates at the University of Gondar Hospital, northwest Ethiopia. BMC Ophthalmol 2017 Aug 22;17(1):151 [FREE Full text] [doi: 10.1186/s12886-017-0548-6] [Medline: 28830451]
99. Javiya VA, Ghatak SB, Patel KR, Patel K. Antibiotic susceptibility patterns of Pseudomonas aeruginosa at a tertiary care hospital in Gujarat, India. Indian J Pharmacol 2008 Oct;40(5):230-234 [FREE Full text] [doi: 10.4103/0253-7613.44156] [Medline: 20040963]
100. Warghane AJ, Wagh GN, Nag BBSP, Jisnani ML, Thaware RR, Kitey HS. Isolation and characterization of Pseudomonas species from Godavari river sample. Asiat J Biotechnol Resour 2011;2(07):862-866.
101. O'Callaghan RJ. The Pathogenesis of Staphylococcus aureus Eye Infections. Pathogens 2018 Jan 10;7(1):2018 [FREE Full text] [doi: 10.3390/pathogens7010009] [Medline: 29320451]
102. Getahun E, Gelaw B, Assefa A, Assefa Y, Amsalu A. Bacterial pathogens associated with external ocular infections alongside eminent proportion of multidrug resistant isolates at the University of Gondar Hospital, northwest Ethiopia. BMC Ophthalmol 2017 Aug 22;17(1):151 [FREE Full text] [doi: 10.1186/s12886-017-0548-6] [Medline: 28830451]
103. How Norovirus Spreads. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/norovirus/about/transmission.html [accessed 2020-04-14]
104. Makison Booth C, Frost G. Potential distribution of viable norovirus after simulated vomiting. J Hosp Infect 2019 Jul;102(3):304-310 [FREE Full text] [doi: 10.1016/j.jhin.2019.02.010] [Medline: 30797885]
105. Machado C, Teixeira S, Fonseca L, Abreu M, Carvalho A, Pereira MT, et al. Evolutionary trends in bacteria isolated from moderate and severe diabetic foot infections in a Portuguese tertiary center. Diabetes Metab Syndr 2020;14(3):205-209. [doi: 10.1016/j.dsx.2020.02.010] [Medline: 32171163]
106. Jacobs JA, Van Ranst M. Biometric fingerprinting for visa application: device and procedure are risk factors for infection transmission. J Travel Med 2008;15(5):335-343 [FREE Full text] [doi: 10.1111/j.1708-8305.2008.00232.x] [Medline: 19006507]
107. Blomeke CR, Elliott SJ, Walter TM. Bacterial Survivability and Transferability on Biometric Devices. 2007 Presented at: 41st Annual IEEE International Carnahan Conference on Security Technology; October 8-11, 2007; Ottawa, ON p. 80-84. [doi: 10.1109/ccst.2007.4373472]
108. Casadevall A. The Pathogenic Potential of a Microbe. mSphere 2017 Feb 22;2(1):1-7. [doi: 10.1128/msphere.00015-17]
109. Donlan RM. Biofilms: microbial life on surfaces. Emerg Infect Dis 2002 Oct;8(9):881-890 [FREE Full text] [doi: 10.3201/eid0809.020063] [Medline: 12194761]
110. Cowling BJ, Ip DKM, Fang VJ, Suntarattiwong P, Olsen SJ, Levy J, et al. Aerosol transmission is an important mode of influenza A virus spread. Nat Commun 2013 Jun 4;4(1):1935 [FREE Full text] [doi: 10.1038/ncomms2922] [Medline: 23736803]
111. Touchless Solutions from TBS. TBS Biometrics. 2020 Mar 09. URL: https://www.tbs-biometrics.com/en/touchless-solutions-from-tbs.html [accessed 2020-04-12]

112. Ikonen N, Savolainen-Kopra C, Enstone JE, Kulmala I, Pasanen P, Salmela A, PANDHUB consortium. Deposition of respiratory virus pathogens on frequently touched surfaces at airports. BMC Infect Dis 2018 Aug 29;18(1):437 [FREE Full text] [doi: 10.1186/s12879-018-3150-5] [Medline: 30157776]
113. Kampf G. Potential role of inanimate surfaces for the spread of coronaviruses and their inactivation with disinfectant agents. IPIP 2020 Jun;2(2):100044. [doi: 10.1016/j.infpip.2020.100044]
114. Olsen MA, Dusio M, Busch C. Fingerprint skin moisture impact on biometric performance. 2015 Presented at: IEEE 3rd International Workshop on Biometrics and Forensics (IWBF); March 3-4, 2015; Gjovik, Norway p. 1-6. [doi: 10.1109/iwbf.2015.7110223]
115. van der Putte T, Keuning J. Biometrical Fingerprint Recognition: Don't Get Your Fingers Burned. In: Domingo-Ferrer J, Chan D, Watson A, editors. Smart Card Research and Advanced Applications. Boston, MA: Springer; 2000:289-303.
116. Stewart RF, Estevao M, Adler A. Fingerprint recognition performance in rugged outdoors and cold weather conditions. 2009 Presented at: 009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems; September 28-30, 2009; Washington, DC p. 1-6. [doi: 10.1109/btas.2009.5339061]
117. Siroli L, Patrignani F, Serrazanetti DI, Chiavari C, Benevelli M, Grazia L, et al. Survival of Spoilage and Pathogenic Microorganisms on Cardboard and Plastic Packaging Materials. Front Microbiol 2017 Dec 22;8:2606 [FREE Full text] [doi: 10.3389/fmicb.2017.02606] [Medline: 29312271]
118. Okereafor KU, Onime C, Osuagwu OE. Enhancing Biometric Liveness Detection Using Trait Randomization Technique. 2017 Presented at: 2017 UKSim-AMSS 19th International Conference on Computer Modelling & Simulation; April 5-7, 2017; Cambridge, UK p. 28-33. [doi: 10.1109/uksim.2017.44]
119. Akhtar Z, Micheloni C, Foresti GL. Biometric Liveness Detection: Challenges and Research Opportunities. IEEE Secur Privacy 2015 Sep;13(5):63-72. [doi: 10.1109/msp.2015.116]
120. Okereafor KU, Osuagwu OE, Onime C. Biometric Anti-spoofing Technique Using Randomized 3D Multi-Modal Traits. IJSSST 2018;19(5):5.1-5.8. [doi: 10.5013/ijssst.a.19.05.05]
121. Espinoza M, Champod C. Using the Number of Pores on Fingerprint Images to Detect Spoofing Attacks. 2011 Presented at: 2011 International Conference on Hand-Based Biometrics; November 17-18, 2011; Hong Kong p. 1-5. [doi: 10.1109/ichb.2011.6094347]
122. Novel Coronavirus (COVID-19)—Fighting Products. American Chemistry Council Center for Biocide Chemistries. 2020 Jul 20. URL: https://www.americanchemistry.com/Novel-Coronavirus-Fighting-Products-List.pdf [accessed 2020-07-31]
123. Abdelbary MA. Exploration of factors affecting adoption of biometric technology by five-star Egyptian hotel employees. Dissertation. Iowa State University 2011. [doi: 10.31274/etd-180810-2593]
124. Sheridan S. Techniques for Collecting and Analyzing Fingerprints. Forensic Resources. 2013 Jun 20. URL: https://ncforensics.wordpress.com/2013/06/20/techniques-for-collecting-and-analyzing-fingerprints/ [accessed 2020-06-24]
125. Nirmal SB, Kinage KS. Contactless Fingerprint Recognition and Fingerprint Spoof Mitigation using CNN. IJRTE 2019 Nov 30;8(4):9271-9275. [doi: 10.35940/ijrte.d9420.118419]
126. Noh D, Lee W, Son B, Kim J. Empirical study on touchless fingerprint recognition using a phone camera. J Electron Imag 2018 May 1;27(03):1. [doi: 10.1117/1.jei.27.3.033038]
127. Lee C, Lee S, Kim J. A Study of Touchless Fingerprint Recognition System. In: Yeung DY, Kwok JT, Fred A, Roli F, de Ridder D, editors. Structural, Syntactic, and Statistical Pattern Recognition. SSPR/SPR 2006. Lecture Notes in Computer Science, vol 4109. Berlin, Germany: Springer; 2006:358-365.
128. Kayser GL, Rao N, Jose R, Raj A. Water, sanitation and hygiene: measuring gender equality and empowerment. Bull World Health Organ 2019 May 14;97(6):438-440. [doi: 10.2471/blt.18.223305]

Abbreviations

ATM: automatic teller machine
ECWA: Evangelical Church Winning All
eHealth: electronic health
EPA: Environmental Protection Agency
FCT: Federal Capital Territory
FCTA: Federal Capital Territory Administration
ICT: Information and Communications Technology
ISO: International Organization for Standardization
NHIS: National Health Insurance Scheme
NMC: National Mirror Committee.
TC: Technical Committee

Edited by G Eysenbach; submitted 25.04.20; peer-reviewed by D Laborde, D Cebo; comments to author 12.06.20; accepted 24.07.20; published 08.09.20.

Please cite as:
Okereafor K, Ekong I, Okon Markson I, Enwere K
Fingerprint Biometric System Hygiene and the Risk of COVID-19 Transmission
JMIR Biomed Eng 2020;5(1):e19623
URL: http://biomedeng.jmir.org/2020/1/e19623/
doi:10.2196/19623
PMID:

Original Paper

Heart Rate and Oxygen Saturation Monitoring With a New Wearable Wireless Device in the Intensive Care Unit: Pilot Comparison Trial

Srinivasan Murali[1*], PhD; Francisco Rincon[1*], PhD; Tiziano Cassina[2], MD; Stephane Cook[3*], MD; Jean-Jacques Goy[3], MD

[1]SmartCardia, Lausanne, Switzerland
[2]Cardiocentro, Lugano, Switzerland
[3]University Hospital Fribourg, Fribourg, Switzerland
[*]these authors contributed equally

Corresponding Author:
Jean-Jacques Goy, MD
University Hospital Fribourg
Rue des Pensionnats 5-7
Fribourg,
Switzerland
Phone: 41 792136465
Email: jjgoy@goyman.com

Abstract

Background: Continuous cardiac monitoring with wireless sensors is an attractive option for early detection of arrhythmia and conduction disturbances and the prevention of adverse events leading to patient deterioration. We present a new sensor design (SmartCardia), a wearable wireless biosensor patch, for continuous cardiac and oxygen saturation (SpO_2) monitoring.

Objective: This study aimed to test the clinical value of a new wireless sensor device (SmartCardia) and its usefulness in monitoring the heart rate (HR) and SpO_2 of patients.

Methods: We performed an observational study and monitored the HR and SpO_2 of patients admitted to the intensive care unit (ICU). We compared the device under test (SmartCardia) with the ICU-grade monitoring system (Dräger-Healthcare). We defined optimal correlation between the gold standard and the wireless system as <10% difference for HR and <4% difference for SpO_2. Data loss and discrepancy between the two systems were critically analyzed.

Results: A total of 58 ICU patients (42 men and 16 women), with a mean age of 71 years (SD 11), were included in this study. A total of 13.49 (SD 5.53) hours per patient were recorded. This represents a total recorded period of 782.3 hours. The mean difference between the HR detected by the SmartCardia patch and the ICU monitor was 5.87 (SD 16.01) beats per minute (bias=–5.66, SD 16.09). For SpO_2, the average difference was 3.54% (SD 3.86; bias=2.9, SD 4.36) for interpretable values. SmartCardia's patch measures SpO_2 only under low-to-no activity conditions and otherwise does not report a value. Data loss and noninterpretable values of SpO_2 represented 26% (SD 24) of total measurements.

Conclusions: The SmartCardia device demonstrated clinically acceptable accuracy for HR and SpO_2 monitoring in ICU patients.

(JMIR Biomed Eng 2020;5(1):e18158) doi:10.2196/18158

KEYWORDS

cardiac monitoring; wireless monitor; wearable; cardiology; ICU; respiratory monitoring

Introduction

Electrocardiogram (ECG) monitoring in hospital units and intensive care units (ICU) has been performed for 50 years [1]. The goals of monitoring have changed over time, from simple tracking of heart rate (HR) and basic rhythm to the diagnosis of complex arrhythmias, detection of myocardial ischemia, identification of prolonged QT interval, or modifications of the QRS complex. Continuous ECG monitoring assists physicians in the evaluation of patients. Vital signs are usually measured

and documented by nurses. The frequency of measurements is increased when necessary, and, in the event of aberrant signs, the physician is informed and nurses manually check patient vital signs, sometimes with poor compliance [2,3]. This practice has several potential weaknesses. The frequency of monitoring by nurses is low, and relevant changes in vital signs may remain undetected, especially when they are subtle or within the normal range. These could lead to adverse events or complications [4].

Over time, major improvements have been made in cardiac monitoring systems, including computerized arrhythmia detection algorithms, ST segment/ischemia monitoring software, improved noise reduction strategies, multileads monitoring, reduced lead sets for monitoring-derived 12-lead ECG, wireless transmission of information and centralized data collection, all with a minimal number of electrodes [5,6]. With the introduction of wireless sensors that allow wireless continuous monitoring, an improvement in patient safety can be achieved [7]. Recently, the accuracy of wireless systems with different sensing principles was validated [8] in high-risk patients. Some of these devices are Food and Drug Administration (FDA)- or Conformité Européene (CE)-approved. It is hoped that these wireless systems can keep the alarm range as low as possible to reduce alarm fatigue that may otherwise result in alarm desensitization. New systems should be able to identify changes in patients' values when an adverse event develops. However, no system presently meets all these criteria. Wireless devices improve continuous monitoring and outcomes in hospital wards, as demonstrated recently [8-10]. Wireless systems now provide HR monitoring and oxygen saturation (SpO_2) measurements. HR monitoring is obtained with adhesive patches. In some instances, these systems can detect arrhythmias at better rates than Holter monitoring. Ventilatory frequency and SpO_2 measurements are of crucial importance in patient monitoring on hospital wards. Clinical deterioration of patients with hypoxemia and hypotension can be missed with typical ward monitoring, which is performed every 2-6 hours. In recent studies, routine checking of vital signs every 4 hours missed 90% of patients with at least 15 minutes of SpO_2 <90% [11]. Ventilatory frequency can also be derived from wireless pulse oximeters and chest patch systems that sense respiratory variation in R-wave amplitude and RR intervals. We designed a pilot trial to assess the safety and validity of an innovative wireless device, SmartCardia, in monitoring patients hospitalized in the ICU.

Methods

Study Design

This is a descriptive trial analysis of vital signs during hospitalization in a subset of patients, most of whom were being treated for cardiac disease. ICU patients at the Cardiocentro Ticino Hospital (Lugano, Ticino, Switzerland) were included in the trial. They were monitored with a medically approved existing monitoring system designated as the gold standard system (Dräger-Healthcare). In addition, the innovative wireless equipment (SmartCardia) was used to record ECG and SpO_2 for the duration of the patient's stay in the ICU. The study was designed as a pilot trial. We chose patients in the ICU because they are monitored 24 hours a day, making comparison with our device easier. This study is a first step prior to testing the device as a telemetry system or for 24-hour outpatient monitoring. Although the device was designed to detect arrhythmia, this feature was not tested in this trial. We deliberately decided to first confirm a high level of reliability and therefore focused on HR and SpO_2 in this study. This corresponds to a validation of the device for HR and SpO_2 monitoring. The next step will be to test the device as an ambulatory or telemetry monitoring device.

Study Population

Patients scheduled for major surgery or hospitalized in the ICU with an indication for cardiac and respiratory monitoring were asked to participate. Patients with an implantable cardiac device as well as those allergic to adhesive or with thorax skin irritation were excluded.

The inclusion criteria were as follows:

1. Patients with cardiovascular disease admitted to Cardiocentro's ICU, and aged 16 years and above
2. Patients who voluntarily signed informed consent forms
3. Postsurgery patients with stable hemodynamic condition
4. Postmyocardial infarction patients with stable hemodynamic condition
5. ICU patients who need continuous monitoring of vital signs for at least 12h. The patients were not consecutive patients, and they were fully immobilized in their bed.

Informed Consent and Data Collection

This protocol was approved by the local ethics committee, and written informed consent of all patients was obtained. Data were collected anonymously and transferred to SmartCardia for a blinded analysis.

System Description

The system is a wireless patch with a low-cost disposable component and a rechargeable/reusable electronic unit (dimensions 65 mm × 125 mm; Figure 1). The patch measures a single-lead ECG, HR, HR variability, respiration rate, oxygen saturation, skin temperature, posture, activity, and blood pressure variations. The data are transmitted by Bluetooth to a mobile phone or router. The measured signals and parameters are also stored on the device. The device is placed on the left upper portion of the patient's chest (Figure 2). The patch-based device offers up to 7-day data storage and 3.5-day real-time connectivity through a smartphone connected to cloud storage on a single battery charge. The ability to receive, store, and interpret a broad range of parameters offers the opportunity to go far beyond monitoring individual parameters. In this study, the ECG-based HR and SpO_2 measurements taken by the SmartCardia device were compared with the measurements obtained by the standard monitoring system used in the ICU.

Figure 1. The appearance and dimensions of the SmartCardia wireless device.

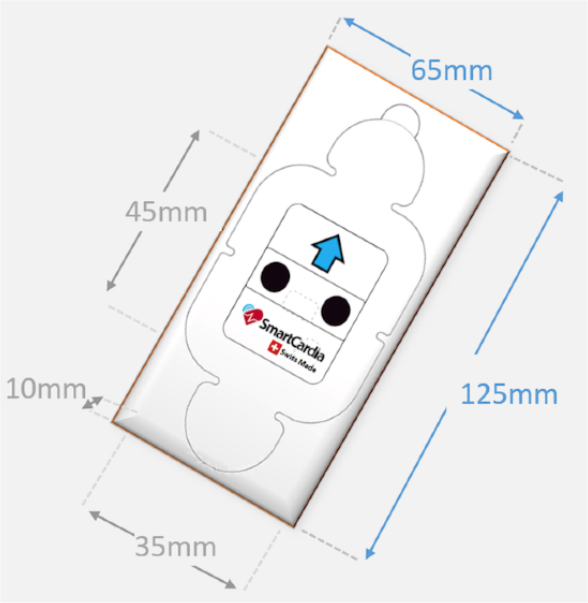
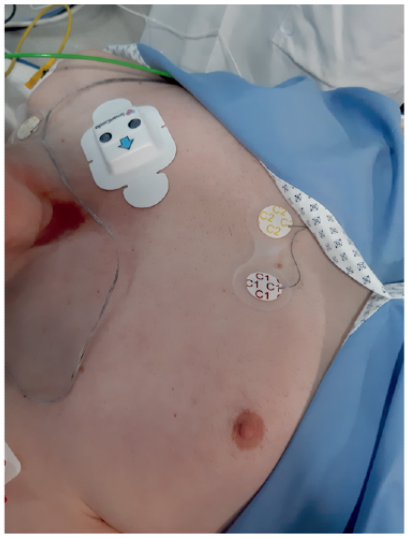


Figure 2. Device on the chest of the patient.

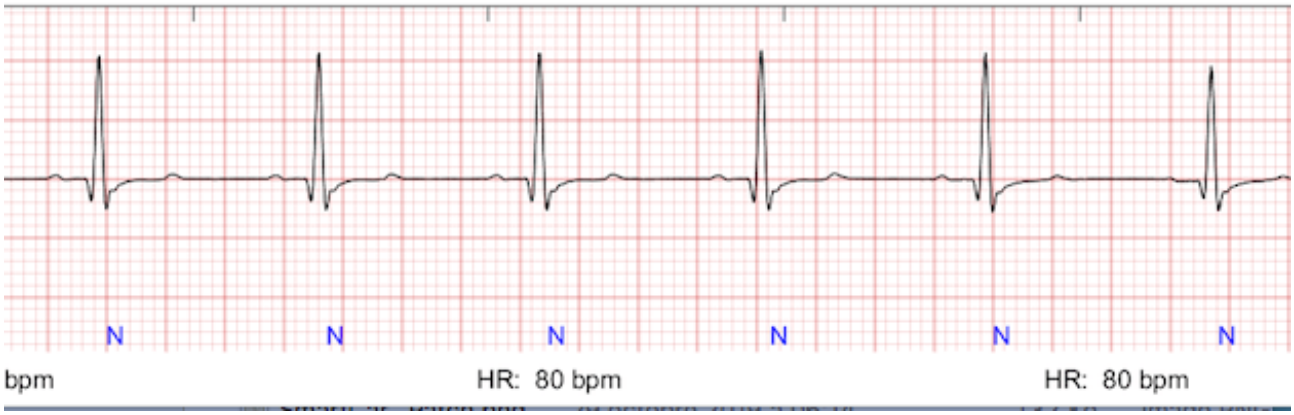


Safety

The materials and the patches that have skin contact meet the ISO 10993 guidelines, specifically the criteria for skin irritation, skin sensitization, and in vitro cytotoxicity.

Two SmartCardia patches were attached to the chest and the left arm, respectively, of the patient (Figure 2). A high-quality ECG signal is recorded with the chest device (Figure 3), while the arm device records the best oxygen saturation signal.

Figure 3. Typical trace recorded with the SmartCardia device.



Endpoints

The endpoints of this pilot trial were the following: safety of the wireless device, tolerance of the device, correlation of HR measurements obtained by the wireless device and the conventional monitoring device, and correlation of SpO_2 measurements obtained by the wireless device and the conventional monitoring device.

Statistics

Sample size could not be calculated, as no power calculations exist in the literature. Data pairs of HR and SpO_2 measurements were analyzed with the Bland and Altman method, as suggested by Breteler et al [7,12]. This method was used to account for within-subject variations by correcting the differences between the average differences. Bias between the reference monitor and the sensor, with its 95% limits, was determined for both HR and SpO_2. The correlation between the two systems was considered optimal when the mean HR differed by <5 beats per minute or <10% and when SpO_2 variations were <4%.

The Clarke Error Grid was used to determine the clinical accuracy of the wireless sensor compared with the reference standard [11]. Performance was analyzed by the percentage of data that was lost or noninterpretable.

Results

During the study period (May 17, 2017, to March 31, 2018), 58 patients (42 men and 16 women), with a mean age of 71 (SD 11) years, were included in this prospective pilot trial (Table 1). Patients mainly had heart problems and were hospitalized in the ICU after cardiac surgery (n=39, 67%), transaortic valvular implantation (n=14, 24%), acute coronary syndrome (ACS, n=3, 5%), aggravation of chronic obstructive pulmonary disease (COPD, n=1, 2%), and out of hospital cardiac arrest (n=1, 2%). The recording could be achieved in all patients. No patients reported side effects (such as skin rash) from the SmartCardia device. There was no disconnection of the device or interruption of the recording due to poor skin contact.

Total monitoring time is displayed in Table 2 and was 782 h 17 min (mean 809 min/patient, SD 332) for the chest device and 794 h 22 min for the arm device (mean 822 min/patient, SD 322).

The correlation rate for the ECG signal between the two systems was 97.6% for the total recorded values (Figure 4). In 10 patients (17%), correlation between the two systems was <95% (Figure 5).

Table 3 shows the bias and precision (95% agreement) between HR and SpO_2 measurements, respectively, by the wireless device and the gold standard. The 95% limit of agreement was calculated with the Bland-Altman method (HR: Figure 6; SpO_2: Figure 7).

The causes for instances of low correlation between the two systems included (1) failure of the gold standard in 2 cases (disconnection of one cable); (2) failure of the SmartCardia device due to poor skin contact or bad positioning in 4 patients (solved by reapplication or repositioning); (3) overdetection of the T wave as a QRS complex, inducing double count in 2 patients (Figures 8 and 9); and (4) technical failure of the SmartCardia device in 2 patients (memory corruption in one case and a casing mechanical issue in the other).

The mean exact correlation rate for SpO_2 measurement (<4% difference) was obtained in 44/58 patients (75%; Table 2). A correlation between 5% and 6% of exact correlation was obtained in 4 patients (7%). Finally, in 10 patients (18%), the correlation rate was considered low (>6% of error) and the largest difference was 14%.

The Clarke Error Grid analysis was used to quantify the clinical accuracy of the HR and SpO_2 measurements obtained by the wireless system. This information is shown in Table 4 and plotted in Figures 10 and 11. The percentage of data pairs in regions A to E are shown in Table 4. For HR, 96.2% of the values are in regions A and B. For SpO_2, 100% of the points are in regions A and B.

Table 1. Demographics of the patients included in the trial.

Demographic	Male (n=42)	Female (n=16)
Age (years), mean (SD)	70 (12)	74 (8)
Trans-aortic valvular implantation, n (%)	10 (24)	4 (25)
Coronary artery bypass grafting, n (%)	21 (50)	8 (50)
Open-heart valvular surgery, n (%)	6 (14)	4 (25)
ST segment elevation myocardial infarction and ischemic heart disease, n (%)	3 (7)	0 (0)
Chronic obstructive pulmonary disease, n (%)	1 (2)	0 (0)
Cardiac arrest, n (%)	1 (2)	0 (0)

Table 2. Recording results and correlation rates.

Parameter	Heart rate monitoring (n=58)	Oxygen saturation monitoring (n=58)
Total monitoring time	782 h 17 min	794 h 22 min
Monitoring time per patient (min)	822 (SD 332)	809 (SD 332)
Exact value (<5% error), n (%)	48 (83)	44 (75)
Values in the 5%-6% error range, n (%)	2 (4)	4 (7)
Values >6% error, n (%)	8 (14)	10 (18)

Figure 4. Recordings by the two monitoring systems, showing extremely good correlation (blue = SmartCardia recording; black = conventional monitoring).

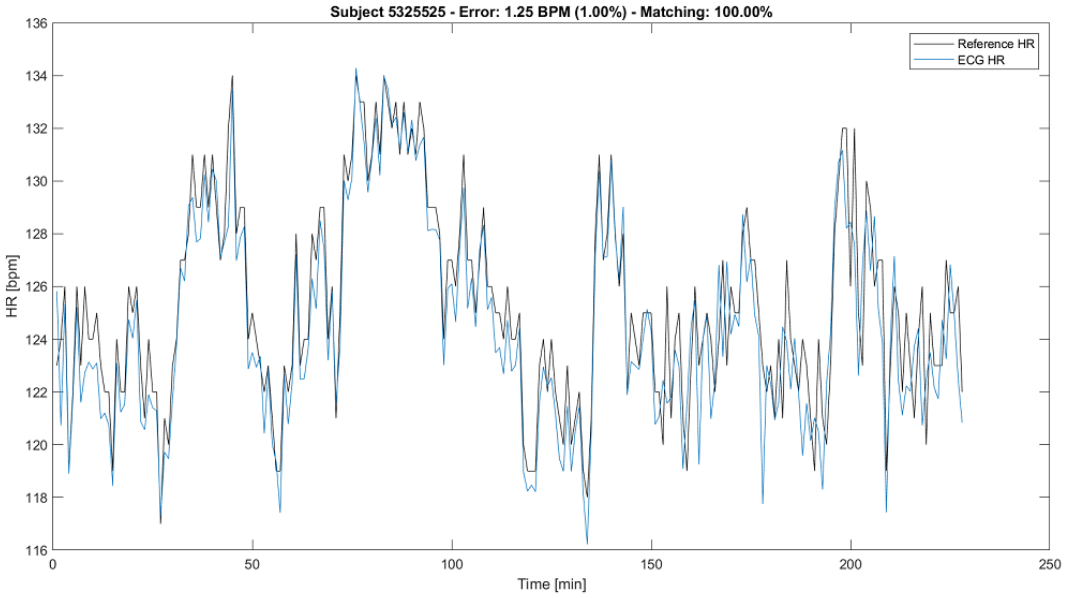


Figure 5. Example of a trace defined as exact. Some PVB (premature ventricular beats) are correctly detected by the wireless system where the gold standard failed.

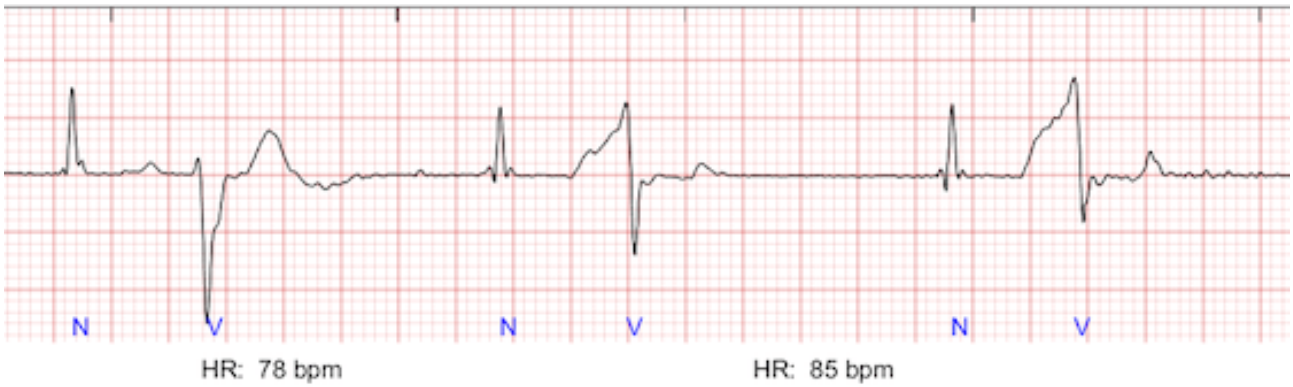


Table 3. Bland-Altman analysis of heart rate and oxygen saturation.

Parameters measured	Number of measurement pairs	Number of patients	Bias, n (SD)	Average absolute error, n (SD)	Lower 95%	Upper 95%
Heart rate	41283	58	–5.66 (16.09)	5.87 (16.01)	–37.19	25.87
Oxygen saturation	58970	58	2.9 (4.36)	3.54 (3.86)	–5.65	11.45

Figure 6. Bland-Altman plots of heart rate for the wireless system. Y-axis = difference between Smart-Cardia device and reference.

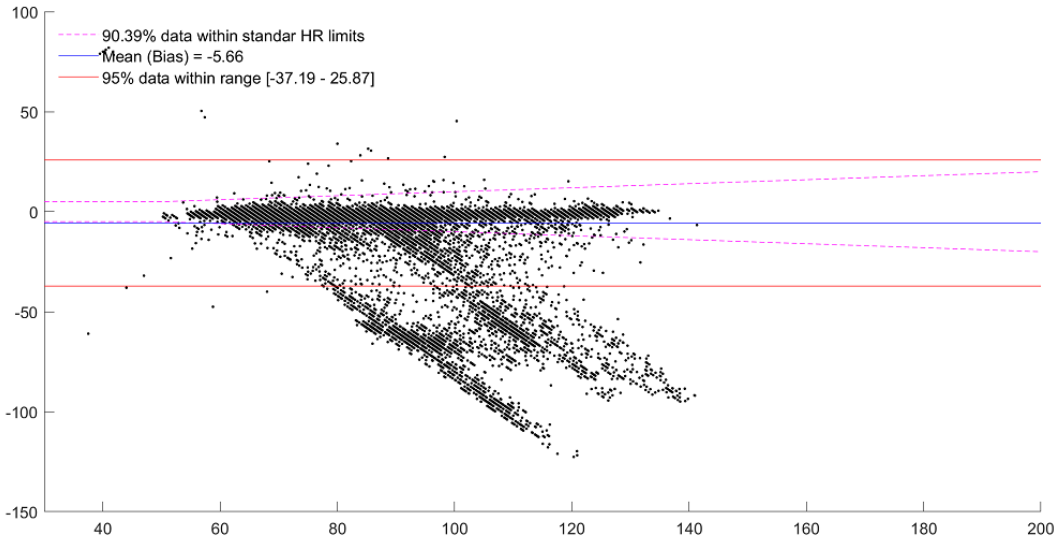


Figure 7. Bland-Altman plots of oxygen saturation measurements by the wireless system. Y-axis = difference between Smart-Cardia device and reference.

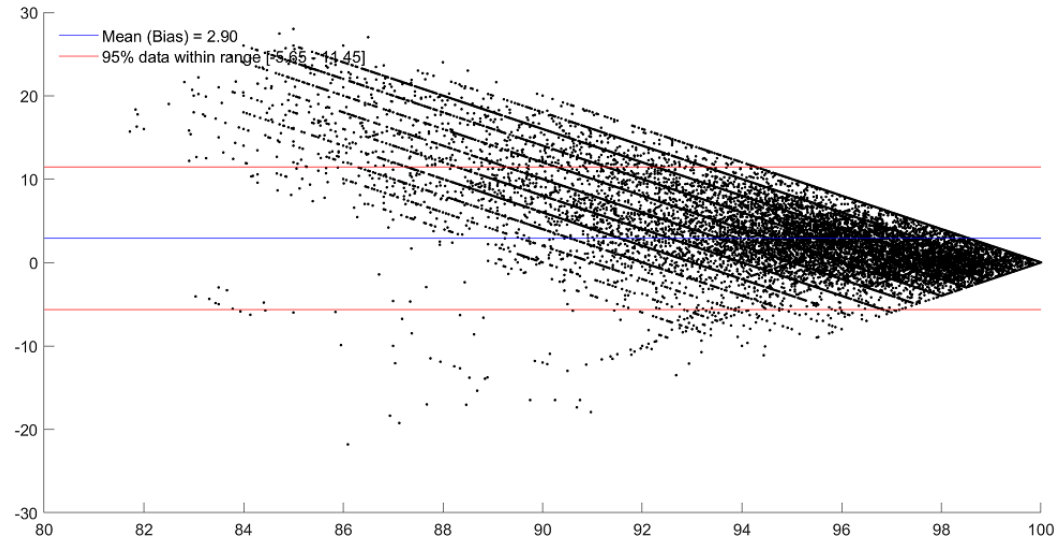


Figure 8. Example of a trace qualified as "non exact". The T wave has been included in the heart rate count, doubling the value to 126 bpm from 63 bpm.

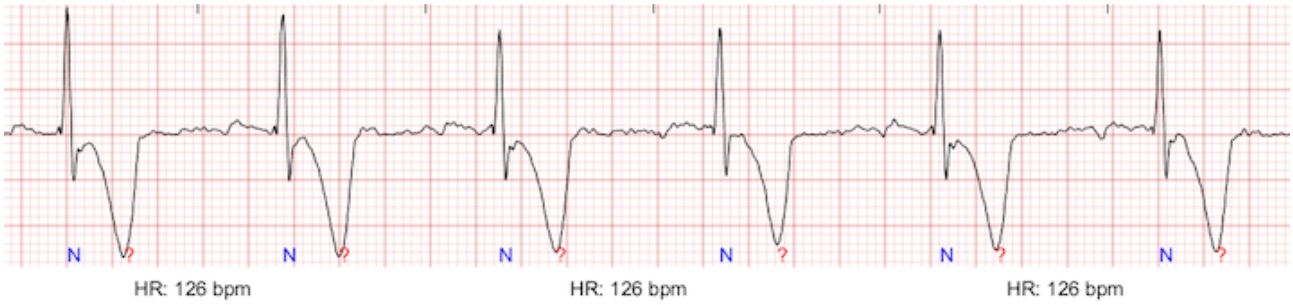


Figure 9. Discrepancy between the two monitoring systems, lasting a few hours, due to double counting by the SmartCardia device (SmartCardia heart rate values were double that of the gold standard). This was due to Q wave detection being counted as a QRS complex.

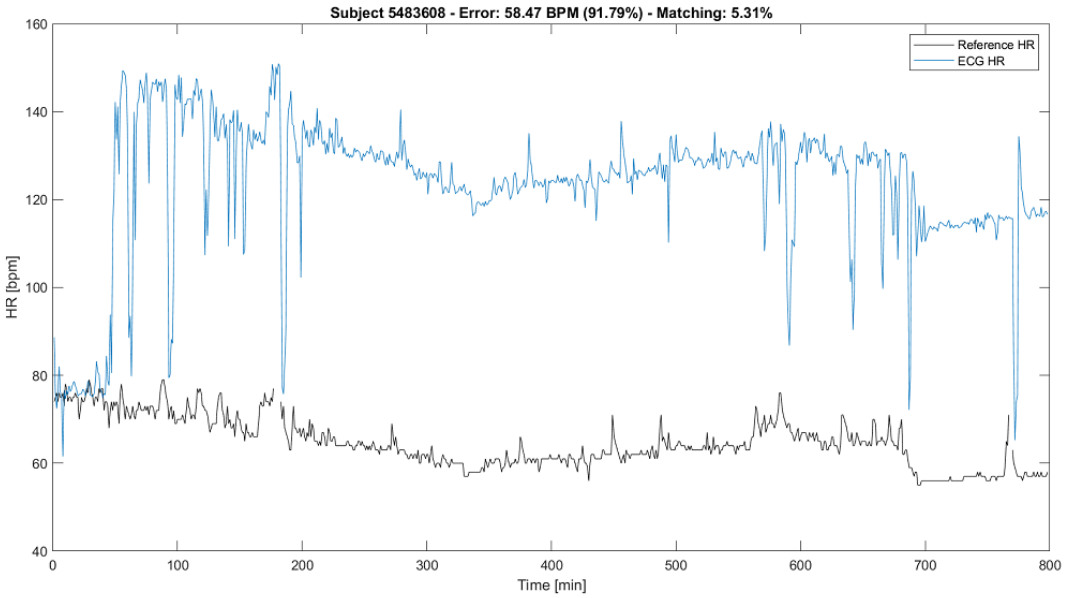


Table 4. Clarke Error Grid analysis to quantify the clinical accuracy of heart rate and oxygen saturation.

Parameters measured	Measurements in region A, n (%)	Measurements in region B, n (%)	Measurements in region C, n (%)	Measurements in region D, n (%)	Measurements in region E, n (%)
Heart rate	38133 (92)	1743 (4.2)	1553 (3.7)	5 (0.1)	0 (0)
Oxygen saturation	58414 (69)	26133 (30.9)	0 (0)	0 (0)	0 (0)

Figure 10. Clarke Error Grid analysis to quantify the clinical accuracy of heart rate measurements by the Smart Cardia device. Region A contains 92% of recorded values, indicating a high accuracy of measurements. Regions D and E represent regions with unacceptable accuracy and only 0.1% of values were in these regions.

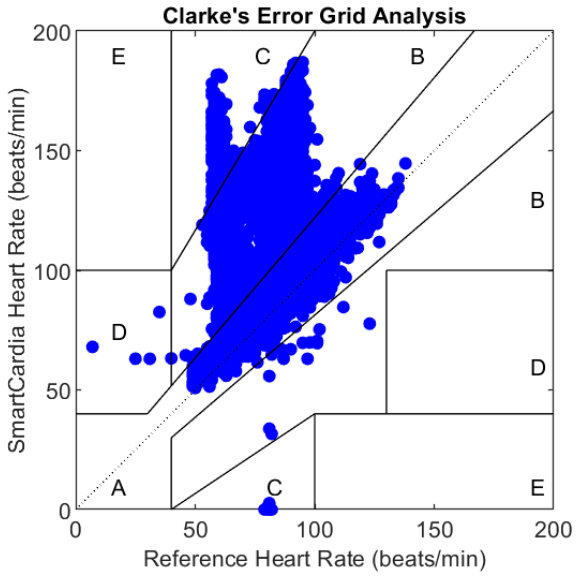
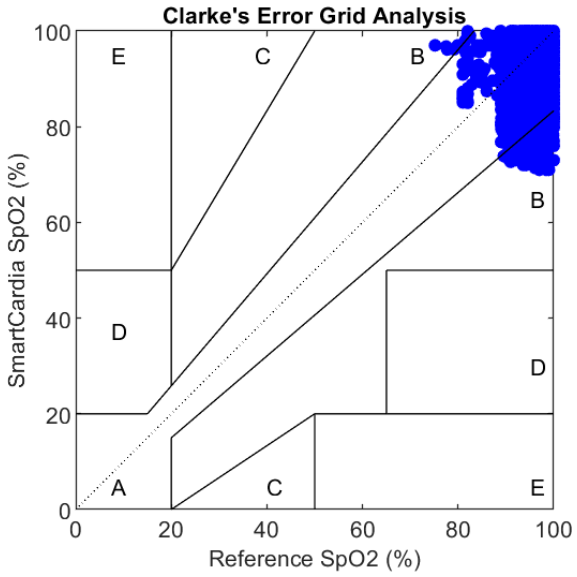


Figure 11. Clarke Error Grid analysis to quantify the clinical accuracy of oxygen saturation measurements taken by the SmartCardia device. Region A and B contain 96% of values within the limit range.



Discussion

Principal Findings

Monitoring is a critical issue in the ICU and intermediate care units during the first days following the intervention. Today, cardiac monitoring is carried out with skin electrodes. The signal is transmitted by wires to a bedside monitor and a central station. In this pilot trial, we tested the ability of a new wireless vital sign sensor to monitor patients. The first generation of wireless sensors was shown to be useful in the detection of abnormalities in vital signs, sometimes preceding adverse events [7-15]. It is well accepted that early detection of arrhythmias is important since they can precede serious complications in patients in the ICU or general ward [14,15]. The quality of the signal is dependent on the physical device connections and good contact between electrodes and the patient's skin. Monitoring of SpO_2 is usually performed using a system attached to the finger of the patient, with a cable transmission to the central station. Recently, wireless monitoring has been confirmed as a valuable tool to detect adverse events in high-risk patients [4-6]. A recent study showed significant differences between different wireless sensors [8]. For example, the Radius-7 (Masimo) underestimates HR, as it calculates HR from the plethysmographic waveform obtained from the pulse oximeter probe. The EarlySense system may also underestimate HR during periods of arrhythmia, as it derives HR from cardiac ballistic movement associated with the ejection of blood with each heart cycle. During a rapid ventricular rate, some beats will not be long enough to allow for ventricular filling and will therefore not result in a detectable peripheral pulse. SmartCardia's patch measures and stores ECG data and parameters for up to 7 days on a single charge. A comparison of ECG patches by Sensium and VitalConnect [8] showed that they both have a limited storage capacity of 10 hours or less. Additionally, these patches do not measure SpO_2, while this study demonstrated the feasibility of measuring both SpO_2 and HR with the SmartCardia patch [16,17].

The benefits of patient monitoring with wireless sensors have been shown [18-21]. To be useful for patient management after major surgical interventions, reliable wireless sensors must be connected to a central system with alarm notification. False alerts should be avoided to increase the reliability of the monitoring system.

The new wireless SmartCardia device, tested in this trial, is a combined system allowing monitoring of both HR and SpO_2. The recording system is in a small patch-box that is adhered to the patient's skin without wires, avoiding undesired problems related to wire disconnection. In addition, it is user-friendly in terms of installation. Transportation of the patient is also facilitated without the need to disconnect and reconnect wires.

This study was mainly designed to validate the sensor accuracy of the SmartCardia system, not to clinically monitor patients in the ICU. Thus, the sample size is rather small, and the number of events was too small to identify specific signs or patterns of adverse events. However, our results do provide insight into the ability of wireless sensors to assist in patient monitoring and early detection of patient deterioration.

In this pilot trial, we showed that the HR data obtained with the SmartCardia device is of good quality as compared with the gold standard monitoring system. We identified that both this new system and the gold standard system can fail. The major problem identified with this first generation of the SmartCardia system is maintaining adequate skin contact. With further technical improvements, this problem has been solved, leading to 100% success in device attachment. We also noted that optimal positioning of the device on the patient's chest is of crucial importance. In the small number of patients with poor recording quality, the device was not positioned correctly. Better placement of the device allowed significant improvement of the recording. We also noted that in 2 patients, a technical problem of the SmartCardia device was the reason for failure. We identified this problem and improvements were made in the most recent generation of devices.

Limitations

The main limitation of this trial is the relatively small number of patients. In addition, it has only been tested in patients hospitalized in the ICU. Tests in postoperative patients in the ward must be realized.

Conclusion

We demonstrated that a new wireless device provides good quality HR measurements in patients in the ICU. Furthermore, SpO_2 monitoring is feasible, although further technical improvements are mandatory. Based on these results, further trials should be performed in an outpatient population to define the place of such devices in daily practice. Finally, it should be compared with conventional Holter monitoring.

Conflicts of Interest

SM and FR are part of the board of SmartCardia.

References

1. Day HW. Preliminary studies of an acute coronary care area. Lancet 1963;83:53-55. [Medline: 14025617]
2. Ludikhuize J, Smorenburg SM, de Rooij SE, de Jonge E. Identification of deteriorating patients on general wards; measurement of vital parameters and potential effectiveness of the Modified Early Warning Score. Journal of Critical Care 2012;27(4):424.e7-424.e13. [doi: 10.1016/j.jcrc.2012.01.003]
3. Cardona-Morrell M, Prgomet M, Lake R, Nicholson M, Harrison R, Long J, et al. Vital signs monitoring and nurse–patient interaction: A qualitative observational study of hospital practice. International Journal of Nursing Studies 2016;56:9-16. [doi: 10.1016/j.ijnurstu.2015.12.007]
4. Dower GE, Yakush A, Nazzal SB, Jutzy RV, Ruiz CE. Deriving the 12-lead electrocardiogram from four (EASI) electrodes. Journal of Electrocardiology 1988;21:S182-S187. [doi: 10.1016/0022-0736(88)90090-8]
5. Drew BJ, Pelter MM, Brodnick DE, Yadav AV, Dempel D, Adams MG. Comparison of a new reduced lead set ECG with the standard ECG for diagnosing cardiac arrhythmias and myocardial ischemia. Journal of Electrocardiology 2002;35(4):13-21. [doi: 10.1054/jelc.2002.37150]
6. Downey C, Randell R, Brown J, Jayne DG. Continuous Versus Intermittent Vital Signs Monitoring Using a Wearable, Wireless Patch in Patients Admitted to Surgical Wards: Pilot Cluster Randomized Controlled Trial. J Med Internet Res 2018 Dec 11;20(12):e10802. [doi: 10.2196/10802]
7. Breteler MJM, KleinJan EJ, Dohmen DAJ, Leenen LPH, van Hillegersberg R, Ruurda JP, et al. Vital Signs Monitoring with Wearable Sensors in High-risk Surgical Patients. Anesthesiology 2020;132(3):424-439. [doi: 10.1097/aln.0000000000003029]
8. Breteler MJM, KleinJan E, Numan L, Ruurda JP, Van Hillegersberg R, Leenen LPH, et al. Are current wireless monitoring systems capable of detecting adverse events in high-risk surgical patients? A descriptive study. Injury 2019 Nov 17 [FREE Full text] [doi: 10.1016/j.injury.2019.11.018] [Medline: 31761422]
9. Michard F, Teboul JL. Predictive analytics: beyond the buzz. Ann Intensive Care 2019 Apr 11;9(1). [doi: 10.1186/s13613-019-0524-9]
10. Sessler DI, Saugel B. Beyond ‘failure to rescue’: the time has come for continuous ward monitoring. British Journal of Anaesthesia 2019 Mar;122(3):304-306. [doi: 10.1016/j.bja.2018.12.003]
11. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating Clinical Accuracy of Systems for Self-Monitoring of Blood Glucose. Diabetes Care 1987 Sep 01;10(5):622-628. [doi: 10.2337/diacare.10.5.622]
12. Altman DG, Bland JM. Measurement in Medicine: The Analysis of Method Comparison Studies. The Statistician 1983 Sep;32(3):307. [doi: 10.2307/2987937]
13. Selvaraj N, Nallathambi G, Moghadam R, Aga A. Fully Disposable Wireless Patch Sensor for Continuous Remote Patient Monitoring. 2018 Presented at: 40th Annual International Conference of the IEEE (IMBS); 20-21 July; Honolulu. [doi: 10.1109/embc.2018.8512569]
14. Selvaraj N, Nallathambi G, Kettle P. A Novel Synthetic Simulation Platform for Validation of Breathing Rate Measurement. 2018 Presented at: 40th Annual International Conference of the IEEE (IMBS); July; Honolulu. [doi: 10.1109/embc.2018.8512352]
15. Seesing M. F. J., Scheijmans, J.C.G., Borggreve, A.S., van Hillegersberg, R., Ruurda, J. P. The predictive value of new-onset atrial fibrillation on postoperative morbidity after esophagectomy. Dis Esophagus Nov 1 2018;31:11. [doi: 10.1093/dote/doy028]
16. Chan AM, Ferdosi N, Narasimhan R. Ambulatory respiratory rate detection using ECG and a triaxial accelerometer. 2013 Presented at: Conf Proc IEEE Eng Med Biol Soc; 2013; Atlanta p. 4058-4061. [doi: 10.1109/embc.2013.6610436]
17. Hernandez-Silveira M, Ahmed K, Ang S, Zandari F, Mehta T, Weir R, et al. Assessment of the feasibility of an ultra-low power, wireless digital patch for the continuous ambulatory monitoring of vital signs. BMJ Open 2015 May 19;5(5):e006606-e006606. [doi: 10.1136/bmjopen-2014-006606]

18. Weenk M, Koeneman M, van de Belt TH, Engelen LJ, van Goor H, Bredie SJ. Wireless and continuous monitoring of vital signs in patients at the general ward. Resuscitation 2019 Mar;136:47-53. [doi: 10.1016/j.resuscitation.2019.01.017]
19. Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The Value of Modified Early Warning Score (MEWS) in Surgical In-Patients: A Prospective Observational Study. Ann R Coll Surg Engl 2006 Oct;88(6):571-575. [doi: 10.1308/003588406x130615]
20. Posthuma LM, Visscher MJ, Hollmann MW, Preckel B. Monitoring of High- and Intermediate-Risk Surgical Patients. Anesthesia & Analgesia 2019;129(4):1185-1190. [doi: 10.1213/ane.0000000000004345]
21. Vincent J, Einav S, Pearse R, Jaber S, Kranke P, Overdyk FJ, et al. Improving detection of patient deterioration in the general hospital ward environment. Eur J Anaesthesiol 2018 May;35(5):325-333 [FREE Full text] [doi: 10.1097/EJA.0000000000000798] [Medline: 29474347]

Abbreviations

ACS: acute coronary syndrome
CABG: coronary artery bypass grafting
COPD: chronic obstructive pulmonary disease
ECG: electrocardiogram
HR: heart rate
ICU: intensive care unit
IHD: ischemic heart disease
SpO_2: oxygen saturation
STEMI: ST segment elevation myocardial infarction

Edited by G Eysenbach; submitted 07.02.20; peer-reviewed by A del Barrio, A Martin; comments to author 26.02.20; revised version received 14.03.20; accepted 21.03.20; published 22.04.20.

Please cite as:
Murali S, Rincon F, Cassina T, Cook S, Goy JJ
Heart Rate and Oxygen Saturation Monitoring With a New Wearable Wireless Device in the Intensive Care Unit: Pilot Comparison Trial
JMIR Biomed Eng 2020;5(1):e18158
URL: http://biomedeng.jmir.org/2020/1/e18158/
doi:10.2196/18158
PMID:

Original Paper

A Contact-Free, Ballistocardiography-Based Monitoring System (Emfit QS) for Measuring Nocturnal Heart Rate and Heart Rate Variability: Validation Study

Ville Vesterinen[1*], PhD; Niina Rinkinen[1*], MSc; Ari Nummela[1*], PhD

KIHU - Research Institute for Olympic Sports, Jyväskylä, Finland

[*]all authors contributed equally

Corresponding Author:
Ville Vesterinen, PhD
KIHU - Research Institute for Olympic Sports
Rautpohjankatu 6
Jyväskylä, 40700
Finland
Phone: 358 505451049
Email: ville.vesterinen@kihu.fi

Abstract

Background: Heart rate (HR) and heart rate variability (HRV) measurements are widely used to monitor stress and recovery status in sedentary people and athletes. However, effective HRV monitoring should occur on a daily basis because sparse measurements do not allow for a complete view of the stress-recovery balance. Morning electrocardiography (ECG) measurements with HR straps are time-consuming and arduous to perform every day, and thus compliance with regular measurements is poor. Contact-free, ballistocardiography (BCG)-based Emfit QS is effortless for daily monitoring. However, to the best of our knowledge, there is no study on the accuracy of nocturnal HR and HRV measured via BCG under real-life conditions.

Objective: The aim of this study was to evaluate the accuracy of Emfit QS in measuring nocturnal HR and HRV.

Methods: Healthy participants (n=20) completed nocturnal HR and HRV recordings at home using Emfit QS and an ECG-based reference device (Firstbeat BG2) during sleep. Emfit QS measures BCG by a ferroelectret sensor installed under a bed mattress. HR and the root mean square of successive differences between RR intervals (RMSSD) were determined for 3-minute epochs and the sleep period mean.

Results: A trivial mean bias was observed in the mean HR (mean –0.8 bpm [beats per minute], SD 2.3 bpm, P=.15) and Ln (natural logarithm) RMSSD (mean –0.05 ms, SD 0.25 ms, P=.33) between Emfit QS and ECG. In addition, very large correlations were found in the mean values of HR (r=0.90, P<.001) and Ln RMSSD (r=0.89, P<.001) between the devices. A greater amount of erroneous or missing data (P<.001) was observed in the Emfit QS measurements (28.3%, SD 14.4%) compared with the reference device (1.1%, SD 2.3%). The results showed that 5.0% of the mean HR and Ln RMSSD values were outside the limits of agreement.

Conclusions: Based on the present results, Emfit QS provides nocturnal HR and HRV data with an acceptable, small mean bias when calculating the mean of the sleep period. Thus, Emfit QS has the potential to be used for the long-term monitoring of nocturnal HR and HRV. However, further research is needed to assess reliability in HR and HRV detection.

(JMIR Biomed Eng 2020;5(1):e16620) doi:10.2196/16620

KEYWORDS

wearable technology; cardiac autonomic regulation; monitoring; validity

Introduction

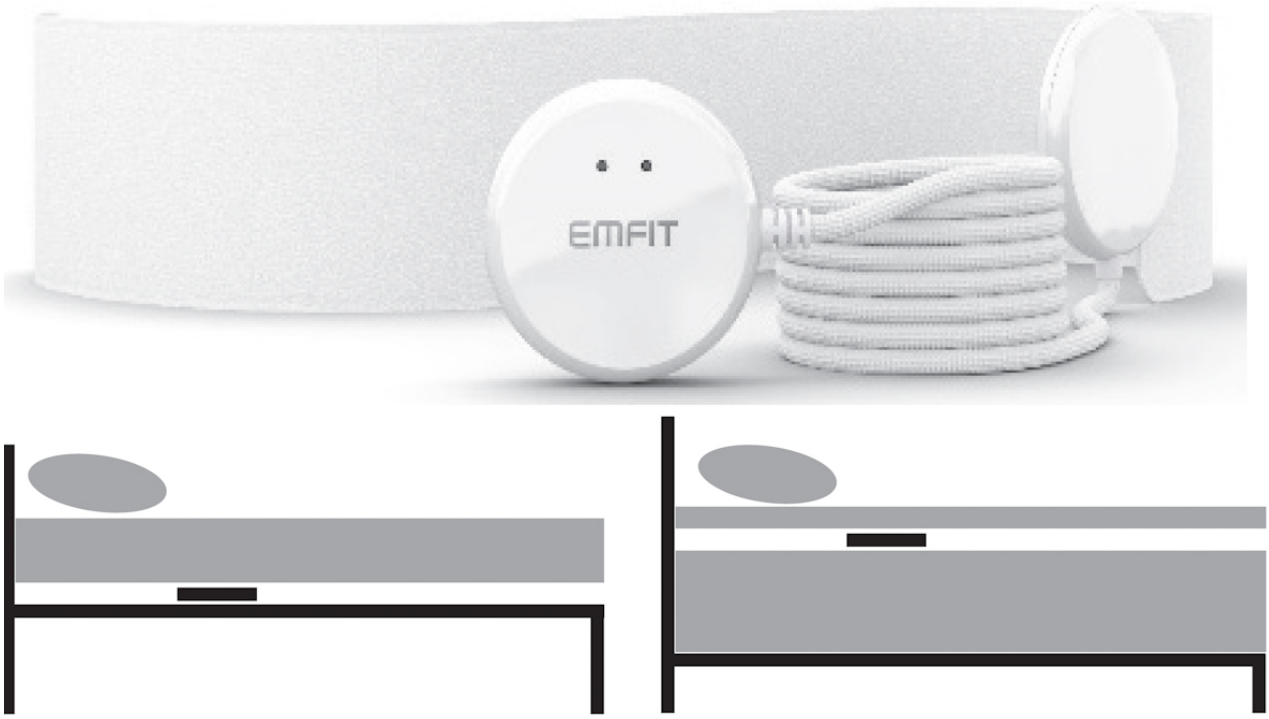
Technological development has brought forth numerous apps, gadgets, and high-tech solutions designed to enhance health, fitness, and performance. Based on a survey by the American College of Sports Medicine, wearable technology is the most popular fitness trend of 2019 [1]. Data from everyday life can be easily recorded by wearable solutions. Because of the growing use of wearable technologies, it has become important to determine their validity. However, more than half of the devices currently used to monitor and improve personal health

and sports performance have not been validated through independent research [2].

Over the past several decades, heart rate (HR) has been used to monitor physiological stress and workload during exercise [3]. Recently, heart rate variability (HRV) measurements has been growing in popularity. HRV reflects the activity of cardiac autonomic regulation and can be used in the monitoring of stress and recovery status [4-6]. Traditionally, HR has been determined through the process of electrocardiography (ECG) by measuring the electrical activity of the heart. A 12-lead ECG, with electrodes attached to the body surface, is widely used in medical examinations. Traditional HR monitors allow for real-time measurement, using a chest strap with wireless ECG sensors [3]. Nowadays, HR can also be measured optically by photoplethysmography (PPG) with wearables, such as watches and mobile phones [7,8]. In addition, HR can be determined by ballistocardiography (BCG), which measures ballistic forces on the heart arising from the sudden ejection of blood into the great vessels with each heart beat [9]. It is a long-established, noninvasive technique that uses several types of sensors like pressure sensors, film-type force sensors, microbend fiber optic BCG sensors, electromechanical film transducer sensors, piezoelectric film sensors, polyvinylidene fluoride sensors, strain gauges, and pneumatic and hydraulic sensors [10]. The novel sensor technologies may detect HR and HRV more accurately. In addition, several different algorithms are being used to detect BCG peaks, each with differing detecting abilities [11]. However, previous studies have been conducted under laboratory conditions. Thus, it is necessary to evaluate the validity of BCG measurements under real-life conditions.

Emfit QS is a BCG-based commercial device for monitoring sleep and recovery. An EMFi sensor (6 cm x 55 cm in size) installed under a bed mattress can detect HR, HRV, breathing, and other body movements (Figure 1 [12]). Emfit QS measurement starts automatically shortly after the user goes to bed and stops recording once they have left the bed in the morning. Data are transferred via a Wi-Fi or 3G network to the internet, and the results can be accessed from a smartphone, tablet, or computer shortly after awakening. Thus, it is a contact-free, effortless, and user-friendly method with the capacity to improve user compliance on a daily basis. However, to the best of our knowledge, there is no previously published data on measuring BCG-based nocturnal HR and HRV under real-life conditions. Thus, the aim of this study was to evaluate the accuracy of Emfit QS in measuring HR and HRV during sleep, alongside an ECG-based device as a reference.

Figure 1. Correct placement of the Emfit QS sensor beneath the mattress or mattress topper under the chest area. Source: Emfit Ltd [12].



Methods

Participants

A total of 20 participants were recruited to the study. Women (n=11; mean age 34 yrs, SD 7 yrs; mean height 1.69 m, SD 0.05 m; mean weight 67 kg, SD 10 kg) and men (n=9; mean age 42 yrs, SD 8 yrs; mean height 1.80 m, SD 0.06 m; mean weight 78 kg, SD 3 kg) were healthy (eg, no disease) and nonsmokers, and did not take medication on a regular basis. The participants were fully informed about the study design and the use of measurement devices before signing an informed consent document. The study complies with the standards set by the Ethics Committee of the University of Jyväskylä, Finland.

Data Collection

Nocturnal recordings were taken at home during sleep. Recordings began shortly after participants went to bed and

stopped once they left their bed in the morning. Before the first recording, Emfit QS's own proprietary cellular ferroelectret sensor was placed beneath the mattress or mattress topper under chest area (Figure 1). The reference RR interval (RRI) data were recorded with Firstbeat Bodyguard 2 (BG2), an ECG-based recorder with two disposable electrodes and a sampling frequency of 1000 Hz. BG2 and its electrodes were set up on a participant's body according to instructions in the user manual. The accuracy of BG2 was previously evaluated in laboratory protocol studies by Parak et al [13] and Bogdány et al [14], and in our unpublished study, which showed perfect agreement in the detection of RRIs (r=1.00) during a 30-minute rest period with Custo Cardio 100BT, a 12-channel ECG device (Custo med GmbH). The participants recorded the data over 3 consecutive nights, and the last recording was used in the analysis. Before measurements were taken, time synchronization was performed on the devices.

Data Analysis

Emfit QS provides HR, the vagal-related HRV index, and the root mean square of successive differences (RMSSD) in RRIs throughout the sleep period in 3-minute epochs. If heart beat detection is disturbed due to a poor signal or artifacts any time during this period, data are not collected. The RRI data of BG2 were analyzed using the Firstbeat Sports software (Firstbeat Technologies Ltd). RRIs were checked by an artifact detection filter of the Firstbeat Sports software [15] and subsequently excluded all falsely detected, missed, and premature heart beats caused by movement artifacts or any other artifacts of unknown origin. HR and RMSSD values were calculated for each 3-minute epoch throughout the measurement. Averages of the whole night period, which were used in the analysis, were calculated from those 3-minute values. Emfit QS and BG2 data were synchronized according to the time stamp. In addition to HR and RMSSD data, the amount of missing data was calculated.

Statistical Analysis

Values are expressed as mean (SD). Averages of the whole night period were calculated from the 3-minute epochs. The Gaussian distribution of the data was assessed with the Shapiro-Wilk goodness-of-fit test. Ln (natural logarithm)-transformation was applied to the RMSSD data in order to meet the assumptions of the parametric statistical analysis. The accuracy of HR and RMSSD measured by Emfit QS was evaluated by determining the amount of missing data, the mean bias (absolute and percentage), and the root mean square error (RMSE) compared with the reference (BG2). Statistical difference between the measurements of BG2 and Emfit QS was analyzed using a paired Student t test. In addition, the magnitude of the differences was expressed as effect size (ES). The difference was considered trivial when ES≤0.2, small when ES≤0.6, moderate when ES≤1.2, large when ES≤2.0, and very large when ES>2.0. A Pearson product-moment correlation and a Bland-Altman plot were used to analyze agreement between the reference and Emfit QS data. A Spearman rank correlation coefficient was calculated to investigate the correlation between the absolute differences (reference minus Emfit QS) and the average of the devices for HR and Ln RMSSD. In addition to the measures of statistical significance, the following criteria were adopted to interpret the magnitude of the correlation between measurement variables: <0.10 (trivial), 0.11-0.30 (small), 0.31-0.50 (moderate), 0.51-0.70 (large), 0.71-0.90 (very large), and 0.91-1.0 (almost perfect) [16]. Significance was accepted as P<.05. Data were analyzed using SPSS Statistics 25 software (IBM Corp).

Results

No significant differences were found in HR (mean difference –1.7%, SD 4.6%, P=.15) and Ln RMSSD (mean difference –2.0%, SD 6.5%, P=.33) between the measurements by ECG (BG2) and BCG (Emfit QS) (Figure 2 and Table 1). A greater amount of erroneous or missing data (P<.001) was found in the 3-minute values of Ln RMSSD by Emfit QS (mean 28.3%, SD 14.4%) compared with the reference device (mean 1.1%, SD 2.3%). Very large correlations were found in HR and Ln RMSSD between the devices (Figure 3). Bland-Altman plots detailing the differences in the mean HR and Ln RMSSD between the reference and Emfit QS are shown in Figure 4. The Spearman rank correlation coefficient between the absolute differences and the average of the devices was r=0.39 (P=.09) for HR and r=0.53 (P=.02) for Ln RMSSD.

No differences were found in the 3-minute averaged HR values between the recordings of the reference device and Emfit QS (Figure 5). Significant differences in Ln RMSSD were found between the devices at 4 time points: 15 minutes (P=.01), 57 minutes (P=.05), 171 minutes (P=.04), and 450 minutes (P=.03). A very large correlation (r=0.72, P<.001) was found in the 3-minute averaged HR, and a large correlation (r=0.58, P<.001) was found in Ln RMSSD between the reference device and Emfit QS.

Figure 2. (A) Mean nocturnal heart rate (HR) and (B) the natural logarithm of the root mean square of successive differences between RR intervals (Ln RMSSD), measured by electrocardiogram (the reference) and ballistocardiogram (Emfit QS) (n=20). Box plots represent median values (solid line), 50th percentile values (box outline), and minimal and maximal values (whiskers).

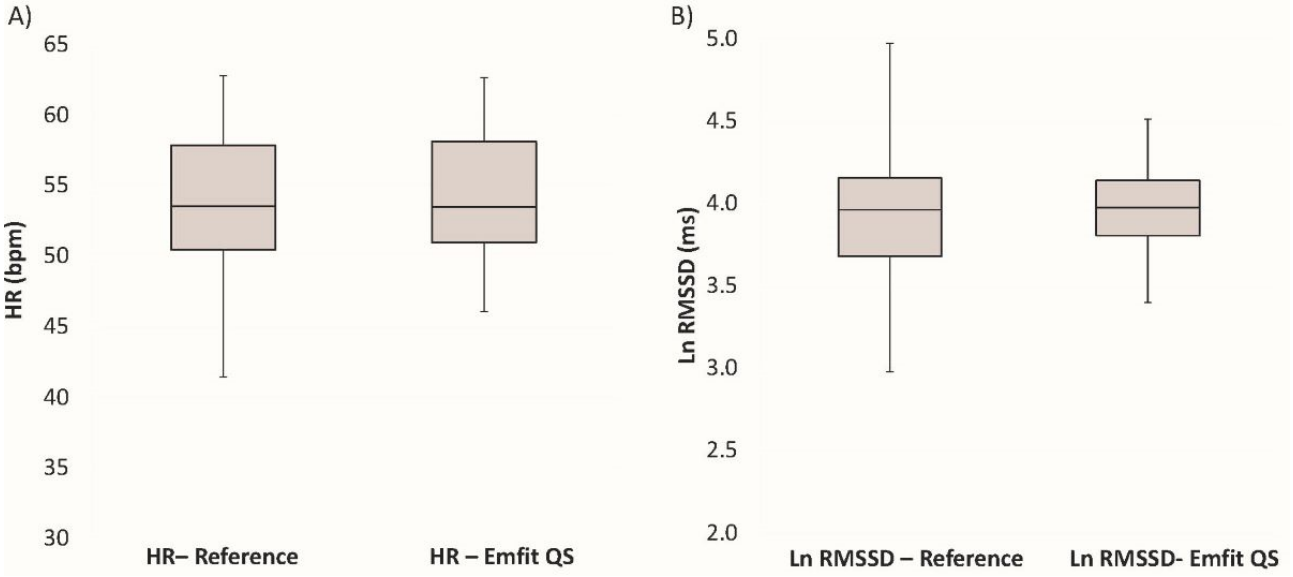


Table 1. Comparison between the mean (SD) values of Emfit QS and the reference device.

Variable	Bias[a] (absolute), mean (SD)	Bias[a] (%), mean (SD)	RMSE[b]	Effect size
HR[c] (bpm)	–0.8 (2.3)	–1.7 (4.6)	2.4	0.16 (trivial)
Ln RMSSD[d] (ms)	–0.05 (0.25)	–2.0 (6.5)	0.25	0.14 (trivial)

[a]Bias: the difference between the reference and Emfit QS.

[b]RMSE: the root mean square error.

[c]HR: heart rate.

[d]Ln RMSSD: natural logarithm of the root mean square of successive differences.

Figure 3. Correlation of the mean (A) heart rate (HR) and (B) the natural logarithm of the root mean square of successive differences (Ln RMSSD) between RR intervals between electrocardiogram (the reference) and ballistocardiogram (Emfit QS). The dashed line represents the line of equivalence (r=1.0).

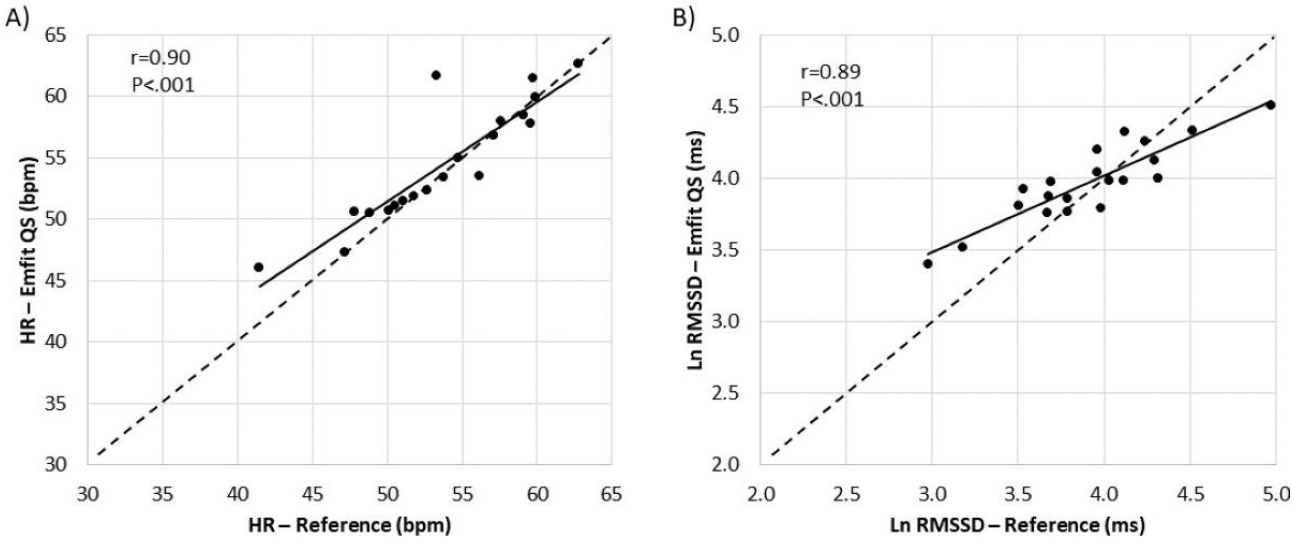


Figure 4. A Bland-Altman plot comparing the reference mean to that of Emfit QS: (A) heart rate (HR) and (B) the natural logarithm of the root mean square of successive differences (Ln RMSSD) between RR intervals.

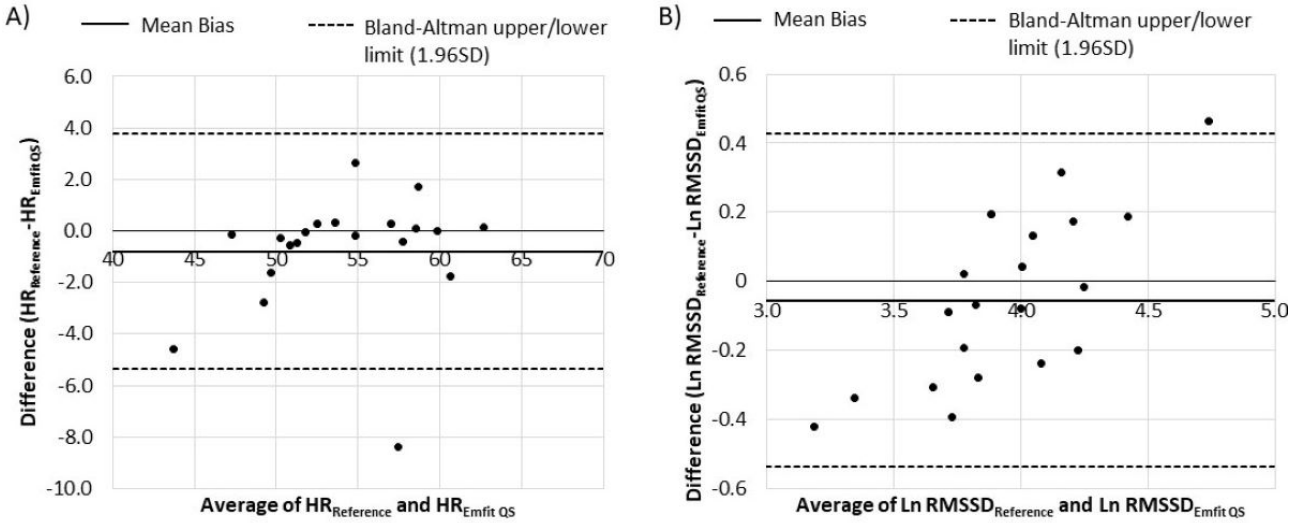
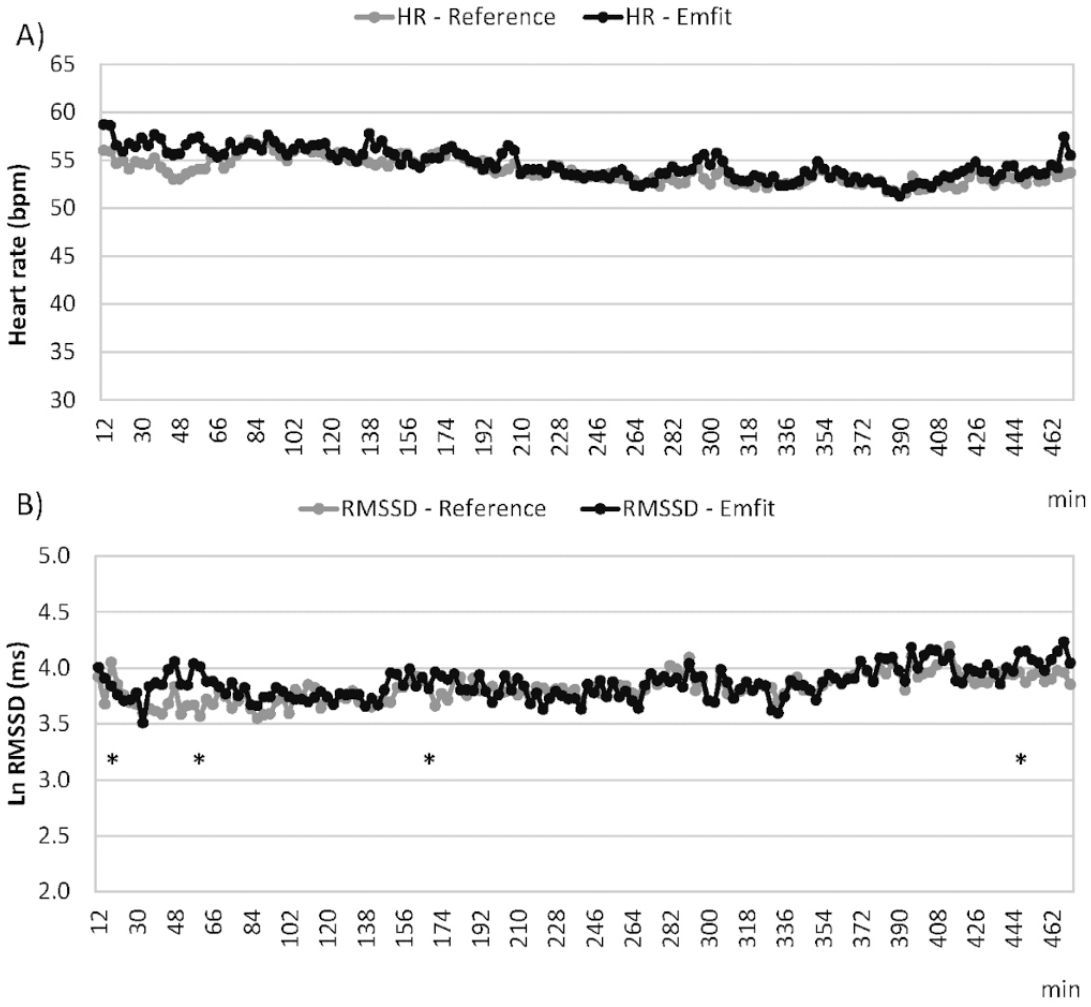


Figure 5. Nocturnal (A) heart rate (HR) (n=17-20) and (B) the natural logarithm of the root mean square of successive differences (Ln RMSSD) between RR intervals (n=9-19) in 3-minute epochs during sleep between recordings of electrocardiogram (the reference) and ballistocardiogram (Emfit QS). Significant difference between the devices noted (P<.05).



Discussion

Principal Results

To the best of our knowledge, this study is the first to evaluate the accuracy of BCG-based nocturnal HR and HRV measurements under real-life conditions. The main finding of this study is that BCG-based Emfit QS showed a trivial mean bias in nocturnal mean HR (1.7%, ES=0.16) and RMSSD (2.0%, ES=0.14) compared with the ECG-based reference device. In addition, the correlation coefficient showed good agreement (r=0.89-0.90) between the devices. Terbizan et al [17] suggested a minimum correlation of 0.9 for heart monitors to be clinically reliable. However, it needs to be acknowledged that a high correlation coefficient does not solely represent good agreement in all cases [18].

Comparison With Prior Work

Choe and Cho [19] found an RMSE of 1.8 bpm (beats per minute) for HR measured over a 15-minute rest period by a piezoelectric sensor (BCG) under laboratory conditions. The error was slightly smaller compared with that in our study (2.4 bpm). In addition, Xie et al [20] reported a 0.90 bpm mean absolute error in HR detection by a BCG sensor placed under a chair during 15-minute recordings and a greater correlation (r=0.98) compared with the present study. In previous studies, the participants were instructed to avoid movement since muscular activity may cause measurement errors in BCG. In this study, the measurements were carried out at home under real-life conditions. During sleep, movement can have interfering effects on HR and HRV detection, which can explain the slightly higher error and the amount of missing data in the present study.

For valid HRV determination, consecutive heart beats must be detected with a high degree of accuracy. Shin et al [21] observed a 5% relative error and a strong correlation (r=0.97) in a time-domain analysis of HRV by BCG and ECG. Wang et al [22] observed perfect correlations (r=0.99-1.00) in HRV variables between BCG and ECG. The authors concluded that HRV can be measured reliably with BCG. In this study, the mean bias in RMSSD was smaller (2.0%) than that in the study by Shin et al [21], but the correlation was weaker (r=0.89).

The Bland-Altman plots showed that 5% (1/20) of the mean HR and Ln RMSSD values were outside the limits of agreement (LoA). Bland and Altman [23] recommended that 95% of the data points should lie within the mean difference (SD 1.96). Our results showed a proportional error in the mean Ln RMSSD determined by Emfit QS. Based on the Bland-Altman plot, a larger error can be found in smaller and larger Ln RMSSD values; it seems that Emfit QS underestimates Ln RMSSD at high HRV levels and overestimates Ln RMSSD at low HRV levels. The LoA for the mean HR is relatively narrow (SD 4.6 bpm, ~8%). It is important that LoA would be narrower than daily changes in long-term monitoring. Al Haddad et al [24] reported a ~12% day-to-day variation in resting Ln RMSSD. In the present study, the LoA for the mean Ln RMSSD was ~12%. Thus, it can be concluded that it is not greater than the day-to-day variation in Ln RMSSD; the LoA is barely acceptable. Unfortunately, this research did not answer whether the biases are stable in repetitive measurements. Reliability is crucial for long-term monitoring of stress and recovery states; if the bias varies within an individual in repetitive measurements, the true changes in daily HR and HRV cannot be detected, and thus, an accurate interpretation of cardiac autonomic regulation will be compromised. In addition, it is good to standardize sleeping and measurement conditions (eg, bed, mattress) in long-term monitoring for reducing biases.

The 3-minute averaged data showed no differences in HR at any time points during the night and few differences in Ln RMSSD (Figure 5). The differences can be explained by the greater amount of missing data of Emfit QS; it is possible that a 3-minute epoch can include only, for example, 15 seconds of Emfit QS data; yet, the system provides a value for the 3-minute epoch. Comparatively, the reference data could include the entire 3-minute data, which could explain the differences in the 3-minute values. Thus, it seems that HR and HRV values for the entire night are more adequate. The overall view of nocturnal HRV is also more important than a single 3-minute value because HRV fluctuates during sleep according to sleep phases [25,26].

Although the bias and LoA analyses appear promising, future development in the accuracy of HRV measurement is needed to decrease the amount of missing data and incorrect 3-minute values. Because of the relatively large amount of missing data in HRV by Emfit QS compared to the reference data (28% vs 1%), the HRV data do not reflect values of the whole night for every individual. Suliman et al [11] observed large differences (77%-95%) in the ability to detect BCG peaks between different algorithms during a short resting period. Thus, the development of these algorithms may decrease missing data. It could be speculated that individuals with a high amount of missing data could also have more differences between the reference and Emfit QS, but our results showed no correlation between the amount of missing data and the mean bias of the measurements. Furthermore, future studies should clarify reliability over time for measuring HR and HRV by Emfit QS.

Practical Applications

Morning ECG measurements with HR straps are time-consuming and arduous to perform every day, and thus, compliance with regular measurements is poor [27]. Nocturnal HRV is a time-efficient method compared to morning HRV measurements taken at waking [28]. Previous studies have mainly focused on BCG measurements under laboratory conditions. This study showed the potential of Emfit QS to serve as a tool for everyday use at home to measure nocturnal HRV. Contact-free and fully automatic analysis by Emfit QS facilitates effortless daily monitoring of stress and recovery status among athletes. Furthermore, it does not require attaching electrodes on the body surface and thus does not disturb participants' typical sleep behaviors, which is advantageous over ECG. Based on our results, Emfit QS provides HR and HRV data with an acceptable, small mean bias compared with ECG. However, due to some large errors in detecting HR and HRV in some individuals, it would be best practice to ensure accuracy by comparing initial results from Emfit QS with ECG.

Conclusions

This study evaluated the accuracy of BCG-based Emfit QS for measuring HR and HRV. Our results showed that Emfit QS provides HR and HRV data with an acceptable, small mean bias compared with ECG. Thus, Emfit QS can be a potential tool for the long-term monitoring of HR and HRV. However, further research is needed to evaluate the reliability of HR and HRV detected by Emfit QS.

Acknowledgments

Funding for this study was provided by Emfit Ltd. The study was conducted by KIHU – Research Institute for Olympic Sports.

Conflicts of Interest

None declared.

References

1. Thompson W. World wide survey of fitness trends for 2018. ACSM's Health & Fitness Journal 2017;21(6):10-19. [doi: 10.1249/FIT.0000000000000341]
2. Peake JM, Kerr G, Sullivan JP. A Critical Review of Consumer Wearables, Mobile Applications, and Equipment for Providing Biofeedback, Monitoring Stress, and Sleep in Physically Active Populations. Front Physiol 2018 Jun 28;9. [doi: 10.3389/fphys.2018.00743]
3. Lambert MI, Mbambo ZH, Gibson ASC. Heart rate during training and competition for longdistance running. Journal of Sports Sciences 2011 Feb 07;16(sup1):85-90. [doi: 10.1080/026404198366713]
4. Bellenger CR, Fuller JT, Thomson RL, Davison K, Robertson EY, Buckley JD. Monitoring Athletic Training Status Through Autonomic Heart Rate Regulation: A Systematic Review and Meta-Analysis. Sports Med 2016 Feb 18;46(10):1461-1486. [doi: 10.1007/s40279-016-0484-2]
5. Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? Front Physiol 2014;5:73 [FREE Full text] [doi: 10.3389/fphys.2014.00073] [Medline: 24578692]
6. Plews DJ, Laursen PB, Stanley J, Kilding AE, Buchheit M. Training Adaptation and Heart Rate Variability in Elite Endurance Athletes: Opening the Door to Effective Monitoring. Sports Med 2013 Jul 13;43(9):773-781. [doi: 10.1007/s40279-013-0071-8]
7. de Zambotti M, Baker FC, Willoughby AR, Godino JG, Wing D, Patrick K, et al. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. Physiology & Behavior 2016 May;158:143-149. [doi: 10.1016/j.physbeh.2016.03.006]
8. Plews D, Scott B, Altini M, Wood M, Kilding A, Laursen P. Comparison of Heart-Rate-Variability Recording With Smartphone Photoplethysmography, Polar H7 Chest Strap, and Electrocardiography. Int J Sports Physiol Perform 2017 Nov 01;12(10):1324-1328. [doi: 10.1123/ijspp.2016-0668] [Medline: 28290720]
9. Eblen-Zajjur A. A simple ballistocardiographic system for a medical cardiovascular physiology course. Adv Physiol Educ 2003 Dec;27(1-4):224-229 [FREE Full text] [doi: 10.1152/advan.00025.2002] [Medline: 14627620]
10. Inan OT, Migeotte P, Park K, Etemadi M, Tavakolian K, Casanella R, et al. Ballistocardiography and Seismocardiography: A Review of Recent Advances. IEEE J Biomed Health Inform 2015 Jul;19(4):1414-1427. [doi: 10.1109/jbhi.2014.2361732]
11. Suliman A, Carlson C, Warren S, Thompson D. Performance evaluation of processing methods for ballistocardiogram peak detection. 2018 Jul 17 Presented at: Conf Proc IEEE Eng Med Biol Soc; 2018; Honolulu, Hawaii p. 502-505. [doi: 10.1109/embc.2018.8512317]
12. Emfit Ltd. Emfit: Sleep sensing solutions [photo] URL: https://www.emfit.com/ [accessed 2020-02-05]
13. Parak J, Korhonen I. Accuracy of firstbeat bodyguard 2 beat-to-beat heart rate monitor. Whitepaper 2013 [FREE Full text]
14. Bogdány T, Boros S, Szemerszky R, Köteles F. Validation of the firstbeat TeamBelt and Bodyguard2 systems. Magyar Sporttudományi Szemle(3) 2016:6-12 [FREE Full text]
15. Saalasti S. Neural networks for heart rate time series analysis. Doctoral thesis. Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä Studies in Computing 2003:33 [FREE Full text]
16. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive Statistics for Studies in Sports Medicine and Exercise Science. Medicine & Science in Sports & Exercise 2009;41(1):3-13. [doi: 10.1249/mss.0b013e31818cb278]
17. Terbizan DJ, Dolezal BA, Albano C. Validity of Seven Commercially Available Heart Rate Monitors. Measurement in Physical Education and Exercise Science 2002 Dec;6(4):243-247. [doi: 10.1207/s15327841mpee0604_3]
18. Zaki R, Bulgiba A, Ismail R, Ismail NA. Statistical Methods Used to Test for Agreement of Medical Instruments Measuring Continuous Variables in Method Comparison Studies: A Systematic Review. PLoS ONE 2012 May 25;7(5):e37908. [doi: 10.1371/journal.pone.0037908]
19. Choe S, Cho W. Simplified real-time heartbeat detection in ballistocardiography using a dispersion-maximum method. Biomedical Research 2017;28(9):3974-3985 [FREE Full text]

20. Xie Q, Wang G, Lian Y. Heart rate estimation from ballistocardiography based on hilbert transform and phase vocoder. Electrical Engineering ans Systems Science:arXiv.0 2018:3174 [FREE Full text] [doi: 10.1109/apccas.2018.8605724]
21. Shin JH, Hwang SH, Chang MH, Park KS. Heart rate variability analysis using a ballistocardiogram during Valsalva manoeuvre and post exercise. Physiol. Meas 2011 Jul 08;32(8):1239-1264. [doi: 10.1088/0967-3334/32/8/015]
22. Wang K, Zhu T, Zhang X, Yu C, Cao X, Tang J, et al. [Comparison of heart rate variability measurements between ballistocardiogram and electrocardiography]. Zhonghua Xin Xue Guan Bing Za Zhi 2015 May;43(5):448-451. [Medline: 26419993]
23. Bland J, Altman D. Measuring agreement in method comparison studies. Stat Methods Med Res 1999 Jun 01;8(2):135-160. [doi: 10.1191/096228099673819272]
24. Al Haddad H, Laursen P, Chollet D, Ahmaidi S, Buchheit M. Reliability of Resting and Postexercise Heart Rate Measures. Int J Sports Med 2011 May 13;32(08):598-605. [doi: 10.1055/s-0031-1275356]
25. Chouchou F, Desseilles M. Heart rate variability: a tool to explore the sleeping brain? Front Neurosci 2014 Dec 11;8:402 [FREE Full text] [doi: 10.3389/fnins.2014.00402] [Medline: 25565936]
26. Stickgold R, Walker MP, editors. The Neuroscience of Sleep. London: Academic Press/Elsevier; 2009.
27. Schäfer D, Gjerdalen GF, Solberg EE, Khokhlova M, Badtieva V, Herzig D, et al. Sex differences in heart rate variability: a longitudinal study in international elite cross-country skiers. Eur J Appl Physiol 2015 May 23;115(10):2107-2114. [doi: 10.1007/s00421-015-3190-0]
28. Herzig D, Testorelli M, Olstad D, Erlacher D, Achermann P, Eser P, et al. Heart-Rate Variability During Deep Sleep in World-Class Alpine Skiers: A Time-Efficient Alternative to Morning Supine Measurements. Int J Sports Physiol Perform 2017 May;12(5):648-654. [doi: 10.1123/ijspp.2016-0257] [Medline: 27768512]

Abbreviations

BCG: ballistocardiography
BG2: Bodyguard 2
bpm: beats per minute
ECG: electrocardiography
ES: effect size
HR: heart rate
HRV: heart rate variability
Ln: natural logarithm
LoA: limits of agreement
PPG: photoplethysmography
RMSE: root mean square error
RMSSD: root mean square of successive differences
RRI: RR interval

Edited by G Eysenbach; submitted 09.10.19; peer-reviewed by C Bellenger, H Durrani; comments to author 26.01.20; revised version received 05.02.20; accepted 01.03.20; published 23.04.20.

