 CONTENTS

Original Papers

Automatic Near Real-Time Outlier Detection and Correction in Cardiac Interbeat Interval Series for Heart Rate Variability Analysis: Singular Spectrum Analysis-Based Approach (e10740)
Michael Lang. ................................................................. 2

An Analytics Framework for Physician Adherence to Clinical Practice Guidelines: Knowledge-Based Approach (e11659)
Jaehoon Lee, Nathan Hulse. ................................................. 19

Perspectives of Orthopedic Surgeons on the Clinical Use of Bioprinted Cartilage: Qualitative Study (e12148)
Ángels Salvador Verges, Luis Fernández-Luque, Meltem Yildirim, Bertran Salvador-Mata, Francesc García Cuyàs. ................................................................. 30

The Impact of Aging and Hand Dominance on the Passive Wrist Stiffness of Squash Players: Pilot Study (e11670)
Taya Hamilton, Stan Durand, Hermano Krebs. ........................................... 41

Determining the Accuracy of Oculus Touch Controllers for Motor Rehabilitation Applications Using Quantifiable Upper Limb Kinematics: Validation Study (e12291)
Leia Shum, Bulmaro Valdés, HF Van der Loos. ........................................... 51
Automatic Near Real-Time Outlier Detection and Correction in Cardiac Interbeat Interval Series for Heart Rate Variability Analysis: Singular Spectrum Analysis-Based Approach

Michael Lang¹, MSc
Graduate School of Excellence Computational Engineering, Technische Universität Darmstadt, Darmstadt, Germany

Corresponding Author:
Michael Lang, MSc
Graduate School of Excellence Computational Engineering
Technische Universität Darmstadt
Dolivostraße 15
Darmstadt, 64293
Germany
Phone: 49 61511624401
Fax: 49 61511624404
Email: michael.lang@ieee.org

Abstract

Background: Heart rate variability (HRV) is derived from the series of R-R intervals extracted from an electrocardiographic (ECG) measurement. Ideally all components of the R-R series are the result of sinoatrial node depolarization. However, the actual R-R series are contaminated by outliers due to heart rhythm disturbances such as ectopic beats, which ought to be detected and corrected appropriately before HRV analysis.

Objective: We have introduced a novel, lightweight, and near real-time method to detect and correct anomalies in the R-R series based on the singular spectrum analysis (SSA). This study aimed to assess the performance of the proposed method in terms of (1) detection performance (sensitivity, specificity, and accuracy); (2) root mean square error (RMSE) between the actual N-N series and the approximated outlier-cleaned R-R series; and (3) how it benchmarks against a competitor in terms of the relative RMSE.

Methods: A lightweight SSA-based change-point detection procedure, improved through the use of a cumulative sum control chart with adaptive thresholds to reduce detection delays, monitored the series of R-R intervals in real time. Upon detection of an anomaly, the corrupted segment was substituted with the respective outlier-cleaned approximation obtained using recurrent SSA forecasting. Next, N-N intervals from a 5-minute ECG segment were extracted from each of the 18 records in the MIT-BIH Normal Sinus Rhythm Database. Then, for each such series, a number (randomly drawn integer between 1 and 6) of simulated ectopic beats were inserted at random positions within the series and results were averaged over 1000 Monte Carlo runs. Accordingly, 18,000 R-R records corresponding to 5-minute ECG segments were used to assess the detection performance whereas another 180,000 (10,000 for each record) were used to assess the error introduced in the correction step. Overall 198,000 R-R series were used in this study.

Results: The proposed SSA-based algorithm reliably detected outliers in the R-R series and achieved an overall sensitivity of 96.6%, specificity of 98.4% and accuracy of 98.4%. Furthermore, it compared favorably in terms of discrepancies of the cleaned R-R series compared with the actual N-N series, outperforming an established correction method on average by almost 30%.

Conclusions: The proposed algorithm, which leverages the power and versatility of the SSA to both automatically detect and correct artifacts in the R-R series, provides an effective and efficient complementary method and a potential alternative to the current manual-editing gold standard. Other important characteristics of the proposed method include the ability to operate in near real-time, the almost entirely model-free nature of the framework which does not require historical training data, and its overall low computational complexity.

KEYWORDS
change-point detection; cumulative sum; forecasting; heart rate variability; R-R series; singular spectrum analysis; ventricular premature complexes

Introduction

Background on Heart Rate Variability
Oscillations in the time interval between successive heart beats, referred to as heart rate variability (HRV), have long been known to allow for insight into the intricate control mechanisms of the autonomic nervous system (ANS) [1-4]. Accordingly, research into HRV has attracted considerable attention over the past 4 decades, as evidenced by an exponential growth of published work [5,6].

In a nutshell, heart rhythm and rate can be said to be governed by the 2 competing divisions of the ANS, that is, the sympathetic nervous system and the parasympathetic nervous system. An increased sympathetic input to the sinoatrial (SA) node yields an increase in heart rate (HR) mediated by the release of adrenaline and noradrenaline and the subsequent activation of β-1-adrenoceptors, resulting in an accelerated diastolic depolarization. On the other hand, an increase in the parasympathetic outflow decreases the HR through the release of acetylcholine by the vagus nerve, which binds to and activates M2 muscarinic acetylcholine receptors in the SA node and eventually slows down the diastolic depolarization rate [4,7-9].

Various HRV measures have been established and are usually categorized as either time domain, frequency domain, or nonlinear analysis techniques [2,4]. To a large extent, the popularity of HRV is because of its easy acquisition and seemingly straightforward interpretation.

The Necessity of Preprocessing the R-R Series
It is crucial to realize that by virtue of its very definition, the HRV encompasses only oscillatory phenomena between heart beats resulting from the SA node depolarization [4,10]. While, ideally, all components of an R-R series originate from the SA node, the actual R-R series in both healthy and diseased subjects are contaminated by outliers due to artifacts and heart rhythm disturbances such as ectopic beats (ie, heart beats whose origin is outside of the SA node). Thus arises the necessity to ensure that the HRV analysis is performed on a series representing only the actual normal sinus rhythm (NSR) interbeat intervals, commonly referred to as N-N series (as in normal-to-normal).

The detrimental impact of ectopics on HRV measures is pronounced and well-documented [6,11-17]. In a recent study Stapelberg et al [18] examined the sensitivity of 38 time domain, frequency domain and nonlinear HRV measures to the addition of artifacts in real and artificial 24-hour recordings. In accordance with previous findings, they concluded short-term time domain HRV measures to be more sensitive to the presence of artifacts than their long-term counterparts. Furthermore, frequency domain measures were found to generally be more sensitive than time domain measures, whereas the less commonly used nonlinear measures were shown to exhibit some inherent robustness properties.

Ectopic heart beats are ubiquitous phenomena and not limited to patients with cardiac disorders and diseases. Hingorani et al [19] retrospectively examined the prevalence of cardiac arrhythmias by scrutinizing 24-hour Holter recordings of 1273 healthy volunteers from 22 phase I clinical trials; note that this specific population is not a representative sample of the overall population. The sample was heavily biased in that, consistent with the requirements of phase I clinical trials, subjects were extensively screened by history, physical examination, and laboratory tests and were therefore significantly healthier than the overall population. Crucially, all subjects underwent 12-lead pretrial electrocardiography (ECG) and were excluded if they had any cardiac conduction disorder or disease, including a personal or familiar history thereof. Those exhibiting ≥2 consecutive ectopics as well as bigeminy, trigeminy and quadrigeminy in their 12-lead ECG examination were also excluded. Despite these rigorous exclusion criteria, Hingorani et al [19], among other findings, found premature atrial complexes in 60.8% of the examined Holter records, followed by premature ventricular complexes (PVCs), which were observed in 43.4% of healthy volunteers. While multifocal PVCs occurred in 5.3%, 3.3% exhibited ≥200 PVCs per 24-hour ECG recordings. A relatively high prevalence of occasional ectopic beats has been reported by numerous other authors as well both in healthy and diseased subjects [20-23].

In summary, the detection and correction of ectopic beats in tachograms are not merely imperative for formal consistency with the HRV definition but arises from the impact of ectopics on HRV measures and the general prevalence of ectopics.

Prior Work and State-of-the-Art
Compared with the overall research interest directed toward HRV applications, the crucial issue of the discrepancy between most real-world R-R series and their respective N-N series has arguably not received as much attention. The actual gold standard of manual R-R series editing, advocated for in the field’s relevant guideline by the European Society of Cardiology and the North American Society of Pacing and Electrophysiology Task Force [4], remains unaltered to this day [11,24].

It should be noted that, aiding and abetting the aforementioned exponential growth of the HRV-related (applied) research, a number of well-designed and widely used software packages for the HRV analysis have been developed and made available to the general public; these include, but are not limited to, Kubios [25,26], Nevrokard’s aHRV [27], and others [28-30]. More often than not, these commercial software packages tend to be closed source and therefore provide only a limited benefit to algorithm developers. On the other hand, it ought to be acknowledged that they are well suited for the needs of practitioners, as they combine advanced preprocessing and analysis features, intuitive graphical interfaces, and support for various file formats.

https://biomedeng.jmir.org/2019/1/e10740/
In automated approaches the task of the detection and correction of outliers is usually tackled as a two-step process. First, outliers in the R-R tachogram are detected, usually by straightforward thresholding based on a fixed percentage difference to previous intervals or their SD [31], although some more sophisticated methods have also been proposed [12].

Second, the detected (regions of) artifacts from step 1 are either discarded (thereby reducing the effective length of the tachogram), substituted by preceding uncorrupted intervals or by intervals obtained through interpolation, with linear and cubic spline interpolation being the most popular techniques [11].

**Objective of This Work**

The objective of this study was to introduce a general framework of low computational complexity that allows for an automatic near real-time detection and correction of outliers in the R-R series based on the singular spectrum analysis (SSA). The main novel contributions of this work include (1) the application of a recently proposed model-free lightweight SSA change-point detection (l-SSA-CPD) algorithm [32]; (2) a modification of l-SSA-CPD through the use of an adaptive control limit sequential ranks (AC-SRC) control chart [33] to drastically reduce detection delays; and (3) upon detection of an anomaly, the substitution of the corrupted tachogram segment by an approximation obtained through recurrent SSA forecasting based on a small outlier-free tachogram segment.

An extensive simulation study comprising 198,000 5-minute long R-R series, obtained by randomly inserting varying amounts of simulated ectopic beats in records taken from the MIT-BIH Normal Sinus Rhythm Database (NSRDB), is conducted to assess the artifact detection and correction performance of the proposed algorithm.

**Fundamentals of Basic Univariate Singular Spectrum Analysis and Singular Spectrum Analysis Based Change-Point Detection**

Recently, this author proposed a low-complexity model-free approach based on SSA and nonparametric cumulative sum (CUSUM) control charts for real-time cardiac anomaly detection, referred to as l-SSA-CPD [32]. It was shown that l-SSA-CPD reliably detects anomalies even when directly applied to unprocessed (ie, no preprocessing was performed) raw ECG and photoplethysmographic records from common databases publicly available through Physiobank [34].

In this study, modifications to the original l-SSA-CPD algorithm will be introduced, and its capabilities will be expanded by adding a recurrent forecasting feature to generate appropriate substitutes of the corrupted tachogram entries. First, however, a brief restatement of SSA and l-SSA-CPD [32] is required.

**Basic Singular Spectrum Analysis Algorithm**

SSA is a powerful technique of time series analysis owing much of its appeal to an inherent model-free concept combining elements of conventional time series analysis, multivariate geometry and statistics, and signal processing [35]. The univariate basic SSA can be considered as the application of the principal component analysis to the Hankel matrix (obtained through an embedding of the original univariate time series) with the subsequent attempt to reconstruct the original series.

Consider $N$ observations $x_{1}, ..., x_{N}$ of a univariate time series and a lag-integer or embedding dimension $M (1 < M < N)$ also commonly referred to as the window length. The basic SSA algorithm then encompasses the following 4 steps:

1. **Embedding** $\tilde{x}_{N} = (x_{1}, ..., x_{N} \rightarrow X \in \mathbb{R}^{M \times K})$

A trajectory matrix $X$ is constructed by mapping $\tilde{x}_{N}$ into a sequence of $K = N - M + 1$ lagged column vectors $X_{j} = (x_{j}, ..., x_{j+M-1})^{T}, j = 1, ..., K$ of size $M$, yielding

$$X = \begin{bmatrix} x_{n+1} & x_{n+2} & \cdots & x_{n+K} \\ x_{n+2} & x_{n+3} & \cdots & x_{n+K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+M} & x_{n+M+1} & \cdots & x_{n+N} \end{bmatrix}.$$  \quad (1)

$X$ is called a Hankel matrix because of its characteristic of having equal elements on the antidiagonals. One can think of $X$ as multivariate data with $M$ characteristics and $K$ observations and accordingly $X_{j}$ of $X$ as column vectors in the $M$-dimensional space $\mathbb{R}^{M}$.

2. **Singular Value Decomposition of X**

Taking the singular value decomposition (SVD) of $X$ decomposes the trajectory matrix into its orthogonal bases and yields a set of $M$ eigenvalues and eigenvectors. Let the eigenvalues of $XX^{T}$ be $\lambda_{1} \geq \cdots \geq \lambda_{M} \geq 0$ and $U_{1}, ..., U_{M}$ be the respective eigenvectors. Then, with $d = \max (i, \text{ such that } \lambda_{i} > 0)$ denoting the rank of $X$, the SVD of $X$ can be expressed as the sum of $d$ elementary matrices

$$X = X_{1} + \cdots + X_{d},$$  \quad (2)

with rank 1 matrices $X_{i} = \sqrt{\lambda_{i}}U_{i}V_{i}^{T}$ and $V_{i} = X^{T}U_{i}/\sqrt{\lambda_{i}}$ being the eigenvectors of $X^{T}X$. Accordingly, the decomposition in Equation (2) is completely characterized by the set of so-called "eigentriples" $(\sqrt{\lambda_{i}}, U_{i}, V_{i})$.

Note that owing to the symmetry of left and right singular vectors, the SVDs of $X$ with lag $M$ and $K = N - M + 1$ are equivalent, entailing the lack of any additional benefit from the use of larger embedding dimensions $M > N / 2$ [32,35-39].

3. **Eigentriple Grouping**

Consider the task of extracting a particular component or signal of interest from observed data contaminated by various artifacts such as noise. To do so, during the third stage of the basic SSA, disjoint subsets of indices $\{1, ..., d\}$ are determined such that the respective systems of eigenvectors span the subspaces associated with those signal components.

In the example of aiming at the separation of a signal from unwanted noise and other disturbances, this entails determining

---

https://biomedeng.jmir.org/2019/1/e10740/
an appropriate subset of indices $I = \{i_1, ..., i_d\}, l < d \leq M$ that span an $l$-dimensional subspace in $\mathbb{R}^M$, denoted as $L_I \subset \mathbb{R}^M = \text{span} \{U_i\} = \text{span} \{U_{i_1}, ..., U_{i_l}\}$, representing the signal whereas the remaining eigentriples with $I = \{i_1, ..., i_d\} / I$ are said to span the noise subspace $L_I \subset \mathbb{R}^M = \text{span} \{U_i\}$.

The trajectory matrix component $X_I$ corresponding to the subset $I$ of eigentriples associated with the signal of interest is then

$$X_I = X_{i_1} + \cdots + X_{i_l}, \quad (3)$$

and the component $X_I$ corresponding to the subset $I = \{i_1, ..., i_l\}$ $I$ associated with the remainder of the observed signal is

$$X_{\bar{I}} = \sum_{i \in \bar{I}} X_i, \quad (4)$$

such that the overall SVD of $X$ can be rewritten as

$$X = X_I + X_{\bar{I}} = \sum_{i \in I} X_i + \sum_{i \in \bar{I}} X_i. \quad (5)$$

In the case of separability (see, e.g., [36] at 17), the contribution of $X_I$ to the entire observed signal $X$ is represented by the respective share of eigenvalues $\frac{\lambda_I}{\sum_{i \in I} \lambda_i}$.

4. Diagonal Averaging

For perfectly separable signal components, all matrices in the expansion of Equation (5) are Hankel matrices, which require that they all have equal entries on their antidiagonals. In real-world problems, however, perfect separability is rarely achievable. While an approximate separability usually suffices, the final fourth step in the SSA algorithm is required, as the matrices $X_i$ from step 3 would have unequal entries on their antidiagonals. Thus, the Hankelization of all $X_i$ is performed, that is, the compliance with the Hankel structure is enforced by taking the average of each antidiagonal and then replacing each element of that antidiagonal with the determined average. This yields

$$\bar{X} = \bar{X}_I + \bar{X}_{\bar{I}} = \sum_{i \in I} \bar{X}_i + \sum_{i \in \bar{I}} \bar{X}_i, \quad (6)$$

with $\bar{X}_i$ being the diagonally averaged matrices.

One can then, for example, easily reconstruct the approximation of the signal of interest through the eigentriples with indices $I$ through the one-to-one correspondence between $\bar{X}_I$ and the respective time series $\bar{x}_N = (\bar{x}_1, \ldots, \bar{x}_N)$ which provides an approximation of either the entire time series $\bar{x}_N$ or some components of it, depending on the particular choice of indices $I$.

**Singular Spectrum Analysis Based Change-Point Detection Rationale**

The sequential application of SSA for change detection is based on Moskvina and Zhigljavsky [40] (see [39] for an exhaustive discussion; see also [35] for an earlier account of the basic idea).

The rationale is to slide 2 (possibly intersecting) moving windows over the time series and to apply the first 3 stages of the basic SSA each time. Think of the first window as adaptively generating a low-dimensional base subspace representation (capturing the main structure in the series) and the second window, which contains at least $M$ new data samples, as generating a test subspace. As both windows are slid over the series, some sort of distance between the base and the test subspace is monitored. If no change-point occurs, the distance remains small, whereas it spikes in the presence of a sudden significant change in the main structure of the series.

The algorithm by Moskvina and Zhigljavsky has been successfully applied to various change detection problems and some improved variations of it have been presented as well (see [32] for a detailed background). Computational complexity, however, ought to be acknowledged as a potentially significant limitation of the said algorithm. Recently, l-SSA-CPD [32] has been proposed as a lightweight alternative. Contrary to the algorithm by Moskvina and Zhigljavsky, which recalculates the SVD of 2 sliding windows each time new observations become available, l-SSA-CPD relies on a nominal low-dimensional signal subspace representation computed only once at startup. As new observations become available, l-SSA-CPD simply allocates them into a vector and calculates a test statistic based on the squared Euclidean distance and the angle between this new data vector and the previously determined nominal reference subspace.

**Singular Spectrum Analysis Parameter Selection**

The behavior of SSA is largely determined by 2 tuning parameters, namely the window length $M$ and the number $l$ of eigenvalues to be used in the embedding and grouping steps, respectively. While it is undisputed that the successful application of SSA hinges on an appropriate choice of $M$ and $l$, practitioners are faced with the nuisance that no clear-cut rules as to how they ought to be determined exist.

For a series of length $N$ and window length (or embedding dimension) $1 < M < N$, as previously discussed, there is no additional benefit from using $M > N / 2$ (see, e.g., [35] at 69 and [36] at 47). It is widely acknowledged that as the appropriate window length $M$ depends on both the underlying data and the particular application, no universal rule for determining it exists [35-39]. General recommendations vary between $M \approx N / 4$ [37] and $M \approx N / 2$ [35,36]. In the case of a periodic signal with period $T$, it is important for $M$ to be, at least, equal to $T$ for the SSA to capture the main structure of the series. Furthermore, $M$ should be proportional to $T$ [35,36,38]. Of note, small windows act like a smoothing filter of width $2M - 1$ [36].

Similarly, it is well known that selecting a proper $l$ is also crucial, for issues are known to arise when eigentriple grouping comprises either an insufficient or an excessive number of eigenvalues. More specifically, if $l$ is too small, the SSA is unable to capture the entire signal whereas if it is too large noise components are unwittingly captured. Since the contribution of a particular series component (associated with the corresponding eigentriple) to the entire series is proportional to the respective eigenvalue, it is common practice to select $l$ by identifying the leading eigenvalues through visual inspection of the eigenvalue
spectra or by setting a predefined eigenvalue share (see, eg, [35,36,38] for a more exhaustive discussion).

