Contents

Original Papers

Effectiveness of the BreatheSuite Device in Assessing the Technique of Metered-Dose Inhalers: Validation Study (e26556)
Meshari Alwashmi, Gerald Mugford, Brett Vokey, Waseem Abu-Ashour, John Hawboldt. ......................................................... 2

Understanding “Atmosome”, the Personal Atmospheric Exposome: Comprehensive Approach (e28920)
Hari Bhimaraju, Nitish Nag, Vaibhav Pandey, Ramesh Jain. ........................................................................................................ 10

Proposal

Tracking the Presence of Software as a Medical Device in US Food and Drug Administration Databases: Retrospective Data Analysis (e20652)
Aaron Ceross, Jeroen Bergmann. ............................................................................................................................................... 39
Effectiveness of the BreatheSuite Device in Assessing the Technique of Metered-Dose Inhalers: Validation Study

Meshari F Alwashmi¹, BN, MSCE, PhD; Gerald Mugford², CMH, BSc, PhD; Brett Vokey¹, BEng; Waseem Abu-Ashour², PharmD, MSc; John Hawboldt², BSP, PharmD

¹BreatheSuite Inc, St. John’s, NL, Canada
²Memorial University, St. John’s, NL, Canada

Corresponding Author:
Meshari F Alwashmi, BN, MSCE, PhD
BreatheSuite Inc
215 Water Street, Suite 301
St. John’s, NL, A1C 6C9
Canada
Phone: 1 7096910728
Email: mfa720@mun.ca

Abstract

Background: The majority of medications used in treating asthma and chronic obstructive pulmonary disease (COPD) are taken through metered-dose inhalers (MDIs). Studies have reported that most patients demonstrate poor inhaler technique, which has resulted in poor disease control. Digital Health applications have the potential to improve the technique and adherence of inhaled medications.

Objective: This study aimed to validate the effectiveness of the BreatheSuite MDI device in assessing the technique of taking a dose via an MDI.

Methods: The study was a validation study. Thirty participants who self-reported a diagnosis of asthma or COPD were recruited from community pharmacies in Newfoundland and Labrador, Canada. Participants used a BreatheSuite MDI device attached to a placebo MDI and resembled taking 3 doses. Pharmacists used a scoring sheet to evaluate the technique of using the MDI. An independent researcher compared the results of the pharmacist’s scoring sheet with the results of the BreatheSuite device.

Results: This study found that the BreatheSuite MDI can objectively detect several errors in the MDI technique. The data recorded by the BreatheSuite MDI device showed that all participants performed at least one error in using the MDI. The BreatheSuite device captured approximately 40% (143/360) more errors compared to observation alone. The distribution of participants who performed errors in MDI steps as recorded by BreatheSuite compared to errors reported by observation alone were as follows: shaking before actuation, 33.3% (30/90) versus 25.5% (23/90); upright orientation of the inhaler during actuation, 66.7% (60/90) versus 18.87% (17/90); coordination (actuating after the start of inhalation), 76.6% (69/90) versus 35.5% (32/90); and duration of inspiration, 96.7% (87/90) versus 34.4% (31/90).

Conclusions: The BreatheSuite MDI can objectively detect several errors in the MDI technique, which were missed by observation alone. It has the potential to enhance treatment outcomes among patients with chronic lung diseases.

(JMIR Biomed Eng 2021;6(4):e26556) doi:10.2196/26556

KEYWORDS
digital health; asthma; smartphone; mHealth; mobile health; effective; observational; treatment; chronic obstructive pulmonary disease; inhaler; app

Introduction

Many medications used in treating asthma and chronic obstructive pulmonary disease (COPD) are administered via inhalation devices. They come in various forms, but the most common form is metered-dose inhalers (MDIs) [1-3]. These inhalers include salbutamol, fluticasone, and ciclesonide. These inhalers are primarily used to manage symptoms and prevent exacerbations. Both the Global Initiative for Asthma (GINA) [4] and the Global Initiative for Chronic Obstructive Lung Disease (GOLD) [5] recommends using inhalers to achieve
good symptom control, minimize future risk of exacerbations, and improve exercise tolerance.

The Aerosol Drug Management Improvement Team, which GINA supports, provided information on the proper way of using an MDI [6]. Furthermore, each drug manufacturer provides instruction on using an MDI and including it in the drug pamphlet. Nevertheless, studies have reported that up to 92% of patients with asthma demonstrate poor inhaler technique [7-9]. Sanchis et al [10] conducted a systematic review in 2016 to assess the most common errors in inhaler use among patients with asthma and those with COPD treated with MDIs. They concluded that incorrect inhaler use is unacceptably high outside clinical trials and does not seem to have improved over the past 40 years [10]. Similarly, Press et al [11] in 2011 examined the rates of inhaler misuse among patients with asthma and those with COPD; they concluded that misuse rates are prevalent among both groups. Improper inhaler technique can significantly affect the amount of medication reaching the lungs, leading to poor symptom control and more emergency department visits [12,13]. Errors in inhaler technique and nonadherence can affect medication delivery and decrease the benefits of taking the medication [14]. A systematic review of errors in the inhaler technique suggests that most reported errors were in coordination, speed of the inhalation, depth of inspiration, and no postinhalation breath-holding [10]. To improve the inhaler technique, researchers recommend frequent assessment of inhalation technique [15,16]. One of the literature’s current gaps concerns the most appropriate method to intervene if patients continue to misuse their inhalers [16]. Innovative technologies have been introduced, which may improve inhaler technique and consequently improve health outcomes [17-19].

Digital Health interventions have been increasing in the past 10 years, with significant advances in mobile apps, web portals, and electronic inhaler sensors. Researchers are now proposing digital health applications for many complex health conditions, including asthma and COPD. These technologies allow patients and health care providers to monitor and manage their symptoms more effectively. Several studies demonstrated improved clinical outcomes from implementing these technologies [17,19-21]. An example includes a meta-analysis that determined electronic reminders can improve patient adherence to inhaled corticosteroids by 19% [22]. One advantage of using digital health applications is having long-term data collection of symptoms, triggers, and inhaler use, which permits the identification of necessary changes to assist patients and their caretakers in understanding if symptoms are exacerbating [23,24].

The BreatheSuite MDI device is an auxiliary, add-on device which is connected to an approved MDI. It passively and quantitatively monitors important inhaler adherence and technique metrics, providing user feedback through a linked mobile app. This information may be used by patients to improve inhaler technique, and may also be shared with health care providers. There are several devices that monitor the adherence of MDIs [25], but there is a paucity in MDIs that monitor both adherence and technique. This study aims to validate the effectiveness of the BreatheSuite MDI device in assessing the technique of taking a dose via an MDI.

**Methods**

**Purpose**

This study aimed to validate the effectiveness of the BreatheSuite MDI device in assessing the technique of taking a dose via an MDI.

**The BreatheSuite MDI**

The BreatheSuite MDI device is an auxiliary, add-on device which is connected to an approved MDI. It is mounted by placing the device over the canister of a standard inhaler. The inhaler is used in the same way that it would be used without the BreatheSuite MDI Device. The device is designed for passive monitoring; it ensures that patients can continue to follow their prescriptions. The BreatheSuite MDI device is approximately 1 inch in diameter and attaches to a standard MDI canister with an elastic sleeve, as shown in the image below (Figure 1).

![Figure 1. Placebo inhaler with the BreatheSuite metered-dose inhaler device.](image-url)

The BreatheSuite MDI device has the potential to improve the technique and adherence of inhaled medications. It uses quantitative measures to assess several inhaler technique metrics. These inhaler technique metrics are based on the inhaler manufacturer instructions as well as quantitative analysis performed on lung deposition as a result of impacts of several
inhaler technique parameters. This has resulted in the quantitative analysis of the following:

1. Shaking duration (if the inhaler was shaken for 3 seconds): The duration of shaking was measured to within 0.1 seconds and rounded to the nearest 0.5 seconds; thus, any duration greater than 2.75 seconds would be considered acceptable.
2. Orientation (whether the device was oriented with the mouthpiece straight toward the back of the throat): The BreatheSuite device measures orientation from +1° to −1° and the acceptable range is from −10° to +20°, with 0° being the location where the mouthpiece is horizontal and points toward the back of the throat.
3. Press timing (if the user began inhaling before actuating the inhaler canister): The press timing is defined as the difference between the actuation time and the inhalation start time (each measured to within +0.1 and −0.1 seconds). The resultant value must be between −1.0 and +1.0 to be considered acceptable.
4. Inhalation duration (if the user has inhaled for at least 3 seconds): This value is defined as the difference between the inhalation end time and the inhalation start time, or actuation time, whichever is shortest. This value is then rounded to the nearest 0.5. Any duration above 3 seconds is considered acceptable.

This information is then transmitted to the BreatheSuite mobile app via Bluetooth and is securely uploaded to a remote server. The user will receive technique-related feedback and subsequent correcting advice by accessing the mobile app. The data are transferred from the app to a secure cloud database such that it is available for analysis and review. Health care providers can view the data to determine if additional training is necessary to improve inhaler technique.

Study Design

The study was a validation study.

Recruitment and Study Setting

We recruited patients from 2 community pharmacies in the province of Newfoundland and Labrador, Canada. One pharmacy was located in an urban area, while the other pharmacy was in a rural area. One pharmacist at each pharmacy was responsible for recruitment and data collection.

A convenience sample of 30 patients was recruited during their routine visits to the pharmacy. The pharmacist invited patients to participate in the study if they met the following eligibility criteria: being able to communicate in English, having a self-reported diagnosis of asthma or COPD, being 18 years or older, and having been prescribed an MDI and having used it in the past. First-time users of MDI were excluded.

Ethical Considerations

Ethical approval for this study was obtained from the Newfoundland and Labrador Health Research Ethics Authority. Before agreeing to participate, all subjects were informed about the nature of the study, potential risks and benefits, and their rights as research subjects. All participants completed a written consent form. They were also given a copy of the consent form.

Each participant and pharmacist were offered a gift card to compensate them for their participation time.

Data Collection

Before enrolling participants, all pharmacists were trained to follow the study protocol, including participant recruitment, baseline questionnaire, and scoring sheet. Each participant completed a baseline study questionnaire about their demographics and smartphone use (Multimedia Appendix 1). After completing the questionnaire, the pharmacist gave participants a BreatheSuite MDI device attached to a placebo MDI (Figure 1). Participants did not receive instructions on how to use the device. They were asked to take 3 doses in the same manner as they would use their inhalers. The BreatheSuite MDI device evaluated the following technique parameters:

1. Was the MDI properly shaken (at least for 3 seconds)?
2. Was the MDI in the upright position before taking the dose?
3. Was the MDI actuated after starting to take a breath?
4. Was the duration of the inhalation more than 3 seconds?

Pharmacists then used a scoring sheet to evaluate the technique of the participant using the MDI. The scoring sheet followed the same parameters that were evaluated by the BreatheSuite MDI (Multimedia Appendix 2). The pharmacists evaluated the technique parameters by using binary answers (yes/no). After the scoring was complete, pharmacists trained participants on how to use inhalers correctly. Each patient and pharmacist and patient will be offered a gift card of Can $20 (US $16.15) to compensate for their participation time.

Blinding

The pharmacist was blinded to the results from the BreatheSuite device. An independent researcher compared the results of the pharmacist’s scoring sheet with the results of the BreatheSuite device.