Please cite as:
Vesterinen V, Rinkinen N, Nummela A
A Contact-Free, Ballistocardiography-Based Monitoring System (Emfit QS) for Measuring Nocturnal Heart Rate and Heart Rate Variability: Validation Study
JMIR Biomed Eng 2020;5(1):e16620
URL: http://biomedeng.jmir.org/2020/1/e16620/
doi:10.2196/16620
PMID:

Original Paper

Effect of Platform Swing Walkway on Locomotor Behavior in Children With Diplegic Cerebral Palsy: Randomized Controlled Trial

Hanaa Mohsen[1], PhD; Omnya Samy[1], MSc

Department of Physical Therapy for Pediatrics and Pediatric Surgery, College of Physical Therapy, Badr University, Cairo, Egypt

Corresponding Author:
Hanaa Mohsen, PhD
Department of Physical Therapy for Pediatrics and Pediatric Surgery
College of Physical Therapy
Badr University
Badr City
Cairo, 022
Egypt
Phone: 20 01009106127
Email: hanaa753@gmail.com

Abstract

Background: Limited attention has been given to the effectiveness of the platform swing walkway, which is a common way to improve gait pattern through activation of sensory stimuli (visual, auditory, vestibular, and somatosensory).

Objective: The objective of this study was to determine the effect of a platform swing walkway on gait parameters in children with diplegic cerebral palsy (CP).

Methods: A total of 30 children of both sexes (aged 6-8 years) with diplegic CP were enrolled in this study. They were randomly assigned into two groups of equal number: the control group (n=15) and the study group (n=15). The control group received the conventional physical therapy plan, whereas the study group received the same conventional physical therapy program in addition to gait training on a platform swing walkway. Temporal parameters during the gait cycle were collected using gait tracker video analysis, and the Growth Motor Function Measure Scale (GMFM-88) was used to assess standing and walking (Dimensions D and E) before and after the treatment program.

Results: A statistically significant improvement in both groups was noted when comparing the mean values of all measured variables before and after treatment ($P \leq .05$). There were significant differences between the control and study groups with respect to all measured variables, which favored the study group when comparing the posttreatment outcomes ($P \leq .05$).

Conclusions: Results suggest that gait training on platform swing walkways can be included as an alternative therapeutic modality to enhance gait parameters and gross motor function in children with diplegic CP.

Trial Registration: ClinicalTrials.gov NTC04246658; https://clinicaltrials.gov/ct2/show/NTC04246658

(JMIR Biomed Eng 2020;5(1):e18232) doi:10.2196/18232

KEYWORDS

cerebral palsy; platform swing walkway; spastic diplegia

Introduction

Cerebral palsy (CP) occurs as a result of prenatal or postnatal lesion in the developing brain of a fetus or infant, which mainly affects motor activity [1]. Because of the complex relationship between primary and secondary motor symptoms in CP, for example, between spasticity and muscle contracture, the diagnosis of gait impairments varies between patients. As a result of irregular muscle activity or bone loading, secondary impairments can develop over time, such as shortened muscles, which limit the joint range of motion. Both primary and secondary impairments manifest in a pathological pattern of CP gait [2].

In children with spastic diplegia, abnormal muscle tone can lead to medial femoral torsion and compensatory external tibial torsion, which result in in-toing and crouch gait, thereby

decreasing the stability during walking [3]. This gait disorder decreases the base of support in the stance phase and increases crossing of the legs in the swing phase that increases the risk of falling and leads to functional balance problems [4].

Locomotion arises from intricate dynamic interactions between a central program and feedback mechanisms. This central program relies on a genetically determined spinal circuit capable of generating basic locomotion patterns, as well as neural drive through various descending pathways that can trigger, stop, and/or steer locomotion. Sensory feedback from muscle and skin afferents, as well as other sensory modalities (vision, audition, vestibular), dynamically adapts the locomotion pattern to the requirements of the environment [5].

Researchers have demonstrated that symmetric weight-bearing training on unstable surface improves patients' performance in activities of daily living, by using, for example, auditory feedback, task-orientation training, and lower extremity elevation method [6]. In particular, unstable surface has been shown to enhance ankle and knee joint stability, lower extremity strength, muscle activation, proprioception, and balance control [7-9]. In addition, unstable surface has been reported to be more effective than other techniques for reducing postural sway while maintaining the standing position [10]. Chaudhuri and Aruin [6] stated that individuals with hemiparesis exposed to platform translations showed improved balance control seen as decreased latency and increased strength of the paretic leg force response.

However, most studies conducted to date have only evaluated the effects of platform swing on balance ability. Little is known about the effects of platform swing on lower extremity function, especially walking performance. Therefore, the purpose of this study was to examine the effects of platform swing walkway on gait parameters and gross motor function in children with diplegic CP.

Methods

Study Design

This is a randomized, single-blind, controlled clinical trial. Eligible children were recruited from the Faculty Physical Therapy Outpatient Clinic, Cairo University, Badr University in Cairo and Prof Dr Kamal Shoukry pediatric rehabilitation center.

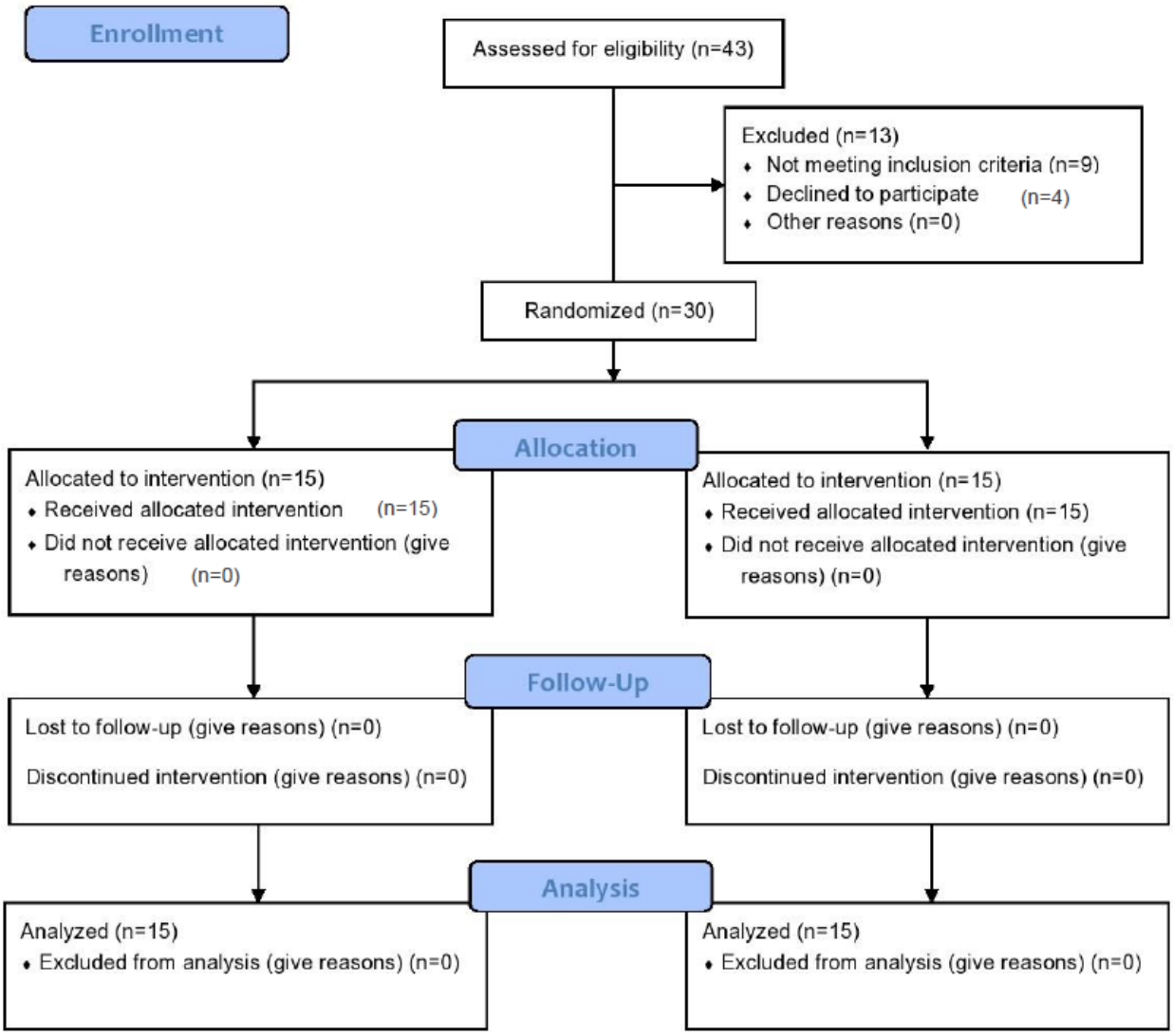
Participants

The study recruited 30 children with diplegic CP of both sexes (18 boys and 12 girls). Children were included only if they met the following inclusion criteria: aged between 6 and 8 years, diagnosed with diplegic CP (which was verified with magnetic resonance images collected from medical records), having spasticity of Grade 1 upper limb to 1+ lower limb on the Modified Ashworth Scale (MAS) [11], in Level II or III according to the Gross Motor Function Classification System [12], used the standard orthotic management according to their individual abilities, and were cognitively capable and competent for interpreting and following instructions. Children who had vision or hearing problems, previous surgery of the lower extremities, botulinum injections of the lower limb muscles within the preceding 6 months, and suffering from cardiopulmonary disorders were excluded. Prior to data collection, the study purposes, procedures, and benefits were fully explained to the parents of the participating children. All parents gave informed consent to have their children participate in the study. This randomized controlled study was approved by the Ethical Review Committee of the Faculty of Physical Therapy, Cairo University (Approval no. P.T.REC/012/002511).

Randomization

A total of 43 children with diplegic CP were recruited for this study; 9 children did not meet the inclusion criteria, and the parents of 4 children refused to give consent. Following the baseline examination, concealed allocation was performed using a computer-generated randomized table of letters created prior to the start of data collection by a researcher not involved in the recruitment or treatment of patients. Individual and sequentially lettered index cards were used to randomly assign participants to the treatment groups. The index cards were folded and placed in sealed, opaque envelopes. A second therapist, blinded to baseline examination findings, opened the envelope and proceeded with treatment according to the group assignment. Each participant received a sealed envelope containing one of the letters A or B. The flowchart of the experimental design is shown in Figure 1.

Figure 1. Participants' flow diagram.



Data Collection Procedures

All procedures were performed at baseline (pretreatment) and at the end of 3 successive months of treatment (posttreatment) in a warm, lighted, and quiet room.

Gross Motor Functional Measurement

The Growth Motor Function Measure Scale (GMFM) is an observational assessment tool incorporating 88 items (GMFM-88), scored on a 4-point ordinal scale (ranging from 0 to 3) [13]. It is further subdivided into 5 domains: (A) lying and rolling; (B) sitting; (C) crawling and kneeling; (D) standing; and (E) walking, running, and jumping. In this study, only the complete dimensions D and E were assessed before and after the treatment program. We calculated the dimensions (%) as follows: (D) Standing (Total Dimension D/39) × 100 = %; (E) Walking (Total Dimension E/72) × 100 = %.

Gait Tracker Video Analysis

In this analysis, a 2D kinematic analysis of the lower limb sagittal plane was performed using a single video frame. It provides unilateral joint kinematics data of the hip, knee, and ankle in the sagittal plane, along with the estimation of gait events and spatiotemporal parameters. The precision of the spatiotemporal parameters estimation was found to be appropriate for clinical use [14]. Participating children were asked to walk at a self-selected velocity along a 2-m walkway. A digital video camera (Canon) was set up to record the children in the sagittal plane on a level tripod, perpendicular to the center of the pathway at a distance of 3 m. As the lower limb will be analyzed, the camera was aligned with the knee and set to view the 2 m of the walkway, which ensured that the calibration area covered the lower limb. The recorded video was analyzed using Tracker software [15].

Intervention

Platform Swing Walkway Features

The platform swing walkway (Figure 2) comprises the following: (1) a durable vestibular metal frame consisting of a ceiling surface and upright metal bars. It has a working load with a floor space of 3 m width × 90 cm depth and a height of 2 m; (2) 6 iron chains (length: 150 cm) with a ceiling hook used for attaching vestibular components to the architectural suspension points; (3) a hard wooden platform swing (floor space: 2.44 m width × 70 cm depth) that is set 50 cm above the ground (floor). The swing is completely coated with Tumble Forms and contains the hardware for attaching the suspension iron chain; (4) 12 ropes of length 1.22 cm with 2 hooks that are used for attaching the iron chains which provide adjustable side rails and safety for children while walking. Height adjustment assembly allows for quick changes in horizontal position. Removable components allow for portability and ease of storage, as well as provide linear acceleration in forward–backward and side-to-side directions.

Figure 2. Dynamic Platform Swing Walkway.

The Control Group

Children in the control group received the conventional physiotherapy program which consisted of facilitation of postural mechanism; proprioceptive training, including weight-bearing activities for upper and lower limbs; strengthening exercise for the back, abdominal muscles, hip flexors, ankle dorsiflexors, knee flexors, knee extensors, hip extensors, hip abductors, and hip internal and external rotators; standing exercise (standing holding on, standing alone with arms free, standing holding on and asking the child to lift one foot, standing on one leg, and standing on balance board); gait training activities in different directions at different speeds (walking up and down the stairs, jumping in place, and broad jumping). The duration of the program was 1 hour, and required to be performed 3 days a week for 12 weeks.

The Study Group

Children in the study group received the same physical therapy program given to the control group for 45 minutes. They additionally received gait training on a platform swing walkway for 15 minutes, which included the following steps: (1) The therapist put the child on the platform, and (2) encouraged the child to stand on the platform in a relaxed state and to hold cords of the walkway; (3) the child was then asked to walk on the platform in forward and backward directions to make sure that he or she adapted to the situation before swinging; (4) the child was asked to overcome simple obstacles such as a stepper and

a separator. This program was performed 3 days a week for 12 weeks.

Data Analysis

G power analysis (power=0.8, α=.05, effect size=0.5) determined a sample size of 30 for this study. Descriptive statistics of mean and standard deviation were calculated for all measured variables. Parametric tests (paired *t* test and unpaired *t* test) were used to analyze the pre- and posttreatment mean values of GMFM-88. In addition, gait cycle time (seconds) for right and left stance and swing phase was calculated. SPSS (version 20 for Windows; IBM Corp) was used for data analysis. Statistical tests were considered significant if $P \leq .05$.

Results

Table 1 presents age, weight, and height of children in the study and control groups. There was no significant difference between both groups in mean (SD) age, weight, and height ($P \leq .05$). Table 2 presents the comparison between groups in the standing and walking tasks. There were no significant differences in pretreatment mean values between the groups in the variables measured: standing ($P=.97$) and walking ($P=.91$). However, when comparing the posttreatment mean values in the standing and walking tasks between the study and control groups, significant differences ($P=.004$ and .001, respectively) in favor of the study group were found.

Table 3 presents the results of within-group comparison in the standing and walking tasks for the two groups. Upon comparing their pre- and posttreatment values, there was a significant improvement in all measured variables in both groups ($P \leq .05$).

Table 4 compares the results between the two groups in the stance and swing phases. There were no significant differences in pretreatment mean values between the study and control groups in the variables measured with regard to the stance of the right ($P=.64$) and left limbs ($P=.66$) nor with regard to the swing of the right ($P=.64$) and left limbs ($P=.61$). When comparing the posttreatment mean values between the study and control groups with regard to the stance and swing phases of the right and left limbs a significant difference in favor of the study group was found ($P \leq .5$).

Table 5 presents the results of within-group comparison in the stance and swing phases of the right and left limbs, respectively, for the two groups. A significant improvement in all measured variables was observed in both groups when comparing their pre- and posttreatment mean values ($P \leq .5$).

Table 1. Demographic characteristics of children in the two groups.

Items	Study group, mean (SD)	Control group, mean (SD)	Comparison	
			t value[a]	*P* value
Age (years)	7.34 (0.7)	7.62 (0.64)	–1.14	.26*
Weight (kg)	29.2 (2.88)	30.13 (2.85)	–0.89	.38*
Height (cm)	126.53 (3.09)	127.06 (2.63)	–0.5	.61*

*Nonsignificant.

[a]From unpaired *t* test.

Table 2. Comparison of standing and walking between both groups.

Period	Control group, mean (SD)	Study group, mean (SD)	*t* value[a]	*P* value*
Pre				
Standing	0.441 (0.099)	0.440 (0.138)	0.030	.97
Walking	0.253 (0.076)	0.250 (0.093)	0.107	.91
Post				
Standing	0.556 (0.105)	0.674 (0.100)	–3.132	.004
Walking	0.310 (0.073)	0.4147 (0.073)	–3.924	.001

*Significant at $P<.05$.

[a]From unpaired *t* test.

Table 3. Comparison of standing and walking within groups.

Group	Standing				Walking			
	Pre, mean (SD)	Post, mean (SD)	*t* value[a]	*P* value*	Pre, mean (SD)	Post, mean (SD)	*t* value[a]	*P* value*
Control	0.441 (0.099)	0.556 (0.105)	–13.06	.001	0.253 (0.076)	0.310 (0.073)	–11.244	.001
Study	0.440 (0.138)	0.634 (0.121)	–10.72	.001	0.250 (0.093)	0.377 (0.119)	–9.964	.001

*Significant at $P<.05$.

[a]From paired *t* test.

Table 4. Comparison of stance and swing between both groups.

Site	Control group, mean (SD)	Study group, mean (SD)	*t* value[a]	*P* value
Right limb				
Pre				
Stance	83.600 (2.898)	82.866 (3.583)	0.616	.64
Swing	16.400 (2.898)	17.1333 (3.583)	–0.616	.64
Post				
Stance	79.8000 (3.569)	69.0667 (6.933)	5.331	.013
Swing	20.2000 (3.569)	30.93 (6.933)	–5.331	.013
Left limb				
Pre				
Stance	77.600 (5.179)	77.6667 (5.839)	–0.033	0.66
Swing	22.400 (5.179)	22.4000 (5.865)	0.001	.61
Post				
Stance	72.3333 (7.752)	67.066 (3.990)	2.339	.12
Swing	27.666 (7.752)	32.933 (3.990)	–2.339	.12

*Significant at $P<.05$.

[a]From unpaired *t* test.

Table 5. Comparison of stance and swing within groups.

Site	Stance				Swing			
	Pre, mean (SD)	Post, mean (SD)	*t* value[a]	*P* value	Pre, mean (SD)	Post, mean (SD)	*t* value[a]	*P* value
Right limb								
Control group	83.6 (2.89)	79.8 (3.57)	3.37	.005	16.4 (2.89)	20.20 (3.569)	–3.37	.005
Study group	82.8 (3.58)	69.1 (6.93)	6.32	.001	17.13 (3.58)	30.93 (6.93)	–6.32	.001
Left limb								
Control group	77.6 (5.18)	72.3 (7.75)	2.280	.039	22.4 (5.18)	27.66 (7.752)	–2.28	.039
Study group	77.6 (5.84)	67.1 (3.99)	6.448	.001	22.4 (5.86)	32.93 (3.9)	–6.37	.001

*Significant at $P<.05$.

[a]From paired *t* test.

Discussion

Principal Results

This study investigated the effectiveness of a platform swing walkway on standing and walking tasks and on temporal gait parameters in children with diplegia which is the main spastic type among the different ones reported. Yokochi [16] also reported that spastic diplegia is the common type of CP that accounts for about 44% of infants with CP and about 80% of those with prematurity.

At the end of the treatment period (12 weeks), all children showed significant improvement in gross motor function and temporal gait parameters after walking on the platform swing walkway, which may be attributed to the increase in muscle strength and endurance. This result is in agreement with Dean and Shuaib [17] who found that an unstable surface increased muscle endurance and postural control compared with training on a typical surface. Our result is also in agreement with findings of Prosser et al [18] who stated that the choice of specific equipment should be based on clinical objectives (eg, enhancing muscle strength, enhancing reciprocal muscle activation, or proper muscle activity while walking) and functional capacities of the child.

Gait training on the platform swing walkway can increase the motility of the lower limb joints and activation of lower limb muscles, leading to neuromuscular involvement that improves gait ability, resulting in increased gait endurance. This result is in agreement with Bohannon [19] who reported that the activation of hip extensors, knee extensors, and ankle plantar flexors on the affected side is significantly related to maintaining or increasing movement velocity, and thus, an increase in gait velocity reflects an improvement in overall gait abilities.

The significant improvement in GMFM-88 and temporal gait parameters could be attributed to the effect of vestibular stimulation over time which tends to enhance sensory integration, standing, and walking in individuals with CP. These

findings support that vestibular stimulation can enhance arousal rates, visual exploratory activity, motor development, balance, and reflex coordination in at-risk infants and in those with developmental retardation disorders [20].

In this study, the improvement in gross motor function and gait parameters may be attributed to the organization of the vestibular system and proprioception. This is in agreement with the finding by Herdman and Clendaniel [21] who stated that peripheral sensory equipment includes a variety of motion sensors (visual, vestibular, and proprioceptive) that send information about head angular velocity and linear acceleration to the CNS, especially to the vestibular nucleus complex and the cerebellum. The CNS interprets these signals and integrates them with other sensory information to accurately predict head and body orientation.

In addition, improvement in motor performance may be attributed to the visual and proprioception input. This fact is supported by findings of Morningstar et al [22], who concluded that visual and vestibular input as well as joint and soft tissue mechanoreceptors play an important role in the regulation of static upright posture. Besides, Prokop et al [23] stated that a combination of visual and proprioceptive information is important for the modulation of walking velocity and indicates that visual information modifies stride length while proprioceptive input maintains a constant stride frequency, leading to a shift in walking velocity.

The findings of this study provide evidence that good sensory integration involves stimulation of the vestibular, proprioceptive, and tactile systems, as a means of exploring new skills. The activities are pitched at a level that stimulates and challenges, yet within the child's capabilities. The efficacy of promoting sensory integration by vestibular stimulation is a key component of integrated sensory therapy. Activities involved in this type of approach lead to the development of a good body scheme, self-image, integration of primitive reflexes, balance, postural stability, motor planning, coordination of both sides of the body, and eye–hand coordination [24]. In this aspect, Lamoth et al [25] found that trunk coordination has an effect on gait parameters and that flexible adaptation in trunk coordination to changes in walking velocity is considered a hallmark of unaffected gait.

Functional training with unstable support is useful for increasing exercise difficulty because unexpected proprietary sensory information and reaction forces in various directions are induced in comparison with training methods on a stable support surface [26-28]. In particular, walking on an unstable support surface such as a balance pad requires more muscle strength and movement around the ankle joint, thereby improving walking speed [28,29].

In this study, functional abilities of children in the study group improved after walking on the platform swing walkway. This result is supported by the work by Jobling and Cuskelly [30] who found that children receiving neurodevelopmental therapy or a combination of neurodevelopmental therapy and other interventions such as vestibular stimulation or sensory integration performed better than those receiving other services.

Some limitations of this study are the relatively small sample size and lack of follow-up. Therefore, larger experimental studies are necessary to define the subcategories of children with CP most likely to benefit from gait training on the platform swing walkway. The strengths of the study are the integration of safety precautions and accommodation during gait training on the platform swing walkway.

Conclusion

Based on our study results, it can be concluded that platform swing walkway could improve gross motor function and gait patterns in children with spastic diplegia. The platform swing walkway was tolerable at the protocol settings applied and provides the basis for stimulation of safety and efficacy. Future studies should examine the long-term effect of platform swing walkway on locomotor behavior in different types of CP.

Conflicts of Interest

None declared.

References

1. Bax M, Goldstein M, Rosenbaum P, Leviton A, Paneth N, Dan B, Executive Committee for the Definition of Cerebral Palsy. Proposed definition and classification of cerebral palsy, April 2005. Dev Med Child Neurol 2005 Aug;47(8):571-576 [FREE Full text] [doi: 10.1017/s001216220500112x] [Medline: 16108461]
2. Õunpuu S, Gorton G, Bagley A, Sison-Williamson M, Hassani S, Johnson B, et al. Variation in kinematic and spatiotemporal gait parameters by Gross Motor Function Classification System level in children and adolescents with cerebral palsy. Dev Med Child Neurol 2015 Oct;57(10):955-962 [FREE Full text] [doi: 10.1111/dmcn.12766] [Medline: 25926016]
3. Ryan DD, Rethlefsen SA, Skaggs DL, Kay RM. Results of tibial rotational osteotomy without concomitant fibular osteotomy in children with cerebral palsy. J Pediatr Orthop 2005;25(1):84-88. [doi: 10.1097/00004694-200501000-00019] [Medline: 15614066]
4. Carmick J. Forefoot mobility in ankle and foot orthoses: effect on gait of children with cerebral palsy. Pediatr Phys Ther 2013;25(3):331-337. [doi: 10.1097/PEP.0b013e31828e30ac] [Medline: 23685740]
5. Rossignol S, Dubuc R, Gossard J. Dynamic sensorimotor interactions in locomotion. Physiol Rev 2006 Jan;86(1):89-154 [FREE Full text] [doi: 10.1152/physrev.00028.2005] [Medline: 16371596]
6. Chaudhuri S, Aruin AS. The effect of shoe lifts on static and dynamic postural control in individuals with hemiparesis. Arch Phys Med Rehabil 2000 Nov;81(11):1498-1503. [doi: 10.1053/apmr.2000.17827] [Medline: 11083355]

7. Borreani S, Calatayud J, Martin J, Colado JC, Tella V, Behm D. Exercise intensity progression for exercises performed on unstable and stable platforms based on ankle muscle activation. Gait Posture 2014;39(1):404-409. [doi: 10.1016/j.gaitpost.2013.08.006] [Medline: 23999147]
8. Carter JM, Beam WC, McMahan SG, Barr ML, Brown LE. The effects of stability ball training on spinal stability in sedentary individuals. J Strength Cond Res 2006 May;20(2):429-435. [doi: 10.1519/R-18125.1] [Medline: 16686575]
9. Verhagen E, Bobbert M, Inklaar M, van Kalken M, van der Beek A, Bouter L, et al. The effect of a balance training programme on centre of pressure excursion in one-leg stance. Clin Biomech (Bristol, Avon) 2005 Dec;20(10):1094-1100. [doi: 10.1016/j.clinbiomech.2005.07.001] [Medline: 16129528]
10. Bayouk J, Boucher J, Leroux A. Balance training following stroke: effects of task-oriented exercises with and without altered sensory input. Int J Rehabil Res 2006 Mar;29(1):51-59. [doi: 10.1097/01.mrr.0000192100.67425.84] [Medline: 16432390]
11. Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth scale of muscle spasticity. Phys Ther 1987 Feb;67(2):206-207. [Medline: 3809245]
12. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. Dev Med Child Neurol 1997 Apr;39(4):214-223 [FREE Full text] [doi: 10.1111/j.1469-8749.1997.tb07414.x] [Medline: 9183258]
13. Russell DJ, Rosenbaum PL, Avery LM, Lane M. Gross Motor Function Measure (GMFM-66 & GMFM-88) User's Manual. London, UK: Mac Keith Press; Nov 2002.
14. Baker R. Gait analysis methods in rehabilitation. J Neuroeng Rehabil 2006 Mar 02;3:4 [FREE Full text] [doi: 10.1186/1743-0003-3-4] [Medline: 16512912]
15. Rodríguez I, Martín J. Effect of marker misplacement on 3D Instrumented Gait Analysis kinematic measurements. Gait & Posture 2014 Jun;39:S112 [FREE Full text] [doi: 10.1016/j.gaitpost.2014.04.154]
16. Yokochi K. Gait patterns in children with spastic diplegia and periventricular leukomalacia. Brain Dev 2001 Mar;23(1):34-37. [doi: 10.1016/s0387-7604(00)00200-x] [Medline: 11226727]
17. Dean N, Shuaib A. Transient ischaemic attacks: unstable, treatable, neglected. Lancet 2007 Oct 20;370(9596):1398-1400. [doi: 10.1016/S0140-6736(07)61449-4] [Medline: 17928045]
18. Prosser LA, Stanley CJ, Norman TL, Park HS, Damiano DL. Comparison of elliptical training, stationary cycling, treadmill walking and overground walking. Electromyographic patterns. Gait Posture 2011 Feb;33(2):244-250 [FREE Full text] [doi: 10.1016/j.gaitpost.2010.11.013] [Medline: 21215636]
19. Bohannon RW. Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants. Age Ageing 1997 Jan;26(1):15-19. [Medline: 9143432]
20. Behm DG, Leonard AM, Young WB, Bonsey WAC, MacKinnon SN. Trunk muscle electromyographic activity with unstable and unilateral exercises. J Strength Cond Res 2005 Feb;19(1):193-201. [doi: 10.1519/1533-4287(2005)19<193:TMEAWU>2.0.CO;2] [Medline: 15705034]
21. Herdman SJ, Clendaniel R. Vestibular Rehabilitation. Philadelphia, PA: F.A. Davis and Company; 2014.
22. Morningstar MW, Pettibon BR, Schlappi H, Schlappi M, Ireland TV. Reflex control of the spine and posture: a review of the literature from a chiropractic perspective. Chiropr Osteopat 2005 Aug 09;13:16. [doi: 10.1186/1746-1340-13-16] [Medline: 16091134]
23. Prokop T, Schubert M, Berger W. Visual influence on human locomotion. Modulation to changes in optic flow. Exp Brain Res 1997 Mar;114(1):63-70. [doi: 10.1007/pl00005624] [Medline: 9125452]
24. An SL. The effects of vestibular stimulation on a child with hypotonic cerebral palsy. J Phys Ther Sci 2015 Apr;27(4):1279-1282 [FREE Full text] [doi: 10.1589/jpts.27.1279] [Medline: 25995606]
25. Lamoth CJC, Stins JF, Pont M, Kerckhoff F, Beek PJ. Effects of attention on the control of locomotion in individuals with chronic low back pain. J Neuroeng Rehabil 2008 Apr 25;5:13 [FREE Full text] [doi: 10.1186/1743-0003-5-13] [Medline: 18439264]
26. Lamoth CJC, Stins JF, Pont M, Kerckhoff F, Beek PJ. Effects of attention on the control of locomotion in individuals with chronic low back pain. J Neuroeng Rehabil 2008 Apr 25;5:13 [FREE Full text] [doi: 10.1186/1743-0003-5-13] [Medline: 18439264]
27. Lee SH. The differences between aero step exercises and weight training on posture, physical fitness, balance, and hormone levels in the elderly. In: Master's thesis. Seoul, South Korea: Ewha Womans University; 2007.
28. Song GB, Park EC. Effect of dual tasks on balance ability in stroke patients. J Phys Ther Sci 2015 Aug;27(8):2457-2460 [FREE Full text] [doi: 10.1589/jpts.27.2457] [Medline: 26357425]
29. Kim EJ. The effects of gait training on treadmill and unstable surface and muscular activity in stroke patient. In: Master's degree thesis. Gyeongsan, South Korea: Daegu University; 2009.
30. Jobling A, Cuskelly M. Young people with Down syndrome: a preliminary investigation of health knowledge and associated behaviours. J Intellect Dev Disabil 2006 Dec;31(4):210-218. [doi: 10.1080/13668250600999186] [Medline: 17178533]

Abbreviations

CNS: central nervous system
CP: cerebral palsy
GMFM-88: 88-item Growth Motor Function Measure Scale
MAS: Modified Ashworth Scale

Edited by G Eysenbach; submitted 13.02.20; peer-reviewed by A Salim, M Abd El-Fattah; comments to author 07.03.20; revised version received 26.04.20; accepted 13.05.20; published 21.07.20.

Please cite as:
Mohsen H, Samy O
Effect of Platform Swing Walkway on Locomotor Behavior in Children With Diplegic Cerebral Palsy: Randomized Controlled Trial
JMIR Biomed Eng 2020;5(1):e18232
URL: http://biomedeng.jmir.org/2020/1/e18232/
doi:10.2196/18232
PMID:

Original Paper

Ease of Use of the Electroconvulsive Therapy App by Its Users: Cross-Sectional Questionnaire Study

Kinza Khan[1], MRCPsy; Chethan Basavarajappa[2], MBBS, DNB, FIPR; Girish Kunigiri[1], MD, DNB, MBA

[1]Bradgate Mental Health Unit, Leicester, United Kingdom

[2]Department of Psychiatry, National Institute of Mental Health And Neuro Sciences, Bengaluru, India

Corresponding Author:
Girish Kunigiri, MD, DNB, MBA
Bradgate Mental Health Unit
Leicester
Leicester, LE3 9EJ
United Kingdom
Phone: 44 1530 453827
Email: Girish.Kunigiri@leicspart.nhs.uk

Abstract

Background: Electroconvulsive therapy (ECT) is one of the oldest, most effective, and potentially life-saving noninvasive brain stimulation treatments for psychiatric illnesses such as severe depression, mania, and catatonia. The decision-making process to use ECT involves well-informed discussion between the clinician and the client. A platform, like an app, which provides this information in an easy-to-understand format may be of benefit to various stakeholders in making an informed decision. Apps developed by clinicians/hospitals taking into consideration user perspectives will filter and provide trustworthy information to the users. In this regard, the ECT app, an app which is freely available for download at the Apple Store, was developed by the Leicestershire Partnership National Health Service (NHS) Trust and Leicestershire Health Informatics Service (LHIS).

Objective: The objective of this study is to evaluate and demonstrate the accessibility of the ECT app to the chosen audiences it was created for, via a paper and electronic questionnaire.

Methods: A survey was conducted between January 2017 and March 2019. A survey questionnaire designed for the study was sent to mental health professionals, medical students, patients, carers, and members of the public via post, email, and SurveyMonkey or informed via posts shared in Psychiatry online groups and face-to-face contact. All participants who were willing to participate in the study were included.

Results: Results were collected via paper forms, email responses, and SurveyMonkey and all were inputted into SurveyMonkey to facilitate analysis. A total of 20 responses were received during the study period (January 2017 to March 2019). The participants of the survey, which included a mixed group of professionals (12/20, 60%), patients (3/20, 15%), and carers (1/20, 5%), opined that the app was easy to download (14/20, 70%) and use (9/20, 45%); contained adequate information (19/20, 95%); they felt more informed after having used the app (9/20, 45%); and they would recommend it to others (19/20, 95%). The participants of the survey also provided suggestions on the app (10/20, 50%).

Conclusions: The ECT app can be beneficial in sharing appropriate information to professionals and the public alike and help in gathering unbiased and nonjudgmental information on the current use of ECT as a treatment option.

(JMIR Biomed Eng 2020;5(1):e20730) doi:10.2196/20730

KEYWORDS

mHealth; depression; apps; electroconvulsive therapy; smartphone; mobile phone; surveys; psychiatry

Introduction

Electroconvulsive therapy (ECT) is one of the oldest, most effective, and potentially life-saving noninvasive brain stimulation treatments for psychiatric illnesses such as severe depression, mania, and catatonia. The National Institute for Health and Care Excellence (NICE) recommends use of ECT only to achieve rapid and short-term improvement in treatment-resistant or life-threatening conditions [1]. The decision-making process to use ECT involves well-informed discussion between the clinician and the client. NICE guidelines also recommend development of leaflets to aid this cause [1].

A platform which provides this information in an easy-to-understand format may be of benefit to various stakeholders in making an informed decision.

Penetration of technology in our daily lives is substantial and advancements in technology and health informatics can be leveraged for better patient care. Mental health care delivery using digital platforms is affordable, available, and accessible to many [2]. Especially, the rapidly expanding numbers of mobile medical apps have the potential to transform the patient–health care provider relationship by improving the turnaround time and reducing costs [3]. In a study conducted in China, it was found that 96.7% (612/633) of participants used mobile devices, among which 38.4% (235/612) used multiple types of health apps [4]. The clinician rather than "prescribing apps" can engage the patient in "problem-solving" discussion, the problem here being "making a treatment choice." There is a vast sea of information available in the market and patients find it difficult to gather information and make informed choices. Clinicians also may lack time to explain a procedure fully and apps will help in the cause [5]. Apps developed by clinicians/hospitals taking into consideration user perspectives will filter and provide trustworthy information to the users.

In this regard, the ECT app was developed by the Leicestershire Partnership National Health Service (NHS) Trust and Leicestershire Health Informatics Service (LHIS) using the mDesign Development Platform to bring awareness about ECT for use by patients, carers, students, professionals, and the general public. It uses a multimedia approach, inclusive of video, text, and audio content, to show all aspects of the treatment.

Although the app was launched in November 2015, as of yet, there has been no formal feedback collated around the effectiveness of its content and accessibility. More recently, the mHealth App Usability Questionnaire (MAUQ) has been developed [6], but was not available during the study period. The aim of this survey is to evaluate and demonstrate the accessibility of the ECT app to the chosen audiences it was created for, via a paper and electronic questionnaire.

Methods

Survey

A survey was conducted between January 2017 and March 2019. The survey questions, which are provided in the Multimedia Appendix 1, were sent to mental health professionals, medical students, patients, carers, and members of the public via post, email, and SurveyMonkey or informed via posts shared in Psychiatry online groups and face-to-face contact. The survey questionnaire was developed for the purpose of this study and was not validated. The survey questionnaire was devised by the corresponding author (GK) in consultation with the ECT team and service users' representatives. The survey questionnaire did not collect personal/demographic details of the participants and hence provided anonymized data. Institute Research and Development opined that "...this proposal is more akin to a customer feedback or satisfaction survey commonly used in product development, and would not, in this form, require approval through R&D." Consent was sought to participate in the survey and the informed consent form is provided in Multimedia Appendix 2.

All participants who were willing to participate in the study were included. There were no exclusion criteria. The participants were chosen in the following manner:

- All service users (and their carers) who were prescribed ECT in the Leicestershire Partnership NHS Trust since July 1, 2016, and the service user/carer group of ECT Accreditation Service (ECTAS) from Royal College of Psychiatrists were contacted.
- All the trainees, consultants, and ward nurses in the Leicestershire Partnership NHS Trust who prescribe or are involved with ECT and the consultants and nurses of the ECTAS were contacted. Consultants and Trainees were also approached through the Deanery of the University of Leicester and the Royal College of Psychiatrists.
- The final-year medical students from the University of Leicester were also approached through the Deanery and the Royal College of Psychiatrists.

Results were collected via paper forms, email responses, and SurveyMonkey and all were inputted into SurveyMonkey to facilitate analysis. The analysis was done using SurveyMonkey [7]. The results are expressed as frequencies and the comments/suggestions are grouped into themes and are expressed verbatim.

Ethics Statement

Institute ethics committee approval was sought and it opined that the survey did not require approval from them.

Consent Statement

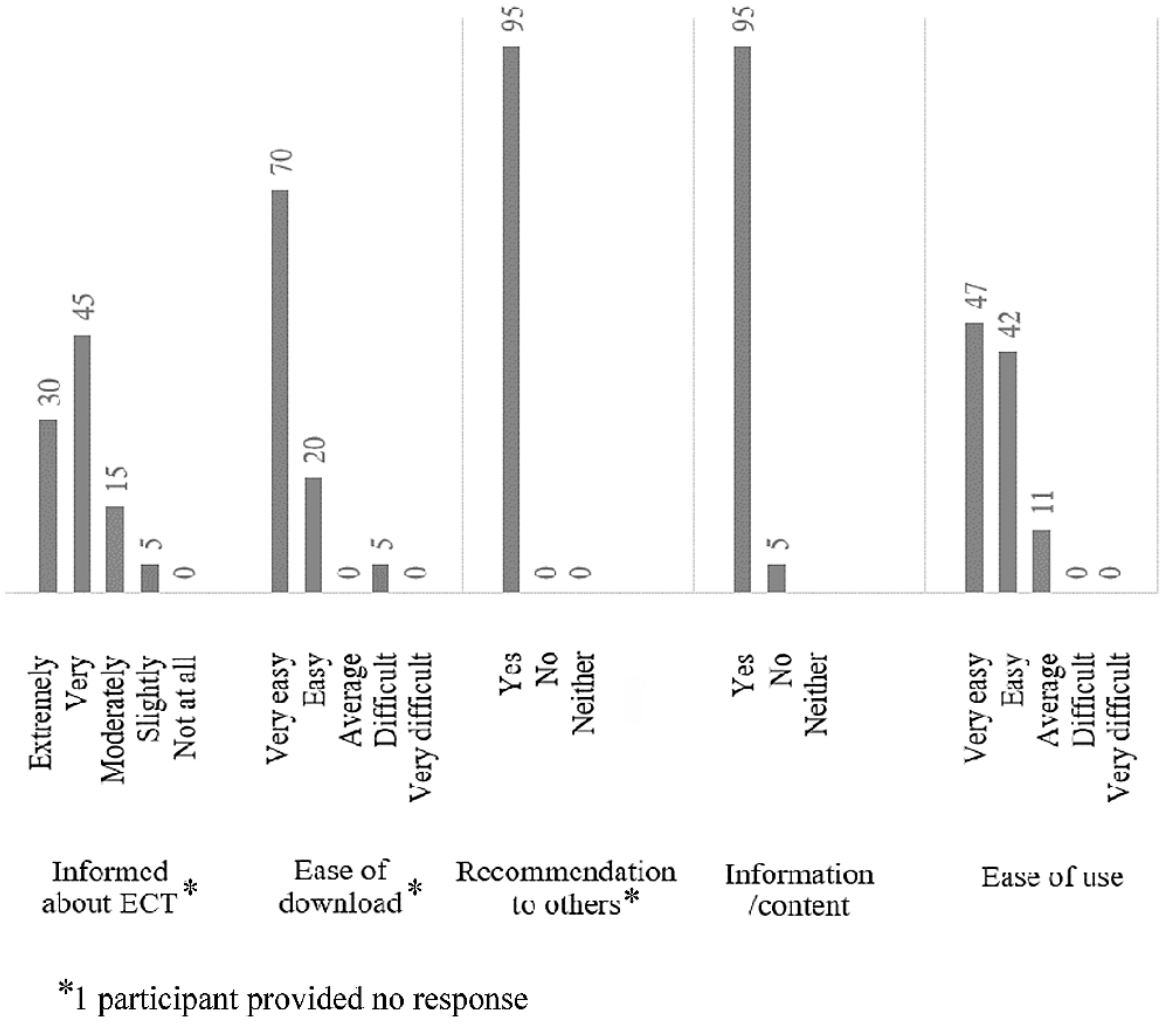
Consent was sought to participate in the survey and the informed consent form is provided in Multimedia Appendix 2.

Results

A total of 20 responses were received during the study period (January 2017 to March 2019). A majority of the participants (12/20, 60%) were professionals; 15% (3/20) were patients; 5% (1/20) were cares; and 20% (4/20) were others, including a member of the public, a friend of a patient, a past patient, and a medical student. The details are provided in Multimedia Appendix 3.

The responses to the questionnaire are provided in Figure 1. Almost everyone (19/20, 95%) (1 person skipped the question) answered that they could find all the information they were expecting to see in the app. Three-fourth of the respondents (15/20, 75%) felt more informed about ECT after the use of the app (Extremely [6/20, 30%] and very [9/20, 45%]). A vast majority felt that the app was very easy to download (14/20, 70%) and use (9/20, 45%). Almost all (19/20, 95%; 1 person did not answer the question) said that they would recommend the app to other patients/professionals.

Figure 1. Survey responses.



Many (10/20, 50%) participants also provided suggestions and comments. Among these, 50% (5/10) commented on information provision, 10% (1/10) complained of network and connectivity issues, 20% (2/10) gave compliments, and 20% (2/10) provided miscellaneous comments to make the app free and that they had no computer or smartphone. The details are provided in Multimedia Appendix 4.

Discussion

The participants of the survey, which included a mixed group of professionals, patients, and carers, opined that the app was easy to download and use, contained adequate information and they felt more informed after having used the app, and they would recommend it to others. The participants of the survey also provided suggestions on the app.

It is pertinent to know that the decision-making process to use ECT involves provision of adequate, appropriate, and accurate information on the method, benefits, and risks of ECT [1]. This discussion is particularly important as ECT is usually shown in a bad light in popular culture and general public have formed myths around ECT. Apps may be particularly beneficial in this regard as they have the advantages of being readily available and mostly free of cost to all users of electronic devices on which apps can be downloaded (eg, mobile phones, tablets, laptops). In addition, in an app, a large amount of information and resources can be condensed into a small space, and can include a wealth of learning material accessible in a number of ways, including audio and video (as opposed to physical material, which can be bulky and one dimensional). Numerous apps are now available to assist health care providers with many important tasks, such as information and time management, health record maintenance and access, communications and consulting, reference and information gathering, patient management and monitoring, clinical decision making, and medical education and training [8]. The ECT app is specially designed to provide information and practical procedure of ECT to patients and carers to make informed choices and to relieve anxiety and myths related to ECT. In addition, it provides detailed information on ECT to both mental health and nonmental health professions to broaden their knowledge and support their patients when appropriate. Patients still rely on their general practitioners and other health professions to help them in making decisions in psychiatry. The app was developed for the use of professionals and public (not for the purpose of this study) by the hospital with inputs from professionals and service users as is also suggested by Chiauzzi and Newell [5]. The app is free to download on the Apple Store online [9].

In this survey, the response rate was low, despite multiple methods of distribution of the survey and reminders being sent to individuals. The exact number of surveys sent cannot be specified due to the use of online forums in which it is unclear how many have seen the link. Not all responders had access to electronic devices which utilize apps. Leaflets would be an alternative option for this population, although not all parts of the app (eg, videos) would be translatable into paper format. Although text is available on the app for all the videos, subtitles for the videos may aid users with hearing disabilities to use the app and may improve accessibility. Not all respondents experienced technical issues; therefore, this is likely to be as a result of the local internet connection availability, rather than a feature of the app itself. Overall, the feedback was positive and encouraging.

The ECT app can be beneficial in sharing appropriate information to professionals and the public alike and help in gathering unbiased and nonjudgmental information on the current use of ECT as a treatment option. The app is specially designed to provide information and practical procedure of ECT to patients and carers to make informed choices and to relieve anxiety and myths related to ECT. In addition, it provides detailed information on ECT to both mental health and nonmental health professions to broaden their knowledge and support their patients when appropriate. Patients still rely on their general practitioners and other health professions to help them in making decisions in psychiatry.

We aim to update the app regularly; to ensure that it is in line with the most recent guidance and advances in ECT practice. In the future, mental health professionals who provide information on ECT can suggest the patients and their carers to go through the app and later have a discussion on their decision. A feedback about the app can be taken from both mental health professional and patients/carers at that point of time. A validated feedback form can also be incorporated in the app. Future studies can aim at using validated survey questionnaires such as MAUQ [6] on targeted population of ECT users. Measuring saved clinician time can help in better utilization of human resources.

Acknowledgments

We thank Dr Jagadisha Thirthalli for critically reviewing the manuscript. The app was developed free by Leicestershire Health Information Services.

Authors' Contributions

All the authors have contributed to the development of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The user survey questionnaire.
[PDF File (Adobe PDF File), 33 KB - biomedeng_v5i1e20730_app1.pdf]

Multimedia Appendix 2
Informed Consent Form.
[PDF File (Adobe PDF File), 55 KB - biomedeng_v5i1e20730_app2.pdf]

Multimedia Appendix 3
Type of Participants.
[PDF File (Adobe PDF File), 29 KB - biomedeng_v5i1e20730_app3.pdf]

Multimedia Appendix 4
Comments and suggestions by the participants.
[PDF File (Adobe PDF File), 56 KB - biomedeng_v5i1e20730_app4.pdf]

References

1. National Institute for Health and Care Excellence. Guidance on the Use of Electroconvulsive Therapy Technology Appraisal Guidance [TA59]. London: National Institute for Health and Care Excellence; 2003 Apr. URL: https://www.nice.org.uk/guidance/ta59/chapter/1-Guidance [accessed 2019-07-13]
2. Basavarajappa C, Chand PK. Digital Platforms for Mental Health-care Delivery. Indian J Psychol Med 2017 Sep;39(5):703-706 [FREE Full text] [doi: 10.4103/IJPSYM.IJPSYM_209_17] [Medline: 29200576]
3. Yetisen AK, Martinez-Hurtado JL, da Cruz Vasconcellos F, Simsekler MCE, Akram MS, Lowe CR. The regulation of mobile medical applications. Lab Chip 2014 Mar 07;14(5):833-840 [FREE Full text] [doi: 10.1039/c3lc51235e] [Medline: 24425070]

4. Xie Z, Nacioglu A, Or C. Prevalence, Demographic Correlates, and Perceived Impacts of Mobile Health App Use Amongst Chinese Adults: Cross-Sectional Survey Study. JMIR Mhealth Uhealth 2018 Apr 26;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]
5. Chiauzzi E, Newell A. Mental Health Apps in Psychiatric Treatment: A Patient Perspective on Real World Technology Usage. JMIR Ment Health 2019 Apr 22;6(4):e12292 [FREE Full text] [doi: 10.2196/12292] [Medline: 31008711]
6. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): Development and Validation Study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
7. SurveyMonkey [Computer Program]. San Mateo, CA: SVMK Inc; 1999. URL: www.surveymonkey.com [accessed 2020-09-13]
8. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
9. Electroconvulsive Therapy (ECT) [computer program], Version 1. Leicester: Leicestershire Partnership NHS Trust; 2015. URL: https://apps.apple.com/us/app/electroconvulsive-therapy-ect/id1072836545

Edited by G Eysenbach; submitted 27.05.20; peer-reviewed by T Alessa, AS Aslam; comments to author 11.07.20; revised version received 19.08.20; accepted 22.08.20; published 22.09.20.

Please cite as:
Khan K, Basavarajappa C, Kunigiri G
Ease of Use of the Electroconvulsive Therapy App by Its Users: Cross-Sectional Questionnaire Study
JMIR Biomed Eng 2020;5(1):e20730
URL: http://biomedeng.jmir.org/2020/1/e20730/
doi:10.2196/20730
PMID:

Original Paper

Measuring Heart Rate Variability in Free-Living Conditions Using Consumer-Grade Photoplethysmography: Validation Study

Emily Lam[1*], BEng; Shahrose Aratia[2*], BSc; Julian Wang[3], MD; James Tung[4], PhD

[1]Possibility Engineering and Research Laboratory, Bloorview Research Institute, Toronto, ON, Canada

[2]Department of Occupational Science and Occupational Therapy, University of Toronto, Toronto, ON, Canada

[3]Michael G Degroote School of Medicine, McMaster University, Hamilton, ON, Canada

[4]Department of Mechanical & Mechatronics Engineering, University of Waterloo, Waterloo, ON, Canada

[*]these authors contributed equally

Corresponding Author:
James Tung, PhD
Department of Mechanical & Mechatronics Engineering
University of Waterloo
200 University Ave W
Waterloo, ON, N2L 3G1
Canada
Phone: 1 519 888 4567 ext 43445
Email: james.tung@uwaterloo.ca

Abstract

Background: Heart rate variability (HRV) is used to assess cardiac health and autonomic nervous system capabilities. With the growing popularity of commercially available wearable technologies, the opportunity to unobtrusively measure HRV via photoplethysmography (PPG) is an attractive alternative to electrocardiogram (ECG), which serves as the gold standard. PPG measures blood flow within the vasculature using color intensity. However, PPG does not directly measure HRV; it measures pulse rate variability (PRV). Previous studies comparing consumer-grade PRV with HRV have demonstrated mixed results in short durations of activity under controlled conditions. Further research is required to determine the efficacy of PRV to estimate HRV under free-living conditions.

Objective: This study aims to compare PRV estimates obtained from a consumer-grade PPG sensor with HRV measurements from a portable ECG during unsupervised free-living conditions, including sleep, and examine factors influencing estimation, including measurement conditions and simple editing methods to limit motion artifacts.

Methods: A total of 10 healthy adults were recruited. Data from a Microsoft Band 2 and a Shimmer3 ECG unit were recorded simultaneously using a smartphone. Participants wore the devices for >90 min during typical day-to-day activities and while sleeping. After filtering, ECG data were processed using a combination of discrete wavelet transforms and peak-finding methods to identify R-R intervals. P-P intervals were edited for deletion using methods based on outlier detection and by removing sections affected by motion artifacts. Common HRV metrics were compared, including mean N-N, SD of N-N intervals, percentage of subsequent differences >50 ms (pNN50), root mean square of successive differences, low-frequency power (LF), and high-frequency power. Validity was assessed using root mean square error (RMSE) and Pearson correlation coefficient (R^2).

Results: Data sets for 10 days and 9 corresponding nights were acquired. The mean RMSE was 182 ms (SD 48) during the day and 158 ms (SD 67) at night. R^2 ranged from 0.00 to 0.66, with 2 of 19 (2 nights) trials considered moderate, 7 of 19 (2 days, 5 nights) fair, and 10 of 19 (8 days, 2 nights) poor. Deleting sections thought to be affected by motion artifacts had a minimal impact on the accuracy of PRV measures. Significant HRV and PRV differences were found for LF during the day and R-R, SDNN, pNN50, and LF at night. For 8 of the 9 matched day and night data sets, R^2 values were higher at night (P=.08). P-P intervals were less sensitive to rapid R-R interval changes.

Conclusions: Owing to overall poor concurrent validity and inconsistency among participant data, PRV was found to be a poor surrogate for HRV under free-living conditions. These findings suggest that free-living HRV measurements would benefit from examining alternate sensing methods, such as multiwavelength PPG and wearable ECG.

(JMIR Biomed Eng 2020;5(1):e17355) doi:10.2196/17355

KEYWORDS

heart rate determination photoplethysmography; wearable electronic sensors; physiological monitoring; ambulatory monitoring; mobile phone

Introduction

Motivation

With the growing ubiquity of commercially available wearable technologies, obtaining long-term physiological measurements under free-living conditions is feasible and permits longitudinal examination of ecologically valid patterns. This presents an opportunity for continuous patient monitoring under free-living conditions, including the potential to identify at-risk individuals (eg, patients with cardiac disease). Heart rate variability (HRV) is a well-established, powerful metric used to assess cardiac health, including autonomic nervous system function regulating cardiac activity. Compared with an individual's heart rate (HR) averaged over a short period, HRV measures variations in HR primarily as an indicator of the efforts of the sympathetic and parasympathetic nervous systems to achieve an optimal cardiac response under constantly changing stimuli [1]. Previous research has explored the use of HRV monitoring in predicting or detecting sleep quality [2], mental stress [3], chronic pain [4], posttraumatic stress disorder [5], bipolar disorder [5], and cardiac health [6].

Measuring HRV

The (gold) criterion standard for measuring HRV is through an electrocardiogram (ECG) to obtain a direct recording of cardiac electrical activity. On ECG, the R wave represents the maximum upward deflection of a normal QRS complex. The duration between two successive R waves defines the R-R interval [7], which is used to measure HR and HRV. Although wearable ECGs exist, they typically require electrodes affixed to the skin, which makes them obtrusive and can cause skin breakdown, and they are also prone to motion artifacts during day-to-day activities [8]. Alternatively, photoplethysmography (PPG) uses an optical sensor widely used to unobtrusively track mean HR, especially in wrist-worn devices (eg, Fitbit).

PPG for Pulse Rate Variability

PPG sensors measure changes in pulsatile blood flow within an individual's vasculature using color intensity signals [9]. Signal peaks associated with the flow of blood are used as indicators of HR, allowing for the calculation of peak-to-peak (P-P) intervals. PPG sensors do not directly measure HRV; instead, they measure pulse rate variability (PRV), the change in vessel pulse periods, from which P-P intervals denote a pulse rate (PR) [10]. PPG sensors can be placed at a variety of measurement sites including the fingers, wrist, brachia, ear, forehead, and esophagus without requiring additional equipment. This makes PPG especially convenient for pervasive cardiac monitoring [11], with well-validated use for mean HR measurements [4]. Although evidence examining PPG capabilities to accurately measure HRV shows promise, studies comparing PPG with gold standard ECG methods under free-living conditions remain limited.

The accuracy of PRV as a measure of HRV has been investigated with clinical devices under controlled, and often stationary, conditions [12-17]. Although these studies indicate that PRV may be a useful as a proxy measure of HRV using medical-grade devices under controlled conditions, studies using wearable consumer-facing devices have shown mixed results. These few studies largely use short-term collections in controlled circumstances, some of which do not simultaneously collect ECG [18,19]. A systematic review by Georgiou et al [20] found that wearable devices can provide accurate measurement of HRV measures at rest; however, accuracy declines as exercise and motion levels increase. The review also showed that heterogeneity in sensor position, detection algorithm, experimental settings, and analysis methods from existing studies limits the evidence. A review by Shäefer and Vagedes [21] found similar results, suggesting that physical activity and mental stressors lead to unacceptable deviations between PRV and HRV. Ultimately, further research is required to determine the efficacy of PRV in estimating HRV during free-living conditions in which individuals are unrestricted and engaging in their daily activities [22].

Limitations of PPG

PPG sensors have been found to be sensitive to motion artifacts, changes in blood flow caused by movement, compression and deformation of the vasculature arising from pressure disturbances at the interface between the sensor and the skin [11], and light leaking between the sensor and the skin [23]. Some studies have examined the removal of motion artifacts from PPG signals using signal processing techniques and acceleration as a reference [23-28]. For example, methods involving accelerometry have shown promise for improving coherence by editing signals likely influenced by motion artifacts [28,29]. Baek and Shin [30] collected PPG measurements over 24 hours using a custom device and filtering method, recommending a subset of HRV metrics as good targets for continuous HRV tracking using commercial devices. Morelli et al [28] conducted a study evaluating the accuracy of a consumer-grade PPG (Microsoft Band 2) for HRV estimation during less restrictive, but controlled, conditions (eg, sitting and walking) over 10-min trials. Errors likely caused by motion artifacts during walking were attenuated by using corresponding accelerometer signals to delete sections of the data corrupted by motion artifacts.

Objectives

Although HR and PR are correlated and closely related, the use of PRV to estimate HRV requires further research, especially under free-living conditions. In this study, the concurrent validity of PRV measurements from a consumer-facing PPG sensor is compared with HRV measurements from a portable ECG under 2 unsupervised conditions up to 4.5 hours each: (1) while engaging in regular activities of daily living and (2) during sleep. A secondary goal of this study is to examine factors

influencing estimation errors of PRV for HRV, including motion artifacts, measurement conditions, and editing approaches.

Methods

Participants

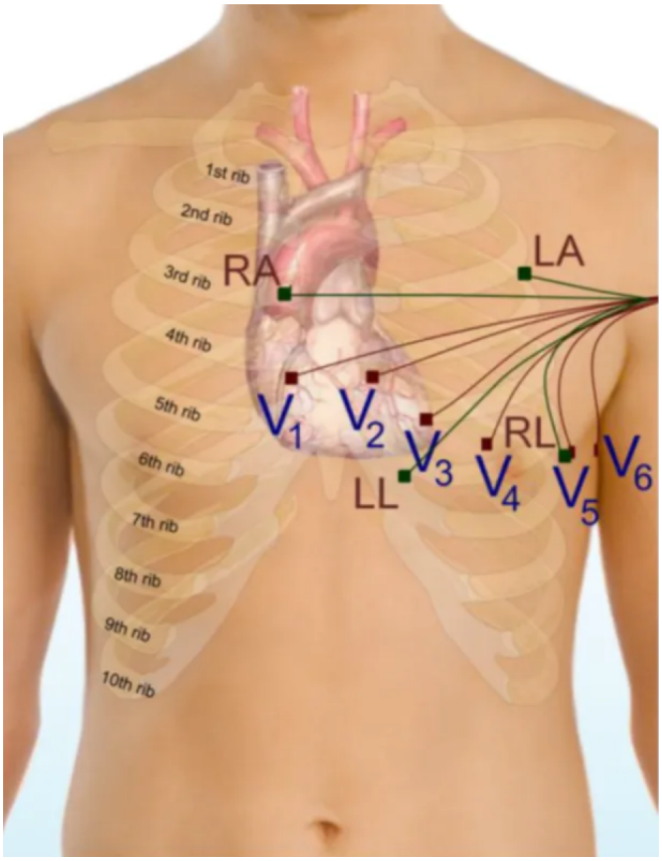
A convenience sample of healthy individuals aged 18-65 years was recruited for the study. Individuals with a history of cardiac and/or sleep disorders were excluded to minimize the collection of irregular cardiac signals. Under these conditions, approval for this study was granted by the University of Waterloo Research Ethics Committee on September 5, 2017, filed under protocol #31197.

Device Setup

A total of 2 wearable devices were used to acquire cardiovascular signals in this study: (1) a commercially available optical PPG wearable device (Microsoft Band 2 or MB2, Microsoft) and (2) a research-grade wearable ECG device (Shimmer3 ECG, Shimmer). Both wearables were recorded simultaneously with signals transmitted via Bluetooth to a smartphone (Pixel or Nexus 3, Google). To synchronize the devices, triaxial accelerations were also recorded with both devices. Participants were asked to wear the devices twice, for at least 90 min each, once during daily activities and a second time when sleeping.

To record ECG, hydrogel electrodes (Kendall 233 Hydrogel, Covidien) were placed in a 4-lead bipolar limb lead configuration (ie, left arm [LA], right arm [RA], left leg [LL], right leg) on the participant's chest as shown in Figure 1 [31]. The participant's skin was prepared by shaving and sanitized using hospital-grade alcohol wipes before electrode placement. Electrodes were connected to a Shimmer3 ECG, worn at the waist with a strap, and all leads were taped to the chest to prevent tangling and static, and minimize motion artifacts. On the smartphone, the Multi-Shimmer Sync mobile app (Shimmer, Dublin, Ireland) was used to record ECG data from the Shimmer3 ECG. The MB2 was worn on the participant's wrist of choice as tightly as possible, without causing discomfort. MB2 size (small or medium) was selected to fit the size of the participant's wrist. A third-party mobile app, Companion for Microsoft Band (released by Pain in My Processor, Google Play Store), was used to log data from the MB2 to the smartphone.

Figure 1. Electrocardiogram 4-lead bipolar limb electrocardiogram configuration on participants' chests. LA: left arm; RA: right arm; V: precordial leads.



Participants' Instructions

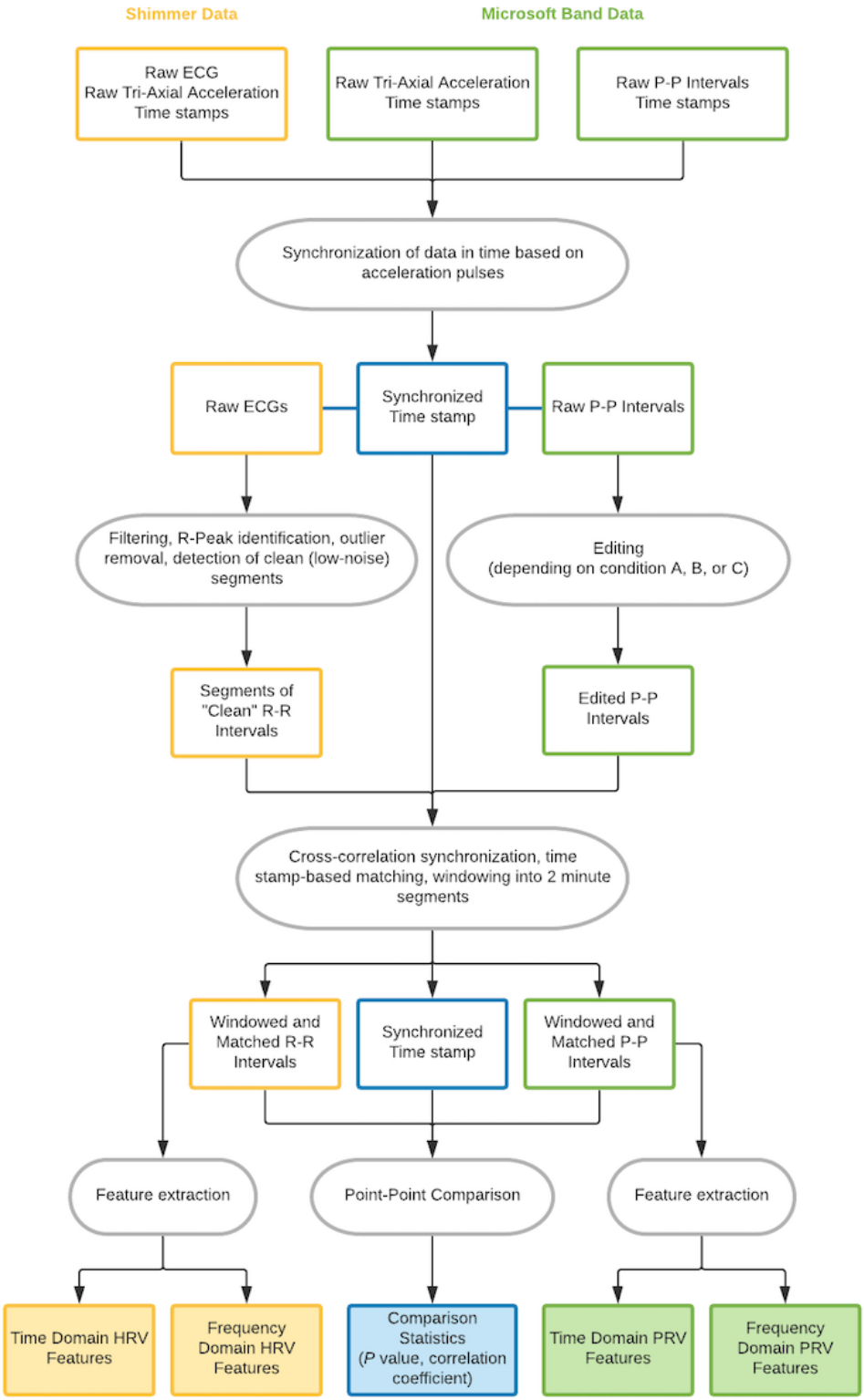
Given the free-living nature of data collection, participants were instructed on how to set up and monitor device connection and logging status to facilitate troubleshooting. To ensure proper electrode placement, a (trained) researcher placed the electrodes in the 4-lead bipolar limb lead configuration (Figure 1) during the first data collection (ie, day). For the second collection (ie, night), electrodes were left on, replaced by the research assistant, and/or marked by location and replaced by the participant. Before the second collection, ECG signals were visually examined to ensure that the QRS complexes were clearly identifiable. Participants were instructed and encouraged to

contact a researcher at any time in case of questions or concerns during data collection.

Postprocessing

Following data collection, all postprocessing and statistical analyses were conducted using MATLAB 2018a (MathWorks). Figure 2 outlines the steps taken in postprocessing.

Figure 2. Postprocessing of data from the Shimmer and Microsoft Band. ECG: electrocardiogram; HRV: heart rate variability; P-P: time between 2 P peaks in a photoplethysmogram or peak-to-peak intervals; PRV: pulse rate variability; R-R: time between 2 R peaks in an ECG.



Synchronizing Devices

Shimmer3 and MB2 were coarsely synchronized by aligning triaxial acceleration peaks from tapping both devices simultaneously on a table. Each device was tapped 3 times in 2 orientations with 10 s of rest between orientations. Fine

synchronization was performed using a cross-correlation method described below (cross-correlation synchronization).

ECG Data Processing

Both LA-RA and LL-RA ECG signals were filtered using a first order bandpass Butterworth filter from 1 to 25 Hz. A maximal overlap discrete wavelet transform with a Daubechies least-asymmetric wavelet with 4 vanishing movements was used to enhance the R peaks in the ECG, followed by a threshold-based peak-finding function used to identify the R-peaks [32,33]. In one sample (participant 2, daytime), the wavelet detection algorithm more accurately and consistently detected the T wave of the ECG signal and was used as a proxy for the QRS complex, previously shown to give results similar to those of R-peak detection [34]. A time series of R-R intervals was extracted from the detected R-peaks, and outlier values outside of the physiological range of values for a healthy individual at rest, walking, or during sleep (R-R<0.3 s or R-R>2.5 s) were removed [35-37]. To remove transients associated with artifacts or noise, segments of at least 15 consecutive R-R intervals were included in the analysis. Longer segment thresholds (30, 60, and 120 consecutive R-R intervals) were tested with negligible effects on the results. On the basis of the signal that provided more R-R intervals, either the LA-RA or LL-RA electrode pair signal was chosen for processing and analysis.

PPG Data Processing

P-P intervals and corresponding time stamps were recorded directly from MB2 outputs as the time interval between 2 continuous heartbeats [38]. Note that the temporal resolution of MB2 is limited to 10 ms. On the basis of existing literature reporting signal processing methods to edit R-R intervals and remove artifacts, three methods were used to identify and delete artifacts in the P-P intervals outputted by the MB2, resulting in 4 conditions of P-P data. Deletion was chosen as the editing technique (as opposed to interpolation) because motion artifacts would likely affect consecutive samples, making interpolation challenging. In addition, the long-term nature of data collection would mitigate one of the major concerns associated with deletion, the loss of samples [39]. The 4 processing conditions were as follows:

- None (condition A): This condition contains the raw P-P intervals.
- Threshold deletion (condition B): Removing implausible P-P interval values for a healthy individual at rest, walking, or sleeping (P-P<0.3 s or P-P>2.5 s) [35-37].
- Moving average deletion (condition C): Threshold deletion (as described in B above) and removing changes in P-P intervals faster than physiologically plausible indicated by a moving average filter. This was done following Morelli et al [28], discarding values for which $|PP_t-\mu_{10}|\geq 0.5\mu_{10}$, where PP_t refers to the P-P interval data and μ_{10} is a 10 s moving average.
- Acceleration-based deletion (condition D): A series of threshold filters, moving average filters (described in C above), and an acceleration filter. Considering that low PPG signal quality may be attributable to movement, Morelli et al [28] removed signal segments affected by motion artifacts by estimating periods of signal quality associated with the corresponding accelerometry time series, W_t, and then removing P-P intervals where W_t was found to exceed a threshold, k. k was identified by examining the correlation between W_t and error, where $W_t$is calculated as an average of w_t over a window of duration τ, and W_t is calculated as follows:

In this study, no significant correlation between W_t and was found. As such, a threshold of =0.02 m/s^2 was used to filter the data with τ=40 s (the same parameters as used by Morelli et al) [28].

Data Synchronization

Following coarse synchronization of MB2 and Shimmer3, consistent delays between the 2 devices were observed. To identify the highest correlation between devices, a cross-correlation between P-P and R-R data was conducted. The estimate of the time-shift was applied to the P-P data, similar to the method used by Pietilä et al [40]. P-P intervals were then matched to R-R intervals by matching data points with the closest time stamps. If a data point did not have a matching interval within 1 s, the interval was deleted. The 1-s delay was chosen to accommodate for delays in Bluetooth transmission and pulse transit time. After matching, the remaining data were divided into 2-min windows from which the HRV and concurrent validity metrics were calculated [41].

HRV Metrics for Analysis

After postprocessing, the following time domain HRV and PRV features were extracted for each trial, where N-N refers to either R-R or P-P:

- Mean N-N: the mean of all N-N intervals
- Mean HR: reciprocal of mean N-N, in beats per minute (bpm)
- SDNN: a measure of overall variability, the SD of all N-N intervals, also known as RRSD
- pNN50: percentage of subsequent differences more than 50 ms
- RMSSD: root mean square of subsequent differences
- LF, HF, LF/HF ratio: low-frequency power (LF), high-frequency power (HF), and the ratio of LF to HF
- SD1 and SD2: SDs of short (x=y) and long (orthogonal to x=y) diagonal Poincaré plot axes [12]

For spectral measures, R-R and P-P intervals were converted to instantaneous HR (60/N-N, where N-N is interval time in seconds) and then interpolated to 4 Hz using a piecewise cubic Hermite interpolation (MATLAB function “pchip”). This ensured regular time intervals between data points, a prerequisite for estimating the Fourier transform and signal power. The Fourier transform was performed (using “fft” function in MATLAB) on the entire data set for each participant. This allowed for the calculation of frequency domain HRV features such as LF (0.04-0.15 Hz) and HF (0.15-0.40 Hz). LF and HF were computed in normalized units by the sum of LF and HF. The ratio of LF to HF was also reported.

Analyses

To quantify the concurrent validity between R-R and P-P intervals, the following metrics were used:

- Root mean square error (RMSE): RMSE between matched R-R and P-P samples
- Pearson correlation coefficient (R^2): The correlation strength between R-R and P-P intervals. R^2 values were categorized as strong ($R^2 \geq 0.7$), moderate ($0.5 \leq R^2 < 0.7$), fair ($0.3 \leq R^2 < 0.5$), and poor ($0.3 < R^2$) [42].

To compare PPG-derived metrics across collection and processing conditions (ie, day- or nighttime collection, filtering condition), two-tailed paired *t* tests were used. Bland-Altman plots were generated to illustrate the agreement between R-R and P-P intervals. In the Bland-Altman plot, the difference between each P-P and R-R measurement is plotted against the mean of each measurement [43].

Results

Overview

This section presents the results of (1) investigating the concurrent validity between R-R and P-P intervals across published filtering methods, (2) a comparison between ECG- and PPG-derived metrics of HRV, and (3) a comparison across free-living data collection conditions (ie, day and night). A total of 10 volunteers were recruited (3 men and 7 women, aged 20-61 years) for this study for a total of 19 trials (1 day and 1 night per participant). One participant's ECG night data were corrupted and therefore not analyzed or reported.

After processing, a large amount of data was lost. The number of matched and windowed N-N intervals is described in Table 1; all comparison statistics were calculated on the basis of these data. The percentages of compared intervals were calculated by dividing the number of matched and windowed samples by the total number of R-R or P-P intervals detected from the ECG or MB2, respectively.

Table 1. Group mean (SD) of data sample sizes used for comparison between R-R and P-P intervals across processing and collection conditions.

Collection condition and processing condition	Number of samples, mean (SD)	Percent R-R intervals compared, mean (SD)	Percent P-P intervals compared, mean (SD)
Day			
A	5168.70 (1683.92)	52.35 (26.89)	47.29 (18.32)
B	4706.5 (1447.57)	48.25 (26.30)	43.91 (18.31)
C	3311.30 (1316.13)	34.68 (24.73)	32.29 (16.88)
D	1847.40 (1334.28)	23.03 (26.28)	21.39 (15.92)
Night			
A	8418.78 (5179.41)	55.05 (27.70)	53.30 (19.26)
B	8197.11 (5060.21)	53.79 (27.31)	52.06 (19.15)
C	7383.00 (4075.76)	46.89 (24.24)	46.66 (19.38)
D	7177.33 (4901.63)	41.15 (27.67)	42.97 (21.23)

A larger data sample was acquired at night than that acquired during the day. Despite formal instructions and training on the operation and charging of the sensor systems, several technical barriers were frequently encountered that limited the number of samples in each trial. These included inadvertent misplacement of ECG electrodes or MB2, insufficient battery charging before night collection, and/or dropped Bluetooth stream to the mobile device.

Concurrent Validity Across the Editing Techniques

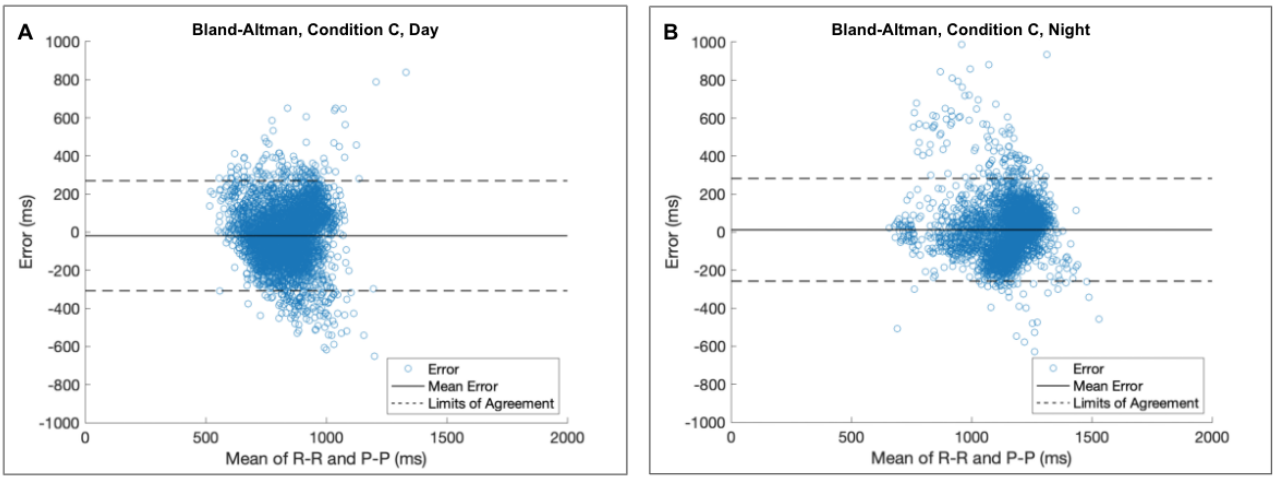
Table 2 compares the concurrent validity of P-P data with that of the R-R data across all processing conditions, including RMSE and R^2. The largest differences were observed in the RMSE between the raw (A) and filtered (B, C, and D) conditions.

Table 2. Group mean root mean square error, concurrent validity (R^2), and number of matched samples across processing and collection conditions.

Processing condition	Day (n=10)	Night (n=9)
Root mean square error (ms), mean (SD)		
A	182 (48)	158 (67)
B	165 (42)	136 (53)
C	144 (39)	120 (45)
D	122 (47)	119 (45)
***R*^{2} , mean (SD)**		
A	0.15 (0.12)	0.28 (0.17)
B	0.14 (0.13)	0.33 (0.19)
C	0.18 (0.13)	0.34 (0.21)
D	0.22 (0.17)	0.34 (0.21)

The RMSE ranged between 46 and 285 ms across all conditions. Increased editing reduced the average error (RMSE). Under condition C, error was further examined by generating Bland-Altman plots comparing the P-P intervals with R-R intervals, as shown in Figure 3. Although the mean error is close to zero for both day and night conditions, the limits of agreement were greater than 200 ms.

Figure 3. Bland-Altman plots for 1 participant under processing condition C for (A) day and (B) night. P-P: time between 2 P peaks in a photoplethysmogram or peak-to-peak intervals; R: time between 2 R peaks in an electrocardiogram.

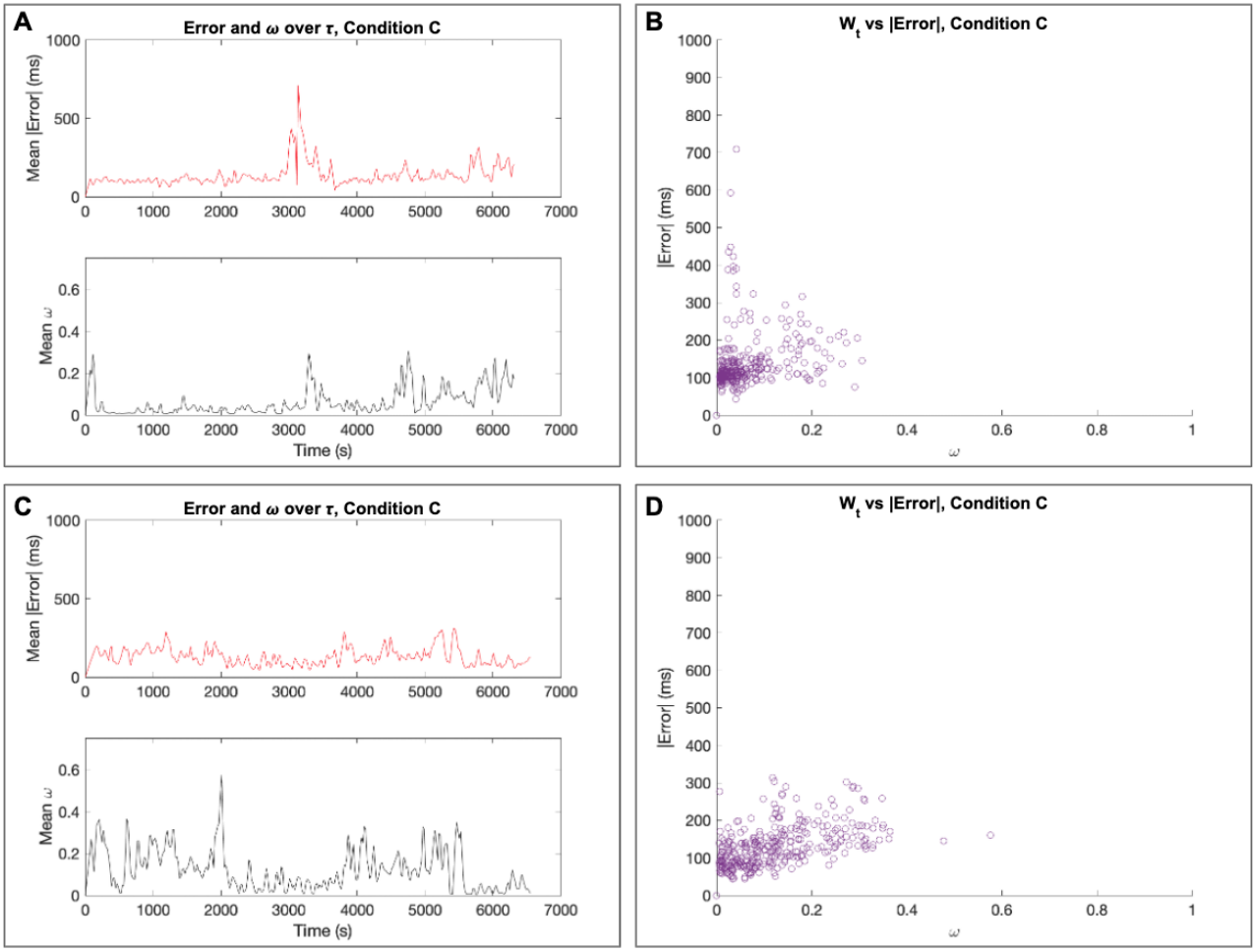


Across all conditions, R^2 values ranged from 0 to 0.66. Editing did not have a large impact on R^2. Although R^2 improved at night, none of the correlations were considered strong; 2 of 19 (all night) were moderate, 7 (2 days, 5 nights) were fair, and 10 (8 days, 2 nights) were poor. Of the 19, 16 (9 days, 7 nights) paired *t* tests between R-R and P-P intervals under condition C yielded *P*=.01, indicating significant differences between ECG- and PPG-based methods.

Under condition D, no data sets showed strong correlations. Only 3 (1 day, 2 nights) were moderate, 7 were fair (1 day, 6 nights), and 9 were poor (7 days, 2 nights). Paired *t* tests between matched R-R and P-P intervals edited under condition D were significant for 12 trials (5 days, 7 nights). Notably, condition D reduced the amount of data available for analysis, especially during the day. From condition C to D, the average sample loss was 40.18% (SD 29.59) during the day and 3.73% (SD 4.37) at night.

Compared with condition C, condition D improved RMSE and R^2 slightly during the day and varied by trial. The mean correlation between error and W_t was 0.28 (SD 0.24), with a range of 0.13 to 0.70 for day data, and 0.29 (0.21), with a range of 0.16 to 0.73 for night data. Figure 4 [44] shows the error and W_t for a sample showing lower correlation between W_t and error (R^2=0.16) and a sample trial with higher observed correlation (R^2=0.50).

Figure 4. Correlation between absolute error and mean change in triaxial acceleration (W_t) under the same conditions as Morelli et al (A) and (B) comparison of |Error| and W_t over time for a sample with low correlation (R^2=0.16) (C), and (D) comparison of |Error| and W_t over time for a sample with higher correlation (R^2=0.50).



Comparison of HRV and PRV Measures

Table 3 compares the HRV and PRV measures across participants under condition C, as this condition yielded the highest concurrent validity for most participants while retaining sample size. The findings in Table 3 are based on a 3311.30 (SD 1316.13) matched samples for day data and 7303.00 (SD 4075.76) for night data. Under condition C, paired *t* tests revealed no significant differences between HRV and PRV measures. At night, SDNN, pNN50, RMSSD, SD1, SD2, LF, HF, and LF/HF ratio metrics were observed to be significantly different. Significant differences between HRV and PRV measures were observed in more measures at night, a condition during which motion artifacts are expected to be lower, allowing for collection of more accurate PRV data. Note that the temporal resolution of MB2 is limited to 10 ms, but many of the observed differences between R-R and P-P intervals are larger.