In addition to the above-referred monographs, interested readers are also referred to some fairly recent contributions by Hassani et al [41] and Khan and Poskitt [42,43], which approach the issue through the concept of separability and information theory, respectively.

In the light of the pragmatic suboptimal approach pursued here, a detailed discussion of the SSA parameter selection is, however, beyond the scope of this study and therefore omitted.

**Methods**

**The Proposed Outlier Detection and Correction Framework**

The proposed framework consisted of 2 separate stages, namely the detection and the correction of outlier corrupted segments of the tachogram.

**Outlier Detection Using Lightweight Singular Spectrum Analysis Change-Point Detection and Adaptive Control Limit Sequential Ranks Control Charts**

Let $M, N, I, p, q$ be fixed integers such that $l < M < N / 2$ and $0 \leq p < q$. Then l-SSA-CPD proceeds as follows:

1. Initialization at $n = 0$

Akin to MZ, the first 3 steps of the basic SSA are applied on the interval $[n + 1, n + N]$ to obtain a low-dimensional nominal subspace $L_l = L_l^{(w=0)}$ which captures the main structure of the series. This involves embedding as in Equation (1) to obtain a trajectory matrix, which, hereafter, we refer to as the base matrix $X_B = X_B^{(0)} = X_B^{(w=0)}$, taking the SVD of $X_B$ and obtaining $L_l$ through an appropriate choice of $I = \{i_1, \ldots, i_l\}, l < d \leq M$ with $d = \max(i, \text{such that } \lambda_i > 0)$.

2. Then, for each $n = 0, 1, \ldots$ l-SSA-CPD proceeds as follows:

- **Construction of a test matrix $X_T^{(n)}$** on the interval $[n + 1, n + q + M - 1]$ as

  
  
  \[
  X_T^{(n)} = \begin{bmatrix}
  x_{n+p+1} & x_{n+p+2} & \cdots & x_{n+q} \\
  x_{n+p+2} & x_{n+p+3} & \cdots & x_{n+q+1} \\
  \vdots & \ddots & \ddots & \vdots \\
  x_{n+p+M} & x_{n+p+M+1} & \cdots & x_{n+q+M-1}
  \end{bmatrix}
  \]  

  \hspace{1cm} (7)

- **Computation of the detection statistic $D_{n,I,p,q}$** as

  
  
  \[
  D_{n,I,p,q} = D_{n,I,p,q}^{11} \circ D_{n,I,p,q}^{12}
  \]

  
  \hspace{1cm} (8)

  with

  
  \[
  D_{n,I,p,q}^{11} = \frac{1}{Q} \sum_{j=p+1}^{q} \left[ (X_J^{(n)})^T X_J^{(n)} - (X_J^{(n)})^T U_I (U_I)^T X_J^{(n)} \right]
  \]

  
  \[
  D_{n,I,p,q}^{12} = 1 - \cos \left( \angle (X_T^{(n)}, L_I) \right)
  \]

  
  and with the angle $\angle (X_T^{(n)}, L_I)$ taking values in $[0, \pi / 2]$ and accordingly $D_{n,I,p,q}^{11} \in [0,1], X_J^{(n)} = [x_{n+p}, \ldots, x_{n+q+M-1}]^T$ are the column vectors of $X_T^{(n)}$ whereas $U_I = [U_{i1}, \ldots, U_{iI}]$ denotes the $M \times l$ matrix of eigenvectors spanning $L_I$ and $\circ$ denotes the Hadamard (or element-wise) product. Note that throughout this study, $p = N, q = N + 1$ is used, that is, the width $Q = q \times p$ of the test matrices $X_T^{(n)}$ used always equals 1. In other words, each test matrix is actually a single column vector containing $M$ new observations. $Q = 1$ is primarily chosen to allow for an agile response time and to minimize the computational burden [32].

- **Monitoring of $D_{n,I,p,q}$ using the AC-SRC Control Chart**

To reduce detection delays, instead of the sequential ranks CUSUM (SRC) by McDonald [44], as used in the conventional l-SSA-CPD [32], a sequential ranks-based CUSUM with adaptive control limits referred to as AC-SRC [33] is used. AC-SRC is inspired by a distribution-free CUSUM proposed by Chatterjee and Qiu [45] where instead of a fixed threshold a sequence of control limits, determined by the bootstrap estimate of the conditional distribution of the CUSUM test statistic given the last time it was zero, is used. AC-SRC carries the approach of Chatterjee and Qiu over to McDonald’s SRC; the main advantage of the SRC (which is that it does not require training data and control limits can thus easily be determined in advance, eg, through Monte Carlo simulations) is retained while detection delays are significantly reduced because of the use of an adaptive sequence of control limits. Furthermore, owing to a trade-off favoring the lower computational burden and simplicity over the optimal performance, AC-SRC lacks some of the refined fine-tuning routines used in the Chatterjee and Qiu method. The author refrains from a detailed discussion of AC-SRC and refers the interested reader to previous studies [33,45].

Let the sequential rank of $D_{n,I,p,q}$ be denoted as

\[
R_n = 1 + \sum_{r=1}^{n-1} \max(0, D_{n,I,p,q} - D_{r,I,p,q})
\]  

\hspace{1cm} (9)

The AC-SRC then starts with the same recursive equation as the SRC, ie,

\[
C_n = \max \left( 0, C_{n-1} + \frac{R_n}{n+1} - k \right), \quad n \geq 1
\]  

\hspace{1cm} (10)

with $C_0 = 0$ and $k$ a reference constant. Let $T_n$ denote the so-called “sprint length” expressing the time elapsed since $C_n$ was last 0, ie,

\[
T_n = 0 \quad \text{if} \quad C_n = 0
\]

\[
T_n = j \quad \text{if} \quad C_{n-j} \neq 0, \ldots, C_{n-j-1} \neq 0, \quad C_{n-j-1} = 0; \quad j = 1, \ldots, n
\]

Furthermore, let $Y_j$ be a random variable with distribution

\[
Y_j \sim C_n | T_n = j
\]  

\hspace{1cm} (11)
Chatterjee and Qiu showed that the conditional distributions of $C_n | T_n$ in Equation (11) are easier to handle than the unconditional distribution of $C_n$ and under some regularity conditions (see [45] and references therein for details), depend only on $j$ and the underlying process generating distribution, but not on $n$. Then, for any positive integer $j_{\text{max}} \leq n$, the distribution of $C_n$ can be approximated by means of the conditional distributions in Equation (11) as

$$C_n \sim \sum_{j=1}^{j_{\text{max}}} Y_j I_{T_n=j} + Y^* I_{T_n>j_{\text{max}}} , \quad (12)$$

with $Y^* \sim C_n | T_n > j_{\text{max}}$ and $I$ being the indicator function.

It can be shown that [44], given the observed process is in-control, the quantities $R_n / (n+1)$ in Equation (10) are independent and discrete uniform on $\left\{ \frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n}{n+1} \right\}$. Herein lies the key advantage of AC-SRC, in that for some fixed $j_{\text{max}}$ and $k$ one can determine a sequence of control limits $(h_{\text{AC-SRC}})_{j=1}^{j_{\text{max}}}$ from

$$Y_j \sim \left[ \mathbb{C}_n | T_n = j \right] , \quad (13)$$

with $\left\{ \frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n}{n+1} \right\}$. The sequence of control limits $(h_{\text{AC-SRC}})_{j=1}^{j_{\text{max}}}$ can then easily and without the need for training data samples be determined by means of Monte Carlo simulations as outlined elsewhere [33].

The AC-SRC signals a change to have occurred if

$$T_n = j \quad \text{and} \quad C_n > h_{\text{AC-SRC}}, \quad \text{for} \quad 1 \leq j \leq j_{\text{max}} ,$$

or if

$$T_n > j_{\text{max}} \quad \text{and} \quad C_n > h_{\text{AC-SRC}}j_{\text{max}} .$$

To limit the computational burden, it is recommended to calculate control limits only up to a reasonably small $j_{\text{max}}$ after which the control limit is kept fixed at $h_{\text{AC-SRC}}j_{\text{max}}$. In addition, following recommendations by Chatterjee and Qiu, the reference constant $k$ is set proportionate to the expected average sprint length $E \{ T_n \}$. Throughout this work, $k$ was empirically calibrated such that $k = \frac{3j_{\text{max}}}{4}$, whereas the average run length (ARL) was set to a nominal $\text{ARL}_0=500$. A detailed discussion as to the empirical calibration of control charts would, however, be beyond the scope of this work [33]. Note that as $k$ increases, accordingly $j_{\text{max}}$ decreases $E \{ T_n \}$, that is, the likelihood of $C_n$ bouncing back to 0 increases [33,45]. For the particular use case examined here, the use of a small $j_{\text{max}}$ to allow for agile change detection is recommended (see Tables 1–4 and the respective discussion in the next section).

Finally, as with the SRC in the conventional l-SSA-CPD algorithm, the control chart is reinitialized after a change has been detected (by setting $C_n=0$ and $T_n=0$) to allow for the detection of multiple and potentially nearby change-points [32].

Monitoring of the l-SSA-CPD test statistic $D_n, L_p, q$ with $N=20$, $M=10$ by means of AC-SRC with $j_{\text{max}}=8$, $E \{ T_n \}=6$ is showcased on a 5-minute tachogram excerpt of record 16,273 MIT-BIH NSRDB artificially contaminated with 2 and 8 ectopics in Figures 1 and 2, respectively. Note the remarkably low detection delay due to the use of AC-SRC.

**Outlier Correction Using Recurrent Singular Spectrum Analysis Forecasting**

Once an outlier has been detected at say $n=\tau$, accounting for the inherent detection delay and accordingly an uncertainty as to the exact outlier location, the interval $[\tau - \Delta_1, \tau]$ is designated as corrupted tachogram segment. Note that $\Delta_1$ depends on the (average) detection delay and should be chosen carefully. However, it need not be exact or otherwise optimal. Then, a second interval of length $\Delta_2$ immediately preceding the corrupted one is used to obtain an anomaly-free forecast to substitute the corrupted segment with.

The rationale of the proposed framework is depicted in Figure 3, with green and yellow rectangles representing the designated uncorrupted and corrupted segments and the detected change highlighted in red.

As the proposed method operates sequentially as new observations arrive, an appropriate uncontaminated data segment (either uncorrupted to begin with or previously cleaned) to serve as a basis for forecasting is always available. While it may be beneficial to use large values for $\Delta_2$ (assuming sufficient past observations were available), keeping $\Delta_1, \Delta_2$ as small as possible is recommended; as doing so allows for near real-time operation and significantly reduces the computational burden. For the purpose of this study, simplicity was prioritized over (more often than not unnecessary) optimality.
Table 1. Detection performance of adaptive control limit sequential ranks with $j_{max}=6$ and lightweight singular spectrum analysis change-point detection with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples.

<table>
<thead>
<tr>
<th>Record</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>16,265</td>
<td>0.9712</td>
<td>0.9827</td>
<td>0.9825</td>
</tr>
<tr>
<td>16,272</td>
<td>0.9415</td>
<td>0.9802</td>
<td>0.9794</td>
</tr>
<tr>
<td>16,273</td>
<td>0.9905</td>
<td>0.9910</td>
<td>0.9909</td>
</tr>
<tr>
<td>16,420</td>
<td>0.9771</td>
<td>0.9778</td>
<td>0.9777</td>
</tr>
<tr>
<td>16,483</td>
<td>0.9953</td>
<td>0.9846</td>
<td>0.9846</td>
</tr>
<tr>
<td>16,539</td>
<td>0.9326</td>
<td>0.9878</td>
<td>0.9870</td>
</tr>
<tr>
<td>16,773</td>
<td>0.9692</td>
<td>0.9832</td>
<td>0.9829</td>
</tr>
<tr>
<td>16,786</td>
<td>0.9722</td>
<td>0.9758</td>
<td>0.9755</td>
</tr>
<tr>
<td>16,795</td>
<td>0.8800</td>
<td>0.9922</td>
<td>0.9905</td>
</tr>
<tr>
<td>17,052</td>
<td>0.9521</td>
<td>0.9830</td>
<td>0.9825</td>
</tr>
<tr>
<td>17,453</td>
<td>0.9768</td>
<td>0.9904</td>
<td>0.9902</td>
</tr>
<tr>
<td>18,177</td>
<td>0.9549</td>
<td>0.9757</td>
<td>0.9754</td>
</tr>
<tr>
<td>18,184</td>
<td>0.9719</td>
<td>0.9805</td>
<td>0.9803</td>
</tr>
<tr>
<td>19,088</td>
<td>0.9711</td>
<td>0.9917</td>
<td>0.9914</td>
</tr>
<tr>
<td>19,090</td>
<td>0.9900</td>
<td>0.9745</td>
<td>0.9745</td>
</tr>
<tr>
<td>19,093</td>
<td>0.9744</td>
<td>0.9837</td>
<td>0.9834</td>
</tr>
<tr>
<td>19,140</td>
<td>0.9689</td>
<td>0.9920</td>
<td>0.9917</td>
</tr>
<tr>
<td>19,830</td>
<td>0.9948</td>
<td>0.9909</td>
<td>0.9909</td>
</tr>
</tbody>
</table>

Table 2. Detection performance of adaptive control limit sequential ranks with $j_{max}=8$ and lightweight singular spectrum analysis change-point detection with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples.

<table>
<thead>
<tr>
<th>Record</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>16,265</td>
<td>0.9335</td>
<td>0.9860</td>
<td>0.9854</td>
</tr>
<tr>
<td>16,272</td>
<td>0.8595</td>
<td>0.9850</td>
<td>0.9828</td>
</tr>
<tr>
<td>16,273</td>
<td>0.9199</td>
<td>0.9922</td>
<td>0.9913</td>
</tr>
<tr>
<td>16,420</td>
<td>0.9141</td>
<td>0.9810</td>
<td>0.9803</td>
</tr>
<tr>
<td>16,483</td>
<td>0.9569</td>
<td>0.9893</td>
<td>0.9888</td>
</tr>
<tr>
<td>16,539</td>
<td>0.7494</td>
<td>0.9889</td>
<td>0.9863</td>
</tr>
<tr>
<td>16,773</td>
<td>0.8781</td>
<td>0.9858</td>
<td>0.9844</td>
</tr>
<tr>
<td>16,786</td>
<td>0.9038</td>
<td>0.9779</td>
<td>0.9767</td>
</tr>
<tr>
<td>16,795</td>
<td>0.7210</td>
<td>0.9936</td>
<td>0.9897</td>
</tr>
<tr>
<td>17,052</td>
<td>0.8259</td>
<td>0.9851</td>
<td>0.9829</td>
</tr>
<tr>
<td>17,453</td>
<td>0.9013</td>
<td>0.9933</td>
<td>0.9921</td>
</tr>
<tr>
<td>18,177</td>
<td>0.8101</td>
<td>0.9785</td>
<td>0.9771</td>
</tr>
<tr>
<td>18,184</td>
<td>0.8508</td>
<td>0.9833</td>
<td>0.9819</td>
</tr>
<tr>
<td>19,088</td>
<td>0.8667</td>
<td>0.9924</td>
<td>0.9910</td>
</tr>
<tr>
<td>19,090</td>
<td>0.9403</td>
<td>0.9793</td>
<td>0.9789</td>
</tr>
<tr>
<td>19,093</td>
<td>0.8920</td>
<td>0.9869</td>
<td>0.9854</td>
</tr>
<tr>
<td>19,140</td>
<td>0.8796</td>
<td>0.9946</td>
<td>0.9933</td>
</tr>
<tr>
<td>19,830</td>
<td>0.9510</td>
<td>0.9945</td>
<td>0.9941</td>
</tr>
</tbody>
</table>
Table 3. Detection performance of adaptive control limit sequential ranks with $j_{\max}=12$ and lightweight singular spectrum analysis change-point detection with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples.

<table>
<thead>
<tr>
<th>Record</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>16,265</td>
<td>0.8816</td>
<td>0.9875</td>
<td>0.9864</td>
</tr>
<tr>
<td>16,272</td>
<td>0.8240</td>
<td>0.9871</td>
<td>0.9844</td>
</tr>
<tr>
<td>16,273</td>
<td>0.8710</td>
<td>0.9940</td>
<td>0.9925</td>
</tr>
<tr>
<td>16,420</td>
<td>0.8741</td>
<td>0.9850</td>
<td>0.9839</td>
</tr>
<tr>
<td>16,483</td>
<td>0.8990</td>
<td>0.9900</td>
<td>0.9889</td>
</tr>
<tr>
<td>16,539</td>
<td>0.7218</td>
<td>0.9908</td>
<td>0.9879</td>
</tr>
<tr>
<td>16,773</td>
<td>0.8202</td>
<td>0.9871</td>
<td>0.9850</td>
</tr>
<tr>
<td>16,786</td>
<td>0.8421</td>
<td>0.9814</td>
<td>0.9794</td>
</tr>
<tr>
<td>16,795</td>
<td>0.6619</td>
<td>0.9943</td>
<td>0.9898</td>
</tr>
<tr>
<td>17,052</td>
<td>0.7753</td>
<td>0.9893</td>
<td>0.9865</td>
</tr>
<tr>
<td>17,453</td>
<td>0.8450</td>
<td>0.9939</td>
<td>0.9921</td>
</tr>
<tr>
<td>18,177</td>
<td>0.7539</td>
<td>0.9824</td>
<td>0.9805</td>
</tr>
<tr>
<td>18,184</td>
<td>0.7952</td>
<td>0.9853</td>
<td>0.9832</td>
</tr>
<tr>
<td>19,088</td>
<td>0.8030</td>
<td>0.9930</td>
<td>0.9909</td>
</tr>
<tr>
<td>19,090</td>
<td>0.8869</td>
<td>0.9824</td>
<td>0.9814</td>
</tr>
<tr>
<td>19,093</td>
<td>0.8379</td>
<td>0.9892</td>
<td>0.9869</td>
</tr>
<tr>
<td>19,140</td>
<td>0.8004</td>
<td>0.9951</td>
<td>0.9931</td>
</tr>
<tr>
<td>19,830</td>
<td>0.9063</td>
<td>0.9948</td>
<td>0.9939</td>
</tr>
</tbody>
</table>

Table 4. Detection performance of adaptive control limit sequential ranks with $j_{\max}=16$ and lightweight singular spectrum analysis change-point detection with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples.

<table>
<thead>
<tr>
<th>Record</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>16,265</td>
<td>0.8623</td>
<td>0.9878</td>
<td>0.9866</td>
</tr>
<tr>
<td>16,272</td>
<td>0.7947</td>
<td>0.9885</td>
<td>0.9854</td>
</tr>
<tr>
<td>16,273</td>
<td>0.8774</td>
<td>0.9942</td>
<td>0.9929</td>
</tr>
<tr>
<td>16,420</td>
<td>0.8585</td>
<td>0.9851</td>
<td>0.9838</td>
</tr>
<tr>
<td>16,483</td>
<td>0.8950</td>
<td>0.9904</td>
<td>0.9893</td>
</tr>
<tr>
<td>16,539</td>
<td>0.7260</td>
<td>0.9920</td>
<td>0.9890</td>
</tr>
<tr>
<td>16,773</td>
<td>0.7951</td>
<td>0.9880</td>
<td>0.9857</td>
</tr>
<tr>
<td>16,786</td>
<td>0.8022</td>
<td>0.9814</td>
<td>0.9790</td>
</tr>
<tr>
<td>16,795</td>
<td>0.6913</td>
<td>0.9941</td>
<td>0.9900</td>
</tr>
<tr>
<td>17,052</td>
<td>0.7866</td>
<td>0.9905</td>
<td>0.9881</td>
</tr>
<tr>
<td>17,453</td>
<td>0.8176</td>
<td>0.9951</td>
<td>0.9932</td>
</tr>
<tr>
<td>18,177</td>
<td>0.7644</td>
<td>0.9841</td>
<td>0.9823</td>
</tr>
<tr>
<td>18,184</td>
<td>0.7846</td>
<td>0.9866</td>
<td>0.9847</td>
</tr>
<tr>
<td>19,088</td>
<td>0.8119</td>
<td>0.9936</td>
<td>0.9917</td>
</tr>
<tr>
<td>19,090</td>
<td>0.8733</td>
<td>0.9837</td>
<td>0.9827</td>
</tr>
<tr>
<td>19,093</td>
<td>0.8299</td>
<td>0.9898</td>
<td>0.9875</td>
</tr>
<tr>
<td>19,140</td>
<td>0.7949</td>
<td>0.9953</td>
<td>0.9934</td>
</tr>
<tr>
<td>19,830</td>
<td>0.8979</td>
<td>0.9952</td>
<td>0.9944</td>
</tr>
</tbody>
</table>
Figure 1. Adaptive control limits sequential ranks cumulative sum (AC-SRC) monitoring with $j_{\text{max}}=8$, $E\{T_n\}=6$, ARL$_0=500$ of $D_{n,I,p,q}$ with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples (top) for a 5-minute tachogram excerpt of record 16273 MIT-BIH Normal Sinus Rhythm Database (NSRDB) with 2 artificially inserted ectopics (bottom).