Statistical Analysis

A database of the questionnaire results was created using unique nonidentifying numbers. The information was password-protected. Before conducting the analysis, data were cleaned, coded, and entered into SPSS (version 25.0, IBM Corp). Unclear or incomplete survey items were flagged for queries. These were brought to the attention of the research team, each item was discussed, and a decision concerning its eligibility and entry was made. Baseline characteristics of participants were summarized with percentages for categorical variables and mean (SD) values for continuous variables. We reported the frequencies and percentages of technique errors from direct observations and the BreatheSuite MDI. Additionally, we calculated Cohen’s κ values to determine if there was agreement between the pharmacist’s observations and the BreatheSuite device.

An independent research analyst, who was not involved in recruiting and training the pharmacists, analyzed the results of the pharmacist scoring sheet with the scoring from the BreatheSuite MDI device. The analyst was able to link both data sets using the time of taking the dose. The pharmacist recorded the time, including seconds, in the scoring sheet. The time was also stored in the database automatically.
Results

Participant Characteristics
A total of 30 patients participated in the study. Approximately 60% (18/30) of participants were male. The mean age of participants was 56.5 (range 33-73) years. The highest level of education for most participants was a high school degree 46.6% (14/30). Half of the participants were living in a rural area 50% (15/30). Although more than half of the participants use mobile apps, only 16.6% (5/30) use health apps. Almost 80% (24/30) of participants did not use a spacer or aerochamber. Table 1 illustrates the participant characteristics.

Table 1. Participant characteristics (N=30).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>56.6 (10.5)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>18 (60)</td>
</tr>
<tr>
<td>Females</td>
<td>12 (40)</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
</tr>
<tr>
<td>General education diploma</td>
<td>3 (10)</td>
</tr>
<tr>
<td>High school</td>
<td>14 (46.6)</td>
</tr>
<tr>
<td>Bachelors</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Masters</td>
<td>1 (3.3)</td>
</tr>
<tr>
<td>PhD/MD/ID</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Other</td>
<td>12 (39.9)</td>
</tr>
<tr>
<td>Locality of residence, n (%)</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>15 (50)</td>
</tr>
<tr>
<td>Small</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Medium</td>
<td>2 (6.6)</td>
</tr>
<tr>
<td>Large</td>
<td>9 (30)</td>
</tr>
<tr>
<td>other</td>
<td>1 (3.3)</td>
</tr>
<tr>
<td>App use, n (%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>13 (43.3)</td>
</tr>
<tr>
<td>Yes</td>
<td>16 (53.3)</td>
</tr>
<tr>
<td>Do not know</td>
<td>1 (3.3)</td>
</tr>
<tr>
<td>Health app usage, n (%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>23 (76.6)</td>
</tr>
<tr>
<td>Yes</td>
<td>5 (16.6)</td>
</tr>
<tr>
<td>Do not know</td>
<td>2 (6.6)</td>
</tr>
<tr>
<td>Spacer/aerochamber use, n (%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>21 (70)</td>
</tr>
<tr>
<td>Yes</td>
<td>6 (20)</td>
</tr>
<tr>
<td>Do not know</td>
<td>1 (3.3)</td>
</tr>
<tr>
<td>I was instructed to use it but I do not use it</td>
<td>2 (6.6)</td>
</tr>
</tbody>
</table>

Comparison of Subjective and Objective Measures of Inhaler Technique
Each participant simulated taking 3 doses using the placebo MDI with the BreatheSuite MDI device, resulting in 90 doses for analysis. For each dose, we measured 4 technique metrics, which resulted in 360 measurements. The data recorded by the BreatheSuite MDI device showed that all participants performed at least 1 error in using the MDI. Among the metrics collected by both the pharmacist and the BreatheSuite MDI, the BreatheSuite MDI device captured 68.3% (246/360) of the errors made by the participant compared to 28.6% (103/360) errors captured by observation alone (Figure 2). The subjective and objective measures of inhaler technique included are shakes, orientation, coordination, and duration. The distribution of technique errors in MDI steps recorded by the BreatheSuite
MDI compared to errors reported by observation alone were as follows: shaking before actuation, 33.3% (30/90) versus 25.5% (23/90); upright orientation of inhaler during actuation, 66.7% (60/90) versus 18.87% (17/90); coordination (actuating after the start of inhalation), 76.6% (69/90) versus 35.5% (32/90); duration of inspiration, 96.7% (87/90) versus 34.4% (31/90).

Figure 2 highlights the percentage of errors in MDI steps recorded by the BreatheSuite MDI compared to errors reported through observation alone. The following are the Cohen $\kappa$ values: shaking (fair agreement, $\kappa = 0.283$), orientation (slight agreement, $\kappa = 0.183$), coordination (slight agreement, $\kappa = 0.071$), and duration (no agreement, $\kappa = -0.03$).

Discussion

Principal Findings

This study has indicated that the BreatheSuite MDI can objectively detect MDI technique errors, and in some cases can be better at assessing various quantitative metrics that are difficult to assess through observation alone. For example, measuring minor inhaler orientation angle deviations by objective observation can be difficult, along with minor differences in inhalation duration. In total, the BreatheSuite MDI device captured approximately 40% (143/360) more errors compared to observation alone.

The low Cohen $\kappa$ values and the objective nature of the BreatheSuite device potentially indicate that the BreatheSuite MDI performs better than observation alone. Objective measures of inhaler technique have the potential to become the new gold standard. Additional research is necessary to assess the relationship between objective improvements in inhaler technique with clinically significant improvements in health outcomes.

Comparison With Previous Work

Using technology to enhance inhaled medication plays a vital role in the management of chronic lung diseases. Several studies have assessed the effectiveness of electronic devices to an MDI to improve medication adherence [17-19]. However, there is a lack of studies and devices that assess the technique of using MDIs.

To further understand how patients use their inhalers, some pharmaceutical companies started manufacturing digital inhalers, such as the Digihaler. The Digihaler captures the inhalation rate when taking a dose, as well as opening and closing the inhaler cap. Studies conducted on Digihaler concluded that it could capture objective technique data [26]. These data may help identify clinically meaningful information early and facilitate physician-patient interventions and conversations [27]. It is important to consider the cost of these digital inhalers, especially when considering the usability and interoperability of using several digital inhalers from different drug manufacturers.

Some of our findings confirm those previously reported in the context of using electronic MDI devices to assess MDI technique. Our findings are in agreement with those of Biswas et al [28], who used an objective measure for inhaler technique and demonstrated that 100% of the patients made at least 1 error in using an MDI. As Biswas et al [28] noted, data recorded by an objective MDI device provides accurate measurements of MDI use, which could help evaluate how effectively patients use their MDIs. The major difference between the device used...
by Biswas et al [28] and the BreatheSuite device is that the latter does not require charging; the battery in the BreatheSuite device can operate for more than a year without charging [28]. This sustained battery life promotes usability through passive data collection.

Strengths and Limitations

The pharmacists were blinded to the BreatheSuite MDI data. An independent analyst compared the data between the BreatheSuite MDI and the pharmacist scoring sheet. Digital applications may be important in geographic locations with relatively large numbers of rural residents, such as Newfoundland and Labrador. Digital applications may enhance care provider access throughout sparsely populated rural areas as they can access information remotely. Half of the study sample was from a rural area, which supports the generalizability of our findings to rural areas.

There were also several limitations of note. Although the sample may not be generalizable to all patients with asthma and those with COPD; the study had broad inclusion criteria to resemble the target population. The BreatheSuite MDI does not track all the steps required to use an inhaler, such as exhaling before taking a dose and holding one's breath after inhalation. However, it takes the majority of the technique steps that have the potential to improve inhaler technique and supplement the teaching offered by health care professionals.

Implications for Practice and Future Research

This study provides insights into the effectiveness of the BreatheSuite MDI drive in capturing errors in MDI use. This information may help a variety of stakeholders (eg, health care providers, patients, administrators, and technology developers) who are planning to use an objective measure of MDI adherence and technique. The BreatheSuite MDI will transmit this information to the BreatheSuite mobile app. Patients can then receive technique-correcting advice by accessing the mobile app. In addition, health care providers can view the data to determine if additional training is necessary to improve the technique or adherence to using MDIs. These data can also be used to identify clinically meaningful information early such as (eg, rescue to controller usage ratios) and facilitate meaningful physician-patient conversations.

Digital applications that assess inhaled medications are increasingly gaining importance in managing chronic lung diseases [17-19]. These applications have the potential to improve health outcomes while reducing health care costs.

Future studies should examine the sustainability of behavior change following the use of the BreatheSuite MDI device. They should also include an experimental design to assess the BreatheSuite MDI’s effectiveness in improving clinical outcomes among patients diagnosed with chronic lung diseases. A larger and more heterogeneous sample with a longer follow-up period could confirm the study findings and expand the knowledge around the effectiveness of devices such as the BreatheSuite MDI.

Conclusions

Our findings potentially indicate that the BreatheSuite MDI can objectively detect several MDI technique errors missed by observation alone. The BreatheSuite MDI has the potential to enhance treatment and therefore improve outcomes among patients with chronic lung diseases. Additional studies are required to examine the effectiveness of the BreatheSuite MDI device on clinical outcomes.

Conflicts of Interest

BreatheSuite Inc funded this study. MFA and BV are full-time employees and shareholders at BreatheSuite Inc. MFA and BV were not involved in the analysis or reporting of the data. WA-A, an independent scientific consultant, was provided with the raw deidentified data to perform statistical analyses.

Multimedia Appendix 1
Baseline Questionnaire.
[PDF File (Adobe PDF File), 47 KB - biomedeng_v6i4e26556_app1.pdf ]

Multimedia Appendix 2
Pharmacist Scoring Sheet.
[PDF File (Adobe PDF File), 47 KB - biomedeng_v6i4e26556_app2.pdf ]

References


6. pDMI. InHalers4U. URL: https://www.inhalers4u.org/index.php/pmdi/ [accessed 2021-03-10]


Abbreviations

- **COPD**: chronic obstructive pulmonary disease
- **GINA**: Global Initiative for Asthma
- **GOLD**: Global Initiative for Chronic Obstructive Lung Disease
- **MDI**: metered-dose inhaler

©Meshari F Alwashmi, Gerald Mugford, Brett Vokey, Waseem Abu-Ashour, John Hawboldt. Originally published in JMIR Biomedical Engineering (http://biomsedeng.jmir.org), 03.11.2021. 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 https://biomsedeng.jmir.org/, as well as this copyright and license information must be included.
Understanding “Atmosome”, the Personal Atmospheric Exposome: Comprehensive Approach

Hari Bhimaraju¹; Nitish Nag¹, PhD; Vaibhav Pandey¹, PhD; Ramesh Jain¹, PhD
Donald Bren School of Information and Computer Sciences, University of California, Irvine, CA, United States

Corresponding Author:
Hari Bhimaraju
Donald Bren School of Information and Computer Sciences
University of California
Donald Bren Hall, 6210
Irvine, CA, 92697
United States
Phone: 1 949 824 7427
Fax: 1 949 824 3976
Email: hari.bhimaraju@columbia.edu

Abstract

Background: Modern environmental health research extensively focuses on outdoor air pollutants and their effects on public health. However, research on monitoring and enhancing individual indoor air quality is lacking. The field of exposomics encompasses the totality of human environmental exposures and its effects on health. A subset of this exposome deals with atmospheric exposure, termed the “atmosome.” The atmosome plays a pivotal role in health and has significant effects on DNA, metabolism, skin integrity, and lung health.