Table 3. Comparison of mean heart rate variability and pulse rate variability metrics under processing condition C.

Features	Day				Night			
	HRV[a]	PRV[b]	|Error|	*P* value[c]	HRV	PRV	|Error|	*P* value[c]
Time domain features								
NN (ms), mean (SD)	829 (70)	833 (51)	19 (15)	.59	967 (151)	960 (142)	10 (9)	.08
SDNN[d] (ms), mean (SD)	90 (36)	98 (25)	25 (20)	.48	87 (37)	69 (25)	20 (10)	.03
pNN50[e] (%), mean (SD)	30.60 (24.51)	39.74 (16.18)	15.90 (11.30)	.14	38.58 (30.59)	21.35 (15.06)	19.48 (14.11)	.02
RMSSD[f] (ms), mean (SD)	104 (58)	116 (38)	42 (36)	.54	101 (57)	67 (20)	34 (36)	.02
SD1[g] (ms), mean (SD)	74 (41)	82 (27)	30 (26)	.54	72 (40)	48 (21)	24 (26)	.02
SD2[h] (ms), mean (SD)	94 (40)	110 (25)	31 (33)	.24	97 (35)	83 (29)	18 (14)	.05
Frequency domain features								
LF[i] (nu), mean (SD)	0.70 (0.03)	0.69 (0.02)	0.03 (0.02)	.43	0.70 (0.03)	0.72 (0.01)	0.02 (0.02)	.02
HF[j] (nu), mean (SD)	0.30 (0.03)	0.31 (0.02)	0.03 (0.02)	.43	0.30 (0.03)	0.28 (0.01)	0.02 (0.02)	.02
LF/HF ratio, mean (SD)	2.39 (0.32)	2.26 (0.22)	0.28 (0.23)	.29	2.43 (0.28)	2.64 (0.13)	0.21 (0.18)	.01

[a]HRV: heart rate variability.

[b]PRV: pulse rate variability.

[c]Results from paired *t* test between HRV and PRV measures.

[d]SDNN: SD of all N-N intervals.

[e]pNN50: percent of subsequent differences more than 50 ms.

[f]RMSSD: root mean square of subsequent differences.

[g]SD1: SD of short (x=y) Poincaré plot axis.

[h]SD2: SD of long (orthogonal to x=y) Poincaré plot axis.

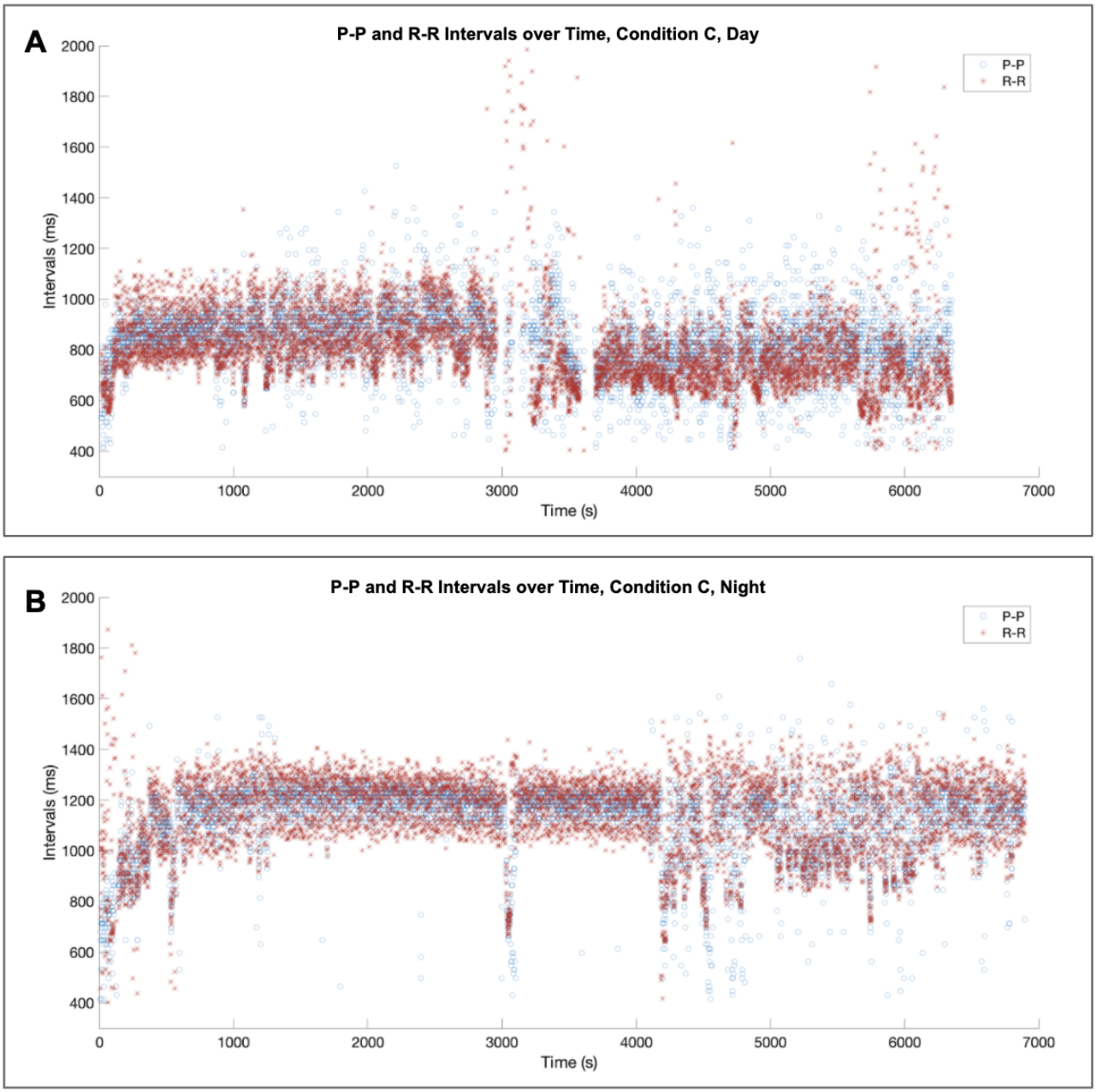
[i]LF: low-frequency power.

[j]HF: high-frequency power.

Compared with processing condition C, similar results were observed in condition D (Multimedia Appendix 1). Under condition D, paired *t* tests revealed significant differences between HRV and PRV measures for no measures during the day, but there were significant differences in R-R and pNN50 at night. Although this may be attributed to condition D using motion artifact editing, the large number of samples edited from condition C to D may partially explain these findings. Given the large sample loss associated with condition D and a lack of strong correlation between W_t and error, the remainder of this study focuses on the results from processing condition C (over D).

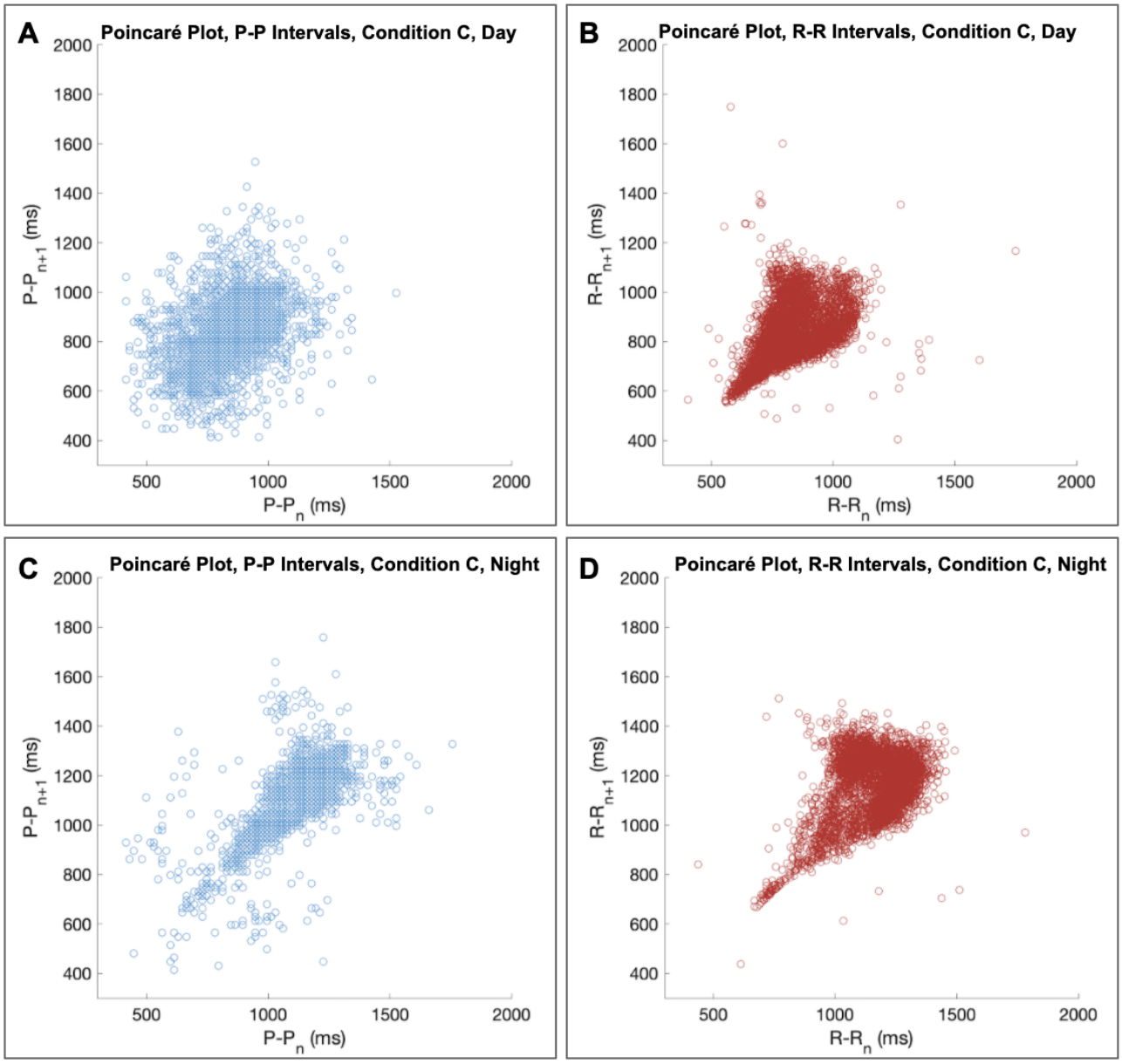
Time series plots of matched and edited R-R and P-P intervals (Figure 5) highlight several differences between the ECG and PPG methods. Similar to the mean N-N results, the data sets follow the same trends on average, but there are notable differences. First, P-P intervals seem to be less sensitive to changes in R-R intervals, as many shorter and longer intervals were not well matched. Fewer artifacts were observed in the R-R intervals that did not appear in the P-P interval signal, which may be attributable to less R-R interval editing.

Figure 5. Time series of matched time between 2 R peaks in an electrocardiogram and time between 2 P peaks in a photoplethysmogram or peak-to-peak intervals for a single participant under processing condition C during (A) day and (B) night. P-P: time between 2 P peaks in a photoplethysmogram or peak-to-peak intervals; R-R: time between 2 R peaks in an electrocardiogram.



Poincaré plots for the same participant under condition C are shown in Figure 6. The P-P and R-R plots during the day appear qualitatively different. Although plots of night data demonstrate more similarities, a greater number of outliers for shorter P-P intervals were observed.

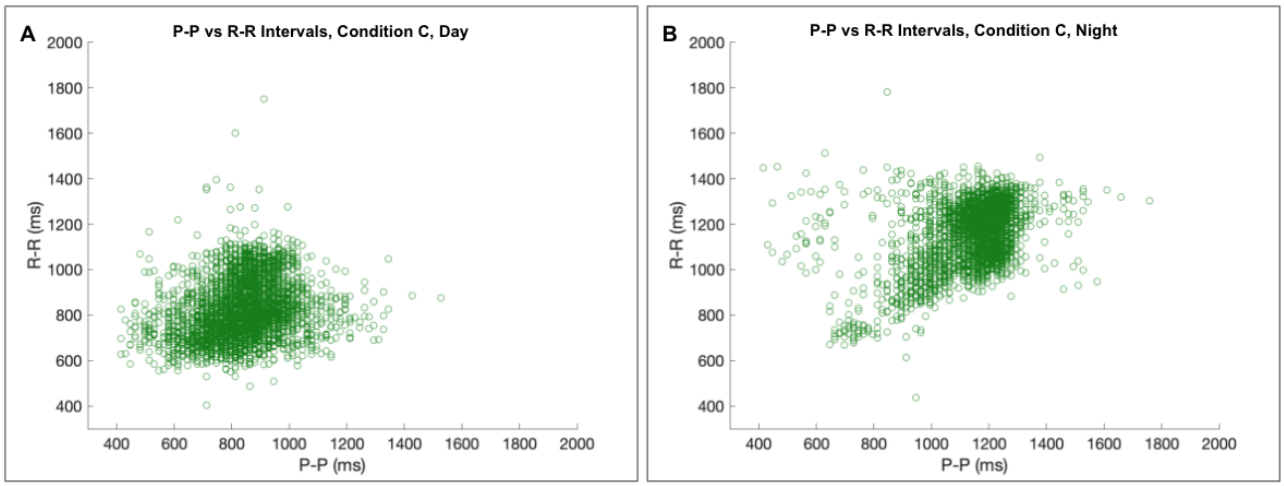
Figure 6. Poincaré plots for a single participant under processing condition C for (a) P-P intervals during the day, (b) R-R intervals during the day, (c) P-P intervals at night, and (d) R-R intervals at night. P-P: time between 2 P peaks in a photoplethysmogram or peak-to-peak intervals; R-R: time between 2 R peaks in an electrocardiogram.



Comparison of Free-Living Data Collection Conditions (Day vs Night)

Table 2 shows the difference in concurrent validity for night data versus day data under condition C. Closer examination of the data reveals further details. For 8 of 9 participants with a day and night data set, the average R^2 values were higher at night. The increase in R^2 is highlighted for one participant in Figure 7, where the R^2_{day}=0.26 and R^2_{night}=0.40. The magnitude of R^2 improvements from day to night differed between participants, ranging from −0.03 to 0.60 with an average improvement of 0.22 (SD 0.31). Paired *t* tests comparing changes in R^2 were significant (*P*=.01).

Figure 7. P-P versus R-R intervals for a participant under processing condition C during (A) day and (B) night. P-P: time between 2 P peaks in a photoplethysmogram or peak-to-peak intervals; R-R: time between 2 R peaks in an electrocardiogram.



Night collections were found to have a slight decrease in RMSE, indicated by a mean decrease in RMSE of 24 (SD 45) ms, ranging from −89 ms to +40 ms difference across participants. For the participant highlighted in Figure 7, $RMSE_{day}$=148 ms and $RMSE_{night}$=138 ms. Paired *t* tests comparing changes in RMSE from day to night approached significance (*P*=.09).

Although night data had more matched samples, an unpaired *t* test revealed that the difference between night and day samples was significant (*P*=.03). The mean percent increase in samples from day to night was 138.61% (SD 159). Differences in percentage loss of data owing to filtering (condition C vs condition A) were slightly higher during the day, averaging 13.31% (SD 11.47), versus night, 8.16% (SD 6.05). This difference did not reach significance under the unpaired *t* test (*P*=.25).

Tables 2 and 3 demonstrate that many PRV estimates of HRV measures were more accurate at night, with $|Error|_{avg}$ decreasing or remaining the same for NN, SDNN, RMSSD, SD1, SD2, and LF/HF ratio. $|Error|_{avg}$ for LF and HF remained approximately the same, whereas $|Error|_{avg}$ increased for pNN50 at night. Although $|Error|_{avg}$ generally decreased, paired *t* tests revealed more differences between PRV and HRV estimates across participants for night samples than day.

Discussion

Principal Findings

This paper examined the accuracy and concurrent validity of PRV measurements from a commercially available PPG sensor against HRV measurements obtained from a portable ECG sensor during unsupervised daytime and nighttime conditions. Accuracy and concurrent validity were examined across different editing methods and day and night collection conditions. In general, concurrent validity and HRV metrics were stronger at night compared with daytime conditions. Although collection during the night was more accurate with a lower mean error, this finding was not generalizable across all participants. Editing to remove outliers was effective in reducing noise, as reflected by the reduced RMSE for conditions B, C, and D. However, efforts to remove samples affected by motion artifacts using accelerometry (ie, condition D) were not as effective in this study compared with previous studies. The implications of these findings on ambulatory measurement of HRV using a commercially available PPG sensor to indicate health are discussed.

Although PPG sensors have strong mean HR measurement capabilities, the results from this study indicate poorer HRV capabilities. As expected, both ECG and PPG methods demonstrated similar mean R-R values with differences of less than 20 ms, reflecting established capabilities to estimate mean HR [17,20]. Examining beat-to-beat intervals using Bland-Altman plots, the mean error is close to zero (Figure 3). However, the wide variability of both under- and overestimated intervals indicates the presence of error-inducing factors, reflected in lower correlation (R^2) and large differences in calculated HRV metrics. Furthermore, Bland-Altman (Figure 3), time series (Figure 5), and Poincaré (Figure 6) plots indicate PPG sensing to trend toward underestimation errors.

The implications of PPG sensing errors on HRV metrics are highlighted in Table 3. pNN50 and LF/HF ratios were particularly sensitive to errors in point-to-point accuracy. PPG-derived estimates of pNN50 were poor, which corroborates previous reports of up to 30% error [12,21]. In addition, LF/HF ratio estimation errors were anticipated to be related to poorer HF estimates during the day arising from larger and more frequent (wrist) motion associated with regular activities of daily living. Across day and night collection conditions, SDNN estimates were similar when comparing ECG and PPG methods. SDNN has been shown to be associated with daytime occupational stress and has been hypothesized to demonstrate the parasympathetic autoregulation of the cardiac system in response to variations in cardiac output [45,46].

Day Versus Night Collection

When comparing day and night collection conditions, concurrent validity and HRV metrics indicate more accurate HRV estimates at night. Improved concurrent validity at night may be attributed to fewer errors related to ambient light changes at night [47,48], as presumably during sleep, the lighting conditions are

consistently darker. An important distinction between day and night was the larger sample size at night, likely owing to a more consistent Bluetooth stream and reduced noise arising from stationary conditions at night (ie, sleeping, lying down). Although mean R^2 values (condition C: mean 0.34, SD 0.21; condition D: mean 0.34, SD 0.21) were highest during night collections, the range of improvements varied across participants. Conversely, paired *t* tests revealed greater differences between PRV and HRV metrics (Table 3) at night. This may be attributed to the larger variability observed during the day, likely associated with a larger set and magnitude of motion during activities of daily living (compared with night). Considering significantly stronger concurrent validity measures (Table 2), coupled with smaller mean differences in HRV measures, we consider PRV estimates to better reflect HRV metrics at night. Although data collected at night may have improved concurrent validity, it is important to note that individuals may not have gone to bed or fallen asleep immediately after beginning the data collection. As a result, metrics taken at night may have captured features of wakefulness, such as shorter N-N intervals. For example, unaccounted for time in bed while remaining awake may have skewed the shape of the Poincaré plots as well as metric SD2.

Impact of Editing

Simple editing methods to improve PPG signals were examined in this study. PPG recordings are known to be affected by motion artifacts, contact force, posture, and ambient temperature [11]. Owing to the free-living nature of the study, the latter factors were not controlled. By adopting methods established by Morelli et al [28], RMSE improved by removing physiologically implausible intervals (condition B) and concurrent validity improved by deleting areas with rapid changes (condition C) at night. However, screening for motion artifacts using accelerometry signals (condition D) was ineffective at improving PPG-derived signals and HRV estimates. This is consistent with the findings by Baek and Shin [30], who were also unsuccessful in obtaining accurate long-term free-living recordings of wrist PPG using a custom device, even when performing deletions based on acceleration and P-P intervals differing by more than 15%. These findings, along with those by Georgiou et al [20], suggest that under unrestricted conditions, PRV is a poor estimator of HRV. Although other studies have looked to improve HRV estimation using alternative editing and correction methods [39,44,49], an exhaustive investigation of correction methods is beyond the scope of this study.

Our finding of relatively ineffective use of motion artifact compensation suggests that other factors affect PPG signals. For example, changes in respiration and peripheral vascular factors (ie, vascular volume, vasomotor activity, and vasoconstrictor waves) are known to affect the AC and DC frequency components of the PPG waveform [50]. In particular, the effect of peripheral vascular factors affects pulse transit time (PTT), or the time delay required for blood to travel between the heart and peripheral tissues [17]. Considering the range of daily activities (eg, body position changes [51], stress, and physical activity) lead to fluctuations in blood pressure [52], an assumed constant PTT is a likely source of error in estimating PRV parameters.

Limitations

The primary limitations of this study were the sample population and technical limitations of the devices. In this study, a convenience sample of 18- to 65-year-old participants with no known cardiac history participated. Although those with known cardiac conditions were excluded, the presence of underlying vascular disease in our cohort is unknown. As such, the findings may not be applicable to target disease populations. The impact of vascular conditions, such as atherosclerosis and cholesterol deposits in the arterial walls leading to decreased vessel compliance, which have been shown to alter pulse waveform from the classic triphasic pattern to mono- or biphasic patterns [53], remains to be examined. Although the number of participants was relatively small (n=10), the large number of within-participant samples (>1500 matched interval points per participant) and analyses supports the overall questions regarding sensor comparisons to estimate HRV.

The devices used in this study were limited in several ways. Both Shimmer3 and MB2 devices logged using separate device clocks, with potential for drift (approximately 1-2 s) over the course of a single trial. The devices were synchronized using an external mechanical stimulus (ie, 3 taps in 2 orientations) and by applying a data-driven delay estimate (ie, cross-correlation). Although these procedures have been used in previous studies with good results coupled with qualitative and quantitative observation of synchronized signals, the potential for dropped samples or desynchronization exists. The publicly available documentation for MB2 offers little to no insight into R-R interval processing or adjustment for when faced with motion artifacts and is no longer commercially available at the time of writing. Of note, signal drops were observed sporadically, including (1) large amplitude arm movements and (2) when MB2 was out of Bluetooth range from the smartphone for long periods (>10 min). We interpret these signal drops as obvious situations where motion artifacts and wireless communication are severely challenged with little to no impact on our findings. Furthermore, the resolution of RR intervals reported by MB2 was 10 ms, limiting accuracy similar to quantization error (ie, round-off errors). Given the large number of samples, resolution limitations are unlikely to affect mean values (eg, mean RR) but may increase variability (eg, RMSSD) estimates. However, the observed underestimation is unlikely to arise from quantization errors and are interpreted as systematic errors associated with the sensing method.

Implications for Future Work

Wearable technologies are becoming more sophisticated with commercially available products capable of providing consumers access to information previously limited to clinical settings, including HRV and ECG data to identify arrhythmias [54]. With this in mind, it is important to understand when and if the data can be considered valid and reliable. This study provides evidence that the relationship between PRV and HRV varies throughout the day, likely attributable to dynamic changes in the peripheral vasculature. The study findings suggest that PPG-derived measures of HRV are reasonable under particular

conditions (ie, at night), wherein this relationship is relatively stable for some HRV metrics (ie, SDNN and Poincaré axes). A deeper examination of factors modifying HRV estimation, particularly vascular factors, is yet to be conducted. As stated previously, our conclusions are drawn primarily from the HR and acceleration data. To further study these windows of high correlation in the future, other variables such as body temperature, cortisol levels, or a cognitive assessment of the participant's mental state may be beneficial.

In future, examining more editing, correction, and interpolation techniques for interbeat intervals may enhance the interpretability and quality of the P-P intervals obtained from commercially available wearables [44,55-57]. This study found that published movement artifact reduction techniques did not significantly improve the quality of our data. As wearable technologies continue to become more advanced, future studies in this field would benefit from the use of improved hardware and more robust sensors. For example, PulseOn and Apple Watch, both commercially available wearable devices, use different strategies to improve the quality of their signals. PulseOn uses multiwavelength PPG to reduce the sensitivity to movement artifacts and ambient light disturbances [47,48], demonstrating 99.57% accuracy during sleep [47]. Considering the lack of peripheral vascular indicators to account for changes in PTT, the Apple Watch approach of directly acquiring R-R intervals using built-in or peripheral ECG sensors (Kardia Band, Alivecor) [58] is justifiable.

Conclusions

The objective of this study was to assess the validity of PRV measurements taken from a PPG sensor by comparing it with the HRV measurements taken from a portable ECG while individuals were engaged in activities of daily living and during sleep. Although PPG sensors demonstrated greater validity at night, overall concurrent validity was poor. HRV metrics pNN50 and LF/HF ratio were especially sensitive to errors in point-to-point accuracy. Increased editing via deletion improved the RMSE but had a small impact on R^2. In comparing editing and deletion methods, screening for motion artifacts using accelerometry signals to remove error-prone signals was largely ineffective in improving HRV estimates. The best results were obtained under condition C (moving average method) at night, with the highest mean R^2 values. Overall, the findings from this study suggest that PRV is a poor surrogate of HRV under free-living conditions. Findings from this study indicate that advances in hardware and wearable technologies, such as multiwavelength PPG sensors, are warranted to unleash the potential of PRV to serve as a proxy measure for HRV.

Acknowledgments

This study was supported by the National Sciences and Engineering Research Council of Canada (NSERC) Discovery grant (RGPIN-2015-05317).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison of mean heart rate variability and pulse rate variability metrics under processing condition D.
[DOCX File , 15 KB - biomedeng_v5i1e17355_app1.docx]

References

1. Heart rate variability. In: Advances in Cardiac Signal Processing. New York, USA: Springer; 2007.
2. Bianchi AM. Signal Processing and Feature Extraction for Sleep Evaluation in Wearable Devices. In: International Conference of the IEEE Engineering in Medicine and Biology Society. 2006 Presented at: IEMBS'06; August 30-September 4, 2006; New York, NY, USA. [doi: 10.1109/iembs.2006.260547]
3. Amoedo A, Martnez-Costa MD, Moreno E. An analysis of the communication strategies of Spanish commercial music networks on the web: http://los40.com, http://los40principales.com, http://cadena100.es, http://europafm.es and http://kissfm.es. J Int Stud 2009 Feb 1;6(1):5-20 [FREE Full text] [doi: 10.1386/rajo.6.1.5_4]
4. Chuang C, Ye J, Lin W, Lee K, Tai Y. Photoplethysmography variability as an alternative approach to obtain heart rate variability information in chronic pain patient. J Clin Monit Comput 2015 Dec;29(6):801-806. [doi: 10.1007/s10877-015-9669-8] [Medline: 25708672]
5. Tan G, Dao TK, Farmer L, Sutherland RJ, Gevirtz R. Heart rate variability (HRV) and posttraumatic stress disorder (PTSD): a pilot study. Appl Psychophysiol Biofeedback 2011 Mar;36(1):27-35. [doi: 10.1007/s10484-010-9141-y] [Medline: 20680439]
6. Liao D, Carnethon M, Evans GW, Cascio WE, Heiss G. Lower heart rate variability is associated with the development of coronary heart disease in individuals with diabetes: the atherosclerosis risk in communities (ARIC) study. Diabetes 2002 Dec;51(12):3524-3531 [FREE Full text] [doi: 10.2337/diabetes.51.12.3524] [Medline: 12453910]
7. Reed M, Robertson C, Addison P. Heart rate variability measurements and the prediction of ventricular arrhythmias. QJM 2005 Feb;98(2):87-95. [doi: 10.1093/qjmed/hci018] [Medline: 15671474]

8. Martin T, Jovanov E, Raskovic D. Issues in wearable computing for medical monitoring applications: a case study of a wearable ECG monitoring device. In: Fourth International Symposium on Wearable Computers. 2000 Presented at: ISCW'00; October 16-17, 2000; Atlanta, GA, USA. [doi: 10.1109/iswc.2000.888463]
9. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
10. Chou Y, Zhang R, Feng Y, Lu M, Lu Z, Xu B. A real-time analysis method for pulse rate variability based on improved basic scale entropy. J Healthc Eng 2017;2017:7406896 [FREE Full text] [doi: 10.1155/2017/7406896] [Medline: 29065639]
11. Tamura T, Maeda Y, Sekine M, Yoshida M. Wearable photoplethysmographic sensors—past and present. Electronics 2014 Apr 23;3(2):282-302. [doi: 10.3390/electronics3020282]
12. Jeyhani V, Mahdiani S, Peltokangas M, Vehkaoja A. Comparison of HRV parameters derived from photoplethysmography and electrocardiography signals. Conf Proc IEEE Eng Med Biol Soc 2015;2015:5952-5955. [doi: 10.1109/EMBC.2015.7319747] [Medline: 26737647]
13. -. Comparison of Heart Rate Variability from PPG with That from ECG. In: The International Conference on Health Informatics. 2014 Presented at: CHI'14; September 15-17, 2014; Verona, Italy. [doi: 10.1007/978-3-319-03005-0_54]
14. Bolanos M, Nazeran H, Haltiwanger E. Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals. Conf Proc IEEE Eng Med Biol Soc 2006;2006:4289-4294. [doi: 10.1109/IEMBS.2006.260607] [Medline: 17946618]
15. Lu G, Yang F, Taylor JA, Stein JF. A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. J Med Eng Technol 2009;33(8):634-641. [doi: 10.3109/03091900903150998] [Medline: 19848857]
16. Russoniello CV, Pougtachev V, Zhirnov E, Mahar MT. A measurement of electrocardiography and photoplethesmography in obese children. Appl Psychophysiol Biofeedback 2010 Sep;35(3):257-259. [doi: 10.1007/s10484-010-9136-8] [Medline: 20552266]
17. Selvaraj N, Jaryal A, Santhosh J, Deepak KK, Anand S. Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. J Med Eng Technol 2008;32(6):479-484. [doi: 10.1080/03091900701781317] [Medline: 18663635]
18. Cropley M, Plans D, Morelli D, Sütterlin S, Inceoglu I, Thomas G, et al. The association between work-related rumination and heart rate variability: a field study. Front Hum Neurosci 2017;11:27 [FREE Full text] [doi: 10.3389/fnhum.2017.00027] [Medline: 28197087]
19. Amoedo A, Martnez-Costa MD, Moreno E. An analysis of the communication strategies of Spanish commercial music networks on the web: http://los40.com, http://los40principales.com, http://cadena100.es, http://europafm.es and http://kissfm.es. radio journal: international studies in 2009 Feb 01;6(1):5-20 [FREE Full text] [doi: 10.1386/rajo.6.1.5_4]
20. Georgiou K, Larentzakis AV, Khamis NN, Alsuhaibani GI, Alaska YA, Giallafos EJ. Can wearable devices accurately measure heart rate variability? A systematic review. Folia Med (Plovdiv) 2018 Mar 1;60(1):7-20. [doi: 10.2478/folmed-2018-0012] [Medline: 29668452]
21. Schäfer A, Vagedes J. How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram. Int J Cardiol 2013 Jun 5;166(1):15-29. [doi: 10.1016/j.ijcard.2012.03.119] [Medline: 22809539]
22. Gil E, Orini M, Bailón R, Vergara JM, Mainardi L, Laguna P. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. Physiol Meas 2010 Sep;31(9):1271-1290. [doi: 10.1088/0967-3334/31/9/015] [Medline: 20702919]
23. Salehizadeh S, Dao D, Bolkhovsky J, Cho C, Mendelson Y, Chon K. A novel time-varying spectral filtering algorithm for reconstruction of motion artifact corrupted heart rate signals during intense physical activities using a wearable photoplethysmogram sensor. Sensors (Basel) 2015 Dec 23;16(1):- [FREE Full text] [doi: 10.3390/s16010010] [Medline: 26703618]
24. Jarchi D, Casson A. Description of a database containing wrist PPG signals recorded during physical exercise with both accelerometer and gyroscope measures of motion. Data 2016 Dec 24;2(1):1 [FREE Full text] [doi: 10.3390/data2010001]
25. Ram MR, Madhav KV, Krishna EH, Komalla NR, Reddy KA. A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter. IEEE Trans Instrum Meas 2012 May;61(5):1445-1457. [doi: 10.1109/tim.2011.2175832]
26. Lee C, Zhang Y. Reduction of Motion Artifacts From Photoplethysmographic Recordings Using a Wavelet Denoising Approach. In: Asian-Pacific Conference on Biomedical Engineering. 2003 Presented at: Kyoto, Japan; October 20-23, 2003; Kyoto, Japan. [doi: 10.1109/apbme.2003.1302650]
27. Kim B, Yoo S. Motion artifact reduction in photoplethysmography using independent component analysis. IEEE Trans Biomed Eng 2006 Mar;53(3):566-568. [doi: 10.1109/tbme.2005.869784]
28. Morelli D, Bartoloni L, Colombo M, Plans D, Clifton DA. Profiling the propagation of error from PPG to HRV features in a wearable physiological-monitoring device. Healthc Technol Lett 2018 Apr;5(2):59-64 [FREE Full text] [doi: 10.1049/htl.2017.0039] [Medline: 29750114]

29. Kos M, Khaghani-Far I, Gordon CM, Pavel M, Jimison HB. Can accelerometry data improve estimates of heart rate variability from wrist pulse PPG sensors? Conf Proc IEEE Eng Med Biol Soc 2017 Jul;2017:1587-1590 [FREE Full text] [doi: 10.1109/EMBC.2017.8037141] [Medline: 29060185]
30. Baek HJ, Shin J. Effect of missing inter-beat interval data on heart rate variability analysis using wrist-worn wearables. J Med Syst 2017 Aug 15;41(10):147. [doi: 10.1007/s10916-017-0796-2] [Medline: 28812280]
31. The ECG Manual: An Evidence-Based Approach. New York, USA: Springer; 2018.
32. R Wave Detection in the ECG. MathWorks. URL: https://www.mathworks.com/help/wavelet/ug/r-wave-detection-in-the-ecg.html [accessed 2019-08-09]
33. Li C, Zheng C, Tai C. Detection of ECG characteristic points using wavelet transforms. IEEE Trans Biomed Eng 1995 Jan;42(1):21-28. [doi: 10.1109/10.362922] [Medline: 7851927]
34. Manurmath JC, Raveendra M. CMATLAB based ECG signal classification. Int J SciEng Technol Res 1946:1950.
35. Agelink MW, Malessa R, Baumann B, Majewski T, Akila F, Zeit T, et al. Standardized tests of heart rate variability: normal ranges obtained from 309 healthy humans, and effects of age, gender, and heart rate. Clin Auton Res 2001 Apr;11(2):99-108. [doi: 10.1007/BF02322053] [Medline: 11570610]
36. Murray MP, Spurr GB, Sepic SB, Gardner GM, Mollinger LA. Treadmill vs floor walking: kinematics, electromyogram, and heart rate. J Appl Physiol (1985) 1985 Jul;59(1):87-91. [doi: 10.1152/jappl.1985.59.1.87] [Medline: 4030579]
37. Snyder F, Hobson JA, Morrison DF, Goldfrank F. Changes in respiration, heart rate, and systolic blood pressure in human sleep. J Appl Physiol 1964 May 1;19(3):417-422. [doi: 10.1152/jappl.1964.19.3.417]
38. 38 L. Reinerman-Jones, J. Harris, and A. Watson, ?Considerations for Using Fitness Trackers in Psychophysiology Research,? Human Interface and the Management of Information: Information, Knowledge and Interaction Design. pp. 598?606 2017:-. [doi: 10.1007/978-3-319-58521-5_47]
39. Giles DA, Draper N. Heart rate variability during exercise: a comparison of artefact correction methods. J Strength Cond Res 2018 Mar;32(3):726-735. [doi: 10.1519/JSC.0000000000001800] [Medline: 29466273]
40. Amoedo A, Martnez-Costa MD, Moreno E. An analysis of the communication strategies of Spanish commercial music networks on the web: http://los40.com, http://los40principales.com, http://cadena100.es, http://europafm.es and http://kissfm.es. J Int Stud 2009 Feb 1;6(1):5-20 [FREE Full text] [doi: 10.1386/rajo.6.1.5_4]
41. Li K, Rüdiger H, Ziemssen T. Spectral analysis of heart rate variability: time window matters. Front Neurol 2019;10:545 [FREE Full text] [doi: 10.3389/fneur.2019.00545] [Medline: 31191437]
42. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018 Sep;18(3):91-93 [FREE Full text] [doi: 10.1016/j.tjem.2018.08.001] [Medline: 30191186]
43. Giavarina D. Understanding bland altman analysis. Biochem Med (Zagreb) 2015;25(2):141-151 [FREE Full text] [doi: 10.11613/BM.2015.015] [Medline: 26110027]
44. Morelli D, Rossi A, Cairo M, Clifton DA. Analysis of the impact of interpolation methods of missing RR-intervals caused by motion artifacts on HRV features estimations. Sensors (Basel) 2019 Jul 18;19(14):- [FREE Full text] [doi: 10.3390/s19143163] [Medline: 31323850]
45. Borchini R, Bertù L, Ferrario MM, Veronesi G, Bonzini M, Dorso M, et al. Prolonged job strain reduces time-domain heart rate variability on both working and resting days among cardiovascular-susceptible nurses. Int J Occup Med Environ Health 2015;28(1):42-51 [FREE Full text] [doi: 10.2478/s13382-014-0289-1] [Medline: 26159946]
46. Billman GE. The effect of heart rate on the heart rate variability response to autonomic interventions. Front Physiol 2013;4:222 [FREE Full text] [doi: 10.3389/fphys.2013.00222] [Medline: 23986716]
47. Parak J, Tarniceriu A, Renevey P, Bertschi M, Delgado-Gonzalo R, Korhonen I. Evaluation of the beat-to-beat detection accuracy of PulseOn wearable optical heart rate monitor. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:8099-8102. [doi: 10.1109/EMBC.2015.7320273] [Medline: 26738173]
48. Renevey PH, Sola J, Theurillat P, Bertschi M, Krauss J, Andries D, et al. Validation of a wrist monitor for accurate estimation of RR intervals during sleep. Conf Proc IEEE Eng Med Biol Soc 2013;2013:5493-5496. [doi: 10.1109/EMBC.2013.6610793] [Medline: 24110980]
49. Lee J, Kim J, Shin M. Correlation analysis between electrocardiography (ECG) and photoplethysmogram (PPG) data for driver’s drowsiness detection using noise replacement method. Procedia Computer Science 2017;116:421-426 [FREE Full text] [doi: 10.1016/j.procs.2017.10.083]
50. Maeda Y, Sekine M, Tamura T. Relationship between measurement site and motion artifacts in wearable reflected photoplethysmography. J Med Syst 2011 Oct;35(5):969-976. [doi: 10.1007/s10916-010-9505-0] [Medline: 20703691]
51. Olufsen MS, Ottesen JT, Tran HT, Ellwein LM, Lipsitz LA, Novak V. Blood pressure and blood flow variation during postural change from sitting to standing: model development and validation. J Appl Physiol (1985) 2005 Oct;99(4):1523-1537 [FREE Full text] [doi: 10.1152/japplphysiol.00177.2005] [Medline: 15860687]
52. Drinnan MJ, Allen J, Murray A. Relation between heart rate and pulse transit time during paced respiration. Physiol Meas 2001 Aug;22(3):425-432. [doi: 10.1088/0967-3334/22/3/301] [Medline: 11556663]
53. Azzopardi YM, Gatt A, Chockalingam N, Formosa C. Agreement of clinical tests for the diagnosis of peripheral arterial disease. Prim Care Diabetes 2019 Feb;13(1):82-86. [doi: 10.1016/j.pcd.2018.08.005] [Medline: 30201222]

54. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: the apple heart study. Am Heart J 2019 Jan;207:66-75 [FREE Full text] [doi: 10.1016/j.ahj.2018.09.002] [Medline: 30392584]
55. Lang M. Automatic near real-time outlier detection and correction in cardiac interbeat interval series for heart rate variability analysis: singular spectrum analysis-based approach. JMIR Biomed Eng 2019 Jan 30;4(1):e10740 [FREE Full text] [doi: 10.2196/10740]
56. Citi L, Brown EN, Barbieri R. A real-time automated point-process method for the detection and correction of erroneous and ectopic heartbeats. IEEE Trans Biomed Eng 2012 Oct;59(10):2828-2837 [FREE Full text] [doi: 10.1109/TBME.2012.2211356] [Medline: 22875239]
57. Tarvainen MP, Niskanen J, Lipponen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV--heart rate variability analysis software. Comput Methods Programs Biomed 2014;113(1):210-220. [doi: 10.1016/j.cmpb.2013.07.024] [Medline: 24054542]
58. Bumgarner JM. Smartwatch algorithm for automated detection of atrial fibrillation. J Am Coll Cardiol 2018;71(21):-.

Abbreviations

ECG: electrocardiogram
HF: high-frequency power (0.15-0.40 Hz)
HR: heart rate
HRV: heart rate variability
LA: left arm
LF: low-frequency power (0.04-0.15 Hz)
LL: left leg
MB2: Microsoft Band 2
N-N: N-N intervals (either R-R or P-P intervals)
pNN50: percent of subsequent differences more than 50 ms
P-P: time between 2 P peaks in a PPG or peak-to-peak intervals
PPG: photoplethysmography
PR: pulse rate
PRV: pulse rate variability
PTT: pulse transit time
R2: Pearson correlation coefficient
RA: right arm
RMSE: root mean square error
RMSSD: root mean square of subsequent differences
R-R: time between 2 R peaks in an ECG
SD1: standard deviation of short (x=y) Poincaré plot axis
SD2: standard deviation of long (orthogonal to x=y) Poincaré plot axis
SDNN: standard deviation of all N-N intervals, also known as RRSD
Wt: mean change in triaxial acceleration

Edited by G Eysenbach; submitted 10.12.19; peer-reviewed by A Vehkaoja, M Lang, C Boodoo; comments to author 10.02.20; revised version received 31.05.20; accepted 26.07.20; published 03.11.20.

Please cite as:
Lam E, Aratia S, Wang J, Tung J
Measuring Heart Rate Variability in Free-Living Conditions Using Consumer-Grade Photoplethysmography: Validation Study
JMIR Biomed Eng 2020;5(1):e17355
URL: http://biomedeng.jmir.org/2020/1/e17355/
doi:10.2196/17355
PMID:

Original Paper

Predictors of Walking Activity in Patients With Systolic Heart Failure Equipped With a Step Counter: Randomized Controlled Trial

Josefine Dam Gade[1], MSc; Helle Spindler[2], MSc, PhD; Malene Hollingdal[3], MD, PhD; Jens Refsgaard[3], MD, PhD; Lars Dittmann[4], MSc, PhD; Lars Frost[5,6], MD, PhD, DMSc; Kiomars Mahboubi[7], MD; Birthe Dinesen[1], MSc, PhD

[1]Laboratory for Welfare Technology - Telehealth & Telerehabilitation, Department of Health Science and Technology, Aalborg University, Aalborg East, Denmark

[2]Department of Psychology and Behavioral Sciences, Aarhus University, Aarhus, Denmark

[3]Cardiology Ward, Regional Hospital in Viborg, Viborg, Denmark

[4]Department of Photonics Engineering, Danish Technical University, Copenhagen, Denmark

[5]Cardiology Ward, Regional Hospital in Silkeborg, Silkeborg, Denmark

[6]Aarhus University, Aarhus, Denmark

[7]Cardiology Ward, Regional Hospital in Randers, Randers, Denmark

Corresponding Author:
Josefine Dam Gade, MSc
Laboratory for Welfare Technology - Telehealth & Telerehabilitation
Department of Health Science and Technology
Aalborg University
Fredrik Bajers Vej 7
Aalborg East, 9220
Denmark
Phone: 45 28342888
Email: jdg@hst.aau.dk

Abstract

Background: Physical activity has been shown to decrease cardiovascular mortality and morbidity. Walking, a simple physical activity which is an integral part of daily life, is a feasible and safe activity for patients with heart failure (HF). A step counter, measuring daily walking activity, might be a motivational factor for increased activity.

Objective: The aim of this study was to examine the association between walking activity and demographical and clinical data of patients with HF, and whether these associations could be used as predictors of walking activity.

Methods: A total of 65 patients with HF from the Future Patient Telerehabilitation (FPT) program were included in this study. The patients monitored their daily activity using a Fitbit step counter for 1 year. This monitoring allowed for continuous and safe data transmission of self-monitored activity data.

Results: A higher walking activity was associated with younger age, lower New York Heart Association (NYHA) classification, and higher ejection fraction (EF). There was a statistically significant correlation between the number of daily steps and NYHA classification at baseline (P=.01), between the increase in daily steps and EF at baseline (P<.001), and between the increase in daily steps and improvement in EF (P=.005). The patients' demographic, clinical, and activity data could predict 81% of the variation in daily steps.

Conclusions: This study demonstrated an association between demographic, clinical, and activity data for patients with HF that could predict daily steps. A step counter can thus be a useful tool to help patients monitor their own physical activity.

Trial Registration: ClinicalTrials.gov NCT03388918; https://clinicaltrials.gov/ct2/show/NCT03388918

International Registered Report Identifier (IRRID): RR2-10.2196/14517

(JMIR Biomed Eng 2020;5(1):e20776) doi:10.2196/20776

KEYWORDS

heart failure; cardiovascular rehabilitation; step counters; physical activity; telerehabilitation

Introduction

Cardiovascular diseases account for 13%-15% of all deaths worldwide and 24.8% in Europe and are thereby the leading cause of mortality [1-4]. Heart failure (HF) is one of the most common cardiovascular causes of mortality [4,5], with a prevalence of 0.4%-2% among the general population and 2.3%-16% among people aged over 75 [6]. Physical activity has been proven to decrease cardiovascular mortality and morbidity [3,7,8] and has therefore been a main focus of rehabilitation programs targeting lifestyle changes [3,7] in order to improve patient recovery. Generally, cardiac rehabilitation includes interventions such as physical activity, improved diet, and weight control, with the aim of improving patients' recovery, functional capacity, psychosocial well-being, and quality of life [3,9,10]. However, participation in rehabilitation programs is often low. In order to increase adherence, rehabilitation programs have been introduced that are more accessible and individualized for the patient, such as home-based cardiac rehabilitation programs involving cardiac telerehabilitation [3,11,12]. Telerehabilitation is defined as rehabilitation activities using information and communication technologies [13].

Walking, a simple physical activity, is an integral part of our daily routine [14] and is suitable for cardiac patients. Walking is both safe and feasible for almost all patients with HF [3]. Today, many self-tracking devices provide information to users regarding their walking activity, such as the number of steps, and may therefore assist patients to monitor their walking. Such self-tracking devices are based on sensor technologies that allow continuous monitoring of physiological data by the walker in the context of everyday life [12]. Telemonitoring concepts including smart device–based monitoring, especially measurements of physical activity, are considered to be beneficial for patients with HF [3,12,15].

People are considered physically active if they perform more than 30 minutes of moderate to intense physical activity per day; this would be equivalent to approximately 7000-10,000 steps per day [3,8]. However, studies of walking activity among cardiac patients have found that these patients walked a mean number of 5889 [3], 7027 [15], and 5869 [12] daily steps. These 3 studies included patients suffering from acute coronary syndrome, coronary artery disease [15], and HF [12], as well as those who underwent coronary artery bypass repair or valve replacement [3].

Bäck et al [15] presented a step index that describes the walking activity for cardiac patients, in which less than 3000 steps per day represent low activity, 3000-9999 steps per day represent medium activity, and over 10,000 steps per day represent high activity. Bäck et al [15] found that 11.18% of the study population had a low activity level, 69.65% a medium activity level, and 19.17% a high activity level. This same tripartite classification was also used by Thorup et al [3] to categorize the walking activity of cardiac patients, where 22% of the study population were classified as having a low activity level, 64% medium, and 14% high.

Based on the initial findings from the research literature and from pilot studies, the Future Patient Telerehabilitation (FPT) program proposed a new approach to telerehabilitation for patients with HF to increase their quality of life and educate them to monitor any worsening of their symptoms. The patients used self-tracking devices for monitoring physical activity, blood pressure, sleep, respiration, and pulse [11]. In the FPT program, recorded data from the daily measurements were available on a shared web platform called the HeartPortal used by patients, their relatives, and health care professionals [11].

In this paper, we report on a substudy of FPT focusing on the walking activity of patients with HF in a telerehabilitation program over 1 year. Our aim is to explore (1) the duration of usage of the step counter devices; (2) the eventual increase in the average number of daily steps; (3) the correlation between the number of daily steps with ejection fraction (EF), New York Heart Association (NYHA) class [16], and age, respectively; and (4) whether baseline EF, NYHA class, age, or gender can be used to predict the daily number of steps.

Methods

Future Patient Telerehabilitation Program

This substudy utilizes the data from the intervention group that received telerehabilitation (TR group) of the FPT (ClinicalTrials.gov: NCT03388918 and the Danish Ethical Committee: N-20160055). The TR group participated in the telerehabilitation program, whereas the control group followed a conventional rehabilitation program at the health care center. The intervention in the FPT consisted of 3 phases: (1) Education and titration of medicine (0-3 months), (2) Telerehabilitation in health care center or call center (approximately 3 months), (3) Daily monitoring via telerehabilitation (approximately 6 months), corresponding to a follow-up of 12 months [11].

During the participation period, the patients in the TR group received a blood pressure device, weight scale, sleep sensor, step counter, and an iPad. In addition, they were also given access to the HeartPortal, which is a digital toolbox developed on the basis of patient feedback [17], that functions as an interactive learning module. The HeartPortal consists of (1) an interactive information site for patient education, (2) a communication platform enabling patients to communicate directly with health care professionals via online messages, (3) visualization of measured values, and (4) patient-reported outcomes data [11].

Based on the study by Munck et al [18], evaluating self-tracking devices for telerehabilitation of patients with HF, Fitbit step counters were used in the FPT, as these received the highest user evaluation and the lowest step count error percentages. The choice of the Fitbit device is consistent with the systematic review by Bunn et al [19], which concludes that Fitbit products

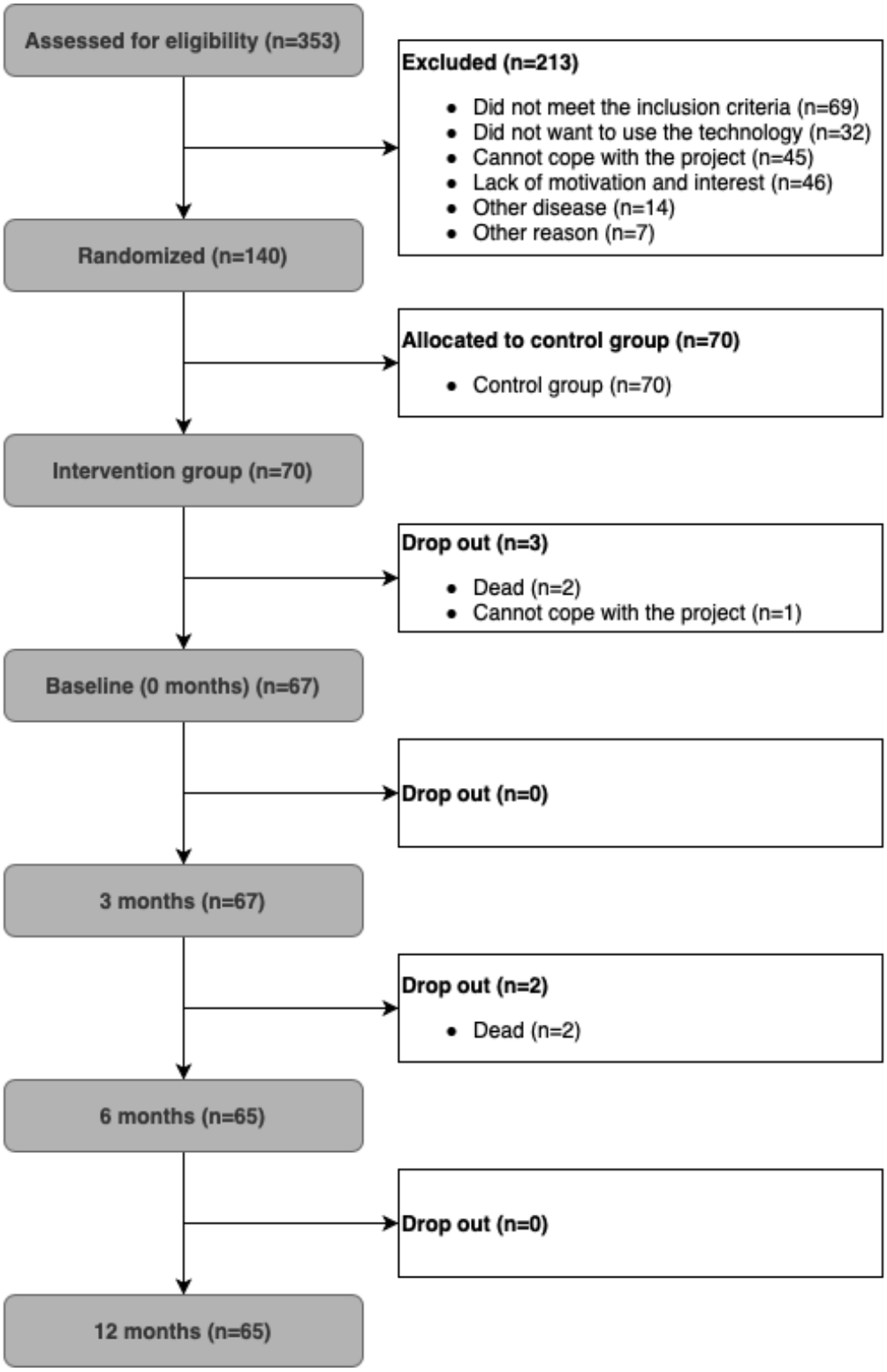
are generally more accurate regarding step counting compared with other wearable physical activity trackers [19].

Participants and Recruitment

The target group of the FPT included patients diagnosed with HF according to the NYHA class I-IV. The patients were recruited from cardiology wards at hospitals in 4 Danish cities, all of which were part of the Regional Hospital Central Jutland [11]. Patients were eligible for the study if (1) they were diagnosed with HF according to NYHA class I-IV, hereof a maximum of 20% of the patients were allowed from NYHA class I, and had experienced an HF-related hospitalization or visit to the outpatient clinic within the past 2 weeks; (2) the patient was 18 years of age or older; (3) the patient had to live in Viborg, Skive, Silkeborg, or Randers municipality; (4) the patient was living at home and capable of caring for himself/herself; (5) they had basic computer skills or a relative who had basic computer skills; and (6) the patient was able to sign an informed consent form. Furthermore, patients were excluded if they (1) were pregnant; (2) had a drug addiction defined as the use of cannabis, opioids, or other drugs; (3) had previous neurologic, musculoskeletal, or cognitive disability or active psychiatric history (as noted in the medical record) other than depression or anxiety related to cardiac or other chronic illness; (4) lacked the ability to cooperate; or (5) did not speak Danish.

In total, 140 patients were included in the FPT, of which 70 patients were randomly allocated to the TR group and 70 to the control group [11]. Of the 70 patients from the TR group, 65 completed the FPT program. Figure 1 (CONSORT [Consolidated Standards of Reporting Trials] diagram) illustrates the randomization procedure, follow-up, and drop out reasons for the TR group. Multimedia Appendix 1 provides CONSORT-EHEALTH (Randomized Controlled Trials of Electronic and Mobile Health Applications and Online TeleHealth) checklist.

Figure 1. CONSORT diagram of the Future Patient trial. CONSORT: Consolidated Standards of Reporting Trials.



Walking Activity

One of the clinical measures used in the FPT was the patients' activity level, which was recorded using a Fitbit step counter, either Fitbit Zip or Fitbit Charge HR [11]. These step counters were selected based on the results presented in Munck et al [18], which found these Fitbit step counters to have the lowest error margin. Both devices calculate steps based on data from the internal 3-axis accelerometer, and the step data were acquired every 20 minutes using application programming interface (API). All patients were asked to wear the step counter during all waking hours, from 8 am to 9 pm, during the course of the project period.

To account for single days of missing values from the step counters, the median value of daily steps during a week has been evaluated and used as the indicator of the general activity level for the patient for that particular week.

The duration of time in which the patients have used the step counters was defined as the number of active days compared to (1) the total number of days enrolled in the study, (2) EF, and (3) NYHA classification. So-called active days have been defined as days with more than 100 steps per day. This 100-step cut-off point was the same as that presented in Thorup et al [3], in order to avoid faulty measurements when the step counter was moved around but not worn. In addition, activity levels were classified as low, medium, or high, following Bäck et al [15].

Data Acquisition

Clinical data (weight, blood pressure, heart rate, EF determined through a standardized transthoracic echocardiography, NYHA classification, and etiology of HF) and sociodemographic data (age, gender, civil status, education level, and work status) were acquired from the patient's medical journal or through self-reporting. Data on the patient's daily activity were acquired from Fitbit every 20 minutes using API and consisted of the number of steps taken.

Statistical Analysis

Prior to analysis, data were examined regarding the normality of the distribution using a Shapiro–Wilk test.

To investigate how long the patients chose to use the step counter and the progression of steps, a one-way analysis of variance (ANOVA) with repeated measures was used.

The association between the activity and clinical parameters was investigated using Pearson correlation coefficient, which determines the covariance of 2 variables divided by the product of their standard deviations.

A linear hierarchical regression analysis was used to determine whether demographic, clinical, and activity parameters can be used to explain a statistically significant amount of variance in the mean number of steps. These were presented in different blocks, of which the first block contained demographic data (age and gender), the second block contained clinical data (baseline EF and baseline NYHA), and the third block contained activity data (baseline number of daily steps).

The statistical analysis was performed using SPSS Statistics version 25 (IBM, Inc.).

Results

Patient Characteristics

The patient characteristics of the TR group are presented in Table 1.

Table 1. Patient characteristics at baseline.

Variable	Values	
	Baseline (N=65)	Follow-up[a] (N=65)
Age (years) by gender		
Men (n=49)		
Mean (SD)	61.71 (10.49)	—
Range	35-81	—
25th-75th percentile	56.5-69	—
Women (n=16)		
Mean (SD)	60.31 (11.31)	—
Range	43-81	—
25th-75th percentile	51.5-69.75	—
Men and women (n=65)		
Mean (SD)	61.37 (10.63)	—
Range	35-81	—
25th-75th percentile	55-69	—
Clinical parameters		
Weight (n=16)		
Mean (SD)	85.19 (20.55)	84.53 (22.37)
Range	56-166	51.60-168
25th-75th percentile	70.03-98.5	69.53-96.65 (n=54)
Systolic blood pressure (mmHg) (n=65)		
Mean (SD)	124.17 (17.62)	116.98 (16.45)
Range	84-172	83-152
25th-75th percentile	112.75-134.25	105-128.25 (n=40)
Diastolic blood pressure (mmHg) (n=65)		
Mean (SD)	79.08 (11.14)	72.45 (10.08)
Range	48-122	46-93
25th-75th percentile	70.75-86	66.50-78.75 (n=40)
Heart rate (beats/minute) (n=65)		
Mean (SD)	78.35 (17.72)	68.11 (16.36)
Range	46-119	41-116
25th-75th percentile	66-90.5	57-77 (n=40)
Ejection fraction (%) (n=65)		
Mean (SD)	31.74 (8.54)	43.50 (7.25)
Range	10-45	20-60
25th-75th percentile	25-40	40-50 (n=65)
NYHA[b], n (%)		
I	10 (15)	26 (40)
II	41 (63)	35 (54)
III	12 (18)	4 (6)
IV	2 (3)	0 (0)
Etiology of heart failure, n (%)		

Variable	Values	
	Baseline (N=65)	Follow-up[a] (N=65)
Ischemic	30 (46)	—
Idiopathic	17 (26)	—
Hypertension	6 (9)	—
Valvular heart disease	8 (12)	—
Alcoholic	0 (0)	—
Postpartum	0 (0)	—
Chemotherapy	0 (0)	—
Other etiology	18 (28)	—
Recruitment, n (%)		
During hospitalization	22 (34)	—
From visit in outpatient clinic	43 (66)	—
Civil status, n (%)		
Single	23 (35)	—
Married/Living with a partner	42 (65)	—
Education, n (%)		
Primary school	4 (6)	—
Unskilled	15 (23)	—
Skilled	30 (46)	—
High school	5 (8)	—
Bachelor’s degree	9 (14)	—
Master’s degree	1 (2)	—
PhD+	1 (2)	—
Work status, n (%)		
Unemployed	0 (0)	—
Sick leave	19 (29)	—
Works under 20 hours/week	5 (8)	—
Works 20-36 hours/week	2 (3)	—
Works full-time 37 hours/week	9 (14)	—
Retired	30 (46)	—

[a]There are missing data from some patients for the clinical parameters at follow-up. The number of patients, for whom the data were available within 2 months prior to follow-up, is stated in parentheses.

[b]NYHA: New York Heart Association.

Usage of Step Counter

Table 2 focuses on activity and its relation to clinical and demographical data. The overall period during which the patients used the step counter in relation to their activity level is shown in Table 2. This duration is presented as the total number of days in which the patients were enrolled in the study and the number of active days in which they were using the step counter. Also presented in the table are their EF and NYHA classifications and gender with regard to the different activity levels.

Table 2. Duration of step counter use with regard to the patients' activity levels.

Variable	All patients	Activity level (steps/day)			*P* value
		Low (<2999)	Medium (3000-9999)	High (≥10,000)	
Number of patients	65	8	47	10	
Days using step counter, mean (SD)					
Total days	358.18 (54.57)	384.13 (61.53)	358.04 (56.77)	338.1 (26.61)	.208
Active days	317.91 (86.34)	310.63 (90.11)	319.32 (92.32)	317.10 (55.42)	.966
Activity days/total days (%)	88.65 (19.75)	81.00 (20.56)	88.74 (20.12)	94.40 (16.94)	.365
Clinical variables					
Ejection fraction (%), mean (SD)	31.74 (8.54)	30.31 (8.50)	32.19 (8.60)	30.75 (8.98)	.788
NYHA[a] classification, n (%)					**.057**
I	10 (100)	0 (0)	8 (80)	2 (20)	
II	41 (100)	4 (10)	31 (76)	6 (15)	
III	12 (100)	3 (25)	7 (58)	2 (17)	
IV	2 (100)	1 (50)	1 (50)	0 (0)	
Sociodemographic characteristics					
Gender, n (%)					**.659**
Male	49 (100)	5 (10)	36 (73)	8 (16)	
Female	16 (100)	3 (19)	11 (69)	2 (13)	

[a]NYHA: New York Heart Association.

Table 2 shows that overall, patients used the step counter 88.65% of the total period in which they were enrolled in the study, and that the time interval in which they use the step counter increased in line with increases in the activity level.

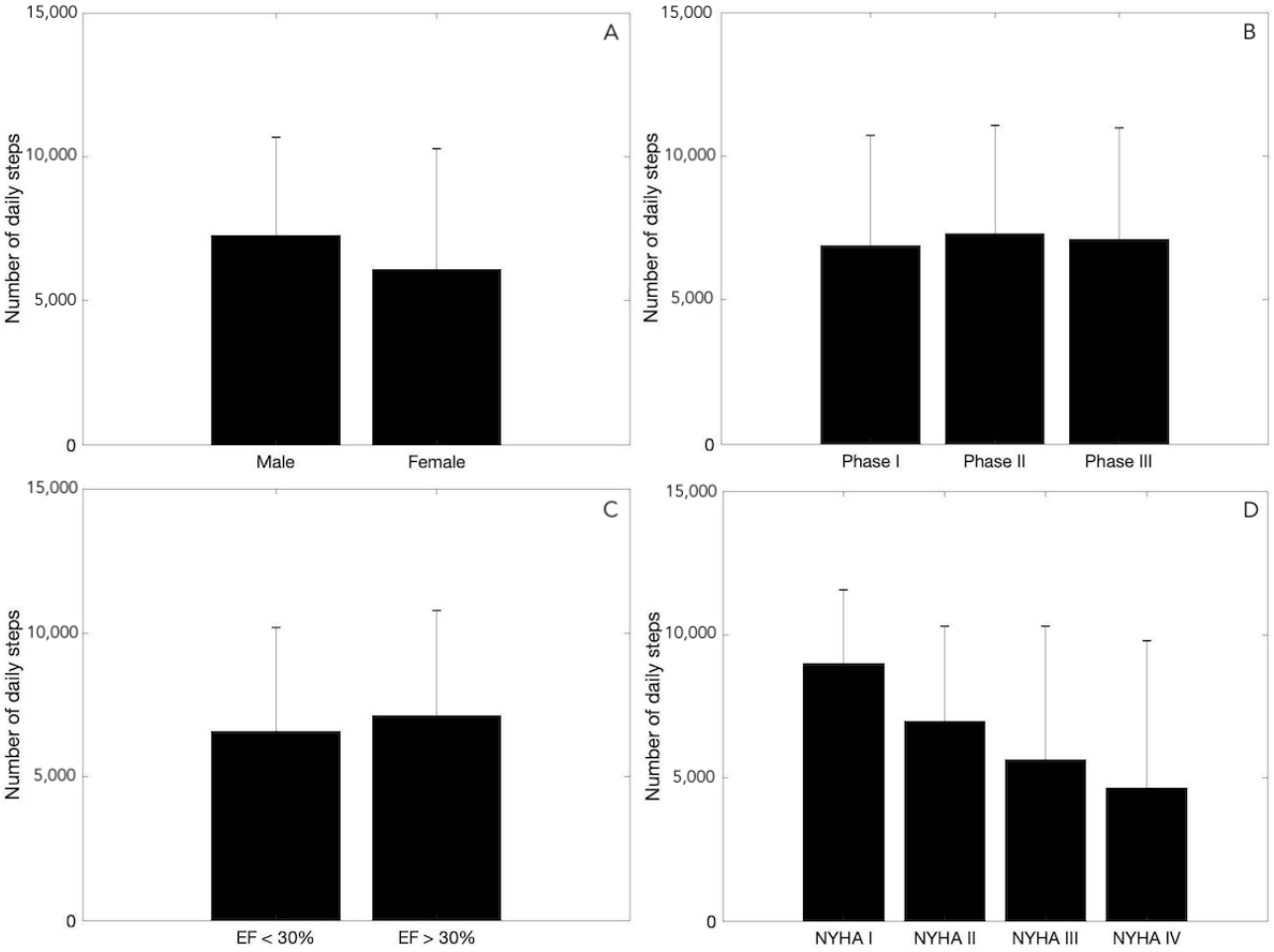
Increase in the Number of Daily Steps

During all 3 rehabilitation phases, the patients walked a mean number of 6962.81 (SD 3630.74) daily steps, with a minimum of 227 daily steps and a maximum of 25,499 daily steps. The mean number of daily steps during the 3 different phases were 5868.56 (SD 3912.34) daily steps in phase I, 7233.89 (SD 4197.36) daily steps in phase II, and 7338.58 (SD 4359.67) daily steps in phase III. These mean values are calculated from the different phases which had a duration of 0-3 months in phase I, 3 months in phase II, and 6 months in phase III.

A one-way ANOVA with repeated measures showed that the mean number of steps was not significantly different across the 3 phases ($F_{2,62}$=1.137, P=.318). Furthermore, a one-way ANOVA with repeated measures using gender ($F_{2,62}$=0.015, P=.978), EF ($F_{2,62}$=2.585, P=.87), and NYHA classification ($F_{2,62}$=2.229, P=.119) as covariates showed that the mean number of steps was not significantly different across the 3 phases.

Figure 2 presents the mean number of daily steps in bar plots with regard to the gender of the patients, the 3 phases of the FPT, EF with a threshold of 30%, and the patient's NYHA classification.

Figure 2. Bar plots of the mean number of steps with regard to (A) gender, (B) phase of the study, (C) EF, and (D) NYHA classification. The SDs are illustrated as error bars. Note: Only 63 of the 65 patients used their step counter in phase III. Hence, the mean numbers of steps illustrated in B is only a mean from these 63 patients. EF: ejection fraction; NYHA: New York Heart Association.



Although none of the differences in means are statistically significant, tendencies appear in the plots with regard to gender, EF, and NYHA. Between males and females, it appears that males have a higher mean number of daily steps than females, patients with an EF lower than 30% walk less than patients with an EF over 30%, and patients with a lower NYHA classification walk more than patients with a higher NYHA classification.

Correlation Between Daily Steps and Clinical Parameters

The correlation between daily steps and demographic and clinical parameters was investigated using Pearson correlation coefficient. The investigated parameters were age, gender, baseline EF, change in EF, baseline NYHA, change in NYHA, mean number of steps, and change in number of daily steps. The correlation values (r) and corresponding P values are presented in Table 3.

Table 3. Pearson correlation coefficients (r) and significance levels.

Variables	Age	Gender	Baseline EF[a]	Change in EF	Baseline NYHA[b]	Change in NYHA	Mean number of daily steps	Change in daily steps
Age	1.000	0.057	–0.105	–0.123	0.093	–0.044	–0.188	–0.112
Gender		1.000	0.033	0.052	–0.187	–0.065	0.140	0.003
Baseline EF			1.000	–0.742[c]	–0.224[d]	–0.198	0.071	–0.427[c]
Change in EF				1.000	0.004	0.096	0.120	0.314[e]
Baseline NYHA					1.000	0.581[c]	–0.288[f]	0.116
Change in NYHA						1.000	0.033	0.065
Mean number of daily steps							1.000	0.053
Change in daily steps								1.000

[a]EF: ejection fraction.

[b]NYHA: New York Heart Association.

[c]*P*<.001.

[d]*P*=.037.

[e]*P*=.005.

[f]*P*=.010.

Based on the correlation coefficients and significance levels in Table 3, a statistically significant correlation of *P*<.05 appears between the following variables: (1) the baseline EF and change in EF, (2) the baseline EF and baseline NYHA, (3) the baseline EF and change in the number of daily steps, (4) change in EF and change in the number of daily steps, (5) the baseline NYHA and change in NYHA classification, and (6) the baseline NYHA classification and mean number of daily steps.

Regression Analysis

Hierarchical regression analyses have been performed with the following purposes: (1) predicting the variation in the mean number of steps and (2) predicting the variation in the change in EF. Both analyses consisted of 3 models in which demographic data were entered in the first block, clinical data in the second block, and activity data in the third block. The results from the analyses are presented in Table 4.

Table 4. Hierarchical regression analyses to predict the mean number of steps, depicted as model statistics and coefficients.

Model and variables	Prediction of variation in the mean number of steps							
	Test statistics				Coefficients			
	R^2	ΔR^2	ΔF (*df*)	*P* value	β	*t* (*df*)	*P* value	pr^2
1	**0.058**	**0.058**	**1.910 (2, 62)**	**.157**				
Age					–0.197	–1.592 (2,62)	.116	0.039
Gender					0.151	1.223 (2,62)	.226	0.024
2	**0.119**	**0.060**	**2.080 (2, 60)**	**.134**				
Age					–0.171	–1.395 (4,60)	.168	0.031
Gender					0.102	0.827 (4,60)	.412	0.011
Baseline EF[a]					–0.008	–0.061 (4,60)	.952	<0.001
Baseline NYHA[b]					–0.255	–2.005 (4,60)	.049[c]	0.063
3	**0.810**	**0.691**	**215.132 (1, 59)**	**<.001[c]**				
Age					–0.089	–1.553 (5,59)	.126	0.039
Gender					0.021	0.362 (5,59)	.719	0.002
Baseline EF					–0.168	–2.832 (5,59)	.006[c]	0.120
Baseline NYHA					0.003	0.052 (5,59)	.959	<0.001
Baseline steps					0.911	14.667 (5,59)	<.001[c]	0.785

[a]EF: ejection fraction.

[b]NYHA: New York Heart Association.

[c]*P*<.05.

The squared correlations (R^2) shown in Table 4 indicate the proportion of variation accounted for when using the variables listed under the model numbers, which in these cases increases with the complexity of the model. The result in model 3 prove that 81% of the variation in the mean steps can be predicted. The correlation for model 3 is highly influenced by the baseline EF and baseline steps variables, as these were the statistically significant predictors (*P*=.006 and *P*<.001, respectively), based on the *P* values.

Discussion

Principal Findings

The aim of this FPT substudy was to investigate the walking activity of patients with HF participating in a telerehabilitation program for 1 year as measured by step counters. There was no statistically significant difference (*P*>.05) between the mean number of daily steps in the 3 phases of the intervention. This may be due to the step counters not being designed to capture different modes of physical activity (eg, cycling and swimming). As a result, not all physical activities may have been documented.

This study has shown a significant correlation between the mean number of daily steps and NYHA classification (*P*=.01), between the increase in daily steps and EF (*P*<.001), and between the increase in daily steps and reduction in EF (*P*=.005). These correlations indicate a relationship between a higher level of physical activity and an improvement in the HF condition. Furthermore, the findings indicate that demographic, clinical, and activity data can be used to predict 81% of the variation in the mean number of daily steps.

A study by Albert et al [20] investigating the requirements from patients with HF regarding devices for monitoring their health and activity found that the patients with HF requested devices that could give them immediate feedback and an overview of data over time [20]. In our study, we used 2 Fitbit step counters, both of which fulfilled these requests. Hence, they gave the patients immediate feedback through the device and an overview of their own data using the HeartPortal [17]. These commercially available devices could therefore be used to encourage the patients to be more active in their daily lives. During this period, the patients participated in a mean total duration of 358.18 (SD 54.57) days, and they used the step counters for mean 88.65% (SD 19.75%) of this time (they were asked to use the Fitbits every day). However, the percentage of time using the step counter differed between the predefined activity levels because patients with a low activity level used the step counter for 81.00% (SD 20.56%) of the participation period, whereas those patients with a high activity level used the step counter for 94.40% (SD 16.94%) of the participation period. These results seem to indicate that a higher adherence is associated with a higher activity level. These findings are consistent with those reported by Thorup et al [3], wherein 72% of the low-activity level cardiac patients walked a minimum of 100 steps with a Fitbit step counter, compared to 88% of those with a medium activity level and 91% of those with a high activity level.

The patients in the FPT walked a mean number of 6962.81 steps daily, which is approximately the same as the mean number of 7027 daily steps reported by Bäck et al [15], whose study also included patients with coronary artery disease. However, the mean number of steps in this study is higher than the mean number of 5889 and 5869 steps reported by Thorup et al [3] and Werhahn et al [12], who either included different cardiac patients (including patients with HF) or only patients with HF.

The hierarchical regression analyses demonstrated that it is possible to predict 81% of the variation in the mean daily steps. However, only the baseline EF (P=.006) and the baseline number of daily steps (P<.001) were statistically significant predictors when predicting variation in the mean number of daily steps.

The relations between clinical variables and daily steps presented in this study can be of value in clinical practice first for the patients. Besides, in collaboration with health care professionals, these data can help facilitate the rehabilitation of patients with HF. This is in alignment with a previous study conducted by the Laboratory for Welfare Technology, which reported that the use of step counters motivated cardiac patients to do more physical activity and made the physical activity visible for the patient [21]. A qualitative study by Andersen et al [22] showed that activity data from wearable devices used by cardiac patients may be a tool for self-care.

To our knowledge, no other studies have included patients with HF and measured and analyzed step counts over a 1-year period. We believe that this study offers a picture of how an activity tracker can be used to document the change in the physical activity over time for patients with HF in their daily lives.

Limitations

Some limitations of this study should be considered. The echocardiograph was not blinded for other clinical information, which may have led to biases. Two different kinds of step counters were used due to patient preferences. However, the 2 types of step counters used in the FPT were among those models with the lowest error margin, as described in Munck et al [18], and are therefore considered valid.

Conclusions

The patients in this study who walked more tended to be of a younger age, had lower NYHA classification, and a higher EF. There was a statistically significant correlation between the mean number of daily steps and the NYHA classification at baseline (P=.01), between the increase in daily steps and EF at baseline (P<.001), and between the increase in daily steps and improvement in EF (P=.005). The patients' demographic, clinical, and activity data can be used to predict 81% of the variation in the mean number of daily steps. These results suggest that a step counter may be a useful tool for patients in helping them to monitor their own physical activity during a telerehabilitation program and a means to help enhance their recovery.

Acknowledgments

The authors would like to acknowledge the Aage and Johanne Louis-Hansen's Foundation for making this study possible. Furthermore, the authors would like to thank all partners in the project: Healthcare centers in Viborg, Skive, Randers, and Silkeborg; the cardiology wards at the regional hospitals in Viborg, Skive, Randers, and Silkeborg; the Danish Heart Association; Technical University of Denmark; Department of Psychology and Behavioral Sciences at Aarhus University; Department of Computer Science at Aalborg University; CIMT at Odense University Hospital; and Laboratory for Welfare Technologies at Aalborg University. The Future Patient project is financed by Aage and Johanne Louis-Hansen's Foundation, Aalborg University, Viewcare, and partners in the project.

Authors' Contributions

The study was designed by BD, HS, MH, and JR. JG and BD have drafted the manuscript. Feedback for the manuscript was provided by all authors. All authors approved the final manuscript before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1088 KB - biomedeng_v5i1e20776_app1.pdf]

References

1. Savarese G, Lund L. Global Public Health Burden of Heart Failure. Card Fail Rev 2017 Apr;3(1):7-11 [FREE Full text] [doi: 10.15420/cfr.2016:25:2] [Medline: 28785469]
2. Cook C, Cole G, Asaria P, Jabbour R, Francis DP. The annual global economic burden of heart failure. Int J Cardiol 2014 Feb 15;171(3):368-376. [doi: 10.1016/j.ijcard.2013.12.028] [Medline: 24398230]

3. Thorup C, Hansen J, Grønkjær M, Andreasen JJ, Nielsen G, Sørensen EE, et al. Cardiac Patients' Walking Activity Determined by a Step Counter in Cardiac Telerehabilitation: Data From the Intervention Arm of a Randomized Controlled Trial. J Med Internet Res 2016 Apr 04;18(4):e69 [FREE Full text] [doi: 10.2196/jmir.5191] [Medline: 27044310]
4. Heidenreich PA, Trogdon JG, Khavjou OA, Butler J, Dracup K, Ezekowitz MD, American Heart Association Advocacy Coordinating Committee, Stroke Council, Council on Cardiovascular RadiologyIntervention, Council on Clinical Cardiology, Council on EpidemiologyPrevention, Council on Arteriosclerosis, ThrombosisVascular Biology, Council on Cardiopulmonary, Critical Care, PerioperativeResuscitation, Council on Cardiovascular Nursing, Council on the Kidney in Cardiovascular Disease, Council on Cardiovascular SurgeryAnesthesia,Interdisciplinary Council on Quality of CareOutcomes Research. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. Circulation 2011 Mar 01;123(8):933-944. [doi: 10.1161/CIR.0b013e31820a55f5] [Medline: 21262990]
5. Gensini G, Alderighi C, Rasoini R, Mazzanti M, Casolo G. Value of Telemonitoring and Telemedicine in Heart Failure Management. Card Fail Rev 2017 Nov;3(2):116-121 [FREE Full text] [doi: 10.15420/cfr.2017:6:2] [Medline: 29387464]
6. Ponikowski P, A. Voors A, D. Anker S, Bueno H, G. F. Cleland J, J. S . Coats A, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Russ J Cardiol 2017 Jan 01;37(1):7-81. [doi: 10.15829/1560-4071-2017-1-7-81]
7. Houle J, Doyon O, Vadeboncoeur N, Turbide G, Diaz A, Poirier P. Innovative program to increase physical activity following an acute coronary syndrome: randomized controlled trial. Patient Educ Couns 2011 Dec;85(3):e237-e244. [doi: 10.1016/j.pec.2011.03.018] [Medline: 21546203]
8. Houle J, Valera B, Gaudet-Savard T, Auclair A, Poirier P. Daily steps threshold to improve cardiovascular disease risk factors during the year after an acute coronary syndrome. J Cardiopulm Rehabil Prev 2013;33(6):406-410. [doi: 10.1097/HCR.0000000000000021] [Medline: 24104407]
9. Louis AA, Turner T, Gretton M, Baksh A, Cleland JG. A systematic review of telemonitoring for the management of heart failure. Eur J Heart Fail 2003 Oct;5(5):583-590. [doi: 10.1016/s1388-9842(03)00160-0] [Medline: 14607195]
10. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, Better Effectiveness After Transition–Heart Failure (BEAT-HF) Research Group. Effectiveness of Remote Patient Monitoring After Discharge of Hospitalized Patients With Heart Failure: The Better Effectiveness After Transition -- Heart Failure (BEAT-HF) Randomized Clinical Trial. JAMA Intern Med 2016 Mar 01;176(3):310-318 [FREE Full text] [doi: 10.1001/jamainternmed.2015.7712] [Medline: 26857383]
11. Dinesen B, Dittmann L, Gade JD, Jørgensen CK, Hollingdal M, Leth S, et al. "Future Patient" Telerehabilitation for Patients With Heart Failure: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Sep 19;8(9):e14517 [FREE Full text] [doi: 10.2196/14517] [Medline: 31538944]
12. Werhahn SM, Dathe H, Rottmann T, Franke T, Vahdat D, Hasenfuß G, et al. Designing meaningful outcome parameters using mobile technology: a new mobile application for telemonitoring of patients with heart failure. ESC Heart Fail 2019 Jun 13;6(3):516-525 [FREE Full text] [doi: 10.1002/ehf2.12425] [Medline: 30868756]
13. Brennan D, Tindall L, Theodoros D, Brown J, Campbell M, Christiana D, et al. A blueprint for telerehabilitation guidelines. Int J Telerehabil 2010 Oct 27;2(2):31-34 [FREE Full text] [doi: 10.5195/ijt.2010.6063] [Medline: 25945175]
14. Savage PD, Ades PA. Pedometer step counts predict cardiac risk factors at entry to cardiac rehabilitation. J Cardiopulm Rehabil Prev 2008;28(6):370-7; quiz 378. [doi: 10.1097/HCR.0b013e31818c3b6d] [Medline: 19008690]
15. Bäck M, Cider Å, Gillström J, Herlitz J. Physical activity in relation to cardiac risk markers in secondary prevention of coronary artery disease. Int J Cardiol 2013 Sep 20;168(1):478-483. [doi: 10.1016/j.ijcard.2012.09.117] [Medline: 23041099]
16. American Heart Association. Classes of Heart Failure Internet. Chicago, IL: American Heart Association; 2020. URL: https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure/classes-of-heart-failure [accessed 2020-11-15]
17. Joensson K, Melholt C, Hansen J, Leth S, Spindler H, Olsen MV, et al. Listening to the patients: using participatory design in the development of a cardiac telerehabilitation web portal. Mhealth 2019;5:33. [doi: 10.21037/mhealth.2019.08.06] [Medline: 31620460]
18. Munck K, Christensen M, Tahhan A, Dinesen B, Spindler H, Hansen J. Evaluation of Self-Trackers for Use in Telerehabilitation. J Usability Stud 2018;13(4):37. [doi: 10.5555/3294033.3294035]
19. Bunn J, Navalta J, Fountaine C, Reece J. Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015-2017. Int J Exerc Sci 2018;11(7):503-515 [FREE Full text] [Medline: 29541338]
20. Albert NM, Dinesen B, Spindler H, Southard J, Bena JF, Catz S, et al. Factors associated with telemonitoring use among patients with chronic heart failure. J Telemed Telecare 2016 Jul 08;23(2):283-291. [doi: 10.1177/1357633x16630444]
21. Thorup CB, Grønkjær M, Spindler H, Andreasen JJ, Hansen J, Dinesen BI, et al. Pedometer use and self-determined motivation for walking in a cardiac telerehabilitation program: a qualitative study. BMC Sports Sci Med Rehabil 2016 Aug 18;8(1):24 [FREE Full text] [doi: 10.1186/s13102-016-0048-7] [Medline: 27547404]
22. Andersen TO, Langstrup H, Lomborg S. Experiences With Wearable Activity Data During Self-Care by Chronic Heart Patients: Qualitative Study. J Med Internet Res 2020 Jul 20;22(7):e15873 [FREE Full text] [doi: 10.2196/15873] [Medline: 32706663]

Abbreviations

ANOVA: analysis of variance
API: application programming interface
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Randomized Controlled Trials of Electronic and Mobile Health Applications and Online TeleHealth
EF: ejection fraction
FPT: Future Patient Telerehabilitation
HF: heart failure
NYHA: New York Heart Association

Edited by G Eysenbach; submitted 28.05.20; peer-reviewed by S Werhahn, L Almeida; comments to author 24.07.20; revised version received 24.09.20; accepted 26.10.20; published 30.11.20.

Please cite as:
Gade JD, Spindler H, Hollingdal M, Refsgaard J, Dittmann L, Frost L, Mahboubi K, Dinesen B
Predictors of Walking Activity in Patients With Systolic Heart Failure Equipped With a Step Counter: Randomized Controlled Trial
JMIR Biomed Eng 2020;5(1):e20776
URL: http://biomedeng.jmir.org/2020/1/e20776/
doi:10.2196/20776
PMID:

Original Paper

Personalized Monitoring Model for Electrocardiogram Signals: Diagnostic Accuracy Study

Rado Kotorov[1], PhD; Lianhua Chi[2], PhD; Min Shen[1], MSc

[1]Trendalyze Inc, Newark, NJ, United States

[2]La Trobe University, Bundoora, Australia

Corresponding Author:
Lianhua Chi, PhD
La Trobe University
Thomas Cherry Bldg, 3rd Fl
La Trobe University
Bundoora, 3086
Australia
Phone: 61 3 94792454
Email: l.chi@latrobe.edu.au

Abstract

Background: Due to the COVID-19 pandemic, the demand for remote electrocardiogram (ECG) monitoring has increased drastically in an attempt to prevent the spread of the virus and keep vulnerable individuals with less severe cases out of hospitals. Enabling clinicians to set up remote patient ECG monitoring easily and determining how to classify the ECG signals accurately so relevant alerts are sent in a timely fashion is an urgent problem to be addressed for remote patient monitoring (RPM) to be adopted widely. Hence, a new technique is required to enable routine and widespread use of RPM, as is needed due to COVID-19.

Objective: The primary aim of this research is to create a robust and easy-to-use solution for personalized ECG monitoring in real-world settings that is precise, easily configurable, and understandable by clinicians.

Methods: In this paper, we propose a Personalized Monitoring Model (PMM) for ECG data based on motif discovery. Motif discovery finds meaningful or frequently recurring patterns in patient ECG readings. The main strategy is to use motif discovery to extract a small sample of personalized motifs for each individual patient and then use these motifs to predict abnormalities in real-time readings of that patient using an artificial logical network configured by a physician.