![AC-SRC monitoring of $D_{n,I,p,q}$](image1)

![MIT-BIH NSRDB record 16273, 2 artificially inserted ectopics](image2)

Figure 2. Adaptive control limits sequential ranks cumulative sum (AC-SRC) monitoring with $j_{\text{max}}=8$, $E\{T_n\}=6$, ARL$_0=500$ of $D_{n,I,p,q}$ with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples (top) for a 5-minute tachogram excerpt of record 16273 MIT-BIH Normal Sinus Rhythm Database (NSRDB) with 8 artificially inserted ectopics (bottom).

![AC-SRC monitoring of $D_{n,I,p,q}$](image3)

![MIT-BIH NSRDB record 16273, 8 artificially inserted ectopics](image4)
Figure 3. Schematic depiction of the proposed R-R series outlier detection and correction framework. Adaptive control limits sequential ranks cumulative sum with low detection delay is used to monitor lightweight singular spectrum analysis (SSA) change-point detection’s test statistic; when it signals (red vertical bar), a segment of length $\Delta_1$ immediately preceding the detection is labeled as being corrupted (yellow rectangle) and a larger, yet still reasonably small segment of length $\Delta_2$ (green rectangle) is used to construct a trajectory matrix for the recurrent SSA forecasting which eventually yields a short-term SSA forecast of length $\Delta_1$ which the corrupted segment is then substituted with.

Time series forecasting plays a crucial role in many scientific fields and arguably also represents one of the most popular (real-world) application areas of the SSA. Note that there exist different approaches for SSA forecasting [46], most notably recurrent and vector forecasting. Refraining from any detailed discussion (see [35-38] for excellent monographs on the subject matter), the recurrent SSA forecasting approach used in this study will briefly be outlined. As with other more advanced capabilities that go beyond the realm of basic SSA, forecasting and missing value imputation (note that the 2 are closely related with the latter being the more general problem incorporating forecasting as a special case) in a way sacrifice some of the model-free beauty and appeal of SSA by invoking certain model assumptions. On the other hand, this appears to be outweighed by the successful application of SSA forecasting in numerous areas [47-52]. Furthermore, Golyandina and Zhigljavsky [36] point out that for short-term forecasts, there is actually only very little use of the imposed model, that is, the series approximately satisfy some linear recurrent formulas.

Consider again a univariate time series $\mathbb{X}_N = (\bar{x}_1, \ldots, \bar{x}_N)$ In addition, given a vector $Y = (y_1, \ldots, y_M)^T \in \mathbb{R}^M$, let $Y^\perp \in \mathbb{R}^{M-1}$ denote the vector consisting of the first $M-1$ components of $Y$ and similarly $Y \Delta \in \mathbb{R}^{M-1}$ the vector consisting of the last $M-1$ components of $Y$. Set $v^2 = \pi_1^2 + \cdots + \pi_r^2$ where $\pi_i|_{i=1, \ldots, r}$ denotes the last component of the eigenvector $U_r$. Assuming $e_M = (0, 0, \ldots, 0, 1)^T \in \mathcal{L}_r$ (ie, $\mathcal{L}_r$ not to be a vertical space), $v^2$ represents the squared cosine of the angle between $e_M$ and $\mathcal{L}_r$ and $v^2 < 1$ holds. The last component $y_M$ of any vector $Y = (y_1, \ldots, y_M)^T \in \mathcal{L}_r$, can then be shown to be a linear combination of its first $M-1$ components $y_1, \ldots, y_{M-1}$ (ie, of $Y^\perp$)

$$y_M = \alpha_1 y_{M-1} + \alpha_2 y_{M-2} + \cdots + \alpha_{M-1} y_1$$

with vector $A = (a_1, \ldots, a_{M-1})^T$ being

$$A = \frac{1}{1-v^2} \sum_{i=1}^{r} \pi_i U_i^\perp.$$ 

Note that the representation in Equation (15) does not depend on the choice of the basis $U_1, \ldots, U_r \in \mathcal{L}_r$.

With the required notation now having been introduced, the time series $\mathbb{Y}_{N+N_F}$ can be defined as

$$y_i = \begin{cases} \tilde{y}_i & i=1, \ldots, N \\ \sum_{j=1}^{M-1} \alpha_j y_i-j & i=N+1, \ldots, N+N_F \end{cases}$$

with $\tilde{y}_i|_{i=1, \ldots, N}$ being the reconstructed series as in basic SSA and $\mathbb{Y}_{N+N_F}$ the recurrent SSA forecast of length $N_F$.

Two examples of the automatic detection and correction of outliers using the method proposed in this study are shown in Figures 4 and 5 and compared with the benchmark approach of replacing the corrupted segment with an immediately preceding block of R-R data, referred to as block replacement [28]. Note how in both cases, while all outliers have been detected and corrected, there are ectopic-free segments of tachogram which have also been corrected because of false alarms.
Figure 4. Ectopic beats detected on a 5-minute tachogram excerpt of record 19093 MIT-BIH Normal Sinus Rhythm Database (NSRDB) with 6 artificially inserted ectopics using lightweight SSA change-point detection with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples and adaptive control limits sequential ranks cumulative sum with $j_{\text{max}}=6$, $E\{T_n\}=4$, $\text{ARL}_0=500$ are corrected by means of the proposed method with $\Delta_1=M$, $\Delta_2=N$ (top) or by replacing the corrupted segment of length $\Delta_1$ with the immediately preceding block of equal length (bottom).

Figure 5. Ectopic beats detected on a 5-minute tachogram excerpt of record 16265 MIT-BIH Normal Sinus Rhythm Database (NSRDB) with 5 artificially inserted ectopics using lightweight SSA change-point detection with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples and adaptive control limits sequential ranks cumulative sum with $j_{\text{max}}=6$, $E\{T_n\}=4$, $\text{ARL}_0=500$ are corrected by means of the proposed method with $\Delta_1=M$, $\Delta_2=N$ (top) or by replacing the corrupted segment of length $\Delta_1$ with the immediately preceding block of equal length (bottom).

Performance Evaluation

The performance and utility of the proposed method are evaluated using records that are publicly available through Physiobank [34]. More specifically, excerpts of all 18 records in the MIT-BIH NSRDB are used. The European Society of Cardiology/North American Society of Pacing and Electrophysiology Task Force guideline [4] recommends to perform the HRV analysis on ECG recordings of at least 5 minutes, and the said length can be considered as the accepted standard for the short-term HRV analysis. Accordingly, for each of the 18 records in MIT-BIH NSRDB excerpts corresponding to the first 5-minute period of NSR are collected. Note that the records in MIT-BIH NSRDB are cardiologist-annotated, that is, they have manually been checked for errors in beat detection and labeling. Thus, one can circumvent the step of QRS complex-detection and use the R-R intervals as provided (and validated by experts) by Physiobank directly. For all 18 records the first 5 minutes were free of ectopic beats and therefore selected. However, it will be pointed out that 4 records contained a couple of extreme outliers (most likely measurement errors), which were removed manually.
Specifically, the 4 affected records were records number 16,272, 18,177, 19,088, and 19,140, for which 2, 1, 2, and 1 data points were removed, respectively. Besides being readily discernible because they deviate by some order of magnitude from surrounding interbeat intervals, these outliers are also marked as artifacts in the respective annotation files.

Next, the effect of a PVC on the R-R interval was simulated by decreasing the respective entry to 2/3 of its actual value and by increasing the following entry to 4/3 of its value (Textbox 1). That is, for a PVC at say \( n = \tau \), one sets

\[
\text{RR}(\tau) = \text{RR}(\tau) - \frac{\text{RR}(\tau)}{3};
\]

which is consistent with the actual signature of a single PVC (see, e.g., [53]) in an R-R series characterized by a premature beat (and accordingly a significantly smaller R-R interval) followed by a compensatory pause (longer R-R interval) followed by a return to baseline (NSR).

Extensive Monte Carlo simulations were then used to determine the performance characteristics of the proposed method, wherein a large number of artificial R-R series based on the 18 5-minute MIT-BIH NSRDB records were created by first randomly drawing an integer between 1 and 6 and then randomly selecting the position(s) within the series where the respective PVCs would be placed (with the constraint that PVCs be at least 5 samples apart).

### Results

With the proposed method consisting of 2 stages (detection and correction), a separate analysis of each stage appears appropriate.

#### Detection Step

In reporting results pertaining to the detection performance of the first stage, established metrics commonly used in the literature are relied upon the sensitivity (\( Se \)), specificity (\( Sp \)), and accuracy (\( Acc \)), which are defined as

\[
Se = \frac{\text{TP}}{\text{TP} + \text{FN}}
\]

\[
Sp = \frac{\text{TN}}{\text{TN} + \text{FP}}
\]

and

\[
Acc = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{FN} + \text{TN})}
\]

with TP, FP, TN, and FN representing the number of true positives, false positives, true negatives, and false negatives, respectively. Intuitively, sensitivity quantifies the ability to correctly detect (actual) anomalies, specificity quantifies the proportion of nonabnormal segments that are correctly identified as such, and accuracy combines both of the aforementioned aspects. More formally, \( Se \) expresses the empirical statistical power, \( 1 - Sp \) the type I (false positive), and \( 1 - Se \) the type II (missed detection/false negative) error rate.

As reported previously [32], for an event occurring at a time instance \( n = \tau \), one allows for a certain detection delay \( \Delta_d \) and considers a signal from the AC-SRC control chart as true positive if it falls in the interval \([\tau, \tau + \Delta_d]\). All results presented in this study were obtained using \( \Delta_d = \Delta_1 = M \).

Note that the design and calibration of a detection algorithm involves an inherent trade-off between false alarms and missed detections and that furthermore reducing the false alarm rate usually entails an increase in the detection delay. For this specific use case, high sensitivity and short detection delays are arguably of the highest priority; for one wants a high likelihood of detecting all actual outliers (to reliably “clean” the R-R series, thereby transforming it into the respective N-N series) with a manageable (i.e., small) detection delay to be able to narrow down the corrupted segments in the series as much as possible.

The latter is particularly relevant because the proposed method substitutes the (allegedly) corrupted segment by an approximation of the outlier-free segment obtained using recurrent SSA forecasting; one aims to keep the forecasting horizon small to obtain acceptable results without having to tweak the SSA parameters extensively. This comes at the expense of higher false alarm rates, as is evidenced by the results presented in Tables 1-4.

As in the considered design, \( j_{\text{max}} \) is linked to the average sprint length \( E \{ T_n \} \), better agility is generally achieved with rather small \( j_{\text{max}} \) in a use case, as the one considered here, were AC-SRC monitors for large deviations [33,45].

While AC-SRC achieved satisfactory performance over a wide range of values \( j_{\text{max}} \) (results not reported here), upon examination of the results reported in Tables 1-4, which are based on 18,000 R-R series, \( j_{\text{max}}=6 \) was chosen to proceed to the second stage of the proposed method (i.e., the correction of corrupted segments of the R-R series).

#### Correction Step

As mentioned earlier this design choice entails higher false alarm rates, as discernible in the 2 examples shown in Figures 4 and 5, which both exhibit actually uncorrupted segments that have been “corrected” by the proposed algorithm because of false alarms, thereby introducing some distortions. More importantly though, both examples also clearly show that all artificially inserted outliers because of PVCs were detected and appropriately corrected.
Results pertaining to the second stage of the proposed method are reported in Table 5, wherein the root mean square error (RMSE) is used to quantify the discrepancy between the outlier-cleaned series and the actual (uncontaminated) series. Let $\mathbf{\hat{x}} = (\hat{x}_1, \ldots, \hat{x}_N)$ denote our allegedly cleaned R-R series.

Calculation of the RMSE is straightforward,

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{x}_i - x_i)^2}. \quad (17)$$

Furthermore, to better compare the RMSE of the proposed method to its competitor (the commonly used block replacement method is used here for benchmarking purposes), consider also the relative RMSE (RRMSE) given by

$$\text{RRMSE} = \frac{\text{RMSE}_{\text{proposed}}}{\text{RMSE}_{\text{benchmark}}}. \quad (18)$$

Note that if the RRMSE, as given in Equation (18) is $<1$, it implies that the novel approach presented in this paper outperforms the competing benchmark approach by $100 \cdot (1 - \text{RRMSE})$. It shall be emphasized, however, that the block replacement method as it is applied here (to correct for outliers in the R-R series) uses the first stage of the proposed method as well, its performance should thus be seen in the context of being applied to pretty tight segments of the series that ought to be substituted, which is all because of the AC-SRC I-SSA-CPD algorithm of the first stage of the framework presented in this study.

That being said, the results reported in Table 5, which are based on 180,000 R-R series, show that the proposed method, which substitutes corrupted tachogram segments by an approximation of the respective outlier-free segment obtained by means of recurrent SSA forecasting, clearly outperforms the competing block replacement method. In fact, it outperforms its competitor for all 18 records (on which the extensive simulations built upon), on average by almost 30%.

### Discussion

#### Background and Relevance

The HRV has long been known to allow for valuable insights into the intricate control mechanisms of the ANS and has a long track record of exponentially growing research interest and output. A crucial prerequisite for any HRV analysis is the exclusion of all artifacts and abnormalities, that is, a clean N-N series is required. Such a clean N-N series is virtually never available, mostly because of various interferences, beats not
detected by the QRS complex-detector, and various highly prevalent heart rhythm disturbances.

Compared with the overall research effort directed toward the HRV, the issue of automatic R-R series (pre)processing has received rather little attention. In fact, manual editing still represents the gold standard. It should be noted that every automatic approach comes with risks and benefits; the former lie in the inherently nonzero type I and type II error rates, whereas the latter include, but are not limited to, considerable savings in both cost and time. Considering that error rates for manual approaches are nonzero as well and that they are most likely an increasing function of time (whereas in automatic approaches they are usually not), the choice of which one or which particular combination to use should always be made on an individual basis, considering all requirements and circumstances of the particular case. Accordingly, while the proposed method could be construed as a potential alternative to manual editing, in the light of the above considerations, this author would rather recommend it as a complementary method and or a preprocessing tool.

The main contribution of this work is the introduction of a general framework of low computational complexity that allows for an automatic near real-time detection and correction of outliers in R-R series based on the SSA. While related work pertaining to the use of the SSA in ECG and R-R data processing exists [54], to the best of the author’s knowledge, this study is the first to propose a general SSA-based framework that handles both detection and correction of (unwanted) anomalies in the R-R series.

Principal Findings

An extensive simulation study comprising 198,000 5-minute R-R series, obtained by randomly inserting varying amounts of simulated ectopic beats in records taken from the MIT-BIH NSRDB was conducted to assess the artifact detection and correction performance of the proposed algorithm.

It was shown that the proposed algorithm reliably detects outliers in the R-R series and achieved an overall sensitivity of 96.6%, specificity of 98.4%, and accuracy of 98.4%. Furthermore, it compares favorably in terms of discrepancies of the cleaned R-R series compared with the actual N-N series, outperforming a block replacement approach on average by almost 30%. It should also be emphasized that a suboptimal pragmatic approach was pursued, deliberately refraining from an optimization of the various (SSA and AC-SRC) tuning parameters, which would most likely have resulted in further performance improvements at the expense of simplicity and computational complexity.

Other important characteristics of the proposed method include the ability to operate in near real-time, the almost entirely model-free nature of the framework, and the low computational complexity. Moreover, it should be pointed out that the proposed method is not limited to the R-R series but can be applied broadly, as all of its components are deliberately kept as general as possible. This entails the additional benefit of not being limited to the removal of PVCs (although only this has extensively been investigated and tested so far) but rather being able to deal with all kinds of anomalies that may contaminate the observed series of interest.

Limitations and Future Research

This work has several limitations, some of which open (new) avenues for future research.

A major limitation is because of the lack of established and generally recognized and validated procedures other than manual editing, which transforms the performance comparison among competing methods in a nontrivial task. Second, to increase the readability and decrease the length, the paper fails to provide general recommendations and guidelines as to how various tuning parameters should be chosen. Finally, the SSA represents a very active research field, as in the last couple of years, and several promising advances and new developments have been reported, but thus far have not been considered in the research reported in this work. This, however, does not infringe on the relevance of this work, as the objective was to introduce a simple, general framework of low computational complexity and to investigate the use of SSA in both the detection and correction step.

Several of the abovementioned recent developments, however, appear to open promising avenues for future research. In particular, the author intends to expand and improve the work presented here by focusing on the issue of the SSA parameter selection and by considering latest developments, especially pertaining to SSA forecasting. For the latter, the interested reader is referred to recent publications by Hassani and Kalantari [55-58].

Acknowledgments

ML was supported by the “Excellence Initiative” of the German Federal and State Governments and the Graduate School of Excellence Computational Engineering at Technische Universität Darmstadt. The views expressed in this article are solely those of the author in his private capacity and do not necessarily reflect the views of Technische Universität Darmstadt or any other organization.

Conflicts of Interest

None declared.

References


Abbreviations

AC-SRC: adaptive control limits sequential ranks cumulative sum
ANS: autonomic nervous system
ARL: average run length
CUSUM: cumulative sum
ECG: electrocardiography
HR: heart rate
HRV: heart rate variability
l-SSA-CPD: lightweight SSA change-point detection
NSR: normal sinus rhythm
NSRB: Normal Sinus Rhythm Database
PVC: premature ventricular complex
RMSE: root mean square error
RRMSE: relative root mean square error
SA: sinoatrial
SRC: sequential ranks cumulative sum
SSA: singular spectrum analysis
SVD: singular value decomposition

©Michael Lang. Originally published in JMIR Biomedical Engineering (http://biomedeng.jmir.org), 30.01.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on http://biomedeng.jmir.org/, as well as this copyright and license information must be included.
An Analytics Framework for Physician Adherence to Clinical Practice Guidelines: Knowledge-Based Approach

Jaehoon Lee¹,², PhD; Nathan C Hulse¹,², PhD

¹Intermountain Healthcare, Salt Lake City, UT, United States
²Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Nathan C Hulse, PhD
Intermountain Healthcare
36 State Street
Salt Lake City, UT, 84111
United States
Phone: 1 3852976482
Email: nathan.hulse@imail.org

Abstract

Background: One of the problems in evaluating clinical practice guidelines (CPGs) is the occurrence of knowledge gaps. These gaps may occur when evaluation logics and definitions in analytics pipelines are translated differently.

Objective: The objective of this paper is to develop a systematic method that will fill in the cognitive and computational gaps of CPG knowledge components in analytics pipelines.

Methods: We used locally developed CPGs that resulted in care process models (CPMs). We derived adherence definitions from the CPMs, transformed them into computationally executable queries, and deployed them into an enterprise knowledge base that specializes in managing clinical knowledge content. We developed a visual analytics framework, whose data pipelines are connected to queries in the knowledge base, to automate the extraction of data from clinical databases and calculation of evaluation metrics.

Results: In this pilot study, we implemented 21 CPMs within the proposed framework, which is connected to an enterprise data warehouse (EDW) as a data source. We built a Web-based dashboard for monitoring and evaluating adherence to the CPMs. The dashboard ran for 18 months during which CPM adherence definitions were updated a number of times.

Conclusions: The proposed framework was demonstrated to accommodate complicated knowledge management for CPM adherence evaluation in analytics pipelines using a knowledge base. At the same time, knowledge consistency and computational efficiency were maintained.