Objective: The aim of this work is to develop a low-cost, comprehensive measurement system for collecting and analyzing atmosomic factors. The research explores the significance of the atmosome in personalized and preventive care for public health.

Methods: An internet of things microcontroller-based system is introduced and demonstrated. The system collects real-time indoor air quality data and posts it to the cloud for immediate access.

Results: The experimental results yield air quality measurements with an accuracy of 90% when compared with precalibrated commercial devices and demonstrate a direct correlation between lifestyle and air quality.

Conclusions: Quantifying the individual atmosome is a monumental step in advancing personalized health, medical research, and epidemiological research. The 2 main goals in this work are to present the atmosome as a measurable concept and to demonstrate how to implement it using low-cost electronics. By enabling atmosome measurements at a communal scale, this work also opens up potential new directions for public health research. Researchers will now have the data to model the impact of indoor air pollutants on the health of individuals, communities, and specific demographics, leading to novel approaches for predicting and preventing diseases.

(JMIR Biomed Eng 2021;6(4):e28920) doi:10.2196/28920

KEYWORDS
exposome; exposomics; personal health; indoor air quality; health state estimation; health informatics; public health policy; epidemiology; embedded systems; internet of things

Introduction

At any moment in time, health is affected by various internal and external factors, such as the genome, microbiome, and exposome. The exposome consists of everything an individual is exposed to across his or her lifespan [1]. It considers lifestyle, occupation, socioeconomic factors, and the environmental conditions in which people live to develop an in-depth understanding of how an individual’s surroundings impact his/her health. The atmospheric exposome, a subset of the complete exposome and which is presented in this work, focuses on the health effects from the air that people breathe.

The term “atmosome” was coined to describe the atmospheric subset of an individual’s exposome. Common indoor air pollutants include PM$_{2.5}$ (particulate matter with diameter of ≤2.5 μm), PM$_{10}$ (particulate matter with diameter of ≤10 μm),
carbon dioxide (CO$_2$), nitrogen dioxide (NO$_2$), carbon monoxide (CO), volatile organic compounds (VOCs), ozone (O$_3$), liquid petroleum gas (LPG), natural gas (NG), formaldehyde (HCHO), and biological contaminants such as bacteria and fungi. Measuring the quality of indoor air can provide insights into the potential adverse effects of poor air quality and preventative measures to keep the exposome cleaner. A cleaner atmosome, in turn, has a positive impact on health and well-being. Thus, there is a need for a portable, real-time, multichannel air measurement system for enabling data-driven analytics and research [2].

A number of studies have been conducted by several federal, state, and local agencies that monitored, collected, and stored outdoor air quality data in the Environmental Protection Agency’s (EPA) Air Quality System database [3,4]. The Environmental Defense Fund in collaboration with Google Earth [5], the World Health Organization Global Urban Ambient Air Pollution Database [6], the World Air Quality Historical Database [7], and many other organizations generated air quality maps. These data are used for various modeling studies, to review policy implementation plans, and to generate reports for the Congress (US) [5]. However, these agencies have overlooked similar quantitative studies of indoor air quality (IAQ).

Previous studies have shown that indoor air is much more polluted than outdoor air and represents a major public health challenge especially in developing countries [8]. Many studies focused on specific or limited indoor air pollutants such as VOCs, CO$_2$, PM$_{2.5}$, PM$_{10}$, and halogen flame-retardants [9-12]. Further, though potentially unexpected, there exist a myriad of well-defined sources of indoor air contamination and, correspondingly, numerous contaminants [13]. To improve air quality and minimize pollution-related disease and mortality, the atmosome must be defined, measured, and analyzed to mitigate adverse environmental conditions and improve health outcomes. Thus, the motivation for this work is to use the atmosome to further personalize an estimation of individual health from multimodal data [14,15].

According to the EPA, indoor air pollution is one of the top 5 environmental risks to public health. Annually, 9 out of every 10 people breathe air containing high levels of pollutants at some point [16,17]. Air pollutants can be in the form of pet dander, mold, dust mites, CO, radon, pests, lead, and secondhand smoke [17]. Americans spend approximately 90% of their time indoors, where the concentrations of some airborne pollutants are 2-5 times higher than those outdoors [13]. This poor IAQ can cause various infections, lung cancer, and chronic lung diseases, including asthma [18]. It can also contribute to the development of atherosclerosis, a root cause of many cardiovascular diseases [19]. In 2020, a long-term study of over 63 million US adults indicated a surprising correlation between PM$_{2.5}$ and hospitalizations for severe neurological diseases [20].

According to the State of Global Air 2020 [21], nearly 500,000 newborns died in 2019 in their first month of life due to exposure to all types of air pollutants described previously. Households inhalation of mold spores and infant pulmonary hemorrhage are found to be linked in some studies [22]. Air pollution even impacts children while they are in their mothers’ womb [23,24], with the effect of air pollution on pregnant women and their fetuses comparable to smoking tobacco [21]. People often assume that indoor spaces are safe from outdoor air pollution, but this is inaccurate. Therefore, IAQ is a significant threat to public health [25]. The Program Needs for Indoor Environments Research (PNIER) document details EPA’s research needs for the indoor environment and recommends that the EPA and other governmental and private sector agencies and organizations address this issue [26].

The environmental research field of IAQ is nascent. Nonetheless, several researchers have recently demonstrated portable gas detection systems using various sensor technologies covering a limited set of analytes. For example, MQ sensors were embedded in a system where VOCs were detected [27]. MQ sensors are well-known to exhibit acceptable selectivity, but low sensitivity. Interestingly, the researchers implemented an artificial neural network and dramatically improved the sensitivity to gas concentrations at single-digit parts per million (ppm). However, the supported analyte set is far too small given the contaminants of concern, as reported by the EPA. A related research effort embedded photoionization detectors (PIDs) into a portable system for the detection of isobutylene, ethanol, propanol, and acetone [28]. PID sensors are well-known to exhibit high sensitivity (ie, on the order of parts per billion [ppb]), but lower selectivity than MQ sensors. Further, that work demonstrated only a small set of detected analytes. Other recent efforts include a portable system with embedded gas chromatography PID sensors [29]. This system detects benzene, toluene, and xylene. The researchers also employed an elegant algorithm using various quantification parameters (eg, pumping time, temperature) and calibration curves to optimize selectivity. This system also requires a pumping time up to 90 seconds and an analysis time of 10 minutes. Nonetheless, as expected, very high sensitivity was achieved, but again the analyte set is small and the system cost is substantially higher than an implementation with low-cost devices, such as MQ sensors. Last, none of these research efforts considered management of the data in an actionable manner. The systems reported were to demonstrate selectivity and sensitivity, the 2 most critical metrics for gas sensors.

This recent research focused on utilizing more common sensor technologies for portable applications. However, more exotic sensors have been developed recently in portable gas detection systems that have been demonstrated for indoor and outdoor use. For example, in [30], a mobile microscopy system (coined as the c-Air device) is presented and utilizes microscopy as a sensing technique and includes machine learning algorithms to increase accuracy. Further, it includes a mobile software app for data display. The device requires a sample of 6.5 L of air and an analysis time of 30 seconds. Also, similar to related studies, the c-Air device supports only a limited set of analytes, including total suspended particulates, PM$_{10}$, and PM$_{2.5}$. Nonetheless, very good results are achieved and it is an advancement that the system is linked to a mobile app, though nothing actionable is reported.
In [31], a so-called portable cyber-physical system is presented for gas detection and is embodied in 2 distinct architectures, a stationary and portable device, each using well-established electronics related to this work, including an Arduino microcontroller and Raspberry Pi system-on-chip (SoC), corresponding to each embodiment. The 2 systems also use MQ sensors. However, the only sensors supported by the system are an MQ-4 (a methane sensor) and an MQ-8 (a hydrogen sensor). Further, although the system supports connectivity to the internet, the functionality is to merely upload the data to cloud storage. No manipulation or presentation of the data is reported.

Basic IAQ commercial-off-the-shelf (COTS) products also exist. Xiaomi, for example, offers a countertop product which detects PM$_{2.5}$, total VOC (tVOC), CO$_2$, temperature, and humidity [32]. It includes a touch screen and Wi-Fi connectivity. The company also offers handheld products that measure individual analytes, such as PM$_{2.5}$ [33]. Further, Xiaomi offers air purifiers that can be controlled by a mobile phone app and are marketed under the brand Mi. There exist many other related COTS products for IAQ monitoring. Although interesting, these products measure a very small set of analytes and offer minimal, if any, actionable information based on the data collected.

In contrast to both the recent research and commercial work, this paper introduces a patent pending and low-cost embedded system called the Atmosome Measurement System (AMS) [34], which can reliably, accurately, and instantaneously monitor and measure significantly more indoor air pollutants such as PM$_{2.5}$, PM$_{10}$, CO$_2$, NO$_2$, CO, VOCs, O$_3$, LPG, NG, equivalent CO$_2$ (eCO$_2$), hydrogen as well as environmental parameters including temperature, humidity, pressure, and altitude. This set of analytes was chosen as it represents the primary sources of indoor air pollution as well as the leading causes of adverse impact on human respiratory health, according to the EPA. Further, the presented work is implemented as a scalable and low-cost embedded system utilizing COTS electronics including an Arduino microcontroller, Raspberry Pi SoC, MQ gas sensors, and simple environmental sensors. MQ sensors were selected for low cost, high selectivity, and sufficient sensitivity. Besides, in contrast to previous studies [10,12], AMS is built with the goal of providing a cloud-based infrastructure that stores, analyzes, and presents insights into IAQ and trends that correlate with personal lifestyles. It displays historic and real-time data from multiple sensors in a user-friendly web application, enabling users to interpret their data, and recommends environmental changes to improve personal atmosome conditions.

Therefore, the development of a system that can evaluate IAQ by using multiple analytes, process and visualize pollutant data, recommend remediation steps, and be built at an affordable price point is the foundation of this research. The system can be configured with a variety of optional customizations including the frequency at which the users would like to monitor their air quality; their geographic location details, including the zip code, city, state, and country; the indoor space details such as home, office, or car; the location within the space such as kitchen, bedroom, garage; and the activity details such as cleaning, cooking, routine. Further, AMS supports representational state transfer (REST) application programming interface (API) to download data for further exploration and analytics. AMS also includes an option for the users to anonymously share their data to further indoor air pollution research. This opens the possibility of developing a public indoor air quality database while maintaining user confidentiality allowing for extended research on indoor air quality, its impacts, and health policy modeling.

Taken together, AMS can be a useful tool for improving public health outcomes as it can provide the necessary data that people need to manage their IAQ in a cost-effective and convenient manner. Moreover, the data can display different atmosomes between 2 neighbors or within neighborhoods of different socioeconomic classes, which can be useful for public health officials or policy researchers that work toward enhancing the health of citizens.