Results: Our approach was tested on 30 minutes of ECG readings from 32 patients. The average diagnostic accuracy of the PMM was always above 90% and reached 100% for some parameters, compared to 80% accuracy for the Generalized Monitoring Models (GMM). Regardless of parameter settings, PMM training models were generated within 3-4 minutes, compared to 1 hour (or longer, with increasing amounts of training data) for the GMM.

Conclusions: Our proposed PMM almost eliminates many of the training and small sample issues associated with GMMs. It also addresses accuracy and computational cost issues of the GMM, caused by the uniqueness of heartbeats and training issues. In addition, it addresses the fact that doctors and nurses typically do not have data science training and the skills needed to configure, understand, and even trust existing black box machine learning models.

(JMIR Biomed Eng 2020;5(1):e24388) doi:10.2196/24388

KEYWORDS

COVID-19; personalized monitoring model; ECG; time series; motif discovery; monitoring; heart disease; electrocardiogram

Introduction

Background

An electrocardiogram (ECG) is a medical test that records the electrical activities of the heart. It is widely used by medical practitioners for diagnosing cardiac conditions by detecting irregular heart rhythms and abnormalities [1]. In some cases, arrhythmic heartbeats can be lethal and the risk of sudden death is significant without remote patient monitoring (RPM) [2]. Therefore, it is highly desirable for patients to have an efficient ECG remote monitoring system that can identify life-threatening situations and send alerts to their health care providers [3]. Lately, the demand for remote ECG monitoring has increased drastically because of the COVID-19 pandemic. To prevent the

spread of the virus and keep individuals with less severe cases out of hospitals, more patients are having heart disease diagnosed and monitored remotely, while at home. The accuracy of the ECG signal classifier is becoming more important because false alarms can overwhelm the system. Therefore, classifying the ECG signals accurately and sending alerts to health care professionals in a timely fashion are urgent problems that need to be addressed.

The classification of ECG signals is an extremely challenging problem as there are no defined optimal classification rules. Many researchers have focused on developing different machine learning models, such as Bayesian framework [4], random forest [5], gradient boosting [6], ensemble boosting [7], and support vector machine [8], among others, and achieved relatively high accuracy. To extract different features for models, various techniques were proposed, such as principal component analysis [9-12], wavelet transform [13], and filter banks [14]. Deep learning methods such as deep neural networks [15], convolutional neural networks [16-20], and recurrent neural networks [21] are also applied extensively to classification problems. Deep learning can be a powerful tool for solving cognitive problems [22], but accurately training such models requires large amounts of labeled data [23]. For clinical applications, due to limited patient contact, variation in medical care, and privacy issues, getting a large amount of high-quality data can be very challenging [24], and the efficacy of deep learning methods can be greatly affected by the lack of training data [25-29]. According to Chen et al's [30] investigation of the training time of deep learning and machine learning methods, deep learning requires a longer training time compared to conventional machine learning algorithms. In addition, building and maintaining the computational infrastructure required for deep learning can be too costly for small health care organizations to implement. Thus, a less computationally expensive method is needed to effectively resolve the issue.

Another limitation of deep learning methods is that the models are not able to capture the individuality of the ECG features and patterns [31]. Most of the deep learning models are generalized models and are not able to be built to the individual level due to a lack of data [26]. However, each patient has unique heartbeats and the waveforms can be completely different on an individual level. Hence, accuracy might be an issue for these models when using real-time data. Traditional machine learning methods usually require more effort related to data preprocessing and feature engineering compared to deep learning models [29]. In addition, they tend to be like black boxes to medical practitioners without a data science background. The lack of interpretability can hinder health care providers' decision making process and communication with patients.

In this paper, we propose a Personalized Monitoring Model (PMM) for ECG monitoring based on motif discovery to address the abovementioned challenges. Motif discovery is a method for analyzing large amounts of time series data. In the health care domain, it has been used for trend analysis and data summarization [32]. Motifs are defined as frequently recurring patterns in certain time series [33]. In a motif discovery process, a similarity search is conducted based on a certain similarity threshold to detect and locate previously defined patterns. In a similarity search, the distances between time series subsequences are calculated, which indicates how similar two subsequences are.

Objectives

The primary aim of this research is to create a robust and easy-to-configure solution for monitoring ECG signals in real-world settings. We developed a technique for building personalized prediction models to address the limitations of generalized models [31]. The main strategy of the model is to extract personalized motifs for each patient and use the motifs to predict the rest of the readings of that patient using an artificial logical network. By performing a systematic analysis and evaluation, we will investigate the hypothesis that the proposed PMM is more accurate and efficient than generalized models. In most cases, doctors and nurses do not have a data science background and the existing machine learning models might be difficult to configure. Hence, a new technique will be required as RPM becomes more common, as has occurred due to the COVID-19 pandemic. The main goal is to develop a technique that allows doctors, nurses, and other medical practitioners to easily configure a personalized model for RPM. The proposed model can be easily understood and configured by medical practitioners, since it requires less training data and fewer parameters to configure.

Methods

In this section, we discuss the proposed PMM for ECG data in detail. The process includes time series sampling, personalized motif discovery, and motif-based prediction using an artificial logical network.

Time Series Sampling

We treated each patient's ECG measures as individual data; the recording is a time series. An ECG is a medical test that detects heart problems by measuring the electrical activity generated by the heart as it contracts. An ECG complex is composed of different components, or waves, that represent the electrical activity in specific regions of the heart. ECG readings from healthy hearts have a characteristic shape. If the ECG reading is a different shape, that could suggest a heart problem.

During this stage, there is an important parameter, t, which represents the training ratio ($0<t<1$) and affects the sampling process. Before starting to discover motifs, each individual patient's ECG time series was sampled as individual training data. L represents the length of each patient's ECG time series. Based on the value of the training ratio t, we used Equation 1 to calculate the length of time series S we should take from the whole ECG reading of each patient:

$$S = L \times t \quad \textbf{(1)}$$

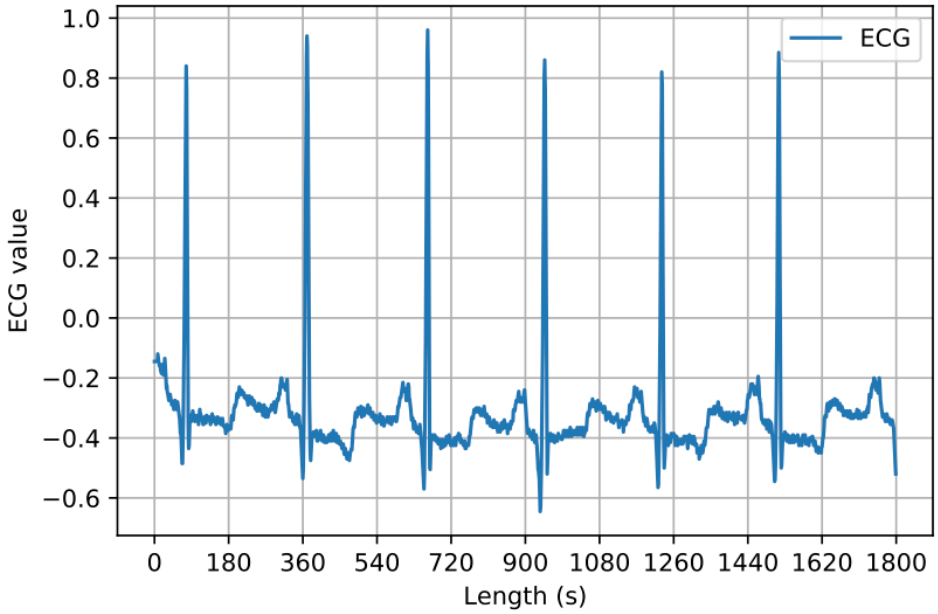
As calculated, we took the first S length of time series from the ECG readings of each individual patient as the individual training sample data to generate the personalized motifs. During sampling, we started from the first point of the patient's ECG time series and sampled until the length of sample reached the expected sample size S. We then divided the sampled ECG data

into *M* subsequences based on Equation 2 and each subsequence was regarded as a pattern unit.

$$M = S/180 \quad \textbf{(2)}$$

In Equation 2, the number 180 is the sampling rate of the ECG recording device. It partitions the ECG into heartbeats with sufficient precision of intervals for heart rate variability analysis [34]. Figure 1 shows an ECG sample with 1800 points from one of the patients.

Figure 1. ECG sample. ECG: electrocardiogram.

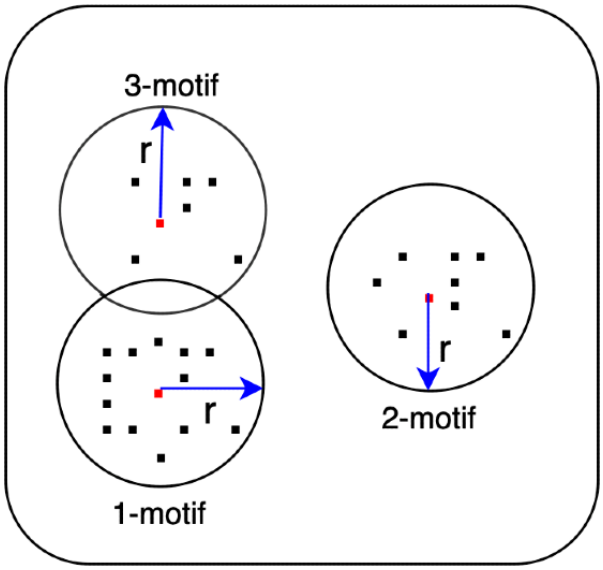


Personalized Motif Discovery

After sampling, we discovered personalized motifs from the *M* subsequences sampled from an individual patient. For this model design, we needed to consider two major parameters that could affect the performance of motif discovery: (1) *r*, the time series similarity threshold and (2) *k*, the number of motifs.

We calculated all Euclidean distances between each subsequence and generated motif candidates. In each motif circle/cluster, the distances from the central subsequence to other subsequences that belong to the same motif circle must be less than *r* and all motif circles cannot share the same subsequence, as proposed by a previous paper [35] and as shown in Figure 1. In Figure 2, each black dot represents each subsequence and each red dot represents the central subsequence in that motif circle. We supposed *k*=3, as we can see 1-motif has the most subsequences, 2-motif has the second most, and 3-motif has the fewest subsequences. All motif circles have the same radius *r*. The distance between the red central subsequence of 1-motif and the red central subsequence of 2-motif is more than *2r*, and these naturally do not share the same subsequence. However, the distance between the central subsequence of 1-motif and the central subsequence of 3-motif is less than *2r*, but they do not share the same subsequence, which is allowable during motif discovery. In this example, the 3 red central subsequences are the 3 extracted motifs.

Figure 2. Subsequence motifs (k=3).

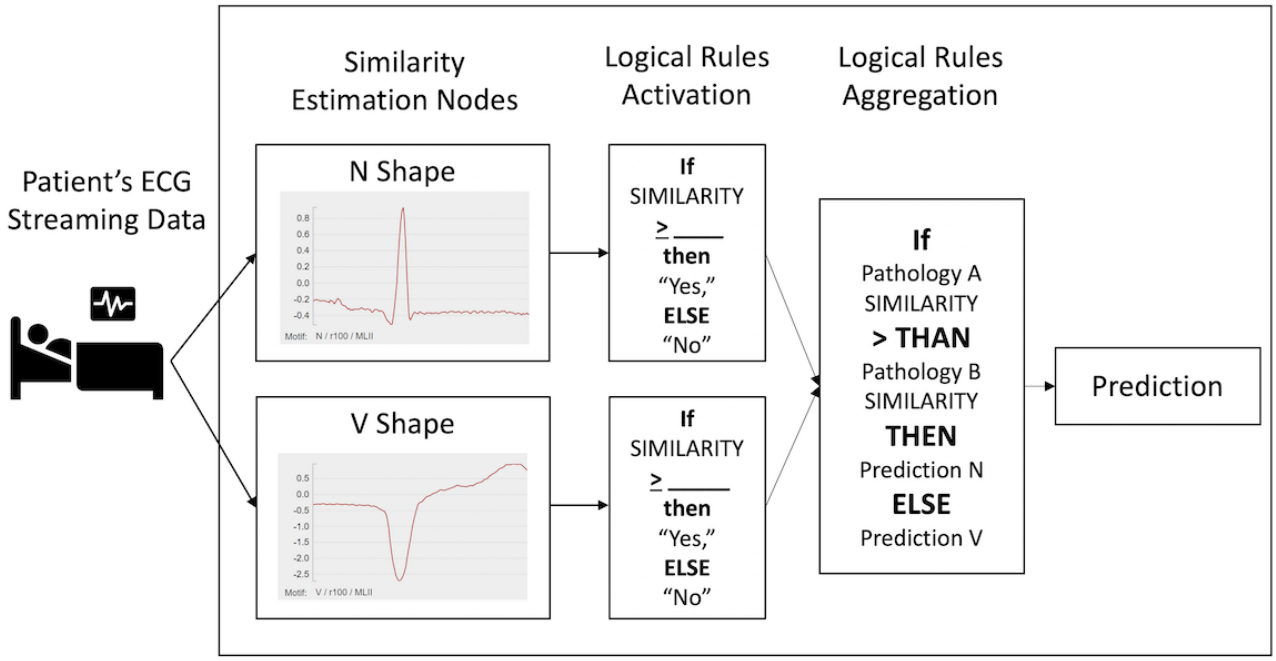


Based on this strategy of motif discovery, we generated all motif circle candidates from the *M* subsequences. Based on the parameter *k*, which is the number of motifs, we only kept the first *k* motif circles and used the central subsequences of all these *k* motif circles as our extracted heartbeat motifs. From this stage, we obtained *k* personalized motifs for each individual patient.

Motif-Based Prediction Using an Artificial Logical Network

We used the generated *k* personalized motifs to predict the rest of the ECG readings for each corresponding patient. For example, suppose we have two types of heartbeats (N and V) and need to predict which type (N or V) the test subsequence belongs to. Following the personalized motif discovery, we generated *k* of N motifs and *k* of V motifs from the sampled subsequences of each individual patient. The generated N and V motifs were organized into an artificial logical network where each N and V motif is a dedicated evaluation node, as shown in Figure 3. For the remaining subsequences of that patient, we then tested each subsequence by comparing its distance to all N motifs and V motifs in the artificial logical network. A simple logical rule was applied to select the closest one as the predicted type of test subsequence. Finally, we predicted all labels for the rest of the subsequences of that patient. If the subsequences did not meet the matching criteria of any of the N and V nodes, the logical rule identified them as new anomalies for future learning. The combination of dedicated motif comparison and logical rules allows us to easily build a prediction system.

Figure 3. Motif-based prediction using an artificial logical network.



Results

Benchmark Data

For this benchmark data, we used 32 patients' ECG measures; each measure contained 30 minutes of ECG readings from each patient. In the data set, there were 5 different types of heartbeats (V, N, A, F, and S). The V pathology is expected to be morphologically different than the normal N. To detect the A pathology, we needed to monitor the frequency of the heartbeats and identify the heartbeats that appeared faster than expected. The F pathology is also expected to have a different morphology than the normal N. The S pathology is related to heart rhythm abnormalities that may not drastically change the morphology, but its occurrence is out of rhythm. Before training and testing the models, we removed the "noisy" heartbeats and only kept N heartbeats and V heartbeats as the two labels the models would identify.

Baseline Models

We evaluated and compared the following 3 models:

1. Generalized Monitoring Model 1 (GMM1): Based on the training ratio *t*, we took the first *t* percent samples from each of 32 patients and combined the samples from those 32 patients together to extract the *k* of N motifs and *k* of V motifs as the N and V heartbeat motifs. During the testing stage, we applied the extracted N motifs and V motifs to the rest of the ECG readings for each patient and predicted the label (N or V).
2. Generalized Monitoring Model 2 (GMM2): We first extracted all N heartbeats and V heartbeats from all 32 patients. Based on the training ratio *t*, we then randomly extracted *t* percent samples from all N heartbeats and *t* percent samples from all V heartbeats. Next, we extracted the *k* of N motifs from the N heartbeat samples and the *k* of V motifs from the V heartbeat samples. The testing was the same as for GMM1. The main difference between GMM1 and GMM2 is how the training data were sampled.
3. Our proposed Personalized Monitoring Model (PMM): In this personalized model, based on the training ratio *t*, we only extracted the first *t* percent of N heartbeats and the first *t* percent of V heartbeats from an individual patient

and then generated a set of personalized *k* of N motifs and *k* of V motifs to test the rest of the ECG readings of that patient. The main strategy here was to individually extract personalized motifs for the current patient and use those extracted motifs to predict the rest of readings for that patient.

There are 3 major parameters that could affect the models' performance, which are compared in detail in this section: (1) *r*, the time series similarity threshold, (2) *k*, the number of motifs, and (3) *t*, the training ratio.

To compare the models and determine which model is the best, we evaluated them based on the following 2 factors:

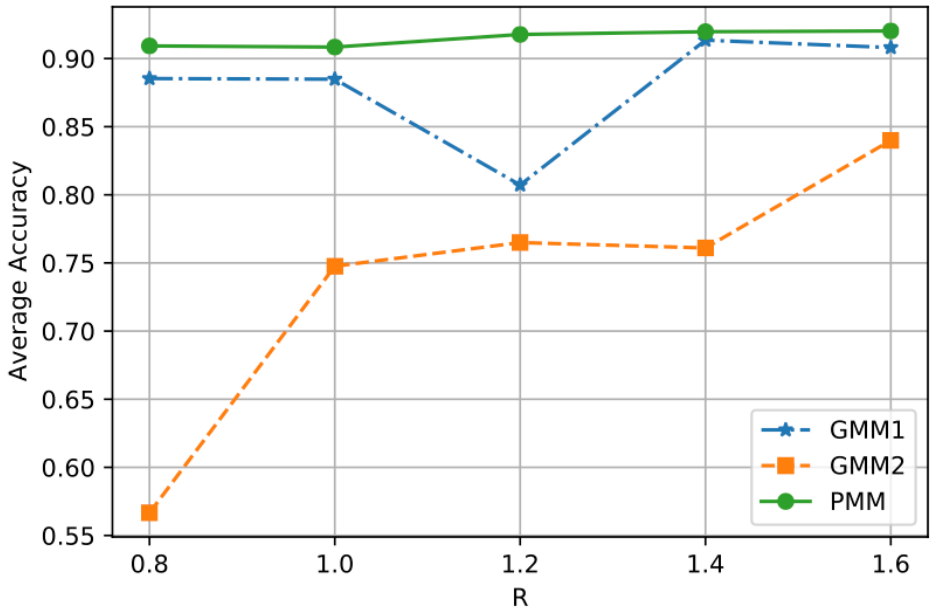
1. Accuracy: We needed to get an estimate of how accurate each model is on unseen/test data. For all tested heartbeats, we have the corresponding ground-truth information, which is the original label. By comparing the predicted label with the original label, we can calculate how many heartbeats were correctly predicted. Therefore, accuracy is calculated by taking the number of heartbeats predicted correctly and dividing it by the number of all heartbeats tested.
2. Running time: To obtain the final prediction results faster, we also needed to guarantee the chosen model was the fastest, including training and testing time.

Effectiveness Evaluation

Considering the 3 main parameters *r*, *k*, and *t*, we designed the effectiveness evaluation by adjusting the values of these 3 parameters and seeing how the performance of each model changed. In figures, we used the corresponding capital letters *R*, *K*, and *T* of *r*, *k*, and *t* for clear representation.

First, we adjusted the similarity threshold *R* values from 0.8 to 1.6, by steps of 0.2, and observed how the average performance changed for each model. The average accuracy was calculated based on the sum accuracy across all 32 patients. Figure 4 shows 3 curves, each representing the average accuracy change of each model when *R* was increased from 0.8 to 1.6. The green line represents the PMM, which performed the best among the 3 models in terms of stability and accuracy.

Figure 4. Performance comparison with different R (similarity thresholds).



Second, we adjusted the *K* values from 2 to 10, by steps of 2, and observed how average performance changed for each model. The PMM still performed the best, even as the *K* varied (Figure 5).

Lastly, we adjusted the training ratio *T* from 0.1 to 0.25, by steps of 0.05. Generally, more training samples result in more accurate prediction; although this is not what we observed with GMM1 and GMM2, it did apply to the PMM (Figure 6). From the curve comparison, we can see the PMM significantly outperformed GMM1 and GMM2.

Figure 5. Performance comparison with different K (number of motifs).

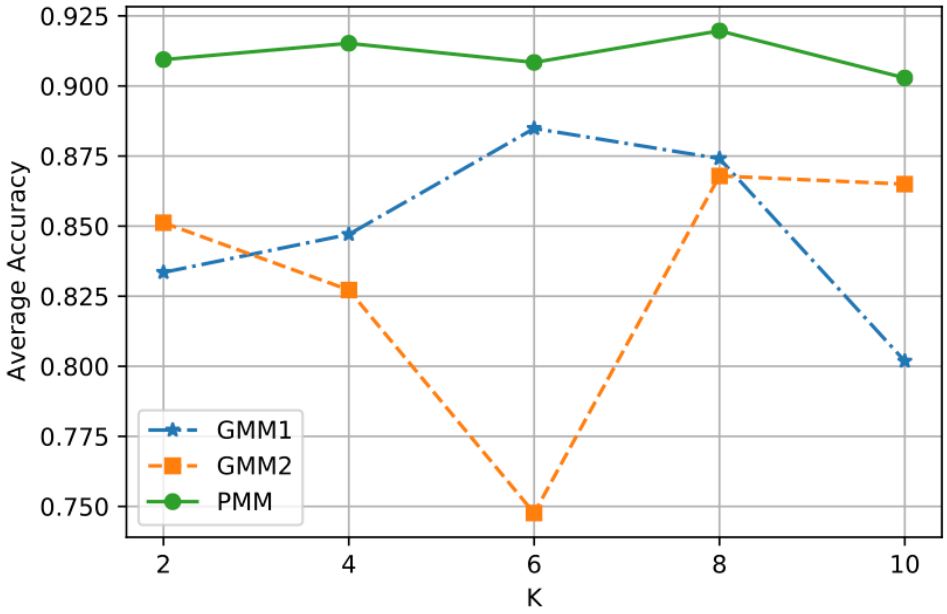
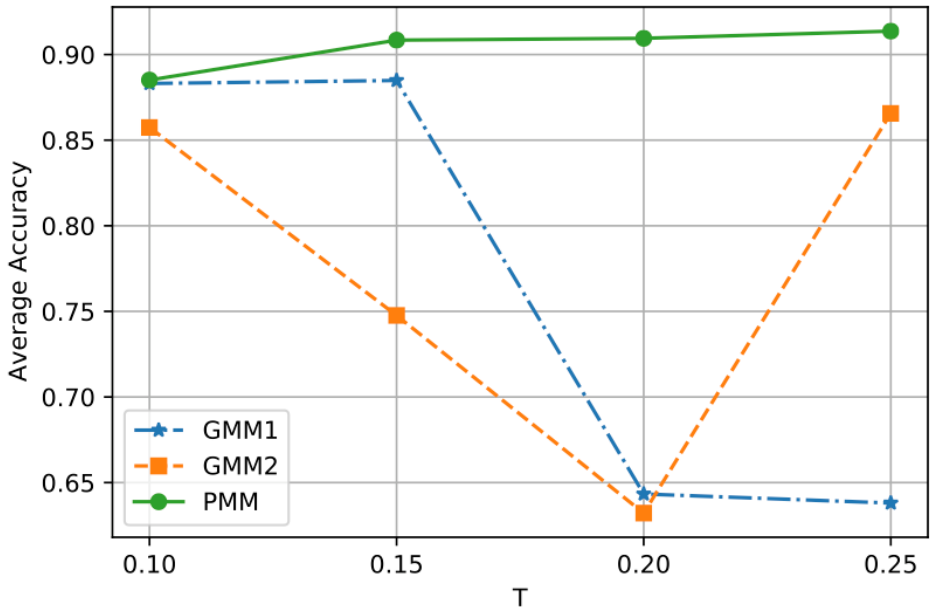


Figure 6. Performance comparison with different T (training ratios).



T is the training ratio and could affect model performance more than *R* and *K* in most cases. Hence, we showed the detailed performance result for each patient as determined by the 3 models based on two sets of values of *R* and *K*: (1) *R*=0.8 and *K*=2 and (2) *R*=1 and *K*=6.

In Figures 7 to 12, we can see GMM2 fluctuated much more than GMM1 and the PMM. However, if we compare GMM1 and the PMM when *T* was increased from 0.1 to 0.25, the PMM gradually performed better on almost all patients, while GMM1 became worse as that parameter changed. In Figures 13 and 14, the two bar charts show the average accuracy of each model with different *T*. As *T* increases, GMM1's performance becomes worse, while GMM2's performance fluctuates the most. The PMM performed the best.

Figure 7. GMM1 results when R=0.8 and K=2: performance comparison on each patient with different T. GMM1: Generalized Monitoring Model 1.

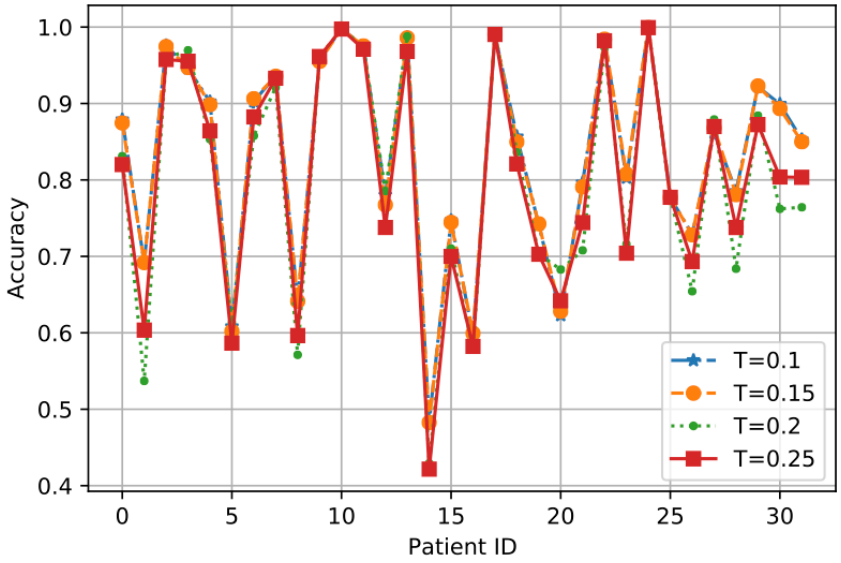


Figure 8. GMM2 results when R=0.8 and K=2: performance comparison on each patient with different T. GMM2: Generalized Monitoring Model 2.

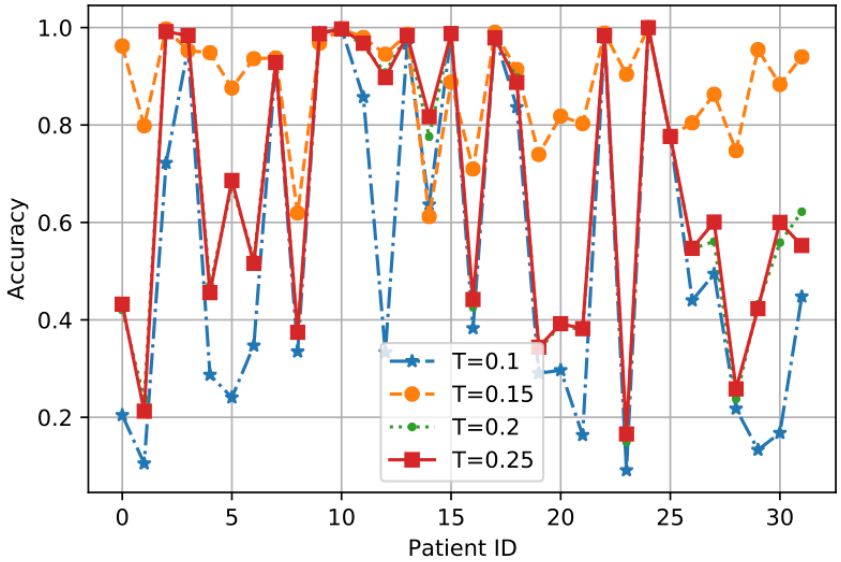


Figure 9. PMM results when R=0.8 and K=2: performance comparison on each patient with different T. PMM: Personalized Monitoring Model.

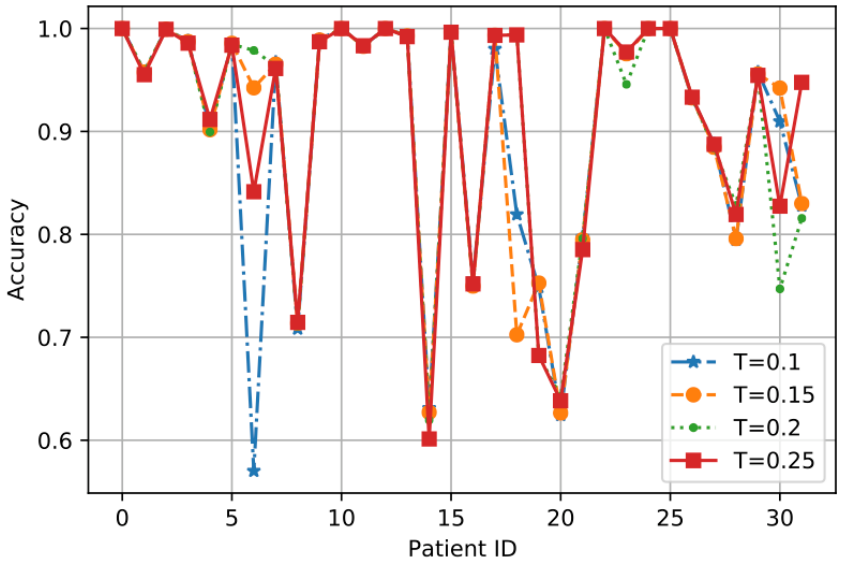


Figure 10. GMM1 results when R=1 and K=6: performance comparison on each patient with different T. GMM1: Generalized Monitoring Model 1.

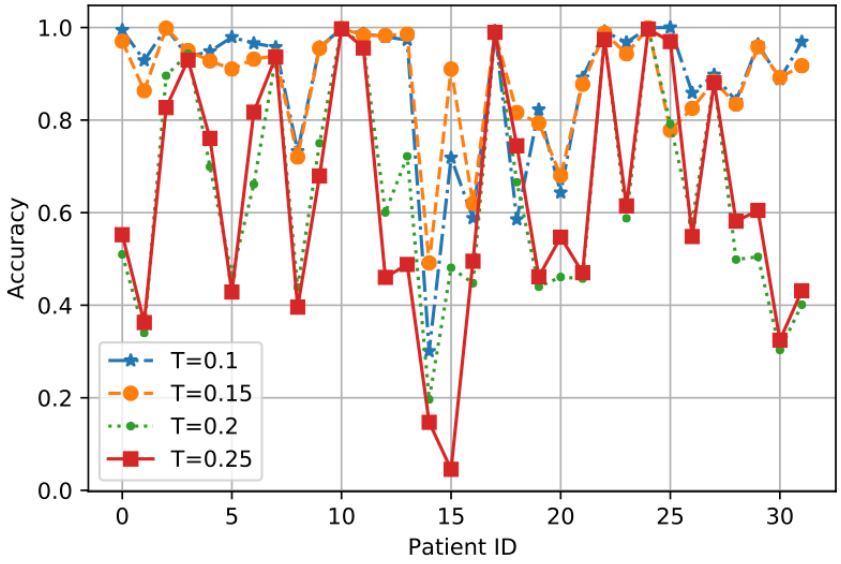


Figure 11. GMM2 results when R=1 and K=6: performance comparison on each patient with different T. GMM2: Generalized Monitoring Model 2.

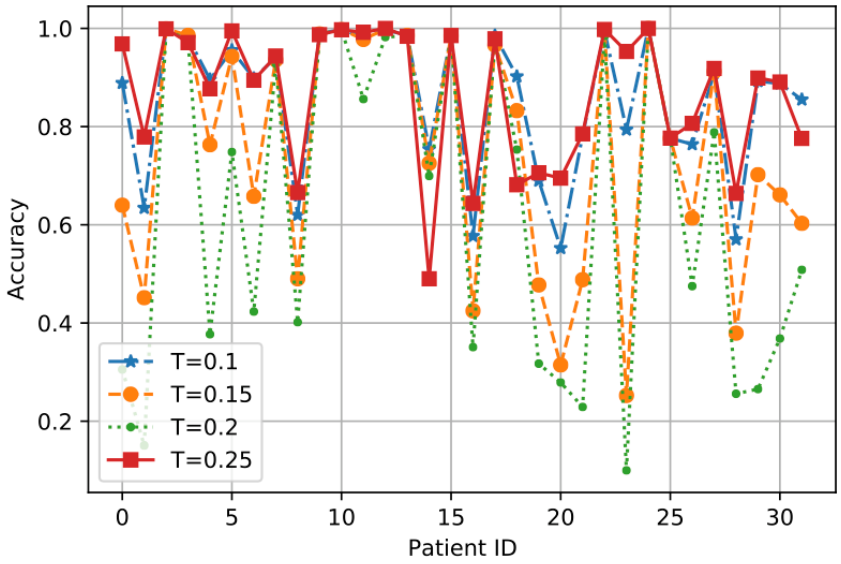


Figure 12. PMM results when R=1 and K=6: performance comparison on each patient with different T. PMM: Personalized Monitoring Model.

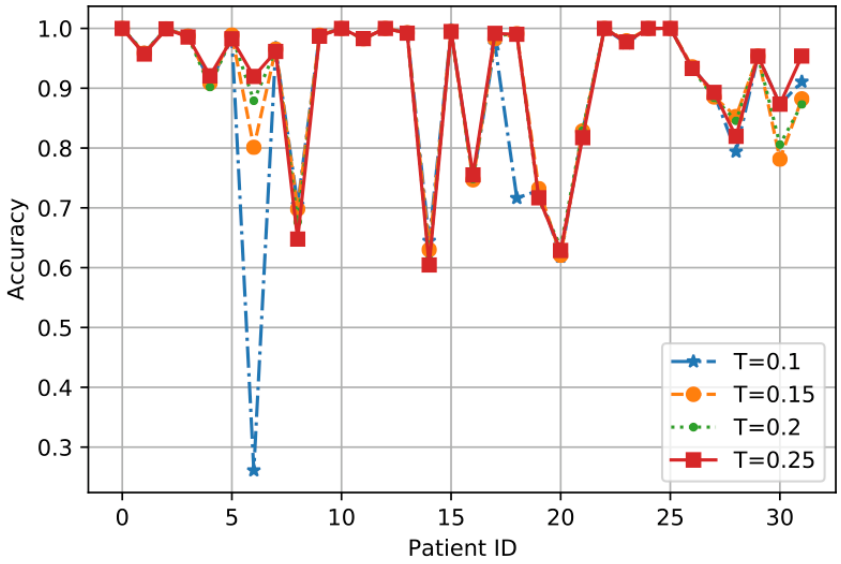


Figure 13. Results from all models when R=0.8 and K=2: average performance comparison with different T. GMM1: Generalized Monitoring Model 1; GMM2: Generalized Monitoring Model 2; PMM: Personalized Monitoring Model.

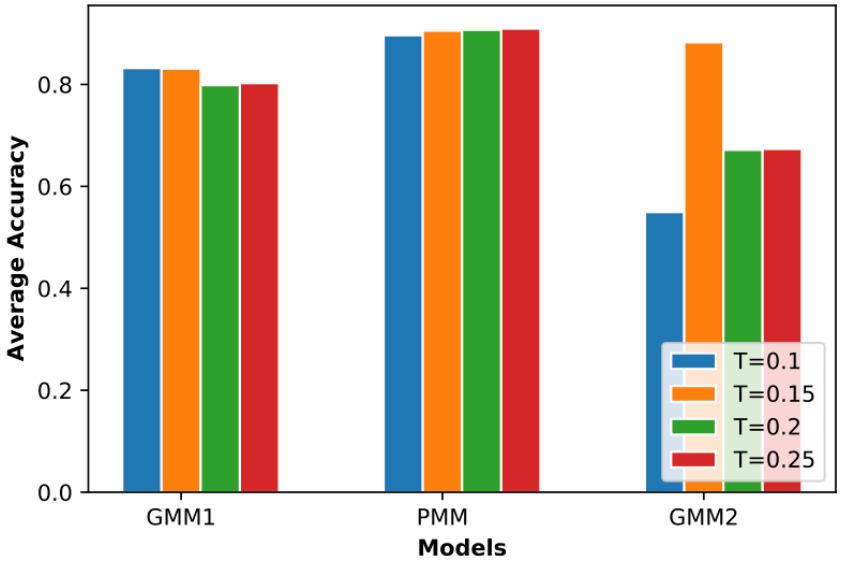
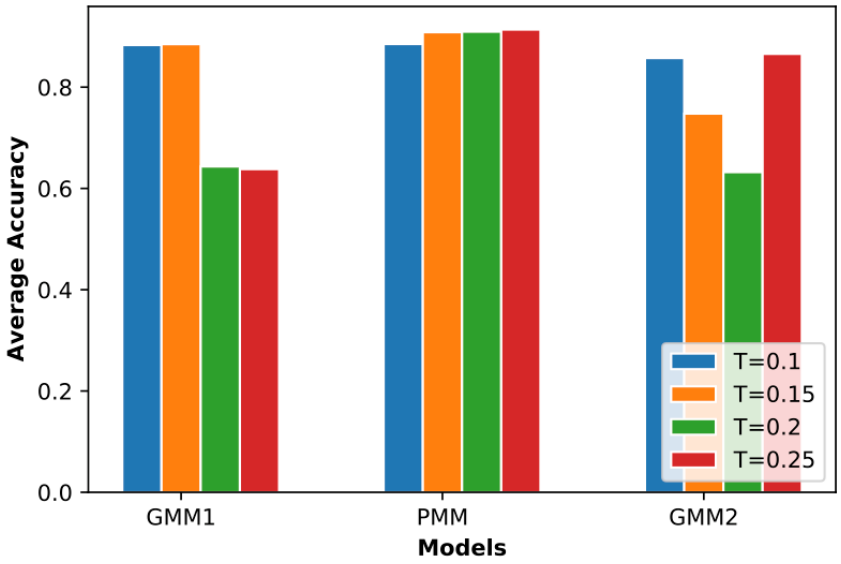


Figure 14. Results from all models when R=1 and K=6: average performance comparison with different T. GMM1: Generalized Monitoring Model 1; GMM2: Generalized Monitoring Model 2; PMM: Personalized Monitoring Model.



Efficiency Evaluation

We evaluated the time efficiency of each model and observed which model runs the fastest. The process consists of two stages: training and testing. We considered all computation time in this evaluation, including the training and testing time. From the 3 evaluated parameters in the previous section, we know that the training ratio *T* is the one that most affects training time. Here, we adjusted the value of *T* and observed the corresponding running time of each model. Figure 15 shows the linear change in running time for each model, while Figure 16 is a bar chart of the running time. The time consumption of GMM1 and GMM2 increased almost exponentially with an increase in *T*. However, the time consumption of the PMM was linear and the lowest among the 3 models.

Figure 15. Linear comparison of time efficiency across models when R=0.8 and K=2. GMM1: Generalized Monitoring Model 1; GMM2: Generalized Monitoring Model 2; PMM: Personalized Monitoring Model.

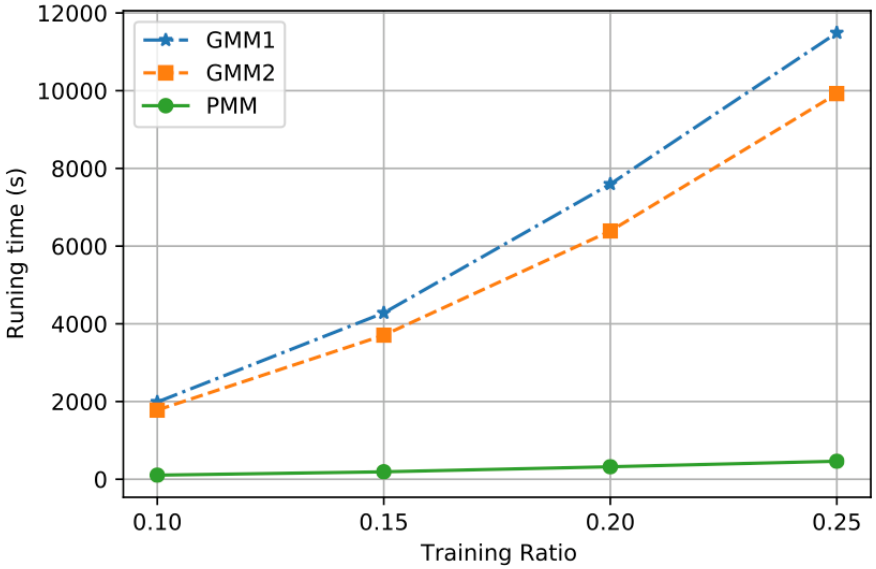
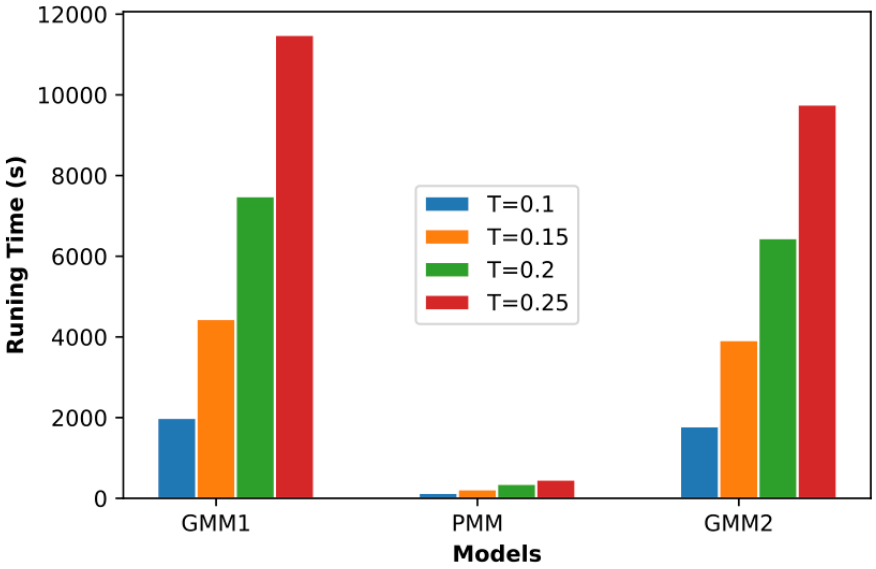


Figure 16. Time efficiency of all models when R=1 and K=6. GMM1: Generalized Monitoring Model 1; GMM2: Generalized Monitoring Model 2; PMM: Personalized Monitoring Model.



Discussion

Principal Results

According to the empirical results, the PMM performed the best in all cases in terms of prediction accuracy and time efficiency.

Figure 4 shows the average classification accuracy (y-axis) with respect to the *R* (x-axis) of all 3 models when using different similarity thresholds *R*. The PMM's performance (green line) becomes better as *R* increases from 0.8 to 1.4, and becomes stable when *R* reaches 1.6. The prediction performance of the PMM is relatively stable for similarity threshold *R*. If we compare GMM1 and GMM2, we find GMM1 outperforms GMM2. However, GMM1's performance (blue curve) was unstable with respect to *R*, and GMM2's performance (orange curve) became better as *R* increases, although the overall performance of GMM2 was worse than GMM1. This is because GMM2 extracts all N heartbeats and V heartbeats from all 32 patients for motif discovery and a greater similarity threshold allows GMM2 to aggregate more similar heartbeats for motif circles, which may help GMM2 find more representative motifs for heartbeat prediction. However, if the similarity threshold *R* is too big, it may introduce different types of heartbeats into a given motif circle, which may result in worse prediction. GMM1 takes the first *T* percent of samples from 32 patients and then extracts motifs based on the similarity distance threshold *R*. An increase in *R* may introduce more heartbeats from different patients and the extracted motifs may result in a fluctuation in accuracy. However, the PMM only extracts the first *T* percent of N samples and V samples from one patient and then generates personalized motifs to test the rest of the ECG readings of that patient. As *R* increases, more similar N heartbeats and V heartbeats will be collected for motif discovery, which could help enhance prediction if *R* is not so large that it introduces another type of heartbeat.

Figure 5 shows the average classification accuracy (y-axis) with respect to the *K* (x-axis) of all 3 models, by using different numbers of motifs *K*.

The PMM was always the best, but the GMM1 and GMM2 curves fluctuated as *K*, which represents the number of motifs, changed. A certain number of motifs can be representative and guide prediction well. However, having too many motifs may introduce unrepresentative motifs, which could hinder prediction. This is why all models fluctuated as *K* changed.

Figure 6 shows the average classification accuracy (y-axis) with respect to *T* (x-axis) on all 3 models by using different training ratios *T* from 0.1 to 0.25. As *T* increases from 0.1 to 0.25 in steps of 0.05, the PMM curve becomes higher. However, the GMM1 and GMM2 curves fluctuate a lot and are both lowest when *T* is 0.2. Generally, in machine learning, more training data should improve a model's performance. However, if more interference or noisy data is introduced into the training data, then performance will degrade. This could explain why GMM1 and GMM2 fluctuate a lot, as these models used training data from different patients and most heartbeats are unique at the patient level. As for the PMM, a larger training ratio could help improve performance because all training heartbeats are from the same patient, which avoids the introduction of interference or noisy data. As *T* changed, the PMM always outperformed GMM1 and GMM2 (Figures 7 to 14).

Based on the overall evaluation results, the PMM significantly outperformed GMM1 and GMM2 in terms of prediction accuracy.

Considering time efficiency, Figures 15 and 16 show the average running time (y-axis) with respect to the training ratio *T* (x-axis) for all 3 models by using different training ratios *T* from 0.1 to 0.25. Compared with GMM1 and GMM2, the PMM required significantly less running time. The average running time of the 3 models increased as the training ratio increased. This is because a larger training ratio results in more training data, which increases the training time accordingly. The overall average time required by GMM1 is close to that required by GMM2 and both increase almost exponentially as the training ratio increases. Therefore, the PMM has more stable and better efficiency as the training ratio increases.

Limitations

A limitation of the PMM is that it might be hard to maintain the models for each patient. However, this limitation is not significant given that retraining the model does not take a lot of time. In addition, the cost of maintenance and computation might be lower in the future as more industries adopt personalized models.

Conclusions

In this paper, we proposed a PMM for ECG recordings for two reasons: (1) the COVID-19 pandemic has accelerated the adoption of remote diagnosis and patient monitoring and (2) personalized care promises better outcomes, especially as it applies to digital health. Digital health care allows for continuous 24/7 care in the home environment while minimizing the risk of fatal accidents and re-admissions. For remote monitoring to gain traction at scale, several requirements must be met. First, monitoring has to be sufficiently automated with fewer false positive alarms to minimize the number of health professionals involved in monitoring. Second, it has to be quickly and easily configurable by the health care professionals themselves. Although traditional machine learning and deep learning approaches can be used in automation, they typically cannot be easily configured or adjusted by health care professionals due to a lack of modeling skills and data needed to build a personalized model. To solve these challenges, we employed a motif discovery algorithm to individually extract personalized motifs for each individual patient and used an artificial logical network for ECG signal prediction. We proposed a personalized model for faster ECG signal detection, which significantly improves the efficiency of ECG prediction; such a model could help satisfy the demand for remote monitoring services, especially during the COVID-19 pandemic. By comparing our proposed model, PMM, with two generalized monitoring models using real-world patient ECG data, we demonstrated that the PMM outperformed the generalized models in both prediction accuracy and time efficiency.

Per our discussions with clinicians, this approach can easily be deployed for outpatient monitoring as outlined below; this is the subject of a forthcoming clinical trial. A wearable 12-channel ECG monitor is sent to a patient or configured during a hospital stay. An augmented reality app or video conference is used to remotely guide the patient to accurately place the electrodes, while simultaneously testing the accuracy of the received signal. During setup, personalized motifs are automatically extracted, and the physician selects the center motifs to be used by the artificial logical network. The artificial logical network is a flexible structure that allows for learning when particular reference motifs are missing from the setup sample. For example, if samples of atrial fibrillation motifs are not recorded during setup, reference motifs from a general library can be used, or general anomaly detection can be applied, alerting a medical professional to review any anomalous occurrences. If the physician determines that the anomaly is atrial fibrillation, they can instantly push the motif to the artificial logical network. These configurations are the subject of the forthcoming clinical trial. By augmenting the expert's knowledge with algorithmic computational power, hospital stays can be significantly reduced, and care can be delivered in the comfort of the patient's home.

Acknowledgments

We would like to thank Dr Mazomenos from the University of Leeds for helping us prepare the data set used in this research. He annotated the heartbeats and pathologies and developed the 180-records heartbeat methodology.

Conflicts of Interest

None declared.

References

1. Mann H. A Method of Analyzing the Electrocardiogram. Arch Intern Med 1920 Mar 01;25(3):283. [doi: 10.1001/archinte.1920.00090320054004]
2. Nattel S. Sudden cardio arrest: when normal ECG variants turn lethal. Nat Med 2010 Jun;16(6):646-647. [doi: 10.1038/nm0610-646] [Medline: 20526318]
3. Guo S, Han L, Liu H, Si Q, Kong D, Guo F. The future of remote ECG monitoring systems. J Geriatr Cardiol 2016 Sep;13(6):528-530 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2016.06.015] [Medline: 27582770]
4. Sayadi O, Shamsollahi M, Clifford G. Robust Detection of Premature Ventricular Contractions Using a Wave-Based Bayesian Framework. IEEE Trans Biomed Eng 2010 Feb;57(2):353-362. [doi: 10.1109/tbme.2009.2031243]
5. Kropf M, Hayn D, Schreier G. ECG classification based on time and frequency domain features using random forests. Computing in Cardiology (CinC) 2017 Sep 24:1-4. [doi: 10.22489/cinc.2017.168-168]
6. Alarsan FI, Younes M. Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. J Big Data 2019 Aug 31;6(1):1-5. [doi: 10.1186/s40537-019-0244-x]
7. Wang H, Wu J. Boosting for Real-Time Multivariate Time Series Classification. 2017 Feb 04 Presented at: AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence; February 4-9 2017; San Francisco, CA, USA p. 4999-5000. [doi: 10.1057/978-1-137-31303-4_2]
8. Li Q, Rajagopalan C, Clifford GD. A machine learning approach to multi-level ECG signal quality classification. Comput Methods Programs Biomed 2014 Dec;117(3):435-447. [doi: 10.1016/j.cmpb.2014.09.002] [Medline: 25306242]
9. Castells F, Laguna P, Sörnmo L, Bollmann A, Roig JM. Principal Component Analysis in ECG Signal Processing. EURASIP J Adv Signal Process 2007 Feb 8;2007(1):1. [doi: 10.1155/2007/74580]
10. Monasterio V, Laguna P, MartÍnez JP. Multilead Analysis of T-Wave Alternans in the ECG Using Principal Component Analysis. IEEE Trans Biomed Eng 2009 Jul;56(7):1880-1890. [doi: 10.1109/tbme.2009.2015935]
11. Martis RJ, Acharya UR, Mandana K, Ray A, Chakraborty C. Application of principal component analysis to ECG signals for automated diagnosis of cardiac health. Expert Systems with Applications 2012 Oct;39(14):11792-11800. [doi: 10.1016/j.eswa.2012.04.072]
12. Kallas M, Francis C, Kanaan L, Merheb D, Honeine P, Amoud H. Multi-class SVM classification combined with kernel PCA feature extraction of ECG signals. 2012 Apr 23 Presented at: 2012 19th International Conference on Telecommunications (ICT); April 23 2012; Jounieh, Lebanon p. 1-5. [doi: 10.1109/ictel.2012.6221261]
13. Kachuee M, Fazeli S, Sarrafzadeh M. ECG heartbeat classification: A deep transferable representation. 2018 Jun 04 Presented at: 2018 IEEE International Conference on Healthcare Informatics (ICHI); June 4 2018; New York City, NY, USA p. 443-444. [doi: 10.1109/ichi.2018.00092]
14. Afonso V, Tompkins W, Nguyen T, Luo S. ECG beat detection using filter banks. IEEE Trans Biomed Eng 1999 Feb;46(2):192-202. [doi: 10.1109/10.740882] [Medline: 9932341]
15. Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nat Commun 2020 Apr 09;11(1):1760 [FREE Full text] [doi: 10.1038/s41467-020-15432-4] [Medline: 32273514]
16. Pyakillya B, Kazachenko N, Mikhailovsky N. Deep Learning for ECG Classification. J Phys Conf Ser 2017 Oct 25;913:012004. [doi: 10.1088/1742-6596/913/1/012004]
17. Zhang J, Tian J, Cao Y, Yang Y, Xu X, Wen C. Fine-grained ECG classification based on deep CNN and online decision fusion. Comput Res Repos 2019 Jan [FREE Full text]
18. Jun T, Nguyen H, Kang D, Kim D, Kim D, Kim Y. ECG arrhythmia classification using a 2-D convolutional neural network. arXiv Preprint posted online on April 18, 2018. [FREE Full text]
19. Rajpurkar P, Hannun A, Haghpanahi M, Bourn C, Ng A. Cardiologist-level arrhythmia detection with convolutional neural networks. arXiv Preprint posted online on July 7 2017. [FREE Full text]
20. Tang W, Long G, Liu L, Zhou T, Jiang J, Blumenstein M. Rethinking 1D-CNN for Time Series Classification: A Stronger Baseline. arXiv Preprint posted online on February 24, 2020. [FREE Full text]
21. Chen W, Wang S, Long G, Yao L, Sheng Q, Li X. Dynamic illness severity prediction via multi-task RNNs for intensive care unit. 2018 Nov 17 Presented at: 2018 IEEE International Conference on Data Mining (ICDM); November 17 2018; Singapore, Singapore p. 917-922. [doi: 10.1109/icdm.2018.00111]
22. Ismail Fawaz H, Forestier G, Weber J, Idoumghar L, Muller P. Deep learning for time series classification: a review. Data Min Knowl Disc 2019 Mar 2;33(4):917-963. [doi: 10.1007/s10618-019-00619-1]
23. Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, et al. Deep Learning: A Primer for Radiologists. Radiographics 2017 Nov;37(7):2113-2131. [doi: 10.1148/rg.2017170077] [Medline: 29131760]
24. Shyu L, Wu Y, Hu W. Using Wavelet Transform and Fuzzy Neural Network for VPC Detection From the Holter ECG. IEEE Trans Biomed Eng 2004 Jul;51(7):1269-1273. [doi: 10.1109/tbme.2004.824131]

25. Wang F, Casalino LP, Khullar D. Deep Learning in Medicine-Promise, Progress, and Challenges. JAMA Intern Med 2019 Mar 01;179(3):293-294. [doi: 10.1001/jamainternmed.2018.7117] [Medline: 30556825]
26. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc 2018 Oct 01;25(10):1419-1428 [FREE Full text] [doi: 10.1093/jamia/ocy068] [Medline: 29893864]
27. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? Heart 2018 Jul 19;104(14):1156-1164. [doi: 10.1136/heartjnl-2017-311198] [Medline: 29352006]
28. Wiens J, Shenoy E. Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. Clin Infect Dis 2018 Jan 06;66(1):149-153 [FREE Full text] [doi: 10.1093/cid/cix731] [Medline: 29020316]
29. Miotto R, Wang F, Wang S, Jiang X, Dudley J. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018 Nov 27;19(6):1236-1246 [FREE Full text] [doi: 10.1093/bib/bbx044] [Medline: 28481991]
30. Chen D, Liu S, Kingsbury P, Sohn S, Storlie CB, Habermann EB, et al. Deep learning and alternative learning strategies for retrospective real-world clinical data. NPJ Digit Med 2019 May 30;2(1):43 [FREE Full text] [doi: 10.1038/s41746-019-0122-0] [Medline: 31304389]
31. Jambukia S, Dabhi V, Prajapati H. Classification of ECG signals using machine learning techniques: A survey. 2015 Mar 19 Presented at: 2015 International Conference on Advances in Computer Engineering and Applications; March 19 2015; Ghaziabad, UP, India p. 714-721. [doi: 10.1109/icacea.2015.7164783]
32. Pathinarupothi R, Rangan E. Discovering vital trends for personalized healthcare delivery. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. 2016 Sep 12 Presented at: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; September 12 2016; Heidelberg Germany p. 1106-1109. [doi: 10.1145/2968219.2972716]
33. Mueen A, Chavoshi N. Enumeration of time series motifs of all lengths. Knowl Inf Syst 2014 Oct 18;45(1):105-132. [doi: 10.1007/s10115-014-0793-4]
34. Kwon O, Jeong J, Kim HB, Kwon IH, Park SY, Kim JE, et al. Electrocardiogram Sampling Frequency Range Acceptable for Heart Rate Variability Analysis. Healthc Inform Res 2018 Jul;24(3):198-206 [FREE Full text] [doi: 10.4258/hir.2018.24.3.198] [Medline: 30109153]
35. Chi L, Chi H, Feng Y, Wang S, Cao Z. Comprehensive and efficient discovery of time series motifs. J Zhejiang Univ - Sci C 2011 Dec 6;12(12):1000-1009. [doi: 10.1631/jzus.c1100037]

Abbreviations

ECG: electrocardiogram
GMM: Generalized Monitoring Model
PMM: Personalized Monitoring Model
RPM: remote patient monitoring

Edited by G Eysenbach; submitted 17.09.20; peer-reviewed by G Long, L Gao, KC Wong; comments to author 05.10.20; revised version received 11.10.20; accepted 21.10.20; published 29.12.20.

Please cite as:
Kotorov R, Chi L, Shen M
Personalized Monitoring Model for Electrocardiogram Signals: Diagnostic Accuracy Study
JMIR Biomed Eng 2020;5(1):e24388
URL: https://biomedeng.jmir.org/2020/1/e24388
doi:10.2196/24388
PMID:33529270

Original Paper

Rhythmic Haptic Cueing for Gait Rehabilitation of People With Hemiparesis: Quantitative Gait Study

Theodoros Georgiou[1*], PhD; Simon Holland[2*], PhD; Janet van der Linden[2*], PhD

[1]School of Mathematical & Computer Sciences, Heriot-Watt University, Edinburgh, United Kingdom

[2]The Open University, Milton Keynes, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Theodoros Georgiou, PhD
School of Mathematical & Computer Sciences
Heriot-Watt University
Edinburgh Campus
Edinburgh
United Kingdom
Phone: 44 131 451 4132
Email: t.georgiou@hw.ac.uk

Abstract

Background: Rhythm, brain, and body are closely linked. Humans can synchronize their movement to auditory rhythms in ways that can improve the regularity of movement while reducing perceived effort. However, the ability to perform rhythmic movement may be disrupted by various neurological conditions. Many such conditions impair mechanisms that control movement, such as gait, but typically without rhythmic perception being affected. This paper focuses on hemiparetic stroke, a neurological condition that affects one side of the body. Hemiparetic stroke can cause severe asymmetries in gait, leading to numerous physical problems ranging from muscle degeneration to bone fractures. Movement synchronization via entrainment to auditory metronomes is known to improve asymmetry and related gait problems; this paper presents the first systematic study of entrainment for gait rehabilitation via the haptic modality.

Objective: This paper explores the gait rehabilitation of people with hemiparesis following a stroke or brain injury, by a process of haptic entrainment to rhythmic cues. Various objective measures, such as stride length and stride time, are considered.

Methods: This study is a quantitative gait study combining temporal and spatial data on haptically cued participants with hemiparetic stroke and brain injury. We designed wearable devices to deliver the haptic rhythm, called Haptic Bracelets, which were placed on the leg near the knee. Spatial data were recorded using a Qualisys optical motion capturing system, consisting of 8 optoelectronic cameras, and 20 markers placed on anatomical lower limb landmarks and 4 additional tracking clusters placed on the right and left shank and thigh. Gait characteristics were measured before, during, and after cueing.

Results: All 11 successfully screened participants were able to synchronize their steps to a haptically presented rhythm. Specifically, 6 participants demonstrated immediate improvements regarding their temporal gait characteristics, and 3 of the 6 improved their gait in terms of spatial characteristics.

Conclusions: Considering the great variability between survivors of stroke and brain injury and the limited number of available participants in our study, there is no claim of statistical evidence that supports a formal experimental result of improved gait. However, viewing this empirical gait investigation as a set of 11 case studies, more modest empirical claims can be made. All participants were able to synchronize their steps to a haptically presented rhythm. For a substantial proportion of participants, an immediate (though not necessarily lasting) improvement of temporal gait characteristics was found during cueing. Some improvements over baseline occurred immediately after, rather than during, haptic cueing. Design issues and trade-offs are identified, and interactions between perception, sensory deficit, attention, memory, cognitive load, and haptic entrainment are noted.

(JMIR Biomed Eng 2020;5(1):e18649) doi:10.2196/18649

KEYWORDS

hemiparetic gait; stroke; technology assisted rehabilitation; quantitative study; gait analysis; gait asymmetry; gait; neurology; hemiparesis; rehabilitation; brain injury

Introduction

Brain injury following an accident or stroke can leave people with life-changing neurological conditions and a general weakness on one side of the body. Motor control of one side of the body can be severely affected with unilateral loss in sensation and muscle coordination of both upper and lower limbs.

Motor control deficiencies can lead to spatial and temporal asymmetries affecting walking in a condition known as hemiparetic gait. The asymmetries can cause sufferers of hemiparetic gait to overuse their nonaffected (nonparetic) leg, exposing it to potentially damaging, higher vertical forces [1,2], while underuse of the paretic (affected) leg can lead to loss of muscle tone and reduction of bone mineral density [3]. These effects, in turn, increase the risk of knee and joint problems, hip and bone fractures, and falls [4].

Walking following an external metronomic rhythm has been shown to improve gait, leading survivors of stroke to walk more symmetrically [5] and to neglect their affected leg less.

Early studies by Prassas et al [6] using rhythmic auditory stimulation (RAS)—a neurological technique using audio rhythms for facilitating rehabilitation, development, and maintenance of movements that are fundamentally biologically rhythmical—have shown immediate spatial benefits, with stride lengths becoming more symmetrical. In addition, the work of Prassas et al [6] found the hip joint range of motion to increase, and the center of mass displacement to decrease, during walking with RAS, making the overall forward movement smoother.

More recent studies found that survivors of stroke could easily synchronize their steps to a rhythmic audio metronome while walking on a treadmill; this led to improvements in various temporal symmetries, including improved step time symmetry between the paretic and the non-paretic leg [7], as well as reduced variability in the paretic step times [8].

These results indicate that RAS and, by extension, walking to a rhythm has clear clinical significance since step asymmetry is a leading cause of many problems associated with neurological conditions like hemiparetic gait [9]. More specifically, improvements in symmetry and variability suggest steps toward functional recovery of gait mechanics, as they are kinematically associated with a healthy gait [9].

However, auditory cues can be impractical to use outside of the clinic, or to use in situations where it is desirable to maintain environmental awareness or social engagement. Moreover, the way in which audio cues are presented to patients during rehabilitation in the clinic introduces particular limitations. In order to maintain communication between patients and physiotherapists, audio cues are usually played out through speakers. In addition, for optimum entrainment results, the tempo must match the patient's walking pace [10]. Therefore, only one patient can receive the audio cues at a time, confining the number of sessions that can run simultaneously in the same space to one. This could be problematic in situations where the same space (ie, a gym inside a rehabilitation clinic) is shared by more than one physiotherapist.

Haptic rhythms, on the other hand, can be directed to more than one patient simultaneously, each with their own tempo to match their cadence, without interfering with each other. This enables more efficient sharing of resources between health professionals and physiotherapists.

Cueing of the steps of each leg has demonstrated stronger auditory-motor synchronization than cueing those of just one leg (either paretic or nonparetic) [10]. However, interpreting an audio rhythm that differentiates cues for the paretic and nonparetic leg can challenge those needing concentration to walk at all.

A pilot study attempting to assign cues of different pitches to each leg [8] identified a number of limitations to this approach, mainly an illusion created by the pitch differences between successive cues causing an isochronous (regular) rhythm to be perceived as irregular. In addition to this acoustic illusion, participants in the same pilot study reported that having a rhythm of two tones is difficult to understand, with one participant withdrawing from the study because they "did not like the dual-tone" [8].

The haptic sense, on the other hand, has the potential to focus attention and proprioception on each leg by simply applying a tactile pulse to each leg in turn, in a more embodied, directly spatial, and less burdensome way.

As a systematic extension of an earlier pilot [11], this study is designed to investigate the effects of rhythmic haptic cueing on diverse gait characteristics associated with healthy kinematics and gait patterns. These include spatial characteristics (ie, stride lengths), temporal characteristics (ie, step, stand, and swing times), and derivative asymmetries, calculated from the spatiotemporal characteristics.

Methods

Procedure

This was a single-session, repeated-measures study. Participants with hemiparetic brain injury were first asked to walk a short distance without intervention to establish a baseline, and then their pace was matched to an isochronous (regular) haptic rhythm whose period matched a symmetric version of their natural cadence, calculated by averaging the time in milliseconds between each successive step. Subsequently, any resulting residual effects (such as rhythm persistence and the ability to walk to the rhythm from memory) were tested by asking the participants to walk to the rhythm from memory shortly afterward.

The default placement of the wearable devices we designed to deliver the haptic rhythm, called Haptic Bracelets, was on the

leg near the knee. This area, chosen due to findings from an earlier pilot study [12], is covered in what is termed “hairy skin,” which has a low concentration of the particular mechanoreceptors (sensory neurons) responsible for perceiving vibrational stimuli, namely Pacinian and Meissner's corpuscles. Hairy skin does, however, contain Merkel's disk and Ruffini endings, which can detect pressure and skin stretch [13]. Generally, tactile cues coming from the vibrotactile actuators of our devices are perceived more as a tap sensation than as a vibration, as reported by several participants in pilot testing [12]. This suggests that the dominant mode of perception will not be vibrational.

During this study, when choosing the intensity of cues, an informal qualitative approach was preferred, ensuring firstly that the participants could feel the cues, and then adjusting the intensity to personal preference. Author TG [14] engages in a full discussion of the technical implementation of the devices, the engineering of the haptic cues, special considerations made to enhance the tap sensation of each cue, and lab testing to refine the device and its placement in Chapter 5 of his work “Rhythmic Haptic Cueing for Gait Rehabilitation of Hemiparetic Stroke and Brain Injury Survivors.”

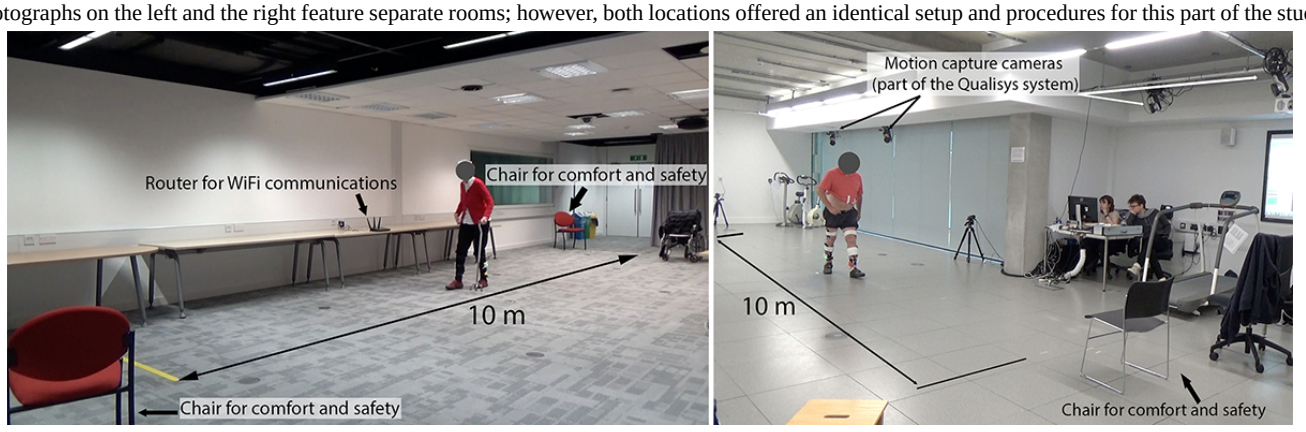
In order to identify quantitative changes in gait, two kinds of data were collected, involving temporal and spatial asymmetry, respectively; these measures often play a role in guiding the clinician’s treatment decisions [15]. This involved measures of paretic and nonparetic step timings and measures of paretic and nonparetic step length.

Temporal data were collected from all participants using the Haptic Bracelets wearable devices. The Haptic Bracelets are part of a wearable prototype system designed and developed at the Open University for the purpose of gait monitoring and gait rehabilitation through rhythmic haptic cueing [14].

Haptic Bracelets work in pairs (one for each leg). Each Bracelet contains sensors integrated in an inertia monitoring unit (IMU) for recording appropriate motion data. This data is used for calculating gait characteristics, including gait asymmetries and step variability in terms of overall temporal asymmetry (OTA) ratio values. Each unit also has a vibrotactile actuator capable of producing a tactile cue of comfortable quality but sufficient intensity to be felt by the user. Having one unit on each leg creates a multi-limb tactile metronome, producing tactile cues on alternating legs. This allows for the cueing of each limb independently, facilitating the investigation of effects related to attention and proprioception in the context of gait rehabilitation. In addition, this multi-limb approach has the additional advantage of motion data being gathered from each leg independently. This makes it possible to easily and clearly distinguish between the motion of the paretic leg (affected by the neurological condition) and the nonparetic leg. This improves the accuracy of the measurement of temporal characteristics and the analysis of asymmetry.

Temporal data was collected for all 11 participants (easy anywhere using the Haptic Bracelets). However, due to limitations in the geographical availability of participants, the collection of spatial data was limited to just 6 participants, as this was possible only at an optical motion capture facility that was local to these 6 participants. Figure 1 shows the lab setup during this study.

Figure 1. Participants during a 10-meter walking trial. The photograph on the right was taken during a trial in the optical motion capture lab. The photographs on the left and the right feature separate rooms; however, both locations offered an identical setup and procedures for this part of the study.



Participants

The participants in this study were 11 community-dwelling, community-ambulant adults (7 men and 4 women) with chronic hemiparesis (“chronic” defined as >6 months since stroke onset). For 10 of the 11 participants, hemiparesis followed a stroke; for the other participant, it was due to brain trauma following an accident. The age range of the participants is shown in Table 1, and the time since the occurrence of stroke varied from 8 months to 12 years. All participants gave written, informed consent to participate.

Participants were recruited through local support groups and the recommendations of private physiotherapists. Health professionals collaborating with our team for this study were responsible for determining the conditions of all of the inclusion and exclusion criteria. Inclusion criteria were (1) walking disability, but with a retained or subsequently recovered ability to stand and ambulate; (2) the ability to walk unsupported (but with a walking aid if needed) for a minimum distance of 10 meters; (3) a Rivermead Motor Assessment (RMA) scale score of more than 8.

Table 1. Participant demographic information (n=11).

Participant characteristics	Values
Age in years, mean (SD)	61.75 (7.85)
Gender, n	
Men	7
Women	4
Paretic side, n	
Right side	9
Left side	2

RMA is a standard and widely used test for assessing functional mobility in gait, balance, and transfers after stroke [16]. A score of 8 and higher was selected as the inclusion criterion following discussions with physiotherapists. RMA scores were collected in private by the health professionals (private physiotherapists and other health professionals working for support groups) handling our recruitment.

Participants were excluded if they had cognitive impairments preventing understanding of the task. Participants could use their assistive devices (ankle-foot orthosis splint or cane) in the trials.

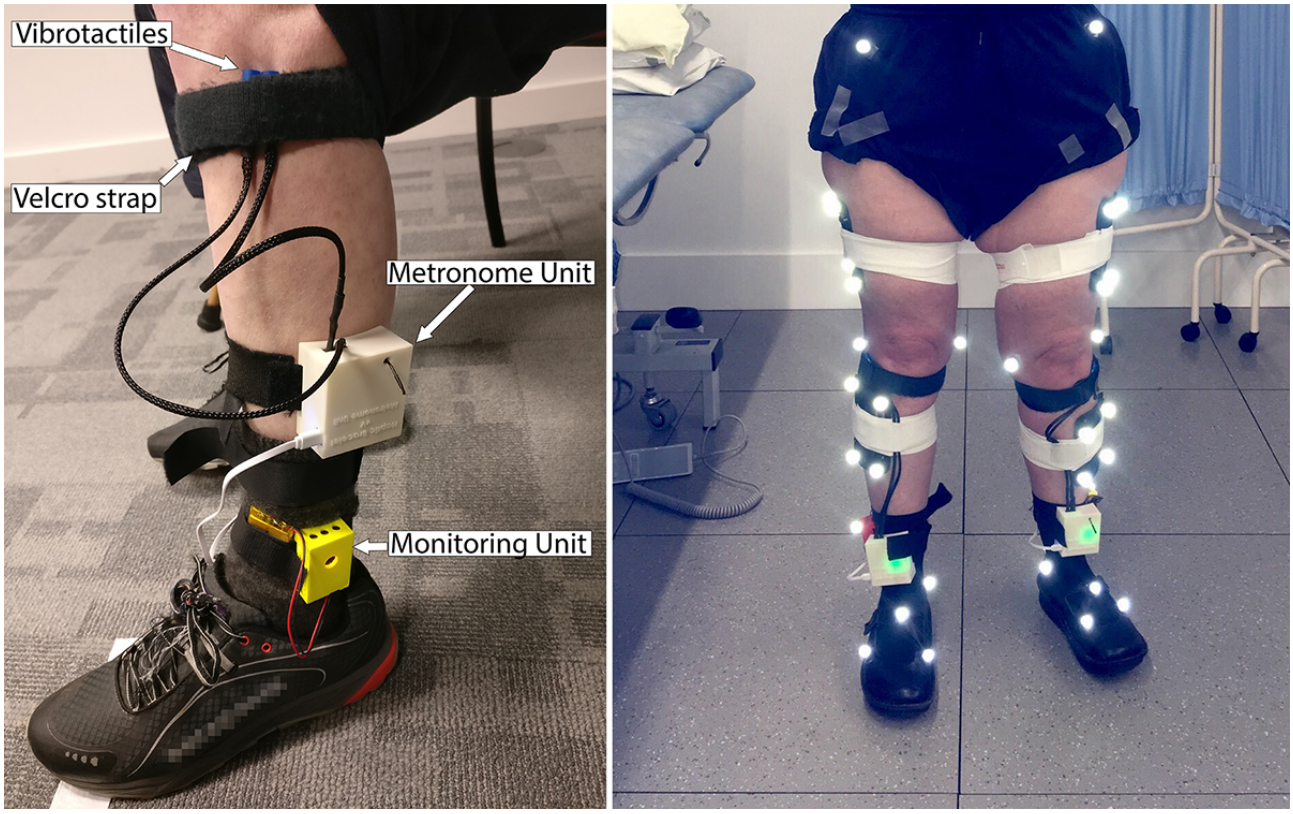
Prestudy Preparation

For the contingent reasons noted above, the study took place in two locations: the 6 local participants attended a kinematics lab in Manchester with an optical motion tracking system, and the 5 remaining participants attended a large observation lab in Milton Keyne. All participants had temporal data collected, but only those attending the optical tracking lab (6 out of the 11) had spatial data collected simultaneously (Table 2). An identical procedure was followed in both locations, except that 3 additional preparatory steps were required to meet the needs of the optical motion tracking system. Namely, participants at the kinematics lab location were asked to change into shorts, 30 to 45 additional minutes were required for a trained technician to place all of the optical tracking markers on their body (Figure 2), and biometric measurements were taken (ie, height and weight).

Table 2. Demographic data for each participant per location. Participants with an MMU code took the study at the Manchester location; those with an OU code took the study at Milton Keynes.

Participant code & locations	Age in years	Gender (M/F)	Years since brain trauma	Paretic side (Left/Right)	Data collected (Spatial/Temporal)
Manchester (kinematics lab)					
MMUP01	53	F	12	Right	Spatial and Temporal
MMUP02	57	F	0 (8 months)	Right	Spatial and Temporal
MMUP03	73	M	2	Left	Spatial and Temporal
MMUP04	68	M	7	Right	Spatial and Temporal
MMUP05	61	M	5	Right	Spatial and Temporal
MMUP06	55	F	4	Right	Spatial and Temporal
Milton Keynes (observation lab)					
OUP01	50	M	21	Right	Temporal only
OUP02	67	M	9	Right	Temporal only
OUP03	60	M	2	Right	Temporal only
OUP04	73	F	1	Right	Temporal only
OUP05	56	M	1	Left	Temporal only

Figure 2. Left: Haptic Bracelets wearable devices (metronome and monitoring units) and vibrotactiles strapped on a participant's leg using Velcro straps. Right: all the optical markers attached to a participant's body.



In both locations, the Haptic Bracelets were attached via Velcro straps onto the tibia of each leg near the ankle. The vibrotactile (the part of the device that gives the haptic cue) was attached using another Velcro strap near the knee (Figure 2).

The placement of the vibrotactile was initially based on the suggestion of physiotherapists participating in an earlier study [12]. However, this decision was later revealed to be based on a conflation by the physiotherapists of stimulus response with entrainment (ie, conflating functional electrical stimulation aimed at a direct muscle response with neural entrainment, where a regular external pulse leads to a matching, predictive internal neural resonance [12]). Nonetheless, it was subsequently decided to maintain this placement for this study for two main reasons: (1) proximity to major nerves, giving the tactile cue a good chance to be felt; and (2) having the vibrotactile unit away from the IMU of the Haptic Bracelet helps to minimize unwanted noise in the gait data.

Participants who did not have their spatial data monitored had a simpler preparation than those that did; they did not have to wear any kinematics markers and could simply wear a pair of Haptic Bracelets over or under their everyday clothes (depending on preference).

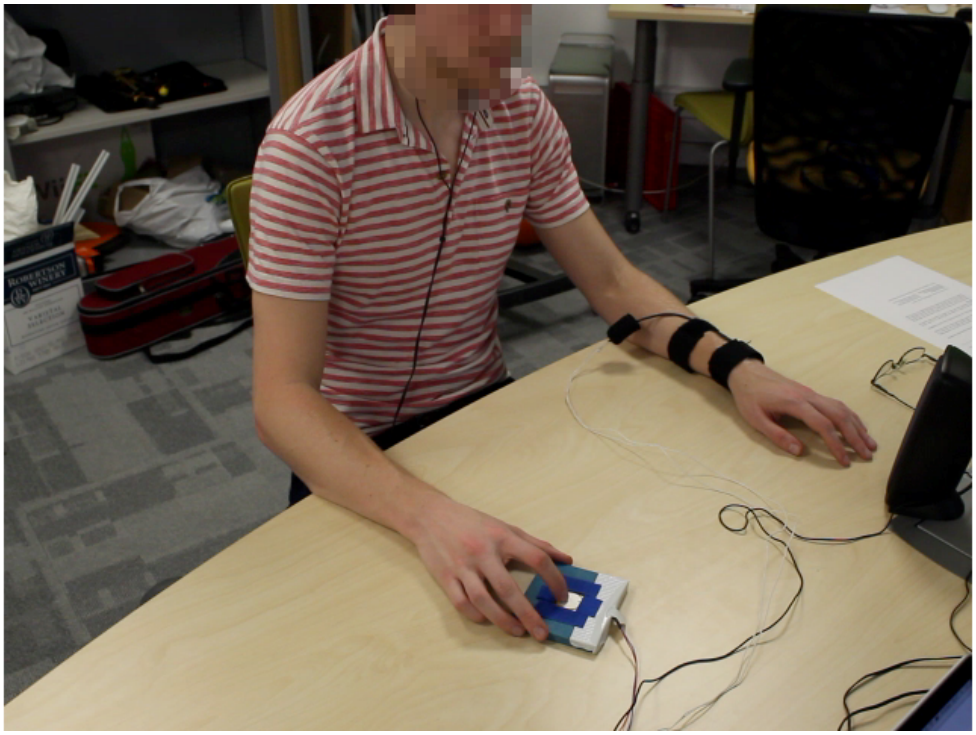
The Haptic Tap Test

Before the first trial, all the participants were asked to take a haptic rhythm perception test, called the haptic tap test. Performance in this tap test was not used as an inclusion criterion for the study but was rather used to investigate the participants' rhythm perception capabilities.

During rhythm-based rehabilitation sessions, it is not easy to distinguish between survivors of stroke who simply have physical difficulties walking to a rhythm from participants who are unable to entrain to a rhythm for other perceptual or cognitive reasons. This ability to perceive a rhythm can be affected by injury to certain parts of the brain [17]. Such a deficit may go undetected by medical professionals and is not identified by standardized gait assessment tests such as the Rivermead Mobility Index [16] and the timed-up-and-go (TUG) test [18].

Therefore, we administered a rhythm perception test. In this test, patients tap a button following a steady rhythm presented to them haptically via a vibrating unit fixed on their wrist (Figure 3). Their ability to synchronize their tap tempo to the rhythm can be used for assessing their ability to perceive and follow, or *entrain* to, the given rhythm. Consequently, this test can be used for identifying patients who are (1) able to entrain to rhythms and (2) not impaired in the ability to entrain to rhythm presented haptically.

Figure 3. Participant performing a haptic tap test by tapping a button with his right index finger. The vibrotactile of the Haptic Bracelet device is secured on his left wrist using a pair of Velcro straps.



The haptic tap test allows for an empirical indication of the participants' cognitive and perceptual abilities to entrain to a rhythm for movement rehabilitation, aiming to assess people's ability to perceive and replicate a rhythm (albeit by tapping) perceived via the haptic modality.

To conduct the haptic tap test, participants were asked to tap with their index finger on a sensor in time to a range of rhythms. The rhythm was delivered haptically on the paretic wrist, using a wired version of the Haptic Bracelets. Tap times were recorded digitally. Tapping was performed by the nonparetic hand in order to avoid any physical constraints due to hemiparesis or the effects of haptic masking [19]. Haptic masking describes situations where haptic sensation is temporarily muted or attenuated by adjoining muscle movement. Participants were all tested with rhythmic periods of 500, 600, and 700 ms. Based on experiences from pilot tests and in a previous, similar study involving able-bodied participants, each trial lasted just 20 seconds to minimize participant fatigue. The ability to tap was observed visually during the task and was also analyzed from the digital data log after the conclusion of each session. Flawless performance was not required; a general ability to keep in time with the beat was sufficient to pass the test.

Walking to the Rhythm Trials

This study followed a repeated-measures design with three conditions: baseline, cued, and post. The participants were asked to walk the length of a 10-meter runway 6 times for each condition. A 5-minute break was offered between each condition.

Baseline Condition

The initial condition allowed each participant's baseline gait to be measured, including mean step time. The mean step time was used to set the period of the haptic metronome for the subsequent cued condition, subject to any final adjustments.

Choosing the Period for the Cued Condition

This approach to choosing the cueing period was motivated by the literature, but with scope for practical adjustments to cope with the realities and comfort of hemiparetic participants. One of the underlying neurological principles defined in the RAS rehabilitation technique is to entrain steps to an external rhythm that, as nearly as possible, matches the preferred (uncued) cadence of the patient [20]. Independently, Roerdink [10], working with elderly able-bodied participants, found optimal performance when cueing with a regular period close to each individual's naturally preferred cadence. Cues in this study were delivered to alternate legs, with evidence from auditory cueing emphasizing the benefits of this approach [21].

Bearing in mind the characteristic temporal asymmetry of hemiparetic gait, the *average* period, calculated from a hemiparetic individual's preferred pace, may be too fast for one leg and too slow for the other; this suggests that some trial-and-error adjustments of the period may be desirable. Also, some participants needed to apply conscious effort to walk naturally, and thus, some might feel less confident when asked to undertake an additional task while walking. For these two reasons, some leeway in the preferred adjustment of the period was allowed.

Familiarization Period

In order to find a comfortable period, close to the preferred period before commencing with cued walking, participants were variously asked to tap their foot while sitting down, or to step in place while standing up, or to walk around following the rhythm, as was most convenient for them. This familiarization period allowed final adjustments to the period and to the

intensity of the haptic cue. The haptic cue intensity was adjusted to meet each participant's personal preference, making sure they could feel it while walking rather than trying to home in on specific frequencies for mechanoreceptor activation.

Findings from an earlier study using a technology-probe approach [12] highlighted the importance of clear instructions on entraining steps to a rhythm. Participants were instructed to time the steps of their nonparetic legs to the beat but not to worry about the cadence of their paretic leg. They were encouraged to feel the rhythm in a similar manner as one would to a song. After testing various instructions through trial and error during the earlier pilot study [12] and with pilot volunteers in the lab, this instruction was found to generally lead to a more balanced gait than asking participants explicitly to time the steps of *both* legs to their respective beats.

Here an interesting issue arises. One might be tempted to engineer a cue where the ratio of step periods between legs lay at some intermediate point between the participant's baseline asymmetry and a perfectly balanced symmetry. However, a wide range of studies [22-25] indicate that rhythms with exactly regular periodicity are best at engaging with human neural mechanisms for entrainment to allow upcoming beats to be precisely predictable, thus facilitating bodily entrainment. Similar considerations apply when recalling previously heard rhythms [14,22,24]. These findings agree with observations during in-lab pilot testing, where participants generally found it much easier to entrain to a regular beat than an irregular or "swung" beat (in musical terminology).

Cued Condition

Once participants confirmed that they understood the instructions, they were asked to walk six 10-meter lengths following the haptic rhythm. Temporal data and, in the case of the participants in the Manchester lab, spatial data were recorded for each walk. A short 5-minute break was offered between the cued and post conditions.