KEYWORDS
clinical practice guidelines; care process model; visual analytics; clinical decision support

Introduction

Clinical practice guidelines (CPGs) are systematically developed statements that assist clinicians in making decisions about appropriate patient care for specific clinical circumstances [1]. Since the nature of CPGs includes complicated clinical knowledge, it is known to be challenging not only to formulate clinicians’ logics into the form of guidelines, but also to translate and implement these guidelines properly into clinical tasks and processes. Therefore, a number of studies have tried to develop systematic ways to implement CPGs [2,3], including computer-aided clinical decision support (CDS)-based approaches that enable personalized and timely implementation of CPGs [4-6] and knowledge-based approaches to systematically transform complicated knowledge of CPGs into clinical decision and practices [7-9].

Since clinical knowledge within CPGs originated from evidence of best practices, realization of CPGs in practice has had a positive impact on clinical workflow and patient outcome [10-14]. Therefore, it is important to measure and evaluate physicians’ adherence to CPGs in order to understand how providers are following guidelines in the postimplementation phase [15,16]. This evaluation may be an interdisciplinary project involving domain experts, knowledge engineers, and...
data analysts, among others. Domain experts should derive evaluation logics and metrics from CPGs and document them by collaborating with knowledge engineers. To perform this evaluation, database engineers should create queries based on the definitions and run them against the clinical database. As a result of data extraction, evaluation outcomes would be delivered to consumers (ie, clinical champions or management leaders), often in the form of Web-based reports.

The problem is that knowledge gaps always exist while transforming logics of CPGs into evaluation-definition documents, developing queries, and generating reports. For example, translating definitions into computationally executable queries may vary by individual knowledge engineers and data developers [17]. Once analytics are delivered, consumers can only see the resulting data, but would lack an understanding of what logics were used to extract the numbers. If analytics are being delivered in a regular manner and evaluation logics are modified, it becomes confusing to know whether a report was made based on an old or new definition. In particular, analytics have become more integrated and automated by integrating data pipelines, report generation, and delivery workflow; this may involve more knowledge translation into the framework and may require a systematic method of management [18,19].

Knowledge gaps are caused by the difficulty of handling too many data points, miscommunication between domain experts and developers, misinterpretation of guidelines, loss of authorship of documents, and revision and update of knowledge sources [20]. This may result in data inconsistency across the analytics pipeline, causing consumers to experience a lack of trust regarding the analytics results. Consumers may be confused and may feel that “these numbers don’t make sense,” but it is difficult to understand the problem with the current analytics process and where it lies in the pipeline.

To address this limitation, we propose an analytics framework whereby data and visualization pipelines are integrated with a knowledge base. A knowledge base is a tool designed to store clinical knowledge content in a systematic way by managing attributes of content authorship, version, and relationship between resources [21,22]. We used a knowledge base as a key component to store CPGs, their adherence definitions, and their executable queries, so that the related documents in their different forms could be managed with metadata and easily shared by domain experts, query developers, and analysts. In addition, we connected data pipelines to executable queries in the knowledge base so that a change of query can be immediately incorporated into the analytics pipelines.

In a pilot study, we adopted locally developed CPGs used in our health care system, resulting in a care process model (CPM). We developed an integrated analytics framework that consists of a knowledge base that employs CPMs in different forms: a commercial data pipelining tool, Alteryx, connected to our clinical database and a commercial visualization tool, Tableau, to generate reports. We built a dashboard that provides views for adherence by physicians to 21 CPMs. Over an 18-month-long proof-of-concept project, we ran a working group to analyze CPM adherence and to manage definition revisions. We investigated how the use of the knowledge base contributed to

**Methods**

**Subject: Care Process Models**

In this study, we used CPMs that were created using locally developed CPGs in our health care organization and that were designed to reduce clinical variation, improve quality, and support local preferences [23]. The CPMs were published and managed by clinical programs, which are made up of clinical expert groups consisting of clinical champions, medical directors, nursing administrators, data managers, and data analysts. Over the last 20 years, clinical programs have developed over 120 CPMs that cover a variety of clinical conditions and procedures, such as hypertension, heart failure, breast cancer, appendicitis, and acute myocardial infarction, among others.

CPMs were originally developed as paper or electronic documents containing descriptions of target problems or procedures, logics of decision-making, and recommended actions. Traditionally, CPM implementation was conducted through the involvement and education of care teams and providers. Over the last four years, we have installed a new enterprise-wide electronic health record (EHR) system in our hospitals and clinics in Utah, USA. We have also started developing computationally executable CPMs inside the EHR system using a variety of decision-support components, including order sets, decision-support rules, and care pathways (ie, decision flow and state-based order recommendation tools), among others.

In addition to CPM implementation, there was a strong need from the clinical leadership to monitor and evaluate how providers are complying with CPMs. Analyzing such data may allow us to understand how well-embedded CPMs are within clinical practices and may allow us to gain insights into how to improve best practices within them. However, it has been challenging to quantitatively measure whether CPMs were used as intended after implementation, since key data points for the evaluation are complicated to define and capture. In addition, time-consuming manual data processing to calculate evaluation metrics was required. To address these problems, clinical programs and informatics specialists initiated an effort to build a framework that creates a systematic approach for data extraction and visual analysis within the evolution cycle of CPM development, implementation, and improvement.

To correctly evaluate adherence to CPMs, three types of information should be defined. First, since CPMs are developed to treat patients with certain conditions, a target population (ie, patient cohort) should be identified. Specifically, a combination of inclusion and exclusion criteria should be defined, including patient demographics, diagnosis, lab results, and medications. Second, metrics to quantitatively measure CPM utilization for the defined target population should be defined. A timeline of when to develop the metrics, typically key concepts extracted from logics and actions in CPMs, should be included since it is unrealistic to capture all of the concepts within the CPMs.
Examples of key concepts include the following: diagnosed with pneumonia, image ordered for d-dimer, or 1-90 days old with a fever(≥38°C). Third, an adherence formulation is needed that would be used to count credits given to providers or care units who utilized CPMs as intended for target patients.

Although CPMs are well-described guidelines, it is often challenging to derive the key information above, as the nature of the CPMs are composed of complicated, domain-specific knowledge and often contain ambiguity. In addition, it is difficult to connect the derived key concepts to data points—often tables or columns in clinical databases—in real-world data sources. Thus, a multidisciplinary collaboration effort may be required in order to transfer knowledge between the various experts below:

1. Domain knowledge expert: an expert, author, or publisher of CPMs. This individual would be responsible for interpreting and defining the highest levels of evaluation criteria.
2. Domain data expert: an individual with both clinical domain and database expertise. This individual would be responsible for translating CPM knowledge and linking concepts to data points in databases.
3. Informatics expert: an expert in clinical knowledge management. This individual would be responsible for communication between domains.
4. Database engineer: a database expert. This individual would be responsible for developing and maintaining database tables for CPM relationship information.
5. Data analyst: a database expert. This individual would be responsible for developing data pipelines for analysis.

**Problem Statement**

The problem we are addressing is that knowledge gaps often exist while translating logics CPMs from domain experts into analytics pipelines. For example, domain experts may define the diagnosis of pneumonia at a high level in the original CPMs. However, the disease should be clearly defined to extract real data from the clinical database, including diagnosis codes in standard terminologies, types of patient visits, clinical context, problem status, and exceptions. Gaps may exist while clarifying such information by mapping incorrect data points, miscommunication, misinterpretation, changes in CPM contents, and changes in data sources. Such inconsistency could result in confusion and distrust of data quality for analytics consumers. In the current analytics environment there is no tool to track and investigate what the gaps are and where they exist.

To address inconsistencies in knowledge translation and improve analytics productivity, we aim to adopt a knowledge base that will help us manage the content across the whole CPM evaluation process. A knowledge base is a component in our EHR system that manages the authoring, review, publication, delivery, and versioning processes that surround clinical knowledge assets for consumption within these frameworks. Typically, clinical knowledge sources may be order sets, decision support rules, nursing protocols, clinical guidelines, and patient education resources. Any knowledge-based content can be a source for a knowledge base, such as concept definitions, formulations, executable queries, and design concepts of visualizations. These assets can be consumed by a human (ie, in the case of narrative clinical care guidelines) or by a computer (ie, rule bases for consumption by inference engines), or they can be aligned for consumption on both ends of the spectrum [21].

There are practical benefits to using a knowledge base for CPM analytics: (1) storing and sharing all the knowledge resources in a centralized place, instead of local storage or exchanging content by emails, to reduce miscommunication and redundancy; (2) being able to manage knowledge resources with unique numeric identifiers and metadata, including author, version, and history of revisions; and (3) making knowledge resources consumable with data analytics tools. In addition, a knowledge base may be useful for end-user data consumers by giving them more contextual information about the data with answers to questions such as "Who defined the definitions of the measures?", "How were the measures calculated?", and "What changes have been made?"

**Selection of Care Process Models and Deriving Key Measures**

As part of a pilot study to evaluate adherence based on the proposed framework, we selected 21 CPMs. Our selection criteria were as follows: (1) clinical utility and popularity of the CPM, (2) whether the CPM was already implemented within our EHR system, (3) the ease with which the CPM allowed the definition of evaluation metrics, and (4) whether data points related to the CPM were fully or partially collected in our clinical information systems. For the selected CPMs, we worked with CPM publishers in clinical programs to derive the three types of information for adherence evaluation defined in a previous section. The adherence rate (%), defined in the equation below, represents physician utilization of designated CPM components for a group of patients who are intended to be treated by the CPM:

Adherence (%) = (number of cases treated using the CPM) / (number of cases intended to be treated by the CPM)

Since the definition above is highly abstracted, there were diverse details regarding how to practically calculate the adherence. For example, to calculate the numerators, we determined what types of decision-support components were used to implement the CPMs (eg, order set) and which were used to connect key concepts to real data points (eg, order set ordering history). We found that 10 of the 21 CPMs (48%) were implemented as physician order sets, or small groups of order sets, whereas the rest were implemented through combinations of care pathways and decision-support rules.
### Table 1. Summary of care process model (CPM) adherence definitions.

<table>
<thead>
<tr>
<th>Clinical program (adherence definitions, n; queries, n)</th>
<th>CPM name</th>
<th>Cohort-defining condition</th>
<th>Decision-support component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oncology, 1, 3</td>
<td>Lung, breast, or colon cancer</td>
<td>Diagnosis and chemotherapy</td>
<td>Order set</td>
</tr>
<tr>
<td>Pediatrics, 1, 1</td>
<td>Febrile infant</td>
<td>Vital sign or a report</td>
<td>Care pathway</td>
</tr>
<tr>
<td>Neuroscience, 1, 1</td>
<td>Acute stroke</td>
<td>Problem and admission time</td>
<td>Order set</td>
</tr>
<tr>
<td>Primary care, 4, 4</td>
<td>Diabetes</td>
<td>Age and diagnosis</td>
<td>Care pathway</td>
</tr>
<tr>
<td>hypertension</td>
<td>Diagnosis</td>
<td>Care pathway</td>
<td></td>
</tr>
<tr>
<td>Acute sinusitis</td>
<td>Diagnosis and medication</td>
<td>Care pathway</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>Diagnosis</td>
<td>Care pathway</td>
<td></td>
</tr>
<tr>
<td>High blood pressure</td>
<td>Diagnosis</td>
<td>Care pathway</td>
<td></td>
</tr>
<tr>
<td>Cardiovascular, 1, 3</td>
<td>Heart failure</td>
<td>Age, admission, and diagnosis</td>
<td>Order set</td>
</tr>
<tr>
<td>Acute myocardial infarction</td>
<td>Age, admission, and diagnosis</td>
<td>Order set</td>
<td></td>
</tr>
<tr>
<td>Coronary artery bypass graft</td>
<td>Age, admission, and diagnosis</td>
<td>Order set</td>
<td></td>
</tr>
<tr>
<td>Musculoskeletal, 1, 1</td>
<td>Total hip or knee surgery</td>
<td>Procedure</td>
<td>Order set</td>
</tr>
<tr>
<td>Surgical service, 1, 2</td>
<td>Appendicitis</td>
<td>Procedure</td>
<td>Order set</td>
</tr>
<tr>
<td>Cholecystectomy</td>
<td>Procedure</td>
<td>Order set</td>
<td></td>
</tr>
<tr>
<td>Behavioral health, 3, 3</td>
<td>Depression</td>
<td>CDS rule and diagnosis</td>
<td>Care pathway</td>
</tr>
<tr>
<td>Suicide prevention</td>
<td>CDS rule</td>
<td>Care pathway</td>
<td></td>
</tr>
<tr>
<td>Mental health integration</td>
<td>Clinical document and clinic visit</td>
<td>Care pathway</td>
<td></td>
</tr>
<tr>
<td>Women and newborn, 1, 2</td>
<td>Jaundice</td>
<td>Newborn</td>
<td>Care pathway</td>
</tr>
<tr>
<td>Neonatal hypoglycemia</td>
<td>Observation, rules, and age</td>
<td>Care pathway</td>
<td></td>
</tr>
<tr>
<td>Intensive medicine, 4, 4</td>
<td>Pneumonia</td>
<td>CDS rule</td>
<td>Care pathway</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>Order, imaging, diagnosis, and care pathway use</td>
<td>Care pathway</td>
<td></td>
</tr>
<tr>
<td>Pediatric sepsis in emergency</td>
<td>CDS rule</td>
<td>Care pathway</td>
<td></td>
</tr>
<tr>
<td>Pediatric minor head trauma</td>
<td>CDS rule</td>
<td>Care pathway</td>
<td></td>
</tr>
</tbody>
</table>

*CDS: clinical decision support.*

The denominators were used to identify CPM-specific patient cohorts. For example, the denominator for the adherence rate of the hypoglycemia CPM was defined as a combination of (1) whether certain decision-support rules related to hypoglycemia were fired, (2) whether a patient has a history of hypoglycemia according to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), and (3) whether specific nursing documents were recorded.

Below are examples of detailed key concepts to be included in the definitions:

1. **Cohort (denominator):** clinical inclusion and exclusion criteria, patient master index, encounter, facility, care unit, decision-support rule firing criteria, clinical form used, etc.
2. **Utilization (numerator):** Order set usage, order set title, version, content, orderable items, customized order sets, care pathway title, components in care pathway used, etc.
3. **Adherence:** adherence definition, aggregation level by patient, encounter, provider, unit and facility, etc.

A group of experts with different backgrounds reviewed these concepts, documented the definitions, and built queries. **Table 1** shows a summary of the derived CPM adherence definitions.

---

**Results**

### Development of Framework

We developed the analytics framework by integrating several homegrown and commercial tools. For the knowledge base, we used Intermountain’s knowledge repository that our organization has used over the last 10 years. We adopted Alteryx as a tool for data analytics pipelines and Tableau as a tool for visual transformation and Web-based dashboard development. We used our enterprise data warehouse (EDW) as the data source, which is an integrated clinical data repository for research and quality improvement that stores over 100 billion records tied to encounters, lab observations, diagnoses, procedures, medications, and billing over 20 years. We also used CPM-centered database tables in our EDW that were developed by clinical programs, which store condition-specific patient cohorts, quality metrics, and outcomes for clinical research and quality improvement.

**Figure 1** depicts the architecture and flow of knowledge in the framework. Domain experts in clinical programs translate paper-based CPMs to adherence definition documents. Structured Query Language (SQL) developers create queries based on the documents and deploy them into the knowledge
base on the Web (see Figure 2). Alteryx imports the SQL information from the knowledge base to run against the EDW; Alteryx then exports the cohort and metrics as a Tableau-specific data file to a Tableau server so they can be visually transformed by Tableau to automatically generate charts on the Web. In the dashboard, links are embedded in the charts, enabling users to view the original CPM documents and queries used in the knowledge base.

Figure 3 shows a CPM in various forms during knowledge transformation: human-readable PDF document (top left); logics of the CPM are implemented as an order set in a computerized physician order entry (CPOE; bottom left); adherence definition document (top right); and computerized adherence logic as a SQL (bottom right). Since these knowledge content components in different forms originated from one source, they are semantically linked to each other with authorship, translation record, and version history.

Figure 1. Architecture diagram of the knowledge-based analytics framework for care process model (CPM) adherence. EDW: enterprise data warehouse; SQL: Structured Query Language.

Figure 2. Screenshot of the care process model (CPM) knowledge content components uploaded into the online knowledge repository.
**Figure 3.** Evolution of care process model (CPM) knowledge. Top left: an original PDF document; bottom left: a surgical order set inside a computerized physician order entry (CPOE) that implemented logics of the CPM; top right: a narrative text of evaluation criteria; bottom right: Structured Query Language (SQL) to extract adherence metrics. The red boxes and arrows represent the transformation of concepts and logics from one form to another. Source: Cerner PowerChart. Used with permission by Cerner Corporation.

### Development of Dashboard

We developed a Tableau-based dashboard in a production environment that is accessible to clinicians and researchers in our organization through secured user access. It employs five detailed views for each representative CPM adherence in different contexts and scales.

1. **Main view:** this view provides an overview of adherence trends at the highest level. It consists of a bar chart representing the average cumulative rates of CPM utilization and a line chart showing the monthly average over time (see Figure 4). Each CPM is marked with different color.

2. **Facility view:** this view is used to monitor the monthly average adherence rate by hospitals. This view consists of a dual chart that represents the percent adherence rate by month and the number of encounters where CPM was used or not used.

3. **Provider view:** this view shows the summary of CPM utilization by individual providers. Users can select a filter to the right of the bar graph narrows down the data by facility. Average adherence rates range from 0% to 100%, with 0% indicating that no physician used any CPM for the specific condition-based cohort and 100% indicating that all physicians used CPMs for all associated relevant cases.
provider in the filter on the right, then the view narrows down by patients treated by the selected provider (see Figure 5).

4. Encounter view: this view provides detailed actions in CPM utilization at the encounter level. A table view includes patient identifier, provider, length of stay, CPM used, and enrolled date and time.

5. Patient view: this view provides a summary of CPM-based treatments for a patient.

**Data Extraction and Early Usage Pattern**

Data pipelines were scheduled to run regularly, with different refresh frequencies depending on CPMs and data sources. As of July 1, 2018, the number of patients eligible for 21 CPMs was 230,669 and the number of encounters was 377,507. For those target patients, 7895 providers utilized CPMs at least once. Total adherence rate across all the CPMs during the period was 8.8%.

(Figure 4. Main view of the Tableau–based care process model (CPM) utilization dashboard (screenshot).

(Figure 5. Provider view of the care process model (CPM) utilization dashboard (screenshot).

Provider details: ALLEN, D.D., BRENT J. (Physician - Surgeon)

Summary of Care Providers: 83
Enrolled Encounter: 61
% Enrollment: 50.0%
Figure 6. Dashboard usage pattern showing number of daily sessions and number of distinct users. The blue bar represents the number of users that accessed the dashboard and the orange line represents the number of distinct users.

Table 2. Care process model (CPM) adherence revision history since dashboard implementation.

<table>
<thead>
<tr>
<th>Clinical program</th>
<th>Number of revisions</th>
<th>Type of revision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oncology</td>
<td>2</td>
<td>Amendment of adherence definition; new data source</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>3</td>
<td>Amendment of adherence definition; new data source</td>
</tr>
<tr>
<td>Neuroscience</td>
<td>2</td>
<td>Amendment of adherence definition</td>
</tr>
<tr>
<td>Primary care</td>
<td>2</td>
<td>Amendment of adherence definition</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>2</td>
<td>Amendment of adherence definition</td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>1</td>
<td>Amendment of adherence definition</td>
</tr>
<tr>
<td>Surgical service</td>
<td>1</td>
<td>Amendment of adherence definition</td>
</tr>
<tr>
<td>Behavioral health</td>
<td>4</td>
<td>Amendment of adherence definition; new data source; new facility</td>
</tr>
<tr>
<td>Women and newborn</td>
<td>2</td>
<td>Amendment of adherence definition; data-quality improvement</td>
</tr>
<tr>
<td>Intensive medicine</td>
<td>4</td>
<td>Amendment of adherence definition; new CPM component</td>
</tr>
</tbody>
</table>

We analyzed user sessions of the dashboard using a Tableau system monitoring report, as represented in Figure 6. Test and administration users were excluded from the analysis. Overall, the usage pattern was stable with minor seasonal effects and spikes.