Comparisons of AMS with various COTS products are available in Multimedia Appendix 1 [35-41]. AMS stood out in both number of analytes and cost compared with the nearest COTS product, Aeroqual, that covers multiple analytes [35]. Comparisons are also made with recent studies and those details are available in Multimedia Appendix 2 [42-47]. Aspects such as data sampling duration and pollutant streams in AMS are found to be much more extensive than similar indoor air quality assessments in a college campus [44] and homes in a temperate region of the United States [46]. Comparisons are not made with the cited research because the analyte sets are dramatically smaller than those supported by AMS and there is no actionable interpretation of the data in the studies, because they focus primarily on sensor selectivity and sensitivity, as described previously. Further, few of those efforts included internet connectivity and none included any actionable information based on the data collected.

Methods

Study Approach and Design

An experimental approach combining internet of things (IoT) hardware and software development was used to measure air pollutants and air quality metrics. The researchers used AMS to nonintrusively monitor air quality through daily indoor life. AMS provided visuals and recorded trends that could indirectly indicate the relationship between lifestyle and observed pollutant values.

In this work, the researchers used AMS to collect air quality data indoors at home (in Cupertino, California; South Lake Tahoe, California; and Hyderabad, Telangana, India) as well as during local commute in the United States and during an inflight journey from the United States to India. These locations were chosen to monitor and evaluate the performance of AMS in environments associated with distinct indoor air quality profiles and climates. The device, similar to a thermostat placed in a room, is completely noninvasive. To initiate readings, it is powered on with a USB cable (or a wall socket) and AMS software is launched by the user. The studies were conducted, intermittently, between January 2020 and January 2021.

After initial calibration, to ensure the continued accuracy of the measurements made by AMS on an ongoing basis, the sensors...
were recalibrated every 3 months. Measurements were repeated over similar activities in several different timeframes to better analyze and predict possible relationships between activities and associated pollutants. To control for bias, measurements were also carried out across several recalibrations.

System Architecture
AMS comprises 3 distinct modules that collect, analyze and store, and display data. As shown in Figure 1, these 3 interconnected modules include the data collection system (DCS), data analysis and storage engine (DASE), and user interface (UI). The DCS is a sensor system that collects the user’s environmental data in user-defined intervals. The data gathered are then analyzed and stored by the system’s DASE, and the air quality metrics obtained are displayed by the system’s web-based UI.

Materials and Software
Materials used to build the DCS and the software used to create DASE and UI are discussed in the following sections.

Data Collection System
Overview
The DCS is an IoT air monitoring sensor system for acquiring an individual’s unique geospatial data to track air quality. It includes 17 environmental sensors that measure 22 different air pollutant data streams, an Arduino Mega, a Raspberry Pi, and a power source. Referring back to Figure 1, the Arduino Mega is a microcontroller that captures sensor data and the Raspberry Pi is an SoC which supports Wi-Fi and enables an interface to the cloud.

Sensors
The DCS monitors environmental conditions such as temperature, pressure, altitude, humidity, and various analytes including PM\textsubscript{2.5}, PM\textsubscript{10}, CO, O\textsubscript{3}, CO\textsubscript{2}, eCO\textsubscript{2}, tVOCs, LPG, methane, hydrogen, flammable gases, aromatic compounds, hydrogen sulfide, ammonia, nitrogen oxide, NG, and HCHO as shown in Table 1. Metal oxide semiconductor (MOS) gas sensors were used to detect the different air pollutants. MOS-based sensors detect the concentration of various kinds of gases by acting as a chemiresistor, where a change in resistance of the metal oxide occurs due to the adsorption of specific gases. These sensors, and specifically MQ series sensors, are ideally suited for low-cost and low-power applications in indoor environments. Selectivity to certain gases is dependent on the specific sensor model, which is indicated numerically. Some MQ sensors are sensitive to multiple gases; for example, both MQ-5 and MQ-6 measure LPG, but MQ-6 exhibits higher selectivity and sensitivity to LPG and is calibrated for that particular gas. As individual sensors are calibrated for their specific gases, they are less selective to other gases. A total of 17 different kinds of sensors were used to detect and measure the level of pollutants as well as environmental parameters (Table 1).
Table 1. List of sensors, including target gas analytes and parameters as well as sensor type, in the data collection system implementation. The system supports 22 data streams from 17 different sensors.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Analyte/parameter</th>
<th>Sensor type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCS811/BME280</td>
<td>Total volatile organic compounds, equivalent carbon dioxide, temperature, humidity, pressure, altitude</td>
<td>Environmental combo sensor</td>
</tr>
<tr>
<td>CO₂</td>
<td>Carbon dioxide</td>
<td>Nondispersive infrared</td>
</tr>
<tr>
<td>PM₂.₅</td>
<td>Particulate matter 2.5</td>
<td>Optical, infrared-emitting diode</td>
</tr>
<tr>
<td>MQ2</td>
<td>Smoke/particulate matter 10</td>
<td>MOS⁴</td>
</tr>
<tr>
<td>MQ4</td>
<td>Methane</td>
<td>MOS gas sensor</td>
</tr>
<tr>
<td>MQ6</td>
<td>Liquefied petroleum gas</td>
<td>MOS gas sensor</td>
</tr>
<tr>
<td>MQ7</td>
<td>Carbon monoxide</td>
<td>MOS gas sensor</td>
</tr>
<tr>
<td>MQ131</td>
<td>Ozone</td>
<td>MOS gas sensor</td>
</tr>
<tr>
<td>MQ3</td>
<td>Alcohol (ethyl alcohol)</td>
<td>MOS gas sensor</td>
</tr>
<tr>
<td>MQ5</td>
<td>Natural gas</td>
<td>MOS gas sensor</td>
</tr>
<tr>
<td>MQ8</td>
<td>Hydrogen</td>
<td>MOS gas sensor</td>
</tr>
<tr>
<td>MQ9</td>
<td>Flammable gases</td>
<td>MOS gas sensor</td>
</tr>
<tr>
<td>MQ135</td>
<td>Aromatic compounds</td>
<td>MOS gas sensor</td>
</tr>
<tr>
<td>MQ136</td>
<td>Hydrogen sulfide</td>
<td>MOS gas sensor</td>
</tr>
<tr>
<td>MQ137</td>
<td>Ammonia</td>
<td>MOS gas sensor</td>
</tr>
<tr>
<td>NOₓ</td>
<td>Nitrogen oxides</td>
<td>MOS gas sensor</td>
</tr>
<tr>
<td>HCHO</td>
<td>Formaldehyde</td>
<td>MOS gas sensor</td>
</tr>
</tbody>
</table>

*MOS: metal oxide semiconductor.

Realized AMS

Figure 2 presents the complete realized AMS including sensors and the Raspberry Pi platform, which are securely mounted to an FR4 printed circuit board (PCB) and an Arduino Mega 2560 microcontroller reverse mounted to the same. The Arduino Mega microcontroller is programmed to capture data from the sensors. All of these system components and associated positions on the PCB are labeled clearly in Figure 2.

Figure 2. A photograph of the populated PCB with the Arduino (reverse mounted) and 17 sensors listed in Table 1. CO: carbon monoxide; CO₂: carbon dioxide; H₂S: hydrogen sulfide; HCHO: formaldehyde; LPG: liquid petroleum gas; NG: natural gas; NH₃: ammonia; PCB: printed circuit board; PM₂.₅: particulate matter with diameter of 2.5 μm or less; PM₁₀: particulate matter with diameter of 10 μm or less; VOC: volatile organic compound.
IoT

The sensor PCB is connected using a USB cable linked to a computer (either a Raspberry Pi or a laptop that runs the DCS software). The computer receives sensor data from Arduino Mega microcontroller and posts it to the cloud (DASE) server or saves it in local memory in the absence of an internet connection. The PCB comes in 2 variations: a portable model that consists of 8 different sensors providing 13 data streams and powered by the USB; and a high-power model that has 17 sensors providing 22 data streams and which must be powered by a wall socket. Future work will aim to make the high-power model portable.

Data Storage and Analysis Engine

DASE runs in the cloud on Amazon Web Services (AWS) and stores the sensor data in a Postgres Database. AWS was selected for simplicity of implementation, low cost, and well-known user data security as shown in a recent case study [48]. Further, no personally identifiable information is collected for this study. The username and zip code are stored in the cloud and password protected. The software implementation is built using the Python Flask Framework and supports REST API to support receiving data from the sensor board at user location (DCS) and for sending data to the UI or alerts to the user. The source module layout and libraries are shown in Figure 3.

Figure 3. DASE (data analysis and storage engine) programming components and functional flow diagram.

The backend of the system is based on Python and Flask and has 3 distinct interconnected components that make up the Atmosome engine: the web application framework, the database model that receives data from the cloud database, and the numerical analysis libraries that operate on the data received from the user’s device and surface them to the user via the front end UI. To accomplish this workflow, DASE employs various Flask web framework components including flaskmigrate (to support modifications to existing DB tables), sqlalchmey (an object-relational mapper that enables reading from/writing to the DB python data objects without the need for using the DB’s SQL), flasksqlalchmey (to manage DB connections and sessions), flaskcors (to enable cross-domain REST API communication between the server and the UI), flasmashmallow (to handle python-json serialization/deserialization for REST API communication), numpy (Python’s multidimensional numerical analysis library), pandas (Python’s DataFrame support much like a database table...
in memory), and matplotlib (Python’s data visualization library). flaskproject.models defines the DB tables in python, flaskproject.config defines the web application configurations, for example, the type of database being used, and flaskproject.app is the code that handles all data computations.

**User Interface**

The UI is built as a progressive web application using the React Framework. This makes it available on any device with a web browser. Additionally, because it is a progressive web app, it automatically adjusts to the size of mobile platforms, and thereby presents a user experience similar to a native app. The source code is structured and modularized into pages and components within pages.

For example, Figure 4 illustrates the 6 javascript (.js) source files that render the various sections of the Atmosome dashboard. header.js is the header of the page, and displays the page title, user location, date, and time. gauge.js displays the AQI gauge which gives an “at-one-glance” state of the current indoor air conditions. weatherParameters.js displays the temperature, humidity, pressure, and altitude. Although weather metrics are not pollutants, they are part of lifestyle conditions and are recorded along with information on pollutants. pollutants.js displays numeric and visual information about each of the pollutants. averageExposure.js displays the user’s average exposure to various pollutants in different intervals of time. timeseries.js displays the quarterly time series graphs of the user’s historic exposure to various pollutants. The UI is designed mainly to render data and does not store data or perform computations. It makes REST API calls to DASE to retrieve the data.

**Procedure**

In this section, the methods involved in calibrating the DCS device, collecting data from a setting, and posting data to DASE will be discussed. Moreover, the analyzed data can be viewed on UI screens for the given set up.
Sensor Calibration

The sensors in AMS’s DCS are first “burned-in,” meaning that they are placed in an environment with clean air and operated with active power for 48 hours. Next, AMS’s DCS is run through a meticulous calibration process to ensure accuracy. A flow diagram of the calibration algorithm is illustrated in Figure 5. AMS DCS is calibrated for several variables such as CO₂, tVOC, LPG, PM₂.₅, PM₁₀, and others as indicated. During the initial set up, this process is repeated daily and accumulated over a period of 15 days to train a linear regression model to predict the values. The calibration code is run on the sensor’s board Arduino Mega microcontroller, and the data are accumulated on the Pi system.