Post Condition

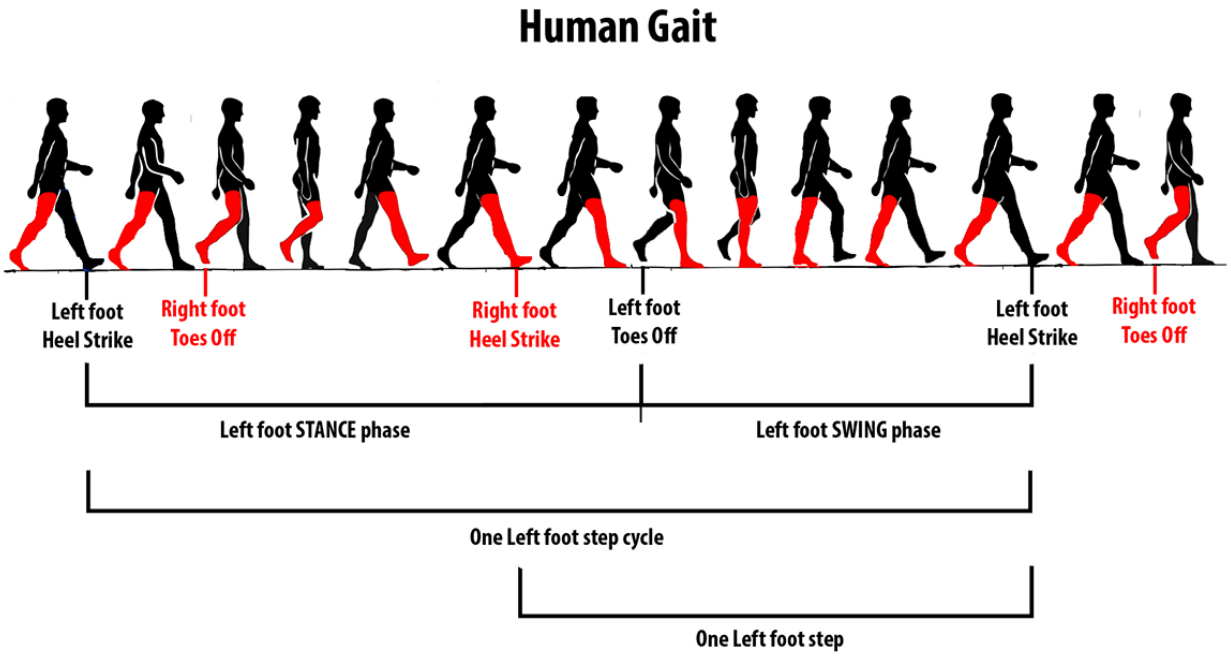
Immediately following the cued condition, and after a short 5-minute break, each participant was asked to repeat a further six 10-meter walks without haptic cueing. The purpose of this post condition was to investigate any residual effects of walking to the rhythm. This was inspired by literature on Parkinson disease reporting rhythm persistence [26-28], and also by comments from a participant in the pilot study [12] who noted how the rhythm remained in their memory; "If it is switched off […], it's still there. […] in my head." (Participant 4 [12]).

In addition to a short 5-minute break scheduled between conditions, participants were told they could take a break at any point during the study. The study concluded with a short, unstructured discussion about the participant's experience walking with the haptic rhythm to help us understand how our gait rehabilitation method was perceived and if any methodological or technological improvements or alterations were necessary. All trials were video-recorded for later review by expert physiotherapists as, even if they were present in the room during the cued and post conditions, they were often concentrating more upon the participant's safety than subtle changes in gait pattern and body movement.

Data Analysis

When assessing gait, it is useful to break walking down into various constituent elements. For each step, the foot lifts off the ground, swings forward, hits the ground, and stands until ready to lift off again, forming one step cycle (Figure 4).

Figure 4. Phases in human gait. The swing phase is defined from the moment the toes of the foot initiating the step lift off the ground and the leg begins to swing forward. This phase completes when the heel of that foot strikes the ground, starting the stance phase. Between two successive step cycles of alternating legs, there is the double support phase, where both legs touch the ground (the time between the heel strike of one foot and toes off of the other foot). This is an integral part of walking; however, double stance time does not affect how gait symmetries are calculated in this paper.



Multiple-step cycles happen for each alternating leg. Therefore, each leg acts as an oscillator performing a fundamentally cyclic process on every step. However, unlike a simple pendulum, the cycle time of normal gait is not purely mechanically determined and can exhibit any of a wide range of periods. Gait patterns can be entrained to a suitable external rhythm by a neurologically mediated process, as described in the following summary adapted from Holland et al [29]. The concept of entrainment, originally from physics, describes how two or more physically connected rhythmic processes can interact with each other to adjust towards, and eventually lock into, a common or closely related periodicity and phase. However, the concept has unexpected application in areas as diverse as perception, cognition, and music. The best current account of the underlying brain mechanism comes from Neural Resonance Theory [22,23]; it proposes that humans have a specialized neural organ consisting of a bank of actively powered oscillators with temporal periods from about 200 ms to 2 seconds. Human entrainment can be explained by the way in which these hypothesized oscillators tend to entrain with sensory input and are able to entrain, in turn, with motor processes such as gait. Given that rhythm and gait rehabilitation are focal points of this study, the capacity of gait to entrain with external rhythms is of vital importance.

In this study, temporal and spatial data were recorded and analyzed. Temporal data refers to times between events (eg, time between subsequent heel strikes) and other gait characteristics that can be calculated from these timings. In particular, temporal data were analyzed and assessed using the OTA ratio.

Overall Temporal Asymmetry (OTA) Ratio

First, the swing-stance ratio is calculated for each leg (1). This is the ratio (1) between the swing (SW) and stance time (ST) for each leg in turn.

The OTA ratio (2) compares the swing-stance ratio of the paretic leg with the nonparetic leg.

For a healthy individual, the value for the OTA should be between 0.9 and 1.1, which is a range described as healthy or normative [15]. In the case of neurological conditions, higher swing-stance ratios are commonly seen. For example, in the case of hemiparesis, one side of the body is affected unilaterally, typically causing the affected leg to swing more slowly. Due to weakness and sensitivity loss, a survivor of stroke often loses trust in the affected leg, doubting it can support them. This further causes them to swing the nonaffected (nonparetic) leg faster in order to minimize the time spent standing on the affected leg. Consequently, the swing time of the *nonparetic* leg decreases (ie, swings faster), causing the stance time of the paretic leg to also decrease.

These combined changes in stance and swing times of both legs raise the OTA value. Generally, while 0.9 to 1.1 describes normal asymmetry, values between 1.1 to 1.5 describe mild asymmetry, while values over 1.5 describe severe asymmetry [15]. As previously noted, entraining to an external rhythm has been shown to reduce gait asymmetry and results in a more symmetric and less variable gait pattern [24].

Spatial Data Analysis

Spatial data are used to determine gait characteristics relevant to space, such as stride length. Spatial data were recorded using a Qualisys optical motion capturing system. The system consists of 8 optoelectronic cameras, with a sampling frequency of 100Hz [30]. The trajectories of 20 markers placed on anatomical lower limb landmarks and 4 additional tracking clusters placed on the right and left shank and thigh (Figure 2) were collected and filtered using a fourth-order, zero-lag, low-pass Butterworth filter with a 6Hz cut off.

Each marker is tracked by the cameras, triangulating its position in space. This allows for the tracking of motion in 3 degrees of freedom, to millimeter accuracy. For spatial data, the movement of the markers in the forward direction was calculated for each step.

Spatial and temporal data collected during the cued and post conditions were compared against the baseline, looking for any effects that entrainment to haptic rhythmic cueing might have on the participant's gait. Comparing the data against the baseline allowed each participant to act as their own control, making clear any walking effects that were caused by the cue.

Results

Outliers

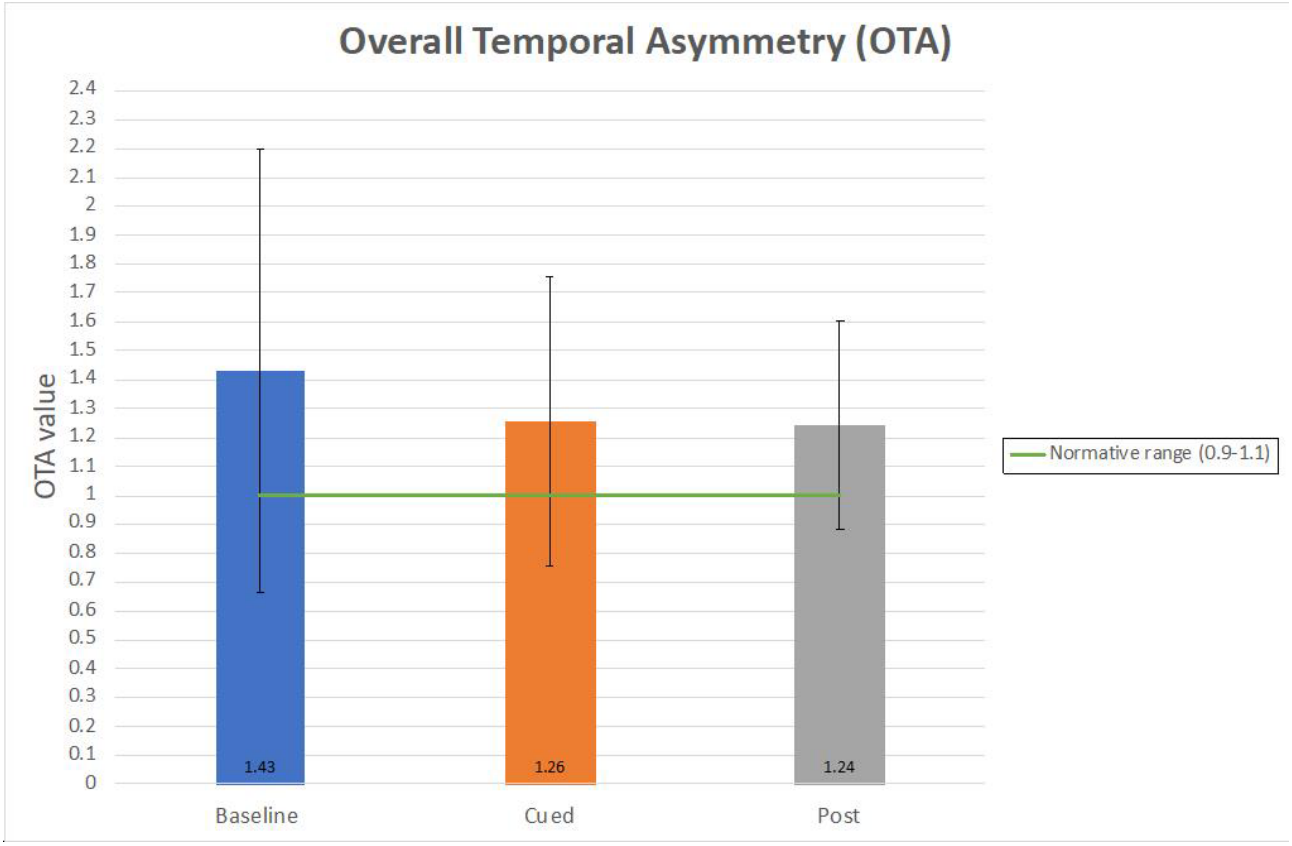
All data were examined for outliers. Outlier values were identified using the Tukey fence method. The Tukey fence method defines outliers as any values lying at a distance greater than 1.5 times the interquartile range. For characteristics that jointly contribute to the step time (ie, swing and stance times), if either value was considered to be an outlier, both values were removed from the analysis.

Temporal Data

As previously discussed, paretic and nonparetic step timings were determined from initial footfall contact by the Haptic Bracelet's onboard sensors. Analyzing data from different sensors and combining the information together allowed for the analysis of stance and swing times. From these, the OTA ratio value was calculated for every participant in every condition (baseline, cued, and post).

The average reduction on the OTA values across all participants indicates an overall improvement, with the OTA value approaching normative asymmetry levels (normative range: 0.9-1.1 [15]). The results are summarized in Figure 5.

Figure 5. Mean overall temporal asymmetry (OTA) values from all participants for all three conditions: baseline, cued, and post. The figure includes normative levels indicating healthy walking symmetry. Error bars show 1 standard error of mean.



Methodological Note: Comparing Baseline, Cued, and Post Conditions

Due to the high degree of variability between survivors of hemiparetic stroke (Figure 5) and the relatively small number of participants, statistics only reveal a limited part of the story; more information can be gleaned from the details of individual cases.

Consideration of Individual Results

Due to the inherently wide variability between individual survivors of stroke and brain injury, changes to the OTA value were normalized by calculating the percentage changes from each individual's baseline value. Negative percentage values indicate beneficial change (ie, approach the normative value range of 0.9 to 1.1). These data are shown in detail in Table 3, graphed in Figure 6, and summarized for clarity in Figure 7.

Table 3. The overall temporal asymmetry (OTA) values of all participants in all three conditions (baseline, cued, and post; n=11).

Participant code	OTA			OTA percentage change from baseline (%)[a]	
	Baseline	Cued	Post	Cued	Post
MMUP01	1.32	1.23	1.51	-6.99[b]	14.37
MMUP02	1.08	1.06	1.04	-1.94[b]	-3.89[b]
MMUP03	2.27	2.46	1.83	8.46	-19.16[b]
MMUP04	1.11	1.04	1.25	-6.66[b]	12.30
MMUP05	1.84	1.58	1.64	-14.11[b]	-10.88[b]
MMUP06	1.16	1.07	1.22	-7.36[b]	5.13
OUP01	0.85	0.92	0.88	-7.72[b]	-3.06[b]
OUP02	1.14	1.14	1.18	0.44	3.80
OUP03	3.32	1.75	1.56	-47.35[b]	-52.93[b]
OUP04	0.79	0.79	0.80	-0.04[b]	0.17
OUP05	0.86	0.76	0.74	11.68	14.01

[a]Represents the change from the baseline toward the normative range of symmetry (0.9 - 1.1), representing the healthy portion of the population.
[b]Negative values indicate an improvement in symmetry.

Figure 6. Graphical representation of OTA values shown in Table 3. Perfect symmetry (1.0) and normative asymmetry levels in the range of 0.9-1.1 are shown by horizontal green lines.

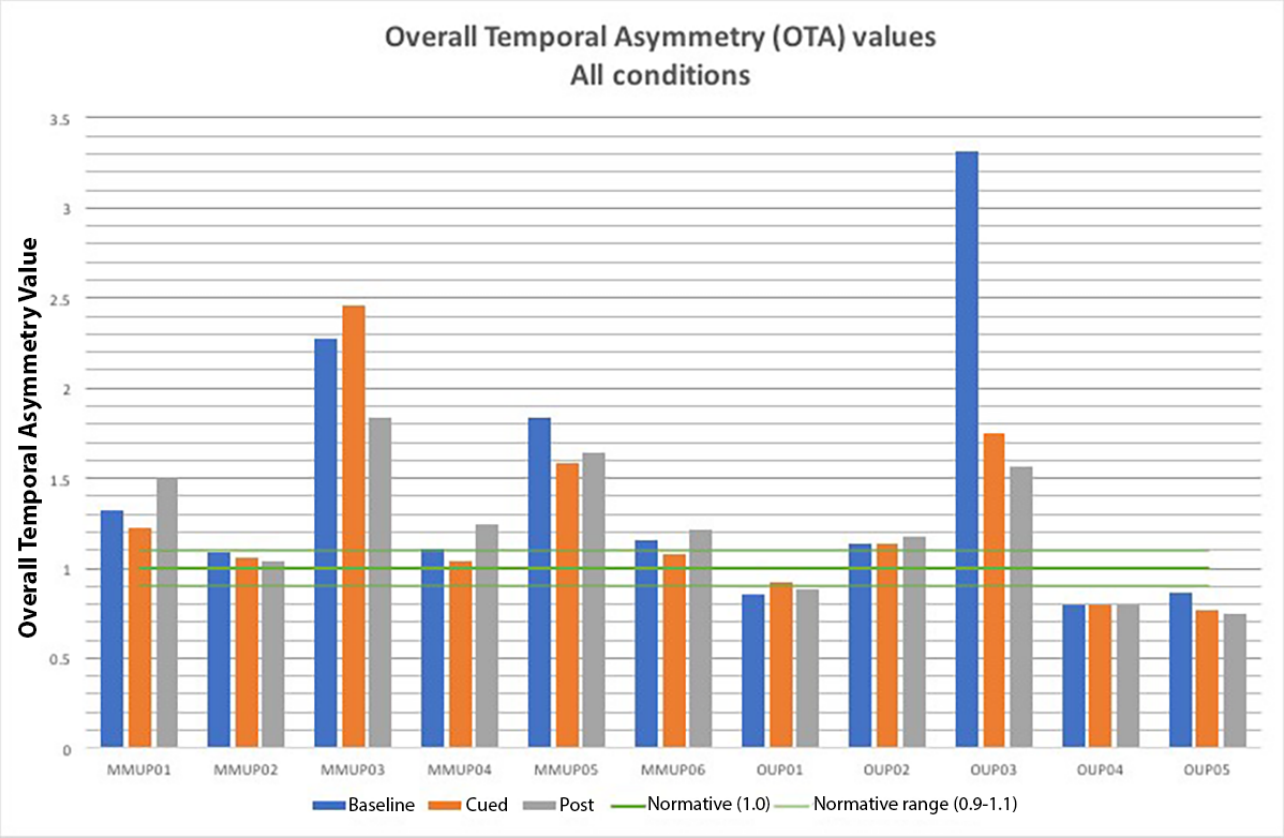
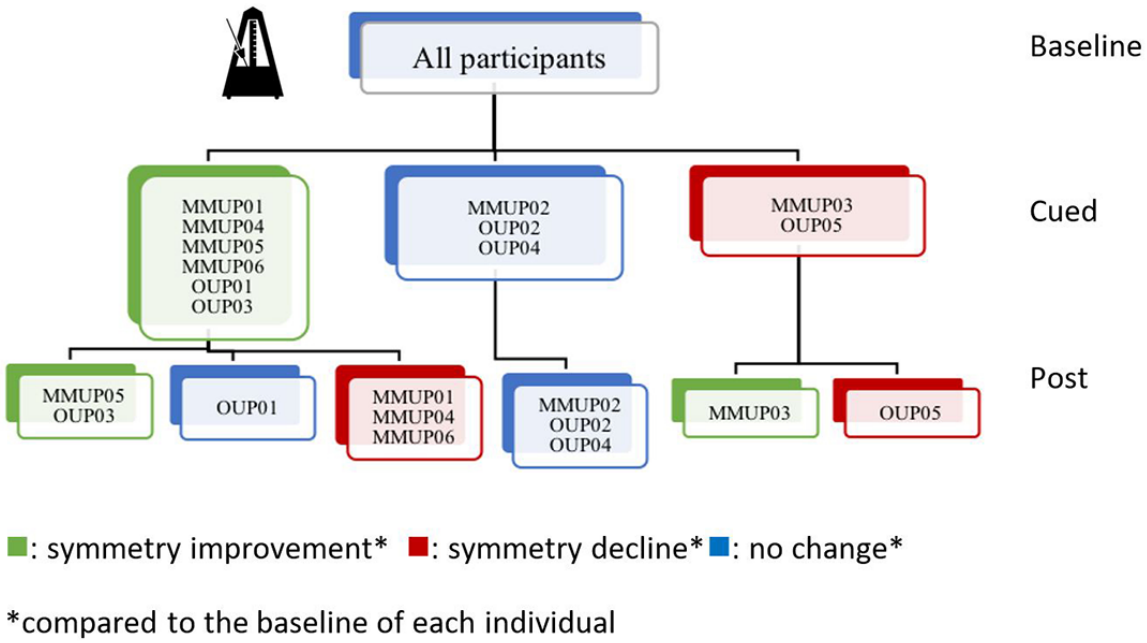


Figure 7. Temporal changes. The participant distribution is based on their overall-temporal-asymmetry percentage change compared to baseline. The three study conditions (baseline, cued, and post) are represented by the three levels of the flowchart.



Temporal Data Results Summary

In this study, three outcomes were possible for the cued and post conditions: Compared to the baseline, a participant's OTA could improve, worsen, or stay the same. Alterations of the OTA were considered meaningful if their magnitude was ±5% compared to each individual's baseline.

As shown in Table 3 and Figure 6, six of the 11 participants with brain injury exhibited immediate improvements in their gait, with their OTA values decreasing toward better symmetry in the cued condition compared to their baseline. Of these 6, 2 maintained a more symmetric OTA value in the post condition compared to their initial baseline, 3 became worse with increasing OTA values, and 1 returned back to the baseline level of asymmetry.

During the cued condition, 2 of the 11 participants became more asymmetric, with their OTA values increasing compared to their baseline. Of these two participants, one (patient code OUP05) maintained a more asymmetric OTA value in the post condition while the other (patient code MMUP03) showed a big improvement compared to the baseline.

Of the 11 participants, 3 did not show any changes in their OTA values, in either the cued or post condition. These data are summarized in Figure 7, where each level on the flowchart represents each condition (baseline, cued, and post). Temporal gait changes for each participant are color-coded: green indicates an improvement, red indicates an increase in asymmetry, and blue indicates no change. All comparisons are made against each individual's baseline values.

Spatial Data

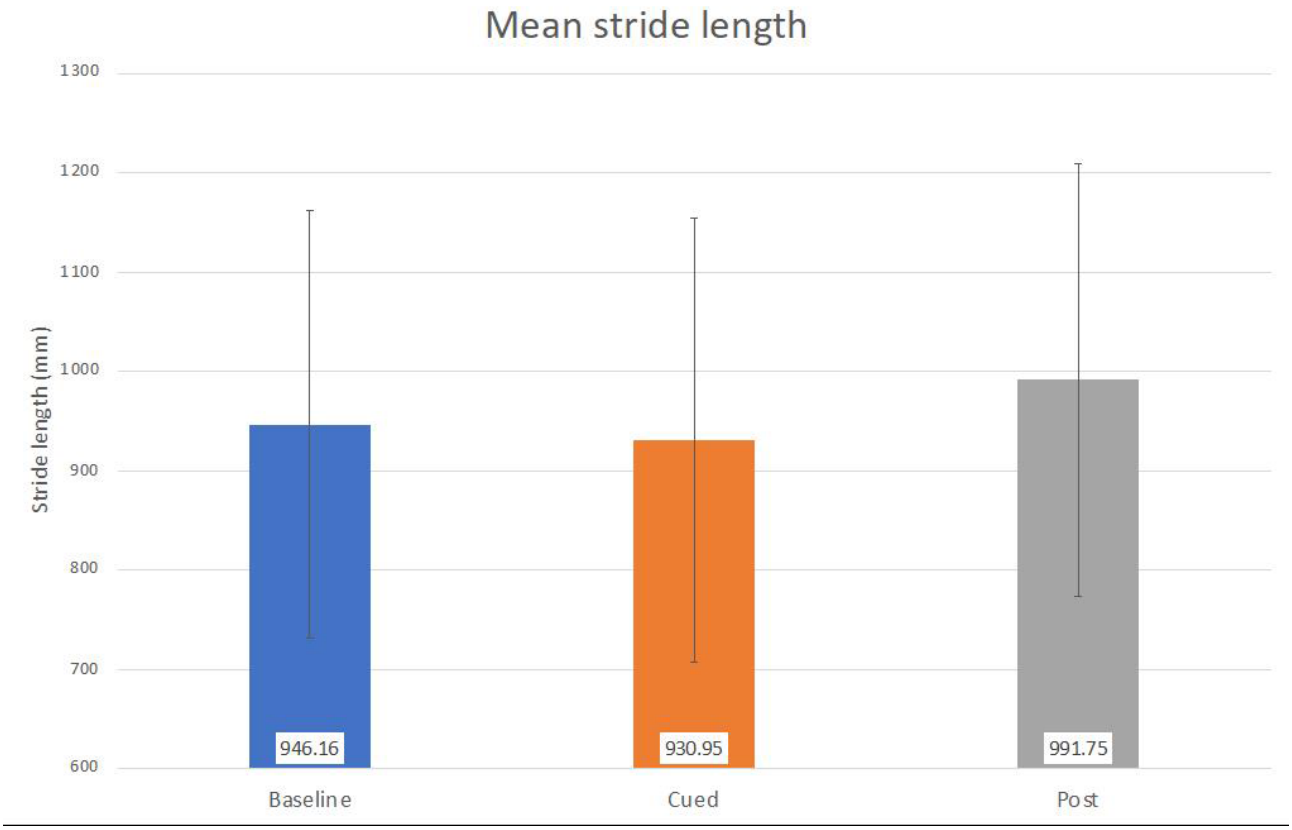
Spatial data usefully complement temporal data for assessing changes in gait symmetry. The step length is a valuable metric for assessing the therapeutic effect of rhythm in gait rehabilitation and can be used for calculating changes in spatial gait asymmetry [24,31,32].

Two aspects of spatial data were analyzed: firstly, step length in the cued and post conditions were compared against the baseline, and secondly, the spatial asymmetry value was calculated using the same formula discussed above for the OTA.

Due to the great variability between participants, useful information can be extracted by considering both changes in mean values and changes in individual cases. Changes in mean values are considered first.

In contrast to the change in symmetry, the average stride length *decreased* by 16 mm in the cued condition compared with the baseline, and increased by almost 50 mm in the post condition over baseline (Figure 8). Bearing in mind that all participants in this study had experienced hemiparetic stroke with various degrees of cognitive and motor control deficiencies, it may be that the increased cognitive load of attending to the rhythm led to shorter stride lengths. A related but slightly different argument might be that asking participants to walk while following the haptic rhythm may have led to more conscious attention to their movement, increasing cognitive load and causing them to take shorter and more careful steps. Further light is cast on these hypotheses by the observation of considerably longer steps in the post condition, where the external haptic rhythm was removed and participants were asked to walk to the rhythm from memory.

Figure 8. Average stride lengths between all participants for all three conditions: baseline, cued, and post. Stride lengths are calculated from optical marker data and analyzed using bespoke algorithms. Error bars show 1 standard error of mean.



The standard error value remained similar for all conditions: SE 215 mm, SE 223 mm, and SE 217 mm. This suggests that, on average, stride length did not become more or less variable during either the cued or the post condition when compared to the baseline. Unsurprisingly, in all conditions, the mean paretic stride length was shorter than the nonparetic stride length (Table 4).

Table 4. Average stride lengths of 6 participants using optical marker data from a Qualisys motion capture system. Each value represents the mean of trials in each condition (baseline, cued, post).

Participant code	Paretic leg, mean stride length (mm)			Nonparetic leg, mean stride length (mm)		
	Baseline	Cued	Post	Baseline	Cued	Post
MMUP01	1106.71	1086.47	1110.63	1105.03	1080.39	1108.21
MMUP02	812.59	1001.08	1092.22	801.57	997.30	1090.41
MMUP03	647.08	586.77	717.40	685.90	621.79	749.60
MMUP04	1146.11	1182.15	1193.77	1148.95	1194.59	1208.98
MMUP05	743.89	709.52	630.48	841.82	727.63	767.68
MMUP06	1159.07	992.25	1117.81	1155.14	991.48	1113.84

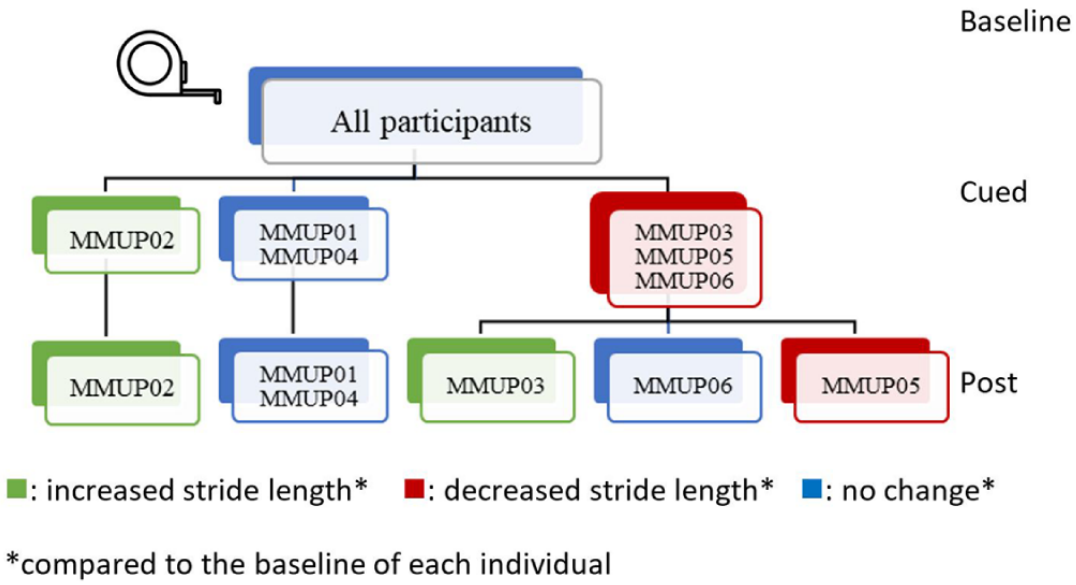
Spatial Data Results Summary

As previously noted, in situations with great individual variation, it is important to consider and analyze individual results from all participants before drawing firm conclusions. In this section, individual changes in spatial data are summarized. For 3 of the 6 participants, the mean cued stride length was shorter than the baseline. When cued, 1 participant walked with longer stride lengths, and 2 did not show any changes when walking with the cue.

When walking to the rhythm from memory in the post condition, the stride lengths of 2 participants increased, 1 decreased, and 3 showed no change. Interestingly, the same 2 participants were unchanged during cueing and post.

This data is summarized in Figure 9; each condition (baseline, cued, and post) is represented by a different level in the flowchart. Spatial changes for each participant are color-coded: green indicates an improvement, red indicates a decrease in stride lengths, and blue indicates no change. All comparisons are made against each individual's baseline values.

Figure 9. Spatial data. The participant distribution is based on their stride-length percentage change compared to baseline. The three study conditions (baseline, cued, post) are represented by the three levels of the flowchart.



Discussion

Interpretation of the Findings

The purpose of this study was to examine the effects of rhythmic haptic cueing on spatial and temporal gait characteristics of people with hemiparesis.

Two kinds of improvement (amongst others) that physiotherapists and other relevant health professionals seek to achieve in the gait of a survivor of hemiparetic stroke are better temporal symmetry and longer stride length. One must be careful in the interpretation of quantitative information in terms of mean results from studies with a relatively small number of participants; however, we did observe overall improvements in symmetry and longer strides, both during and immediately after cueing.

Nevertheless, a more nuanced and arguably more interesting story unfolds when considering the results of individuals. Individual temporal and spatial results for each participant are presented separately in Figures 5 and 6.

Temporal Walk Symmetry and Temporal Gait Pattern

Of the 11 participants, 6 showed immediate improvements in their OTA values during the cued condition (Figure 7). This indicates a more symmetric walking pattern, with the participants spending more time on their paretic leg, reducing the burden on the nonparetic leg. Of these 6 participants, 2 retained near normative OTA value in the post condition, in which they were asked to walk to the rhythm from memory. This suggests rhythm persistence and continuing entrainment influencing their motor response after the cue is removed. An analogous phenomenon has been noted in the past by Thaut [24] where patients with hemiparesis were given an audio-rhythmic stimulus for long periods of time before asking them to walk to the rhythm from memory. Our study appears to be the first to show a similar persistence effect with a haptic cue and to demonstrate that this can take effect with relatively short exposure.

Of the 4 participants who exhibited improvements in the cued condition, 1 returned to their baseline symmetry and the other 3 showed evidence in the post condition of becoming *more* asymmetric than their baseline. This increase in asymmetry may be attributable to fatigue, as the post condition was at the end of the study, after the participants had already walked roughly 120 meters (2x6x10 meters). While such a distance may not tax an able-bodied person, it can be considerably taxing for people with hemiparesis. Comments from 2 participants shared that this was the longest they had walked since their stroke. Fatigue may have negatively affected the performance of other participants in the post condition without bringing it below baseline.

Moreover, 3 participants (participant codes MMUP02, OUP02, and OUP04; Figure 7) showed no change in either the cued or post conditions. This may be either because they did not understand or engage with the task or simply because they had far less room for improvement than other participants; these 3 participants had OTA values that approached normative in the baseline condition (1.08, 1.14, and 0.79, respectively, compared to the normative range of 0.9-1.1). However, 1 of these 3 participants (participant MMUP02) showed improvement in step lengths for both conditions, compared to her baseline. This indicates the possibility of therapeutic gait-related benefits and shows the change in confidence a simple intervention can make, as well as its effect on healthier kinematics, even without a change in temporal gait pattern. Unfortunately, no spatial data were collected for the other 2 participants (participants OUP02 and OUP04).

Stride Lengths Changes

In this study, only 1 out of the 6 participants showed an immediate improvement in stride length and also maintained it during the post condition (Figure 9). Participant MMUP02, who normally uses a walking stick to ambulate, requested to do the trials without her stick, against the recommendations of one of the present physiotherapists. When asked immediately after the trials, participant MMUP02 commented, “it’s when I’ve gone

outside [my house] I've got no confidence, so I have the walking stick. [...] I felt a lot stronger walking in there without the stick. It made a difference in that I had a bit more confidence. I felt confident without the stick." When asked if this confidence arose from having the physiotherapist team present in the room with her, she initially agreed but then suggested further reasons: "Yes, I felt confident when she [physiotherapist] was there [next to her], but when she wasn't there, I still felt that confidence" [Participant MMUP02].

During the trials, the clinical gait analyst attending the study also commented on how the posture of this participant had changed from being crouched over her walking stick to walking upright and looking ahead. "Look how tall she [MMUP02] stands." Interestingly, participant MMUP02 is one of the participants that showed no change in her temporal data. This outcome is not entirely surprising as there is evidence in the literature to suggest that spatial and temporal parameters need not be directly related [33].

Video footage of the study was informally shown to the expert physiotherapists collaborating in this research who assisted with the initial participant recruitment for this study. As some of the study participants were their regular patients, they were also asked to comment on anything they saw as noteworthy. One of the physiotherapists remarked that participant OUP02 (who is also one of her regular physiotherapy patients) changed her habits after this study to include more walks in empty supermarket corridors during early mornings, when it is less busy. "[Participant OUP02] became more confident…(and) started using a lighter walking stick," which the therapist interpreted as another sign of increased confidence in her capabilities. While watching a video of participant OUP02 walking in one of the cued trials, the physiotherapist commented that "it's so nice to see her [OUP02] not thinking too much about her right (paretic) leg. Just let it go; let it flow."

After discussing the above findings with the physiotherapists assisting with the study, an interesting link between mood and one particular spatial (as opposed to temporal) gait characteristic was highlighted, with a potential bearing on the above results. Physiotherapists and gait rehabilitation experts routinely expect a confident mood to be associated with long strides. In contrast, in situations where hemiparetic walkers are insecure and being mindful of how they walk, shorter strides are expected. This behavior is typically observed independent of cadence (ie, walking tempo). Therefore, due to the specific link of confidence with stride lengths, the relatively short exposure to this novel approach to walking may have had a crucial role in the spatial changes observed.

Walking to a haptic rhythm mediated via a multi-limb metronome is unusual and not often encountered in everyday life. The novelty of this approach may have decreased the confidence of some participants when they walked with shorter stride lengths, as was observed in the cued condition.

However, with increased familiarity with the procedure, and having removed the task of feeling the rhythm while walking, the confidence for most participants increased and they walked with longer stride lengths (compared to their baseline) in the post condition when asked to walk to the rhythm from memory (Figure 9).

Conclusions

This paper presents a detailed empirical study of gait rehabilitation using haptic entrainment.

This study appears to be the first systematic study of haptic cueing for gait rehabilitation. Consequently, there is not a lot of previous work available for direct comparison. However, while audio-based rhythmic cueing for rehabilitation is typically conducted differently (eg, using treadmills), previous research on it has typically used similar metrics, such as stride length and stride time [34-36], and it has been associated with a similar degree of improvement [20,36].

Given the great variability of survivors of stroke and the limited number of available participants, there is no claim here of statistical evidence able to support a formal experimental result of improved gait. However, with this participant variability in mind, this empirical gait investigation may be viewed as a set of 11 case studies and more modest empirical claims can be made. At a case study level, improvements were seen in the temporal data of 6 of the 11 study participants.

An overall beneficial decrease in the hemiparetic participants' OTA values was found, indicating an immediate improvement and a more symmetric and healthier gait pattern for 6 out of the 11 participants.

Spatial data were gathered from 6 of the 11 participants using a motion capture system; it showed 3 of the 6 participants with shorter strides with the cue than during their baseline. These spatial changes may be linked to confidence during the walk, with shorter strides indicating a temporary decrease in confidence levels or possibly a higher cognitive load. Temporal asymmetry remained improved during the post condition.

Results from the study indicate some short-term rhythm persistence in memory (post-cue condition) and entrainment to the rhythm from memory. This is not the first time this phenomenon was observed [26,27]; however, it is the first time it has been observed in the context of gait entrainment using haptically presented rhythms. This rhythm-persistence phenomenon may have significant implications for certain patients with hemiparesis who suffer from attention deficits that make it challenging to perceive the rhythm and walk at the same time. This suggests a range of studies to investigate this phenomenon in more depth in the future. This study has also shown empirically that improvements to spatiotemporal gait characteristics can occur immediately after haptic cueing, leading to a more symmetric and healthier gait pattern for a substantial portion of participants. However, this was a lab-based study, with participants with hemiparesis receiving limited exposure to the haptic rhythm. The effects of longer exposure to a haptic rhythm for gait rehabilitation in a home setting are currently unknown. A longitudinal study can be designed to investigate the effects of the haptic rhythm on patients with hemiparetic gait over longer exposure times. In addition, a longitudinal study could also include possible usability metrics (eg, increased confidence, technology acceptance) that were observed in this study but not formally investigated.

Finally, the approach of rhythmic haptic cueing can be extended to include other neurological conditions that lead to motor and gait deficits. Literature in rhythmic auditory stimulation provides evidence of the benefits of entrainment for neurological conditions such as Parkinson disease, cerebral palsy, and Huntington disease. An approach similar to the one adopted in this study for survivors of hemiparetic stroke and brain injury could be applied to investigate these other conditions. An initial case study with a single participant living with Huntington disease provided some encouraging results [37].

Acknowledgments

We would like to thank our participants, physiotherapists, and health professionals who unreservedly donated their time to help us in our research. Dr Theodoros Georgiou is funded by The Open University. Access to some participants and partial data for one study was assisted by funding from the Greater Manchester Academic Health Science Network. Favorable ethical approval was granted by the Open University Human Research Ethics Committee (HREC/2015/2551/Georgiou/1) and the Manchester Metropolitan University Ethics Committee (Ref: 1368). The study was carried out in accordance with the principles laid out by the Declaration of Helsinki.

Conflicts of Interest

None declared.

References

1. Mercer VS, Freburger JK, Chang SH, Purser JL. Measurement of paretic-lower-extremity loading and weight transfer after stroke. Phys Ther 2009 Jul;89(7):653-664 [FREE Full text] [doi: 10.2522/ptj.20080230] [Medline: 19465370]
2. Bohannon RW, Larkin PA. Lower extremity weight bearing under various standing conditions in independently ambulatory patients with hemiparesis. Phys Ther 1985 Sep;65(9):1323-1325. [doi: 10.1093/ptj/65.9.1323] [Medline: 4034666]
3. Min DG, Lee JH, Choe HS, Kim EJ, Shin SH, Lee JH. Comparison of bone density on the dominant and nondominant sides between healthy elderly individuals and stroke patients. J Phys Ther Sci 2016 Sep;28(9):2533-2536 [FREE Full text] [doi: 10.1589/jpts.28.2533] [Medline: 27799687]
4. Verghese J, Ambrose AF, Lipton RB, Wang C. Neurological gait abnormalities and risk of falls in older adults. J Neurol 2010 Mar 26;257(3):392-398 [FREE Full text] [doi: 10.1007/s00415-009-5332-y] [Medline: 19784714]
5. Thaut M, McIntosh G, Rice R. Rhythmic facilitation of gait training in hemiparetic stroke rehabilitation. Journal of the Neurological Sciences 1997 Oct;151(2):207-212. [doi: 10.1016/s0022-510x(97)00146-9]
6. Prassas S, Thaut M, McIntosh G, Rice R. Effect of auditory rhythmic cuing on gait kinematic parameters of stroke patients. Gait & Posture 1997 Dec;6(3):218-223. [doi: 10.1016/s0966-6362(97)00010-6]
7. Roerdink M, Lamoth CJC, Kwakkel G, van Wieringen PCW, Beek PJ. Gait coordination after stroke: benefits of acoustically paced treadmill walking. Phys Ther 2007 Aug;87(8):1009-1022. [doi: 10.2522/ptj.20050394] [Medline: 17553922]
8. Wright RL, Masood A, MacCormac ES, Pratt D, Sackley CM, Wing AM. Metronome-Cued Stepping in Place after Hemiparetic Stroke: Comparison of a One- and Two-Tone Beat. ISRN Rehabilitation 2013;2013:1-5. [doi: 10.1155/2013/157410]
9. Thaut M, Leins A, Rice R, Argstatter H, Kenyon G, McIntosh G, et al. Rhythmic Auditor y Stimulation Improves Gait More Than NDT/Bobath Training in Near-Ambulatory Patients Early Poststroke: A Single-Blind, Randomized Trial. Neurorehabil Neural Repair 2007 Mar 16;21(5):455-459. [doi: 10.1177/1545968307300523]
10. Roerdink M, Bank PJ, Peper CE, Beek PJ. Walking to the beat of different drums: Practical implications for the use of acoustic rhythms in gait rehabilitation. Gait & Posture 2011 Apr;33(4):690-694. [doi: 10.1016/j.gaitpost.2011.03.001]
11. Holland S, Wright R, Wing A, Crevoisier T, Hödl O, Canelli M. A gait rehabilitation pilot study using tactile cueing following hemiparetic stroke. 2014 May Presented at: The 8th International Conference on Pervasive Computing Technologies for Healthcare; 2014; Oldenburg, Germany p. 402-405. [doi: 10.4108/icst.pervasivehealth.2014.255357]
12. Georgiou T, Holland S, van der Linden J, Tetley J, Stockley R, Donaldson G, et al. A blended user centred design study for wearable haptic gait rehabilitation following hemiparetic stroke. 2015 May Presented at: The 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth); 2015; Istanbul, Turkey.
13. Albright TD, Jessell TM, Kandel ER, Posner MI. Cell and Molecular Biology of the Neuron. In: Kandel ER, Schwartz JH, Jessell TM, Siegelbaum SA, Hudspeth AJ, editors. Principles of Neural Science. New York: McGraw-Hill; 2012.
14. Georgiou T. Rhythmic Haptic Cueing for Gait Rehabilitation of Hemiparetic Stroke and Brain Injury Survivors. Milton Keynes: The Open University; 2018.
15. Patterson KK, Parafianowicz I, Danells CJ, Closson V, Verrier MC, Staines WR, et al. Gait Asymmetry in Community-Ambulating Stroke Survivors. Archives of Physical Medicine and Rehabilitation 2008 Feb;89(2):304-310. [doi: 10.1016/j.apmr.2007.08.142]
16. Williams G. Rivermead Mobility Index. In: Kreutzer JC, DeLuca J, Caplan B, editors. Encyclopedia of Clinical Neuropsychology. New York: Springer; 2011:2186.

17. Kobinata N, Ueno M, Imanishi Y, Yoshikawa H. Immediate effects of rhythmic auditory stimulation on gait in stroke patients in relation to the lesion site. J Phys Ther Sci 2016 Sep;28(9):2441-2444 [FREE Full text] [doi: 10.1589/jpts.28.2441] [Medline: 27799666]
18. Podsiadlo D, Richardson S. J Am Geriatr Soc 1991 Feb 27;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
19. Bouwer A, Holland S, Dalgleish M. The Haptic Bracelets: Learning Multi-Limb Rhythm Skills from Haptic Stimuli While Reading. In: Holland S, Wilkie K, Mulholland P, Seago A, editors. Music and Human-Computer Interaction. London: Springer; 2013:101-122.
20. Thaut C, Rice R. Rhythmic auditory stimulation. In: Thaut MH, Hoemberg V, editors. Handbook of neurologic music therapy. Oxford: Oxford University Press; 2016:95-105.
21. Roerdink M, Lamoth CJC, van Kordelaar J, Elich P, Konijnenbelt M, Kwakkel G, et al. Rhythm perturbations in acoustically paced treadmill walking after stroke. Neurorehabil Neural Repair 2009 Sep 23;23(7):668-678. [doi: 10.1177/1545968309332879] [Medline: 19307435]
22. Clayton M, Sager R, Will U. In time with the music: the concept of entrainment and its significance for ethnomusicology. European Meetings in Ethnomusicology 2005;11:3-142.
23. Angelis V, Holland S, Upton PJ, Clayton M. Testing a Computational Model of Rhythm Perception Using Polyrhythmic Stimuli. Journal of New Music Research 2013 Mar;42(1):47-60. [doi: 10.1080/09298215.2012.718791]
24. Thaut MH. Neurologic Music Therapy in Sensorimotor Rehabilitation. In: Thaut MH, editor. Rhythm, Music, and the Brain: Scientific Foundations and Clinical Applications. London: Routledge; 2005:137-164.
25. Merker B. Groove or swing as distributed rhythmic consonance: introducing the groove matrix. Front Hum Neurosci 2014 Jun 23;8:454 [FREE Full text] [doi: 10.3389/fnhum.2014.00454] [Medline: 25002843]
26. Thaut MH, McIntosh KW, McIntosh GC, Hoemberg V. Auditory rhythmicity enhances movement and speech motor control in patients with Parkinson's disease. Funct Neurol 2001;16(2):163-172. [Medline: 11495422]
27. Nieuwboer A, Kwakkel G, Rochester L, Jones D, van Wegen E, Willems AM, et al. Cueing training in the home improves gait-related mobility in Parkinson's disease: the RESCUE trial. J Neurol Neurosurg Psychiatry 2007 Feb;78(2):134-140 [FREE Full text] [doi: 10.1136/jnnp.200X.097923] [Medline: 17229744]
28. Nieuwboer A. Cueing for freezing of gait in patients with Parkinson's disease: a rehabilitation perspective. Mov Disord 2008 Jul 30;23 Suppl 2(S2):S475-S481. [doi: 10.1002/mds.21978] [Medline: 18668619]
29. Holland S, Hödl O. Haptics for the Development of Fundamental Rhythm Skills, Including Multi-limb Coordination. In: Papeti S, Saitis C, editors. Musical Haptics. London: Springer; 2018:238.
30. Qualisys. Qualisys: Technical specification. Qualisys Hardware. URL: https://www.qualisys.com/hardware/ [accessed 2020-05-17]
31. Patterson KK, Gage WH, Brooks D, Black SE, McIlroy WE. Evaluation of gait symmetry after stroke: a comparison of current methods and recommendations for standardization. Gait Posture 2010 Feb;31(2):241-246. [doi: 10.1016/j.gaitpost.2009.10.014] [Medline: 19932621]
32. Thaut MH, Abiru M. Rhythmic Auditory Stimulation in Rehabilitation of Movement Disorders: A Review Of Current Research. Music Perception 2010;27(4):263-269. [doi: 10.1525/mp.2010.27.4.263]
33. Krasovsky T, Levin MF. Review: Toward a Better Understanding of Coordination in Healthy and Poststroke Gait. Neurorehabil Neural Repair 2009 Oct 12;24(3):213-224. [doi: 10.1177/1545968309348509]
34. Roerdink M, Lamoth CJC, van Kordelaar J, Elich P, Konijnenbelt M, Kwakkel G, et al. Rhythm perturbations in acoustically paced treadmill walking after stroke. Neurorehabil Neural Repair 2009 Sep;23(7):668-678. [doi: 10.1177/1545968309332879] [Medline: 19307435]
35. Pelton TA, Johannsen L, Huiya Chen, Wing AM. Hemiparetic stepping to the beat: asymmetric response to metronome phase shift during treadmill gait. Neurorehabil Neural Repair 2010 Jun;24(5):428-434. [doi: 10.1177/1545968309353608] [Medline: 19952366]
36. Wright RL, Bevins JW, Pratt D, Sackley CM, Wing AM. Metronome Cueing of Walking Reduces Gait Variability after a Cerebellar Stroke. Front Neurol 2016 Jun 01;7:84 [FREE Full text] [doi: 10.3389/fneur.2016.00084] [Medline: 27313563]
37. Georgiou T, Islam R, Holland S, van der Linden J, Price B, Mulholland P, et al. Rhythmic Haptic Cueing Using Wearable Devices as Physiotherapy for Huntington Disease: Case Study. JMIR Rehabil Assist Technol 2020 Sep 14;7(2):e18589 [FREE Full text] [doi: 10.2196/18589] [Medline: 32924955]

Abbreviations

IMU: inertia monitoring unit
OTA: overall temporal asymmetry
RAS: rhythmic auditory stimulation
RMA: Rivermead Motor Assessment

Edited by G Eysenbach; submitted 10.03.20; peer-reviewed by L Shum; comments to author 27.04.20; revised version received 19.06.20; accepted 11.11.20; published 24.11.20.

Please cite as:
Georgiou T, Holland S, van der Linden J
Rhythmic Haptic Cueing for Gait Rehabilitation of People With Hemiparesis: Quantitative Gait Study
JMIR Biomed Eng 2020;5(1):e18649
URL: http://biomedeng.jmir.org/2020/1/e18649/
doi:10.2196/18649
PMID:

Original Paper

Longitudinal Magnetic Resonance Imaging as a Potential Correlate in the Diagnosis of Alzheimer Disease: Exploratory Data Analysis

Afreen Khan[1*], MCA; Swaleha Zubair[1*], PhD

Aligarh Muslim University, Aligarh, India

[*]all authors contributed equally

Corresponding Author:
Swaleha Zubair, PhD
Aligarh Muslim University
Department of Computer Science
Adjacent Computer Centre
Anoopshahr Road
Aligarh, 202002
India
Phone: 91 9410059635
Email: swalehaowais123@gmail.com

Abstract

Background: Alzheimer disease (AD) is a degenerative progressive brain disorder where symptoms of dementia and cognitive impairment intensify over time. Numerous factors exist that may or may not be related to the lifestyle of a patient that result in a higher risk for AD. Diagnosing the disorder in its beginning period is important, and several techniques are used to diagnose AD. A number of studies have been conducted on the detection and diagnosis of AD. This paper reports the empirical study performed on the longitudinal-based magnetic resonance imaging (MRI) Open Access Series of Brain Imaging dataset. Furthermore, the study highlights several factors that influence the prediction of AD.

Objective: This study aimed to correlate the effect of various factors such as age, gender, education, and socioeconomic background of patients with the development of AD. The effect of patient-related factors on the severity of AD was assessed on the basis of MRI features, Mini-Mental State Examination (MMSE), Clinical Dementia Rating (CDR), estimated total intracranial volume (eTIV), normalized whole brain volume (nWBV), and Atlas Scaling Factor (ASF).

Methods: In this study, we attempted to establish the role of longitudinal MRI in an exploratory data analysis (EDA) of AD patients. EDA was performed on the dataset of 150 patients for 343 MRI sessions (mean age 77.01 [SD 7.64] years). The T1-weighted MRI of each subject on a 1.5-Tesla Vision (Siemens) scanner was used for image acquisition. Scores of three features, MMSE, CDR, and ASF, were used to characterize the AD patients included in this study. We assessed the role of various features (ie, age, gender, education, socioeconomic status, MMSE, CDR, eTIV, nWBV, and ASF) on the prognosis of AD.

Results: The analysis further establishes the role of gender in the prevalence and development of AD in older people. Moreover, a considerable relationship has been observed between education and socioeconomic position on the progression of AD. Also, outliers and linearity of each feature were determined to rule out the extreme values in measuring the skewness. The differences in nWBV between CDR=0 (nondemented), CDR=0.5 (very mild dementia), and CDR=1 (mild dementia) are significant (ie, $P<.01$).

Conclusions: A substantial correlation has been observed between the pattern and other related features of longitudinal MRI data that can significantly assist in the diagnosis and determination of AD in older patients.

(JMIR Biomed Eng 2020;5(1):e14389) doi:10.2196/14389

KEYWORDS

Alzheimer disease; dementia; longitudinal; magnetic resonance imaging; exploratory data analysis; Mini-Mental State Examination; Clinical Dementia Rating; Atlas Scaling Factor

Introduction

Alzheimer disease (AD) is a degenerative brain ailment characterized by the development of dementia and other related cognitive impairments [1-3]. It is a heterogeneous, irreversible neurodegenerative disorder that may find an association with genetic complexity in the individual. The Alzheimer's Association describes dementia as a syndrome comprising a cluster of symptoms that encompass several features including age, gender, education, and the Mini-Mental State Examination (MMSE) of the inflicted patients [4].

There has been a significant increase in the number of AD cases in recent years. It has been reported that it is the sixth most diagnosed disease in the Unites States. As of 2018, 5.7 million Americans of all ages have been diagnosed with AD [4]. Approximately 44 million people worldwide are living with AD or an associated kind of dementia [5].

With the advancement of technology pertaining to treatment methodologies and development of novel diagnostic tools, many of the modern age diseases are being diagnosed earlier and treated successfully. In contrast, AD still remains a poorly diagnosed ailment with little success in treatment.

In the information technology era, machine and deep learning tools have found a wide scope in medical diagnosis [6]. Although medical expert opinion, disease symptom, and other related data from the patient remain the prime parameters that help in the diagnosis of a particular disease, machine learning predictions, data analytics visualizations, and other artificial intelligence techniques have emerged as alternate ways to predict diseases and help the current state of the medical world in a great way [7,8].

The occurrence of cognitive disorders is a common feature observed in elderly people, and this can be considered a primary indication of a growing dementing syndrome like AD [9]. Individuals with cognitive disorders experience mild cognitive impairment (MCI) [10-12]. Various biomarkers or related parameters may evolve that can help in the diagnosis of AD in patients. Similarly, techniques like magnetic resonance imaging (MRI) studies, positron emission tomography scans, and neurochemical testing of the cerebrospinal fluid can also help in the diagnosis of AD [13,14].

In this study, we systematically examined the distinct and interactive impact of age, gender, education, socioeconomic status (SES), Mini Mental State Examination (MMSE), Clinical Dementia Rating (CDR), estimated total intracranial volume (eTIV), normalized whole brain volume (nWBV), and Atlas Scaling Factor (ASF) on the basis of several longitudinal MRI sessions of various patients. The information was retrieved from the Open Access Series of Brain Imaging (OASIS-2) dataset. We performed exploratory data analysis (EDA) to understand the correlation between various feature sets. Consistent with the literature, we predicted that men were more likely to be diagnosed with AD compared with women. The gender bias can be correlated to the dataset dependency. The ε4 allele of the apolipoprotein E gene (APOE-ε4) has also been reported to play a major role in the occurrence of AD. We did not include APOE-ε4 data in the study in order to avoid the complexity. A significant relationship has been observed among educational background and SES of the patients and emergence of dementia. Anomalies and linearity of each of the features were resolved to remove extreme values in determining the skewness.

Methods

Subjects

The dataset used in this study consists of a longitudinal collection of MRI data in demented and nondemented older adults. A total of 150 subjects aged 60 to 96 years participated in 373 MRI sessions. The data included in this study were based on the subjects reported to a longitudinal collection of MRI scans at the Washington University Alzheimer Disease Research Center [15].

The T1-weighted MRI acquisition of each subject was performed on a 1.5 Tesla Vision scanner (Siemens). Related technical details are as follows: sequence of magnetization prepared rapid acquisition gradient echo, repetition time=9.7 msec, echo time=4.0 msec, flip angle=10°, inversion time=20 msec, delay time=200 msec, orientation is sagittal, thickness=1.25 mm, gap=0 mm, slice number=128, and resolution=256×256 (1×1 mm) [15].

Methodology

In this analysis, we cataloged previous EDA. The general objective of the study was to report the relative association between the target group (demented or nondemented) and other features that play a major role in the diagnosis of AD. Furthermore, we examined the risk of AD induction in inflicted patients. We analyzed longitudinal MRI data of both healthy patients and patients with AD [15].

Scoring Rules

In this study, we used the following instruments to determine the state of the healthy versus inflicted brain.

- SES: according to the Hollingshead Index of Social Position, the SES is classified into groups of highest status (1) and lowest status (0) [16]
- MMSE: values range from 0 to 30; 0 to 9 indicates extreme impairment, 10 to 18 demonstrates moderate dementia, 19 to 23 mild dementia, and 24 to 30 is considered normal [17]
- CDR: scored after a semistructured discussion with the patient, with scores ranging from 0 to 3 (ie, 0=none, 0.5=very mild, 1=mild, 2=moderate, 3=extreme dementia) [18]

Experiment Environment

Empirical analysis of the dataset described in this paper was performed using Python libraries conducted on the Jupyter platform of Anaconda Navigator. The Jupyter platform presents a well-defined skeleton for developers to process, develop, and assess their models. Python is an interpreted and high-level programming language comprising dynamic semantics. It includes Seaborn, a visualization library through which statistical graphs can be plotted with the aim of performing univariate and multivariate analyses.

Exploratory Data Analysis

EDA is a data analysis methodology using techniques that are usually graphical. It maximizes understanding of the dataset, reveals underlying structure, detects anomalies and outliers, extracts imperative features, and ascertains ideal factor settings [19]. EDA is not similar to statistical graphics despite the fact that the two terms are used interchangeably. It is a more direct approach that allows the data to reveal the underlying model and its structure [19].

In this study, we focused on establishing a correlation between attributes of MRI tests and patient classification groups. The primary objective of performing this exploratory analysis was to determine the association of data among the features before performing the data analysis or data extraction process. It was supposed to assist in understanding the data subclassification and facilitate choosing the proper analysis technique for the model later.

Dataset Description

The dataset comprised 373 observations and 15 attributes, out of which group was the target variable while the rest were the independent variables in this empirical study. Table 1 provides a description of the dataset attributes.

Figure 1 outlines the dataset attributes in terms of the total count of each attribute for 15 columns on the basis of null/nonnull and data type of respective attributes. It can be seen from the figure that SES and MMSE consist of values less than the total 373 MRI sessions, marked by the red right bracket in the figure. This is what missing values relates to. The rest of the features, marked by the blue right brackets, do not contain any missing values (ie, for the total 373 sessions, all recorded MRI features emerged as nonnull and without any missing value).

The *P* values used for comparison in the study are shown in Table 2.

Table 1. Detail of dataset attributes.

Number	Attribute name	Attribute description
1	Subject ID[a]	Patient's identification number
2	MRI[b] ID	Patient's imaging identification number
3	Group	Demented, nondemented, or converted
4	Visit	Number of visits of each patient
5	MR[c] Delay	Magnetic resonance delay is the delay time given before the image procurement in real time
6	M/F[d]	Patient's gender
7	Hand	Right-handed or left-handed
8	Age	Patient's age at the scanning
9	EDUC[e]	Educational level of the patient
10	SES[f]	Socioeconomic status of the patient
11	MMSE[g]	Mini-Mental State Examination score
12	CDR[h]	Clinical Dementia Rate score
13	eTIV[i]	estimated total intracranial volume result
14	nWBV[j]	normalized whole brain volume result
15	ASF[k]	Atlas Scaling Factor

[a]ID: identification.
[b]MRI: magnetic resonance imaging.
[c]MR: magnetic resonance.
[d]M/F: male or female.
[e]EDUC: educational level of the patient.
[f]SES: socioeconomic status.
[g]MMSE: Mini-Mental State Examination.
[h]CDR: Clinical Dementia Rating score.
[i]eTIV: estimated total intracranial volume.
[j]nWBV: normalized whole brain volume.
[k]ASF: Atlas Scaling Factor.

Figure 1. Dataset Information.

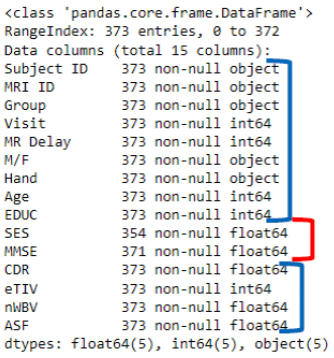


Table 2. *P* value for the corresponding attribute.

Attribute name	*P* value
EDUC: educational level of a patient	<.001
MMSE: Mini Mental State Examination	<.001
CDR: Clinical Dementia Rating	<.01
nWBV: normalized whole brain volume	<.01

Summary Statistics

Statistical information includes count, mean, standard deviation, first quartile, second quartile (median), third quartile, and minimum and maximum values of each attribute as shown in Table 3.

From the data depicted in Table 3, we can infer that the mean value is less than the median on some features and greater than the median value on certain another sets of features. The median value is represented by 50% (50th percentile) in the index column. The median value of each feature aids in the data preprocessing when dealing with the imputation step. There is a large difference in the 75th percentile and maximum values of predictors in MR delay, CDR, and eTIV. The observation suggests the occurrence of extreme values (ie, outliers) in the dataset.

Table 3. Summary statistics of each attribute.

Attribute	Count	Mean (SD)	Min-max[a]	Quartiles		
				25%	50%	75%
Visit	373.00	1.88 (0.92)	1.00-5.00	1.00	2.00	2.00
MR[b] delay	373.00	595.10 (635.49)	0-2639.00	0	552.00	873.00
Age	373.00	77.01 (7.64)	60.00-98.00	71.00	77.00	82.00
EDUC[c]	373.00	14.60 (2.88)	6.00-23.00	12.00	15.00	16.00
SES[d]	354.00	2.46 (1.13)	1.00-5.00	2.00	2.00	3.00
MMSE[e]	371.00	27.34 (3.68)	4.00-30.00	27.00	29.00	30.00
CDR[f]	373.00	0.29 (0.37)	0-2.00	0	0	0.50
eTIV[g]	373.00	1488.13 (176.14)	1106.00-2004.00	1357.00	1470.00	1597.00
nWBV[h]	373.00	0.73 (0.04)	0.64-0.84	0.70	0.73	0.76
ASF[i]	373.00	1.20 (0.14)	0.88-1.59	1.10	1.19	1.29

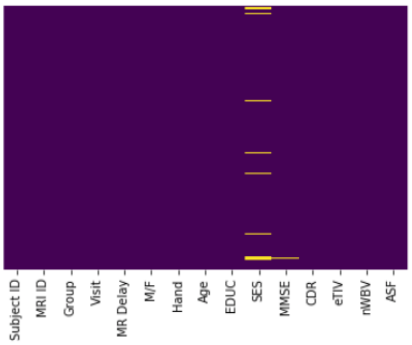
[a]Min-max: minimum and maximum values.
[b]MR: magnetic resonance.
[c]EDUC: educational level of the patient.
[d]SES: socioeconomic status.
[e]MMSE: Mini-Mental State Examination.
[f]CDR: Clinical Dementia Rating.
[g]eTIV: estimated total intracranial volume.
[h]nWBV: normalized whole brain volume.
[i]ASF: Atlas Scaling Factor.

Data Exploration

Initially, the dataset consisted of 373 MRI sessions out of which there were nondemented (n=72), demented (n=64), and converted patients (n=14). On the first visit, patients were grouped as nondemented and were categorized as demented at a later visit. The 14 converted patients are those patients which were found to be nondemented in the first visit, but in their second and third visits, they were diagnosed with dementia. Therefore, only the subjects of the first visit are being considered throughout the study, and total of 150 subjects have been explored under this analysis.

The dataset consists of many missing values (ie, some of the rows of certain attributes consist of no value, which is determined during the EDA step). To locate exactly which column comprises missing values, a heat map is plotted for all 373 MRI sessions initially, consisting of all the patient visits (Figure 2A). The SES and MMSE columns contain missing values (represented by yellow lines on a purple background). Figure 2B delineates the count of missing values in numeric form for all attributes. Figures 3A and 3B highlight the heat map and count of missing values for the 150 subjects for visit 1. SES is the only feature that consists of 8 missing values, while the rest of the features have all values filled.

Figure 2. Illustration of missing values for 373 magnetic resonance imaging sessions for all patient visits.

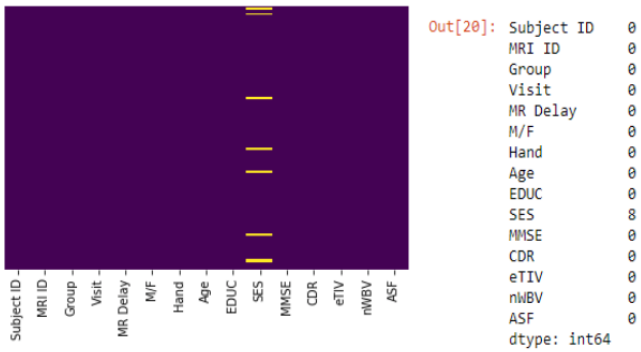


(a) Heat map for missing values

```
Out[10]: Subject ID     0
         MRI ID         0
         Group          0
         Visit          0
         MR Delay       0
         M/F            0
         Hand           0
         Age            0
         EDUC           0
         SES           19
         MMSE           2
         CDR            0
         eTIV           0
         nWBV           0
         ASF            0
         dtype: int64
```

(b) Count of missing values

Figure 3. Outline of missing values for 150 patients for the first visit.



(a) Heat map for missing values (b) Count of missing values

Results

In this section, the results of the EDA are reported. Subsequent to applying the preprocessing and data preparation strategies, we attempted to break down the data outwardly and make sense of the dispersion of features as far as adequacy and effectiveness are concerned. By breaking down data, we have tried to make it more simple and meaningful. This helped in increasing the efficiency of the analysis.

Patient Demographic Profiles

The study comprised 62 males and 88 females within the age range of 60 to 96 years. Table 4 illustrates the demographic summary of patients who were examined for AD.

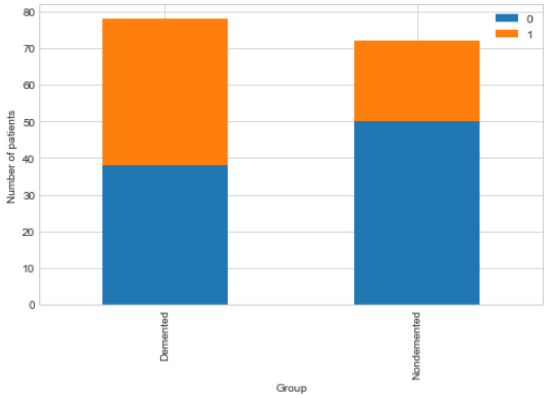
Table 4. Demographic profile of the study population (n=150).

Characteristic	Values
Gender, n (%)	
Male	62 (41.3)
Female	88 (58.7)
Age in years, mean (SD)	77.01 (7.64)
Age in years, median	77

Gender and Demented Proportion

The bar chart as demonstrated in Figure 4 confirms that men are more prone to dementia than women. The blue color, coded as 0, represents female, while the orange color, coded as 1, represents male. Of the 150 patients, 78 are in the demented category. Of the 78 demented patients, 40 are male.

Figure 4. Gender and demented proportion (female=0, male=1).

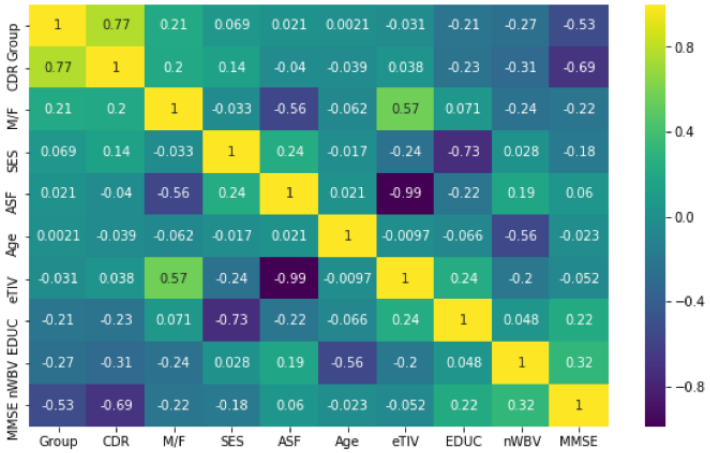


Correlation Matrix With a Heat Map

In order to build the model, an essential condition is to eliminate the correlated variables. Correlations were obtained by applying the Python Pandas corr() function, which aided us in visualizing the correlation grid built using a heat map.

The correlation matrix with heat map is illustrated in Figure 5. The dark shades represent positive correlation while lighter shades represent negative correlation. We exclude the target variable (ie, group) and then checked for the correlated independent variables. Thus we can infer that eTIV has a strong positive correlation with male/female (M/F) whereas it has a strong negative correlation with ASF among all.

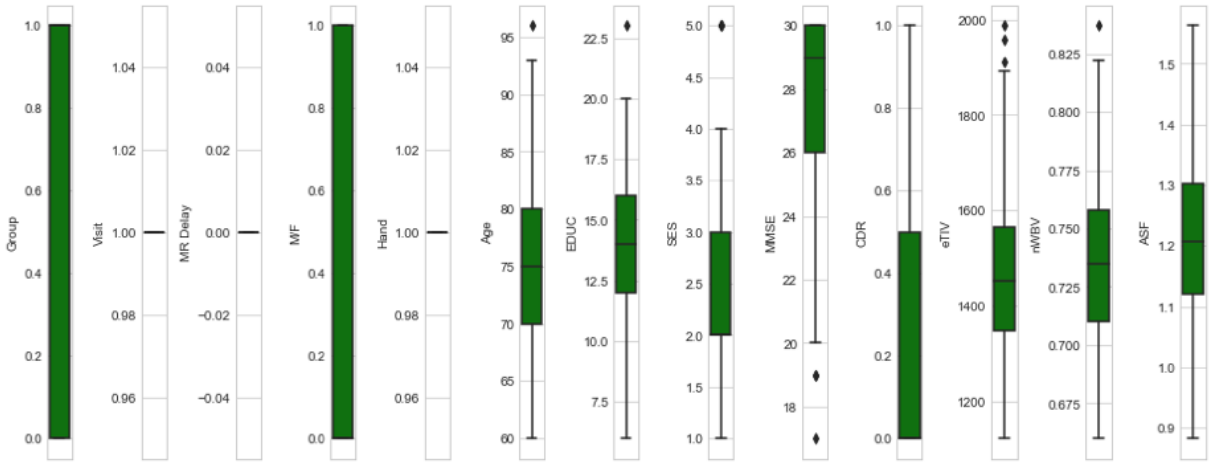
Figure 5. Heat map illustrating the correlations among the dataset features.



Outliers Check With Box-Whisker Plot

A box-whisker plot displays the spread of quantitative data in a manner that facilitates comparisons between attributes. In Figure 6, the box illustrates the dataset's quartiles whereas the whiskers stretch out to demonstrate whatever remains of the dispersion. The box-whisker schema is a standardized method for displaying the data distribution, which is dependent on 5 major aspects: minimum value, first quartile, median value (second quartile), third quartile, and maximum value. The middle rectangle traverses the first quartile to the third quartile, known as interquartile range (IQR). A fragment inside the rectangle demonstrates the median value. Whiskers above and beneath the rectangle demonstrate the areas of the minimum and greatest value. Outliers are either 3×IQR or progressively over the third quartile or 3×IQR or more beneath the first quartile. Thus, we can infer from Figure 6 that age, patient education level (EDUC), SES, MMSE, eTIV, and nWBV feature columns show outliers.

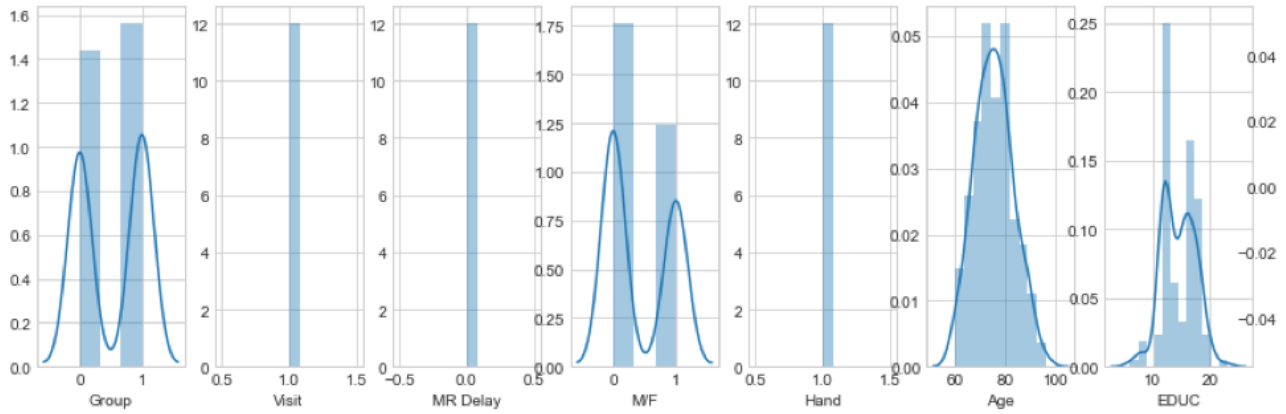
Figure 6. Box-whisker plot demonstrating outliers.



Skewness and Distribution Plot

The linearity of the attributes was determined by plotting a distribution graph. The graph was used to study the skewness of both the target variable and the independent variables. From Figure 7, it can be concluded that group, visit, MR delay, M/F, hand, and age feature columns appear to be normally distributed while all the remaining independent variables are discovered to be experiencing skewness.

Figure 7. Distribution plot of the dataset features.

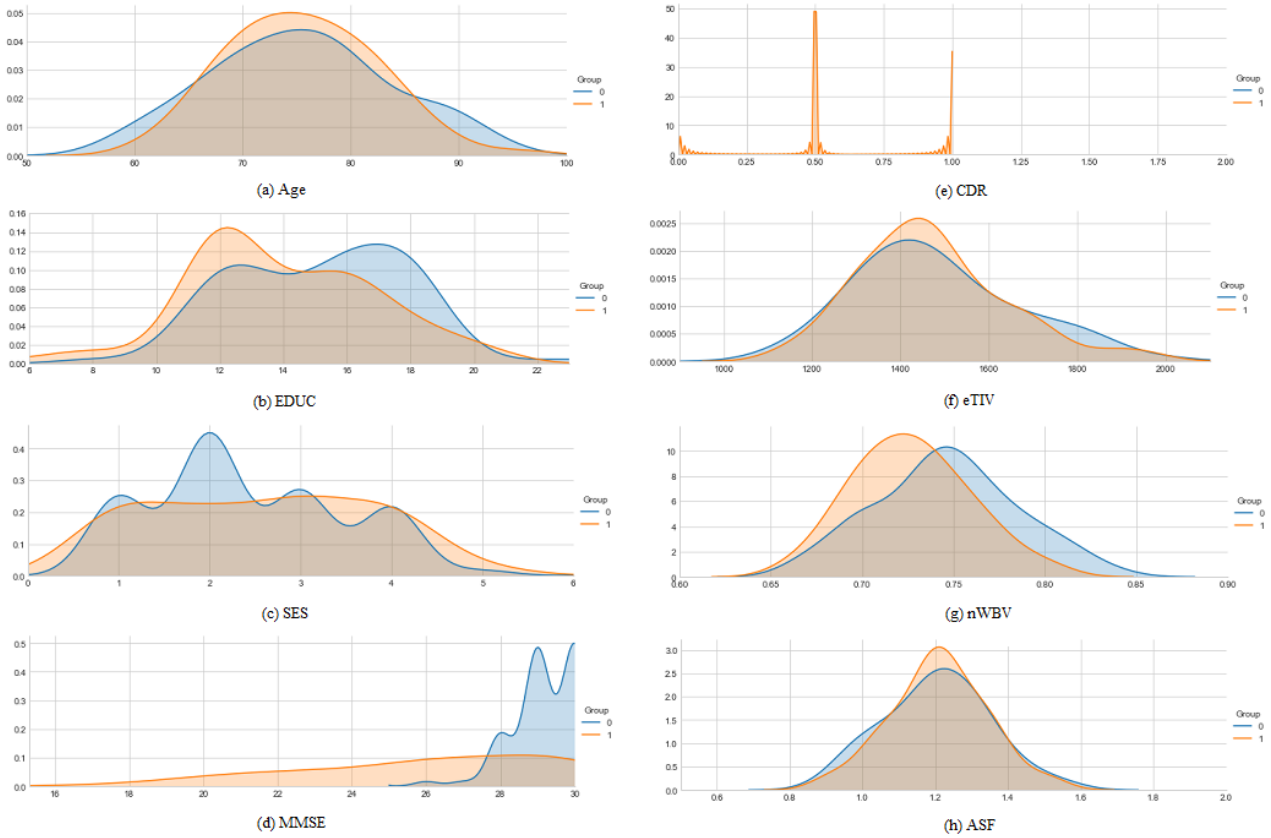


Effect of Independent Variable on Dependent Variable

A graph was plotted between the target variable (ie, group: demented/nondemented) and independent variables. We plotted 8 such graphs, for age, EDUC, MMSE, ASF, eTIV, nWBV, SES, and CDR, shown in Figure 8.

The following features were inferred: (1) age: between 60 and 90 years; (2) EDUC: demented patients were less educated; (3) SES: considerable increment in the prevalence of dementia as we move from highest status to lowest status; (4) MMSE: nondemented group got much higher MMSE scores than the demented group; (5) CDR: more individuals with a score of 0.5 (ie, very mild dementia), fewer individuals with a score of 1 (ie, mild dementia), and very few with a score of 0 (ie, no dementia); (6) eTIV: higher for demented patients; (7) nWBV: nondemented group has higher brain volume ratio than demented group; and (8) ASF: demented patients have higher score than nondemented ones. The differences in nWBV between CDR=0 (nondemented), CDR=0.5 (very mild dementia), and CDR=1 (mild dementia) comes out to be significant (ie, $P<.01$).

Figure 8. A plot between the target variable and each independent variable (nondemented=0, demented=1).

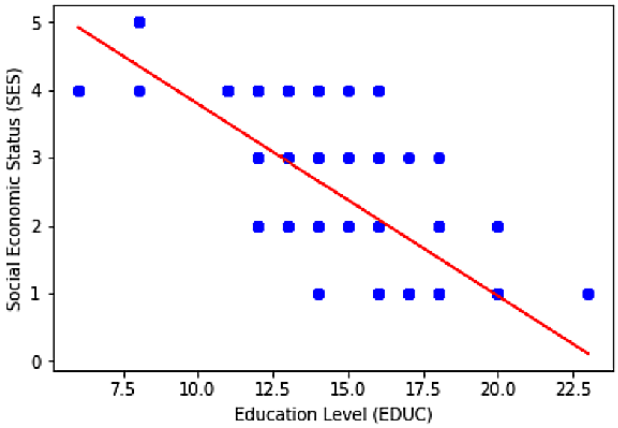


Impact of Socioeconomic Status and Education Level in the Demented Group

The relationship between SES and EDUC on dementia can be inferred from Figure 9, which shows that individuals with the highest status (1) exhibit higher education levels while individuals with the lowest status (5) exhibit lower education level. Thus, years of education have an immense effect on dementia. The scatter plot with linear regression lines for SES

and EDUC display a positive correlation among EDUC and SES.

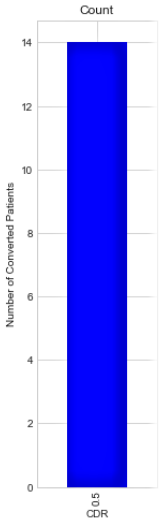
Figure 9. Scatter plot for socioeconomic status and level of education.



Correlation Between Converted Patients and Clinical Dementia Rating

The data shown in Table 3 suggest that 14 patients converted. These patients were earlier classified as nondemented and in a later visit found to have dementia. We tended to draw a relationship among these 14 converted patients with their respective CDR values on subsequent (second and third) visits. For developing a correlation between dementia and other related factors, we focused on changes incurred in CDR values. For earlier visits, it was 0.0, signifying that the patient was nondemented, while at a later visit, it changed to 0.5, indicating the patient had very mild dementia. Figure 10 shows a correlation between converted patients with their CDR values.

Figure 10. Distribution plot for converted patients and their Clinical Dementia Rating value.



Discussion

Principal Findings

This study provides an understanding of attributes related to AD in older adults. We observed that men are more likely to have AD compared with women. There are several major differences that frequently appear between men and women in the occurrence, presentation, and development of psychiatric disorders [20]. Earlier studies suggested that women are more prone to develop AD since they are at greater risk of depression compared with men [21]. The genetic factor APOE-ε4 has also been reported to affect men and women differently [21]. Riedel et al [22] stated that age, APOE-ε4, and sex are the most serious risk factors in the development of AD. Further, the rate of AD is practically identical in women and men until late age when the frequency becomes more prominent in women [22].

We performed an empirical analysis on the dataset comprising longitudinally obtained T1-weighted MRI data of 150 patients aged between 60 to 96 years. Among the 15 studied features, we found that only gender, age, educational years, SES, MMSE, CDR, eTIV, and nWBV were significantly associated with making an impact on the occurrence of AD in both demented and nondemented subjects. Our analysis demonstrated that patients aged between 70 and 90 years exhibit a higher clustering of dementia than nondemented patients. Since AD has a lower survival rate, it is the reason why data available in the aged patient is scarce. All patients examined were right-handed, thus handedness doesn't have an effect in this analysis. Of the 150

patients, demented patients were found to be less educated compared with nondemented patients (Figure 8B).

We found an independent link between various features in both demented and nondemented groups and found that there were numerous correlated indicators of AD. Unfortunately, this study lacks an adequate feature set that could have helped in uncovering related associations efficiently.

We observed that over the change from higher (score 1) to lower (score 5) SES, there was a considerable decrease in the prevalence of dementia. In general, education has been found to be directly associated with SES. In fact, there seems to be a high to moderate level of association between education and occupation-based SES [23]. Social epidemiology relates education with SES by defining education as "the transition from a socioeconomic position largely received from parents to an achieved socioeconomic position as an adult" [16]. Various components of SES, viz education, income and occupational status, can influence AD development in the aged patients [24].

The MMSE, a complete measure of cognitive impairment, has been widely used in the detection of AD. Arevalo-Rodriguez et al [25] performed an analysis to determine the MMSE accuracy for the detection of AD in people with mild MCI. In fact, the MMSE score cannot aid in categorizing people as demented or nondemented [25]. In contrast to this, we identified that the nondemented study group got a much higher MMSE score than the demented group.

The scoring of the CDR have been widely used in clinical trials and longitudinal studies to determine the state of dementia [18]. We found the CDR peaks at 0.5 (very mild dementia), followed by 1 (mild dementia) and 0 (no dementia). Unsurprisingly, our results are in agreement with those illustrated by Marcus et al [15], which states that patients who were categorized to be nondemented in the first visit were found to be demented in later visits with a CDR of greater than 0.

The plot for eTIV summarizing various data shows that demented patients have more eTIV compared with nondemented patients (Figure 8F). The intracranial volume, describing brain size, is found to be less in AD patients. Earlier, Tate et al [26] reported that there were certain patients for which the total intracranial volume emerged to have an impact on dementia prediction when the data were examined in a nonparametric manner.

In line with an earlier study using a subset of the data [27], we found that the nondemented group had a higher nWBV than the demented group. This could be attributed to the fact that AD may lead to shrinkage of neuronal tissues of the brain. Marcus et al [15] exploited nWBV as an approach to evaluate the anatomical features of the brain to determine the level of dementia. Several other studies suggested that nWBV declines upon advancement of AD stage and growing age of the patients [27-31].

Our findings suggest that demented patients have a higher ASF when compared with nondemented ones. The scaling factor changes the skull and native-space brain to the atlas target, which is determined by calculating the determinant of the transform matrix [32].

On the basis of data analysis, we infer that there was no correlation between the repeated measures. In longitudinal data analysis, it seems to be an easy and straightforward approach but an unrealistic alternative. To this end, we can justify it as a fair approach to assess the relationship among covariates irrespective of the visits. This structure was chosen at the commencement of the analysis, and we suggested that it bears a resemblance to the experimental correlations for improved estimate of standard errors.

Limitation

More feature set brain mapping is required to strengthen the robustness of the results and discover the causal methods underlying the relation between distinct features of both longitudinal and cross-sectional MRI data and the consequence on the late-life health.

Conclusion and Future Work

This study highlights the relationship between the target and the independent features in MRI sessions of AD patients. It can be argued that whatever effect the independent features have on the prediction of the target variable (demented/nondemented), it is unlikely to be dependent on the sample size relationship. We infer that men are more likely to suffer from AD than women. The study also finds that attributes such as eTIV, nWBV, and ASF have a greater correlation in the prevalence of AD in women compared with men. Finally, we conclude that imaging biomarkers play a major role in the diagnosis of AD.

Acknowledgments

The data used in the preparation of this article were obtained from the OASIS database [33], made available by the Washington University Alzheimer Disease Research Center. Longitudinal MRI data were retrieved from the following published NIH grants: P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, and R01 MH56584.

Conflicts of Interest

None declared.