Data Provenance: Tracking Revisions of Care Process Model Adherence Definitions

Since implementation, we set up monthly meetings involving clinical programs, CPM implementation leadership, data analysts, and knowledge engineers. These meetings were meant to (1) monitor CPM adherence data, (2) review current adherence definitions and discuss room for improvement, and (3) discuss ways to encourage providers to use CPMs. During the pilot study period, a number of revisions to the definitions were made (see Table 2). Many of the revisions included amendments to the definitions, while some included the addition of new data sources to the EDW or resolving data-quality issues. Definitions were updated with revisions in the knowledge base and the links to the Alteryx pipelines were automatically renewed.

As seen in the bottom portion of Figure 4, the adherence rate steadily increased since implementation. We believe there are two practical reasons for this increase. One is that clinical programs have encouraged their clinicians to utilize CPMs, through CPOE training, physician education, etc. The other reason is that some clinical programs revised their adherence definitions. For example, the oncology program expanded CPM-designed order sets, which resulted in an increase of the numerator. The intensive medicine program revised the definitions of their target patients by narrowing them down to specific facilities, which resulted in a decrease of the denominator.

Discussion

Comparison With Prior Work

Several studies have used knowledge engineering tools for translation of CPG logics to implement them into practice for the purpose of knowledge standardization and computational automation [4-9]. Unlike in those studies, we used knowledge engineering tools for managing knowledge transformation within analytics processes. Compared with prior work, our original results included (1) validating the usefulness of knowledge management tools within analytics processes, (2) validating the feasibility of integrating knowledge management tools and an analytics framework, and (3) demonstrating the proposed approach using empirical clinical data from local EHR systems.
Limitations
Although the main contribution of this study is the use of knowledge management, we did not quantitatively analyze the improvement in the consistency of knowledge in transformation or the productivity of analytics pipelines. Rather, we demonstrated it qualitatively, including determining which functionalities of the knowledge base were able to support the consistency of CPM-related knowledge in the development and maintenance phases. We will conduct further analyses in the future as we collect additional data. This will include adding more CPMs with complex clinical settings and practices, including chronic conditions and comorbidities that span multiple encounters, locations, and providers.

In this study, we simplified the adherence definitions to include whether a designated CPM component is used or not used for a target patient, although there may be many variations. We will continue to add more detailed definitions of adherence, including how specific order items or content within CPM components are used. By doing so, we can investigate the mechanism of CPM utilization (ie, use of standard clinical guidelines) and how this can change patient care or clinical workflow.

Conclusions
This case study demonstrated that the proposed analytics framework could accommodate complicated knowledge management and data pipelining for CPM evaluation using a knowledge base, while maintaining computational efficiency. It is expected that the benefits of using a knowledge base will be more significant as we add complicated clinical guidelines into the analytics framework.

Authors' Contributions
JL developed the framework, built the dashboard, performed data analysis, and designed the study. NCH helped design the study, provided knowledge base consultation, collaborated with clinicians, and validated the feasibility of the architecture and the results of data analysis. The authors do not have any funding or financial support associated with this project.

Conflicts of Interest
None declared.

References


Abbreviations

- CDS: clinical decision support
- CPG: clinical practice guideline
- CPM: care process model
- CPOE: computerized physician order entry
- EDW: enterprise data warehouse
- EHR: electronic health record
- ICD-10: 10th revision of the International Statistical Classification of Diseases and Related Health Problems
- SQL: Structured Query Language

©Jaehoon Lee, Nathan C Hulse. Originally published in JMIR Biomedical Engineering (http://biomedeng.jmir.org), 27.02.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information is available at http://biomedeng.jmir.org/2019/1/e11659/.
information, a link to the original publication on http://biomedeng.jmir.org/, as well as this copyright and license information must be included.
Perspectives of Orthopedic Surgeons on the Clinical Use of Bioprinted Cartilage: Qualitative Study

Ángels Salvador Verges1, MD; Luis Fernández-Luque2, PhD; Meltem Yildirim3, PhD; Bertran Salvador-Mata4, BTech; Francesc Garcia Cuyàs5, MD, PhD

1Digital Care Research Group, Universitat de Vic – Universitat Central de Catalunya, Barcelona, Spain
2Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
3Research Group on Methodology, Methods, Models and Outcomes of Health and Social Sciences, Centre for Health and Social Care Research, Faculty of Health Science and Welfare, Universitat de Vic – Universitat Central de Catalunya, Barcelona, Spain
4Scientific Communication, Universitat Pompeu Fabra, Barcelona, Spain
5ICT and Health, Catalan Society of Digital Health, Universitat de Vic – Universitat Central de Catalunya, Barcelona, Spain

Corresponding Author:
Ángels Salvador Verges, MD
Digital Care Research Group
Universitat de Vic – Universitat Central de Catalunya
11, Magí Casanovas
Barcelona, 08870
Spain
Phone: 34 630887995
Email: angels.salvador@uvic.cat

Abstract

Background: Over the past 60 years, no technique used for treating cartilage disorders has been completely successful. Bioprinting provides a highly anticipated, novel alternative solution to this problem. However, identifying barriers to this new technology is crucial in order to overcome them when bioprinting reaches the implementation stage. This kind of research has been declared essential because clinical efficacy and safety studies alone do not always lead to successful implementation.

Objective: This qualitative study aimed to explore the stance of orthopedic surgeons on the use of bioprinted cartilage grafts for cartilaginous lesions. The study sought to summarize and classify the barriers and facilitators of this technique and to identify the key factors for successful implementation of bioprinted cartilage in routine clinical practice.

Methods: A qualitative thematic analysis method was used to evaluate data obtained from semistructured interviews and from focus groups. Data were collected between June 2017 and February 2018. Interviews focused on the collection of expert opinions on bioprinted cartilage.

Results: The perceived barriers to the adoption of this technology were (1) awareness of a lack of information on the status and possibilities of this technology, (2) uncertainty regarding compliance with current health care regulations and policies, and (3) demands for clinical evidence. The facilitators were (1) lack of surgical alternatives, (2) the perception that research is the basis of the current health system, and (3) the hope of offering a better quality of life to patients.

Conclusions: The results of this study are preliminary in nature and cannot be generalized without a broader group of participants. However, the key factors identified provide a frame of reference to help understand the challenges of bioprinted cartilage and help facilitate the transition toward its clinical use. These findings will also provide information for use at multidisciplinary meetings in scientific societies; create bridges between researchers, orthopedic surgeons, and regulators; and open a debate on the funding of this technique and the business model that needs to be developed.


KEYWORDS
bioprinting; orthopedic surgeons; qualitative research; cartilage; expert testimony
Introduction

Background

Traumatic cartilage lesions and arthritis are two of the most prevalent chronic diseases worldwide. According to data from the Global Burden of Disease Study [1], the number of people suffering from disorders caused by such diseases has increased from 140 million in 1990 to 242 million in 2013. Cartilage is a highly hydrated and specialized tissue providing a low-friction surface and resistance to erosion and diarthrodial joint load, allowing for effective articular movements. Unfortunately, the function and structure of cartilage are often damaged by trauma or ageing, adding to the fact that cartilage has a low capacity to self-heal.

Treatment of these diseases is still a challenge, and an effective solution remains to be found. These defects or lesions can last for years and can lead to arthritis [2]. Current repair techniques for cartilage lesions can be divided into two main groups: bone marrow stimulation and transplantation techniques [3,4].

The potential of regenerative medicine and tissue engineering is now recognized worldwide. These new techniques are responsible for “shifting the paradigm in health care from symptomatic treatment in the 20th century to curative treatment in the 21st century” [5-7]. Currently, three-dimensional (3D) printing is used for several applications in the medical field, for example, in the printing of patient-specific osteotomy guides. Other surgical specialties use 3D printing to study the disease pathology in a patient and practice with a 3D-printed model before surgery [8].

Bioprinting refers to the use of 3D printing to combine cells, growth factors, and biomaterials to create tissues and organs mimicking the features of their natural counterparts [9]. Bioprinting generally uses the extrusion-based method, which consists of the layer-by-layer deposition of cells through bio-ink, creating a structure similar to the natural tissue that can be used in tissue engineering and medicine. Bioprinting, which emerged in 2004 with the use of additive manufacturing, combines cells, gels, and several biocompatible elements in a single scaffold, which can replace injured tissue with a complex structure that contains several components, including structural and cellular constituents. The external shape and internal architecture can be modeled based on clinical images. Ideally, cartilage creations aimed to fill cartilage defects should be similar to the extracellular matrix to keep cells in their place and preserve a space for the tissue that will grow there [10]. Levato et al. published that although the most suitable types of cells for bioprinting are well known, more research needs to be conducted regarding zonal organization of cartilage [11]. There is also the need to study the complex mechanical behavior of cartilage under compression, as a result of sliding and shear [12].

However, barriers and challenges for implementing a new technology must not be underestimated, and it is essential that they are addressed in advance to guarantee the widespread application of bioprinting once it has reached its maturity. Research focused on this has been highlighted as crucial, since clinical efficiency and safety do not always lead to successful implementation. A recent editorial [13] encouraged implementation research at the beginning of development.

Objectives

This qualitative study aimed to explore the stance of orthopedic surgeons on the use of bioprinted cartilage grafts for cartilaginous lesions. We sought to classify the barriers and facilitators of this new technology and identify key factors that need to be considered for successfully implementation of bioprinted cartilage in routine clinical practice.

Methods

Study Design

The applied design consisted of a hybrid inductive and deductive thematic analysis, which allowed for interpreting gross data extracted from in-depth, semistructured interviews with orthopedic surgeons. This methodology was chosen to best reflect the perspective of interviewees.

Inclusion Criteria

For inclusion, orthopedic surgeons had to (1) have more than 5 years of experience in the field, (2) be currently working in a hospital, and (3) be actively performing surgery. No contacted participants were excluded.

Ethical Compliance

All participants were volunteers and agreed to participate in the interview or focus group. All were provided with the Ethics Research Committee document and signed the informed consent form. Authorization by the University Research Committee of University of Vic – Central of Catalonia (Spain) was granted (record number 28/2017).

All participants were informed in advance about the nature of the project, risks, advantages, and alternatives and their rights as research subjects. Measures were taken to ensure the data collected remained confidential; participants’ safety and privacy were protected during and after the study.

Participant Selection

All participants were contacted via email, signed the informed consent form, and authorized the recording of the interview. Participants consisted of 18 orthopedic surgeons, ages 35-67. The group included 15 men and 3 women. Six were heads of units and 12 were specialists.

Interview Structure for Data Generation

In 2001, Patton [14] created a list of 6 question types that could be formulated based on behavior or experience, opinion or values, feelings, knowledge, and perception; those questions aimed to obtain demographic or background data. Our guideline included the following themes: 3D printing, bioprinted cartilage, cell origin, current needs, rejection, expectations, and suggestions.
The interviews began by exploring the participants’ knowledge regarding the medical applications of 3D printing and bioprinted cartilage. Questions on the use of stem cells were an important element of the interview, as much research is currently being conducted on mesenchymal cells obtained from umbilical cord tissue, adipose tissue, and bone marrow. The use of induced pluripotent stem (iPS) cells was also explored.

Questions related to current needs were aimed to corroborate the lack of efficiency in existing surgical techniques and the importance of research to find new practices. Questions on the expectations and reluctance of surgeons regarding the use of bioprinted cartilage sought to understand reasons for and against usage if the opportunity arose. The final section of the interview allowed for them to analyze the current situation and talk about future possibilities.

The interviews were always done with the same system (see the flowchart in Figure 1).

**Recording the Interviews**

The interviews and focus group discussions were conducted between June 2017 and February 2018. To protect the identity of all participants, each participant was codified to a randomly generated number that was then used in all study documentation. Their information was kept in a password-protected virtual folder of the university. Interviews were recorded digitally and transferred to the computer, where they were saved with the interviewee number and date of the recording. Informed consent forms were also stored at the university.

A single interview was conducted for each of the 18 participants, with the introduction providing context for the interview. The shortest and longest interviews were 25 minutes, 7 seconds and 43 minutes, 11 seconds, respectively. In total, we recorded 10 hours and 18 minutes of interviews.

Most interviews were conducted in the workplace of the interviewee, except for 4 participants who chose to have the interview in a coffee shop.

**Recording the Focus Groups**

Two focus groups were put together, and participant privacy was guaranteed in the same way as for the interviews. The first group consisted of 8 people, and the recording lasted 45 minutes, 23 seconds. The second group consisted of 5 people, and the recording lasted 74 minutes, 49 seconds.

**Global Data Analysis**

To ensure thematic integrity, this study used only data obtained from orthopedic surgeons. This analysis aimed to generate a list of relevant concepts that could be extrapolated and categorized. This leads to an inductive approach where topics have been identified by contextual information.

The integrity of the analysis was ensured by the directives established by Shenton [15], which included iterative questioning in data collection dialogues and the construction of an “audit trial,” among others. The iterative process of grouping and subgrouping questions and answers led to a series of abstract constructions that were used to create a model to understand the context.

**Inductive and Deductive Analysis of the Data**

The analysis used mixed elements of inductive and deductive methods to interpret the gross data [16] and explore the attitude and experiences of the orthopedic surgeons interviewed. The flexibility of the approach helped analyze qualitative data from the interviews. The approach was useful in this study due to its large quantity of data.

Codification was reached through discussion and consent. Three researchers continued their discussion until consensus was reached regarding categorization and subcategorization of topics. Braun and Clarke’s [17] methodology, which identifies, analyzes, and describes reporting patterns, was used as a basis for thematic analysis. Due to the exploratory approach, this practical method was thought to fit the needs of the study perfectly. The process of thematic analysis is developed through 6 phases [18]. Phase 1 is becoming familiar with the data; Phase 2 is generating initial codes; Phase 3 is searching for themes and depuration of codes; Phase 4 is reviewing themes and finding those that are important either for reiteration or relevance to the research question; Phase 5 is defining and naming themes; and Phase 6 is producing the report.

Issues regarding trustworthiness were approached as described by Shenton [15], who provides a description of research, collection, and analysis design. The strategies used to ensure
honesty in the interviews include encouragement to be candid and the assurance of the voluntary nature of the interview and the right to withdraw at any chosen moment. Transferability was accomplished by providing contextual in-depth information on the study and the role of the researcher. The researcher guaranteed confidentiality [19]. For data analysis and figure generation, ATLAS.ti version 8.2.34 was used.

Results

The analysis of the interviews and focus groups is presented in Table 1 with the aim of describing the current stance of orthopedic surgeons on cartilage grafting. The Table 1 list is what ATLAS.ti denominates as a “frequency count.” It represents the number of times these concepts were identified in the texts. Each point was given a code denoting different levels of classification and abstraction, which were later linked to the established categories.

Within the discursive pattern of clinicians, two argumentative groups were identified, which were classified as facilitators and barriers. From these two groups of codes, the most relevant were selected to establish the key factors that will provide a general perspective on the stance of orthopedic surgeons.

Table 1. Collected concepts and translation to numbers and codes.

<table>
<thead>
<tr>
<th>Concepts</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need for clinical trials</td>
<td>34</td>
</tr>
<tr>
<td>Implantation techniques</td>
<td>28</td>
</tr>
<tr>
<td>Viability and traceability of the graft</td>
<td>27</td>
</tr>
<tr>
<td>Characteristics of the cartilaginous tissue</td>
<td>25</td>
</tr>
<tr>
<td>Small lesions (focal defects and osteochondritis)</td>
<td>24</td>
</tr>
<tr>
<td>Durability</td>
<td>14</td>
</tr>
<tr>
<td>Safety</td>
<td>14</td>
</tr>
<tr>
<td>It’s the future</td>
<td>13</td>
</tr>
<tr>
<td>Costs</td>
<td>13</td>
</tr>
<tr>
<td>Cell therapies</td>
<td>12</td>
</tr>
<tr>
<td>Need to wait for clinical results</td>
<td>12</td>
</tr>
<tr>
<td>Cell types</td>
<td>12</td>
</tr>
<tr>
<td>Current techniques</td>
<td>11</td>
</tr>
<tr>
<td>Uncertainty regarding the future</td>
<td>10</td>
</tr>
<tr>
<td>Regulation</td>
<td>10</td>
</tr>
<tr>
<td>Stem cells</td>
<td>8</td>
</tr>
<tr>
<td>Technical difficulties in some articulations</td>
<td>8</td>
</tr>
<tr>
<td>Cell viability</td>
<td>8</td>
</tr>
<tr>
<td>Teratogenesis</td>
<td>7</td>
</tr>
<tr>
<td>Biocompatibility</td>
<td>7</td>
</tr>
</tbody>
</table>

Barriers

The barriers consist of the arguments and opinions put forward by the orthopedic surgeons that reflect the perceived challenges or the lack of information with regards to adopting the technology. Figure 2 shows the links between several components generated by ATLAS.ti. It shows the groupings and connections between the codes.

Lack of Information

The first barrier to be identified, which was coded as “lack of information,” had an impact on the following aspects.

Cell Therapy

Orthopedic surgeons admitted their lack of knowledge regarding the acquisition and cell origin of chondrocytes. The participants showed great reluctance regarding the origin of the cells. Furthermore, if cell therapy implied the manipulation of unknown-origin cells, their stance was of total rejection. However, the level of acceptance was considerably higher if they knew the cells originated from the patient, even if they knew that they had to be manipulated (numbers in parentheses after quotes correspond to interviewee identity).

*These constructors imply cell manipulation.* [#D 13]

We don’t know if bone marrow or adipocyte is better, it seems like bone marrow could be useful, but it’s not so clear. [#23]
The better-known stem cell origins were mesenchymal, adipose, umbilical cord, and bone marrow as they are currently being used in other types of therapies.

*I think mesenchymal cells are the way to go.* [23]

There is a huge quantity of umbilical cord stored at the blood and tissue bank...Of umbilical cord cells, of adipose cells, cells of peripheral blood, from the skin; we can obtain cells from many places. [5181]

However, when trying to get the participants to discuss the topic in more detail, they appeared confused, especially when discussing iPS cells, which they were not aware of or did not fully understand.

The safety of iPS is not clear; there is an infinite number of complications—you can ask Yamanaka or Arnold Caplan. [23]

### Patient Safety

All interviewees raised concerns regarding issues related to patient security, the graft, and associated diseases. With regards to the patient and the graft, the concerns focused on teratogenesis and the genetic predisposition of the cells in the graft, as well as the long-term behavior of the graft. Participants also showed a concern regarding graft implantation in patients with severe associated diseases, even though this fear does not have scientific merit.

*Two types of safety: Safety for the patients’ lives, of course, and safety in knowing that the graft will grow into cartilage, that you know for certain that this thing will create cartilage.* [2341]

*Three certainties: One, that these cells behave as we expect them to behave, like cartilaginous cells with no marginalization at all; two, that these cells are viable in the long term; and third, durability. If I am to implant cartilage, I’ll want it to last.* [204]

### Difficulties With the Surgical Technique

Participants anticipated difficulties with regards to the shape of the graft, as they were unsure if the printing process could comply with the exact measurements provided by doctors. They also cited the place of injury as a possible difficulty.

*Again, there’s the problem of the three-dimensional structure of the cartilage.* [2341]

*Not all places are the same. For example, the knee: I think it’s viable to insert it on the articular surface of the tibial plateau...Another thing is how it would anchor to the bone, right? But technically I don’t see a difficulty here. Now then, it’s another thing to insert it on the hip bone, between the cotyloid cavity and the femoral head.* [86]

The lack of knowledge regarding the shape and manipulation of the graft, together with its characteristics led to a third kind of uncertainty, which we see as a barrier related to the surgical difficulties. Similarly, not being able to visualize the graft as part of cartilage that would adapt to the host left participants doubtful as to whether the graft would be able to anchor itself and stay in place.

*I would use it now, for young people with osteochondritis of the talus or the knee, where you have a two- or three-millimeter.* [5690]
The size is a factor with the cartilage will it stay in place? [#2901]
How will you fix it there? How? How does it stay there? [#2341]
Another thing I worry about is that this tissue that we insert stays anchored. [#6]

**Graft Characteristics**

Regarding graft characteristics, the main factors mentioned were viability, durability, integration with the host, and mechanical characteristics of cartilage. For example, participants doubted that the graft would become functional cartilage or develop chondrogenic hypertrophy, which is what happens with current techniques such as the matrix-induced autologous chondrocyte implantation (MACI) procedure.