Figure 5. DCS calibration algorithm flow diagram. CO: carbon monoxide; CO₂: carbon dioxide; DCS: data collection system; LPG: liquefied petroleum gas; NG: natural gas; O₃: ozone; PM₂.₅: particulate matter with diameter of 2.5 μm or less; PM₁₀: particulate matter with diameter of 10 μm or less; tVOC: total volatile organic compound.

To ensure the continued accuracy of the new system, the sensors were recalibrated every 3 months, and new training improves the accuracy of the data correction algorithms. Upon continued comparison of the DCS postcalibration values with precalibrated COTS devices, over 90% accuracy postcalibration was achieved when compared with precalibrated COTS devices. Alternative calibration approaches, such as calibrating outdoors against the values of the nearest outdoor weather station, were explored and similar accuracy was noted.

Data Collection

Once the user receives the portable DCS, they need to power it and connect it to a Wi-Fi network, if available. Then they need to launch the application. Once launched, the user can change any of their default settings, or retain the defaults, and initiate collection of air quality data.

Although AMS is intended to measure IAQ, part of the calibration routine was performed outdoors. This decision was made to compare results easily with the nearest, trusted outdoor air pollution station. Further, the environmental differences between the indoor and outdoor settings were not notable enough to introduce a substantial error. Certainly, an entirely indoor calibration routine would yield higher sensor accuracy, but precise indoor calibration techniques are complex and not suitable for consumer use. For example, precision routines to calibrate gas sensors are typically performed in the presence of a high concentration source of each analyte. Such an approach was deemed unrealistic and prohibitively expensive.

The researcher (HB) used AMS to monitor air quality through daily life at home, such as in the kitchen, living room, bedroom, home office, and during commute. The researcher also measured air quality during air travel to India, and at home in India and the United States.

Data Transmission

Local Storage and Transmission to Cloud

In the presence of Wi-Fi, as the DCS program collects air quality data from the sensors at the frequency specified by the user, it concurrently transmits the data to the cloud and requires no user interaction. In the absence of Wi-Fi, the DCS stores the data in local memory. Once a Wi-Fi connection is available, the user can connect the DCS to the Wi-Fi, and select an icon in the UI that indicates “Upload AMS data to the Cloud.” This loads the data to the cloud, and once complete, deletes the locally stored data.
data, automatically removing the no-longer required data from local storage.

**Post/Receive Data to/From Cloud**

The DCS software wraps the readings from each of its sensors into JSON, a simple format, with a series of key value pairs, used to store and transmit data to DASE, which is running in the cloud. Once a JSON payload with the values of each pollutant, atmospheric data, and sampling location information has been assembled (a portion of which is shown in Figure 6) it is ready to be transmitted via the internet to DASE.

**Figure 6.** Data transmission: REST API POST data from DCS to DASE containing sensor types and values. API: application programming interface; DASE: data analysis and storage engine; DCS: data collection system; REST: representational state transfer.

The DCS software uses REST APIs to interact with DASE. REST APIs use HTTP requests to interact with a remote web server. The HTTP GET method is used to receive data from the server, and the HTTP POST method is used to send data to the remote server.

The DCS software sends this JSON payload by making a REST API POST request to DASE, which receives the data, and subsequently analyzes and stores them. Upon receiving an HTTP GET request from the UI to display data, DASE formats the information required by the UI into the JSON payload and sends it, as illustrated in Figure 7. The UI then displays these data.
Data Storage

DASE stores the received sensors data in its Postgres database. Postgres is a free database and is seamlessly integrated into DASE’s Python Flask framework to store and retrieve data using REST API calls. Currently, there is no limit on the amount of time the user’s history data are stored in DASE and there is no user action involved in this step.

AQI Calculation

General Equation

The Indoor Air Quality Index is calculated using the weighted mean formula. The contribution of each pollutant is multiplied by its weightage, whose calculation is explained in the next section, and divided by the sum of pollutant weightages.

This can be represented by the formula:

\[
AQI = \frac{\sum_{i} W_i P_i}{\sum_{i} W_i}
\]

where \( W_i \) is weightage of pollutant \( i \) and \( P_i \) is the reading of pollutant \( i \).

Calculation of Pollutant Weightages

Weightages for each of the pollutants have been calculated based on the concentrated means and their contribution to different AQI levels.

Individual AQIs (Table 2) and the breaking points for the concentration mean of different pollutants in a fixed cycle were used to arrive at the weights of each pollutant at the respective AQI levels. Weightage of each pollutant’s mean concentration at each AQI level was calculated by measuring the fractional contribution to the AQI.
Table 2. Individual Air Quality Indexes and the breaking points for the concentration mean of pollutants [49]a,b.

<table>
<thead>
<tr>
<th>Indoor Air Quality Index</th>
<th>Sulfur dioxide 24 hours</th>
<th>Sulfur dioxide 1 hour c</th>
<th>Nitrogen dioxide 24 hours</th>
<th>Nitrogen dioxide 1 hour c</th>
<th>PM2.5 24 hours</th>
<th>PM10 24 hours</th>
<th>Carbon monoxide 24 hours</th>
<th>Carbon monoxide 1 hour c</th>
<th>Ozone 1 hour</th>
<th>Ozone 8 hours</th>
<th>PM2.5 e 24 hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>150</td>
<td>40</td>
<td>100</td>
<td>50</td>
<td>2</td>
<td>5</td>
<td>160</td>
<td>75</td>
<td>100</td>
<td>35</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>500</td>
<td>80</td>
<td>200</td>
<td>150</td>
<td>4</td>
<td>10</td>
<td>200</td>
<td>75</td>
<td>300</td>
<td>115</td>
</tr>
<tr>
<td>150</td>
<td>475</td>
<td>650</td>
<td>180</td>
<td>700</td>
<td>250</td>
<td>14</td>
<td>35</td>
<td>300</td>
<td>115</td>
<td>215</td>
<td>265</td>
</tr>
<tr>
<td>200</td>
<td>800</td>
<td>800</td>
<td>280</td>
<td>1200</td>
<td>350</td>
<td>24</td>
<td>60</td>
<td>400</td>
<td>150</td>
<td>265</td>
<td>250</td>
</tr>
<tr>
<td>300</td>
<td>1600</td>
<td>f</td>
<td>565</td>
<td>2340</td>
<td>420</td>
<td>36</td>
<td>90</td>
<td>800</td>
<td>800</td>
<td>100</td>
<td>350</td>
</tr>
<tr>
<td>400</td>
<td>2100</td>
<td>f</td>
<td>750</td>
<td>3090</td>
<td>500</td>
<td>48</td>
<td>120</td>
<td>1000</td>
<td>g</td>
<td>350</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>2620</td>
<td>f</td>
<td>940</td>
<td>3840</td>
<td>600</td>
<td>60</td>
<td>150</td>
<td>1200</td>
<td>g</td>
<td>500</td>
<td></td>
</tr>
</tbody>
</table>

aData presented are mean values.
bSulfur dioxide (not collected by AMS) and nitrogen dioxide, which are primarily outside pollutants, are excluded from the indoor Air Quality Index calculation. Carbon dioxide and volatile organic compounds are much more common indoors and are more relevant and considered in the indoor Air Quality Index calculation.
cThe concentration means of 1-hour sulfur dioxide, nitrogen dioxide, and carbon monoxide just adapt to the real-time calculation for Indoor Air Quality Index, but the concentration means of 24-hour sulfur dioxide, nitrogen dioxide, and carbon monoxide were used to calculate for a whole day.
dPM10: particulate matter with diameter of ≤10 μm.
ePM2.5: particulate matter with diameter of ≤2.5 μm.
fThe concentration mean of 1-hour sulfur dioxide higher than 800 μg/m³ is calculated with the concentration mean of 24-hour sulfur dioxide.
gThe concentration mean of 8-hour ozone higher than 800 μg/m³ is calculated with the concentration mean of 1-hour ozone.

tVOC Information
Values from Table 3 have been considered for the corresponding AQI windows, with the midpoint of concentration breaking points chosen as pollutant representation, which is used as the denominator, and the breaking point of the IAQI range as the numerator, similar to calculations performed for other pollutants.

Table 3. Individual AQIs and the breaking points for the concentration mean of VOCs and others [50].

<table>
<thead>
<tr>
<th>Level</th>
<th>AQI a range</th>
<th>VOC b (µg/m³) concentration (BP LO–BP HI)e</th>
<th>CO d (µg/m³) concentration (BP LO–BP HI)</th>
<th>PM2.5 (µg/m³) concentration (BP LO–BP HI)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0-50</td>
<td>0-200</td>
<td>0-4.99</td>
<td>0-30</td>
<td>Good</td>
</tr>
<tr>
<td>B</td>
<td>51-100</td>
<td>201-350</td>
<td>5-9.99</td>
<td>31-90</td>
<td>Moderate</td>
</tr>
<tr>
<td>C</td>
<td>101-250</td>
<td>351-500</td>
<td>10-14.99</td>
<td>91-140</td>
<td>Unhealthy</td>
</tr>
<tr>
<td>D</td>
<td>251-400</td>
<td>501-757</td>
<td>15-2000</td>
<td>141-750</td>
<td>Very unhealthy</td>
</tr>
</tbody>
</table>

aAQI: Air Quality Index.
bVOC: volatile organic compound.
cBP: breaking point (LO: low; HI: high).
dCO: carbon monoxide.
ePM: particulate matter.

CO₂ Information
Values from Table 4 have been considered for the corresponding AQI windows, with the midpoint of pollutant concentration calculated and chosen as pollutant representation, which is used as the denominator, and the breaking point of the IAQI range as the numerator, as is the case with others.
Table 4. Individual Air Quality Indexes and the pollutant concentration ranges of carbon dioxide and others [51].

<table>
<thead>
<tr>
<th>Pollutants</th>
<th>Indoor Air Quality Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon dioxide (ppm)</td>
<td>0-0.2, 0.21-2, 2.1-9, 9.1-15.4, 15.5-30.4, 30.5-50.4</td>
</tr>
<tr>
<td>Carbon monoxide (ppm)</td>
<td>0-379, 380-450, 451-1000, 1001-5000, 5001-30,000, 30,001-40,000</td>
</tr>
<tr>
<td>Hydrogen (ppm)</td>
<td>0-24, 1.1-2, 2.1-3, 3.1-5, 5.1-8, 8.1-10</td>
</tr>
<tr>
<td>Ammonia (ppm)</td>
<td>0-0.49, 0.5-10, 11-49, 51-100, 101-400, 401-500</td>
</tr>
<tr>
<td>Ethanol (ppm)</td>
<td>0-0.00033, 0.0034-1.5, 1.6-5, 11-49, 101-400, 401-500</td>
</tr>
<tr>
<td>Hydrogen sulfide (ppm)</td>
<td>0-0.0247, 0.0248-0.6, 0.7-1.6, 11-49, 101-400, 401-500</td>
</tr>
<tr>
<td>Toluene (ppm)</td>
<td>19-20.9, 19-20.9, 15-19, 12-15, 10-12, &lt;10</td>
</tr>
<tr>
<td>Oxygen (%)</td>
<td>20.95, 19-20.9, 15-19, 12-15, 10-12, &lt;10</td>
</tr>
<tr>
<td>Indoor Air Quality Index</td>
<td>0-50, 51-100, 101-150, Unhealthy for sensitive individuals, 151-200, Very unhealthy, 201-300, Hazardous</td>
</tr>
</tbody>
</table>

The final weightage factor of each pollutant was calculated by taking the arithmetic mean as shown in Tables 5-10.