References

1. Wilson RS, Segawa E, Boyle PA, Anagnos SE, Hizel LP, Bennett DA. The natural history of cognitive decline in Alzheimer's disease. Psychol Aging 2012 Dec;27(4):1008-1017 [FREE Full text] [doi: 10.1037/a0029857] [Medline: 22946521]

2. Barker WW, Luis CA, Kashuba A, Luis M, Harwood DG, Loewenstein D, et al. Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. Alzheimer Dis Assoc Disord 2002;16(4):203-212. [doi: 10.1097/00002093-200210000-00001] [Medline: 12468894]
3. Kim H. Understanding internet use among dementia caregivers: results of secondary data analysis using the us caregiver survey data. Interact J Med Res 2015;4(1):e1 [FREE Full text] [doi: 10.2196/ijmr.3127] [Medline: 25707033]
4. 2018 Alzheimer's disease facts and figures. URL: https://www.alz.org/media/homeoffice/facts%20and%20figures/facts-and-figures.pdf [accessed 2019-02-17]
5. Alzheimer’s disease statistics. 2018. URL: https://alzheimersnewstoday.com/alzheimers-disease-statistics/ [accessed 2019-02-19]
6. Cleret de Langavant L, Bayen E, Yaffe K. Unsupervised machine learning to identify high likelihood of dementia in population-based surveys: development and validation study. J Med Internet Res 2018 Dec 09;20(7):e10493 [FREE Full text] [doi: 10.2196/10493] [Medline: 29986849]
7. Farhan W, Wang Z, Huang Y, Wang S, Wang F, Jiang X. A predictive model for medical events based on contextual embedding of temporal sequences. JMIR Med Inform 2016 Nov 25;4(4):e39 [FREE Full text] [doi: 10.2196/medinform.5977] [Medline: 27888170]
8. Celi LA, Davidzon G, Johnson AE, Komorowski M, Marshall DC, Nair SS, et al. Bridging the health data divide. J Med Internet Res 2016 Dec 20;18(12):e325 [FREE Full text] [doi: 10.2196/jmir.6400] [Medline: 27998877]
9. Khan A, Zubair S. Machine learning tools and toolkits in the exploration of big data. ijcse 2018 Dec 31;6(12):570-575. [doi: 10.26438/ijcse/v6i12.570575]
10. Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001 May 08;56(9):1133-1142. [doi: 10.1212/wnl.56.9.1133] [Medline: 11342677]
11. Maroco J, Silva D, Rodrigues A, Guerreiro M, Santana I, de Mendonça A. Data mining methods in the prediction of dementia: a real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. BMC Res Notes 2011 Aug 17;4:299 [FREE Full text] [doi: 10.1186/1756-0500-4-299] [Medline: 21849043]
12. Bott N, Kumar S, Krebs C, Glenn JM, Madero EN, Juusola JL. A remote intervention to prevent or delay cognitive impairment in older adults: design, recruitment, and baseline characteristics of the Virtual Cognitive Health (VC Health) study. JMIR Res Protoc 2018 Aug 13;7(8):e11368 [FREE Full text] [doi: 10.2196/11368] [Medline: 30104186]
13. Portet F, Ousset PJ, Visser PJ, Frisoni GB, Nobili F, Scheltens P, MCI Working Group of the European Consortium on Alzheimer's Disease (EADC). Mild cognitive impairment (MCI) in medical practice: a critical review of the concept and new diagnostic procedure. Report of the MCI Working Group of the European Consortium on Alzheimer's Disease. J Neurol Neurosurg Psychiatry 2006 Jun;77(6):714-718 [FREE Full text] [doi: 10.1136/jnnp.2005.085332] [Medline: 16549412]
14. Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol 2007 Aug;6(8):734-746. [doi: 10.1016/S1474-4422(07)70178-3] [Medline: 17616482]
15. Marcus DS, Fotenos AF, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. J Cogn Neurosci 2010 Dec;22(12):2677-2684 [FREE Full text] [doi: 10.1162/jocn.2009.21407] [Medline: 19929323]
16. Lynch J, Kaplan G. Socioeconomic position. In: Berkman L, Kawachi I, editors. Social Epidemiology. New York: Oxford University Press; 2000:13-35.
17. Magni E, Binetti G, Padovani A, Cappa SF, Bianchetti A, Trabucchi M. The Mini-Mental State Examination in Alzheimer's disease and multi-infarct dementia. Int Psychogeriatr 1996;8(1):127-134. [doi: 10.1017/s1041610296002529] [Medline: 8805093]
18. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993 Nov;43(11):2412-2414. [doi: 10.1212/wnl.43.11.2412-a] [Medline: 8232972]
19. NIST/SEMATECH e-handbook of statistical methods. 2012 Apr. URL: http://www.itl.nist.gov/div898/handbook/ [accessed 2019-02-10]
20. Sukel K. BrainFacts/SfN. 2018 Nov 15. Figuring out why Alzheimer’s disease strikes more women than men URL: https://www.brainfacts.org/diseases-and-disorders/topic-center-alzheimers-and-dementia/figuring-out-why-alzheimers-disease-strikes-more-women-than-men-1115183 [accessed 2020-03-06]
21. Hara Y. How does Alzheimer’s affect women and men differently?. 2018 Jul 02. URL: https://www.alzdiscovery.org/cognitive-vitality/blog/how-does-alzheimers-affect-women-and-men-differently [accessed 2020-03-06]
22. Riedel BC, Thompson PM, Brinton RD. Age, APOE and sex: triad of risk of Alzheimer's disease. J Steroid Biochem Mol Biol 2016 Jun;160:134-147 [FREE Full text] [doi: 10.1016/j.jsbmb.2016.03.012] [Medline: 26969397]
23. Karp A, Kåreholt I, Qiu C, Bellander T, Winblad B, Fratiglioni L. Relation of education and occupation-based socioeconomic status to incident Alzheimer's disease. Am J Epidemiol 2004 Jan 15;159(2):175-183. [doi: 10.1093/aje/kwh018] [Medline: 14718220]

24. Evans DA, Hebert LE, Beckett LA, Scherr PA, Albert MS, Chown MJ, et al. Education and other measures of socioeconomic status and risk of incident Alzheimer disease in a defined population of older persons. Arch Neurol 1997 Nov;54(11):1399-1405. [doi: 10.1001/archneur.1997.00550230066019] [Medline: 9362989]
25. Arevalo-Rodriguez I, Smailagic N, Roqué IFM, Ciapponi A, Sanchez-Perez E, Giannakou A, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). Cochrane Database Syst Rev 2015 Mar 05(3):CD010783. [doi: 10.1002/14651858.CD010783.pub2] [Medline: 25740785]
26. Tate DF, Neeley ES, Norton MC, Tschanz JT, Miller MJ, Wolfson L, et al. Intracranial volume and dementia: some evidence in support of the cerebral reserve hypothesis. Brain Res 2011 Apr 18;1385:151-162 [FREE Full text] [doi: 10.1016/j.brainres.2010.12.038] [Medline: 21172323]
27. Fotenos AF, Snyder AZ, Girton LE, Morris JC, Buckner RL. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. Neurology 2005 Mar 22;64(6):1032-1039. [doi: 10.1212/01.WNL.0000154530.72969.11] [Medline: 15781822]
28. Storandt M, Grant EA, Miller JP, Morris JC. Longitudinal course and neuropathologic outcomes in original vs revised MCI and in pre-MCI. Neurology 2006 Aug 08;67(3):467-473. [doi: 10.1212/01.wnl.0000228231.26111.6e] [Medline: 16894109]
29. Killiany RJ, Gomez-Isla T, Moss M, Kikinis R, Sandor T, Jolesz F, et al. Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease. Ann Neurol 2000 Apr;47(4):430-439. [Medline: 10762153]
30. Killiany RJ, Hyman BT, Gomez-Isla T, Moss MB, Kikinis R, Jolesz F, et al. MRI measures of entorhinal cortex vs hippocampus in preclinical AD. Neurology 2002 Apr 23;58(8):1188-1196. [doi: 10.1212/wnl.58.8.1188] [Medline: 11971085]
31. Fox NC, Freeborough PA. Brain atrophy progression measured from registered serial MRI: validation and application to Alzheimer's disease. J Magn Reson Imaging 1997;7(6):1069-1075. [doi: 10.1002/jmri.1880070620] [Medline: 9400851]
32. Buckner RL, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, et al. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage 2004 Oct;23(2):724-738. [doi: 10.1016/j.neuroimage.2004.06.018] [Medline: 15488422]
33. Open Access Series of Brain Imaging (OASIS) database. URL: https://www.oasis-brains.org [accessed 2020-03-06]

Abbreviations

AD: Alzheimer disease
APOE-ε4: ε4 allele of the apolipoprotein E gene
ASF: Atlas Scaling Factor
CDR: Clinical Dementia Rating
EDA: exploratory data analysis
EDUC: patient education level
eTIV: estimated total intracranial volume
IQR: interquartile range
MCI: mild cognitive impairment
M/F: male/female
MMSE: Mini-Mental State Examination
MRI: magnetic resonance image
nWBV: normalized whole brain volume
OASIS: Open Access Series of Brain Imaging
SES: socioeconomic status

Edited by G Eysenbach; submitted 15.04.19; peer-reviewed by Y Hu, F Lanfranchi; comments to author 16.07.19; revised version received 08.09.19; accepted 09.02.20; published 14.04.20.

Please cite as:
Khan A, Zubair S
Longitudinal Magnetic Resonance Imaging as a Potential Correlate in the Diagnosis of Alzheimer Disease: Exploratory Data Analysis
JMIR Biomed Eng 2020;5(1):e14389
URL: http://biomedeng.jmir.org/2020/1/e14389/
doi:10.2196/14389
PMID:

Original Paper

Diagnosis of Type 2 Diabetes Using Electrogastrograms: Extraction and Genetic Algorithm–Based Selection of Informative Features

Paramasivam Alagumariappan[1*], PhD; Kamalanand Krishnamurthy[2*], PhD; Sundravadivelu Kandiah[3*], MBBS; Emmanuel Cyril[4*], PhD; Rajinikanth V[5*], PhD

[1]Department of Electrical and Electronics Engineering, Chennai, India

[2]Department of Instrumentation Engineering, MIT Campus, Anna University, Chennai, India

[3]Shree Balaji Clinic, Ezhil Nagar, Selaiyur, Chennai, India

[4]Global Hospitals & Health City, Cheran Nagar, Perumbakkam, Chennai, India

[5]Department of Electronics and Instrumentation Engineering, St. Joseph's College of Engineering, Chennai, India

[*]all authors contributed equally

Corresponding Author:
Paramasivam Alagumariappan, PhD
Department of Electrical and Electronics Engineering
B. S. Abdur Rahman Crescent Institute of Science and Technology
Chennai, 600048
India
Phone: 91 9843780801
Email: parama.ice@gmail.com

Abstract

Background: Electrogastrography is a noninvasive electrophysiological procedure used to measure gastric myoelectrical activity. EGG methods have been used to investigate the mechanisms of the human digestive system and as a clinical tool. Abnormalities in gastric myoelectrical activity have been observed in subjects with diabetes.

Objective: The objective of this study was to use the electrogastrograms (EGGs) from healthy individuals and subjects with diabetes to identify potentially informative features for the diagnosis of diabetes using EGG signals.

Methods: A total of 30 features were extracted from the EGGs of 30 healthy individuals and 30 subjects with diabetes. Of these, 20 potentially informative features were selected using a genetic algorithm–based feature selection process. The selected features were analyzed for further classification of EGG signals from healthy individuals and subjects with diabetes.

Results: This study demonstrates that there are distinct variations between the EGG signals recorded from healthy individuals and those from subjects with diabetes. Furthermore, the study reveals that the features Maragos fractal dimension and Hausdorff box-counting fractal dimension have a high degree of correlation with the mobility of EGGs from healthy individuals and subjects with diabetes.

Conclusions: Based on the analysis on the extracted features, the selected features are suitable for the design of automated classification systems to identify healthy individuals and subjects with diabetes.

(JMIR Biomed Eng 2020;5(1):e20932) doi:10.2196/20932

KEYWORDS

electrogastrograms; genetic algorithm; feature extraction; feature selection; diabetes

Introduction

Digestion is the breakdown of food into small water-soluble molecules that can be absorbed by the intestinal epithelium [1]. During digestion, food enters the mouth and undergoes mechanical and chemical processes that result in the breakdown of food and absorption of nutrients [1].

Electrogastrography is a noninvasive technique used to measure and record the gastric myoelectrical activity associated with the process of digestion [2]. Electrogastrograms (EGG) are the recordings of the electrical signals originating from the stomach

muscles. Several cutaneous electrodes are placed on the upper abdomen, over the stomach, for the acquisition of EGG signals [3]. The dominant frequency of the EGG signal is identical to the frequency of the electrical activity of the stomach. The frequency of a healthy EGG signal ranges from 2.6 to 3.7 cycles per minute (cpm), which is produced by the interstitial cells of Cajal located in the muscular wall of the gastric corpus and antrum [4].

Diabetic gastropathy is defined as a spectrum of neuromuscular abnormalities of the stomach. In diabetic gastropathy, the normal average EGG signal (3 cpm) is disrupted by bradygastrias, tachygastrias, and other mixed dysrhythmias [5]. Several studies have identified neuromuscular abnormalities in subjects with diabetes with upper gastrointestinal symptoms for the diagnosis of gastric dysrhythmias [6-10]. Koch et al (2001) [6] discussed the clinical applications of electrogastrography in diabetic gastropathy. Altintop et al (2016) [7] proposed the use of parametric methods such as Cramer-Rao lower bound and power spectral density for the analysis of EGG signals obtained from subjects with gastroparesis and healthy volunteers using cutaneous electrodes [7]. Additionally, the authors extracted several features from the power spectral density functions, which were utilized to identify subjects with gastroparesis and healthy subjects [7].

The frequency spectra of healthy and diabetic EGG signals often show an exponential increase of power toward the very low frequency range (<1 cpm) [5]. These frequencies are not likely to originate from the stomach or other parts of the human body. These ultralow frequency components of EGG signals may be caused by factors such as low-frequency electrode noise due to variations in electrode potential, and movement artifacts [5]. Therefore, it is necessary to filter frequencies <1 cpm to avoid false interpretation. In recent years, the empirical mode decomposition (EMD) technique has been used to preprocess or filter several biosignals with high accuracy [11-15]. Furthermore, a study has proposed the use of the noise-assisted multivariate empirical mode decomposition for multichannel electromyography signal processing [11].

Feature extraction is a technique used to extract useful information that is hidden in biosignals. The selection of the appropriate feature is important, as it leads to precise analysis and high classification accuracy [16]. Additionally, during the feature selection process, potentially informative features can be selected for future classification processes and analysis. Furthermore, the performance of the classifier is highly sensitive to efficient informative features [17]. Several studies have proposed various feature extraction methods such as time domain features, frequency domain features, and time frequency domain features for the analysis of biosignals [16-18]. A study reviewed feature extraction methods on EEG signals using linear analysis in frequency and time-frequency domains and showed that the frequency domain methods provided more detailed information on EEG signal analysis than the time-frequency methods [15]. Another study about textile image classification based on its texture used the feature extraction methods Gray level co-occurrence matrix (GLCM), linear binary pattern, and a moment invariant [18]. The study found that the best result was achieved using a combination of GLCM and linear binary pattern features [18].

The objective of this work was to extract features from EGG signals from healthy individuals and subjects with diabetes to select useful and highly informative features for the diagnosis of diabetes. Additionally, we aimed to evaluate the correlation between the selected features and the process of digestion in both groups of individuals.

Methods

Participants

A total of 30 healthy individuals and 30 subjects with diabetes participated in this study. Participants ranged in age from 20 to 50 years.

The ethical clearance (HR/2017/MS/002) to conduct this research study was obtained from Global Hospitals & Health City, Chennai.

EGG Signal Acquisition

An EGG measurement system with 3 surface electrodes was developed and used to record the EGGs from healthy individuals and subjects with diabetes. Of the 3 electrodes, 2 electrodes were positioned on the outer curvature (fundus) and on the inner curvature (mid corpus) of the stomach with a separation distance of 5 cm between the electrodes, in accordance with the standard electrode placement protocol [19-21]. For isolation purposes, the third electrode was placed as ground, away from the stomach area. The acquired EGGs were amplified with an amplification system developed using IC AD624 [2] and were logged using LABVIEW hardware and software.

EGGs from all participants were acquired for a period of 10 minutes (Figures 1 and 2). All EGGs were preprocessed and analyzed using custom made functions in MATLAB R2011b [2].

Figure 1. Block diagram of the electrogastrogram acquisition system.

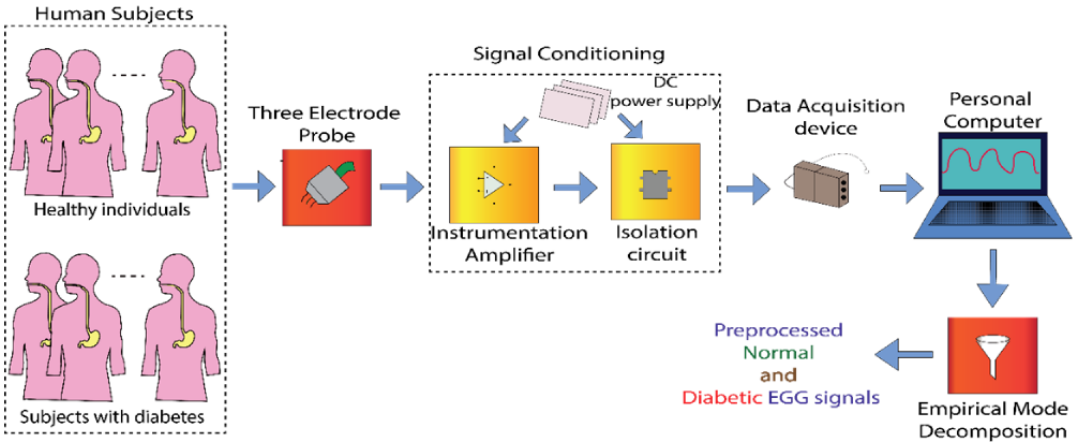
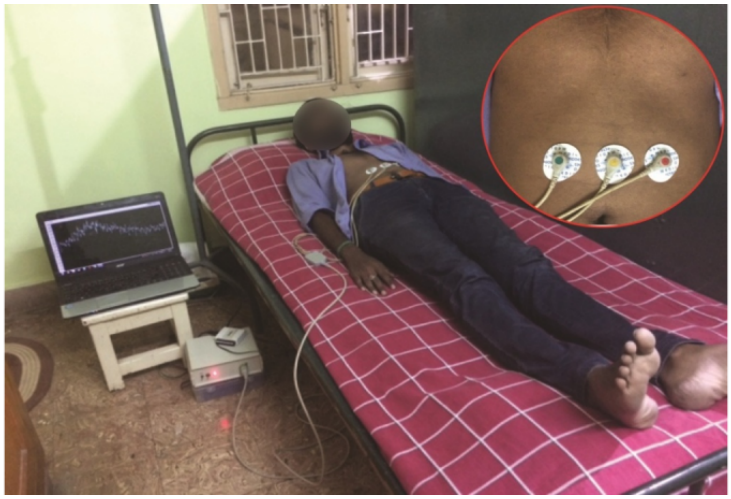


Figure 2. Acquisition of electrogastrogram from a participant.



Preprocessing of EGG Signals

The EMD analysis was used to decompose the input EGG signal into different frequency components called intrinsic mode functions (IMFs) [11]. The number of IMFs can be extracted following two fundamental requirements, the number of extrema or zero-crossings must be the same or differ by at most 1, and the mean of upper and lower envelopes of IMFs should be 0. By applying the EMD algorithm, the EGG signal ($x[n]$) can be represented as follows [11,12]:

where IMF_i $[n]$ is the i^{th} IMF, r_k $[n]$ is the residue, and k is the total number of IMFs. The length, nonlinearity, and nonstationarity of the EGG signal determines the number of IMFs to be generated [11]. The EMD filter is well established and has been described in detail in the available literature [13-15]. In this study, the IMFs holding the ultralow frequency components <1 cpm were removed and the rest of the IMFs were added to obtain the filtered EGG signal. Further, the preprocessed EGG signals from healthy individuals and subjects with diabetes were subjected to feature extraction methods.

Feature Extraction

The feature extraction technique plays a vital role in achieving high classification accuracy in the analysis of biosignal processing. The process of feature extraction involves the transformation of raw EGG signals into a feature vector [22]. The 30 EGG signal features including descriptive statistics (mean, median, mode, minimum value, maximum value, standard deviation, skewness, and kurtosis), Hjorth parameters (activity, mobility, and complexity), entropy measures (Renyi entropy, Tsallis entropy, spectral entropy, and image entropy), fractal dimensions (Maragos fractal dimensions, MFD; and Hausdorff box count fractal dimension, HFD), the fast Fourier transform (FFT) peak, and the GLCM (contrast, correlation, energy, and homogeneity) were extracted from preprocessed healthy and diabetic EGG signals.

FFT Peak

The peak frequency of healthy and diabetic EGG signals was extracted using the FFT. By taking the FFT for recorded healthy and diabetic EGG signals, the frequency components present in the EGG signal were plotted against an amplitude spectrum of a single side. Further, the frequency component with maximum amplitude was considered as the peak frequency of an EGG signal.

Hjorth Parameters

Hjorth parameters are used to characterize the information on the temporal dynamics of the measured biosignals. In this work,

the Hjorth features activity, mobility, and complexity were extracted from healthy and diabetic EGG signals.

Activity represents the measurement of variance or the average power of an EGG signal. Activity is given as follows [23]:

$$Activity = \text{var}(y(t)) \quad \textbf{(2)}$$

where $(y(t))$ is the input EGG signal.

Mobility represents the average frequency of an EGG signal. The mobility parameter is defined as the square root of the ratio of the variance of the first derivative of the signal and the variance of the signal. Mobility of an EGG signal is defined as follows:

The mobility parameter has a proportion of standard deviation of the power spectrum.

Complexity represents a measure of variability of an EGG signal. Complexity of an EGG signal is defined as follows:

The complexity parameter indicates the similarity between input EGG signals to a pure sine wave. The value of complexity converges to 1 as the shape of the signal gets more similar to a pure sine wave.

Entropy Measures

Entropy is defined as a measure of disorder associated with a system, and hence, it is a measure of information content, uncertainty, and complexity of the system.

The Rényi entropy of the sample $H(\alpha)$ is given by the following equation [24]:

where p_i is the probability that a random variable takes a given value of n values and alpha is the order of the entropy measure. As alpha increases, the Rényi entropy increases. The Rényi entropy is an effective measure of the complexity of the signal [24-26]. The complexity of the EGG signals recorded from healthy individuals and subjects with diabetes were extracted using the Rényi entropy with 5 different orders of the entropy measure (alpha=0.2, 0.4, 0.6, 0.8, and 0.9).

The Tsallis entropy is one of the most promising information theoretic methods for biosignal analysis. The Tsallis entropy (H_R) is defined as follows [25]:

where p_i is a given set of probabilities and alpha is a real number. As alpha increases, the Tsallis entropy decreases. The information content of the EGG signals recorded from healthy individuals and subjects with diabetes were extracted using the Tsallis entropy with 5 different orders of the entropy measure (alpha=0.2, 0.4, 0.6, 0.8, and 0.9).

Time domain and frequency domain are the two different possible ways in which the entropy of a biosignal can be computed. The spectral entropy of EGG signals shall be computed in frequency domain [26]. The spectral components can be evaluated using the FFT. The concept of spectral entropy originates from a measure of information called Shannon entropy when applied to the power spectrum of a signal, spectral S is given as follows [27]:

where  , N=number of frequencies region, and p_k are spectral amplitudes of k frequencies region.

Fractal Dimension

Fractals are mathematical sets with a high degree of geometrical complexity, which can model many classes of time series data as well as images [28]. Maragos and Sun [29] developed an approach for estimating the fractal dimension of time dependent signals using morphological erosion and dilation operations to create covers around a signal's graph at multiple scales. Maragos and Sun [29] proposed the "morphological covering method," which utilizes multiscale morphological operations with varying structuring elements that improve other covering methods. Experimental investigations on the morphological covering method demonstrate good performance with low estimation errors.

Spectrogram

The preprocessed EGG signals recorded from healthy individuals and subjects with diabetes were converted into a time corrected instantaneous frequency spectrogram using a spectrogram method. The spectrogram was plotted as an image with the intensities encoding the levels. The spectrogram had time on the x-axis and frequency on the y-axis [30]. Further, image entropy and HFD as well as the four GLCM features, contrast, correlation, energy, and homogeneity were extracted from converted healthy and diabetic spectrograms.

Image entropy is defined as a scalar value that represents the entropy of a grayscale image. Entropy is a measure of disorder or randomness that can be used to characterize the texture of the input image. Images with lesser entropy have lot of black sky, less contrast, and a large number of pixels. Image entropy is expressed by the equation [31]:

where P_i is the probability that the difference between 2 adjacent pixels is equal to i, and log_2 is the base 2 logarithm.

The HFD is a descriptor of the complexity of the geometry of a given set. The set can be the trajectory of any dynamical system and can be reconstructed from the measured data. Suppose that A is the set whose dimension is to be calculated. Let $C(r,A) = \{B_1, B_2 \ldots B_K\}$ be a finite cover of the set A by sets whose diameters are less than r. Then, the following function defines a measure of the set A [30]:

For most values of *D*, the limit leads to a degenerate measure, either or The box-counting dimension estimate can be written as follows:

with sufficiently small *r*. The problem is determining if a given box of grid contains a point (or points) of trajectory over all boxes in grids.

The GLCM is a sum of the number of times that the pixel with the gray level value *i* occurred in the specified spatial relationship to a pixel with the value *j*. The spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent). The size of the GLCM is proportional to the number of gray levels in the image [32,33]. In addition, the GLCM exposes certain properties about the spatial distribution of the gray levels in the texture image. The features contrast, correlation, energy, and homogeneity were extracted from the GLCM matrix of healthy and diabetic EGG signals.

Contrast is a measure of the intensity (contrast) between a pixel and its neighbor pixel over the whole image. Contrast is 0 for a constant image. In general, the property contrast is also known as variance and inertia [32]. Correlation is a measure of the correlation between a pixel and its neighbor pixel over the whole image. The correlation value shall be 1 or –1 for a perfectly positively or negatively correlated image, respectively. Energy is the sum of squared elements in the GLCM matrix. Energy is 1 for a constant image. In general, the property energy is also known as uniformity and uniformity of energy. Homogeneity is a measure of closeness of the distribution of elements in the GLCM to the GLCM diagonal. The homogeneity value shall be 1 for a diagonal GLCM.

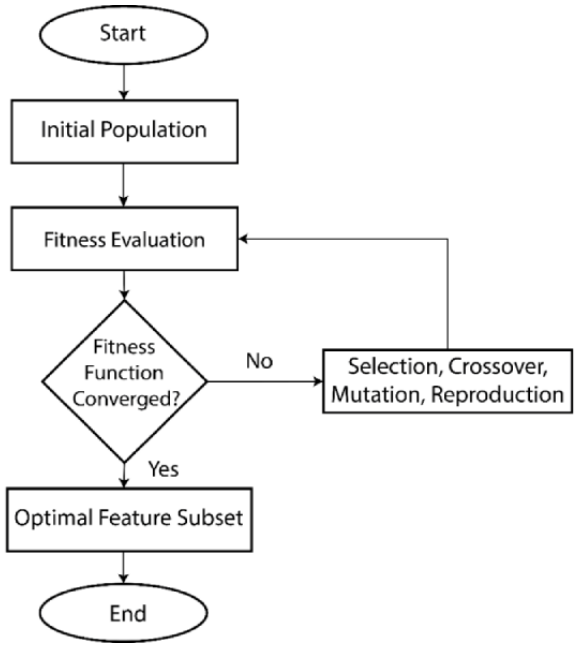
Feature Selection Using a Genetic Algorithm

Using different feature extraction methods, a number of features can be extracted and, from them, effective informative features can be selected [34]. Further, the performance of a classifier is highly sensitive to the efficiency of the feature selection methods. Genetic algorithms are search-optimization techniques based on Darwin's principle of natural selection [34,35].

In this work, a genetic algorithm–based feature selection method was adapted to search, identify, and select potentially informative features from extracted healthy and diabetic EGG signal features for feature analysis. The flowchart of the genetic algorithm is shown in Figure 3. If *F* is the total number of features, then 2^F possible feature subsets can be created. The initial set of possible solutions or populations with a fixed population size is randomly constructed and fitness of each individual is evaluated with its fitness function. In this work, classification accuracy was adopted as the fitness measure. By adopting a genetic algorithm, the optimization was performed to select the optimal subset of features [34,35].

Of the 30 features extracted from preprocessed healthy and diabetic EGG signals, the 20 best features were chosen using a genetic algorithm–based feature selection method.

Figure 3. Flowchart of the genetic algorithm.



Results

Different patterns of EGG signals were observed in healthy individuals and subjects with diabetes. Figures 4A and 4B show a typical EGG signal recorded from a healthy individual and the single-sided amplitude spectrum of a healthy EGG signal, respectively.

Figures 5A and 5B show a typical EGG signal recorded from a subject with diabetes and the single-sided amplitude spectrum of a diabetic EGG signal, respectively.

Figure 4. (A) Typical electrogastrogram signal recorded from a healthy individual. (B) The single-sided amplitude spectrum of a healthy electrogastrogram signal.

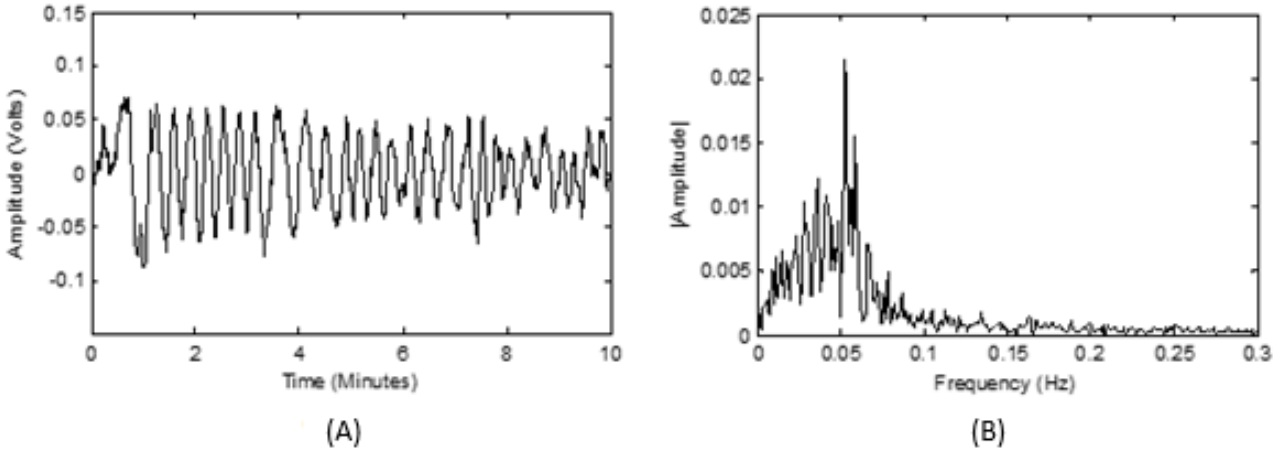
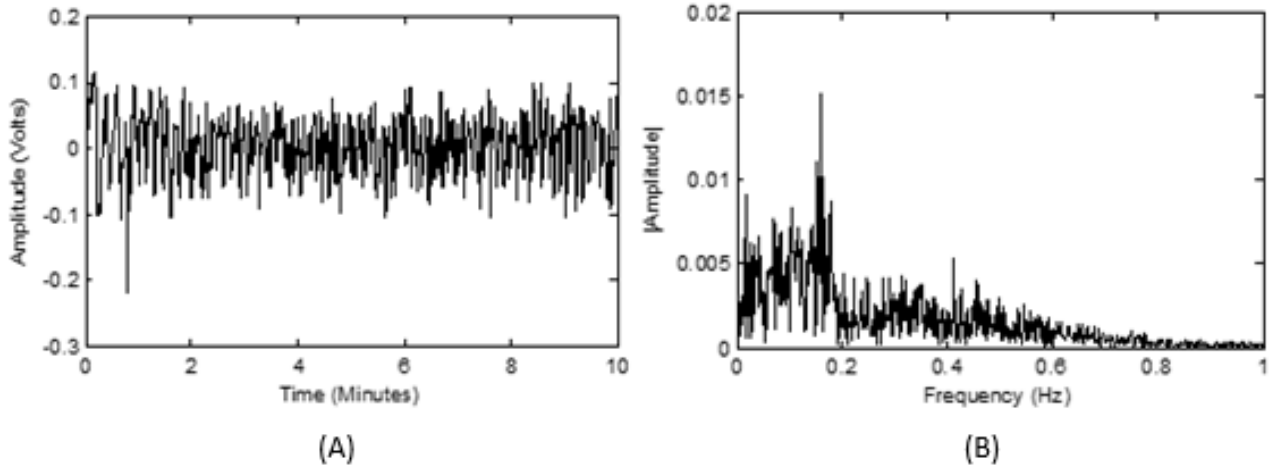


Figure 5. (A) Typical electrogastrogram signal recorded from a subject with diabetes. (B) The single-sided amplitude spectrum of a diabetic electrogastrogram signal.



The variation of spectral entropy values was evaluated as a function of mobility of EGG signals recorded from healthy individuals (Figure 6A) and subjects with diabetes (Figure 6B). We found that the spectral entropies extracted from healthy EGG signals (R=0.96741) and diabetic EGG signals (R=0.90993) had a high correlation with mobility.

The variation of HFD values was investigated as a function of mobility of EGGs recorded from healthy subjects (Figure 7A) and subjects with diabetes (Figure 7B). HFD values extracted from healthy EGGs had a high degree of correlation (R=0.91737) with mobility. HFD values extracted from diabetic EGGs had a good correlation (R=0.77178) with mobility.

Figure 6. Variation of spectral entropy values as a function of mobility of electrogastrogram signals. (A) Healthy individuals. (B) Subjects with diabetes.

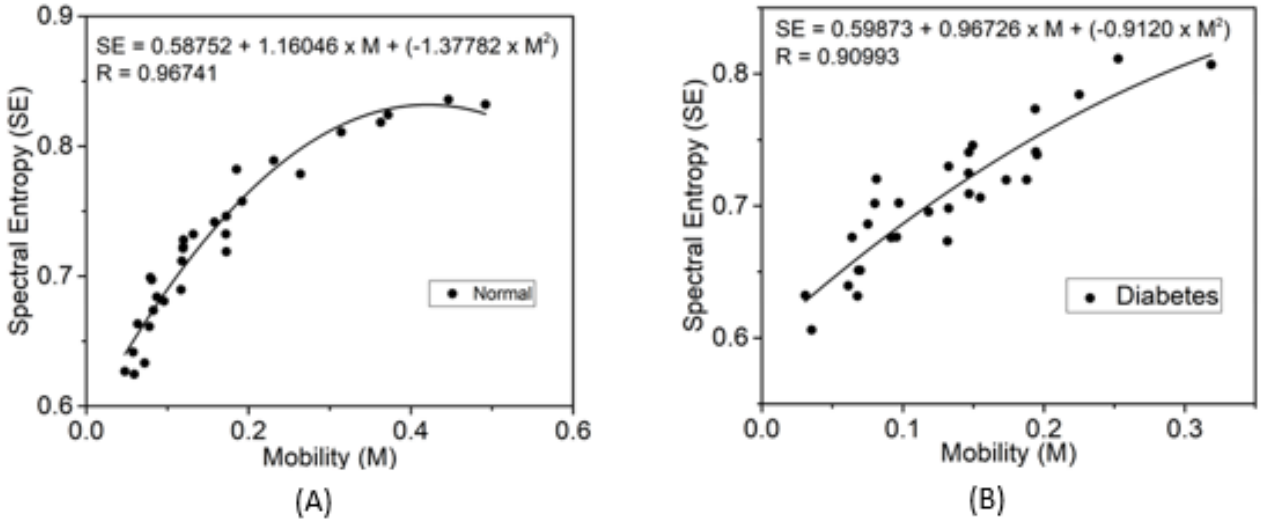
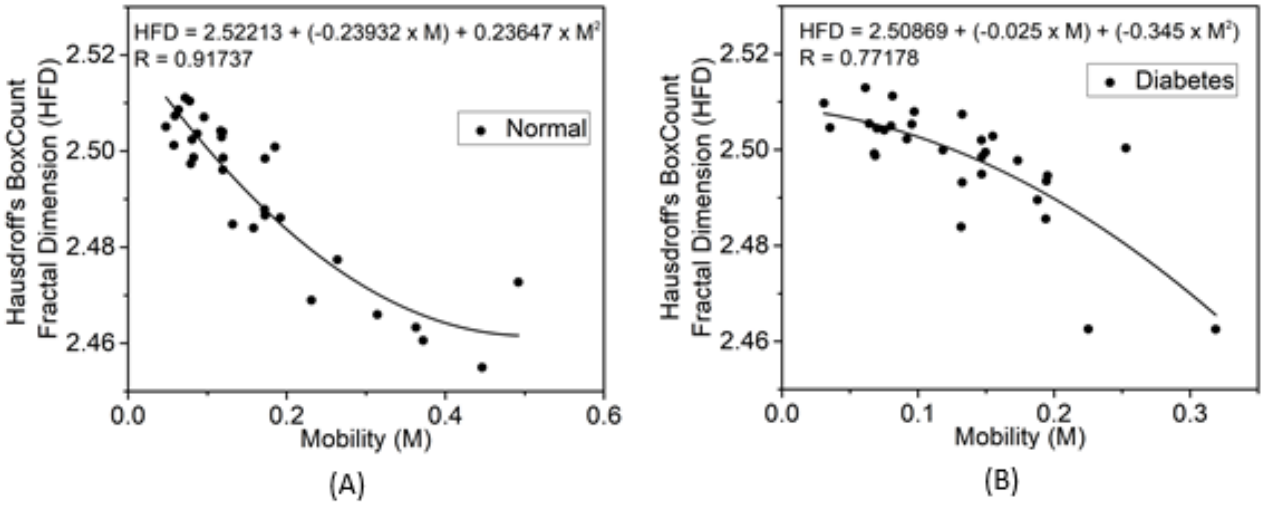


Figure 7. Variation of the Hausdorff box-counting fractal dimension values as a function of mobility of electrogastrograms. (A) Healthy individuals. (B) Subjects with diabetes.



The values of MFD were evaluated as a function of mobility of EGGs acquired from healthy subjects (Figure 8A) and subjects with diabetes (Figure 8B). MFD values extracted from healthy EGGs have a high degree of correlation (R=0.88976) with mobility. Similarly, MFD values extracted from diabetic EGG signals have a good correlation (R=0.8077) with mobility.

The average Hjorth parameters activity, mobility, and complexity of EGG signals were recorded from healthy individuals and subjects with diabetes (Figure 9). The average mobility and complexity of the EGG signals recorded from healthy individuals are higher than the average mobility and complexity of the EGG signals recorded from subjects with diabetes. Further, the average activity of the EGG signals recorded from subjects with diabetes is higher than the average activity of the EGG signals recorded from healthy individuals.

The MFD and HFD values of the EGG signals were recorded from healthy individuals and subjects with diabetes (Figures 10A and 10B, respectively). The average MFD of EGG signals recorded from healthy individuals is higher than the average MFD of EGG signals recorded from subjects with diabetes. Furthermore, the HFD of EGG signals recorded from healthy individuals are lower than the average HFD of EGG signals recorded from subjects with diabetes.

Figure 8. Variation of the Maragos fractal dimension values as a function of mobility of electrogastrograms. (A) Healthy individuals. (B) Subjects with diabetes.

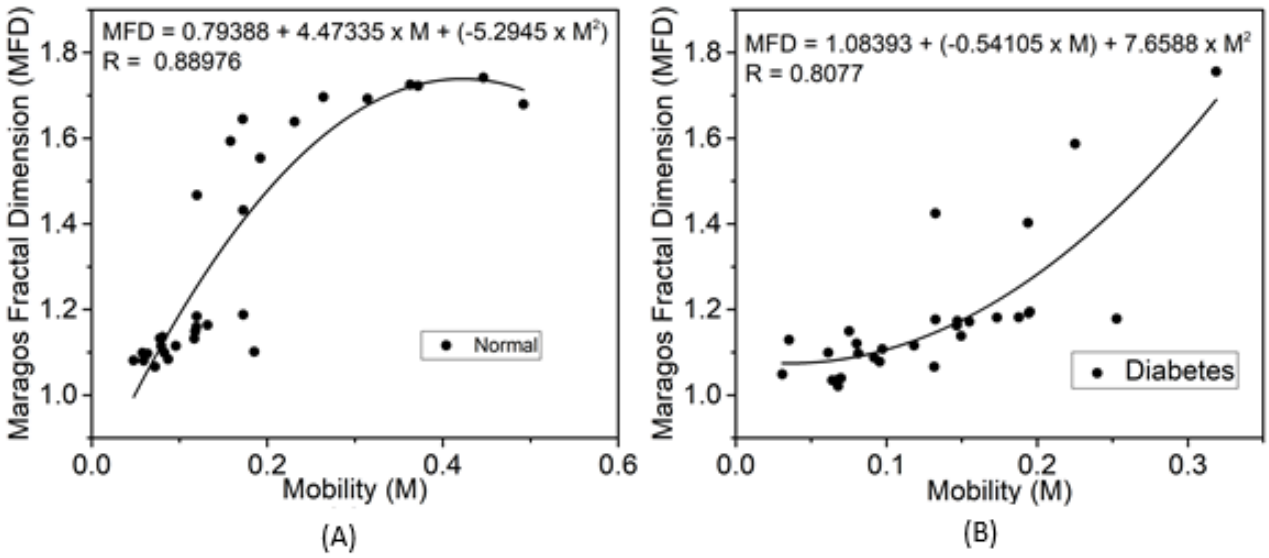


Figure 9. Hjorth parameters (mean) of electrogastrogram signals recorded from healthy individuals and subjects with diabetes. Error bars indicate standard error.

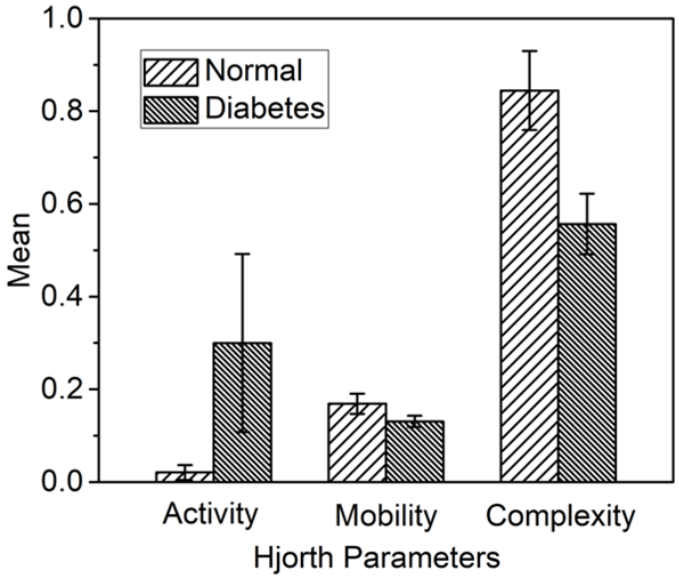
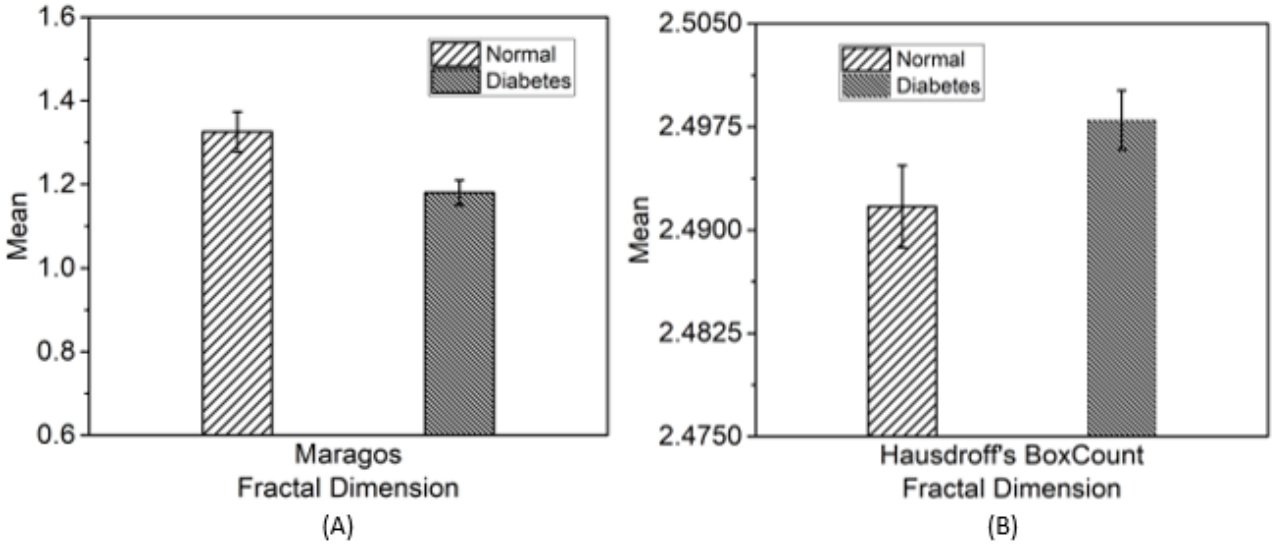


Figure 10. The fractal dimension (mean) of electrogastrogram signals from healthy individuals and subjects with diabetes. Error bars indicate standard error. (A) Maragos fractal dimension. (B) Hausdorff box-counting fractal dimension.



Discussion

Principal Findings

Type 2 diabetes is a chronic disease that prevents the physiological system from using insulin efficiently. It is expected that the global number of type 2 diabetes cases will reach around 450 million by 2030. Undiagnosed diabetes is often associated with complications such as cardiovascular and kidney diseases. However, these risk factors are preventable by the early detection and diagnosis of diabetes [36]. In this regard, a method for the early detection of type 2 diabetes is of high value. The method needs to be simple, self-applicable, noninvasive, and safe. This study aimed to develop a device for mass screening of diabetes. The results confirmed that the frequency of 3 cpm is dominant in the EGG signals acquired from healthy individuals. However, a frequency of 9.6 cpm was dominant in the EGG signals acquired from subjects with diabetes. It was demonstrated that the EGG signals with diabetic complexities cannot be visualized or examined by naked eyes. Mobility is the average frequency of an EGG; therefore, it was highly correlated with the dynamic process of digestion. Additionally, the extracted features of healthy and diabetic EGGs were found to be well correlated with the physiological process of digestion. Further, it was demonstrated that the features spectral entropy, energy, HFD, and MFD provide information about abnormalities in the EGGs.

Conclusion

Human gastric myoelectrical activity can be measured using a noninvasive technique known as EGG. However, although frequency characteristics are one of the most significant parameters, the visual analysis of EGG signals is very difficult. Subjects with diabetes who have poorly controlled diet habits are often suspected of diabetic gastroparesis. In this work, features such as time domain features, frequency domain features, and time-frequency domain features were extracted from EGG signals recorded from healthy individuals and subjects with diabetes. Further, potentially informative features were selected using a genetic algorithm–based feature selection method. Additionally, the correlation of the extracted features with the mobility of the digestive system was analyzed. Results demonstrate that the extracted features grasp individual informative characteristics that can be used for analysis. Further, the features MFD and HFD have a high degree of correlation with the mobility of healthy and diabetic EGG signals. Additionally, the spectral entropy of EGG signals recorded from healthy individuals is highly correlated with the mobility of EGG signals recorded from healthy individuals and subjects with diabetes. This work appears to be of high clinical significance, as these extracted potentially informative features can be used for the analysis and classification of digestive system disorders. In the future, deep learning techniques can be utilized for the automated classification of healthy and diabetic EGG signals.

Conflicts of Interest

None declared.

References

1. DeSesso J, Jacobson C. Anatomical and physiological parameters affecting gastrointestinal absorption in humans and rats. Food Chem Toxicol 2001 Mar;39(3):209-228. [doi: 10.1016/s0278-6915(00)00136-8]
2. Alagumariappan P, Krishnamurthy K, Kandiah S, Ponnuswamy MJ. Effect of electrode contact area on the information content of the recorded electrogastrograms: An analysis based on Rényi entropy and Teager-Kaiser Energy. Polish J Medical Phys Eng 2017;23(2):37-42. [doi: 10.1515/pjmpe-2017-0007]

3. Yin J, Chen JDZ. Electrogastrography: Methodology, validation and applications. J Neurogastroenterol Motil 2013 Jan 31;19(1):5-17. [doi: 10.5056/jnm.2013.19.1.5]
4. Dong K, Yu XJ, Li B, Wen EG, Xiong W, Guan QL. Advances in mechanisms of postsurgical gastroparesis syndrome and its diagnosis and treatment. Chin J Dig Dis 2006 May;7(2):76-82. [doi: 10.1111/j.1443-9573.2006.00255.x] [Medline: 16643334]
5. Verhagen MA, Van Schelven LJ, Samsom M, Smout AJ. Pitfalls in the analysis of electrogastrographic recordings. Gastroenterology 1999 Aug;117(2):453-460. [doi: 10.1053/gast.1999.0029900453] [Medline: 10419928]
6. Koch KL. Electrogastrography: physiological basis and clinical application in diabetic gastropathy. Diabetes Technol Ther 2001 Mar;3(1):51-62. [doi: 10.1089/152091501750220019] [Medline: 11469708]
7. Altıntop Ç, Latifoğlu F, Çelikzencir E, Sezgin GC, Yurci MA. Determining of gastroparesis disease from electrogastrogram signals using Cramer-Rao lower bound and power spectral density. 2017;17(1):3061-3067.
8. Ouyang A. Gastrointestinal motility disorders: Diagnosis and treatment. Gastroenterology 1990 Sep;99(3):898-899. [doi: 10.1016/0016-5085(90)90995-d]
9. Ajumobi AB, Griffin RA. Diabetic gastroparesis: evaluation and management. Hosp Physician 2008;44(3):27-35.
10. Bradshaw LA, Cheng LK, Chung E, Obioha CB, Erickson JC, Gorman BL, et al. Diabetic gastroparesis alters the biomagnetic signature of the gastric slow wave. Neurogastroenterol Motil 2016 Jun 03;28(6):837-848 [FREE Full text] [doi: 10.1111/nmo.12780] [Medline: 26839980]
11. Zhang Y, Xu P, Li P, Duan K, Wen Y, Yang Q, et al. Noise-assisted multivariate empirical mode decomposition for multichannel EMG signals. Biomed Eng Online 2017 Aug 23;16(1):107 [FREE Full text] [doi: 10.1186/s12938-017-0397-9] [Medline: 28835251]
12. Jain S, Bajaj V, Kumar A. Riemann Liouvelle fractional integral based empirical mode decomposition for ECG denoising. IEEE J Biomed Health Inform 2018 Jul;22(4):1133-1139. [doi: 10.1109/jbhi.2017.2753321]
13. Krupa N, Ali M, Zahedi E, Ahmed S, Hassan FM. Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine. Biomed Eng Online 2011 Jan 19;10:6 [FREE Full text] [doi: 10.1186/1475-925X-10-6] [Medline: 21244712]
14. Phinyomark A, Phukpattaranont P, Limsakul C. Feature reduction and selection for EMG signal classification. Expert Syst Appl 2012 Jun;39(8):7420-7431. [doi: 10.1016/j.eswa.2012.01.102]
15. Al-Fahoum AS, Al-Fraihat AA. Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. ISRN Neurosci 2014;2014:730218 [FREE Full text] [doi: 10.1155/2014/730218] [Medline: 24967316]
16. Mohamad M, Deris S, Yatim SM, Othman MR. Feature selection method using genetic algorithm for the classification of small and high dimension data. 2004 Presented at: First International Symposium on Information and Communications Technologies; 7-8 October 2014; Putrajaya p. 13-16 URL: https://www.myexperiment.org/files/747/download
17. Riezzo G, Russo F, Indrio F. Electrogastrography in adults and children: the strength, pitfalls, and clinical significance of the cutaneous recording of the gastric electrical activity. Biomed Res Int 2013;2013:282757 [FREE Full text] [doi: 10.1155/2013/282757] [Medline: 23762836]
18. Pawening RE, Dijaya R, Brian T, Suciati N. Classification of textile image using support vector machine with textural feature. : Institute of Electrical and Electronics Engineers; 2015 Presented at: International Conference on Information & Communication Technology and Systems (ICTS); 16 Sept 2015; Surabaya p. 119-122. [doi: 10.1109/icts.2015.7379883]
19. Alagumariappan P, Krishnamurthy K. An approach based on information theory for selection of systems for efficient recording of electrogastrograms. In: Proceedings of the International Conference on Computing and Communication Systems.: Springer Nature Singapore Pte Ltd; 2018 Presented at: International Conference on Computing and Communication Systems; 11-13 Nov 2016; Shillong p. 463-471. [doi: 10.1007/978-981-10-6890-4_45]
20. Nazmi N, Abdul Rahman M, Yamamoto S, Ahmad S, Zamzuri H, Mazlan S. A review of classification techniques of EMG signals during isotonic and isometric contractions. Sensors (Basel) 2016 Aug 17;16(8) [FREE Full text] [doi: 10.3390/s16081304] [Medline: 27548165]
21. Oh S, Lee Y, Kim H. A Novel EEG Feature Extraction Method Using Hjorth Parameter. Int J Electron Electr Eng 2014;2(2):106-110. [doi: 10.12720/ijeee.2.2.106-110]
22. Bromiley PA, Thacker NA, Bouhova-Thacker E. Shannon entropy, Renyi entropy, and information. Statistic and segmentation series.: Imaging Science and Biomedical Engineering Division. The University of Manchester; 2004. URL: https://bit.ly/3nnkpig [accessed 2020-09-25]
23. Alagumariappan P, Rajagopal A, Krishnamurthy K. Complexity analysis on normal and abnormal electrogastrograms using Tsallis entropy. : Multidisciplinary Digital Publishing Institute; 2016 Presented at: 3rd International Electronic and Flipped Conference on Entropy and Its Applications; 1-10 Nov 2016; e-conference. [doi: 10.3390/ecea-3-a003]
24. Gonzalez Andino S, Grave de Peralta Menendez R, Thut G, Spinelli L, Blanke O, Michel C, et al. Measuring the complexity of time series: an application to neurophysiological signals. Hum Brain Mapp 2000 Sep;11(1):46-57 [FREE Full text] [Medline: 10997852]
25. Nunes RR, Almeida MPD, Sleigh JW. Spectral entropy: a new method for anesthetic adequacy. Rev Bras Anestesiol 2004 Jun;54(3):404-422. [Medline: 19471748]

26. Tang YY. Document analysis and recognition with wavelet and fractal theories. Toh Tuck Link: World Scientific; 2012.
27. Fulop SA, Fitz K. Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications. J Acoust Soc Am 2006 Jan;119(1):360-371. [doi: 10.1121/1.2133000] [Medline: 16454291]
28. Brink A. Using spatial information as an aid to maximum entropy image threshold selection. Pattern Recognit Lett 1996 Jan;17(1):29-36. [doi: 10.1016/0167-8655(95)00096-8]
29. Maragos P, Sun F. Measuring the fractal dimension of signals: Morphological covers and iterative optimization. IEEE Trans Signal Process 1993 Jan;41(1):108-121. [doi: 10.1109/TSP.1993.193131]
30. Schleicher D. Hausdorff dimension, its properties, and its surprises. Am Math Mon 2018 Jan 31;114(6):509-528. [doi: 10.1080/00029890.2007.11920440]
31. Dani LM, Dhod R. Non-invasive glucose level measurement in blood by GLCM technique. International Journal of Advanced Engineering Research and Science 2015 Jul;2(7):50-52.
32. Chaikla N, Qi Y. Genetic algorithms in feature selection. In: IEEE SMC'99 Conference Proceedings.: Institute of Electrical and Electronics Engineers; 1999 Presented at: IEEE International Conference on Systems, Man, and Cybernetics; 12-15 Oct 1999; Tokyo p. 538-540. [doi: 10.1109/icsmc.1999.815609]
33. Ludwig O, Nunes U. Novel maximum-margin training algorithms for supervised neural networks. IEEE Trans Neural Netw 2010 Jun;21(6):972-984. [doi: 10.1109/TNN.2010.2046423] [Medline: 20409990]
34. Aziz S, Khan MU, Aamir F, Javid MA. Electromyography (EMG) data-driven load classification using empirical mode decomposition and feature analysis. : Institute of Electrical and Electronics Engineers; 2019 Presented at: International Conference on Frontiers of Information Technology (FIT); 16-18 Dec 2019; Islamabad p. 272-277. [doi: 10.1109/fit47737.2019.00058]
35. Sun C, Xu J, Xu L, Wang G, Xu Y. Research on locomotive motor bearing diagnosis method based on wavelet threshold denoising and correlation analysis combined with EMD decomposition and LSTM neural network. : Institute of Electrical and Electronics Engineers; 2020 Presented at: Chinese Control And Decision Conference (CCDC); 22-24 Aug 2020; Hefei p. 2770-2775. [doi: 10.1109/ccdc49329.2020.9164259]
36. Narayan KMV, Chan J, Mohan V. Early identification of type 2 diabetes: policy should be aligned with health systems strengthening. Diabetes Care 2011 Jan;34(1):244-246 [FREE Full text] [doi: 10.2337/dc10-1952] [Medline: 21193623]

Abbreviations

EGG: electrogastrograms
EMD: empirical mode decomposition
FTT: fast Fourier transform
GLCM: Gray Level co-occurrence matrix
HFD: Hausdorff box-counting fractal dimension
IMF: intrinsic mode function
MFD: Maragos fractal dimensions

Edited by G Eysenbach; submitted 02.06.20; peer-reviewed by E Priya, F Palmieri; comments to author 22.06.20; revised version received 16.08.20; accepted 21.08.20; published 07.10.20.

Please cite as:
Alagumariappan P, Krishnamurthy K, Kandiah S, Cyril E, V R
Diagnosis of Type 2 Diabetes Using Electrogastrograms: Extraction and Genetic Algorithm–Based Selection of Informative Features
JMIR Biomed Eng 2020;5(1):e20932
URL: http://biomedeng.jmir.org/2020/1/e20932/
doi:10.2196/20932
PMID:

Original Paper

Interpretation of Maturity-Onset Diabetes of the Young Genetic Variants Based on American College of Medical Genetics and Genomics Criteria: Machine-Learning Model Development

Yichuan Liu[1*], PhD; Hui-Qi Qu[1*], MD, PhD; Adam S Wenocur[1], MSc; Jingchun Qu[1,2], BSc; Xiao Chang[1], PhD; Joseph Glessner[1], PhD; Patrick Sleiman[1,3,4], PhD; Lifeng Tian[1], PhD; Hakon Hakonarson[1,3,4], MD, PhD

[1]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, United States

[2]Texas A&M University, College Station, TX, United States

[3]Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, United States

[4]Division of Human Genetics, Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[*]these authors contributed equally

Corresponding Author:
Hakon Hakonarson, MD, PhD
Center for Applied Genomics
Children's Hospital of Philadelphia
3615 Civic Center Blvd
Philadelphia, PA, 19104
United States
Phone: 1 267 426 0088
Fax: 1 267 426 0363
Email: hakonarson@email.chop.edu

Abstract

Background: Maturity-onset diabetes of the young (MODY) is a group of dominantly inherited monogenic diabetes, with *HNF4A*-MODY, *GCK*-MODY, and *HNF1A*-MODY as the three most common forms based on the causal genes. Molecular diagnosis of MODY is important for precise treatment. Although a DNA variant causing MODY can be assessed based on the criteria of the American College of Medical Genetics and Genomics (ACMG) guidelines, gene-specific assessment of disease-causing mutations is important to differentiate among MODY subtypes. As the ACMG criteria were not originally designed for machine-learning algorithms, they are not true independent variables.

Objective: The aim of this study was to develop machine-learning models for interpretation of DNA variants and MODY diagnosis using the ACMG criteria.

Methods: We applied machine-learning models for interpretation of DNA variants in MODY genes defined by the ACMG criteria based on the Human Gene Mutation Database (HGMD) and ClinVar database.

Results: With a machine-learning procedure, we found that the weight matrix of the ACMG criteria was significantly different between the three MODY genes *HNF1A*, *HNF4A*, and *GCK*. The models showed high predictive abilities with accuracy over 95%.

Conclusions: Our results highlight the need for applying different weights of the ACMG criteria in relation to different MODY genes for accurate functional classification. As proof of principle, we applied the ACMG criteria as feature vectors in a machine-learning model and obtained a precision-based result.

(JMIR Biomed Eng 2020;5(1):e20506) doi:10.2196/20506

KEYWORDS

genetics; genomics; machine learning; medicine; maturity-onset diabetes of the young

Introduction

Monogenic diabetes results from DNA mutations in a single gene, which accounts for about 1%-4% of all cases of diabetes in the United States [1]. The most common form of monogenic diabetes is maturity-onset diabetes of the young (MODY), an autosomal dominant disease that most commonly occurs in adolescence or early adulthood [2]. Genetic sequencing is needed to identify the causal mutations and to diagnose different subtypes of MODYs [3]. The DNA variant causing MODY can be specifically assessed using the criteria established by the American College of Medical Genetics and Genomics (ACMG), as published in their guidelines [4]. Although the ACMG guidelines can be universally applied for all human DNA variants, our previous study suggested that a gene-specific assessment is important for identifying disease-causing mutations in different MODY genes [5]. In addition, contradictory evidence is commonly seen in functional classification of genetic variations when using the ACMG guidelines [6]. The ACMG guidelines may suggest a variant of uncertain significance; however, classification of the variant may have contradictory evidence, and some variants with contradictory evidence may turn out to have a reliable definite classification.

Machine learning has been advocated as an important tool for both clinical and research purposes in human diseases [7,8]. In this study, we aimed to develop machine-learning models for interpretation of DNA variants using the ACMG criteria, with a focus on DNA variants of three MODY genes (*HNF1A*, *HNF4A*, and *GCK*) underlying the three most common types of MODYs [9].

Methods

Data Collection for Machine-Learning Procedures

Known DNA variants of the three MODY genes *HNF1A*, *HNF4A*, and *GCK* were acquired from the dbSNP [10], the ClinVar database [11], and the Human Gene Mutation Database (HGMD) 2019 professional version [12]. Among the multihundred variants reported in these genes, approximately half have a classification of pathogenic/likely pathogenic (P/LP) according to the annotation in ClinVar or HGMD. According to the HGMD, the three genes were curated by Professor Andrew Hattersley, a leading genetic expert in MODYs [13]. The classification of variants as benign/likely benign (B/LB) varies between the different databases according to the annotation of ClinVar or dbSNP. Overall, for the three genes, there are 899 unique variants reported in *HNF1A*, including 569 P/LP sites and 330 B/LB sites; 1037 unique variants for *HNF4A*, including 182 P/LP sites and 855 B/LB sites; and 1664 unique variants for *GCK*, including 1065 P/LP sites and 599 B/LB sites. However, several of these variants have different annotation features between the different databases.

Feature Vector Generation

The feature vectors used for machine-learning modeling were the criteria based on the ACMG guidelines [14]. The criteria terms were generated based on InterVar [15], a computational tool that uses a preannotated or variant call format file as an input and generates automated interpretation based on the ACMG criteria. It should be noted that not all 33 ACMG criteria can be computationally scored. For example, the PS3 criterion requires well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product. As a result, the following 15 ACMG criteria were used, which was also the length of feature vectors for the three MODY genes: PVS1, PS1, PS4, PM1, PM2, PM4, PM5, PP2, PP3, PP5, BA1, BS1, BP4, BP6, and BP7.

Using machine-learning regression procedures, we normalized the weights for the evidence of different categories in accordance with the ACMG guidelines, assuming that the weight coefficient of PVS1 is 1, that of PS is 1/2, that of PM is 1/6, and that of PP is 1/12. We additionally assumed that the weight coefficient of BA1, BS, and BP is –1, –1/2, and –1/4, respectively. As the ACMG criteria were not originally designed for machine learning, these criteria are not true independent variables. Multicollinearity among feature vectors is commonly seen within each gene, which is the case for the PM1 and PP2 criteria. By definition, a PM1 hit means that the variant is located in a mutational hotspot or in a critical and well-established functional domain without benign variation, and a PP2 hit means that there is a missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease. In many situations, PM1 and PP2 are consistent with each other, which increases the risk of inappropriate weighting of the two criteria because of multicollinearity. To detect the collinearity among feature vectors, we calculated the variance inflation factor (VIF) and pairwise correlation coefficient for the ACMG criteria. Feature vectors with a VIF greater than 10 or a correlation coefficient larger than 0.8 were removed before the learning procedures.

Learning Procedures and Predictive Modeling

The machine-learning procedure used in this study was a typical logistic regression based on the Scikit-learn package in Python [16]. For detection of the weight matrix of the ACMG criteria, all variants, including P/LP and B/LB variants, were taken into account. For predictive modeling, we split the data based on 2-fold random shuffle processes. In other words, the P/LP and B/LB variants were split randomly into equally sized sets, with one set serving as training data and the other set serving as testing data, to determine the predictive capabilities of the model. This process was repeated 20 times to obtain the mean and standard deviation for accuracy measures, including sensitivity and specificity.

Results

Variation in the Weight Matrix of ACMG Criteria Among the Three MODY Genes

Based on the machine-learning procedure, we found that the weight matrix of the ACMG criteria was significantly different between the three MODY genes *HNF1A*, *HNF4A*, and *GCK* (Table 1, Figure 1). The differences are nontrivial and must be taken into consideration in clinical interpretation of DNA variants for genetic diagnosis.

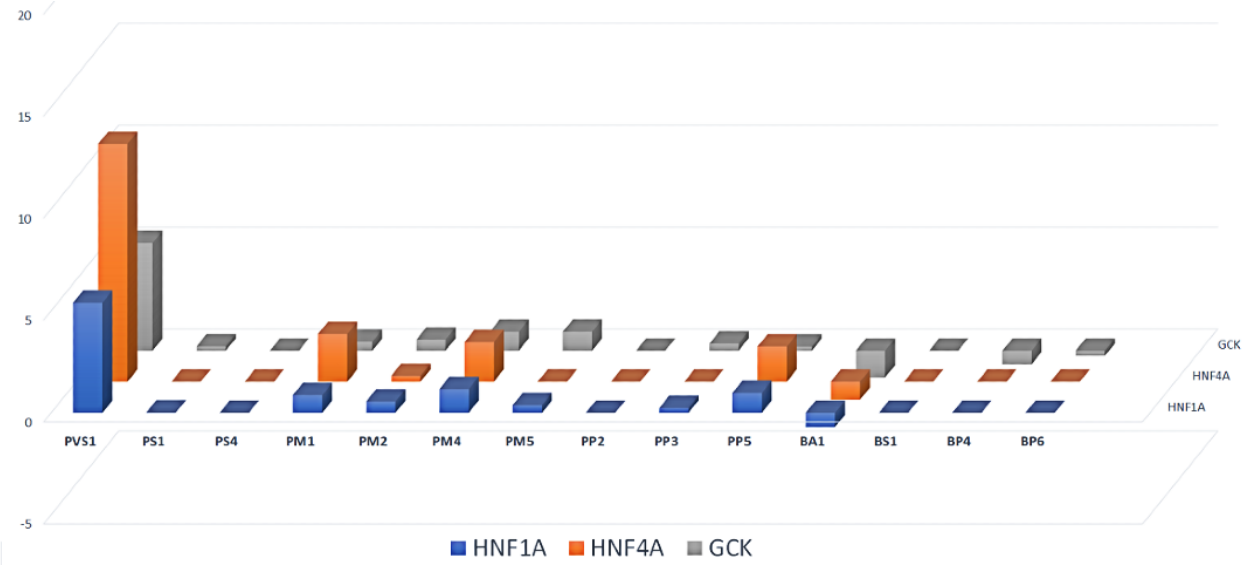
Table 1. Weight matrix of the American College of Medical Genetics and Genomics criteria for three maturity-onset diabetes of the young genes.

Criteria	Frequency						Weight		
	HNF1A P/LP[a]	*HNF1A* B/LB[b]	*HNF4A* P/LP	*HNF4A* B/LB	*GCK* P/LP	*GCK* B/LB	*HNF1A*	*HNF4A*	*GCK*
PVS1	0.3040	0.0030	0.2582	0.0000	0.2183	0.0033	5.401	11.665	5.275
PS1	0.0018	0.0000	0.0000	0.0000	0.0039	0.0000	0.018	0.000	0.210
PS4	0.0000	0.0030	0.0000	0.0000	0.0000	0.0000	0.000	0.000	0.000
PM1	0.5149	0.0030	0.4835	0.0000	0.6326	0.0083	0.866	2.340	0.442
PM2	0.9772	0.4273	0.7802	0.1450	0.9922	0.5092	0.555	0.290	0.542
PM4	0.0053	0.0000	0.0055	0.0000	0.0049	0.0000	1.167	1.944	0.923
PM5	0.0264	0.0000	0.0055	0.0000	0.1335	0.0000	0.375	0.007	0.927
PP2	0.0000	0.0000	0.5440	0.0012	0.0000	0.0000	0.000	0.000	0.000
PP3	0.6151	0.0061	0.5769	0.0000	0.7437	0.0050	0.235	0.000	0.347
PP5	0.1195	0.0000	0.1923	0.0000	0.1540	0.0033	0.975	1.727	0.202
BA1	0.0105	0.2091	0.0165	0.1708	0.0039	0.1703	-0.696	–0.880	–1.312
BS1	0.0228	0.5667	0.0440	0.5895	0.0078	0.4891	0.000	0.000	0.000
BP4	0.0193	0.0727	0.0165	0.0222	0.0458	0.0234	0.000	0.000	–0.665
BP6	0.0123	0.1182	0.0220	0.0480	0.0029	0.0234	0.000	0.000	–0.253
BP7	0.0123	0.0636	0.0110	0.0211	0.0400	0.0200	0.000	0.000	0.000

[a]P/LP: pathogenic/likely pathogenic variant.

[b]B/LB: benign/likely benign variant.

Figure 1. Weight matrix of three MODY genes, HNF1A, HNF4A and GCK: normalized weight for American College of Medical Genetics and Genomics criteria for three most common maturity-onset diabetes of the young genes.



Evidence for PS is rarely observed for the MODY variants. By contrast, evidence for PS4 (ie, the prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls) is commonly observed but is often misclassified. As an example, the *HNF1A* variant 12:121420807-G-A (rs1183910) was reported to be associated with C-reactive protein, a marker of inflammation, in a genome-wide association study [17]. However, as a common single nucleotide polymorphism with a minor allele frequency of 0.292 in European populations, this cannot be a variant causing the rare and dominantly inherited form of *HNF1A*-MODY.

With respect to evidence for PM criteria, PM1, which is defined as a variant located in a mutational hotspot or in a critical and well-established functional domain (eg, active site of an enzyme) without benign variation, and PM2 (absent from controls or at extremely low frequency if recessive in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium) are both commonly observed, in support of pathogenic variants in the three MODY genes. However, PM2

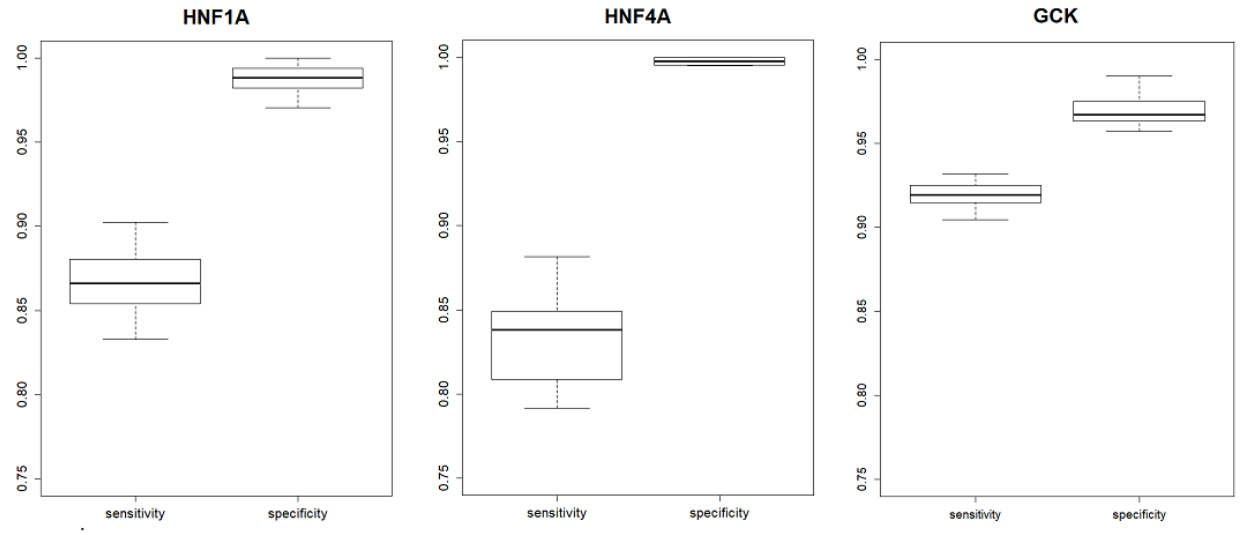
is also commonly seen among B/LB variants in these three genes, thus lacking specificity for functional classification. In this study, PM2 showed a VIF of 79.0 in *HNF1A* and a VIF of 247 in *GCK*. Therefore, although PM2 is much more common than PM1 for the three MODY genes, the weight of PM2 in *HNF1A* is lower than that of PM1.

With respect to the evidence for PP criteria, PP2 (missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease) is absent in *HNF1A* and *GCK*, but is commonly seen in *HNF4A*. However, PP2 showed a correlation coefficient of 0.932 with PM1, and therefore does not add substantial weight to the classification of P/LP variants in *HNF4A*.

Highly Accurate Predictive Ability for MODY Gene Pathogenicity

HNF4A-MODY (MODY1), *GCK*-MODY (MODY2), and *HNF1A*-MODY (MODY3) are the three most common types of MODYs, accounting for ~70% of all MODY genes [18]. Therefore, a predictive model that can accurately recognize pathogenic variants would be useful for the diagnosis of novel mutations in these genes. As described in the Methods section, we used 2-fold random shuffle testing with 50% of the 3600 mutations as training data and the other 50% as testing data, and repeated the analysis 20 times. The logistic regression machine-learning model showed overall accuracy above 95% (1676/1786) for MODY gene mutations (Figure 2). Both *HNF1A* (true negatives=163, false positives=2) and *HNF4A* (true negatives=428, false positives=0) had a specificity close to 100%, and the specificity in *GCK* was also above 95% (true negatives=289, false positives=10). This lower specificity is also consistent with the benign phenotype and mild clinical expression of *GCK*-MODY.

Figure 2. Overall accuracy based on logistic regression machine learning. The boxplot represents the sensitivity and specificity for 2-fold random shuffle tests.



These results proved the principle that ACMG criteria could be applied as meaningful feature vectors in a machine-learning model, and such a model based on ACMG criteria could provide accurate pathogenic classification for other Mendelian disease genes in a gene-specific manner.

Discussion

Our results highlight the need for applying different weights of the ACMG criteria in the functional classification of DNA variants of different MODY genes. In the past decade, sequencing technologies have evolved rapidly with the advance of high-throughput next-generation sequencing (NGS). By adopting NGS, clinical laboratories are now performing an ever-increasing volume of genetic testing for genetic disorders. However, increased complexity in genetic testing has been accompanied by new challenges in sequence interpretation, and multiple new standards have been implemented for physicians and genetic counselors regarding the interpretation and reporting of sequence variants at different levels of pathogenicity. Currently, there are multiple computational tools available based on different algorithms and databases that are being used to predict the pathogenicity of DNA variants, such as SIFT [19], MutationTaster [20], likelihood ratio test [21], FATHMM by a supervised machine-learning model [22], GERP++ by maximum-likelihood evolutionary rate estimation [23] for coding variants, and DANN for both coding and noncoding variants using a deep neural network [24]. However, all of these computational tools assess each gene with a common rule, which is not based on biology, whereas this study proposes that a gene-specific assessment for pathogenicity is required, at least for MODY genes [5].

The evolutionary selection pressures on MODYs vary across different genes, and is considered to be the lowest in the case of *GCK*-MODY [25]. Similar issues exist with functional classification based on the ACMG criteria, which are globally applied for all human genes. The ACMG criteria contain 33 terms that lead to five categories of mutations (“pathogenic,”

"likely pathogenic," "uncertain significance," "likely benign," and "benign"), as one of the most commonly used standards.

MODY represents a group of dominantly inherited monogenic diabetes, and *HNF4A*-MODY (MODY1), *GCK*-MODY (MODY2), and *HNF1A*-MODY (MODY3) are the three most common subtypes of MODY. These MODY genes are involved in different molecular pathways. MODY variants of different genes show different clinical features and thus require different treatments. For example, *HNF1A*-MODY is characterized by a reduced beta cell mass or impaired function, and has been treated with sulfonylureas for decades with excellent results [26]. Patients with *HNF1A*-MODY are highly sensitive to sulfonylurea treatment and may be susceptible to developing hypoglycemia during the treatment [26]. *HNF4A*-MODY has similar clinical features with *HNF1A*-MODY, and the affected transcription network plays a role in the early development of the pancreas. The pancreatic beta cells produce adequate insulin in infancy but the capacity for insulin production declines thereafter [27]. The beta cells in *GCK*-MODY have a normal capacity to make and secrete insulin, but do so only above an abnormally high glucose threshold, which results in a chronic, mild increase in blood sugar that is usually asymptomatic [25]. Accordingly, treatment of *GCK*-MODY can be achieved by a healthy diet and exercise, while oral hypoglycemic agents or insulin is of no benefit for these patients [25]. Therefore, accurate molecular diagnosis of these MODYs is important for precise treatment.

In conclusion, we applied a computational machine-learning method together with the ACMG criteria for functional classification of genetic variants of the three most common MODY genes, *HNF1A*, *HNF4A* and *GCK*. Our results show that a typical machine-learning model using 15 computational ACMG criteria as the feature vector has predictive abilities that are highly accurate (>95% accuracy) for hundreds of annotated variants in three MODY genes. Therefore, this model could serve as a fast, gene-specific method for physicians or genetic counselors assisting with diagnosis and reporting, especially when confronted by contradictory ACMG criteria. Moreover, we show that the weight of the ACMG criteria exhibits gene specificity, which advocates for the application of machine-learning methods with the ACMG criteria to capture the most relevant information for each disease-related variant.

Acknowledgments

The study was supported by Institutional Development Funds from the Children's Hospital of Philadelphia to the Center for Applied Genomics, and The Children's Hospital of Philadelphia Endowed Chair in Genomic Research to HH. We thank the Center for Applied Genomics staff and support from Children's Hospital of Philadelphia.

Authors' Contributions

YL and HQ conceptualized and designed the study, drafted the initial manuscript, and reviewed and revised the manuscript. AW, JQ, XC, JG, PS, and LT collected data, carried out the initial analyses, and reviewed and revised the manuscript. HH conceptualized and designed the study, and critically reviewed the manuscript.

Conflicts of Interest

None declared

References

1. National Institute of Diabetes and Digestive and Kidney Diseases. URL: https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/monogenic-neonatal-mellitus-mody [accessed 2020-05-20]
2. Shields BM, Hicks S, Shepherd MH, Colclough K, Hattersley AT, Ellard S. Maturity-onset diabetes of the young (MODY): how many cases are we missing? Diabetologia 2010 Dec 25;53(12):2504-2508. [doi: 10.1007/s00125-010-1799-4] [Medline: 20499044]
3. Rubio-Cabezas O, Hattersley AT, Njølstad PR, Mlynarski W, Ellard S, White N, International Society for Pediatric Adolescent Diabetes. ISPAD Clinical Practice Consensus Guidelines 2014. The diagnosis and management of monogenic diabetes in children and adolescents. Pediatr Diabetes 2014 Sep 03;15(Suppl 20):47-64. [doi: 10.1111/pedi.12192] [Medline: 25182307]
4. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015 May;17(5):405-424 [FREE Full text] [doi: 10.1038/gim.2015.30] [Medline: 25741868]
5. Li Q, Liu X, Gibbs RA, Boerwinkle E, Polychronakos C, Qu H. Gene-specific function prediction for non-synonymous mutations in monogenic diabetes genes. PLoS One 2014;9(8):e104452 [FREE Full text] [doi: 10.1371/journal.pone.0104452] [Medline: 25136813]
6. Qu H, Wang X, Tian L, Hakonarson H. Application of ACMG criteria to classify variants in the human gene mutation database. J Hum Genet 2019 Nov 26;64(11):1091-1095. [doi: 10.1038/s10038-019-0663-8] [Medline: 31451714]

7. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
8. Pande A, Mohapatra P, Nicorici A, Han JJ. Machine Learning to Improve Energy Expenditure Estimation in Children With Disabilities: A Pilot Study in Duchenne Muscular Dystrophy. JMIR Rehabil Assist Technol 2016 Jul 19;3(2):e7 [FREE Full text] [doi: 10.2196/rehab.4340] [Medline: 28582264]
9. Pihoker C, Gilliam LK, Ellard S, Dabelea D, Davis C, Dolan LM, SEARCH for Diabetes in Youth Study Group. Prevalence, characteristics and clinical diagnosis of maturity onset diabetes of the young due to mutations in HNF1A, HNF4A, and glucokinase: results from the SEARCH for Diabetes in Youth. J Clin Endocrinol Metab 2013 Oct;98(10):4055-4062 [FREE Full text] [doi: 10.1210/jc.2013-1279] [Medline: 23771925]
10. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001 Jan 01;29(1):308-311 [FREE Full text] [doi: 10.1093/nar/29.1.308] [Medline: 11125122]
11. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 2014 Jan;42(Database issue):D980-D985 [FREE Full text] [doi: 10.1093/nar/gkt1113] [Medline: 24234437]
12. Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet 2017 Jun;136(6):665-677 [FREE Full text] [doi: 10.1007/s00439-017-1779-6] [Medline: 28349240]
13. Human Gene Mutation Database. URL: http://www.hgmd.cf.ac.uk/docs/new_back.html [accessed 2020-05-20]
14. Bahcall OG. Genetic testing. ACMG guides on the interpretation of sequence variants. Nat Rev Genet 2015 May;16(5):256-257. [doi: 10.1038/nrg3940] [Medline: 25854183]
15. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am J Hum Genet 2017 Feb 02;100(2):267-280 [FREE Full text] [doi: 10.1016/j.ajhg.2017.01.004] [Medline: 28132688]
16. Pedregosa F, Varoquaux G, Gramfort A, Vincent M, Thirion B. Scikit-learn: Machine learning in Python. J Machine Learn Res 2011;12:2825-2830 [FREE Full text]
17. Ligthart S, Vaez A, Hsu Y, Inflammation Working Group of the CHARGE Consortium, PMI-WG-XCP, LifeLines Cohort Study, et al. Bivariate genome-wide association study identifies novel pleiotropic loci for lipids and inflammation. BMC Genomics 2016 Jun 10;17:443 [FREE Full text] [doi: 10.1186/s12864-016-2712-4] [Medline: 27286809]
18. Ellard S, Bellanné-Chantelot C, Hattersley AT, European Molecular Genetics Quality Network (EMQN) MODY group. Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young. Diabetologia 2008 Apr 23;51(4):546-553 [FREE Full text] [doi: 10.1007/s00125-008-0942-y] [Medline: 18297260]
19. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003 Jul 01;31(13):3812-3814 [FREE Full text] [doi: 10.1093/nar/gkg509] [Medline: 12824425]
20. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 2014 Apr;11(4):361-362. [doi: 10.1038/nmeth.2890] [Medline: 24681721]
21. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res 2009 Sep;19(9):1553-1561 [FREE Full text] [doi: 10.1101/gr.092619.109] [Medline: 19602639]
22. Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. Bioinformatics 2018 Feb 01;34(3):511-513 [FREE Full text] [doi: 10.1093/bioinformatics/btx536] [Medline: 28968714]
23. Davydov E, Goode D, Sirota M, Cooper G, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol 2010 Dec 02;6(12):e1001025 [FREE Full text] [doi: 10.1371/journal.pcbi.1001025] [Medline: 21152010]
24. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics 2015 Mar 01;31(5):761-763 [FREE Full text] [doi: 10.1093/bioinformatics/btu703] [Medline: 25338716]
25. Chakera AJ, Steele AM, Gloyn AL, Shepherd MH, Shields B, Ellard S, et al. Recognition and Management of Individuals With Hyperglycemia Because of a Heterozygous Glucokinase Mutation. Diabetes Care 2015 Jul 23;38(7):1383-1392. [doi: 10.2337/dc14-2769] [Medline: 26106223]
26. Pearson ER, Starkey BJ, Powell RJ, Gribble FM, Clark PM, Hattersley AT. Genetic cause of hyperglycaemia and response to treatment in diabetes. Lancet 2003 Oct 18;362(9392):1275-1281. [doi: 10.1016/S0140-6736(03)14571-0] [Medline: 14575972]
27. Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, et al. Control of pancreas and liver gene expression by HNF transcription factors. Science 2004 Feb 27;303(5662):1378-1381 [FREE Full text] [doi: 10.1126/science.1089769] [Medline: 14988562]

Abbreviations

ACMG: American College of Medical Genetics and Genomics
B/LB: benign/likely benign

HGMD: Human Gene Mutation Database
MODY: maturity-onset diabetes of the young
NGS: next-generation sequencing
P/LP: pathogenic/likely pathogenic
VIF: variance inflation factor

Edited by R Kukafka; submitted 30.05.20; peer-reviewed by C Doogan, F Palmieri; comments to author 22.06.20; revised version received 19.07.20; accepted 03.11.20; published 01.12.20.

Please cite as:
Liu Y, Qu HQ, Wenocur AS, Qu J, Chang X, Glessner J, Sleiman P, Tian L, Hakonarson H
Interpretation of Maturity-Onset Diabetes of the Young Genetic Variants Based on American College of Medical Genetics and Genomics Criteria: Machine-Learning Model Development
JMIR Biomed Eng 2020;5(1):e20506
URL: http://biomedeng.jmir.org/2020/1/e20506/
doi:10.2196/20506
PMID:

Original Paper

Challenges and Opportunities in Collecting and Modeling Ambulatory Electrodermal Activity Data

Donna L Coffman[1], PhD; Xizhen Cai[2], PhD; Runze Li[3], PhD; Noelle R Leonard[4], PhD

[1]Department of Epidemiology and Biostatistics, Temple University, Philadelphia, PA, United States
[2]Department of Mathematics and Statistics, Williams College, Williamstown, MA, United States
[3]Department of Statistics, Pennsylvania State University, State College, PA, United States
[4]Rory Meyers College of Nursing, New York University, New York City, NY, United States

Corresponding Author:
Donna L Coffman, PhD
Department of Epidemiology and Biostatistics
Temple University
1301 Cecil B. Moore Ave
Ritter Annex, 9th floor
Philadelphia, PA, 19122
United States
Phone: 1 2152044420
Email: dcoffman@temple.edu

Abstract

Background: Ambulatory assessment of electrodermal activity (EDA) is an emerging technique for capturing individuals' autonomic responses to real-life events. There is currently little guidance available for processing and analyzing such data in an ambulatory setting.

Objective: This study aimed to describe and implement several methods for preprocessing and constructing features for use in modeling ambulatory EDA data, particularly for measuring stress.

Methods: We used data from a study examining the effects of stressful tasks on EDA of adolescent mothers (AMs). A biosensor band recorded EDA 4 times per second and was worn during an approximately 2-hour assessment that included a 10-min mother-child videotaped interaction. The initial processing included filtering noise and motion artifacts.

Results: We constructed the features of the EDA data, including the number of peaks and their amplitude as well as EDA reactivity, quantified as the rate at which AMs returned to baseline EDA following an EDA peak. Although the pattern of EDA varied substantially across individuals, various features of EDA may be computed for all individuals enabling within- and between-individual analyses and comparisons.

Conclusions: The algorithms we developed can be used to construct features for dry-electrode ambulatory EDA, which can be used by other researchers to study stress and anxiety.

(JMIR Biomed Eng 2020;5(1):e17106) doi:10.2196/17106

KEYWORDS

electrodermal activity; functional data analysis; ambulatory stress assessment

Introduction

Background

Electrodermal activity (EDA), commonly known as galvanic skin response, is a measure of sympathetic nervous system activity that is used to assess physiological arousal. Measurement of and methods for examining EDA in the laboratory setting have been well described [1], but there is less guidance for ambulatory EDA. The aim of this study was to describe and illustrate several analytic methods, specifically functional data analysis (FDA) [2] and local polynomial regression with autoregressive (AR) errors [3], neither of which have previously been applied to EDA data. We also addressed the challenges of collecting ambulatory EDA, denoising the data, and detecting meaningful features common across participants. We concluded with opportunities and limitations of the future use of ambulatory EDA data.

Electrodermal Activity

EDA has been measured in the laboratory using gel electrodes typically placed on the palm. Entire books [1] and publication standards (eg, Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures [4]) have been written about the measurement and the use of EDA data collected in the laboratory. However, much less has been written about the measurement of EDA in ambulatory settings [5], which has recently become popular because of the availability of dry-electrode wrist-worn devices that can be used while participants go about their daily activities. These devices have dry electrodes that are placed on the ventral wrist as opposed to laboratory placement on the palm. Emotional effects on EDA can be more sensitively recorded with electrodes on the palm rather than the wrist, but obviously, electrodes on the palm would interfere with normal activities. However, dry electrodes used in ambulatory measurement have uncertain contact with the skin in comparison with gel electrodes used in laboratory-based measurements. Important issues in the measurement of ambulatory EDA are the influence of temperature (both ambient and skin) and physical activity [4]. Several wrist-worn devices have been developed for the ambulatory collection of EDA as well as temperature and three-axis accelerometry [6,7] (eg, Affectiva Q sensor) and, more recently, blood volume pulse (eg, Empatica E4). This study used the iCalm device [8], which measures temperature and accelerometry as well as EDA (the *Methods* section provides a detailed description about this).