Needs to have all the characteristics of the original cartilage. [#6]
We need proof that long term there will still be cartilage and not fibrous tissue. [#2901]
Needs to behave biologically like the host’s cartilage. We’re talking about live cells, right? [#204]

Orthopedic surgeons also questioned the viability and durability of the graft. They were unaware that the aim of the graft is to become integrated to the native cartilage and that the graft’s behavior would mimic the patient’s native cartilage.

How long will it last? [#6099]
And the viability of these cells, and their possible side effects. [#1753]
Then, what I understand that these cells are viable themselves, it’s not that they need to be invaded by the periphery, but that they are viable and live by themselves. [#5181]

**Health Policies and Regulation**

Bioprinting, like any other product of tissue engineering will have to comply with the current Good Manufacturing Practice regulations enforced by the Food and Drug Administration and/or the European Medicines Agency.

The clinicians stated that they felt there would be a timeframe in which health policies will not be able to provide an effective answer to their questions, which would be problematic for all practitioners using them.

Who will guarantee the manufacturing process until its arrival in the operating room? [#2083]
It’s a legal aspect within the framework of drugs, implants, of techniques. We need to see this technique, legally the European guidelines on the use of tissues and cells. [#4821]

**Need for Clinical Evidence**

As with any scientific innovation, orthopedic surgeons demanded hard clinical evidence be available before they would use bioprinted cartilage. In most cases, this demand materializes as clinical trials and independent clinical research. However, this may be a barrier since clinical trials are not scheduled to take place in the imminent future.

I want more evidence, that is to say, scientific studies that support their efficacy; independent scientific studies. [#2901]
Basically, that there are appropriate clinical trials. [#204]

**Facilitators**

Facilitators include all entities that encouraged orthopedic surgeons to be more open to new surgical possibilities to improve the lives of patients. In this group, three themes were identified that were essential to the clinicians to catalyze implementation of cartilage grafts (Figure 3).
Clinical Need

Given that current surgical techniques are not able to provide a definitive solution, orthopedic surgeons are open to innovative techniques that can fill a surgical need. However, as they were not aware of the possibilities bioprinting would bring to field, most of the interviewees stated that research into finding new solutions was critical. Several arguments emphasize this need:

*Especially as we don’t have anything that works for these patients right now... if it lasts, say, 40 years, I’d say it’s marvelous. [#2901]*

*Nowadays this is something that doesn’t have any solution, so of all the things I’ve heard about maybe this very innovative technique works since no other offer is effective. [#2083]*

*If this works, it’s very promising. [#23]*

In this way, the main facilitator identified was the lack of current alternatives, as was to be expected.

Taking this into consideration, orthopedic surgeons, despite the barriers mentioned, are open to this new technology.

We also identified the type of patients that surgeons would be willing to consider treating with bioprinted cartilage implementations. Younger patients were perceived to be better candidates, as they are susceptible to high-risk sports injuries, which often become chronic and difficult to solve in the long term.

*That is, with young people with partial cartilage lesions, I see it very clearly; with bigger lesions, I am less optimistic.* [#23]

*I see it as a solution to young people’s pathologies, athletes, that have damage due to chondral lesions and which can happen at any age, but they hinder young people’s activity.* [#6]

Additionally, we identified the specific characteristic within this population that significantly improved the acceptance of bioprinting technology, namely the size of the lesion. Orthopedic surgeons were distinctly in favor of using grafts in small lesions (1 or 2 cm at most), in order to accelerate integration with the host. However, they did not show the same certainty with larger lesions.

*I can see it being used with partial cartilage lesions.* [#2901]

*If the lesions are small, and the joints are not loaded.* [#6356]

Perception of Bioprinting as a Future Treatment

Clinicians agreed that medicine depends on constant research to find solutions to unresolved problems. In other words, they perceived scientific research as a positive entity. Another argument identified the perception of bioprinting as a future solution, with participants being optimistic about graft bioprinting.

*It is a future solution to important problems for orthopedic surgeons.* [#6346]

*When we talk about the medicine of the future, which is not so far away now, to be able to reproduce the tissue of the patient.* [#204]

Expectations

Faced with an unresolved clinical need and the perception that bioprinting could be a solution in the future, orthopedic surgeons expect to hear about the benefits of this technique. Biocompatibility is not only a favorable factor but is essential to surgical practice. Many current techniques already have these characteristics, and therefore clinicians demand that future solutions meet or exceed these standards.

*To find a three-dimensional structure that holds the cells, that holds what they must have, and that this three-dimensional structure is biocompatible, degradable, and easy to manipulate.* [#2341]

At the same time, participants emphasized that this technique had the ability to improve patients’ quality of life significantly, either by alleviating their pain, improving their mobility, or by preventing lesions from developing into arthritis in younger patients. If these were to be accomplished with the new technology, surgeons expect the need for total prosthesis to diminish significantly.

Key Factors

Where barriers and facilitators meet, key factors emerge. Key factors function as the theoretical framework for the perspectives of orthopedic surgeons on bioprinted cartilage. In general, their belief is grounded on clinical need and expectations for effective solutions. Despite this, a reluctance to adopt the technology was detected among the interviewees, with reasons ranging from (conscious and unconscious) lack of information to clinical demands. Figure 4 offers a complex concept map, which is the first attempt to represent the stance of clinicians with an aim to help direct future research.

Apart from the elements present in both barriers and facilitators, two more factors were considered key factors and coded as such. They included costs and the identity of companies that would manage the product. It is impossible to address these uncertainties now; hence, they could not be labeled as either facilitators or barriers, only as relevant factors that need to be addressed.
Discussion

Principal Considerations

As has been proven in a previous bibliographical review [20,21], the research and acquisition of bioprinted cartilage is still in a premature state. Other researchers have already highlighted that despite the growing number of solutions coming from tissue engineering that are being transitioned to clinical use, the success of considerably sized scaffolding with personalized geometries is still a significant challenge. Therapies based on mesenchymal stem cells (MSC), despite having been successful in renovating the cartilage and alleviating pain, have not provided enough evidence on original hyaline cartilage restoration that would improve osteoarthritis in the long term.

The goal of this research is to understand how trauma surgeons perceive this situation and define main barriers and facilitators to develop strategies favoring the future implementation of bioprinted cartilage. The data collected and organized into either barriers or facilitators as detailed in our results will help future discussions focus on the most fundamental aspects of this technology.

One of the main needs identified is improving communication with orthopedic surgeons, particularly regarding 3D printing. The lack of knowledge was evident in two ways: conscious and unconscious. Of the two, the latter will be more difficult to address, as it requires further research to better identify the knowledge gaps. When conscious of their lack of knowledge, clinicians have no problem asking questions. However, the lack of knowledge was unconsciously displayed when assumptions were made regarding terms, techniques, or solutions leading to misinterpretation and confusion. To lead and conduct successful translational research, it is necessary to study and solve problems transversally. An unconscious deficit of knowledge was driven by reading publications or listening to conversations that dealt with cell therapy in a generic and unscientific manner. Thus, the lack of background knowledge was significant, resulting in misperceptions and unfounded reluctance in adopting the technology.

Furthermore, we discovered a lack of knowledge on current applications of 3D printing in medicine, with many of the interviewees having no knowledge of this facet. To address this issue, organization of specific multidisciplinary seminars to discuss the current applications of 3D printing in medicine should be undertaken. This could contribute to orthopedic surgeons becoming more proactive in the implementation of bioprinted cartilage grafts.

In parallel, efforts should be made to help promote an understanding of the fundamentals of cell therapy. This issue was identified as an unconscious knowledge gap. This was also observed in the case of gene therapy for cancer treatment [22]; participants were aware of the treatment but had no deep understanding of it. This was evident from the fact that they used cartilaginous cells provided by laboratories [23] but were reluctant to consider using bioprinted cartilage made of unknown-origin cells, or other cell therapies.

These findings highlight the evident need to develop formative strategies. These strategies would need to be based initially on the fundamentals of cell therapy, escalating toward the future possibilities this technology could offer. Thus, new channels of communication could be created in the medical community.

Figure 4. Where barriers and facilitators meet, key factors emerge.
While it is important for surgeons to have basic knowledge regarding the future applications of cartilage grafting, it is equally important for researchers to recognize and acknowledge the practical needs of clinicians and strive to meet their expectations. Some of the factors that caused the orthopedic surgeons concern included issues such as viability, traceability, and durability. Tissue and skin banks for allografts have established their reliability by ensuring traceability and establishing manipulation standards. Orthopedic surgeons now demand the same degree of reliability from bioprinted cartilage.

In addition to the characteristics of the graft, this study has identified important factors that would help direct research in the initial stages. First, by focusing research on specific lesions, such as 1 or 2 cm lesions found in the knee or the ankle, clinicians would have access to a site that is easier to access and operate on. Second, younger patients should be established as the primary recipients of the graft, with the aim of avoiding long-term joint deterioration.

Another issue detected during this study was concern regarding the business model for producing bioprinted grafts. Orthopedic surgeons feared it may not align with current production models. This situation, coded as a key factor, is one of the main issues identified as a barrier in the implementation of this technique.

By analyzing the stance of orthopedic surgeons, at least two possible lines of action can be suggested. If production was handled by private companies, the main demands from surgeons would be for the pieces to be individually customized, with a short production-delivery timeframe. In this instance, the biosafety and tissue traceability could be controlled. Another issue that would need addressing is the cost of the graft. This model would also need to address the patent issue and comply with the ethical requirements and, more importantly, with the current regulations and legislation. The Spanish company Regemat is an example of this. They use Hoffa’s fat pad and chondrocytes as described by Lopez Ruiz [25] and induced differentiation of autologous MSC to develop and commercialize cartilage.

Another possibility would be to establish public centers, possibly in a public-private collaboration model, that would have the human and technical resources necessary to produce their own 3D bioprinted pieces. The foundations for such a model already exist in hospitals where 3D printers are already in use. As mentioned above, these hospitals have trained staff who are already competent in the use of 3D printers for a wide variety of uses ranging from the printing of fractures, surgical planning, and creating customized guides for the patient. This system, which would be integrated into hospitals, would allow for constant communication between the medical and technical teams. The hospital as a meeting point ensures that as the technology becomes widely used, more potential applications will be detected, thereby improving the learning curve for both sides—the medical team exploring new and better applications for the technique and the engineering team designing context-specific solutions. This solution would mean bioprinting is the next logical step, born from the growing needs of all medical specialties.

Cell therapy has stirred a debate within the scientific community. Cell therapy can be individually customized, is expensive and innovative, and might help bring a change in health regulation and health care policies. Our research has shown that the demands for scientific evidence for bioprinting will be more stringent than what was required for previous techniques. This is the case of platelet-rich plasma, which has been used by doctors for more than 20 years despite the lack of evidence for its effectiveness [26], with information on clinical trial outcomes having only recently been published [27].

Communication, not only among medical professionals, but among policy makers and health care authorities, is essential to start a debate to define the level and form of evidence required. In this manner, one of the main barriers highlighted by orthopedic surgeons, namely the need for clinical trials, could be surmounted.

**Limitations**

This study needs to be interpreted in the context of its limitations. There are inherent limitations to the number of participants and the number of focus groups. Only the data extracted from the orthopedic surgeons’ interactions is legitimate; however, it is their opinion that focuses the research in this context.

**Conclusions**

These study results are preliminary in nature and therefore they cannot be generalized without a broader demographic. However, the preliminary literature review confirms the lack of research on clinical applications of bioprinted cartilage. Orthopedic surgeons are willing to accept that this new technology has the potential to solve a clinical need and to recognize bioprinting as the technology of the future. However, clear scientific evidence is required before bioprinted cartilage can be used and a debate regarding the optimal business model will be necessary.

We also believe it is necessary to develop a communication strategy and a forum for multidisciplinary discussion to discuss the need for regulation and define the necessary scientific evidence that is required to promote the acceptance of grafts as a viable therapeutic option. From our perspective, this study serves as a first step in the clinical translation of bioprinting cartilage research.

**Acknowledgments**

We would like to thank all participants in the interviews and focus group; Professor Juan Muñoz from the Social Psychology Department of the University Autonomy of Barcelona for his help with the methodology and the analysis; and Marina Mustieles for her help in the translation of this text.
Authors' Contributions
The first author completed the field work and the analysis of the data obtained. The other authors contributed to the content, writing, and editing of the manuscript. All the authors read and approved the final version of the manuscript.

Conflicts of Interest
None declared.

References
14. Salvador Verges et al. JMIR Biomedical Engineering 2019 | vol. 6 | iss. 1 | e12148 | p.39 http://biomedeng.jmir.org/2019/1/e12148/ (page number not for citation purposes) [FREE Full text] [doi: 10.4012/18-0018] [Medline: 29950870]


Abbreviations

iPS: induced pluripotent stem
MSC: mesenchymal stem cells

©Àngels Salvador Verges, Luis Fernández-Luque, Meltem Yildirim, Bertran Salvador-Mata, Francesc Garcia Cuyàs. Originally published in JMIR Biomedical Engineering (http://biomedeng.jmir.org), 28.02.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on http://biomedeng.jmir.org/, as well as this copyright and license information must be included.
The Impact of Aging and Hand Dominance on the Passive Wrist Stiffness of Squash Players: Pilot Study

Taya Hamilton¹, PT, GDNR; Stan Durand¹,², MSc, MEng; Hermano Igo Krebs¹,³, PhD

¹The 77 Lab, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, United States
²Institut de Biomécanique Humaine, Georges Charpak Arts et Métiers, ParisTech, Paris, France
³Division of Rehabilitation Medicine, Department of Neurology, School of Medicine, University of Maryland, Baltimore, MD, United States

Corresponding Author:
Taya Hamilton, PT, GDNR
The 77 Lab
Department of Mechanical Engineering
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA, 02139
United States
Phone: 1 857 204 7456
Email: tayavn1@gmail.com

Abstract

Background: Passive joint stiffness can influence the risk of injury and the ability to participate in sports and activities of daily living. However, little is known about how passive joint stiffness changes over time with intensive repetitive exercise, particularly when performing unilateral activities using the dominant upper limb.

Objective: This study aimed to investigate the difference in passive wrist quasi-stiffness between the dominant and nondominant upper limb of competitive squash players, compare these results with a previous study on young unskilled subjects, and explore the impact of aging on wrist stiffness.

Methods: A total of 7 healthy, right-side dominant male competitive squash players were recruited and examined using the Massachusetts Institute of Technology Wrist-Robot. Subjects were aged between 24 and 72 years (mean 43.7, SD 16.57) and had a mean of 20.6 years of squash playing experience (range 10-53 years, SD 13.85). Torque and displacement data were processed and applied to 2 different estimation methods, the fitting ellipse and the multiple regression method, to obtain wrist stiffness magnitude and orientation.

Results: Young squash players (mean 30.75, SD 8.06 years) demonstrated a stiffer dominant wrist, with an average ratio of 1.51, compared with an average ratio of 1.18 in young unskilled subjects. The older squash players (mean 64.67, SD 6.35 years) revealed an average ratio of 0.86 (ie, the nondominant wrist was stiffer than the dominant wrist). There was a statistically significant difference between the magnitude of passive quasi-stiffness between the dominant and nondominant wrist of the young and older squash player groups (P=.004).

Conclusions: Findings from this pilot study are novel and contribute to our understanding of the likely long-term effect of highly intensive, unilateral sports on wrist quasi-stiffness and the aging process: adults who participate in repetitive sporting exercise may experience greater joint quasi-stiffness when they are younger than 45 years and more flexibility when they are older than 60 years.


KEYWORDS
wrist; exercise; aging
Introduction

Background
Joint stiffness is a biomechanical feature of human anatomy that is both essential and potentially detrimental to participation in sports, leisure activities such as music, and activities of daily living. Reduced joint stiffness may increase the risk of injury because of poor stabilization of joints and an inability to maintain joint postures [1,2]. Increased joint stiffness has been associated with reduced joint range of motion (ROM) and inflexibility, such as during the aging process [3,4], following surgical intervention [5], or because of abnormal muscle tone secondary to neurological injury or disease [6,7], which can cause both injury and functional limitations.

The term joint stiffness has various definitions depending on the discipline and nature of the research. In physical education, sports medicine, and allied health disciplines, joint stiffness is more commonly referred to as the flexibility, or ROM, of a joint or group of joints [8,9]. In the biomechanical literature, joint stiffness is referred to as the ratio of the change in joint torque to the change in joint angle [10]. Regardless of the definition of the term, joint stiffness is understood to be multidimensional, made up of (1) the elastic properties of noncontractile tissue (tendons, ligaments, and the joint capsule), (2) the elastic properties of intrinsic muscle cross-bridges, and (3) the reflex action of a muscle following change in length [11]. The different methods of assessment aim to distinguish between these components of joint stiffness, altering both the primary tissue structure under evaluation and the magnitude of joint stiffness reported. This study focuses on the measurement of passive joint quasi-stiffness, defined as the rate of change of resistance torque during a slow angular displacement of the joint, in the absence of muscle contraction [1,12].

Extensive research has been conducted on passive joint quasi-stiffness of the lower limb and the impact on gait [6,7], with less research focusing on the upper limb and the equally important impact on activities of daily living, leisure activities, and sports. Of the joints within the upper limb, the neuromuscular control of the wrist has been identified as being dominated by joint stiffness [13,14]. It is, therefore, of paramount importance that we develop our understanding of the properties and variables of passive wrist stiffness in different populations to better comprehend how stiffness impacts the planning and coordination of wrist movements by the neuromuscular system during functional and sporting activities.

Within the literature on wrist joint stiffness, there is large variability in the magnitude and orientation of passive joint quasi-stiffness, likely because of differences in study methodology (alignment and orientation of the starting position, ROM, number of degrees of freedom assessed, and the method of data analysis) as well as subject characteristics (including sex, hand dominance, participation in sporting and leisure activities, and age). Previous studies have investigated the impact of study methodology [15-17], sex [14-17], and hand dominance [17]. A recent study [17] demonstrated that there was increased passive wrist quasi-stiffness in the dominant upper limb compared with the nondominant upper limb of healthy young men and women. This finding suggests that there are biomechanical factors associated with increased use of the dominant upper limb that influence the passive stiffness of the wrist. To date, the impact of both participation in a highly unilateral sporting activity and aging on passive wrist quasi-stiffness has not been investigated, despite the vital function of the wrist in common sporting activities such as tennis, bowling, golf, badminton, and squash.

Objectives
Squash players were chosen as the target population for this study because of the highly repetitive, intensive, unilateral nature of the sport and for the vital role the wrist and forearm play in generating high racquet head speeds during the forehand and backhand stroke actions [18,19]. Research has demonstrated that exercise causes adaptations in the properties of muscle tissue such as muscle fiber size, type, and muscle cross-sectional area [20,21], which have been associated with changes in passive joint quasi-stiffness [1,22]. As increased mechanical stimulation of the wrist likely leads to adaptation in muscle tissue during competitive squash play, we hypothesize that (1) squash players will have a larger difference in wrist quasi-stiffness between the dominant (playing side) and nondominant upper limb, regardless of age, and (2) the orientation of wrist stiffness would be equivalent for all study participants. The purpose of this study was to further develop our understanding of joint stiffness by determining the impact of intensive unilateral exercise, in this case playing competitive squash, and the process of aging on wrist properties.