**Pollutant Weightage Formula**

\[
W_p = \frac{\sum \text{Indoor Air Quality Index} \times \text{PM} \text{ or CO}}{\sum \text{Indoor Air Quality Index}}
\]

where \(W_p\) is the weightage of pollutant; IQAI, is the individual AQI pollution level; and PCM, is the corresponding concentration threshold mean of a pollutant.

**Weightage Worksheet Tables by Pollutant**

**Table 5. Weightage calculation for PM\textsubscript{10} a.**

<table>
<thead>
<tr>
<th>Indoor Air Quality Index</th>
<th>PM\textsubscript{10} 24-hour mean</th>
<th>PM\textsubscript{10} weightage</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>0.666666667</td>
</tr>
<tr>
<td>150</td>
<td>250</td>
<td>0.6</td>
</tr>
<tr>
<td>200</td>
<td>350</td>
<td>0.571428571</td>
</tr>
<tr>
<td>300</td>
<td>420</td>
<td>0.714285714</td>
</tr>
<tr>
<td>400</td>
<td>500</td>
<td>0.8</td>
</tr>
<tr>
<td>500</td>
<td>600</td>
<td>0.833333333</td>
</tr>
<tr>
<td>Final weightage</td>
<td></td>
<td>0.740816327</td>
</tr>
</tbody>
</table>

\(a\text{PM}_{10}:\) particulate matter with diameter of \(\leq 10\) \(\mu\)m.

**Table 6. Weightage calculation for CO a.**

<table>
<thead>
<tr>
<th>Indoor Air Quality Index</th>
<th>CO 1-hour mean</th>
<th>CO weightage</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>150</td>
<td>35</td>
<td>4.28571429</td>
</tr>
<tr>
<td>200</td>
<td>60</td>
<td>3.33333333</td>
</tr>
<tr>
<td>300</td>
<td>90</td>
<td>3.33333333</td>
</tr>
<tr>
<td>400</td>
<td>120</td>
<td>3.33333333</td>
</tr>
<tr>
<td>500</td>
<td>150</td>
<td>3.33333333</td>
</tr>
<tr>
<td>Final weightage</td>
<td></td>
<td>5.37414966</td>
</tr>
</tbody>
</table>

\(a\text{CO:}\) carbon monoxide.
### Table 7. Weightage calculation for O$_3^a$.

<table>
<thead>
<tr>
<th>Indoor Air Quality Index</th>
<th>O$_3$ 1-hour mean</th>
<th>O$_3$ weightage</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>160</td>
<td>0.3125</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>0.5</td>
</tr>
<tr>
<td>150</td>
<td>300</td>
<td>0.5</td>
</tr>
<tr>
<td>200</td>
<td>400</td>
<td>0.5</td>
</tr>
<tr>
<td>300</td>
<td>800</td>
<td>0.375</td>
</tr>
<tr>
<td>400</td>
<td>1000</td>
<td>0.4</td>
</tr>
<tr>
<td>500</td>
<td>1200</td>
<td>0.416666667</td>
</tr>
</tbody>
</table>

| Final weightage          | 0.429166667       |

$^a$O$_3$: ozone.

### Table 8. Weightage calculation for PM$_{2.5}^a$.

<table>
<thead>
<tr>
<th>Indoor Air Quality Index</th>
<th>PM$_{2.5}$ 24-hour mean</th>
<th>PM$_{2.5}$ weightage</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>35</td>
<td>1.428571429</td>
</tr>
<tr>
<td>100</td>
<td>75</td>
<td>1.333333333</td>
</tr>
<tr>
<td>150</td>
<td>115</td>
<td>1.304347826</td>
</tr>
<tr>
<td>200</td>
<td>150</td>
<td>1.333333333</td>
</tr>
<tr>
<td>300</td>
<td>250</td>
<td>1.2</td>
</tr>
<tr>
<td>400</td>
<td>350</td>
<td>1.142857143</td>
</tr>
<tr>
<td>500</td>
<td>500</td>
<td>1</td>
</tr>
</tbody>
</table>

| Final weightage          | 1.248920438          |

$^a$PM$_{2.5}$: particulate matter with diameter of ≤2.5 μm.

### Table 9. Weightage calculation for tVOC$^a$.

<table>
<thead>
<tr>
<th>Indoor Air Quality Index</th>
<th>tVOC mean</th>
<th>tVOC weightage</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>100</td>
<td>0.5</td>
</tr>
<tr>
<td>100</td>
<td>275</td>
<td>0.363636364</td>
</tr>
<tr>
<td>250</td>
<td>425</td>
<td>0.588235294</td>
</tr>
<tr>
<td>400</td>
<td>628</td>
<td>0.636942675</td>
</tr>
</tbody>
</table>

| Final weightage          | 0.522203583     |

$^a$tVOC: total volatile organic compound.

### Table 10. Weightage calculation for CO$_2^a$.

<table>
<thead>
<tr>
<th>Indoor Air Quality Index</th>
<th>CO$_2$ mean</th>
<th>CO$_2$ weightage</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>341</td>
<td>0.146627566</td>
</tr>
<tr>
<td>100</td>
<td>747</td>
<td>0.133868809</td>
</tr>
<tr>
<td>150</td>
<td>1305</td>
<td>0.114942529</td>
</tr>
<tr>
<td>200</td>
<td>5400</td>
<td>0.037037037</td>
</tr>
<tr>
<td>300</td>
<td>31,500</td>
<td>0.00952381</td>
</tr>
<tr>
<td>400</td>
<td>63,000</td>
<td>0.006349206</td>
</tr>
</tbody>
</table>

| Final weightage          | 0.074724826 |

$^a$CO$_2$: carbon dioxide.
Data Analysis

The data analysis flowchart is depicted in Figure 8. Analysis of the collected data from the DCS is automatically executed in the background and is transparent to the user. Upon receiving a request from the UI, or through the background user alert process, the database is queried to retrieve data. The data are validated against thresholds predetermined by environmental safety limits [52,53]. If the user is registered to receive alerts and the values exceed safe thresholds, an alert is sent to the user. Additionally, the values and corresponding qualitative metrics of each pollutant are determined. Finally, AQI is computed and the data are returned.

Figure 8. DASE data flow on the web server used to populate the Atmosome UI or send a user alert. DASE: data analysis and storage engine; DB: database; UI: user interface.

The authors considered including an adaptive threshold for each analyte to account for spatial, temporal, and environmental variations. However, AMS is recalibrated in each new usage location and is intended for indoor use. Thus, environmental variations are small and no substantial error between AMS and COTS sensors was observed when analyzing the data. Nonetheless, it is well-known that MQ sensors exhibit high temperature coefficients and sensitivity to humidity [54]. Drift
of these sensors is low and ongoing recalibration every 3 months is more than sufficient to address such drift [54]. Considering these factors, the ranges at which the AQI considered are absolute, but the unit itself is adjusted to account for location. Further, the sample rate is considered sufficient for the application. In the data that follow, it does not appear that the sensors are undersampled, so temporal variation is not considered in this embodiment. Nonetheless, one of the goals for the future work is to automate the calibration process and also allow customization of the calibration interval according to the user’s choice.

Data Alerts

If a user opts to be alerted when pollutants are above the optimal threshold, DASE sends an email and SMS text message alerts. This could be invaluable in preventing accidents or calling for emergency services in the case of an NG leakage or similar emergency.

Data Display

Current UI Module and Plans for the Next Version

The UI module displays the air quality information to users. The user accesses the UI at the atmosome website [55]. There is a dropdown menu to select the user’s zip code.

The next version of the atmosome UI will include a login screen for the user, instead of the current zip code selection. The various parts of the UI displayed to the user are described below.

Hourly Dashboard

The dashboard (Figure 9) presents the user with an “at-a-glance” state of their indoor air quality using a gauge that reflects AQI. The gauge conforms to the conventions of the US AQI gauge. The AQI is computed based on the pollutants that show the most variability due to user lifestyle, including PM$_{2.5}$, tVOC, CO$_2$, O$_3$, CO, and PM$_{10}$. The past hourly averages sample of 10 pollutant analytes is presented visually and quantitatively. Thresholds for each analyte are predetermined by environmental safety limits. Relative assessments of levels are color coded and reported as GOOD, MODERATE, POOR, or BAD. Four environmental parameters, including temperature, pressure, humidity, and altitude, are also shown quantitatively.

![Figure 9. Representative example of hourly dashboard of user’s atmosome data displayed by AMS UI. The values of 4 environmental parameters are also quantitatively presented below the AQI gauge. AMS: Atmosome Measurement System; AQI: Air Quality Index; PM$_{10}$: particulate matter with diameter of 10 μm or less; UI: user interface.](image)

Recommendations

Each pollutant in the dashboard is associated with detailed information about acceptable thresholds and specific suggestions on how to manage it to be within healthy limits. This is shown in Figures 10 and 11.
Figure 10. Lifestyle recommendations to improve atmospheric quality, provided to the user in AMS UI. Clicking on the “Tips” link of each of the pollutants in the UI dashboard, as shown, opens up a popup box showing the quantitative thresholds associated with the pollutant and gives the user recommendations on managing it to maintain healthy levels. AMS: Atmosphere Measurement System; ppb: parts per billion; UI: user interface; VOC: volatile organic compound.

Figure 11. Examples in the AMS UI that show the recommendations and thresholds for NG and PM2.5. AMS: Atmosphere Measurement System; NG: natural gas; PM$_{2.5}$: particulate matter with diameter of 2.5 μm or less; UI: user interface.

Average Exposure
The cumulative average values of different pollutants indicate the overall exposure across various periods as shown in Figure 12. These data provide potential insights into the underlying causes of poor AQI, which could be correlated with specific living conditions.
Figure 12. An example of AMS UI showing the average exposure statistics for different pollutants across various periods, ranging from 8 hours to 1 week. AMS: Atmosome Measurement System; CO$_2$: carbon dioxide; O$_3$: ozone; ppb: parts per billion; ppm: parts per million; UI: user interface; VOC: volatile organic compound.

Quarterly Temporal Graphs
The UI time-series data display quarterly data from the different pollutant data streams (Figure 13). This enables users to visualize trends over time and gain deeper insights into which pollutants are affecting their air quality most substantially. For example, tVOC measurements are noted to be highly variable due to indoor sources resulting from occupant lifestyle, including exposure to cosmetics, cleaning products, room refreshers, cooking fumes, and more.

Figure 13. An example of AMS UI showing graphs of the user’s quarterly trends of tVOC, CO$_2$, O$_3$, and humidity data streams. AMS: Atmosome Measurement System; CO$_2$: carbon dioxide; O$_3$: ozone; tVOC: total volatile organic compound; UI: user interface.