All these devices record the overall skin conductance (SC), which can be decomposed into what is referred to as tonic and phasic activity components [9-11]. Tonic activity varies slowly and is also referred to as skin conductance level (SCL). Individuals have different levels of tonic activity, that is, some individuals have a higher level of tonic activity and others have a lower level of tonic activity. In contrast, phasic activity varies rapidly in response to stimuli such as stress and is also referred to as the skin conductance response (SCR). The SCR amplitude is considered to be a measure of sympathetic activity. SCR is characterized by a rapid incline to a peak and then a slower decline back to the individual's SCL [1]. Typically, the goal is to detect these peaks and compute their features (eg, amplitude), which can serve as predictors or outcomes in statistical models. Although the features can be detected by visualizing the data and coding the peaks, the amount of data that are typically collected in ambulatory research studies makes this time consuming and impractical. Next, we introduce the study from which the data were obtained before returning to approaches for analyzing ambulatory EDA data.

Power Source Parenting Study

Many adolescent mothers (AMs) have challenges effectively regulating emotion, which interferes with their use of positive parenting skills and places them at risk of maltreating their children. Data for this study were collected from participants in a pilot, group-randomized controlled trial of the Power Source Parenting (PSP) intervention for high-risk AMs [12,13]. The AMs were homeless and lived with their children in transitional living programs (TLPs) in a Northeast state. Each participant received a baseline and 2 postintervention follow-up assessments (at 3 and 6 months). Each assessment lasted between 1.5 and 2 hours, and the AMs wore the iCalm biosensor band that measured EDA throughout each of the 3 assessments, which included 2 potentially stressful tasks: (1) a timed Stroop task and (2) a 10-min videotaped mother-child interaction. We examined EDA at each assessment for the Stroop and video interaction tasks. We chose these 2 tasks because we expected that they induce different types of stressors (eg, Stroop task is a cognitive stressor) and they have differing degrees of structured activity. The Stroop task was more structured than the videotaped interaction task, and it more closely resembles the measurement of EDA in the laboratory (ie, AMs complete the computerized task seated alone without her child present). In contrast, the semistructured mother-child videotaped interaction does not have the constraints of carefully controlled laboratory settings but provides more ecologically valid information about maternal and child behaviors concurrent with measurement of AMs' EDA that can be examined across participants.

In our study, different AMs had quite different features in their EDA profiles (ie, the EDA level over the time span of a single task), making it difficult to compare them directly. Some of the differences are noise, such as those from movement artifacts. Despite innovations in devices, ambulatory assessment of EDA is inherently noisier than laboratory assessment because of these movement artifacts and the dry electrodes. Thus, the first step in processing EDA data is to identify these artifacts so that these data can be discarded. The step typically involves applying a signal processing filter, such as a Hanning filter with a 1-second window [14], a zero-phase order 10 low-frequency Butterworth filter [5], or a finite impulse response (FIR) low-pass filter [15]. The application of a signal processing filter is not mandatory, as the next step, identifying valid versus invalid data, can be applied directly to the raw, unfiltered data [15]. Taylor et al [15] developed an algorithm using support vector machines (SVMs) to identify valid versus invalid data and applied it to both the raw SC data and SC data filtered using an FIR low-pass filter. SVMs have a disadvantage in that it is not transparent what rules are being implemented to classify data as valid versus invalid. In contrast, a rule-based algorithm was proposed [16] for classifying data as valid or not. For example, values of less than 0.05 μS could be excluded [5]. A disadvantage of this type of approach is that it is less flexible; however, it does not have the reproducibility problems that the Taylor et al [15] approach may have [16]. Usually, a combination of a signal processing filter and either a rule-based approach or a machine learning approach is used to identify and discard invalid data [5,15]. Given the valid data, the next step is to separate the tonic and phasic components.

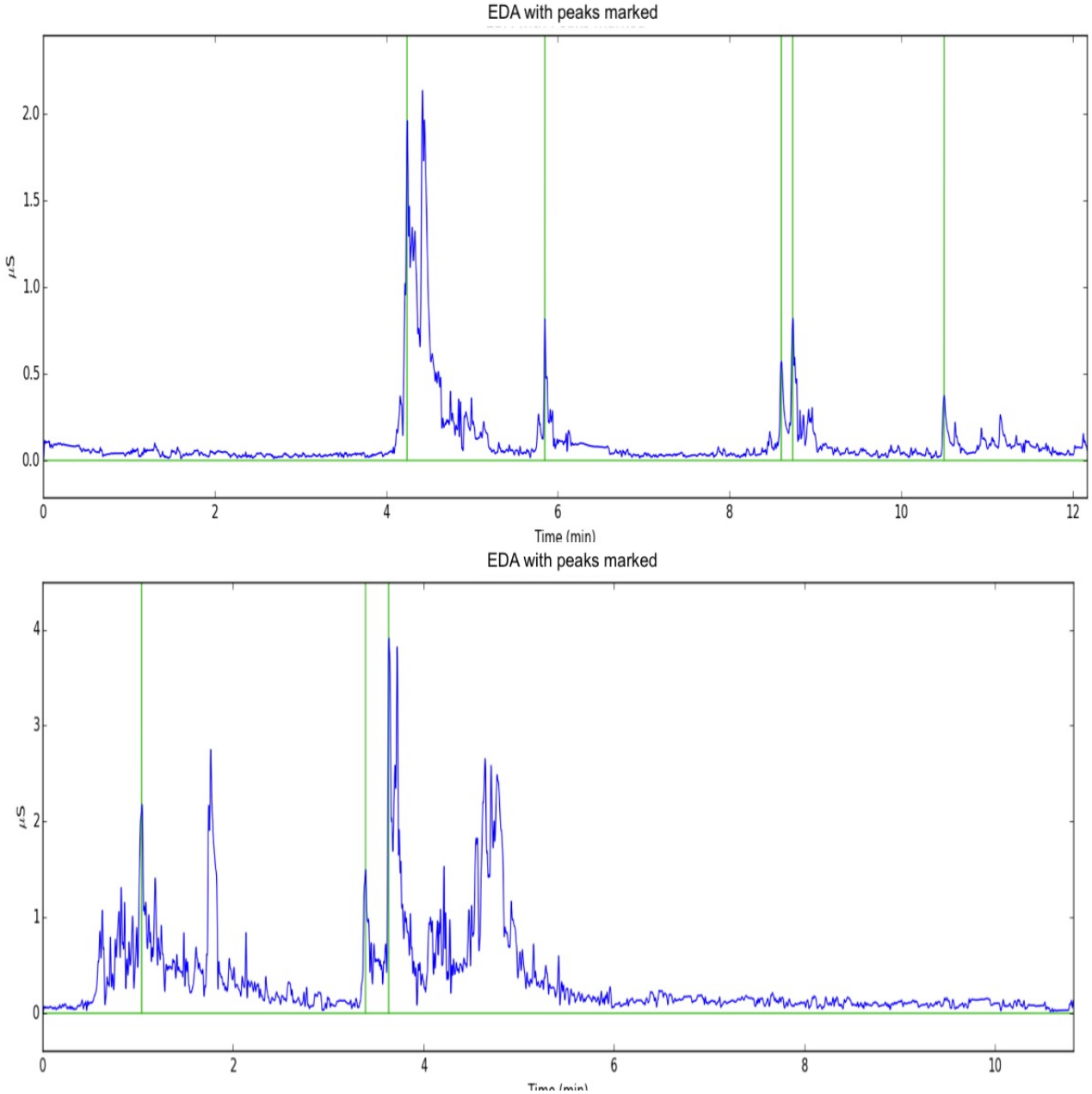
Many of the previous methods for separating the components are based on EDA data collected in the laboratory in which the participant is exposed to a stimulus, such as a loud noise, at regular intervals (eg, interstimulus interval of 8 seconds). These methods include the sigmoid exponential model [11] and negative deconvolution [9]. Although these methods were developed for data collected in the laboratory, the negative deconvolution method has been applied to ambulatory data [14].

There is a question as to whether it is important to separate these components in ambulatory data because the stimuli are typically more gradual (typically referred to as nonspecific SC responses [1]) as opposed to the abrupt startle response stimuli typically used in laboratory research. Hernandez et al [14] analyzed the data without decomposing the tonic and phasic components and by decomposing them using nonnegative deconvolution and found that the decomposition provided more discriminative information. Thus, they recommended decomposing the EDA. Regardless of whether the EDA is decomposed or not, the next step is to identify features of the data, typically peaks, and compute the characteristics of these features.

One automated approach to detecting peaks and constructing features is available on an interactive website at the MIT Media Lab [15] and uses SVMs, although the focus was primarily on identifying valid versus invalid data. Kleckner et al [16] also did not focus on detecting peaks but only on identifying valid versus invalid data. Thus, there is very little guidance on detecting peaks once data have been identified as valid. We initially used the Taylor et al [15] approach and discovered that the approach did not identify many of the peaks and marked much of the data as invalid. The results of applying this approach are presented in Figure 1 for one AM as an example and are what motivated us to try to find a better approach to detecting peaks in the ambulatory setting.

In general, our approach was to (1) use the raw SC data or SC data filtered with an FIR low-pass filter, (2) identify valid versus invalid data using rules established in the literature, (3) decompose the signal into its tonic and phasic components using negative deconvolution, (4) apply a spline-based smoothing technique (eg, B-spline regression or local polynomial regression with AR errors) to the EDA profile, and (5) compute various features of the peaks identified from the smoothed curve. The latter 2 steps, particularly the spline-based approaches, are the unique contributions of this study. Comparisons between different groups of individuals can then be performed on the peak features, if desired.

Figure 1. Electrodermal activity peak detection results for the free-play (first 5 min) and teaching (second 5 min) videotaped interaction tasks for one control group mother (patient ID=102) and at baseline and 3-month follow-up assessments. EDA: electrodermal activity.



Methods

iCalm Sensor Band

The iCalm applies a constant voltage to the skin through 2 silver/silver chloride metal electrodes and measures the resulting current [8,13]. A transconductance amplifier is used to convert the current to a voltage, which is sampled at a rate of 4 Hz using a 12-bit analog-to-digital converter. Three-axis accelerometry and temperature were sampled at the same rate. The iCalm has been validated against laboratory measures of EDA [6]. The iCalm connects via Bluetooth to Android-based mobile phones.

Participants

The data were obtained from the aforementioned PSP intervention study of 71 mothers. The EDA values for some mothers were very low, which was primarily because of a lack of contact between the skin and the electrodes. The general guidance in the literature is to discard data in which the EDA is approximately 0 [16], as it indicates that the electrodes were not in good contact with the skin. Details regarding these invalid data are given in section Data Preprocessing. Finally, there were 5 AMs who did not return at any of the follow-up assessments and several AMs who had data at one but not both of the follow-up assessment periods; thus, we used 6-month follow-up data if an AM did not have data at the 3-month follow-up. The final sample size numbers are given following discussion of the identification of valid versus invalid data.

Procedures

The AMs were interviewed at the TLP in a private room by trained interviewers. The interviews included a survey questionnaire examining demographic and background characteristics; questions regarding mental health, parenting,

and risk-taking behaviors; as well as activities described in the *Measures* section. AMs wore the iCalm band on their right wrist throughout the duration of each interview. Following the interview, mothers in the intervention TLPs received the PSP intervention, whereas the control TLPs received standard care [12].

Measures

Each assessment included the computer-delivered survey, a variety of computerized tasks including the Stroop task, and the 10-min videotaped interaction of mothers with their child. In this analysis, we focused on the Stroop task and the videotaped interaction task. The Stroop task involved reading the written color names independent of the color name that is written as quickly as possible in 1 min. For the videotaped interaction task, the mothers were asked to play with their child as they usually do for the first 5 min, which we will refer to as free play. For the second 5 min of the interaction task, the mothers were asked to teach their child a concept that was slightly above their developmental level. This latter task, which we will refer to as teaching, was intended to be slightly frustrating for the mother and the child. In addition, a physical task (going up and down stairs for approximately 5 min) was only performed at the baseline assessment and only for the intervention group AMs. We examined these data as a comparison task in which there is a lot of movement.

Data Analysis

The data analysis for analyzing and comparing the EDA profiles entails several steps described earlier, including preprocessing, smoothing, identifying peaks, computing features of the peaks, and, finally, statistical modeling. In this section, we present the details of these steps with regard to our data. The R code for implementing the analysis is provided in Multimedia Appendix 1. To date, software for identifying invalid data, decomposing SC, detecting peaks, and constructing features has been written in either MATLAB or Python. To our knowledge, there are no R packages for handling these tasks.

Data Preprocessing

Data preprocessing involved computing the elapsed time since the beginning of the Stroop and video interaction tasks for each mother at the baseline and follow-up assessments so that both assessments and all AMs would be on the same time scale. Next, we constructed plots of EDA for each mother by task to visualize motion artifacts and loose band connections. For the filtered data, we generated an FIR low-pass filter using the function *fir1* in the R package *signal* and chose the number of coefficients as 16 (ie, the sample rate×4). We considered EDA values that were approximately 0 as invalid, as these usually indicate that the band had a loose connection and did not accurately record EDA [5]. These data were removed from both the filtered and raw SC data. We also examined plots of the EDA during the physical task to ensure that, as expected, EDA increased with physical activity and that the band had a good connection. Finally, each mother at each of the baseline and follow-up assessments had differing ranges of EDA values (eg, refer to the y-axis in Figure 1). Hence, in accordance with standard practice [14], we normalized EDA values by the range and distance from the maximum value:

where the one minus is to correct the sign of the normalized EDA values to be the same as the original EDA values. Next, we used 2 methods for *data smoothing*: FDA with B-splines and local polynomial regression for data with AR errors.

Functional Data Analysis

FDA considers the EDA profile as a single entity rather than as a collection of data points. Missing data or unequally spaced measurements are smoothed over. The target function is estimated by a linear combination of basis functions, specifically spline functions (hence the name B-splines) of a specific polynomial order *m*. The entire range of the target function is divided into subintervals at breakpoints called knots. Splines are then fitted in each interval and are adjusted to join the knots to estimate the underlying smoothed function. We also added a penalty on the roughness of the fitted curve, with the tuning parameter chosen by the generalized cross-validation (GCV) index. In addition, the number of knots was chosen by the method proposed by Ruppert [17]. We implemented this approach using the *fda* R package to smooth the data and create functional profiles. Specifically, we used the function *create.bspline.basis* to generate B-spline basis functions and then used the functions *fdPar* and *smooth.basis* to produce the fitted functional form as well as the corresponding value for the GCV index. The first 2 derivatives of the estimated function were calculated and saved for future use. The estimated smoothed curve captures the general features of the normalized EDA profile. The results could change significantly for the choice of location of knots and the number of knots. We chose the knots to be equally spaced on the time span, but it is not necessary that they be equally spaced [18]. If they are not equally spaced, then the spacing may need to be customized for each individual. For example, knots could be placed more closely together when the EDA profile is changing significantly and placed further apart when the EDA profile does not change as much (ie, when the EDA profile is flat).

Local Polynomial Regression

Local polynomial regression with AR errors is another technique to obtain a smooth estimate of the EDA profile. We applied a version of local linear regression for data with AR errors [3] because the EDA data in a profile is longitudinal and from the same individual (hence correlated). This approach assumes that the response variable is the sum of a smoothed unknown function of time and an error term, where the error term follows an AR model with d terms:

For a fixed *d*, the estimation procedure starts by obtaining an initial estimate  by assuming working independence among observations and then calculating the corresponding residuals  . The initial estimate  is calculated by local linear regression with the plug-in bandwidth selector. The problem can be converted to a partial linear problem as follows:

and  and  are estimated by profile least squares. More specifically, we first estimate  by the so-called *difference-based method* [19] and then derive a local linear smoother on the difference:

as an estimate of the final  . The number of AR terms, *d*, is selected by a penalization method, with the tuning parameter selected by the Bayesian information criterion [20]. The local linear approach may not work well in cases where the fitted curve is lower than the actual EDA profile or is influenced by some outliers with extreme values. The local linear regression may also over smooth the fitted curves and therefore miss capturing peaks in the EDA profiles. In these cases, local quadratic regression or local cubic regression may yield better results. We applied both the local quadratic and local cubic regressions, but the fit was similar to the local linear regression; hence, we present only the local linear regression results in the Results section. Other literature [15] on analyzing EDA data suggests using a low-pass filter to denoise data first. Thus, we also tried the combination of an FIR low-pass filter and local linear regression as well as using local linear regression alone.

Identifying Peaks

Using the smoothed EDA profiles, our interest lies in identifying features related to significant local maxima (ie, peaks). Correct peaks should increase quickly and then decrease more slowly, with large enough rise and drop [1]. To identify peaks, we used the first and second derivatives saved earlier during the spline smoothing estimation. The local maxima are locations where the first derivative is 0 and the second derivative is negative. We first found all local maxima and then filtered them by the magnitude of the rises/drops. Specifically, for each local maximum, we identified the nearest local minimum on each side and then calculated the corresponding drop. If both drops are larger than a preset threshold, the local maximum is identified as a peak. The threshold was set to 0.1 [1]. There could be an incomplete peak at the right limit of the EDA profile, which we did not identify as a useful peak even if its drop satisfies the 0.1 criterion.

Computing Features of the Identified Peaks

To make a comparison between different individuals, we computed the characteristics or features of the identified peaks, including the number of identified peaks, time to the first peak, time to the highest peak, and maximum value (ie, amplitude) at the first peak. Initially, we also considered the area under the entire smoothed curve. However, the actual observed time span for each individual was not exactly the same, although the actual task was designed to last the same amount of time, making the area under the curve (AUC) not very comparable across individuals for a given task. Nevertheless, for data in which the time span is the same for each individual, AUC may be a meaningful feature. For each task, we calculated the abovementioned peak features for each AMs' EDA profile. The R code for identifying and computing the features of the peaks is provided in Multimedia Appendix 1.

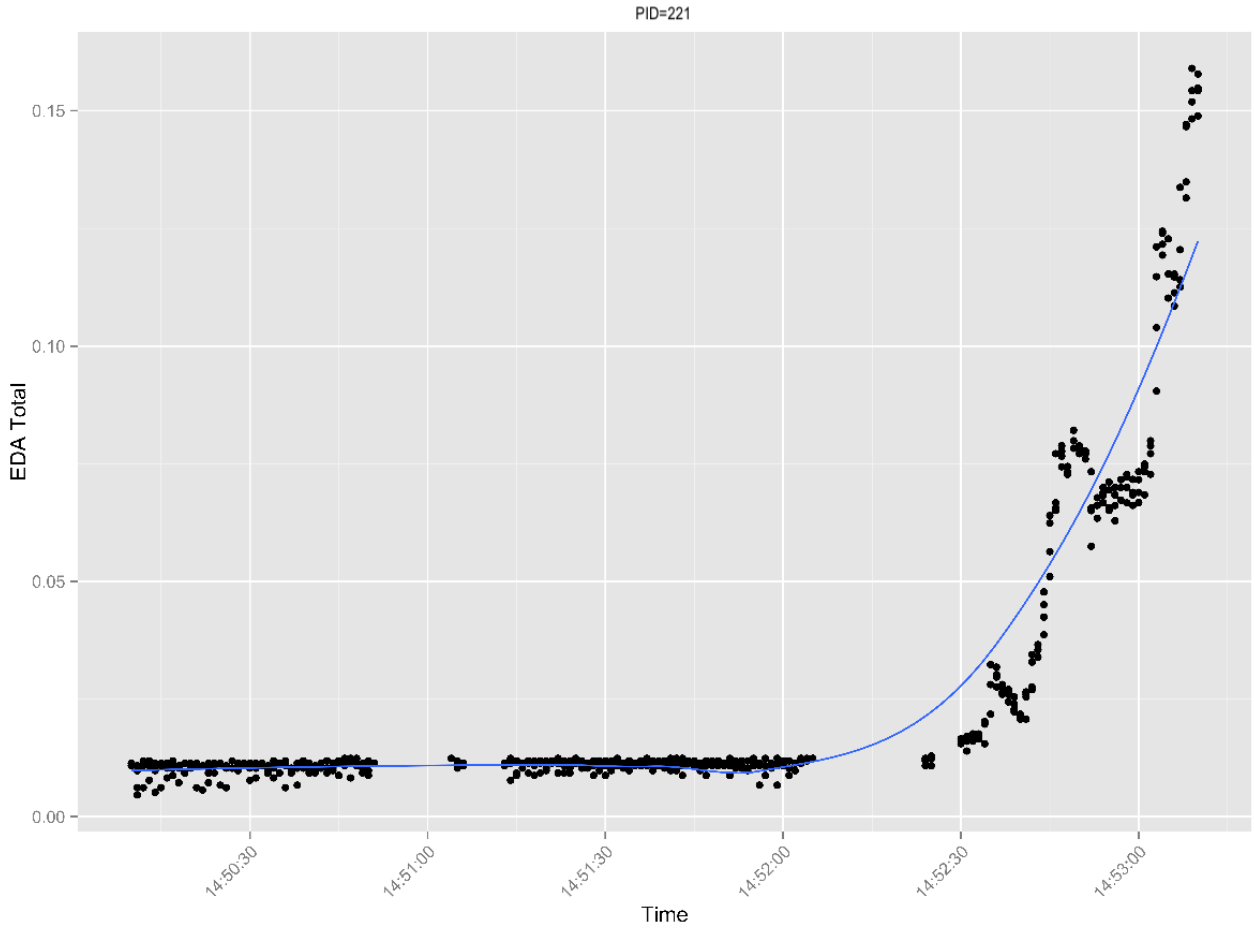
The final step is statistical modeling using the features of EDA peaks. This step depends, of course, on whether there are groups to be compared and repeated features on the same individual (eg, baseline and postintervention). Thus, any statistical model could potentially be used depending on the characteristics of a particular study.

Results

We began by examining plots of the EDA for each mother by task. Graphical display of the data showed EDA varied both between individuals and within individuals during both tasks. Owing to space, we did not present figures for all the AMs. Instead, we focused on one control group AM (patient ID [PID] 102) and one intervention group AM (PID 202).

We also examined plots of the EDA data for the physical task for the intervention AMs at baseline. The examination of these plots revealed that the AMs' EDA increased, as it should (Figure 2 shows an example plot of one mother) during the physical task.

Figure 2. Electrodermal activity for physical task at baseline for one intervention group mother. A loess curve has been fit to the data. EDA: electrodermal activity; PID: patient ID.



Local Polynomial Regression

Next, we present the results of the local linear regression with AR errors. Figure 3 shows the smoothed function overlaying the normalized EDA values for the baseline and 3-month follow-up assessments for the teaching task for one control group AM. Figure 4 presents the smoothed function overlaying the normalized EDA values for baseline and 3-month follow-up assessments for the free-play task for the same AM, as presented in Figure 3. In the case of this AM, the smoothed function does not fit the actual peaks in the data very well. Specifically, the smoothed peaks are dampened in comparison with the actual peaks, which creates a problem for identifying peaks because they are defined by their increase/decrease. Even if peaks are identified, some features of the peaks (eg, amplitude) would be dampened in comparison with the actual data. We also tried using an FIR low-pass filter before smoothing with local linear regression for AR errors. However, this did not substantially improve the fit of the estimated functions to the actual data. Figure 5 shows the smoothed function overlaying the normalized EDA values after applying the FIR low-pass filter for the baseline and 3-month follow-up assessments for the teaching task for the same AM, as presented in Figure 3. The free-play task is presented in Figure 6. It is evident that the FIR low-pass filter did not change the estimated EDA profiles substantially from the raw SC data. We next attempted smoothing using B-splines to estimate the functional EDA profiles.

Figure 3. Local linear regression with autoregressive errors for the teaching task at baseline and 3-month follow-up assessments for one mother (patient ID=102). EDA: electrodermal activity; PID: patient ID.

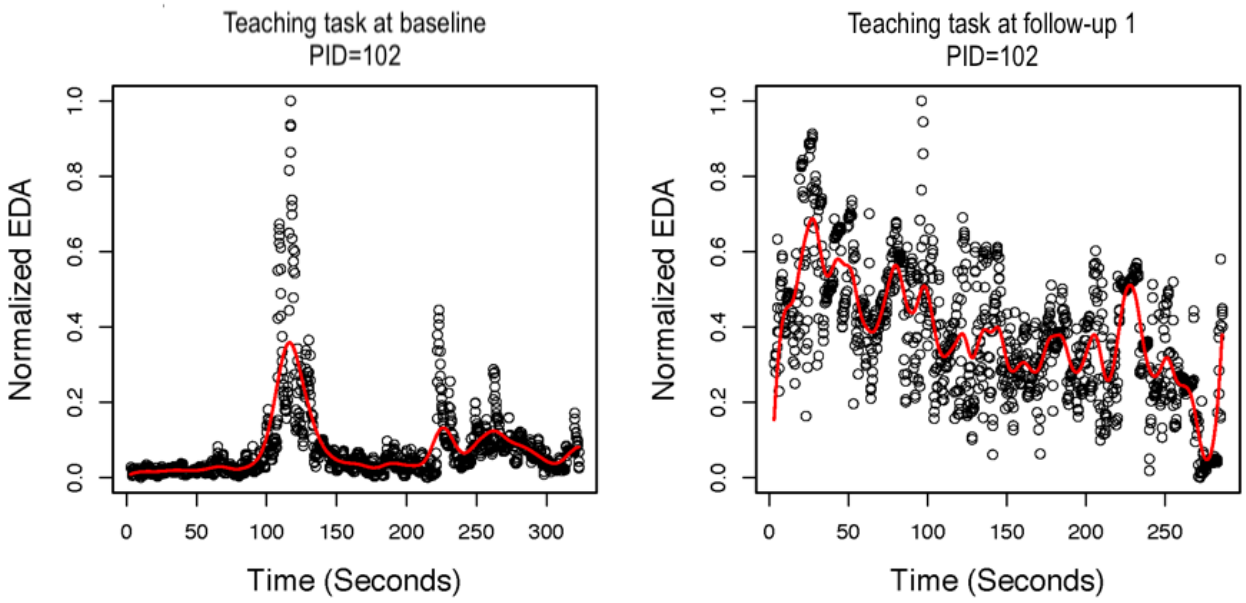


Figure 4. Local linear regression with autoregressive errors for the free-play task at baseline and 3-month follow-up assessments for one mother. EDA: electrodermal activity; PID: patient ID.

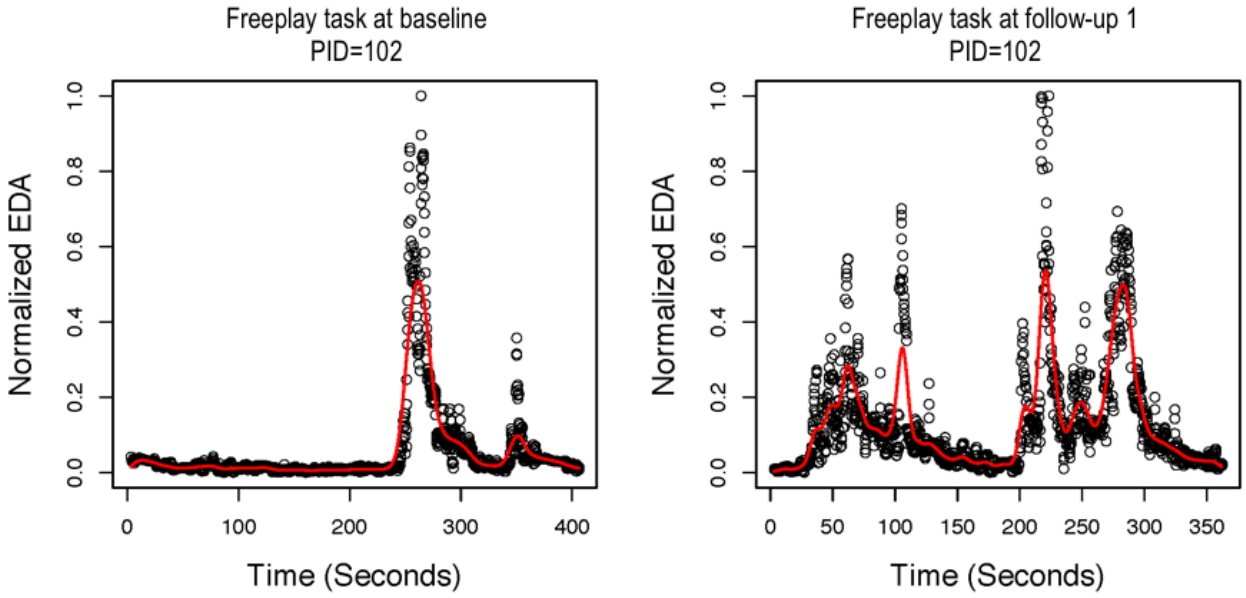


Figure 5. Local linear regression with autoregressive errors for the teaching task at baseline and 3-month follow-up assessments for one mother after applying a low-pass filter. EDA: electrodermal activity; PID: patient ID.

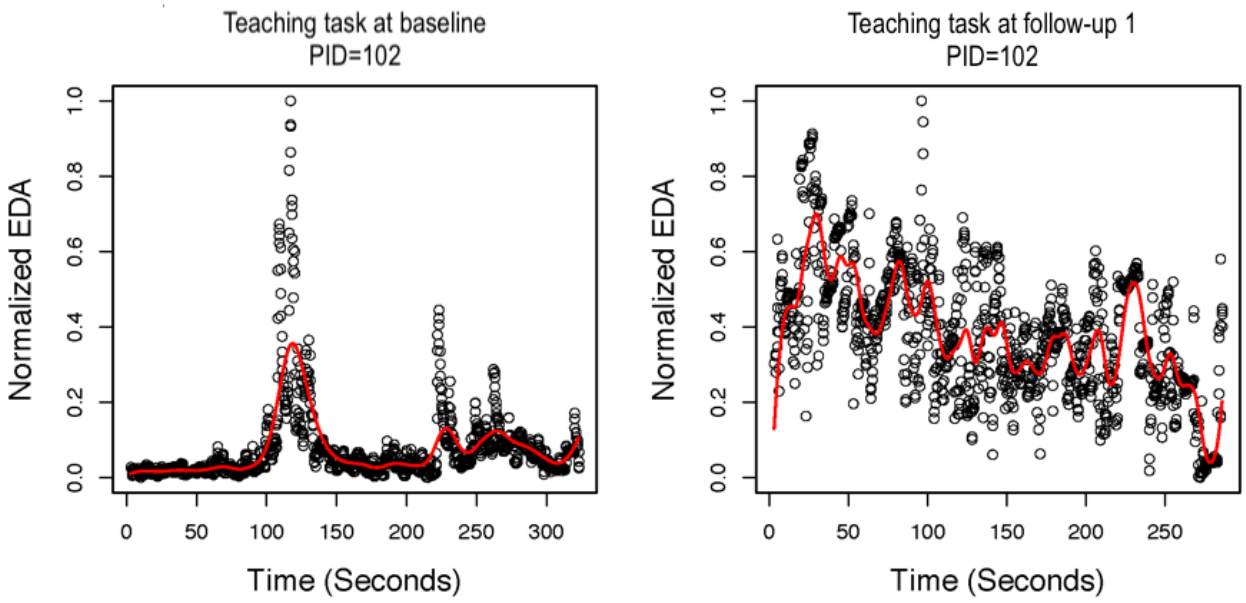
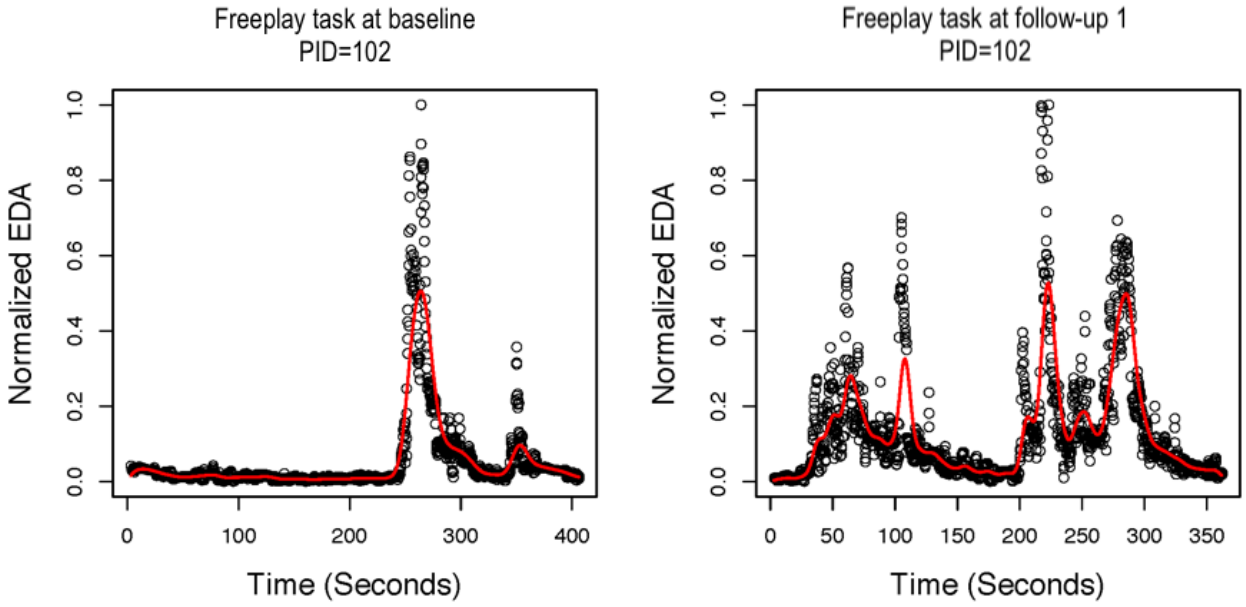


Figure 6. Local linear regression with autoregressive errors for the free-play task at baseline and 3-month follow-up assessments for one mother after applying a low-pass filter. EDA: electrodermal activity; PID: patient ID.



Functional Data Analysis

In implementing the smoothing via B-splines, we used 13 basis functions. We plotted the smoothed versus normalized EDA for each AM, for each task, and for each assessment. The plots for the same AM illustrated in Figures 3 and 4 are shown in Figure 7, left column. The figure illustrates a potential problem with the B-spline approach, which relies on a cubic polynomial. To fit the first peak, the cubic function, by definition, must first increase and then decrease before increasing again. In so doing, the function imposes a false peak just before the actual first peak, and therefore, the function does not fit the actual data well (Figure 7). A different polynomial could be chosen for fitting the B-splines, but it would also impose a particular functional form that may not fit the actual data well. Figure 8 shows the EDA profiles for the teaching task for the same AM (PID 102), as shown in Figure 7. In comparison with the local linear regression approach illustrated in Figure 3, the B-spline approach illustrated in Figure 8 appears to fit at least as well for PID 102. In addition, the B-spline approach fits some individuals better than the local linear regression (not shown). For example, EDA for the free-play task for one intervention group AM is presented in Figure 9, and EDA for the teaching task for the same AM is presented in Figure 10. For this AM, the B-spline approach seemed to work well. We will discuss the 2 smoothing approaches further in the *Discussion* section.

Figure 7. Electrodermal activity (EDA) for the free-play task at baseline and 3-month follow-up assessments for one mother (patient ID=102). The left column shows the estimated B-spline smoothed function overlaying the normalized EDA values. The right column shows the identified peaks and the values for peak features. The red dashed vertical lines demarcate the first peak, the black dashed vertical lines demarcate the highest peak, and the blue dashed vertical lines demarcate all other identified peaks. AUC: area under the curve; EDA: electrodermal activity; PID: patient ID.

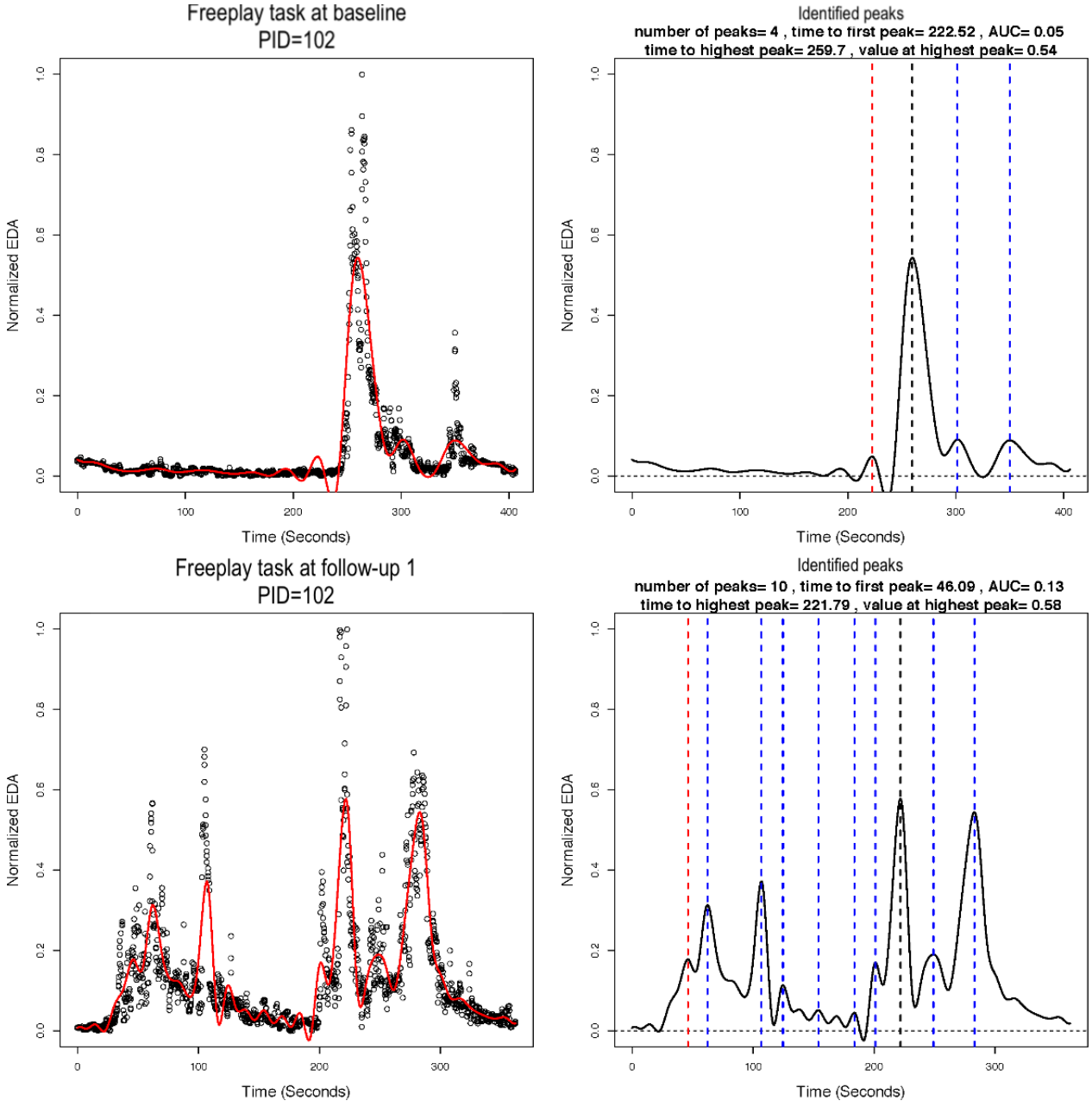


Figure 8. Electrodermal activity (EDA) for the teaching task at baseline and 3-month follow-up assessments for one mother (patient ID=102). The left column shows the estimated B-spline smoothed function overlaying the normalized EDA values. The right column shows the identified peaks and the values for peak features. The red dashed vertical lines demarcate the first peak, the black dashed vertical lines demarcate the highest peak, and the blue dashed vertical lines demarcate all other identified peaks. AUC: area under the curve; EDA: electrodermal activity; PID: patient ID.

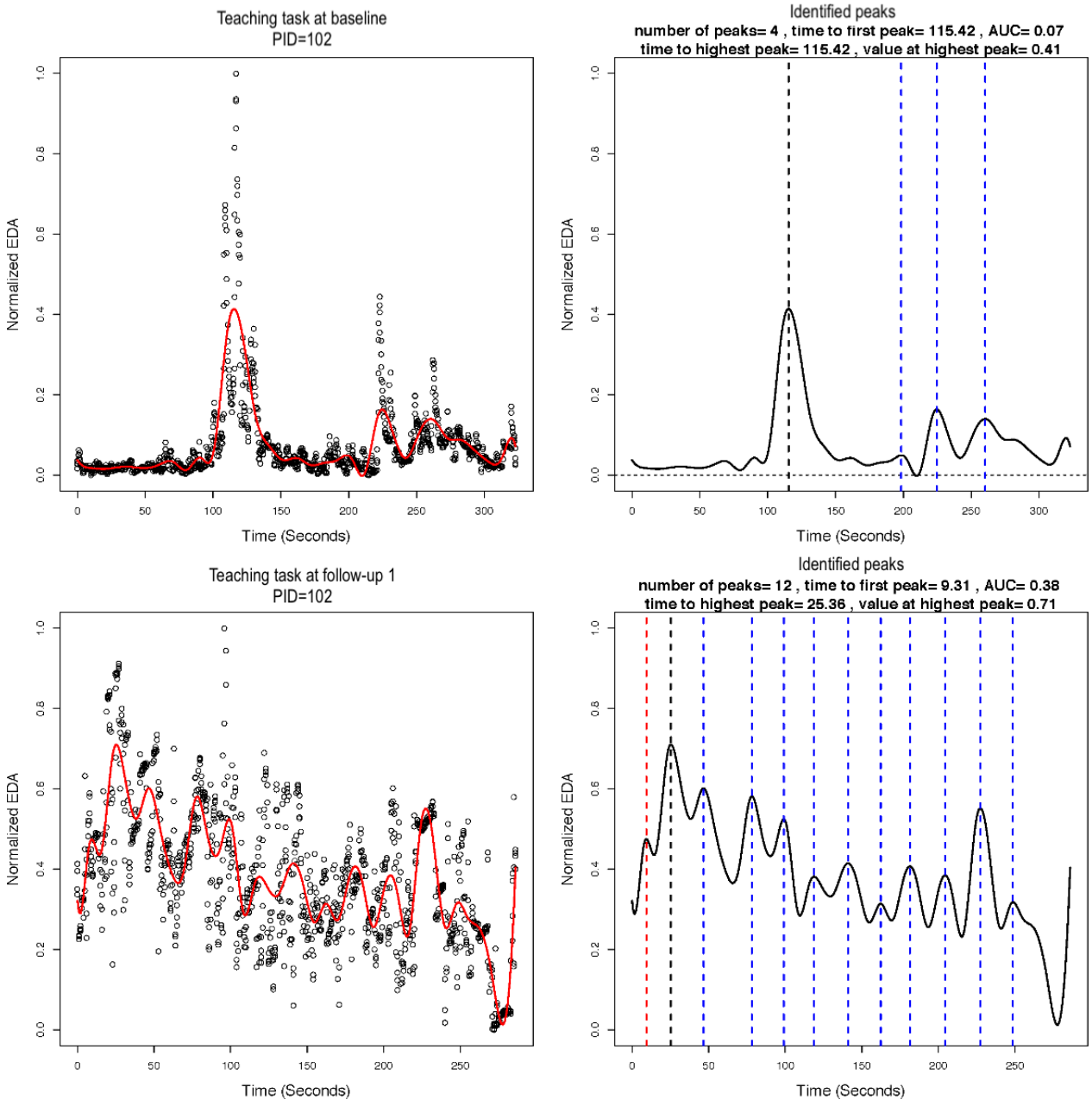


Figure 9. Electrodermal activity (EDA) for the free-play task at baseline, 3-month, and 6-month assessments for one mother (patient ID=202). The left column shows the estimated B-spline smoothed function overlaying the normalized EDA values. The right column shows the identified peaks and the values for peak features. The red dashed vertical lines demarcate the first peak, the black dashed vertical lines demarcate the highest peak, and the blue dashed vertical lines demarcate all other identified peaks. AUC: area under the curve; EDA: electrodermal activity; PID: patient ID.

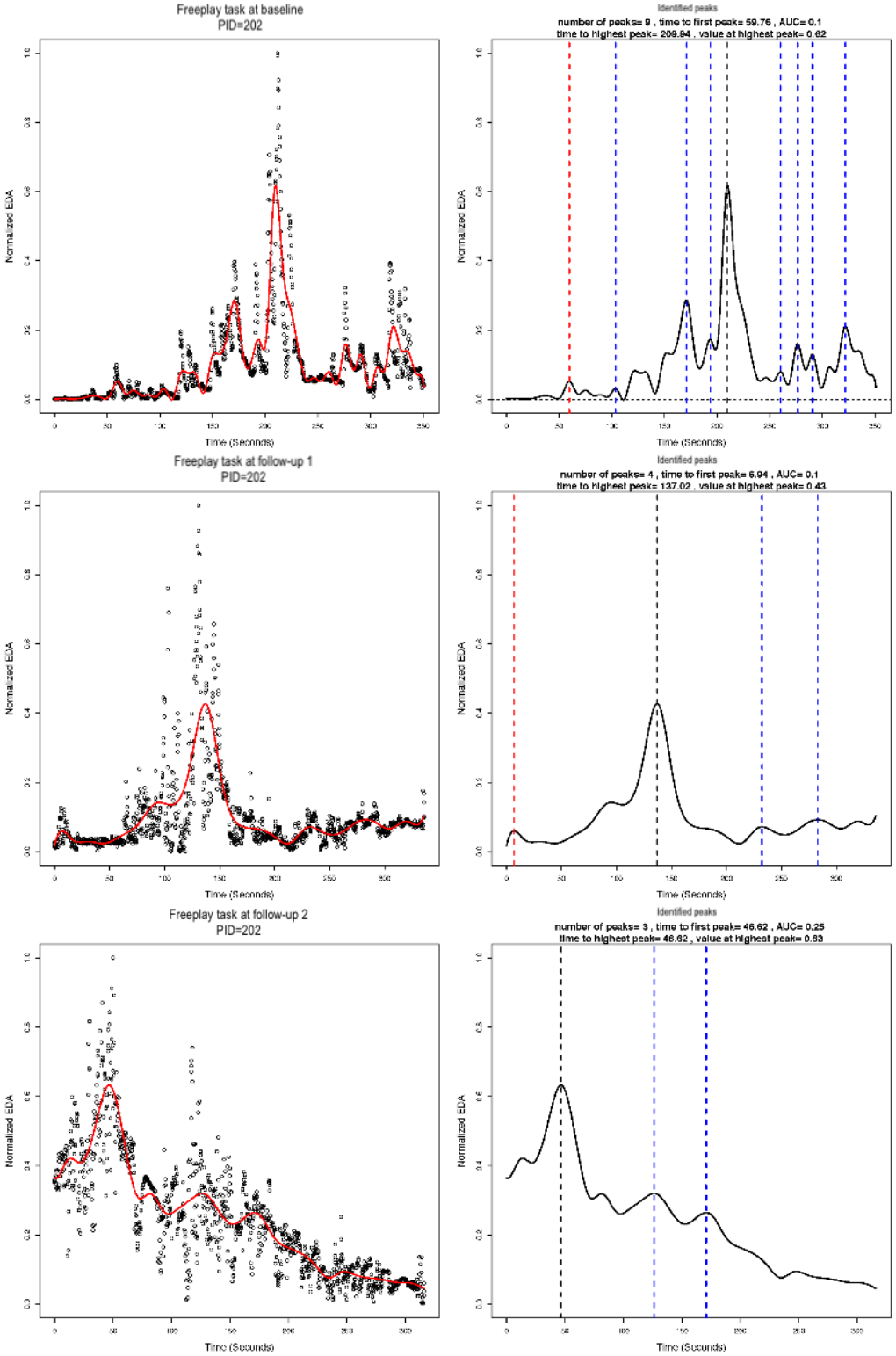
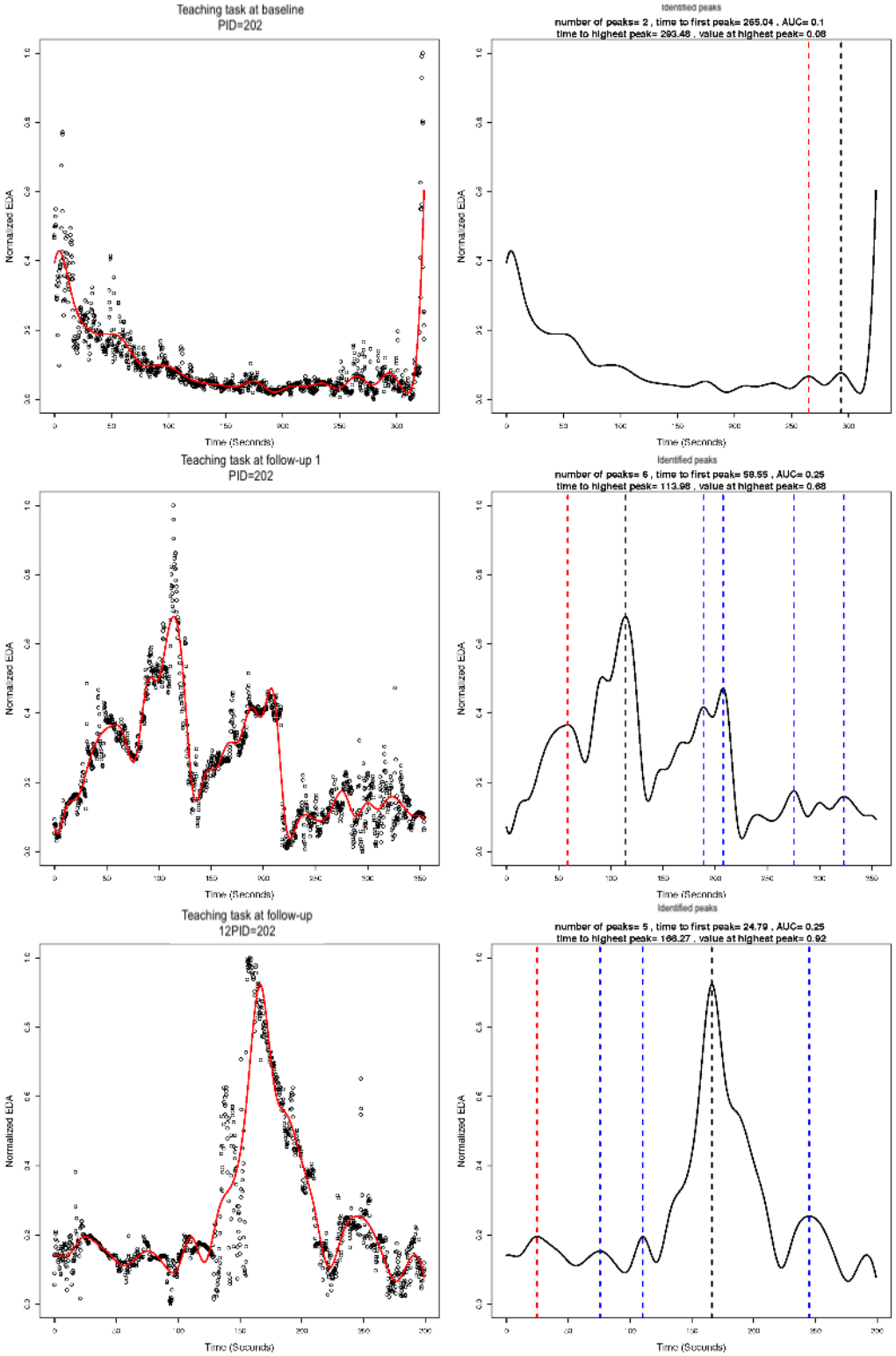


Figure 10. Electrodermal activity (EDA) for the teaching task at baseline, 3-month, and 6-month assessments for one mother (patient ID=202). The left column shows the estimated B-spline smoothed function overlaying the normalized EDA values. The right column shows the identified peaks and the values for peak features. The red dashed vertical lines demarcate the first peak, the black dashed vertical lines demarcate the highest peak, and the blue dashed vertical lines demarcate all other identified peaks. AUC: area under the curve; EDA: electrodermal activity; PID: patient ID.



Computing Features of the Identified Peaks

Next, we identified peaks and computed various features of the peaks from the estimated EDA profiles. Figure 7 (right column) illustrates the identified peaks and presents the values for the peak features. The red dashed vertical lines demarcate the first peak, the black dashed vertical lines demarcate the highest peak, and the blue dashed vertical lines demarcate all other identified peaks. The time to first peak and the highest peak (both in seconds), the normalized EDA value at the highest peak (ie, amplitude), the drop from the highest peak (ie, reactivity), and the number of peaks are all presented in Figure 7, right column for one AM. The peaks were identified, and the features were computed for all AMs. These may be saved in a file for

statistical analysis. Our sample size was ultimately too small for statistical significance testing between the intervention and control groups (ie, between-subjects), but we present descriptive statistics of the peak features in Table 1.

Table 1. Descriptive statistics for peak features by task among those adolescent mothers with identified peaks.

Peak feature/task	Baseline, mean (SD)	Follow-up[a], mean (SD)
Time to the highest peak (seconds)		
Stroop	122.14 (63.42)	103.30 (61.45)
Teaching	144.29 (87.07)	123.55 (94.04)
Free play	201.79 (87.07)	215.10 (93.17)
Value at highest peak (μS)		
Stroop	0.67 (0.21)	0.75 (0.21)
Teaching	0.50 (0.29)	0.67 (0.22)
Free play	0.59 (0.26)	0.48 (0.25)
Drop from the highest peak (seconds)		
Stroop	13.18 (10.76)	17.91 (18.87)
Teaching	30.12 (17.49)	27.90 (43.01)
Free play	29.86 (24.71)	27.76 (18.41)
Number of peaks		
Stroop	6.39 (3.34)	5.13 (3.05)
Teaching	4.14 (3.51)	6.86 (3.61)
Free play	5.59 (2.95)	5.71 (3.33)
Time to the first peak (seconds)		
Stroop	45.66 (36.14)	51.49 (49.47)
Teaching	90.61 (86.08)	27.73 (18.91)
Free play	95.78 (69.93)	81.25 (90.94)

[a]Note that follow-up is at 3 months unless it was missing, in which case, the 6-month follow-up assessment for that individual was used (n=23 adolescent mothers [AMs] for the Stroop task, n=14 AMs for the teaching task at both baseline and follow-up, and n=17 AMs for the free-play task at both baseline and follow-up).

Discussion

Our goal was to present 2 approaches for smoothing and summarizing EDA data by identifying EDA peaks and computing features of those peaks, including the number of peaks, time to first peak, time to highest peak, amplitude of the highest peak, and the time it takes to drop from the highest peak (ie, reactivity). These are certainly not all the features that could be constructed; rather, these are the ones that we thought may be most informative based on the laboratory EDA literature. It is quite possible that other features might be more interesting for ambulatory EDA. Similarly, the approaches we presented for smoothing the data are not the only 2 options [15,16]. Future research needs to examine other possible smoothing methods because the 2 approaches we used did not fit some individuals very well. On the other hand, there may be data for which these methods will fit better than others, as it really depends on the characteristics of a particular dataset. Even within our own data, we found instances where one approach fits the data better than the other and vice versa. Smoothing, identification, and construction of peak features are dependent on data quality.

A larger issue with regard to the use of EDA data for measuring stress, particularly in the ambulatory setting, is the degree to which it is an ecologically valid measure of stress. In the ambulatory setting, it is particularly difficult to distinguish emotional arousal from, for example, hand motions. It may be that the 2 of these are correlated to some degree because if people are emotionally aroused while talking, they may be moving their hands around a lot. Although the device contains an accelerometer, this does not entirely solve this problem. Other devices have been proposed for assessing stress that do not involve EDA. For example, the cStress device triangulates on stress by measuring respiration and heart rate variability via a chest strap [21]. On the other hand, a chest strap device is not as appealing as a wrist-worn device for most participants. Nevertheless, as the technology for the measurement of ambulatory EDA improves, the longitudinal measurement of EDA may have significant clinical implications, as providers and patients can review patterns of arousal that correspond to particular environmental stressors at specific times of day and help patients plan strategies for dealing with these stressors [22]. Similarly, the measurement of EDA may enable better stress management in the workplace, particularly in highly stressful occupations [23].

Limitations

As mentioned previously, a lot of data were lost because the electrodes did not have a good connection with the skin, which resulted in EDA values that were approximately 0. Thus, the first lesson is to make sure that there is a good connection between the dry electrodes on the device and the skin. This means that the electrodes must be touching the skin on the ventral side of the wrist, which means that the device must be fairly tight on the wrist. People often do not wear watches, bands, or bracelets as tight as is necessary for the EDA device. Therefore, we suggest real-time monitoring of incoming EDA data so that if all the values are approximately 0, the participant could be contacted to inform them to tighten the device. Note that simply detecting that data are not being recorded in real time is not enough; there must be real-time detection of data that is being recorded but is essentially 0. Some of our EDA data were not usable because the band lost contact with the skin during some tasks, particularly early in the interview. By *loose connection*, we mean that the electrodes on the band did not have a good connection with the skin. We do not mean that the band was not connected in terms of recording the EDA. If the band was not recording, then EDA was missing at that particular measurement. This did happen sometimes but not enough to present a problem, particularly because we were able to smooth over it using the proposed smoothing approaches. It should be noted that the loose connection problem is not limited to the particular device, the iCalm, that we used. Empatica E4 also has this problem, as we have used it in a pilot study [24] and lost data because of a loose connection. Other studies using wrist-based ambulatory measures of EDA with the Affectiva Q Sensor have reported greater success [25,26] in both controlled and uncontrolled settings. However, other studies using the Affectiva Q sensor have reported data lost to electrode connection problems (eg, 31% [14] and 17% [22]). A total of 22.5% of data were lost because of electrode connection problems for the free-play task and 24% for the teaching task. Owing to the loose connection problems during data collection, our sample size was substantially reduced, which limits our ability to draw conclusions about the between-subject effects of the intervention on the features constructed from the EDA data. Nevertheless, we think it is important to draw attention to data collection challenges as the measurement of ambulatory EDA has become more popular. In addition, researchers currently collecting EDA data have limited guidance on how to preprocess the data before statistical analysis.

A second challenge in EDA data collection is motion artifacts. These are identifiable as sharp spikes or drops in EDA data. One concern with the free-play task is that there were many more potential motion artifacts than there were in the teaching task because the AMs may have been moving around more during the free-play task. Another issue that may sometimes occur is that the EDA during the current task could be affected by the EDA response to the previous task. In our study, we tried to minimize this by having recovery periods between tasks, but, of course, there is no guarantee that the recovery periods were long enough. An individual's EDA does not always return to the previous baseline, so it can be difficult to assess if the recovery period is long enough.

Conclusions

In summary, in light of the increase in wearables for the continuous, ambulatory collection of EDA and associated challenges collecting data outside of the laboratory, analytic methods for computing various features of EDA are needed to help meet these challenges. We described 2 methods for smoothing and summarizing EDA data using FDA and local polynomial regression with AR errors. As the technology of wearable devices continues to advance, future research opportunities abound for improvement in device design that may improve the electrodes and their connection with the skin and facilitate in the development of improved methods for smoothing the data, identifying peaks, and constructing features.

Acknowledgments

This study was supported by the National Institute on Drug Abuse (NIDA) Grant R34 DA032960 (principal investigator [PI]: NL) and the National Institute on Environmental Health Sciences (NIEHS) K01 ES025437 (PI: DC). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA, the NIEHS, or the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
R code for conducting the functional data analysis with B-splines and the local polynomial regression with autoregressive errors.
[DOCX File , 27 KB - biomedeng_v5i1e17106_app1.docx]

References

1. Boucsein W. Electrodermal activity. 2nd edition. New York: Springer; 2012.
2. Ramsay J, Silverman B. Functional Data Analysis. Hoboken, New Jersey: John Wiley & Sons, Inc; 2005.
3. Li R, Li Y. Local linear regression for data with AR errors. Acta Math Appl Sin 2009 Jul 1;25(3):427-444 [FREE Full text] [doi: 10.1007/s10255-008-8813-3] [Medline: 20161374]

4. Boucsein W, Fowles DC, Grimnes S, Ben-Shakhar G, Roth WT, Dawson ME, Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures. Publication recommendations for electrodermal measurements. Psychophysiology 2012 Aug;49(8):1017-1034. [doi: 10.1111/j.1469-8986.2012.01384.x] [Medline: 22680988]
5. Doberenz S, Roth WT, Wollburg E, Maslowski NI, Kim S. Methodological considerations in ambulatory skin conductance monitoring. Int J Psychophysiol 2011 May;80(2):87-95 [FREE Full text] [doi: 10.1016/j.ijpsycho.2011.02.002] [Medline: 21320551]
6. Poh MZ, Swenson NC, Picard RW. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. IEEE Trans Biomed Eng 2010 May;57(5):1243-1252. [doi: 10.1109/tbme.2009.2038487]
7. Zangróniz R, Martínez-Rodrigo A, Pastor JM, López MT, Fernández-Caballero A. Electrodermal activity sensor for classification of calm/distress condition. Sensors (Basel) 2017 Oct 12;17(10):- [FREE Full text] [doi: 10.3390/s17102324] [Medline: 29023403]
8. Fletcher RR, Dobson K, Goodwin MS, Eydgahi H, Wilder-Smith O, Fernholz D, et al. iCalm: wearable sensor and network architecture for wirelessly communicating and logging autonomic activity. IEEE Trans Inf Technol Biomed 2010 Mar;14(2):215-223. [doi: 10.1109/TITB.2009.2038692] [Medline: 20064760]
9. Benedek M, Kaernbach C. Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology 2010 Jul 1;47(4):647-658 [FREE Full text] [doi: 10.1111/j.1469-8986.2009.00972.x] [Medline: 20230512]
10. Benedek M, Kaernbach C. A continuous measure of phasic electrodermal activity. J Neurosci Methods 2010 Jun 30;190(1):80-91 [FREE Full text] [doi: 10.1016/j.jneumeth.2010.04.028] [Medline: 20451556]
11. Lim CL, Rennie C, Barry RJ, Bahramali H, Lazzaro I, Manor B, et al. Decomposing skin conductance into tonic and phasic components. Int J Psychophysiol 1997 Feb;25(2):97-109. [doi: 10.1016/s0167-8760(96)00713-1] [Medline: 9101335]
12. Leonard NR, Casarjian B, Fletcher RR, Praia C, Sherpa D, Kelemen A, et al. Theoretically-based emotion regulation strategies using a mobile app and wearable sensor among homeless adolescent mothers: acceptability and feasibility study. JMIR Pediatr Parent 2018;1(1):e1 [FREE Full text] [doi: 10.2196/pediatrics.9037] [Medline: 30637376]
13. Rajan S, Leonard N, Fletcher R, Casarjian B, Casarjian R, Cisse C, et al. Ambulatory autonomic activity monitoring among at-risk adolescent mothers. J Mob Technol Med 2012;1(3):25-31 [FREE Full text] [doi: 10.7309/jmtm.19] [Medline: 23626657]
14. Hernandez J, Riobo I, Rozga A, Abowd GD, Picard RW. Using Electrodermal Activity to Recognize Ease of Engagement in Children During Social Interactions. In: Proceedings of the International Conference on Ubiquitous Computing. 2014 Presented at: UbiComp '14; September, 2014; Seattle Washington URL: https://dl.acm.org/doi/proceedings/10.1145/2632048
15. Taylor S, Jaques N, Chen W, Fedor S, Sano A, Picard R. Automatic Identification of Artifacts in Electrodermal Activity Data. In: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2015 Presented at: EMBC'15; August 25-29, 2015; Milan, Italy. [doi: 10.1109/embc.2015.7318762]
16. Kleckner IR, Jones RM, Wilder-Smith O, Wormwood JB, Akcakaya M, Quigley KS, et al. Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data. IEEE Trans Biomed Eng 2018 Jul;65(7):1460-1467 [FREE Full text] [doi: 10.1109/TBME.2017.2758643] [Medline: 28976309]
17. Ruppert D. Selecting the number of knots for penalized splines. J Comput Graph Stat 2012 Jan;11(4):735-757. [doi: 10.1198/106186002853]
18. Tan X, Shiyko MP, Li R, Li Y, Dierker L. A time-varying effect model for intensive longitudinal data. Psychol Methods 2012 Mar;17(1):61-77. [doi: 10.1037/a0025814] [Medline: 22103434]
19. Fan J, Li R. An overview on nonparametric and semiparametric techniques for longitudinal data. Front Stat 2006:277-303. [doi: 10.1142/9781860948886_0013]
20. Schwarz G. Estimating the dimension of a model. Ann Statist 1978 Mar;6(2):461-464. [doi: 10.1214/aos/1176344136]
21. Hovsepian K, al'Absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. cStress: towards a gold standard for continuous stress assessment in the mobile environment. Proc ACM Int Conf Ubiquitous Comput 2015 Sep;2015:493-504 [FREE Full text] [doi: 10.1145/2750858.2807526] [Medline: 26543926]
22. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. J Med Internet Res 2018 Jun 8;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]
23. Furberg RD, Taniguchi T, Aagaard B, Ortiz AM, Hegarty-Craver M, Gilchrist KH, et al. Biometrics and policing: a protocol for multichannel sensor data collection and exploratory analysis of contextualized psychophysiological response during law enforcement operations. JMIR Res Protoc 2017 Mar 17;6(3):e44 [FREE Full text] [doi: 10.2196/resprot.7499] [Medline: 28314707]
24. Leonard NR, Silverman M, Sherpa DP, Naegle MA, Kim H, Coffman DL, et al. Mobile health technology using a wearable sensorband for female college students with problem drinking: an acceptability and feasibility study. JMIR Mhealth Uhealth 2017 Jul 7;5(7):e90 [FREE Full text] [doi: 10.2196/mhealth.7399] [Medline: 28687533]
25. Carreiro S, Fang H, Zhang J, Wittbold K, Weng S, Mullins R, et al. IMStrong: deployment of a biosensor system to detect cocaine use. J Med Syst 2015 Dec;39(12):186 [FREE Full text] [doi: 10.1007/s10916-015-0337-9] [Medline: 26490144]

26. Quick JA, Bukoski AD, Doty J, Bennett BJ, Crane M, Barnes SL. Objective measurement of clinical competency in surgical education using electrodermal activity. J Surg Educ 2017;74(4):674-680. [doi: 10.1016/j.jsurg.2017.01.007] [Medline: 28373078]

Abbreviations

AMs: adolescent mothers
AR: autoregressive
AUC: area under the curve
EDA: electrodermal activity
FDA: functional data analysis
FIR: finite impulse response
GCV: generalized cross-validation
NIDA: National Institute on Drug Abuse
NIEHS: National Institute on Environmental Health Sciences
PI: principal investigator
PID: patient ID
PSP: Power Source Parenting
SC: skin conductance
SCL: skin conductance level
SCR: skin conductance response
SVM: support vector machine
TLP: transitional living program

Edited by G Eysenbach; submitted 19.11.19; peer-reviewed by T Kichline, T Aslanidis; comments to author 12.12.19; revised version received 29.01.20; accepted 13.05.20; published 12.06.20.

Please cite as:
Coffman DL, Cai X, Li R, Leonard NR
Challenges and Opportunities in Collecting and Modeling Ambulatory Electrodermal Activity Data
JMIR Biomed Eng 2020;5(1):e17106
URL: http://biomedeng.jmir.org/2020/1/e17106/
doi:10.2196/17106
PMID:34888487

Viewpoint

Video Cloud Services for Hospitals: Designing an End-to-End Cloud Service Platform for Medical Video Storage and Secure Access

Piotr Pawałowski[1*], MA; Cezary Mazurek[1*], DPhil; Mikołaj Leszczuk[2*], DPhil; Jean-Marie Moureaux[3*], DPhil; Amine Chaabouni[4*], DPhil

[1]Poznan Supercomputing and Networking Center, Poznań, Poland
[2]AGH University of Science and Technology, Kraków, Poland
[3]Centre de Recherche en Automatique de Nancy, University of Lorraine, Nancy, France
[4]DeepRiver, Nancy, France
[*]all authors contributed equally

Corresponding Author:
Piotr Pawałowski, MA
Poznan Supercomputing and Networking Center
ul Z Noskowskiego 12/14
Poznań, 61-704
Poland
Phone: 48 693919937
Email: astagor@man.poznan.pl

Abstract

The amount of medical video data that has to be securely stored has been growing exponentially. This rapid expansion is mainly caused by the introduction of higher video resolution such as 4K and 8K to medical devices and the growing usage of telemedicine services, along with a general trend toward increasing transparency with respect to medical treatment, resulting in more and more medical procedures being recorded. Such video data, as medical data, must be maintained for many years, resulting in datasets at the exabytes scale that each hospital must be able to store in the future. Currently, hospitals do not have the required information and communications technology infrastructure to handle such large amounts of data in the long run. In this paper, we discuss the challenges and possible solutions to this problem. We propose a generic architecture for a holistic, end-to-end recording and storage platform for hospitals, define crucial components, and identify existing and future solutions to address all parts of the system. This paper focuses mostly on the recording part of the system by introducing the major challenges in the area of bioinformatics, with particular focus on three major areas: video encoding, video quality, and video metadata.

(JMIR Biomed Eng 2020;5(1):e18139) doi:10.2196/18139

KEYWORDS

medical video; telemedicine; medical cloud platforms

Background

In recent years, technologies and services in the areas of electronic health (eHealth) and telemedicine have been evolving more and more rapidly. In many European countries, the law is being changed so that remote medical treatments conducted using information and communications technology (ICT) systems are now recognized as valid medical procedures. The physical presence of the doctor and the patient at the same location is often no longer necessary, and private as well as state-owned insurance companies now reimburse doctors for performing remote consultations with the patient. A remote interaction between the doctor and the patient can be conducted via phone or internet chats, but as video plays a significant role in current communication, solutions based on the real-time transmission of audio-video are considered to provide better service with higher trustworthiness. In many such scenarios, there is a need to record remote treatment sessions for further reference, follow-up analysis, or for legal reasons.

Furthermore, in many European countries, there is a growing need for recording medical procedures conducted in hospitals or health care centers. There is a general trend toward increasing transparency when it comes to medical treatment, and thus to improve the quality of services provided to the patient. Medical

errors do happen, but sometimes a dissatisfied patient decides to prosecute even when no mistake has been made. Therefore, hospitals need to have a way to settle such disputes based on proof. Some countries are preparing laws to make it obligatory for hospitals to record all treatment procedures. Moreover, incoming law changes will qualify such recordings as medical data, forcing hospitals to store these data for decades. Insurance companies also offer much lower liability insurance fees to hospitals recording their procedures, because in the case of a lawsuit, there is clear evidence to determine whether or not the hospital is at fault.

Taking this context into account, hospitals, health care centers, and doctors in Europe will soon need to implement recording of treatment and storage of these recordings for treatment conducted both locally and using telemedicine systems. However, most hospitals and health care centers are not ready to introduce such services as they do not have the necessary recording devices, are overwhelmed with the infrastructural needs of storing thousands of hours of recordings per year, and cannot guarantee the necessary security of the repositories.

To resolve these issues, an end-to-end solution for the recording, secure storage, and access to records of medical procedures conducted locally in hospitals and health care centers is needed along with using telemedicine systems. This requires new approaches to video coding, image quality, transmission, and security.

In this paper, we propose a generic platform architecture for medical video recording and storage. We discuss current and upcoming challenges, and identify the possible technologies to tackle these challenges. We also outline the existing solution that can be incorporated into a holistic platform. This paper focuses mostly on the recording part of the system, as it introduces the major challenges in the area of bioinformatics from three major areas: video encoding, video quality, and video metadata.

The remainder of this paper is organized as follows. The next section describes the current situation and upcoming challenges. This is followed by a proposal of the generic architecture and possible technologies to be used to resolve these challenges. The following section outlines how existing solutions based on such a platform could be built. Finally, we provide an overview and future prospects.

Current Situation and Upcoming Challenges

Telemedicine services, which allow for remote treatment and remote consultations using videoconferencing or other ICT systems, are increasingly being recognized as valid treatment solutions by national health care legislation. For example, as of 2018 in Germany [1], France, and Poland, the physical presence of the doctor and the patient at the same location is no longer required. This allows for new concepts of health care services to be introduced, and solutions providing such services are already available on the market (eg, hopimedical.com, myvirtualdoctor.com, or vividoctor.com). As in the United States, such services have been in place for some time now, and the European players who want to compete for the market share need to provide solutions tuned to European legislation, taking into consideration the diversity of the health care system landscape in European countries. The telemedicine market worldwide is increasing, which expanded from US $9.8 billion in 2010 to US $11.6 billion in 2011, and continued to grow to US $27.3 billion in 2016, representing an annual growth rate of 18.6% [2].

Telemedicine services can be divided into three basic types: (1) enabling patient-doctor contact, (2) supporting remote collaboration between medical professionals, and (3) enhancing medical education. These three types, although they involve different users and technical solutions, are substantially based on the transmission of video and thus require a secure solution to provide video recording and storage.

University clinics, which often use videoconferencing systems for medical education, also face the problem of how to store and publish recordings of educational sessions. A typical setup with H.323 videoconferencing equipment does not allow for recording sessions without the use of an additional dedicated recording device or cloud service. Other devices introduce extra costs and require management by the hospital. Using an external cloud service may entail not following the legal requirements, as most providers of videoconferencing solutions that are currently used are developed by US-based companies.

A similar problem applies to recording eHealth video connections between the patient and the doctor. Although the amount of data is much lower, web-based or mobile app–based solutions do not fulfil the legal requirements when it comes to the recording and storing of medical videos. Records of treatment procedures conducted remotely contain private patient data and need to be processed and stored accordingly. In many European countries, medical data cannot leave the given country; therefore, the medical videos must be stored securely by a cloud service that is located within the country.

In addition to telemedicine systems, medical videos can be produced by modern operating rooms (ORs), which contain several devices that are also video sources, including microscopes, endoscopes, surgical robots, and macroscopic cameras. Owing to the increasing complexity of the surgical working environment, technical solutions must be found to help relieve the surgeon [3]. Only a limited number of hospitals are equipped with integrated ORs that provide a common video management platform for all of these sources.

In most cases, medical devices record video on local hard drives, DVDs, or USB flash drives, if they provide that possibility at all. This is a highly unmanageable situation, and is a great hassle for the information technology department to manage and store the recordings. Even if there is an integrated video system available in the hospital, the infrastructure needed to save the videos, especially for many years, is not affordable for many hospitals, and storing the videos in hospital is not cost-effective. With the development of video technologies, where full high-definition systems are now standard and 4K or even 8K resolution systems are available or upcoming, the needed storage space for medical video has proliferated. For example, radiological imaging such as computed tomography scans,

magnetic resonance images, or positron emission tomography scan images can reach several hundreds of megabytes per exam, while stereoscopic video related to high-definition surgery is based on two streams of 1.5 GB per second each. In the United States, the storage of mammograms required 2.5 PB in 2009, and 30% of all the images stored in the world in 2010 were medical images [4].

Moreover, in many countries, the retention period of medical records is very long. In Poland and France, medical images associated with a patient must be kept for 20 years after their last visit under the current law. As an example, the Department of Radiology and Imaging of the University Hospital in Nancy, France produced 55 TB of image data in 2015. Regarding videos, a full high-definition endoscopic camera (used routinely in endoscopic surgery) can provide 2.6 TB of data during a 2-hour operation. These figures highlight the challenge to be faced in terms of archiving capacity.

Consequently, medical images and video should be encoded (ie, compressed), as long as the encoding does not affect the therapeutic quality of the data for regular use by health professionals. Regardless of whether the data need to be transmitted or stored, it is of high importance to ensure that the compression step introduces no significant loss (degradation, or any other processing such as watermarking) at the encoder stage. In other words, it is crucial to maintain an acceptable visual quality of the video stream for health care professionals.

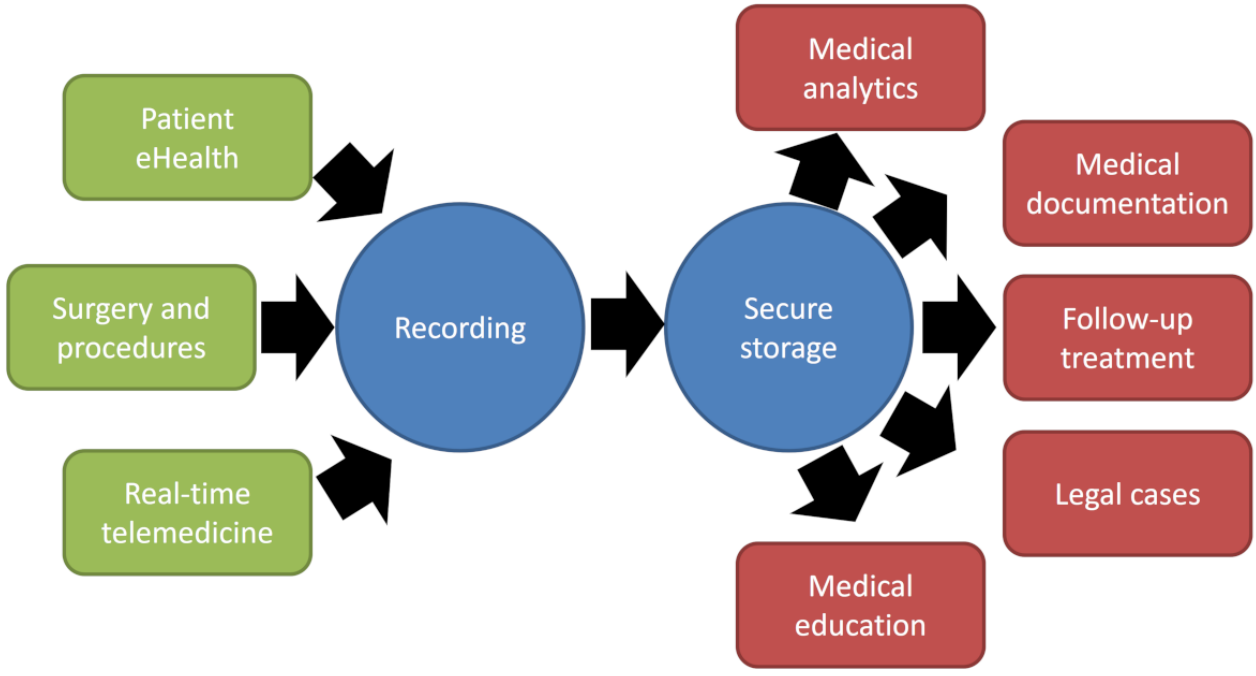
There are many solutions on the market providing recording functionality in the hospital environment. Companies such as Storz, Olympus, or Stryker have medical recorders in their offer. These systems integrate with the proprietary storage solutions provided by their producers, which have to be deployed within the information technology infrastructure of the hospital. However, this is not a scalable solution when all video recordings have to be kept as medical data over many years. Standard videoconferencing solutions, available from Cisco or Polycom, do not have a built-in recorder but utilize separate server-based recorders to be installed in the information technology infrastructure of the hospital. In addition, this is not a scalable solution and these products are focused on video conferencing in general. Therefore, they do not fulfill the requirements of medical video recording and storage when it comes to security and access rights.

Possible Solutions

Currently, European hospitals are not yet obliged to record all medical procedures; however, many do so for in-house usage or to increase surgeons' efficiency [5]. If it becomes obligatory by law to record all surgical procedures, the amount of data produced will be enormous. Although it may constitute a burden, recording of medical procedures also brings new possibilities of utilizing the stored videos.

Our group, composed of research centers and companies from Poland, France, and Germany, has designed a generic architecture (Figure 1) for a holistic recording and storage platform for hospitals, which is outlined below. We have also identified components that can be used as parts of such a platform.

Figure 1. Generic architecture of the holistic recording and storage platform for hospitals. eHealth: electronic health.



Scenarios

The high-level scenarios that need medical video recording (marked with green in Figure 1) are: (1) patient eHealth, involving telemedical situations in which the patient is connected with the doctor using the web or mobile solutions providing videoconferencing services; (2) surgery and procedures, which include scenarios in which the hospital or health care center needs to record medical procedures or surgeries; and (3) real-time telemedicine, which includes

scenarios focusing on hospitals and health care centers using telemedical solutions to provide remote consultations, surgery supervision, or medical education.

Technical Constituents

The two major technical constituents of the system (marked with blue in Figure 1) are recording and secure storage. Recording includes components of the system responsible for video acquisition, encoding, and uploading to the secure storage. These components should provide novel video encoding mechanisms, maximize video quality of experience for medical applications, be ready for 4K and 8K medical devices, allow to record stereoscopic (three-dimensional) video, and provide an innovative approach to a metadata description of the videos. The second main conceptual component of our system is the safe storage platform providing storage, transcoding, and access to the medical video recordings. The room should be centralized, in terms of the provider, for each country, but it should be possible to be distributed over various data centers. National regulations should be implemented for the deployment in each state, providing the needed security standards. The envisioned solution should provide measures for the secure upload of recorded videos from the devices and recording applications listed above. The platform will also offer various means of access to the videos depending on user privileges and use case scenarios.

Use Case Scenarios

The different use case scenarios of the recordings are marked with red in Figure 1. The access to the records should be granted through a web portal or an application programming interface (API).

Medical Analytics

The platform should provide access to recordings that have been marked as public and anonymized to enable medical analytics and research using the stored video content.

Medical Documentation

Videos are also an essential part of the medical literature. The recordings provided through the platform should be linked to medical documentation files.

Follow-Up Treatment

Recordings stored on the platform should be accessible by doctors to review the previous treatment or procedures to provide better follow-up treatment. The videos should either be available only to the doctor/hospital that created the recording or made available to other health care providers, according to varying laws in European countries.

Legal Cases

The recorded videos could also be used by courts or insurance companies in legal matters. As mentioned in the Introduction, with a general trend to increase transparency when it comes to medical treatment, recoding medical procedures reduces the liability of hospitals. This is especially important as telemedical services, allowing remote treatment using videoconferencing, are being recognized as valid treatment solutions by national health care legislation.

Medical Education

The insufficient number of adequately educated medical personnel is one of the main obstacles to providing universal health care at the highest level. Medical school is especially time- and money-consuming in areas such as surgery, in which students and young doctors have to observe several operations during their education. However, because of spatial limitations in ORs, only a small number of students can be present during a given surgery, especially if they need to see the essential details from up close. This access scenario is an integral part of modern medicine, as audio-video content highly increases the effectiveness of medical education, which in turn benefits the whole society.

Service

Another critical aspect of medical procedures recording that must be considered is that the service delivered to the hospitals has to be an end-to-end service. This means that, from the hospital perspective, the service should not require additional infrastructure except for medical recorders. These recorders have to be able to acquire audio and video from the OR equipment, record and encode it, and then automatically transfer it to the secure cloud infrastructure outside the hospital. It is desired that a standard open API be designed to allow companies producing medical recorders and other video-based equipment to integrate with the secure cloud, avoiding vendor lock-in. Furthermore, the legal framework for each country may differ, so that the process of deployment of the service must be conducted with the participation of legal and health care system experts. There is a long list of norms that has to be taken into account, including ISO 13485 (Medical Devices) as a quality management system [6], ISO 14971:2007 (Part 1: Application of Risk Management to Medical Devices) [7], and IEC 60601-1-11:2015 (Medical Electrical Equipment, Part 1-11: General Requirements for Basic Safety and Essential Performance; Collateral Standard: Requirements for Medical Electrical Equipment and Medical Electrical Systems Used in the Home Health Care Environment) [8].

In the following subsections, we will focus on video encoding, video quality, and video metadata, as these are the areas introducing the greatest challenges with respect to the topic of bioinformatics.

Video Encoding

It is crucial to choose an appropriate set of codes that will minimize the storage space requirements and maintain the required quality for medical purposes. We have created a model to calculate the amount of storage needed for surgery recording while fulfilling the requirements of being treated as medical data. In the model, we assumed that we record one full high-definition (1920×1080) video image from an OR, with the H.264/AVC video codec at 6 Mbps. This codec was chosen for the model as it is currently the most widespread video codec and provides adequate quality for medical purposes. This has been confirmed by carrying out subjective tests, as they are an excellent way to assess video/image quality for a group of users. These tests allowed medical specialists such as surgeons to give their opinions about specific encoded sequences [9]. The

example calculations were performed for Poland. In 2017, there were 951 general hospitals in Poland. On average, there are 8 ORs in a hospital, with 3.2 surgeries being performed per day in each of them [10]. Based on this, we can calculate that one hospital recording all surgical procedures produces 189.8 GB of video data per day and 67.7 TB per year. Therefore, for the whole of Poland, this adds up to 62.9 PB per year. If these recordings are treated as medical data, they have to be stored for 20 years since the last visit of the patient. As an example, if surgery of a 20-year-old patient is recorded in 2019 and the patient lives until 80 years old, regularly going back to see the doctors, this recorded surgery will have to be stored until 2099. For calculations in our model, we used the average life expectancy for Poland, which is 75 years [11], and the average age of a patient of 56.87 years [12]. This means that, on average, a recorded surgery has to be stored for 38.13 years, and before any recorded video can be deleted, the storage needs of Polish hospitals will add up to 2325.3 PB. Such volumes will be a severe challenge for the information technology infrastructures of the hospitals. The above calculations assume recording only one full high-definition video. However, there are already medical devices (eg, endoscopes) available on the market that provide 4K or even 8K resolution [13]. The usage of these devices will increase the amount of data produced by a factor of 4 or 16. To compensate for this, new ways of video encoding are needed that will provide higher compression, while maintaining the quality at a medical grade.

Current video compression systems can be roughly divided into two categories: strong and lightweight compression. Strong compression is mainly used for the final distribution of the video content, which provides compression ratios from 10× to more than 100×. High compression is obtained at the expense of losing the information in the video (lossy compression) and usually requires substantial computational effort on the encoder side. However, in the newest video codecs, the compression ratio and resulting visual quality can be controlled by the encoder for different application scenarios. For telemedicine, we can achieve a video quality that is visually not distinguishable from the source for humans. Video coding standards such as H.264 (ie, advanced video coding [AVC]) and H.265 (ie, high-efficiency video coding [HEVC]) are part of the strong compression category.

Lightweight compression is used during the production process to reduce the data size for transmission or storage while maintaining the quality as close as possible to the original (ie, lossless or visually lossless compression). Compression ratios between 2× and 6× are common, and implementations of encoders and decoders require much less computational resources compared to firm compression. Codecs such as JPEG-2000, JPEG-XS, and VC-2 are part of this group.