Methods

Recruitment
Volunteer competitive squash players were recruited with a study flyer posted on a local Web-based squash association newsletter. The inclusion criteria were as follows: (1) male, (2) aged 18 years or older, (3) no prior wrist surgery or injury, (4) English-speaking, and (5) a minimum of 5 years of playing experience. Only male subjects were recruited to reduce the risk of data variability, as previous work had demonstrated a difference in the magnitude and direction of wrist quasi-stiffness between the male and female sex [14-17]. Subjects were asked not to participate in any upper limb exercise in the 24 hours preceding their evaluation session to decrease the risk of reduced wrist ROM because of muscle swelling [11] and muscle thixotropic behavior [23] associated with eccentric wrist exercise. To ensure the study participants were competitive players, their squash skill level was recorded using the United States Squash Rating Algorithm (USSRA) [24]. The USSRA calculates a measure of each player’s squash skill and ability using the data collected from all matches played in the last 45 months. Further details regarding the calculation of the algorithm can be found elsewhere [24].
A total of 7 healthy male competitive squash players aged between 24 and 72 years (mean 43.7, SD 16.57) volunteered to participate (Table 1). All 7 subjects identified as right hand-dominant for squash play, with 1 subject (subject 3 in Table 1) identifying as ambidextrous (left-hand dominant for activities of daily living, but right-hand dominant during squash play). The sample subject group had a mean of 20.6 years of squash playing experience (range 10-53 years, SD 13.85), a USRSL of 4.2 (range 3-5.5, SD 0.94—possible ranking ranges from 2 to 6, with 6 being the highest ranking), and reported playing between 2 and 5 times per week (median=4, mode=4). A total of 4 subjects were enrolled in the young squash group (mean age 30.75, SD 8.06 years), and 3 subjects were enrolled in the older squash group (mean age 64.67, SD 6.35 years). The Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects approved the study, with all volunteers providing written informed consent before participation.

Data from 7 right-hand dominant males (mean age 28.57, SD 12.11, range 19-55 years) reported in the study by Durand et al [17] (using the same experimental set-up) were used to compare wrist quasi-stiffness of unskilled subjects with the subjects participating in this trial who participate in regular, intensive, unilateral upper limb activity.

**Evaluation Method: Massachusetts Institute of Technology Wrist-Robot**

Wrist quasi-stiffness was evaluated using a 3 degree of freedom Wrist Robot (Figure 1, InMotion 3.0, Interactive Motion Technologies, Watertown, MA, USA), described elsewhere [25]. The robot forearm support positions the wrist joint so it aligns with the rotation axes of the robot. The Wrist Robot generates torques and simultaneously records the angular displacement produced into wrist flexion-extension (FE) and radial-ulnar deviation (RUD) [25]. A strap was used to lock forearm pronation-supination of the robot to eliminate confounding forearm movements during the trial. A gravity compensator was included in the robotic set-up to reduce the influence of gravity on the data collected. The gravity compensator was constant for each direction and equivalent to the sum of an average hand mass and the robot handle mass.

![Massachusetts Institute of Technology Wrist-Robot and experimental forearm, wrist, and hand position (excluding finger strap).](http://biomedeng.jmir.org/2019/1/e11670/)

---

### Table 1. Subject demographics and age group allocation.

<table>
<thead>
<tr>
<th>Group allocation</th>
<th>Young players</th>
<th>Older players</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subject 3</td>
<td>Subject 5</td>
</tr>
<tr>
<td>Age (years)a</td>
<td>31</td>
<td>42</td>
</tr>
<tr>
<td>Years of playing experience</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Squash sessions/week (n)</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>US squash rating criteria by skill level</td>
<td>5.5</td>
<td>5</td>
</tr>
<tr>
<td>Hours post upper limb exercise</td>
<td>24</td>
<td>&gt;48</td>
</tr>
</tbody>
</table>

*aSubject demographics were collected by interview before commencing the experiment.*
Experimental Procedure
Subject demographics were collected in the form of an interview (Table 1). The experiment commenced randomly with the left or right upper limb (using the MATLAB randi X function, MathWorks, Natick, 2016, MA) [26]. The reference position for the upper limb was in keeping with the set-up of Durand et al [17] and Drake and Charles [14] to allow accurate analysis and comparison of study results; the elbow was flexed to 30°, the third metacarpal was aligned with the forearm, and the wrist was positioned in 0° wrist extension and 7° of ulnar deviation (UD). Although previous studies have used an almost neutral RUD wrist position (0° along wrist UD), this study and Durand et al [17] deliberately selected 7° of wrist UD as this initial wrist position was more comfortable, allowing subjects to remain in a passive muscle state and avoid unwanted muscle activity. The wrist and forearm position was verified with goniometry [27]. Hand grip remained relaxed and the forearm in a neutral position by wrapping a strap over the fingers, securing the hand to the Wrist Robot handle. Subjects were instructed to remain relaxed throughout the experiment, which was monitored by palpating the muscle groups responsible for wrist FE-RUD and checking the data collected between trials. The robot applied a torque of up to 1.95 Nm to reach a predefined target (from 0-20° along each direction defined through the 2D FE-RUD space) at a predefined speed (between 0.1 and 0.2 rad/s to inhibit muscle reflexes). Data were collected at a rate of 200 Hz. Each trial consisted of 36 movements (inbound and outbound movements) along 12 equally spaced directions through the space defined by FE-RUD. The movements started in pure wrist extension for the right upper limb (pure flexion for the left side) and proceeded counterclockwise, with each of the 12 targets reached once. This cycle was repeated 3 times during each trial to reduce the influence of any artifacts (reflex or small muscle contraction). A trial was conducted on the left and right side before repeating the sequence a second time. Out of the 2 trials for each subject, the trial with the least data noise (unwanted muscle activity) on the left and right wrist was used in the data analysis.

Statistical Analysis
Torque and displacement data were processed using a customized program in MATLAB 2016b [26]. Data collected before commencing each trial of 36 movements were removed, knowing the time to complete each movement, the number of movements, and the acquisition frequency. The processed data were then applied to 2 different estimation methods, the fitting ellipse [16] and the multiple regression (MR) method [28], to obtain wrist stiffness magnitude and orientation among the 4 parameters commonly used (listed below) to characterize a stiffness ellipse [16,28]:

- Size: stiffness magnitude (ellipse surface [Nm/ rad]²)
- Orientation: stiffness orientation (angle in degrees between radial deviation (RD) direction and ellipse major axis direction toward RD, counterclockwise angles are considered positive)
- Shape: the ratio of the major axis of the stiffness ellipse to the minor axis
- Equilibrium position: the offset of the ellipse center corresponding to the FE and RUD offset angles.

The fitting ellipse method (Figure 2) calculates the torque and angular displacement parallel and perpendicular to the direction of each of the 36 perturbations. Stiffness was then estimated (separately for outbound and inbound movements) by running a linear regression of the torque and the angle parallel to the perturbation direction. The mean stiffness values for the estimates of each of the outbound and inbound 36 perturbation directions were then used to fit a stiffness ellipse. Previous research indicates that the major weakness of the fitting ellipse approach is that it only considers the components of torque and angle parallel to the perturbation direction and does not include stiffness effects perpendicular to the perturbation direction. Of note, the fitting ellipse method allows for asymmetry of the elastic field with respect to the neutral position and is susceptible to data noise [17]. Although we fitted the stiffness values with a least square condition to keep consistent ellipse shapes, an excessive stiffness value along 1 direction will tend to stretch the ellipse and increase the size of the ellipse.

The MR method (Figure 2) determines the 4 elements of the stiffness matrix by multiple linear regression (using MATLAB’s regress function [26]). Separate stiffness matrices were estimated for the inbound, outbound, and the composite of both movements. Each matrix was separated into the symmetric and asymmetric parts. Only the symmetric part of each stiffness matrix was displayed as a stiffness ellipse [28].

A 1-sample Kolmogorov-Smirnov test was used to confirm the normal distribution of the data. A 1-way analysis of variance was calculated to determine the statistical difference between the wrist quasi-stiffness of the left and the right arm in the young and older squash player groups and to compare the magnitude of the wrist quasi-stiffness and orientation of the 4 young squash players with the results of the 7 young right-handed dominant, unskilled male subjects [17]. The level of statistical significance for comparisons was set to P<.05. A Pearson r calculation was performed to determine the correlation between the magnitude of passive wrist quasi-stiffness in both the young and older squash groups and (1) age, (2) years of squash play, and (3) frequency of squash play.
**Results**

**Magnitude of Wrist Stiffness in Squash Players: Hand Dominance and Age Group Analysis**

For the older adult group, the mean passive wrist stiffness magnitude of the left upper limb was 11.91 Nm/rad² (SD 1.26) and 9.99 Nm/rad² (SD 2.12) for the right upper limb. The ratio between right (dominant playing upper limb) and the left arm for the older adult group was 0.86. The older adult demonstrated a higher mean stiffness in the left (nondominant, nonplaying) upper limb (Figure 3). For the young adult group, the mean passive wrist stiffness magnitude of the left upper limb was 4.58 Nm/rad² (SD 1.39) and 6.75 Nm/rad² (SD 4.44) for the right upper limb. The ratio between right and left upper limb for the young squash group was 1.51. This playing group demonstrated a higher mean stiffness in the right (dominant playing) upper limb (Figure 4). There was a statistically significant difference between the magnitude of passive quasi-stiffness between the dominant and nondominant wrist of the young and older squash player groups (P=.004).
**Magnitude of Wrist Stiffness in Young Adults: Unskilled and Squash Player Analysis**

Results from the comparison of the 4 young squash players (ratio right over left stiffness magnitude of 1.51) with the 7 young unskilled males (ratio right over left stiffness magnitude of 1.18) showed a stiffer dominant wrist for the squash players. There was no statistically significant difference between the young unskilled and the young squash player groups ($P=0.98$).

**Correlations: Magnitude of Wrist Stiffness and Subject Characteristics**

The correlation between the passive wrist stiffness of all 7 players and subject characteristics revealed an interesting difference between the dominant and nondominant upper limb. There was a strong positive correlation between passive stiffness of the nondominant left wrist and age ($R^2=0.87$), whereas there was no clear correlation between passive wrist stiffness of the dominant right wrist and age ($R^2=0.09$) or years of squash play ($R^2=0.01$). When analyzing the correlation of the subject characteristics and the age subgroups, the young squash player group showed a strong positive correlation between passive wrist stiffness of the right wrist and the frequency of play ($R^2=0.76$). The older adult players demonstrated a strong negative correlation between passive wrist stiffness of the right wrist and the age of the player ($R^2=0.88$) and the number of years of playing experience ($R^2=0.96$).

**Orientation of Wrist Stiffness: Unskilled and Squash Player Analysis**

For the young unskilled group, the orientation of highest quasi-stiffness followed a “dart throwing” pattern, with a mean angle of 14.42° (SD 4.45) for the left wrist and a mean angle of −13.96° (SD 7.06) for the right [17]. For the young squash player group, the orientation of highest quasi-stiffness had a mean angle of 16.11° (SD 12.80) for the left wrist and a mean angle of −20.95° (SD 15.20) for the right. Both groups showed symmetric values for right and left wrists; however, a statistically significant difference was found between the unskilled and squash player groups ($P<0.001$).
Discussion

Passive Wrist Quasi-Stiffness and Exercise

The magnitude of passive wrist quasi-stiffness of the young players confirmed the study hypotheses that competitive squash players demonstrate greater quasi-stiffness in the dominant playing upper limb compared with the nondominant upper limb. Although the ratio of dominant over nondominant wrist quasi-stiffness for young squash players compared with unskilled young subjects did not reach statistical significance, there was a trend favoring the squash player group. Studies investigating other sporting activities have shown that muscle strength, orientation of joint stiffness, and pattern of ROM between the playing and nonplaying upper limbs do not consistently increase or decrease on the playing side. Klinge et al [29] studied the effect of strength training on flexibility and demonstrated that a 43% increase in isometric muscle strength resulted in a 25% increase in passive joint stiffness. Borsa et al [30] studied the difference in glenohumeral joint stiffness and ROM between the pitching and nonpitching upper limb in professional baseball players. Their results showed a difference in the ROM of the pitching and nonpitching upper limb but not a significant increase in passive stiffness on the pitching side. Indeed, the correlation between the measure of joint ROM and passive joint stiffness remains unclear. Some studies claim that joint ROM and passive joint stiffness (K) can be considered 2 components of the same phenomenon. Pando et al [15] demonstrated that the pattern of wrist stiffness was inversely related to the ROM (K_{Radial Deviation} > K_{Ulnar Deviation} > K_{Extension} ~ K_{Flexion}), whereas ROM_{Radial Deviation} < ROM_{Ulnar Deviation} < ROM_{Extension} ~ ROM_{Flexion}). However, in the study by Gleim and McHugh [31], an increase in ROM of a joint did not correspond with a decrease in passive stiffness of a joint or muscle. This study claimed that changes in ROM can be attributed to increased stretch tolerance without a change in stiffness magnitude [31].

There have been few studies, to the authors' knowledge, investigating the effect of exercise or sporting activities on passive wrist stiffness. Leger and Milner [11] investigated passive and active joint stiffness following wrist extensor muscle injury caused by a single session of intensive eccentric exercise. The study concluded that passive wrist stiffness (measured at a single neutral joint position) did not change following eccentric exercise-induced muscle injury. These findings suggest that this form of exercise does not cause mechanical changes to the noncontractile tissues in a neutral wrist position following a single session of such exercise. It is, therefore, likely that both differences in sporting activities as well as the biomechanics of joints will limit the comparison of our study results. Instead, an evaluation of the magnitude of quasi-stiffness in this study (of the young players), when compared with similar studies using unskilled subjects [15,17], reveals that passive wrist stiffness is slightly higher for competitive squash players. This trend in passive wrist stiffness is likely attributed to the greater cross-sectional area of the surrounding forearm muscles and tensile strength of the passive wrist structures (ligaments, tendons, and bone geometry), which are due to increased mechanical loading of the wrist during frequent, repetitive, and intensive squash play [32-35]. This conjecture is supported by the strong positive correlation ($R^2=0.76$) between the frequency of play and passive wrist stiffness in the young player group. The increased mechanical loading of the wrist associated with squash play may also explain the significant difference in the orientation of wrist quasi-stiffness between the unskilled and young squash playing groups. Nonetheless, as there is inconsistency in the orientation of wrist quasi-stiffness reported in the literature, it is possible that both differences in study subject's physical activities and study methodology contribute to the variance seen [14-17].

Influence of Aging and Exercise on Quasi-Stiffness

Although age-related increases in passive joint stiffness may seem clinically obvious, how exercise impacts passive joint stiffness in older adults has not been widely studied and results are inconsistent. Inconclusive results are likely because of variances in the joints evaluated, the study methodology, and the definition of flexibility and stiffness measurements. In research on flexibility, active older tennis players were shown to maintain shoulder flexibility on their playing side [36], whereas older soccer players were shown to have less flexibility in lumbar flexion and hip rotation than younger players [37]. Investigations into the effects of aging and immobility have shown that reduced physical mobility can be associated with increased passive stiffness. Lapier et al [38] and Gillette and Fell [39] found in animal studies that immobilization leads to increased intramuscular connective tissue and increased joint stiffness. In this study, the negative correlation between wrist stiffness on the dominant playing side and both the years of squash play ($R^2=0.96$) and age ($R^2=0.88$) in the older adult group strongly suggest that participating in high-intensity exercise could, in fact, slow or even prevent increases in joint stiffness during the aging process. This theory explains the higher stiffness magnitude on the nondominant (nonplaying) wrist in the older squash group and the significant difference in the magnitude of quasi-stiffness between the older and young squash players.

Changes to Muscle Fiber During Aging

The effect that squash play appears to have on the passive stiffness of the wrist with aging may partly be because of changes in the physiological properties of muscle fibers. The wrist flexor and extensor muscles are composed of approximately 50% type I and 50% type II muscle fibers in young adults aged 17 to 30 years [11,40]. Type II muscle fibers appear to be the most affected by aging, with reports of a 15 to 26% reduction in type IIa and IIb cross-sectional area and a preferential denervation of type II fibers between the age of 20 and 80 years [41]. The increasing proportion of smaller, slow-contracting type I muscle fibers with aging is thought to be an adaptive response to minimize fiber loss. The changes to type II muscle fibers and the increasing percentage of type I fibers are thought to be largely responsible for the reduced muscle mass, force generation of muscle tissue [41], and possibly the increased passive joint stiffness of older adults [42].
Type I muscle fibers have been reported to have greater passive stiffness, likely because of increased collagen concentration and cross-linking of collagen compared with type II muscle fibers [8,42]. Human and animal studies have demonstrated that an increase in type I muscle fibers leads to increased passive stiffness, whereas when physical exercise induces an increase in type II muscle fibers, there is a corresponding decrease in passive stiffness [8,42]. These observations might explain the pattern of passive wrist joint stiffness seen in the older squash playing group, where the left nonplaying wrist showed a higher passive stiffness (probably owing to an increased percentage of type I muscle fibers during the aging process) and the right dominant playing wrist demonstrating less passive stiffness (likely because of an exercise-induced relative increase or sparing of type II muscle fibers.)

**Limitations**

One must take this study with the appropriate caveats; several features of our study design limit the impact and generalizability of the results. Notably, the small sample size of this pilot trial significantly limits the power of the results. Other papers within this field of study have also reported small sample sizes ranging from 6 to 15 subjects [13-17]. The volunteer method of recruitment may have introduced self-selection bias, whereby the volunteer study subjects are not representative of the broader population that participate in squash or other unilateral upper limb sporting activities. Although the study recruited subjects with a wide age range, there were not enough subjects within each age group, and hence, for the most part, we only observed statistical trends favoring some of our conclusions. To reliably determine that there was a significant difference in wrist quasi-stiffness between the dominant (playing side) and nondominant upper limb of the older squash player group, a power analysis (GPower 3.1) indicates that (assuming an alpha value of .05, power of 0.80, and an effect size (f) of 0.32, calculated from the results of this pilot study) a total of 82 subjects would be required. Additional demographic data such as the duration of squash play, as well as the frequency of play per week, would better represent the subject’s playing intensity and may enable more accurate characterization of the influence of playing intensity on passive joint quasi-stiffness. In addition, collecting data such as grip strength and electromyography would have provided further insight into the subject’s muscle characteristics and muscle state during the trials to validate the measure of passive joint quasi-stiffness. Nevertheless, we believe our results provide additional information on a sparse landscape of limited literature on the impact of intensive unilateral exercise on passive joint stiffness. Indeed, the absence of comparative studies indicates that our study’s findings are novel and potentially influential and that further research is required in this field.

**Conclusions**

This study provides a valuable initial insight into the possible effect that highly intensive, repetitive, unilateral sports may have over time on wrist quasi-stiffness and reducing the impact of the aging process. Further studies are required to investigate this relationship with a larger sample size and age group analysis. This field would also benefit from the study of passive wrist stiffness in young and older subjects who participate in intensive bilateral sporting activities such as upper limb weight or grip training to determine the magnitude and effect of changes in muscle fiber type across the lifespan. Our findings confirm that the evaluation of passive joint stiffness has relevance and far-reaching value in many fields, from sporting activities to the rehabilitation of the older adults, following surgical interventions, or those with neurological impairments.

**Acknowledgments**

The authors would like to thank the study volunteers for their time and participation. The authors would also like to acknowledge Joseph Davidson, Rogerio Goncalves, and the members of The 77 Lab at the Mechanical Engineering Department, Massachusetts Institute of Technology (MIT) for their technical support during the planning and data collection phases of this study.

**Conflicts of Interest**

HIK is a co-inventor of several MIT-held patents for robotic therapy. He was the founder of Interactive Motion Technologies and Chairman of the Board (1998-2016). He successfully sold Interactive Motion Technologies to Bionik Inc. He founded 4Motion Robotics, Inc in 2017. TH is now employed by Bionik Inc, the company that purchased Interactive Motion Technologies.


Abbreviations
- FE: flexion-extension
- MR: multiple regression
- RD: radial deviation
- ROM: range of motion
- RUD: radial-ulnar deviation
- UD: ulnar deviation
- USSRA: United States Squash Rating Algorithm

©Taya Hamilton, Stan Durand, Hermano Igo Krebs. Originally published in JMIR Biomedical Engineering (http://biomedeng.jmir.org), 07.05.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on http://biomedeng.jmir.org/, as well as this copyright and license information must be included.
Determining the Accuracy of Oculus Touch Controllers for Motor Rehabilitation Applications Using Quantifiable Upper Limb Kinematics: Validation Study

Leia C Shum¹, BASc; Bulmaro A Valdés¹, BEng, MPE, PhD; HF Machiel Van der Loos¹, PEng, PhD

RREACH (Robotics for Rehabilitation Exercise and Assessment in Collaborative Healthcare) Lab, Department of Mechanical Engineering, The University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Leia C Shum, BASc
RREACH (Robotics for Rehabilitation Exercise and Assessment in Collaborative Healthcare) Lab
Department of Mechanical Engineering
The University of British Columbia
ICICS X015
2335 Engineering Rd
Vancouver, BC, V6T 1Z4
Canada
Phone: 1 (604) 822 3147
Email: lcshum@alumni.ubc.ca

Abstract

Background: As commercial motion tracking technology becomes more readily available, it is necessary to evaluate the accuracy of these systems before using them for biomechanical and motor rehabilitation applications.