Pollutants Information Page
For each category of pollutant, upon selecting its name in the UI, the user is taken to a new page that contains a brief description of the pollutant; associated health risks at different concentrations; and examples of how such data can be collected, graphed, and studied further are shown. This provides users further insights into each of the pollutants and the possibility of enabling additional research. An example description page is shown in Figure 14 for PM$_{2.5}$.
**Figure 14.** An example of AMS UI individual pollutant page that is displayed on a separate web page for every pollutant and can be found by selecting the pollutant name in the dashboard. These pages provide the user with greater detail about each pollutant. This includes information such as more details about the pollutant, common sources of the pollutant, possible research and data analysis that have been/could be done on the pollutant using the corresponding AMS sensor, and more. This figure is an example of the PM2.5 pollutant page. Besides providing more details about PM$_{2.5}$, it shows how the AMS PM$_{2.5}$ sensor and extensive statistical analysis of its data was used for new, internationally published research on low-cost enhancement of facial mask filtration. AMS: Atmosome Measurement System; PM$_{2.5}$: particulate matter with diameter of 2.5 μm or less; UI: user interface.

**Particulate Matter (PM2.5)**

Particulate matter refers to mixtures of microscopic solid and liquid particles suspended in air. There are two types of particulate matter that are most relevant to air pollution: PM10 and PM2.5. PM10 refers to particles that are between 2.5 and 10 microns; some examples of these include dust, pollen, and particles of mold. PM2.5 consists of fine particles that are 2.5 microns in diameter or less; fuel combustion, cigarette smoke, aerosols, and more can form them. Particulate matter is a health risk because it is small enough to be inhaled and deposits itself in airways of the human body.

The AMS’s PM 2.5 sensor can also be used to spin off new research as done by the same researchers in “Low-cost enhancement of facial mask filtration to prevent transmission of COVID-19” published in the BBG Journal.

A system as shown below was developed to compare the efficacies of different facial masks, based on their abilities to filter PM 2.5 particles and the data was analyzed using a nested two-way ANOVA model, besides many other statistical techniques.

---

**REST API**

An extended REST API is also available for advanced users and developers interested in conducting further research or data analysis using the DCS measurements. The REST API enables users to download their data in a .csv format. These data can be mined to gain deep insights into the dynamics of the various indoor air pollutants across time, address extreme or alarming conditions by taking appropriate corrective actions, and exploring possible connections between air pollution and various health conditions.

**Results**

**Study Purpose**

The purpose of this work was to measure various pollutants and other air quality metrics that affect individual environmental atmosomes. The following section presents the results of a variety of air quality metrics in selected environments and relates them with the conditions in their atmosome.

**Temperature and Humidity**

Figures 15 and 16 show the quantity and value of readings of relative humidity and temperature, respectively, taken at 3 different indoor locations: a home in Cupertino, California (shown in blue); a home in Hyderabad, Telangana, India (shown in orange); and an airplane economy cabin during a nonstop flight of 17 hours (shown in green). The low humidity readings from the airplane correlate with the dryness and discomfort often experienced by airplane passengers and align with the United States Centers for Disease Control and Prevention’s (CDC) air travel yellow book [56].
Figure 15. Relative Humidity (%): 30-50 marks the ideal range. Conditions within the airplane journey were low in humidity and conditions in Hyderabad went above the recommended range at times. The Cupertino home had ideal humidity values.

![Relative Humidity Graph](image)

Figure 16. Temperature (°C): 20-27 marks a comfortable and healthy range. Conditions within the airplane journey and in Hyderabad were high in temperature. The Cupertino home had mostly ideal temperature values.

![Temperature Graph](image)

Carbon Dioxide

Figure 17 shows a graph of CO₂ measurements taken in the same environments as above. The readings in the home in the United States showed much higher indoor CO₂ levels than the readings in a more polluted area in India. Further analysis has revealed, however, that the closed windows and doors throughout the day during winter in the United States reduced ventilation and increased CO₂ concentration. Studies show that higher CO₂ exposure can cause drowsiness [57]. These data highlight the importance of ventilation during the winter. The readings on the lengthy airplane journey confirmed that the DCS readings and published values by the airline were within the range of each other. Figure 18 illustrates CO₂ readings in 2 rooms and shows the role AMS had in enabling the researcher to take corrective actions to improve indoor air quality.

![Carbon Dioxide Graph](image)

Figure 17. CO₂ (ppm): 250-1000 is the safe range for typical indoor spaces with good air ventilation. Higher than 1000 leads to a range of adverse effects from drowsiness to headaches, nausea, and increased heart rate among other conditions. Conditions within the airplane journey and in Hyderabad were mostly ideal to moderate, while conditions in the Cupertino home were outside healthy limits. CO₂: carbon dioxide; ppm: parts per million.
Figure 18. Temporal data for CO\textsubscript{2} pollution in 2 different rooms in the Cupertino house: (A) Room 1; (B) Room 2. In both rooms, air conditions were initially unhealthy and then improved drastically. In Room 1, conditions became unhealthy again after some hours before improving once more. CO\textsubscript{2}: carbon dioxide.

![Graph showing CO\textsubscript{2} pollution data for Room 1 and Room 2.]

(A) Room 1

(B) Room 2

**Particulate Matter 2.5 µm (PM\textsubscript{2.5})**

Figure 19 depicts the variation in PM\textsubscript{2.5} during the course of a typical home activity (cooking food). The onset of cooking is associated with a sharp, transient increase in the levels of PM\textsubscript{2.5} followed by a sustained period of unhealthy PM\textsubscript{2.5} levels. In this particular scenario, the user responded to the elevated concentration of PM\textsubscript{2.5} by activating the exhaust fan, deactivating the stove, and opening all windows for improved cross-ventilation.

Figure 19. PM\textsubscript{2.5} (µg/m\textsuperscript{3}) readings in the kitchen while cooking food. Much of the time spent cooking was in very unhealthy air conditions. PM\textsubscript{2.5}: particulate matter with diameter of 2.5 µm or less.

This set of mitigating actions lowered the PM\textsubscript{2.5} levels but unhealthy concentrations (defined by the horizontal orange and purple lines in Figure 19) persisted even after cooking ceased. These observational data highlight the power and utility of AMS in identifying and mitigating substantially unhealthy levels of indoor air quality. A longitudinal analysis of such data has the potential to offer rich insights into the interaction between indoor air pollutants and respiratory health outcomes, and aid data-driven health policy research.

**Volatile Organic Compounds and Ozone**

Raw data stored in DASE were downloaded in .csv format using its REST API for advanced users. The data were graphed in a python notebook using matplotlib. As shown in Figure 20, there is a substantial difference between Indian and American household air pollution levels in terms of tVOC measurements. The difference in results could be attributed partly to the substantially higher levels of air pollution in India compared with those in the United States. The current results are consistent with previous results suggesting vehicle exhaust as one of the leading sources of VOC-related pollution in India [58].

![Graph showing tVOC measurements in India and America.]

https://biomedeng.jmir.org/2021/4/e28920
**Figure 20.** tVOCs (ppb): 0-200 is the safe range. Values above 2200 ppb are extremely unhealthy. Conditions in the airplane journey and in the Cupertino home were typically ideal, whereas conditions in Hyderabad were often very unhealthy. ppb: parts per billion; tVOC: total volatile organic compound.

**Figure 21** displays a strong correlation between tVOC and ground-level $O_3$ measurements. For the most part, spikes in tVOC result in increases in $O_3$ levels that cause nitrogen oxides to react with tVOC in the presence of sunlight to create ground-level $O_3$.

**Figure 21.** Ozone (ppb): Ozone levels from February to April 2020 at a home in the Sierra Mountains, El Dorado County, CA, USA. $O_3$: ozone; ppb: parts per billion; tVOC: total volatile organic compound.

**Figure 22** displays tVOC readings recorded during various daily activities such as house cleaning, cooking at night, and commuting by car. The tVOC values were 1549, 2008, and 2868 ppb, respectively, for the aforesaid activities, well above the accepted moderate range of 220-660 ppb (presented in Figure 8), and demonstrate the high levels of tVOC during innocuous common activities [59]. The results also show that tVOC readings impact the indoor AQI value.

**Figure 22.** tVOC readings during (A) home cleaning, (B) cooking, and (C) commuting by car. These statistics are displayed on AMS user interface. AMS: Atmosome Measurement System; tVOC: total volatile organic compound.
Figure 23. Personal atsmere biological modeling. This scatterplot and line-of-best-fit display how ambient temperature affects resting heart rate reflected by data from the GoldenCheetah OpenData Project.

Studies carried out in 2003 by the California Air Resources Board show that cleaning products alone account for the release of 7.4 t of VOCs per day and that their various health effects include asthma attacks and eczema [60]. AMS can furnish necessary information to take precautions in such situations (eg, a spike in tVOC levels). Homes with infants or expecting parents often undergo more cleaning than typical. AMS, however, suggests that a presumably “well-cleaned” house may, in fact, pose a heightened risk of childhood asthma to infants [60], delays in child language development [61], or prenatal exposure to the fetuses that may impact their postnatal growth [62].

In summary, these results show that IAQ may be at unhealthy levels while conducting typical daily activities. AMS is able to track air quality through its sensors and indicate the impact of harmful levels of indoor air pollutants such as $O_3$, tVOC, $CO_2$ levels, and the presence of PM$_{2.5}$ inside homes.

Discussion

Principal Findings

This work has aimed to determine the performance of a low-cost AMS in indoor spaces in terms of gathering sensor data for a variety of air pollutants, transferring data reliably to an analysis engine in the cloud, displaying air quality monitoring results to the user, and sending alerts when the pollutants exceed safe thresholds. This system also provided a way for users to choose to share their IAQ data anonymously for further health research, and for users and researchers to retrieve data using REST APIs for further analysis and data analysis. This can be a foundation for building a public IAQ database across geographical regions.

The results establish that AMS is effective in analyzing multipollutant data streams in multiple settings, displaying AQI value, hourly visual and quantitative data of 18 pollutants, average exposure statistics of various pollutants, and the user’s quarterly pollutant exposure graphs. The dashboard also includes tips for each pollutant that display the GOOD, MODERATE, or POOR thresholds, enabling the user to better interpret the values and graphs shown in the UI and receive recommendations for keeping various pollutants in check with simple mitigation. The individual pollutant pages give the user further insights into the pollutant and show novel research in which AMS sensors and data collection have been employed. The system is also effective in sending alerts when pollutants exceed safe thresholds. The following sections discuss the findings of the work, which are corroborated by previous work and studies of a similar nature.

Particulate Matter

Particulate matter refers to mixtures of microscopic solid and liquid particles suspended in the air. There are 2 types of particulate matter that are most relevant to air pollution: PM$_{10}$ and PM$_{2.5}$. PM$_{10}$ consists of particles that are between 2.5 and 10 μm (diameter); some examples of these include dust, pollen, and particles of mold. PM$_{2.5}$ consists of fine particles that are 2.5 μm in diameter or less; some examples of these include fuel combustion, cigarette smoke, and aerosols. Particulate matter is a health risk because it is small enough to be inhaled and deposits itself in the airways of the human respiratory system. Smaller particles can lodge themselves deep in the lungs or enter the bloodstream. Even short-term exposure to PM$_{2.5}$ has
been associated with worsening respiratory diseases and can lead to emergency care. Long-term exposure (ie, months to years) has been linked to premature death, especially in people with chronic conditions, and leads to reduced lung function in children [63]. The EPA set the maximum 24-hour exposure limit to PM$_{2.5}$ to 35 µg/m$^3$ and annual exposure limit to 12 µg/m$^3$ [64].