The H.265/HEVC video coding standard [14] ratified in 2013 represents the state of the art in strong video compression. Compared to H.264/AVC, H.265/HEVC is capable of reducing the bit rate by 50% while maintaining the same quality. Compared to uncompressed video, HEVC can achieve compression factors between 250× and 500×, depending on the target bit rate and video resolution, more specifically 249× for full high-definition (1920×1080), 30 frames/s at 3 Mbits/s; 478× for 4K (3840×2160), 60 frames/s at 25 Mbits/s; and 398× for 8K (7680×4320), 60 frames/s at 120 Mbits/s. Unlike H.264 Main Profile (MP), HEVC MP supports bit depths up to 10 bit, resulting in better color fidelity and a noticeable reduction of banding artefacts in uniform areas with a continuous gradation of the color tone and luminosity.

In 2014, version 2 of the HEVC standard (HEVCv2) was approved, which specifies, among others, the so-called range extensions (RExt) addressed toward video production and contribution applications that require high-quality color formats such as 4:2:2, 4:4:4, and RGB, up to 12 bits per pixel. HEVC RExt is suitable for medical applications as the extended chroma formats and high bit depths allow for better preservation of the original content.

A new generation of video codecs is being developed to cope with the demands of emerging applications such as UHD TV in 4K and 8K resolutions, 360° video, and new quality formats such as high dynamic range (HDR), high frame rate (HFR), and wide color gamut (WCG).

These next-generation video codecs include AV1 and H.266. The AV1 format has recently been ratified by the Alliance of Open Media (AOM) and is designed for web apps such as video on demand (VoD) [15]. The AV1 codec is intended to be open and royalty-free. Studies conducted in 2017 concluded that AV1 produces an average bit rate reduction between 17% and 22% compared to H.265/HEVC [16]. In terms of encoding speed, the AOM's AV1 reference implementation seems to be too heavy computationally and, as far as we know, there is no efficient AV1 encoder implementation available at present. Therefore, the potential compression gain versus encoding speed of AV1 compared to HEVC does not justify the implementation and optimization of this codec for usage in the context of medical video storage.

The Joint Video Exploration Team on Future Video coding (JVET) was founded in October 2015 by the International Telecommunications Union Telecommunication Standardization Sector (ITU-T) Video Coding Experts Group (VCEG) and International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) Moving Picture Experts Group (MPEG) to analyze whether there is sufficient need for the development of a new codec with compression capabilities beyond HEVC. Test cases have been defined for various types of video, including HD, UHD, HDR, and 360° video, and investigations on advanced compression tools using a common software platform have been performed [17]. JVET plans to finalize a new codec standard by the end of 2020 (ITU-T will probably name this codec H.266).

Currently, there are optimized HEVC/H.265 software encoders available, such as the open-source x265 encoder [18] (in particular, x265 has inherited its basic algorithms from its predecessor x264 [19], H.264/AVC) and commercial encoders offered by companies. Independent comparisons have shown that x265 has better compression-quality-speed performance than other industrial encoders [20], especially for full high-definition video. Currently, x265 is commonly used as a baseline for benchmarking analyses performed by video codec companies.

Real-time 4K and 8K hardware encoders are also already available on the market. Nvidia released a new generation of GPUs called Pascal that supports 8K hardware-accelerated HEVC encoding using NVENC technology. Advantech is commercializing the VEGA 3300 Series for 4K/8K HEVC encoding, decoding, and transcoding. NEC also launched its 4K video codec for ultra-low delay applications. Intel with Quick Sync and AMD with VCE technology are quite active in the area of hardware-accelerated HEVC processing as well. There are also companies and universities providing field programmable gate array (FPGA) accelerators for HEVC-based solutions. However, hardware encoders are much more expensive to create and are often much less feature-rich than software encoders.

Although firm compression delivers a substantial reduction of bandwidth used, this requires relatively high implementation complexity to achieve. In many usage scenarios, especially those closer to the video source, the quality is more important than the compression rate. In these scenarios, the source videos are ideally stored in an uncompressed raw format. Nevertheless, due to the increase in resolution, frame rate, and pixel format quality, this is becoming more and more challenging from both a technical and economic point of view.

Lightweight compression, visually lossless compression, intermediate codecs, and mezzanine codecs are keywords that refer to codecs for these scenarios. The main characteristics of such codecs are a relatively small compression ratio, often somewhat configurable between 2× and 8× compression compared to the raw format; intra-only codec tools, to allow for uncomplicated integration in editing workflows; focus on low latency and low complexity real-time implementation; and focus on multigeneration performance. Codecs are designed to reduce quality degradation with multiple transcoding steps.

JPEG XS is a recent promising effort to create a standardized intermediate codec. A strong focus is placed on achieving visually lossless quality with low latency and low-complexity implementations. The standardization efforts also target the integration of file and transmission formats [21]. The target use cases are video transport over professional video links (SDI, IP, ethernet), real-time video storage, memory buffers, omnidirectional video capture and rendering, and on-sensor compression.

In our opinion, considering the large amounts of data that will be produced by the recording of medical procedures, only firm compression should be taken into account. This is why it is crucial that during the standardization of new robust compression codecs, the aspect of quality for medical purposes should be one of the main focus areas.

There are several publications presenting research results on storing compressed medical video sequences. However, it should be stressed that research on medical multimedia data compression is usually performed within narrow medical specializations such as oncology [22] (the EUROPATH project), radiology [22,23] (the EURORAD project), ophthalmology [24], pediatrics [25], surgery [26-31], nursing [32], dentistry [27], internal medicine [33-35], and emergency medicine [36-38]. The effects of this research have influenced the creation and development of standards for storing and compressing medical video sequences [22].

The most important among them is the digital imaging and communication in medicine (DICOM) standard [39]. The DICOM standard has become the primary standard used for the storage of compressed surgery multimedia data. DICOM allows for storing both still images and video sequences [39], although it has not been designed to be used for long video sequences. Compression algorithms to be chosen for reliable storage of source data were one of the critical aspects taken into account during the development process of the standard. Currently, most of the codecs specified in the DICOM standard are lossless codecs. Usage of lossless codecs results in low compression ratios that are now achieved when compressing and storing video sequences (eg, in the DICOM format). Therefore, using these codecs entails superfluous requirements on both the archive memory and the throughput of the streaming network.

Moreover, it is essential to note that the DICOM standard is continuously under development, and it is likely to include more lossy compression methods in the future. For example, with respect to the storage of long bronchoscopic video recordings, the DICOM standard is usually not used. However, the newest version of the standard introduces the first extensions toward storage of long medical records. For diagnostic purposes, the most popular format is AVI. The most commonly used video codec is the (almost lossless) MJPEG, which results in broadband streams that are not suitable for streaming.

By contrast, lossy codecs generally produce desirable narrow-band video streams. Unfortunately, as mentioned above, lossy codecs are hardly ever used for compression of surgery multimedia data if the decompressed images are meant to support a diagnostic process. These codecs are usually designed with the assumption of introducing a significant loss if used for effective compression (the decompressed image may differ significantly from the original). In the case of using visibly degraded images for diagnostic purposes, there is a severe danger of an unacceptable influence on the diagnosis.

Usually, if a presentation of medical multimedia data does not have to support a diagnostic process, but is used for other purposes (eg, educational purposes), popular consumer (usually lossy) codecs are used. In the case of still images, the JPEG compression standard [40] and the JFIF format are used [41]. If video sequences are being compressed, popular codecs such as MPEG-1, MPEG-2, MPEG-4 [42], H.264, H.265, and others are typically used [43].

Video Quality

In many visual applications, the quality of the moving image is not as important as the ability of the optical system to perform specific tasks for which it is created based on these video sequences. Such sequences are called Target Recognition Video (TRV). Regardless of the different ways in which the concept of TRV quality is understood, its verification is necessary to perform dedicated quality testing. The basic premise of these tests is to find TRV quality limits for which the task can be achieved with the desired probability or accuracy. Such tests are usually subjective psychophysical experiments with a group

of subjects. Unfortunately, due to the issue complexity and relatively poor understanding of human cognitive mechanisms, satisfactory results of TRV quality computer modeling have not yet been achieved beyond minimal application areas.

Given the use of TRV, qualitative tests do not focus on the subjects' satisfaction with the video sequence quality but instead measure how the subject uses TRV to accomplish specific tasks. The purposes of this may include: video surveillance such as recognition of vehicle license plate numbers, telemedicine/remote diagnostics for a correct diagnosis, fire safety and fire detection, rear backup cameras such as for parking the car, and games such as spotting and correctly reacting to a virtual enemy.

Since the human factor has a significant influence on the subjective assessment of TRV quality, it is necessary to ask questions on the procedures that need to be complied with to perform it. In particular, problems arise on: the method of selecting the TRV source on which the test TRV (with degraded quality) is based; subjective testing methods and the general manner of conducting the psychophysical experiment; the process of selecting a subjects group in the psychophysical experiment, especially the identification of any prior task knowledge; training subjects before the start of the operation; conditions in which the test will be carried out; and methods of statistical analysis and presentation of results [44].

Metrics for general quality of experience [45], both full-reference methods (eg, peak signal-to-noise ratio or structural SIMilarity [46]) and no-reference methods (eg, video quality indicators, developed by the AGH University research team [47]), conventionally used in video processing systems for video quality evaluation are not appropriate for recognition tasks in video analytics (ie, TRV). Therefore, the correct assessment of video processing pipeline performance is still a significant research challenge.

Video Metadata

Besides the medical data itself, some information (ie, related to the video, patient, pathology, etc) should be recorded in parallel. These metadata are essential to identify the hospital, surgeon, patient, and all information relevant to the surgery. These additional data increase the size of the total data to be recorded or transmitted. Moreover, these data are subject to specific constraints, as they must be processed using standard algorithms. These are also often sensitive data and must be compliant with regulations such as General Data Protection Regulation. Finally, the data are coming from different sources and suffer from a lack of integration, making it difficult for the expert to access the data in their daily practice in some cases.

Despite all of these constraints, exploitation of these metadata offers several advantages and should be of interest for medicine at present and in the future. In particular, this would enable improvement of the everyday practice of experts during the identification of the disease, diagnosis, and follow up.

One way to circumvent the constraints mentioned above, especially the increase in data size and the lack of integration, could be to use data-hiding techniques to hide metadata in the medical video without increasing the total amount of data. In this way, the metadata would never become separated from the recording.

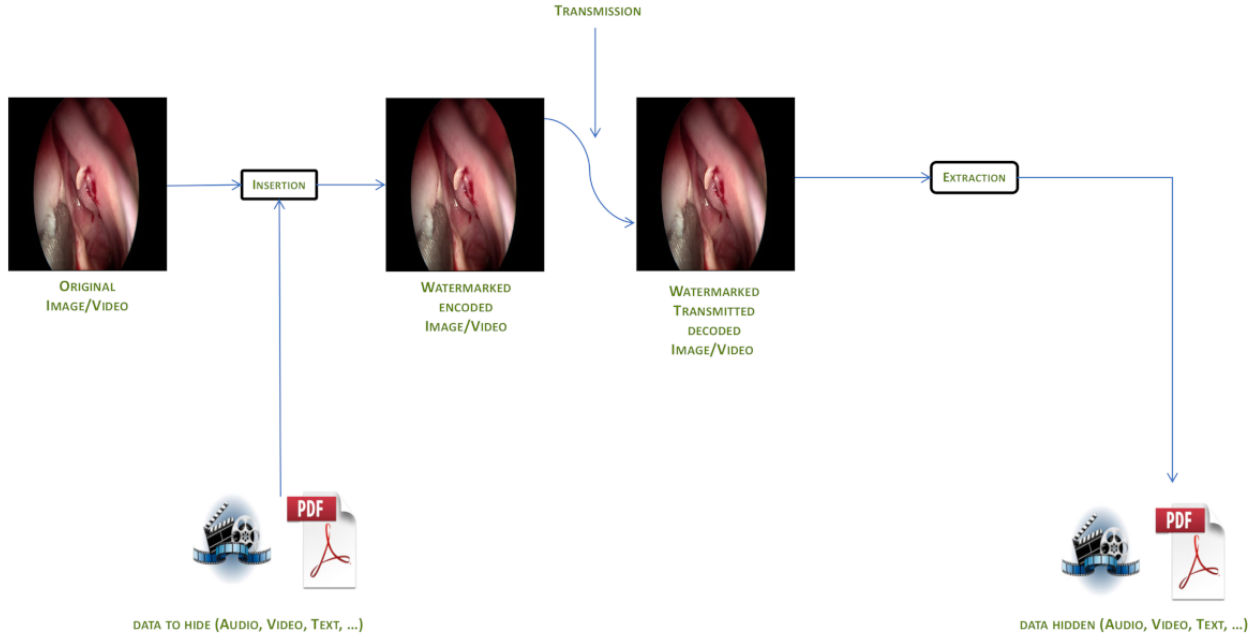
Data hiding can be used to hide different types of information such as text, signature, code, image, video, or audio. It is crucial to include the patient metadata relevant to the recorded procedure in the video, including preexamination results, disease history, and linking to the patient data in a national or hospital database. Such an approach would enable having the data that are most important for follow-up treatment always available with the video. Of course, as these are sensitive medical data, a second version of the recording with anonymized patient data must be prepared for research and educational purposes.

The data-hiding technique was initially conceived in the 1990s to fight against digital piracy and protect the intellectual property of identical copies [48] (especially after the explosion in the use of digital media) during storage and transmission via very open information networks susceptible to attacks. Since its creation, data hiding has been developed in several areas such as telecommunications, video coding, VoD services, medical imaging [49], and computer security.

Figure 2 presents an example of data hiding used to embed information in an ear nose and throat medical video during its transmission. There are two main parts in a data-hiding procedure: data insertion and data detection (or extraction). The mark is inserted within the medical image before transmission. At the reception, the hidden target is detected owing to a secret key and then extracted.

For each application, there is always a relationship between the method of data hiding and the following three main constraints: the capacity of insertion (in bits), imperceptibility/invisibility, and robustness. The size is the amount of information you can insert or hide in the media. Protecting information can distort video quality. It is therefore essential that the mark remains invisible to stay as faithful as possible to the original condition. Data-hiding techniques are sensitive to various distortions that occur during transmission or coding. Robustness to some of these attacks is strongly dependent on the application constraints, while the capacity and invisibility are more related to the structure of the host video itself. Regardless, using a data-hiding technique inevitably leads to a compromise among size, invisibility, and robustness, taking into account the constraints of the application. As an example, in the medical domain, we must place the main effort on invisibility, especially for diagnostic applications.

Figure 2. Schematic depiction of a data-hiding procedure.



Existing Solutions

Our group is already providing services that can be treated as components of the envisioned platform or come close to it.

The medVC platform [50] is a remote collaboration tool for medical professionals that allows for real-time audio-video communication and the usage of specialized medical services. This platform is designed to be installed in ORs, conference rooms, and doctors' offices. It makes it possible to send multiple high-definition video streams coming from cameras, microscopes, endoscopes, and other medical equipment. The platform has a built-in recording system providing high-quality videos encoded with the H.264 or HEVC codec. The system can also automatically upload the recorded videos to cloud platforms such as Medtube [51] or the Interactive Medical University (IMU) [52]. These platforms are dedicated medical video portals, the content of which can be used for medical education. Medtube is a platform allowing any doctor to publish recordings of their surgeries, uploaded using a web browser. IMU is an approach to an end-to-end solution, where videos recorded by dedicated medical recorders are automatically uploaded to a doctor's account, which can then be edited and published.

The deep RIVER [53] company is a spin-off of the University of Lorraine, providing libraries that enable metadata hiding in video content. This novel approach is currently being integrated into the medVC platform, which will allow embedding critical medical data into the video synchronously.

Spin Digital Video Technologies GmbH [54], a small and medium-sized enterprise based in Germany, is the developer of an H.265/HEVC codec implementation, achieving extremely high video quality with reasonable bit rates. Spin Digital software solutions enable media applications that require the latest image and video processing enhancements, including very high resolution (4K, 8K, and 16K), HDR, HFR, WCG, 360° video, and virtual reality. Based on this core technology, Spin Digital has developed a complete solution that includes applications (media player and transcoder) as well as a software development kit that is ready to be integrated into custom applications.

Researchers of the AGH University of Science and Technology [55] working in our group are experts on video quality assessment methods and members of the Video Quality Experts Group [56]. Thanks to their work, recommendations for video coding parameters for medical use have been created, and in the future, emerging codecs will be assessed in this context.

Poznań Supercomputing and Networking Center (PSNC) [57] is one of the leading research and computing centers in Poland. PSNC focuses on using, innovating, and providing high-performance computing, grid, cloud, portal, and network technologies to enable advances in science and engineering. PSNC is also a cloud infrastructure provider, and its network interconnects all clinical hospitals in Poland. Based on this experience and activity, PSNC can become the first operator to introduce secure cloud storage services on the Polish market.

Summary and Prospects

Recording of medical procedures and the requirement to treat these recordings as medical data will be a significant challenge for health care systems around the world. Exabytes of data will be created and will have to be stored for decades. Most hospitals will not be able to handle this from an infrastructural point of view. Therefore, there is a need to provide solutions on national or even international levels. Standardization efforts are needed to define procedures of medical video storage in certified data centers supporting national health care systems. This must be done together with the specification of an API to be implemented by medical recording devices that will automatically upload the recordings to the repositories. In parallel, the standardization of high compression codecs for

medical purposes must progress, and the quality parameters of the codecs, required to ensure medical video quality, must be defined by law. Data hiding can provide an answer to the problems of trustworthiness and integrity of medical data. Thanks to this technique, it is possible to send and store more digital metadata by hiding it in the media. In addition, data hiding offers more advantages for medicine applications by enriching the medical data with useful metadata that can help practitioners improve their diagnosis during the identification of the disease or enhance the level of training for future doctors.

To some extent, there are technologies available or emerging that can be used to build the desired medical video recording and storage platform. With respect to video encoding, there is HEVC [14]; the Joint Collaborative Team on Video Coding is maintaining that the ISO/IEC and ITU-T are conducting ongoing work on new HEVC extensions that can be applied for high-quality medical content. Moreover, the JVET [17] codec of the ITU-T VCEG and ISO/IEC MPEG, which is expected to be completed at the end of 2020, should provide extensions for medical usage. Automatic video quality assessment is more problematic, as conventional video quality assessment metrics are not appropriate for recognition tasks in video analytics; thus, there is still a significant research challenge in this area. New metrics and methods have to be defined, providing an adequate approach to assessing video quality for telemedicine. For video watermarking and trustworthiness of video metadata, novel approaches to data hiding within the video media itself are already available [49]. There are also commercial providers of solutions offering end-to-end services for hospitals, such as medVC [50] and IMU [52], enabling automatic recording, uploading, and storage of medical procedure recordings. However, as there is no legal definition of the quality parameters for highly compressed medical video, these solutions are used mainly for medical education and knowledge sharing at present. Having these standards defined would allow for a broader spectrum of usage scenarios such as getting a second opinion on the treatment, support for follow-up treatment, or analysis in legal cases.

Conflicts of Interest

None declared.

References

1. Telemedizin in Deutschland - Fernbehandlung läuft schleppend. Krankenkassen-Zentrale. URL: https://www.krankenkassenzentrale.de/magazin/telemedizin-in-deutschland-fernbehandlung-laeuft-schleppend-97525 [accessed 2019-08-28]
2. EHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century. European Commission. URL: https://ec.europa.eu/health/sites/health/files/ehealth/docs/com_2012_736_en.pdf [accessed 2019-08-28]
3. Rockstroh M, Franke S, Neumuth T. Requirements for the structured recording of surgical device data in the digital operating room. Int J Comput Assist Radiol Surg 2014 Jan 21;9(1):49-57. [doi: 10.1007/s11548-013-0909-4] [Medline: 23793584]
4. Hey T, Trefethen A. The Data Deluge: An e-Science Perspective. In: Berman F, Fox G, Hey T, editors. Grid Computing: Making the Global Infrastructure a Reality. Hoboken, NJ: John Wiley & Sons; May 30, 2003:809-824.
5. Bergström H, Larsson L, Stenberg E. Audio-video recording during laparoscopic surgery reduces irrelevant conversation between surgeons: a cohort study. BMC Surg 2018 Nov 06;18(1):92 [FREE Full text] [doi: 10.1186/s12893-018-0428-x] [Medline: 30400860]
6. ISO 13485: 2016 Medical Devices - Quality Management Systems - Requirements for Regulatory Purposes. International Organization for Standardization. URL: https://www.iso.org/standard/59752.html [accessed 2019-08-28]
7. ISO 14971: 2007 Medical Devices - Application of Risk Management to Medical Devices. International Organization for Standardization. URL: https://www.iso.org/standard/38193.html [accessed 2019-08-28]
8. EC 60601-1-11:2015 Medical electrical equipment - part 1-11: general requirements for basic safety and essential performance - collateral standard: requirements for medical electrical equipment and medical electrical systems used in the home healthcare. International Organization for Standardization. URL: https://www.iso.org/standard/65529.html [accessed 2019-08-28]
9. Chaabouni A, Gaudeau Y, Lambert J, Moureaux J, Gallet P. H.264 medical video compression for telemedicine: A performance analysis. IRBM 2016 Feb;37(1):40-48. [doi: 10.1016/j.irbm.2015.09.007]
10. In-patient health care. SWAID. URL: http://swaid.stat.gov.pl/en/ZdrowieOchronaZdrowia_dashboards/Raporty_predefiniowane/RAP_DBD_ZDR_3.aspx [accessed 2019-08-28]
11. Life expectancy. SWAID. URL: http://swaid.stat.gov.pl/EN/Demografia_dashboards/Raporty_predefiniowane/RAP_DBD_DEM_17.aspx [accessed 2019-08-28]
12. Health and health care in 2016. Glowny Urzad Statystyczny (Central Statistical Office). URL: https://stat.gov.pl/files/gfx/portalinformacyjny/pl/defaultaktualnosci/5513/1/7/1/zdrowie_i_ochrona_zdrowia_w_2016.pdf [accessed 2019-08-28]
13. Minai H. 8K endoscopes making surgery safer and less painful. Nikkei Asian Review. 2017. URL: https://asia.nikkei.com/Business/Biotechnology/8K-endoscopes-making-surgery-safer-and-less-painful [accessed 2019-08-28]
14. Sullivan GJ, Ohm JR, Han WJ, Wiegand T. Overview of the High Efficiency Video Coding (HEVC) Standard. IEEE Trans Circuits Syst Video Technol 2012 Dec 01;22(12):1649-1668. [doi: 10.1109/TCSVT.2012.2221191]

15. The Alliance for Open Media Kickstarts Video Innovation Era with AV1. Alliance for Open Media. 2018. URL: https://aomedia.org/press%20releases/the-alliance-for-open-media-kickstarts-video-innovation-era-with-av1-release/ [accessed 2019-08-28]
16. MSU Codec Comparison 2017 Part V: High Quality Encoders. CS MSU Graphics & Media Lab, Video Group. 2018 Jan 17. URL: http://compression.ru/video/codec_comparison/hevc_2017/MSU_HEVC_comparison_2017_P5_HQ_encoders.pdf [accessed 2019-08-28]
17. Segall C, Baroncini V, Boyce J, Chen J, Suzuki T. JVET-H1002: Joint Call for Proposals on Video Compression with Capability beyond HEVC. 2018 Jul 20 Presented at: Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 8th Meeting; October 18-24, 2017; Macao, CN.
18. x265 HEVC/H.265 Encoder. x265. URL: http://x265.org [accessed 2019-08-28]
19. x264 H.264/AVC Encoder. VideoLAN. URL: http://www.videolan.org/developers/x264.html [accessed 2019-08-28]
20. Zvezdakov S, Antsiferova A, Kondranin D, Kulikov D, Vatolin D, Erofeev M. HEVC Video Codecs Comparison 2018 (Part IV: FullHD Content, High-Quality Use Case). MSU Codec Comparisons. 2018 Apr 04. URL: http://compression.ru/video/codec_comparison/hevc_2018/ [accessed 2019-08-28]
21. Overview of JPEG-XS. Joint Photographic Experts Group (JPEG). URL: https://jpeg.org/jpegxs/ [accessed 2019-08-28]
22. Rubis Project. 4th Research and Development Framework Programme 1994-1998. In: Healthcare Telematics Projects Final Report. Brussels: European Commission, DG XIII, Directorate B; 2001:17-19.
23. Przelaskowski A. Wavelet-based Image Data Compression (Falkowe metody kompresji danych obrazowych) DSc dissertation. Warsaw: Publishing House of the Warsaw University of Technology (Oficyna Wydawnicza PW); Sep 01, 2002:149-190.
24. Miron H, Blumenthal EZ. Bridging analog and digital video in the surgical setting. J Cataract Refract Surg 2003 Oct;29(10):1874-1877. [doi: 10.1016/s0886-3350(03)00252-9] [Medline: 14604705]
25. Hamilton NM, Frade I, Duguid KP, Furnace J, Kindley AD. Digital video for networked CAL delivery. J Audiov Media Med 1995 Jun;18(2):59-63. [doi: 10.3109/17453059509022995] [Medline: 7494101]
26. Kumar A, Pal H. Digital video recording of cardiac surgical procedures. Ann Thorac Surg 2004 Mar;77(3):1063-1065; discussion 1065. [doi: 10.1016/S0003-4975(03)01258-X] [Medline: 14992927]
27. Reynolds PA, Mason R. On-line video media for continuing professional development in dentistry. Comput Educ 2002 Aug 01;39(1):65-98. [doi: 10.1016/S0360-1315(02)00026-X]
28. Greene PS. Streaming Video for The Annals Internet Readers. Ann Thorac Surg 1998 Apr;65(4):1188-1189. [doi: 10.1016/s0003-4975(98)00219-7]
29. Gandsas A, McIntire K, Palli G, Park A. Live streaming video for medical education: a laboratory model. J Laparoendosc Adv Surg Tech A 2002 Oct;12(5):377-382. [doi: 10.1089/109264202320884135] [Medline: 12470413]
30. Malassagne B, Mutter D, Leroy J, Smith M, Soler L, Marescaux J. Teleeducation in surgery: European Institute for Telesurgery experience. World J Surg 2001 Nov;25(11):1490-1494. [doi: 10.1007/s00268-001-0135-z] [Medline: 11760754]
31. Rosser J, Herman B, Ehrenwerth C. An overview of videostreaming on the Internet and its application to surgical education. Surg Endosc 2001 Jun 19;15(6):624-629. [doi: 10.1007/s004640000338] [Medline: 11591955]
32. Green SM, Voegeli D, Harrison M, Phillips J, Knowles J, Weaver M, et al. Evaluating the use of streaming video to support student learning in a first-year life sciences course for student nurses. Nurse Educ Today 2003 May;23(4):255-261. [doi: 10.1016/s0260-6917(03)00014-5] [Medline: 12727092]
33. Strom J. Streaming Video: Overcoming Barriers for Teaching and Learning. 2002 Jan 01 Presented at: International Symposium Educational Conferencing; 2002; Banff, AB.
34. Wiecha JM, Gramling R, Joachim P, Vanderschmidt H. Collaborative e-learning using streaming video and asynchronous discussion boards to teach the cognitive foundation of medical interviewing: a case study. J Med Internet Res 2003 Jun 27;5(2):e13 [FREE Full text] [doi: 10.2196/jmir.5.2.e13] [Medline: 12857669]
35. Zollo SA, Kienzle MG, Henshaw Z, Crist LG, Wakefield DS. Tele-education in a telemedicine environment: implications for rural health care and academic medical centers. J Med Syst 1999 Apr;23(2):107-122. [doi: 10.1023/a:1020589219289] [Medline: 10435242]
36. Levitan RM, Goldman TS, Bryan DA, Shofer F, Herlich A. Training with video imaging improves the initial intubation success rates of paramedic trainees in an operating room setting. Ann Emerg Med 2001 Jan;37(1):46-50. [doi: 10.1067/mem.2001.111516] [Medline: 11145770]
37. Leung J. Apply Streaming Audio and Video Technology to Enhance Emergency Physician Education. Acad Emerg Med 2002 Oct 01;9:1059. [doi: 10.1197/aemj.9.10.1059]
38. Gisondi MA. Emergency Department Orientation Utilizing Web-based Streaming Video. Acad Emerg Med 2003 Aug 01;10(8):920. [doi: 10.1197/aemj.10.8.920]
39. Digital Imaging and Communications in Medicine (DICOM). National Electrical Manufacturers Association. URL: http://medical.nema.org [accessed 2019-08-28]
40. ISO/IEC 2022:1994: Information technology - Character code structure and extension techniques. International Organization for Standardization. 1994. URL: https://www.iso.org/standard/22747.html [accessed 2019-08-29]

41. Lowe HJ. he New Telemedicine Paradigm: Using Internet-Based Multimedia Electronic Medical Record Systems To Support Wide-Area Clinical Care Delivery. 2001 Jul 06 Presented at: Telemedicine and Telecommunications: Options for the New Century; March 13, 2001; Bethesda.
42. Cuggia M, Mougin F, Le Beux P. Indexing method of digital audiovisual medical resources with semantic Web integration. Int J Med Inform 2005 Mar;74(2-4):169-177. [doi: 10.1016/j.ijmedinf.2004.04.027] [Medline: 15694622]
43. Duplaga M, Leszczuk M, Papir Z, Przelaskowski A. Compression evaluation of surgery video recordings retaining diagnostic credibility (compression evaluation of surgery video). Opto Electron Rev 2008 Dec 01;16(4):428-438. [doi: 10.2478/s11772-008-0041-0]
44. Leszczuk M. Revising and Improving the ITU-T Recommendation. J Telecom Inf Technol 2015 Jan 01:912.
45. Le Callet P, Möller S, Perkis A. Qualinet White Paper on Definitions of Quality of Experience. 2013 Mar 01 Presented at: Fifth Qualinet Meeting European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003); March 12, 2013; Novi Sad.
46. Wang Z, Bovik A, Sheikh H, Simoncelli E. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004 Apr;13(4):600-612. [doi: 10.1109/tip.2003.819861] [Medline: 15376593]
47. Leszczuk M, Hanusiak M, Farias MCQ, Wyckens E, Heston G. Recent developments in visual quality monitoring by key performance indicators. Multimed Tools Appl 2014 Sep 6;75(17):10745-10767. [doi: 10.1007/s11042-014-2229-2]
48. Petitcolas F, Anderson R, Kuhn M. Information hiding-a survey. Proc IEEE 1999 Aug 06;87(7):1062-1078. [doi: 10.1109/5.771065]
49. Puech W, Rodrigues JM. A new crypto-watermarking method for medical images safe transfer. 2004 Sep 06 Presented at: 12th European Signal Processing Conference; September 6-10, 2004; Vienna p. 1481-1484.
50. medVC Remote Collaboration Tool for Medical Professionals. URL: https://medvc.eu/ [accessed 2019-08-28]
51. MedTube. URL: https://medtube.net/ [accessed 2019-08-28]
52. Interactive Medical University - A platform of medical education based on video materials. medVC. URL: https://edu.medvc.eu/ [accessed 2019-08-28]
53. deep RIVER. URL: http://deepriver.fr/ [accessed 2019-08-28]
54. Spin Digital Video Technologies GmbH. URL: https://spin-digital.com/ [accessed 2019-08-28]
55. AGH University of Science and Technology. URL: https://www.agh.edu.pl/en/ [accessed 2019-08-28]
56. Video Quality Experts Group (VQEG). URL: https://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx [accessed 2019-08-28]
57. Poznan Supercomputing and Networking Center (PSNC). URL: https://psnc.pl [accessed 2019-08-28]

Abbreviations

AOM: Alliance of Open Media
API: application programming interface
AVC: advanced video coding
DICOM: digital imaging and communication in medicine
eHealth: electronic health
FPGA: field programmable gate array
HDR: high dynamic range
HEVC: high-efficiency video coding
HFR: high frame rate
ICT: information and communications technology
IEC: International Electrotechnical Commission
IMU: Interactive Medical University
ISO: International Organization for Standardization
ITU-T: International Telecommunications Union Telecommunication Standardization Sector
JVET: Joint Video Exploration Team on Future Video coding
MP: Main Profile
MPEG: Moving Picture Experts Group
OR: operating room
PSNC: Poznań Supercomputing and Networking Center
RExt: range extensions
TRV: target recognition video
VCEG: Video Coding Experts Group
VoD: video on demand
WCG: wide color gamut

Edited by G Eysenbach; submitted 05.02.20; peer-reviewed by M Johanson, E Moro-Rodríguez; comments to author 12.04.20; revised version received 31.05.20; accepted 13.06.20; published 29.07.20.

Please cite as:
Pawałowski P, Mazurek C, Leszczuk M, Moureaux JM, Chaabouni A
Video Cloud Services for Hospitals: Designing an End-to-End Cloud Service Platform for Medical Video Storage and Secure Access
JMIR Biomed Eng 2020;5(1):e18139
URL: https://biomedeng.jmir.org/2020/1/e18139
doi:10.2196/18139
PMID:

Original Paper

Robust Feature Engineering for Parkinson Disease Diagnosis: New Machine Learning Techniques

Max Wang[1], BAdvComp (R&D) (Hons); Wenbo Ge[1], BE (Hons); Deborah Apthorp[1,2], PhD, BPsych (Hons); Hanna Suominen[1,3,4], MSc, Docent, SFHEA, PhD

[1]Research School of Computer Science, College of Engineering and Computer Science, The Australian National University, Canberra, ACT, Australia
[2]School of Psychology, Faculty of Medicine and Health, University of New England, Armidale, NSW, Australia
[3]Machine Learning Research Group, Data61, The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, Australia
[4]Department of Future Technologies, Faculty of Science and Engineering, University of Turku, Turku, Finland

Corresponding Author:
Hanna Suominen, MSc, Docent, SFHEA, PhD
Research School of Computer Science
College of Engineering and Computer Science
The Australian National University
ANU Research School of Computer Science, Hanna Neumann Building, Room 2.35
145 Science Road
Canberra, ACT, 2600
Australia
Phone: 61 431 913 826
Email: hanna.suominen@anu.edu.au

Abstract

Background: Parkinson disease (PD) is a common neurodegenerative disorder that affects between 7 and 10 million people worldwide. No objective test for PD currently exists, and studies suggest misdiagnosis rates of up to 34%. Machine learning (ML) presents an opportunity to improve diagnosis; however, the size and nature of data sets make it difficult to generalize the performance of ML models to real-world applications.

Objective: This study aims to consolidate prior work and introduce new techniques in feature engineering and ML for diagnosis based on vowel phonation. Additional features and ML techniques were introduced, showing major performance improvements on the large mPower vocal phonation data set.

Methods: We used 1600 randomly selected /aa/ phonation samples from the entire data set to derive rules for filtering out faulty samples from the data set. The application of these rules, along with a joint age-gender balancing filter, results in a data set of 511 PD patients and 511 controls. We calculated features on a 1.5-second window of audio, beginning at the 1-second mark, for a support vector machine. This was evaluated with 10-fold cross-validation (CV), with stratification for balancing the number of patients and controls for each CV fold.

Results: We showed that the features used in prior literature do not perform well when extrapolated to the much larger mPower data set. Owing to the natural variation in speech, the separation of patients and controls is not as simple as previously believed. We presented significant performance improvements using additional novel features (with 88.6% certainty, derived from a Bayesian correlated t test) in separating patients and controls, with accuracy exceeding 58%.

Conclusions: The results are promising, showing the potential for ML in detecting symptoms imperceptible to a neurologist.

(JMIR Biomed Eng 2020;5(1):e13611) doi:10.2196/13611

KEYWORDS

machine learning; mobile phone; nonlinear dynamics; Parkinson disease; signal processing, computer-assisted; speech; biomarkers

Introduction

Background

Parkinson disease (PD) affects approximately 1% of the population by the age of 70 years. It is characterized by the deterioration of dopamine-producing neurons in the brain, resulting in symptoms such as abnormal gait, speech, and tremor [1]. Current treatments can provide temporary relief from symptoms and slow its progression [2]; however, these treatments cannot repair damage from the disease. Thus, obtaining an accurate early diagnosis is of high importance.

PD is currently diagnosed with a standardized but subjective test administered by a neurologist or a clinician, the *Unified Parkinson's Disease Rating Scale* (UPDRS) [3]. It is not easy to diagnose, as only a subset of symptoms is present in any one patient [4] and there are many diseases with similar symptoms [5]. PD is especially difficult to diagnose in its early stages, as it is believed that most symptoms only manifest once 20% to 40% of dopamine-producing neurons have deteriorated [6]. Autopsy is one of the only reliable ways to confirm diagnosis, and studies have shown a misdiagnosis rate ranging between 9% and 34% [3,7].

Therefore, the search for a more objective measure for diagnosis is a timely topic in the research community. Discovering more quantifiable biomarkers from sources such as gene expression [8] and bodily fluids [9] is a promising option; however, it is likely that costs will be prohibitive for most early-stage patients uncertain about diagnosis. *Machine learning* (ML) on data from more accessible sources is another viable option, potentially offering an objective and low-cost tool to assist the neurologist in diagnosis through a smartphone-derived version of the UPDRS [10].

Objectives

In this study, we consolidated feature engineering techniques with advances in ML to develop a strong model for PD diagnosis on a large vocal phonation data set. We explored the challenges involved in training ML models on noisy, crowdsourced data and delved into the field of signal processing for audio data. We also found that the experimental setup for classifying healthy and diagnosed individuals does not match the diagnosis process of a neurologist. In addition to merely presenting results as a performance metric, we provided insights into the behavior of these models and showed that it is possible for ML to exceed the performance of clinicians in precise phonation analysis and potentially uncover new biomarkers for PD.

There is a major interest in using ML to assist in PD diagnosis, and current results are positive—often the reported accuracy percentage is in the high 90s using only speech or accelerometer data [11]. However, these results should not be taken at face value, as the experimental setup involves differentiating between (potentially incorrectly) diagnosed PD patients and *healthy controls* (HCs). This oversimplifies the complexities involved in a neurologist's diagnosis in a clinical situation, where the clinician must exclude a number of other causes for the symptoms and handle early-stage patients exhibiting minimal symptoms. Furthermore, the data sets associated with these publications generally consist of fewer than 40 subjects. For such small samples, it is difficult to control for bias and to prevent the overfitting of ML models.

Considering the issues with current data sets, there is an open question as to how ML models should be evaluated. One obvious option is to create a data set by monitoring subjects before any Parkinsonian symptoms until they pass, where the existence of PD can be confirmed through autopsy. With such a longitudinal data set, we could directly compare the performance of ML and neurologists. However, such a data set would be very costly and logistically difficult to collect. To advocate for its collection, there needs to be some evidence of ML effectiveness [10].

Consequently, the question we are most interested in is whether ML techniques can extract more information than observations from a trained clinician. As neurological diagnosis relies on judgment from observation, it is possible that some symptoms are imperceptible but detectable with ML on high-resolution sensor data. Specifically, our research question was to examine whether speech symptoms imperceptible to the human ear can be detected in microphone data using ML.

A larger data set than those that tend to be used in most studies is required to ensure that the results are statistically robust and not influenced by bias. This is not to say that having a larger data set will ensure that there is less influence of bias, but rather that from a larger data set, we can take subsets that are less influenced by biases. We chose the *mPower data set* [12], which contains phonations of the vowel */aa/* from 6000 patients as recorded by an iPhone microphone at 44,100 Hz.

Methods

Features

In this subsection, we describe the methodological background of our study and its feature engineering and signal processing. The subsections after this explain the novel features and materials in our experimental setup. The 2 objectives of these experiments were (1) to consolidate and replicate prior work on a much larger scale and (2) to introduce additional features for dysphonia signal processing to better understand how these features relate to a clinician's diagnosis with speech.

Vocal phonation is the prolonged pronunciation of a sound such as */aa/*. It is an interesting task for diagnosing PD, with evidence symptoms present earlier than other motor symptoms [13]. It also avoids the complexities involved in modeling speech and is, therefore, an easier task in the context of ML. Prior works have shown promising performance, with accuracies of up to 91.4% [14] and 98.6% [11].

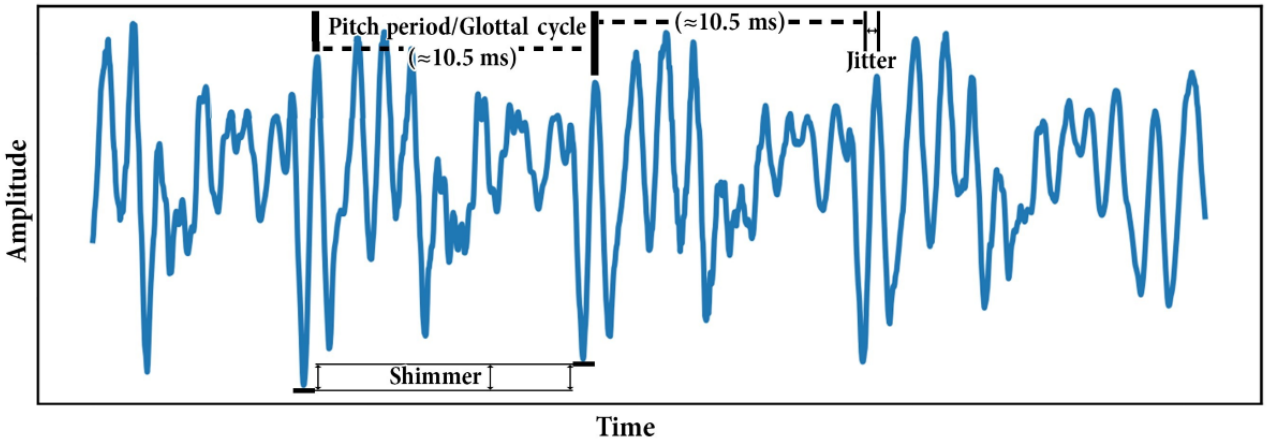
Biologically, phonation is produced by 2 components: the vocal folds and the vocal tract. The vocal folds consist of a flap called the *glottis*, which can be opened and closed. During phonation, air expelled from the lungs causes the glottis to oscillate, producing sound at a range of frequencies. The lowest of these frequencies—the *fundamental frequency*, f_0—represents the duration of 1 oscillation and is often denoted as the *pitch period*. The vocal tract comprises components such as the mouth and nose, and it *shapes* the sound by amplifying and attenuating

certain frequencies. The vocal folds and tract can be viewed as a *source-filter model*, where the source of vocal folds generates the sound signal, shaped (or filtered) by the vocal tract. With PD, the impairment of fine motor control reduces control of the glottis, causing incomplete vocal fold closure. The turbulent airflow around the glottis causes a sound described as a *breathy* or *hoarse* voice and results in increased variation over each glottal cycle. This is termed as *dysphonia*. A similar phenomenon occurs when the vocal folds are irritated by physical causes, such as colds, and it is currently unknown whether differentiation between neurological and physical dysphonia is possible. It is also worth noting that airway diseases as well as muscular diseases and other psychological disorders can also affect the process of sound generation.

People with PD also experience hesitant speech from reduction of cognitive ability and slurred or imprecise articulation from loss of motor control over the vocal tract. This is termed as *dysarthria*. Although dysarthria is very noticeable to humans, it is difficult to quantify it computationally. Spoken language has a wide variety of accents and styles, and it has been shown that models trained on English speakers do not generalize well on German speakers, and vice versa [15].

Dysphonia can be measured using *sustained vowel phonations*, which are easier to model with traditional signal processing approaches and are suitable for small data sets. Dysphonia in vowel phonation is shown in Figure 1. Optimally, both dysphonia- and dysarthria-related features would be used to build models; however, this study focused specifically on dysphonia because of limitations of the data set under consideration.

Figure 1. /aa/phonation from an individual with dysphonia. Variations in jitter (variation in glottal cycle periods) and shimmer (variation in glottal cycle amplitude) are common features for the detection of dysphonia. Algorithms computing each glottal cycle are imperfect, especially for heavy dysphonia.



Features for Dysphonia

Data from a microphone are represented as a stream of values, corresponding to the value of the recorded sound wave at each point in time. A basic microphone samples at approximately 44,100 Hz and quantizes the wave to 2^{16} possible values. The difference between higher sampling rates or more detailed quantization (such as 24 bit) is minimally perceptible to human ears; however, a major problem with low-quality microphones is additional noise and imprecise wave encoding.

Approaches based on signal processing use features that estimate the qualities associated with dysphonia. We divided them into 3 groups: *general* techniques, which are suitable for any time series signal; *dysphonia-specific* techniques, which have been used in prior work specifically to model dysphonia; and *novel* features, which we have shortlisted from other applications as being potentially effective in quantifying dysphonia.

General Signal Processing

Moments are basic statistical descriptors of a signal, with the first 3 moments representing mean, variance, and skewness. For speech, the mean is largely uninformative, and variance corresponds to volume. The zero-crossing rate measures how quickly the signal oscillates around 0 and is a measure of the signal frequency.

Entropy describes the amount of information in a piece of data, if it were modeled by a Bernoulli scheme. It is a simple measure of the complexity/information content of a signal.

The *Fourier* transform decomposes a signal into the amplitudes of frequencies that compose it. This is referred to as mapping from the *time* domain to the *frequency* or *spectral* domain. For speech, this enables us to determine the frequency bands with more energy, corresponding to the fundamental frequency and harmonics. After performing a Fourier transform, the *spectral entropy* [16] can be calculated, which can measure how *sharp* the f_0 and harmonics of speech are. People with PD are expected to have a lower spectral entropy because of the less precise frequency control, causing a more *blurred* Fourier transform.

Squared energy operators, *Teager-Kaiser* energy operators, or other energy operators can obtain the instantaneous frequency and amplitude of a signal [17]. Statistics such as mean or SD and other measures can be computed after applying an energy operator.

Dysphonia Signal Processing

The major feature present in dysphonic speech is the increased variation between each glottal cycle (Figure 1). *Jitter* measures the variation in the length of each glottal cycle, and *shimmer* [18] is the variation in amplitude. These often rely on detecting each glottal cycle, which is not very accurate with current algorithms [19]. The *harmonics-to-noise ratio* (HNR) [20] measures the amount of noise in a signal, which correlates with the *hoarseness* or *breathiness* of speech. The HNR has been improved with a more robust *glottal-to-noise excitation* ratio [21].

The *vocal fold excitation ratio* (VFER) is another extension of the HNR developed in the study by Tsanas et al [22], which also introduced the *Glottal Quotient* (GQ), a measure of the SD duration while the glottis is opened and closed. Both VFER and GQ are built upon concepts of the fundamental frequency estimation algorithm [23].

Mel-frequency cepstral coefficients (MFCCs) are one of the most effective features for speech recognition models, so it is no surprise that they are shown to be similarly effective for dysphonia. Speech recognition involves computing the MFCC at short time intervals and using a Markov model or structured neural network to model temporal information, whereas statistical descriptors have been shown to be effective in detecting dysphonia [11]. There still exists a gap in understanding the relationships between the coefficients and dysphonia.

A recent study [24] showed that *detrended fluctuation analysis*, originally introduced as a measure of the autocorrelation (*autocorrelation* describes the similarity of a signal to itself when offset by a given interval) of a signal, changes with the amount of turbulent airflow in speakers with dysphonia. This study has also proposed the *recurrence period density entropy* (RPDE), which characterizes the periodicity of a signal. These measures are expected to be lower for speakers with dysphonia because of the noise introduced by turbulent airflow. Another study [14] has built upon RPDE to develop *pitch period entropy* as a better measure of the impaired control of pitch experienced by PD patients.

Novel Features

Although existing features have achieved good results on their respective data sets, we have obtained improved performance with this additional set of features. These novel features may not directly relate to dysphonia but have been effective in other signal processing applications, most commonly *electroencephalogram* (EEG), which, similar to speech, is difficult to characterize.

A number of these features relate to *chaos theory*—a field based on understanding the behavior of dynamical systems sensitive to initial conditions [25]. One can imagine the generation of a speech signal as a system, where parameters involve the state of components in the vocal tract. Given no change in parameters such as vocal fold tension, regular */aa/* phonation can be modeled with much fewer dimensions in *phase* space [26]—all possible states of a dynamic system.

The *Lyapunov exponents* quantify the divergence of 2 systems with similar initial parameters. The *largest Lyapunov exponent* (λ^*) characterizes the chaos in a system and is commonly estimated with the algorithm described in the study by Rosenstein et al [26], which reconstructs the system dynamics using a time delay technique. The inverse of  is the Lyapunov time, which defines how long the behavior of a system can be predicted.

The *fractal dimension* is also commonly used in the analysis of dynamical systems. It represents the ratio of the log change in detail to the log change in scale of a signal [27]. This is similar to the coastline paradox, where measuring a coastline with smaller sticks results in an apparent increase in length. The fractal dimension correlates to the complexity of a signal. It has been shown that the fractal dimension of elderly individuals balancing on a force plate is greater than that of younger individuals [28] and, more importantly, that distinct patterns seen in deterministic recurrence quantification analysis of sway (similar to fractal dimension) distinguish patients with PD from controls [29].

General entropy will not differentiate 2 sequences where the frequency of each variable is the same; however, the sequences 0, 0, 0, 0, 1, 1, 1, 1 and 0, 1, 0, 0, 1, 1, 0, 1 are clearly generated by different stochastic processes. The *approximate* and *sample entropy* aims to quantify this [30]. This is extended with multiscale sample entropy [31], which is an especially powerful tool in the analysis of biological signals.

Although signals may appear to have high information content in the time domain, they may be easier to represent in others. For example, the JPEG image compression format primarily relies on human vision, which is less sensitive to high-frequency image details. Images are compressed by taking a Fourier transform and downsampling the high-frequency information. *Spectral entropy* measures the information content of the signal in its frequency-domain representation. The *singular value decomposition* (SVD) factorizes a matrix into orthogonal matrices and *singular values*. *SVD entropy* [16] measures the entropy of the singular values obtained when the signal is embedded with the algorithm described in the study by Rosenstein et al [26].

Many of these features are not simple to interpret; however, our testing shows that they provided significant improvements over prior features used in the literature.

Materials

We chose the mPower data set [12]—a crowdsourced data set consisting of 65,000 */aa/* phonation samples from 6000 participants, of which 1200 were participants with PD.

The primary issue with mPower is quality. Owing to crowdsourcing, participants who are dishonest or perform the task incorrectly can skew results. There are a number of subjects with young-onset PD in mPower; young-onset PD is very uncommon in the general population, but this representation may be possible because of the younger target audience of smartphone apps. Vowel phonation has been captured with a single channel iPhone/iPod microphone at 44,100 Hz, with the

different microphone technology in each generation introducing another variable to the model. Hesitation and phonation of vowels other than /aa/ are common, and the distance between the phone and the user varies: with some speaking directly into the microphone, creating *wind noise*, and others at a distance, introducing significant environmental noise.

We evaluated 1600 randomly selected phonation samples for performing the task correctly, rejecting approximately 40% by using the aforementioned wind, distance, and environmental noise. Using short time energies extracted at 0.1-second intervals with OpenSMILE [32], simple metrics such as variance, mean, and ranges were calculated to rank and filter the samples. Placing a threshold on these gave rise to hand-crafted rules to filter the remaining samples. After filtering, 4100 users remained, 900 with PD, each with a single corresponding sample. We then attempted to balance the joint age and gender distribution within the PD and control groups. Every PD participant had a gender and age *twin* (±2 years; the age-matching was conducted in a way that the twin was at most 2 years younger or older than the respective participant) selected; if a *twin* could not be found, that PD participant was discarded. This resulted in 1022 participants, 511 with PD. Standard libraries for computing the features used to build the modeling included the PyREM package for sleep staging from EEG data [33], nolds [34], and pypsr [35] (Textbox 1). The average f_0 for males and females was assumed to be 120 Hz and 190 Hz, respectively. Sound bit depth was binned to  values for entropy-related calculations. In this feature experiment, we used 6 as the embedding dimension for the time delay embedding methods and calculated τ following the study by Rosenstein et al [26]. Ethical aspects of this research have been approved by the ANU Human Research Ethics Committee with protocol number 2018/108.

Textbox 1. Summary of the features used to build the model.

Dysphonia features

- Prior dysphonia features from the study by Tsanas et al [11]
- Wavelet transform–based features [36]

Novel features

- Higuchi and Petrosian fractal dimension
- Hurst exponent
- Time delay (τ)
- Lyapunov exponents (up to 6)
- Sample and approximate entropy (with r selected by Lu et al [37]) and Fisher information [38]
- Spectral entropy
- Singular value decomposition entropy

Results

Initial Model Validation: Replicating the Previous Results

In a previous study, Tsanas et al [11] used the National Crime Victimization Survey data set, which consists of 263 phonations from 33 people with PD and 10 HCs. A set of 132 features was calculated, and 100 times repeated 10-fold cross-validation (CV) was used to evaluate a *support vector machine* (SVM). On all features, they achieved 97.7% accuracy, and on a 10-feature subset selected by ReliefF, they achieved 98.6% accuracy. They also released an open-source toolbox to assist in replicating the results.

Using this toolbox on our data set, we calculated features on a 1.5-second window of audio beginning at the 1-second mark. An SVM was evaluated with 10 times repeated 10-fold CV over the full feature set and the 10-feature ReliefF subset (Table 1). The data set was stratified by random sampling such that there were equal numbers of PD and control subjects in each fold of CV to better highlight performance. Note that this effectively reduced the data set to approximately 1700 phonations. A grid search was used to select the optimal hyperparameters for the SVM. Of the kernels tested, the *radial basis function* (RBF) was dominant in all models. A Bayesian correlated t test on accuracy was used to determine if a model dominated another [39]. Accuracy was chosen, despite the recent popularity of the *area under the receiver operating characteristic curve* (AUROC) because of criticisms of its bias toward certain predictors [40]. However, for the purposes of comparison with previously published work, we also reported sensitivity, specificity, and AUROC.

Table 1. Cross-validation results of a support vector machine using the full feature set and 10-feature ReliefF subset.

Performance measures	Full, mean (SD)	ReliefF subset[a], mean (SD)
Accuracy (%)	55.6 (4.9)	55.6 (4.5)
Sensitivity (%)	52.3 (7.0)	47.4 (6.6)
Specificity (%)	58.8 (7.8)	63.8 (8.1)
Area under the receiver operating characteristic curve (%)	57.7 (5.7)	58.2 (5.2)

[a]There is a 50.7% probability that the average performance of the ReliefF subset outperforms the full feature set.

Initial Model Validation: Improving Previous Results With Novel Features

The performance of the features by Tsanas et al [11] on the mPower data was evidently worse than that of the NCVS data set. It is possible that this was a consequence of the noisier and lower-quality mPower data. However, overfitting in the original study by Tsanas et al [11] is also highly likely, especially as evidenced by the fact that the MFCC-heavy ReliefF feature subset performed better in experiments by Tsanas et al [11], whereas it performed substantially worse in our testing, achieving 55.6% accuracy rather than 98.6%. Another consideration is the number of observations within each group; having more participants within one group can arbitrarily increase accuracy. It is also ambiguous whether the authors had stratified the phonations on a per-subject scale—failure to do so introduces the *digital fingerprinting* effect. The analysis provided in a recent study [41] exemplifies this, showing that allowing phonations from a subject to appear in both training and testing results in an AUROC of approximately 96%, compared with only 59% in a correctly stratified example.

The features used in the prior literature were clearly insufficient to perform an accurate diagnosis. To improve the model, we introduced an additional range of features that have been effectively applied in the analysis of other biological signals. These features are not directly related to human hearing or speech and thus may be especially useful in detecting symptoms unnoticeable by an expert. We repeated the experiments, adding the novel features, with the results presented in Table 2 showing significant improvement.

Table 2. Cross validation results of the support vector machines using different feature sets.

Performance measures	ReliefF subset, mean (SD)	Novel only, mean (SD)	Combined[a], mean (SD)
Accuracy (%)	55.6 (4.5)	55.3 (4.8)	58.2 (5.1)
Sensitivity (%)	47.4 (6.6)	45.1 (6.5)	56.4 (6.8)
Specificity (%)	63.8 (8.1)	65.5 (7.6)	60.0 (7.0)
Area under the receiver operating characteristic curve (%)	58.2 (5.2)	58.2 (5.2)	60.7 (5.9)

[a]There is an 88.7% probability that the average performance of the combined subset outperforms the ReliefF subset.

Further Improvements: Ensemble Models and Data Augmentation

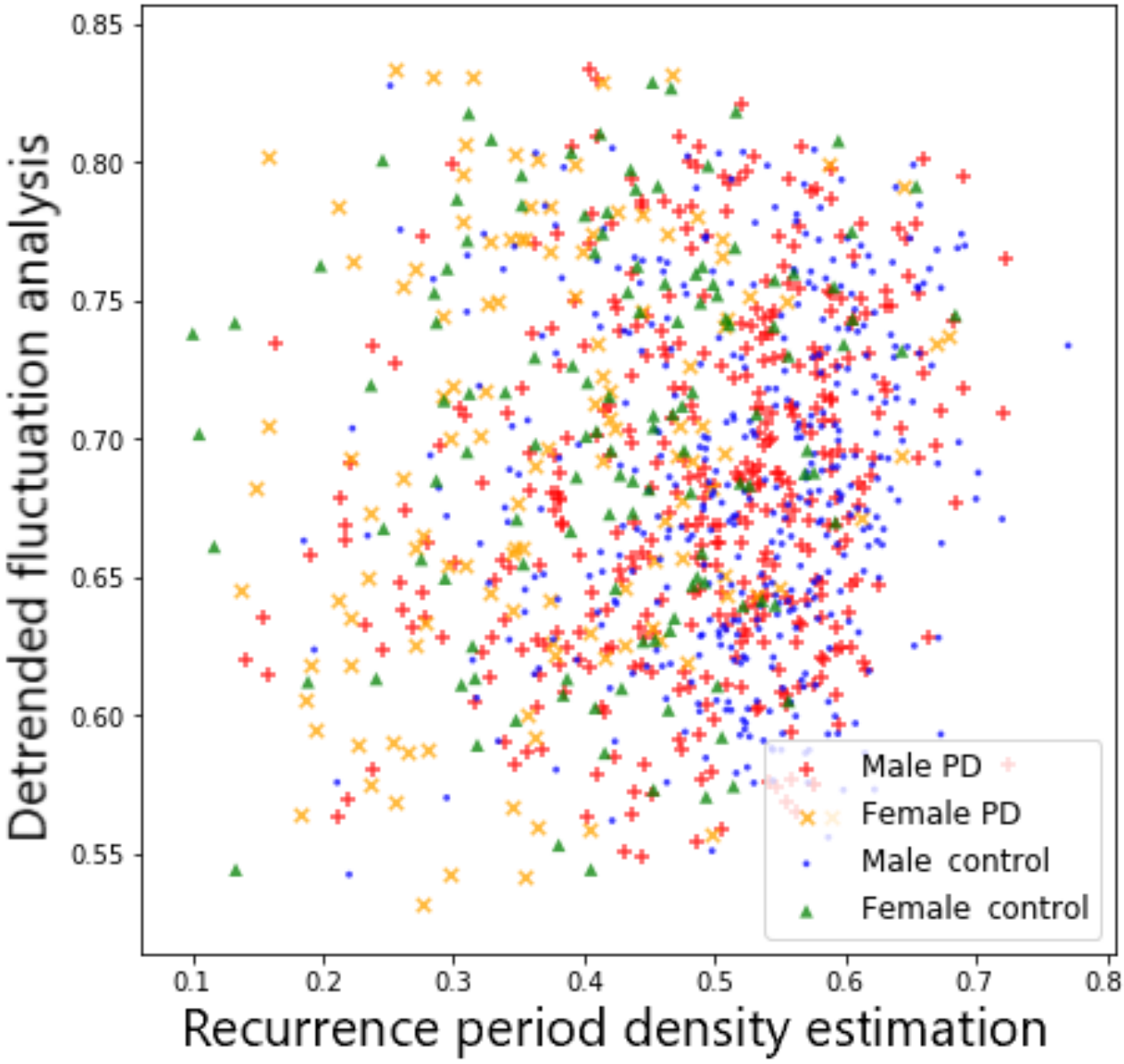
Although the performance was improved, a 58.2% diagnosis accuracy is still far below the requirements for clinical use. In this section, we explore the reasons for the below-expected performance and methods of improving it.

Many dysphonia features rely on estimating the precise length of each glottal cycle, for which the f_0 algorithm of Camacho [42] was used. A preliminary investigation showed that the SD of f_0 exceeded 10 Hz for 18% of the phonations. This was indicative of a substantial failure of the algorithm or a poorly executed recording.

Ensemble Models

Visualizing the features as in Figure 2, it is evident that there are distinct distributions for individuals who are healthy and have PD as well as for males and females. However, there is a large amount of variance and overlap over these distributions, making it infeasible to perform diagnosis over one feature alone.

Figure 2. Visualization of detrended fluctuation analysis against recurrence period density entropy in male and female participants with PD and controls. Although the groups follow distinct distributions, there is heavy overlap. It is evidently difficult to classify PD based on any single feature, and powerful machine learning models are required to make sense of their relationships. PD: Parkinson disease.



Typically, an SVM only performs well if the features are good, which may require a lot of prior knowledge. This is because SVMs can only model linear relationships and, even with the use of nonlinear kernels, are restricted to that particular kernel and are often simple. Other algorithms, such as neural networks, can potentially learn to exploit complex nonlinear relationships between the features to increase performance.

Given the complex relationships between features, it is possible that an ensemble model may achieve better performance. Ensemble models utilize multiple ML models and combine them to create a more powerful one. These ML models can typically be combined by averaging (ie, *voting*) or with other methods (eg, *stacking*), making the combined prediction more powerful than its individual parts.

Building on stacking, *feature-weighted linear stacking* (FWLS [43]) assigns a weight to each feature and model combination. This is motivated by different models being more suitable for using certain features. FWLS was the technique used to ensemble the hundred-model winner of the prestigious Netflix Prize [44].

We ensembled an SVM; Gaussian process; random forest; *k*-nearest neighbor classifier (with k=3); and simple dense neural networks with 3, 5, and 7 hidden layers. All models were suited to the task, diagnosing PD within 3% accuracy of each other. An RBF Gaussian process was chosen to aggregate the models in the stacking and FWLS-based ensembles, as Gaussian processes are inherently probabilistic and suitable for making decisions in situations of high uncertainty. The results are presented in Table 3.

Table 3. Cross validation results of different ensemble methods compared to the support vector machine-only model, using the combined feature set.

Performance measures	Support vector machine only, mean (SD)	Voting, mean (SD)	Stacking, mean (SD)	Feature-weighted linear stacking[a], mean (SD)
Accuracy (%)	58.2 (5.1)	56.5 (5.0)	59.1 (5.1)	5.92 (5.3)
Sensitivity (%)	56.4 (6.8)	54.9 (9.1)	57.4 (7.8)	57.4 (7.6)
Specificity (%)	60.0 (7.0)	58.1 (8.2)	60.9 (7.4)	61.1 (7.4)
Area under the receiver operating characteristic curve (%)	60.7 (5.9)	59.1 (5.6)	62.2 (5.8)	62.3 (5.9)

[a]There is a 70.6% probability that the average performance of the feature-weighted linear stacking ensemble outperforms the support vector machine-only model.

Training ensembles is more computationally expensive; however, in practice, making a prediction from a pretrained model will take a negligible amount of time. Larger data sets may require the use of localized approximations for polynomially computable models such as SVMs and Gaussian processes. However, a major disadvantage of ensembles is the *black box* effect, where their predictions are effectively uninterpretable.

Data Augmentation

Some features were highly sensitive to minor fluctuations in the signal, with their value changing drastically depending on the segment used. Many of these features were not length invariant, and 10% varied by over 0.5 SDs when computed over the same phonation, offset by 0.1 seconds. We computed seven 1.5-second samples from each phonation, ranging from the 1.5- to 4.5-second mark with a 0.5-second step size. We experimented with taking the mean over all 7 values as well as augmenting the data by using the additional samples as extra CV input (ensuring that phonations from the same participant appeared in the same training set or test set). The results are presented in Table 4.

Table 4. Cross validation results of basic data augmentation techniques with the combined feature set and the support vector machine-only model.

Performance measures	Original, mean (SD)	Mean,[a] mean (SD)	Augmented, mean (SD)
Accuracy (%)	58.2 (5.1)	58.7 (4.5)	57.3 (3.7)
Sensitivity (%)	56.4 (6.8)	56.5 (6.8)	54.8 (5.8)
Specificity (%)	60.0 (7.0)	60.8 (7.5)	59.8 (5.8)
Area under the receiver operating characteristic curve (%)	60.7 (5.9)	61.1 (5.1)	60.2 (4.4)

[a]There is a 59.3% probability that the performance with the mean augmented features is better than the nonaugmented features.

Augmentation did not seem to prove useful in overcoming low-quality data and unstable features and seemed to decrease performance. It is possible that sampling from a shorter window exacerbates the instability of the features and that even if one of the windows was more informative than others, that window might not be the same for all recordings. This may essentially introduce more noise into the training data and therefore have an overall negative effect on model performance.

Augmentation proved beneficial when taking the mean over all segments; however, it was not useful when simply including the additional segments as CV data. This is likely because taking the mean reduces the instability and noise of the features, whereas including all the segments exacerbates the problem and introduces more noise into the entire feature set.

Final Results: Performance in Participants With No Speech Difficulties

After developing a decent model to classify PD based on speech, we investigated what it could offer to real-world diagnosis. The raw performance is clearly insufficient to replace neurologists; however, we have shown that the performance bound has not been reached—additional, higher quality data will greatly improve results, along with better and more informative features.

First, we investigated whether it would be possible to detect symptoms imperceptible to a neurologist. If this is possible, the combination of both could potentially increase the accuracy of diagnosis. To do this, a set of PD participants who had *no speech difficulties* (NSDs) was first removed from the data set before any experimentation (including those mentioned in the previous sections) to prevent overfitting during the model evaluation stage. This relied on a field of the mPower UPDRS survey, which defaulted to zero, which we are aware relies on the honestly/reliability of participants. This created a subset of 81 participants (of the previous 4112) who had PD and NSDs.

From the HC participants not used in CV/model evaluation stage, 81 participants were chosen such that they exactly matched the gender of each NSD participant and as closely matched the age as possible. We combined the 2 to form a test set of 162 participants and used the FWLS ensemble trained on the CV set to evaluate the test set. No data augmentation was performed. This is presented in Table 5. Surprisingly, there was an increase in sensitivity between diagnosis of participants with perceptible dysphonia (as indicated by participants having speech difficulties) compared with the diagnosis of those with imperceptible dysphonia (as indicated by participants having NSDs). This implies that there must exist some features that

capture information regarding the disease state of a person but does not relate to audible dysphonia.

Table 5. Performance of the feature-weighted linear stacking ensemble on participants with no audible symptoms and previously unseen data.

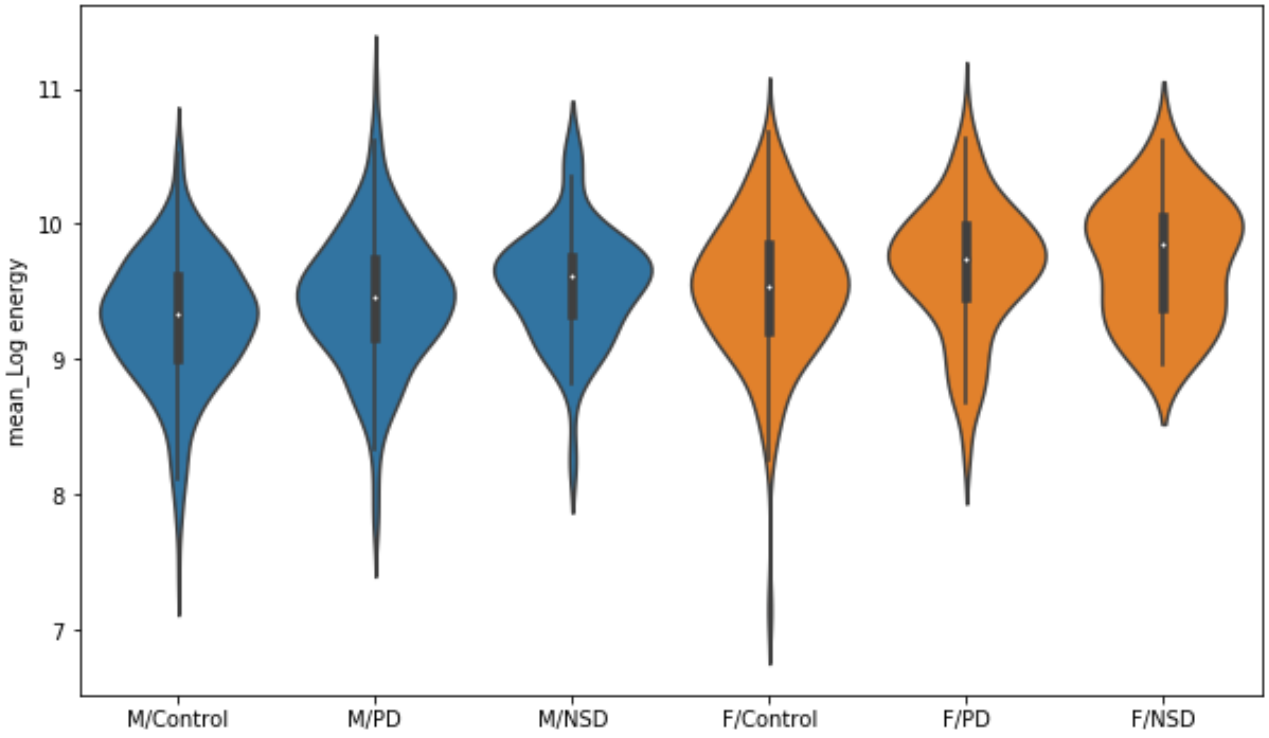
Performance measures	Cross-validation mean	Cross-validation SD	No speech difficulty
Accuracy (%)	59.2	5.3	59.3
Sensitivity (%)	57.4	7.6	63.0
Specificity (%)	61.1	7.4	55.6
Area under the receiver operating characteristic curve (%)	62.3	5.9	66.7

We hypothesized that this is possible because of the more abstract features, which are not related to audible symptoms. To investigate, for each feature, we checked two matters: first, whether the distribution over NSD and PD was similar (ie, captured information about disease state), and second, whether the distribution between NSD and control was different (ie, did not relate to audible dysphonia), implying that the feature is likely discriminative yet inaudible.

To perform significance testing, we first employed a normality test on all features across all groups. If a single feature was normal across all groups (ie, PD, NSD, and HC), then we employed a one-way analysis of variance (ANOVA) to test for significant differences, followed by a Tukey honestly significant difference post hoc. Groups were considered to be sampled from different distributions of a *P* value of .05. If that feature was not normal across all groups, then a Kruskal-Wallis ANOVA was used, followed by a Dunn-Bonferroni posthoc test. Similarly, a value of *P*=.04 is required to say that the feature was sampled from different distributions. Note that because we are only interested in 2 posthoc comparisons (between NSD-PD and NSD-HC), the posthoc tests overcorrected the *P* values. This meant that we can be more confident in any significant difference detected.

We expected that abstract features such as those that try to capture complexity (eg, sample entropy) were likely to be inaudible. These were not biologically motivated and more sensitive to high-frequency information to which the human auditory system was not. This was confirmed in Figure 3, where a significant difference was detected for mean log energy between the NSD and control groups, yet no significant difference was detected between the NSD and PD groups, for males only.

Figure 3. Violin plot of mean_Log energy for males (blue) and females (orange). For males, a significant difference was detected between the no speech difficulty and control groups (*P*<.00002), but not between the no speech difficulty and Parkinson disease groups (*P*=.13). F: female; Log: logarithmic; M: male; NSD: no speech difficulties; PD: Parkinson disease.



In the full results, we found that features such as energy, entropy, and their corresponding coefficients were inaudible, whereas features such as jitter, those derived from fundamental frequency, MFCC, and their corresponding coefficients were audible.

Discussion

Principal Findings

In this study, we have shown that features used in prior literature on small data sets do not perform well when extrapolated to more real-world, much larger data sets such as mPower. Owing to natural variation in speech, the separation of healthy individuals and those with PD is not as simple as previously believed. We have presented significant performance improvements with an additional set of features and techniques such as assembling and data augmentation. Importantly, we have demonstrated this in a robust environment that is well guarded against overfitting.

Strengths and Limitations

Although performance is currently insufficient to substitute for a clinician's diagnosis, significant improvements from simple data augmentation, ensemble models, and additional features imply that peak performance has not yet been achieved. The power of ML only increases as more data become available.

We have also shown that some features used are capable of detecting symptoms of the PD state, yet are *imperceptible to human hearing*. This is a very promising result for the field, with the possibility that a large robust model may eventually outperform humans. For now, these models are a low-cost tool for clinicians to validate their diagnosis; a positive result may be a flag to perform further checks on the patient. This also suggests that vocal phonation features could be used as part of a set of noninvasive biomarkers for PD.

The biggest barrier to the introduction of ML in the clinical setting is trust. This relies on *either* a strong understanding of the models and features *or* good performance on a substantially sized data set with robust evaluation approaches. We believe that developing a deep understanding of ML in clinicians will become increasingly difficult, as models and features become more complex. Thus, it may be best to focus on providing larger data sets. The mPower data set makes great strides in size and availability, but it lacks data control. The field would benefit greatly from a standardized data set against which the performance of models could be empirically evaluated.

Many avenues still remain unexplored. The diagnosis of PD often involves testing a patient's response to medication, such as levodopa. Additional information from phonation samples before and after levodopa consumption may greatly improve performance. We focused only on traditional ML approaches based on signal processing. A notable weakness of feature engineering is information loss, as it is difficult to perfectly describe a signal. Structured neural networks are suitable for making predictions from a raw data stream, such as computer vision and neural networks. In our testing, simple convolutional and recurrent neural networks have achieved similar performance to our best models and greatly improved performance when combined with feature engineering. However, they are more difficult to interpret.

Further improvements could be seen in multidimensional data sets using data from more than one source (eg, postural sway, gait, and finger tapping, all of which could be measured on smartphone devices). Measuring these during clinical visits in a controlled setting under standardized conditions would produce substantially cleaner data at minimal cost, but it is essential to have the cooperation of the clinical and patient communities. Our research provides a first step in establishing the great value these tools could provide, not just in PD.

Acknowledgments

This research was supported by the *Australian Government Research Training Program* Scholarship and delivered in partnership with *Our Health in Our Hands*, a strategic initiative of the Australian National University, which aims to transform health care by developing new personalized health technologies and solutions in collaboration with patients, clinicians, and health care providers. The authors would like to thank Alex Smith, Mehika Manocha, and Professor Christian Lueck for their helpful comments and suggestions.

Conflicts of Interest

None declared.

References

1. Savitt JM, Dawson VL, Dawson TM. Diagnosis and treatment of Parkinson disease: molecules to medicine. J Clin Invest 2006 Jul;116(7):1744-1754 [FREE Full text] [doi: 10.1172/JCI29178] [Medline: 16823471]
2. Fahn S, Parkinson Study Group. Does levodopa slow or hasten the rate of progression of Parkinson's disease? J Neurol 2005 Oct;252(Suppl 4):IV37-IV42. [doi: 10.1007/s00415-005-4008-5] [Medline: 16222436]
3. Tolosa E, Wenning G, Poewe W. The diagnosis of Parkinson's disease. Lancet Neurol 2006 Jan;5(1):75-86. [doi: 10.1016/S1474-4422(05)70285-4] [Medline: 16361025]
4. Thenganatt MA, Jankovic J. Parkinson disease subtypes. JAMA Neurol 2014 Apr;71(4):499-504. [doi: 10.1001/jamaneurol.2013.6233] [Medline: 24514863]
5. Quinn N. Parkinsonism--recognition and differential diagnosis. Br Med J 1995 Feb 18;310(6977):447-452 [FREE Full text] [doi: 10.1136/bmj.310.6977.447] [Medline: 7646647]
6. Brooks DJ. Parkinson's disease: diagnosis. Parkinsonism Relat Disord 2012 Jan;18:S31-S33. [doi: 10.1016/S1353-8020(11)70012-8]

7. Jankovic J, Rajput AH, McDermott MP, Perl DP. The evolution of diagnosis in early Parkinson disease. Parkinson study group. Arch Neurol 2000 Mar;57(3):369-372. [doi: 10.1001/archneur.57.3.369] [Medline: 10714663]
8. Scherzer CR, Eklund AC, Morse LJ, Liao Z, Locascio JJ, Fefer D, et al. Molecular markers of early Parkinson's disease based on gene expression in blood. Proc Natl Acad Sci U S A 2007 Jan 16;104(3):955-960 [FREE Full text] [doi: 10.1073/pnas.0610204104] [Medline: 17215369]
9. Hong Z, Shi M, Chung KA, Quinn JF, Peskind ER, Galasko D, et al. DJ-1 and alpha-synuclein in human cerebrospinal fluid as biomarkers of Parkinson's disease. Brain 2010 Mar;133(Pt 3):713-726 [FREE Full text] [doi: 10.1093/brain/awq008] [Medline: 20157014]
10. Zhan A, Mohan S, Tarolli C, Schneider RB, Adams JL, Sharma S, et al. Using smartphones and machine learning to quantify Parkinson disease severity: the mobile Parkinson disease score. JAMA Neurol 2018 Jul 1;75(7):876-880 [FREE Full text] [doi: 10.1001/jamaneurol.2018.0809] [Medline: 29582075]
11. Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE Trans Biomed Eng 2012 May;59(5):1264-1271. [doi: 10.1109/TBME.2012.2183367] [Medline: 22249592]
12. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 3;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
13. Rusz J, Cmejla R, Tykalova T, Ruzickova H, Klempir J, Majerova V, et al. Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task. J Acoust Soc Am 2013 Sep;134(3):2171-2181. [doi: 10.1121/1.4816541] [Medline: 23967947]
14. Little M, McSharry P, Hunter E, Spielman J, Ramig L. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng 2009 Apr;56(4):1015 [FREE Full text] [doi: 10.1109/TBME.2008.2005954] [Medline: 21399744]
15. Hazan H, Hilu D, Manevitz L, Ramig LO, Sapir S. Early Diagnosis of Parkinson's Disease via Machine Learning on Speech Data. In: 27th Convention of Electrical and Electronics Engineers in Israel. 2012 Presented at: EEEI'12; November 14-17, 2012; Eilat, Israel URL: https://ieeexplore.ieee.org/document/6377065 [doi: 10.1109/eeei.2012.6377065]
16. Roberts SJ, Penny W, Rezek I. Temporal and spatial complexity measures for electroencephalogram based brain-computer interfacing. Med Biol Eng Comput 1999 Jan;37(1):93-98. [doi: 10.1007/BF02513272] [Medline: 10396848]
17. Kaiser JF. ICASSP 90. 1990 International Conference on Acoustics, Speech and Signal Processing. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing. 1990 Presented at: ICASSP'90; April 3-6, 1990; Albuquerque, NM URL: https://ieeexplore.ieee.org/document/115702 [doi: 10.1109/icassp.1990.115547]
18. Horii Y. Jitter and shimmer differences among sustained vowel phonations. J Speech Hear Res 1982 Mar;25(1):12-14. [doi: 10.1044/jshr.2501.12] [Medline: 7087413]
19. Tsanas A, Zañartu M, Little MA, Fox C, Ramig LO, Clifford GD. Robust fundamental frequency estimation in sustained vowels: detailed algorithmic comparisons and information fusion with adaptive Kalman filtering. J Acoust Soc Am 2014 May;135(5):2885-2901 [FREE Full text] [doi: 10.1121/1.4870484] [Medline: 24815269]
20. Yumoto E, Gould WJ, Baer T. Harmonics-to-noise ratio as an index of the degree of hoarseness. J Acoust Soc Am 1982 Jun;71(6):1544-1549. [doi: 10.1121/1.387808] [Medline: 7108029]
21. Michaelis D, Gramss T, Strube HW. Glottal to-noise excitation ratio - a new measure for describing pathological voices. Acta Acustica United with Acustica 1997;83(4):700-706 [FREE Full text]
22. Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. J R Soc Interface 2011 Jun 6;8(59):842-855 [FREE Full text] [doi: 10.1098/rsif.2010.0456] [Medline: 21084338]
23. Kounoudes A, Naylor PA, Brookes M. The DYPSA Algorithm for Estimation of Glottal Closure Instants in Voiced Speech. In: International Conference on Acoustics, Speech, and Signal Processing. 2002 Presented at: ICASSP'02; May 7-11, 2002; Orlando, FL URL: https://ieeexplore.ieee.org/document/5743726 [doi: 10.1109/icassp.2002.1005748]
24. Little MA, McSharry PE, Roberts SJ, Costello DA, Moroz IM. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomed Eng Online 2007 Jun 26;6:23 [FREE Full text] [doi: 10.1186/1475-925X-6-23] [Medline: 17594480]
25. Hegger R, Kantz H, Matassini L. Denoising human speech signals using chaoslike features. Phys Rev Lett 2000 Apr 3;84(14):3197-3200. [doi: 10.1103/PhysRevLett.84.3197] [Medline: 11019046]
26. Rosenstein MT, Collins JJ, de Luca CJ. A practical method for calculating largest Lyapunov exponents from small data sets. Physica D 1993 May;65(1-2):117-134. [doi: 10.1016/0167-2789(93)90009-P]
27. Mandelbrot B. How long is the coast of Britain? Statistical self-similarity and fractional dimension. Science 1967 May 5;156(3775):636-638. [doi: 10.1126/science.156.3775.636] [Medline: 17837158]
28. Doyle TL, Dugan EL, Humphries B, Newton RU. Discriminating between elderly and young using a fractal dimension analysis of centre of pressure. Int J Med Sci 2004;1(1):11-20 [FREE Full text] [doi: 10.7150/ijms.1.11] [Medline: 15912186]
29. Schmit JM, Riley MA, Dalvi A, Sahay A, Shear PK, Shockley KD, et al. Deterministic center of pressure patterns characterize postural instability in Parkinson's disease. Exp Brain Res 2006 Jan;168(3):357-367. [doi: 10.1007/s00221-005-0094-y] [Medline: 16047175]

30. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000 Jun;278(6):H2039-H2049 [FREE Full text] [doi: 10.1152/ajpheart.2000.278.6.H2039] [Medline: 10843903]
31. Costa M, Goldberger AL, Peng C. Multiscale entropy analysis of biological signals. Phys Rev E Stat Nonlin Soft Matter Phys 2005 Feb;71(2 Pt 1):021906. [doi: 10.1103/PhysRevE.71.021906] [Medline: 15783351]
32. Eyben F, Wöllmer M, Schuller B. OpenSmile: The Munich Versatile and Fast Open-Source Audio Feature Extractor. In: Proceedings of the 18th ACM international conference on Multimedia. 2010 Presented at: ACM'10; October 25-29, 2010; Florence, Italy URL: https://dl.acm.org/doi/10.1145/1873951.1874246 [doi: 10.1145/1873951.1874246]
33. Geissmann Q. Python Package for Sleep Scoring From EEG Data. GitHub. 2017. URL: https://github.com/gilestrolab/pyrem [accessed 2017-10-01]
34. Scholzel C. Nonlinear Measures for Dynamical Systems (NoLDS). GitHub. URL: https://github.com/CSchoel/nolds [accessed 2017-10-01]
35. Harrison H. Phase Space Reconstruction: PyPSR. GitHub. 2017. URL: https://github.com/hsharrison/pypsr [accessed 2017-10-01]
36. Tsanas A. Accurate Telemonitoring of Parkinson's Disease Symptom Severity Using Nonlinear Speech Signal Processing and Statistical Machine Learning. Oxford University Research Archive: ORA. 2012. URL: https://ora.ox.ac.uk/objects/uuid:2a43b92a-9cd5-4646-8f0f-81dbe2ba9d74 [accessed 2020-06-09]
37. Lu S, Chen X, Kanters J, Solomon I, Chon K. Automatic selection of the threshold value R for approximate entropy. IEEE Trans Biomed Eng 2008 Aug;55(8):1966-1972. [doi: 10.1109/TBME.2008.919870] [Medline: 18632359]
38. Martin M, Perez J, Plastino A. Fisher information and nonlinear dynamics. Physica A 2001;291(1):523-532 [FREE Full text] [doi: 10.1016/S0378-4371(00)00531-8]
39. Benavoli A, Corani G, Demsar J, Zaffalon M. Time for a change: a tutorial for comparing multiple classifiers through Bayesian analysis. arXivc 2016 Jun 14 epub ahead of print - 1606.04316 [FREE Full text]
40. Peterson AT, Papeş M, Soberón J. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol Model 2008;213(1):63-72 [FREE Full text] [doi: 10.1016/j.ecolmodel.2007.11.008]
41. Neto EC, Perumal TM, Pratap A, Bot BM, Mangravite L, Omberg L. On the analysis of personalized medication response and classification of case vs control patients in mobile health studies: the mPower case study. Arxiv 2017 epub ahead of print [FREE Full text]
42. Camacho A, Harris JG. SWIPE: a sawtooth waveform inspired pitch estimator for speech and music. J Acoust Soc Am 2008 Sep;124(3):1638-1652. [doi: 10.1121/1.2951592] [Medline: 19045655]
43. Sill J, Takacs G, MacKey L, Lin D. Feature-weighted linear stacking. Arxiv 2009:- epub ahead of print(0911.0460) [FREE Full text]
44. Koren Y. The BellKor Solution to the Netflix Grand Prize. The Netflix Prize. 2009. URL: https://www.netflixprize.com/assets/GrandPrize2009_BPC_BellKor.pdf [accessed 2020-06-09]

Abbreviations

ANOVA: analysis of variance
AUROC: area under the receiver operating characteristic curve
CV: cross-validation
EEG: electroencephalogram
FWLS: feature-weighted linear stacking
GQ: Glottal Quotient
HC: healthy control
HNR: harmonics-to-noise ratio
MFCC: Mel-frequency cepstral coefficient
ML: machine learning
NSD: no speech difficulty
PD: Parkinson disease
RBF: radial basis function
RPDE: recurrence period density entropy
SVD: singular value decomposition
SVM: support vector machine
UPDRS: Unified Parkinson's Disease Rating Scale
VFER: vocal fold excitation ratio

Edited by G Eysenbach; submitted 04.02.19; peer-reviewed by A Korchi, E Chaibub Neto; comments to author 02.10.19; revised version received 27.02.20; accepted 14.05.20; published 27.07.20.

Please cite as:
Wang M, Ge W, Apthorp D, Suominen H
Robust Feature Engineering for Parkinson Disease Diagnosis: New Machine Learning Techniques
JMIR Biomed Eng 2020;5(1):e13611
URL: https://biomedeng.jmir.org/2020/1/e13611
doi:10.2196/13611
PMID:

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org