Objective: This study aimed to evaluate the relative position accuracy of the Oculus Touch controllers in a 2.4 x 2.4 m play-space.

Methods: Static data samples (n=180) were acquired from the Oculus Touch controllers at step sizes ranging from 5 to 500 mm along 16 different points on the play-space floor with graph paper in the x (width), y (height), and z (depth) directions. The data were compared with reference values using measurements from digital calipers, accurate to 0.01 mm; physical blocks, for which heights were confirmed with digital calipers; and for larger step sizes (300 and 500 mm), a ruler with hatch marks to millimeter units.

Results: It was found that the maximum position accuracy error of the system was 3.5 ± 2.5 mm at the largest step size of 500 mm along the z-axis. When normalized to step size, the largest error found was 12.7 ± 9.9% at the smallest step size in the y-axis at 6.23 mm. When the step size was <10 mm in any direction, the relative position accuracy increased considerably to above 2% (approximately 2 mm at maximum). An average noise value of 0.036 mm was determined. A comparison of these values to cited visual, goniometric, and proprioceptive resolutions concludes that this system is viable for tracking upper-limb movements for biomechanical and rehabilitation applications. The accuracy of the system was also compared with accuracy values from previous studies using other commercially available devices and a multicamera, marker-based professional motion tracking system.

Conclusions: The study found that the linear position accuracy of the Oculus Touch controllers was within an agreeable range for measuring human kinematics in rehabilitative upper-limb exercise protocols. Further testing is required to ascertain acceptable repeatability in multiple sessions and rotational accuracy.


KEYWORDS
upper extremity; kinematics; physical medicine and rehabilitation; validation studies; virtual reality

Introduction

Current gaming and virtual reality platforms [1] that use motion-controlled interfaces offer an affordable and accessible method of tracking human kinematics. However, given that consumer-grade platforms are originally intended for playing video games and to immerse players in virtual environments, their tracking performance should be evaluated before they are employed as tools for biomechanical or clinical analysis [2]. Previously tested rehabilitation protocols using commercial
gaming technology such as Wii Motes (Nintendo Co, Ltd, Kyoto, Japan) to provide positional feedback for trunk compensation [3] or a Kinect (Microsoft Corporation, Redmond, United States) to measure range and speed of motion for upper-limb exercises [4,5] have shown potential to be used as rehabilitation tools that could provide quantifiable changes in clients’ kinematic motor abilities to therapists. Other studies using accelerometers to track patterns in functional upper-limb movements were able to capture differences similar to those measured by clinical scales [6] and found benefits from objective quantitative evaluations of changes in motor ability during therapy regimens, which can be collected from in-game progress reports [7]. In addition, success has been found in translating kinematic upper-limb metrics to clinical Fugl-Meyer scoring [8] and in detecting exercise repetitions via kinematic monitoring for telerehabilitation and at-home programs [9]. Current clinical assessments for upper-limb motor function, such as the Fugl-Meyer Assessment and Wolf Motor Function Test, only provide low-resolution point-scores rated qualitatively by therapists, and kinematic analysis of upper-limb motion has been reported to be a useful addition to these clinical assessments [10]. When measuring range of motion in a clinical setting, the goniometer is considered a gold-standard clinical measurement tool used by therapists [11]. However, only static joint angles can be measured, and typically with some visual estimation and multiple testers [12].

One of the latest (released December 2016) devices to be developed for interacting with virtual environments is the Oculus Touch (Oculus VR, LLC, Menlo Park, CA, United States) controller set. The controllers are peripheral accessories of the Oculus Rift virtual reality headset and are employed to track users’ hand movements. Their tracking system employs a proprietary algorithm that collects data from infrared sensors via constellation tracking [13] and inertial measurement units (IMUs). Given that the controllers are wireless, lightweight, low-cost devices that can be used to track a user’s hand position and orientation in 3-dimensional (3D) space, they could have the potential to be employed in rehabilitative and biomechanical motion-tracking applications. At the time of this study, there was no sufficient information about the tracking performance of the controllers provided by the manufacturer, and there is currently a lack of scientific papers employing a systematic approach to test their potential application as tools for motion-tracking data capture. As a result, in this study, we evaluated the tracking accuracy of the Oculus Touch controllers to present a preliminary evaluation that could be informative to the biomechanical and rehabilitation research community. The specific aim of the experiment was to quantify the relative positional accuracy of the Oculus Touch controllers in 3 spatial dimensions. As the controllers are intended for hand-held motion control, the evaluation setup was centered around the movement size for standing/sitting upper-limb reaching tasks.

**Methods**

**Technical Setup**

An Oculus Touch controller (Figure 1), 2 Oculus Sensors, an Oculus Rift headset, and a computer running Windows 10 (Microsoft Corporation) were employed in this study. A custom computer application was developed in Unity 2017 (Unity Technologies, San Francisco, United States) to capture and log the controller’s position during the experiment. The data capture was performed at the headset’s native frequency of approximately 90 Hz, using the Unity OVR Plugin package to access controller data. The virtual environment was set up over a 2.4 m x 2.4 m play-space in the x-z plane to be within the recommended manufacturer play area. This space consists of 16 commercial 600 mm square force/torque plates professionally installed on a subfloor of auto-levelling epoxy and flat to within 0.5 mm (Figure 2). The y-axis was only bounded by the camera sensors’ field of view limitations.

To ensure consistency, the Oculus Sensors were placed on the floor at 0.3 m along the front edge of the space and 1.2 m apart, equidistant from the centre line, for the entire experiment. The sensor heads were manually leveled and visually aligned to have parallel, front-facing fields of views. Both the sensors and controllers maintained an initial y-position of 0 at the floor—this would be equivalent to placing the sensors at table height and the controllers at hand height.

All measurements were taken by securing the right-hand Oculus Touch controller to a flat L-shaped jig (Figure 2) and resting it on the floor for 5 seconds. Initial calibration of floor height and play-space size and orientation was done through the official commercial Oculus setup client.

**Figure 1.** The right-side Oculus Touch controller. Left: front view. Right: top-down view.
Figure 2. The experimental setup and coordinate frame. The play-space was divided into 16 squares.

Experimental Procedure

Measurements were taken along each of the 3 spatial axes (x: width, y: height, z: depth) in a single session. The x and z axes were measured in increments of 5, 10, 50, 150, 300, and 500 mm steps relative to a recorded 0 value. The estimate for a 500 mm largest step was attributed to an approximate lower arm and hand length from human anthropometric data in Huston [14]. This length should replicate the size of a simple outward reach from the elbow. Each set of steps was taken from the zero line of each axis in both the positive and negative direction and then taken in the positive direction at +600 mm and in the negative direction at –600 mm along the same axis (Figure 3; left). Graph paper with millimeter unit markings was used to define the step sizes to the relative 0 point of each set. The graph paper step sizes were verified using a Mitutoyo 500-196 digital calipers, accurate to 0.01 mm, visually aligned to the edge of the unit markings within the third significant digit. A ruler with half-millimeter unit markings was used for steps larger than 150 mm. The bottom left corner of the jig was used as an origin for the 3 axes with respect to the controller. The x and z axes edges were aligned with the graph paper visually. The L-shaped jig was checked for orthogonality using a calibrated 90-degree ruler in all 3 directions before its use. To test the repeatability of the L-shaped jig alignment on graph paper, the controller was moved at least 300 mm away from and then toward a single point near the centre of the play-space on each axis 3 times.

In the x-axis, 4 sets of steps were taken at 4 different depths for a total of 16 sets of steps to measure the accuracy of the controllers over the play-space area (Figure 3). The same configuration was used for the z-axis but using 4 sets of steps along 4 different x-axis values, 600 mm apart.
Figure 3. Left: A top-down visual representation of the expected spacing of the data points in the x-axis. Right: The x-z points at which the y step sets were taken.
Figure 4. Close-up of the Oculus Touch controller in the L-shape jig over an aluminium block used to measure y-axis steps. Both controller (system) and graph paper (physical world) axes are represented and were aligned visually during the experiment.

The 16 y-axis step sets were taken 600 mm apart from each other in the x-z plane starting 300 mm from the 0 line in each perpendicular axis in the x-z plane. This was the approximate centre of each of the 16 tiles seen in Figure 3 (right). The y steps were measured by placing the controller and jig on level aluminum blocks of specific heights, which were measured using zeroed, calibrated digital calipers accurate to 0.01 mm. The L-shaped jig was used as a physical origin point and axes was aligned with the top left corners of the aluminum blocks (Figure 4).

The aluminum blocks were chosen to allow the y-axis step sizes to approximately match the same x and z step sizes, and had heights of 6.23, 12.56, 50.82, and 152.5 mm. The x and z positions were monitored and recorded but not analyzed for alignment accuracy.

Data Analysis

To remove motion artifacts from pressing the buttons on the controller to start and stop the data recording, the first 1.5 seconds (135 samples) were removed and the next 2 seconds (180 samples) were used as the sample data for each measured point. After those 2 seconds, the rest of the data recording (approximately 1.5 seconds or 135 samples) was also discarded, regardless of length. The data were averaged to calculate the measured value at each point.

For each sample point, the 3D position of the controller was measured. The variation was calculated for each data point and used to determine the static precision of the system. The position error for each point was calculated by subtracting the measured displacement (Euclidean distance) by the expected step size. The Euclidean distance was used to account for any misalignment of the Oculus tracking coordinate system with respect to the physical grid. The error values were then averaged to generate values for expected displacement error in a specific area of the play-space as well as for a specific step size over the entire play-space. The percent error was calculated to normalize the error to the step size.

Results

Positional Accuracy

The average and percent errors for all step sizes in the x, y, and z directions are presented in Tables 1 and 2.

The largest absolute error was found to be 3.5 mm in the z-axis for a step size of 500 mm, which normalizes to a 0.7% error. The largest normalized error was found to be 12.7% for the smallest step of 6.23 mm in the y-direction. The largest percent errors for the x and z axes were 4.7% and 3.5%, respectively, also at the smallest step size (5 mm).
Table 1. Position error for different step sizes at all areas of the defined play-space measured with 0.01 mm accuracy using digital calipers. The percent error was calculated using nonrounded values of error in millimeter.

<table>
<thead>
<tr>
<th>Directional axis and step size (mm)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average (SD), mm</td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
</tr>
<tr>
<td>5.00</td>
<td>0.23 (0.19)</td>
</tr>
<tr>
<td>10.00</td>
<td>0.25 (0.18)</td>
</tr>
<tr>
<td>50.00</td>
<td>0.39 (0.29)</td>
</tr>
<tr>
<td>150.0</td>
<td>0.76 (0.50)</td>
</tr>
<tr>
<td><strong>z</strong></td>
<td></td>
</tr>
<tr>
<td>5.00</td>
<td>0.17 (0.15)</td>
</tr>
<tr>
<td>10.00</td>
<td>0.25 (0.22)</td>
</tr>
<tr>
<td>50.00</td>
<td>0.28 (0.22)</td>
</tr>
<tr>
<td>150.0</td>
<td>0.72 (0.46)</td>
</tr>
<tr>
<td><strong>y</strong></td>
<td></td>
</tr>
<tr>
<td>6.23</td>
<td>0.79 (0.62)</td>
</tr>
<tr>
<td>12.56</td>
<td>0.48 (0.82)</td>
</tr>
<tr>
<td>50.82</td>
<td>0.41 (0.62)</td>
</tr>
<tr>
<td>152.5</td>
<td>0.93 (1.10)</td>
</tr>
</tbody>
</table>

Table 2. Additional position error for larger step sizes at all areas of the defined play-space measured with ruler and graph paper markings. The percent error was calculated using nonrounded values of error in millimeter.

<table>
<thead>
<tr>
<th>Directional axis and step size (mm)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average (SD), mm</td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
</tr>
<tr>
<td>300.5</td>
<td>1.5 (1.0)</td>
</tr>
<tr>
<td>500.5</td>
<td>2.5 (1.0)</td>
</tr>
<tr>
<td><strong>z</strong></td>
<td></td>
</tr>
<tr>
<td>300.5</td>
<td>2.0 (1.0)</td>
</tr>
<tr>
<td>500.5</td>
<td>3.5 (2.0)</td>
</tr>
</tbody>
</table>

When the average percent error was calculated for each step size across the entire play-space, it was found that the error decreased nonlinearly with increasing step size (Figure 5).

Step sizes with values of ≤10 mm had an accuracy error greater than 2%. Normalized error averages in step sizes larger than >10 mm were fairly uniform.

The variation was calculated for static data points at the 16 different locations in the play-space and averaged across all points to find an average noise value of ±0.036 mm (x: ±0.025 mm, y: ±0.024 mm, z: ±0.055 mm). The single-point repeatability test found that the controller was able to return to the same x-y-z point with an output measurement variation slightly above the average noise value (x: 0.080 ± 0.546 mm, y: 0.088 ± 0.063 mm, z: 0.044 ± 0.326 mm).

To investigate how accuracy varied over the entire play-space, the displacement and percent error were averaged over each area set. These average errors over the play-space area for x, y, and z are presented in Multimedia Appendix 1 as exploratory analysis to see if there were any patterns in accuracy based on distance away from the Oculus sensors. A large error (14.0%) occurred in 1 area of the y-axis measurements. This was a result of a 2 mm error at a 6.23 mm step, resulting in a large normalized value of approximately 40% despite being a small error in absolute distance (millimeters). No distinct pattern of position accuracy based on x-z location in the play-space was observed.
Discussion

Principal Findings

The study team found a maximum positional accuracy across measured step sizes less than 150 mm for the Oculus Touch controller of $0.76 \pm 0.50$ mm in the lateral x-axis, $0.72 \pm 0.46$ mm in the anteroposterior z-axis, and $0.93 \pm 1.10$ mm in the vertical y-axis direction. Larger step sizes found lower positional accuracies of $2.5 \pm 1.0$ mm in the x-axis and $3.5 \pm 2.5$ mm in the z-axis. The largest error in percent when normalized to step size ($12.7 \pm 9.9\%$) was found in the smallest step size in the y-axis set at 6.23 mm.

The error values found are considered within an acceptable range of error for the measurement of biomechanical movement as the human perception of the just noticeable difference (JND) even for fine motor function (finger distance/position) is larger than this value (13.0% for young subjects and 16.1% for older persons [15]). In addition, in a different study [16], it was reported that the JND of the fully extended human shoulder when moved passively was found to be 0.8º. Using a 50th percentile female arm length of 702 mm [14], a 0.8º change in joint angle would cause an arc length of 9.8 mm, which is larger than both the average noise and accuracy error of the system. Therefore, the error in relative distance should not be noticeable to the user.

In studies on the accuracy of visual assessments of angular joint positions done by physical therapists and other health professionals, it was determined that joint positions could only be determined with an error of approximately 5º within the referenced radiometry measurement [17] of wrist angle and 7.4º in shoulder abduction compared with goniometry of nonmoving subjects [18]. Measurement of glenohumeral range of motion using a goniometer was reported to have an SE between 4.4 to 9.9º [19]. Moreover, the Oculus Rift headset has a visual field of view of 100º and a resolution of 2160x1200 pixels [20]. This results in a 0.046º change in the object edge to show up as a single pixel change. With a noise level of 0.069 mm as an arclength, it is expected that visual jitter would not occur until the controller is less than 8.59 mm away from the user in the headset’s point of view.

It was found that although normalized percent error decreased as step size increased, the absolute error (millimeter) was found to be largest when the largest step was measured. This could be the effect of an inherent scaling phenomenon found in the Oculus sensor tracking, which uses infrared image processing as 1 of its main sources of position tracking for the Oculus Touch controllers.

We expected that the outer edges of the defined play-space would provide areas with the largest error; however, no discernable pattern was found to occur over the x-z plane. This provides evidence to support a consistent accuracy of the controllers within the documented [21] x-z field of view of the Oculus Sensors. That is, regardless of where in the play-space a user might stand, a reach of 500 mm outwards from the body would be measured sufficiently accurately.

Comparison With Prior Work

On the basis of the results from this study, the Oculus Touch controllers should be an adequate motion tracking alternative for biomechanics applications, as similar low-cost, commercially available systems that have been employed to measure joint movements, such as the Kinect V1 and V2, have displacement accuracies on the order of centimeters [22,23]. A previous study
evaluating the tracking accuracy of the Oculus Rift head-mounted display found similar accuracy values for the larger step sizes [24]. On the other hand, more expensive and complex motion capture systems such as the Vicon-460 (Vicon Motion Systems Ltd, Oxford, United Kingdom) with submillimetric accuracy [25] allow researchers to measure movements with higher accuracy. In addition, wearable inertial sensors were able to quantify upper-limb positioning within 1 mm when custom sensor algorithms were applied [7]; however, wearable research-grade systems trade ease of use and commercial availability for higher accuracy. For kinematic tracking that does not require submillimetric accuracy, such as for monitoring changes in gross motor upper-limb movements over time in a rehabilitation program [26], for training neural networks to detect the number of repetitions during rehabilitative exercises [9], or in cases where there would otherwise be no quantitative measures [8], the Oculus Touch controllers could provide a cost-effective alternative. Comparison with the Oculus Touch controllers by using other tracking tools simultaneously during upper-limb exercises may provide better insight into which level of accuracy is optimal for different use cases, such as for automated repetition counting as opposed to for measuring joint angles for digital goniometry. In cases where the tracking technology is used to facilitate virtual environments for engaging upper-limb exercises, a higher position accuracy may provide better visual fidelity to movement in the real world, and therefore, better transfer of improved motor functions from game tasks to real-world tasks [27].

Limitations

Occlusion of some of the controllers’ infrared light-emitting diodes could have occurred when placing the controllers close to the floor, which might have increased the measured error. Moreover, as the system requires an initial calibration of the user’s approximate height, this also could have acted as an additional source of error. As a result, future studies should investigate the accuracy of the system away from the floor, as well as the accuracy of the controller’s 3D orientation measurement, as we only measured position error in this study. Reproducibility of the measurements made by the Oculus system should also be investigated by having multiple experimenters perform the same procedure and by comparing measurements from different tracking sessions. Standardized measurement system analysis procedures should be followed in terms of the number of repetitions used as listed in analysis of variance gage repeatability and reproducibility documentation [28]. Dynamic conditions should also be evaluated before use in clinical kinematic analysis to assess the interaction between IMU sensor drift and camera sensor correction while in motion. A limitation of the Oculus Touch controllers is that it is only capable of measuring the position and orientation of a single point as a proxy for hand position. Future studies should directly compare the Oculus Touch absolute point-position with a professional marker-based motion-tracking system to ensure the elimination of error because of the use of visual and physical measuring tools. These absolute point-position studies should also evaluate larger step size accuracies to encompass bigger movements. Additional studies could include the evaluation of inverse kinematic algorithms that employ the hand and head positions (from headset) to generate a model of the user’s arms [29,30]. This would allow direct comparison against other devices that digitally measure goniometric angles.

Acknowledgments

The authors would like to acknowledge the support of the University of British Columbia’s Collaborative Advanced Robotics and Intelligent Systems and Robotics for Rehabilitation Exercise and Assessment in Collaborative Healthcare Lab members. Research reported in this publication was financially supported by the Kids Brain Health Network (TotTech).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Exploratory results of accuracy based on x-z location: displacement in millimeters (left) and in percent (right) accuracy error for the Oculus Touch controller in the x (top), y (centre), and z (below) directions. Error is represented by the width of the circle for each step set area, however the circle size scale was magnified for visualization purposes and it is not to scale with the rest of the chart. Standard deviation is shown in parentheses.

[PNG File, 232KB - biomedeng_v4i1e12291_app1.png ]

References


Abbreviations

3D: 3-dimensional
IMU: inertial measurement unit
JND: just noticeable difference