Quite interestingly, results of this work clearly indicate that when cooking is involved, the level of PM$_{2.5}$ increases well beyond recommended exposure limits and these levels remain high even after cooking has ceased. While it was observed that ventilation could mitigate levels to some extent, it is imperative that cooking methodologies are modified to reduce the emission of PM$_{2.5}$. A study that focused on emission of PM$_{2.5}$ from cooking in homes found that the levels were raised 20-40-fold to 160 µg/m$^3$ in the kitchen and 10-fold in the nearby living room to 60 µg/m$^3$ [65], clearly corroborating the results of this work. Measuring particulate matter in indoor air can direct researchers toward finding easily implementable corrective actions to reduce exposure levels.

Carbon Dioxide
CO$_2$ is considered to be a dominant air pollutant. Moderately high concentrations of CO$_2$ in indoor air can lead to drowsiness, fatigue, and headaches. Increasing amounts can cause dizziness and nausea [66]. In the experimental data, CO$_2$ levels increased dramatically in poorly ventilated rooms as shown in the temporal data in Figure 18. In both the rooms, air conditions were initially unhealthy and then improved drastically. In room 1, conditions became unhealthy again after some hours due to a room heater that did not circulate air. Taking advice furnished by AMS and adding a fan, along with regular ventilation, improved conditions again. These results corroborate the findings of a previous work in the Texas elementary schools that monitored CO$_2$ levels in 120 randomly selected classrooms in 2 school districts [67]. The simple process of improving ventilation and air circulation can reduce high concentrations of CO$_2$ and improve comfort.

Volatile Organic Compounds
The experimental results show that even seemingly routine daily activities can induce significant tVOC release. A variety of household items, such as candles, cooking fumes, room fresheners, cosmetics, cleaning products, and paints, emit VOCs. These are organic chemicals that are usually in gaseous form at room temperature and are photo-chemically active. Short-term exposure to VOCs can cause optic or respiratory irritation, headaches, memory lapses, and dizziness. Long-term exposure can cause nausea, fatigue, organ damage, and cancer [68]. Results shown in Figure 22 based on an AMS report of activities, such as cleaning, cooking, and commuting, corroborate the findings of previous work [69-72].

Ozone
Readings taken while cleaning, cooking, and commuting, as shown in Figure 22, highlighted a correlation between activities that lead to elevated VOC levels in the presence of sunlight and elevated O$_3$ levels. In support, studies in Los Angeles have found as many VOCs being emitted from household products as from vehicle exhaust pipes, which then react in the presence of sunlight to produce ground-level O$_3$ [73]. These results match the EPA’s report that mentions the adverse effects of synthetic chemicals and emissions caused by cars, power plants, and other industrial setups that react in the presence of sunlight to increase ozone levels [74]. The results of this work show that the same levels of VOCs do not elevate O$_3$ as much at night, as there is no sunlight to facilitate such a chemical reaction.

Ground-level O$_3$ can trigger a variety of health problems, including chest pain, coughing, throat irritation, and airway inflammation. It can also worsen bronchitis, emphysema, and asthma [74]. This can lead to the need for increased medical care [75]. In 2016, 90% of noncompliance to the national ambient air quality standards in the United States was due to O$_3$. Both short- and long-term exposures to O$_3$ at concentrations below the current regulatory standards are associated with increased mortality from respiratory and cardiovascular diseases [76].

Temperature
The DCS can measure both events induced by high temperature (eg, the formation of ground-level O$_3$ and low-temperature triggers [77]) and brown adipose tissue metabolism [78]. The direct impact of temperature reflected in a user’s heart rate data stream during their sleep was characterized. An increasing trend in ambient temperature even in the 68-80°F (20-27°C) range and most definitely beyond (>27°C; Figure 16) could be unhealthy. This observation is supported by a previous study highlighted in Figure 23, which shows data collected for 1 user over 424 nights using an iPhone with the Sleep Cycle app and a Garmin Fenix 5 Watch [79]. Clear increases in the user’s circulation flow (ie, reflected by resting heart rate) were observed and appear to be correlated with increases in ambient temperature.

Humidity
Lower humidity levels can cause skin dryness, corneal dryness, dry nasal passages, and sinusitis [80], while higher relative humidity levels can promote the growth of mold, bacteria, and viruses. According to previous studies, low humidity levels lead to increased aqueous tear evaporation [81,82]. The low humidity values in the flight shown in Figure 15 corroborated with the researcher’s discomfort due to dry eye symptoms during the flight.

Significance
Previous studies have shown that several air pollutants, including PM$_{2.5}$, PM$_{10}$, tVOC, CO, CO$_2$, and more, can be present at much higher quantities in indoor air than in the outdoors [83]. There has been an increasing concern among the scientific community about the connection between IAQ and the impact of personal health [84].

The study results are significant in several ways. Data access in AMS is both simple and comprehensive. Real-time AMS data can be obtained and analysis performed from any device through a progressive web application. The website also includes historic data logs and recommendations on how to improve the
user’s current air quality. Because it uses a cloud-based API, it opens the possibility to integrate new types of analytical graphs or recommendations into the system without the users needing to update their device. With all the air quality measurements collected and posted in tandem, there is an increased knowledge about the interconnections and impacts of pollutants on an individual, community, or specific demographic. Other researchers can also tap into the system to download data and include their own analyses for their unique studies.

Limitations
Because the DCS unit’s calibration is not performed on a large scale, the calibration training set is limited when developing data correction algorithms. Another limitation is that the data were collected only by the researcher, and not external participants, which indicates that the data set is small and does not provide insight into various indoor pollutants affecting various living situations. This limits the development of training sets for learning algorithms to generate optimal individual weights for each of the pollutants to calculate AQI and to generate customized optimal recommendations for managing the pollutants to maintain healthy indoor air quality. While the DCS unit can be USB powered and portable when providing 13 data streams, due to increased power consumption for the comprehensive 22 data streams model, it requires to be plugged into a wall socket, thereby limiting its portability.

Future Work
Future research could implement AMS on a large scale to safeguard public and personal health. Users can anonymously share their data to help researchers draw connections between indoor air pollution and public health and to further research in the development of new modeling techniques for public health, low-cost sensor data correction algorithms, and sensor modeling. The larger vision of AMS is to build a public database for indoor air, like EPA’s Air Quality System database [11,12] for ambient air. The REST APIs in the system can be used to access the open database populated by users willing to share their data.

AMS could include dynamic recalibration of the devices installed at various settings over the internet. With this auto-calibration, sensors would maintain the highest accuracy possible and users would not have to perform manual calibration themselves.

AMS can be researched for reduction of power consumption, while still providing the comprehensive set of all 22 data streams, thus maximizing its portability.

AMS can be used within households to compare indoor air quality levels between neighbors, or on a grander scale within communities. Large-scale distribution of AMS can lead to an expansion of preventative health approaches from an individual level to a community level and can provide valuable insights into public health.

Other possible enhancements include making the high-power model portable, and adding physical indicators, such as sound and light on AMS hardware, which can help indicate the hazardous levels of pollutants in a more noticeable way.

Cybernetic and navigational health approaches enable individuals to be in control of their health throughout their lives so that they have the necessary information to always maintain an ideal state of health [85,86]. Continued research in this field will focus on expanding into other health domains, improving quality metrics, and developing methods to combine atmosome with other data streams to provide uniquely tailored lifestyle recommendations.

Last, AMS could implement more sophisticated software algorithms, such as an artificial neural network, as demonstrated in [27], and machine learning, as demonstrated in [30], and adaptive thresholding. In particular, the simplest algorithm to implement would be to account for environmental variation (eg, temperature and humidity) using an adaptive threshold to improve sensor accuracy. Specifically, and referring to [54], an adaptive threshold or the computed gas concentration could be computed as a function of temperature and humidity. This would significantly increase the sensitivity of the MQ sensors. Other algorithms likely exist and could be explored with the intention of maintaining a low-cost and small sensor solution as opposed to transitioning to an alternative sensor.

Conclusions
In this work, the concept of an atmospheric exposome (atmosome) was presented and a low-cost approach to leverage multimodal sensors and cloud data storage in building a personal atmosome was proposed. This work shows that AMS offers both the concept and system to quantify and measure the atmosphere with continuous real-time data streams, which provide users and researchers access to pertinent air quality data through a cloud computing architecture. These data can be used by both members of the public and researchers. For example, AMS could be used to alert an elderly user with dementia about a gas leakage.

Further, AMS can find the effect of other atmospheric streams such as temperature and humidity on the user’s observed health and behavioral outcomes. This system monitors and analyzes VOCs, which could help ensure that pregnant women breathe safe air, safeguarding not only their own health, but also that of their fetus(es). Particulate matter detection and warnings can help users to act on remediating the environment to avoid pulmonary complications in all age groups. Overall, the findings indicate that the multidimensional model of AMS is a step closer in considering a variety of pollutants and atmospheric characteristics that guide users toward a healthy lifestyle.

While the system can predict respiratory triggers, such as asthma attacks in vulnerable people, it is yet to be tethered to other health conditions (eg, proneness of a patient with diabetes to blood glucose variations due to air quality). AMS could certainly provide insight into such correlations and dramatically improve personal health.
Acknowledgments

This research was conducted by HB under the supervision of NN and RJ. This research received no external funding.

Authors' Contributions

Data from GoldenCheetah OpenData Project were retrieved by VP. HB and NN were responsible for study conceptualization and formal analysis. HB was also responsible for methodology, software implementation, data validation, investigation, procuring resources, writing—original draft preparation, and writing—review and editing. Both HB and VP took care of data visualization. NN and RJ provided project supervision and administration. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison of features and performance of COTS (Commercial-Off-The-Shelf) AQI sensor systems to AMS.

[PDF File (Adobe PDF File), 102 KB - biomedeng_v6i4e28920_app1.pdf]

Multimedia Appendix 2
Summary of recent AQI studies as compared to the study presented in this work.

[PDF File (Adobe PDF File), 85 KB - biomedeng_v6i4e28920_app2.pdf]

References

33. Mi PM 2.5 Detector - Your Own Eco-station. URL: https://xiaomi-mi.com/air-and-water-purifiers/xiaomi-mi-pm25-detector-white/ [accessed 2021-08-21]
37. Temtop M2AQMPM. 5 PM10 Detector HCHO Tester Data Export. Temtop. URL: https://tinyurl.com/5ckbu4yy [accessed 2020-12-28]
38. BRWISSEN A18 Air Quality Monitor Indoor Pollution Tester for Co2 Dust Meter PM1.0 PM2.5 PM10 Particulate Matter Analyzer HCHO Formaldehyde TVOC Detector Data Record Storage Particle Counter. URL: https://www.amazon.com/gp/product/B08K7DTKN2/ [accessed 2020-12-28]
41. uHoo Indoor Air Quality Sensor - 9 in 1 Smart Air Monitor. URL: https://getuhoo.com/home [accessed 2021-01-18]


55. Atmosone Live Data. URL: https://www.atmosone.com [accessed 2021-11-10]


63. California Air Resources Board. Inhalable Particulate Matter and Health (PM2.5 and PM10). 2020. URL: https://www2.arb.ca.gov/resources/inhalable-particulate-matter-and-health [accessed 2020-08-20]

64. US Environmental Protection Agency. What are the Air Quality Standards for PM?. URL: https://tinyurl.com/3s4mcntr [accessed 2020-09-15]


Abbreviations

- AMS: Atmosome Measurement System
- ANN: artificial neural network
- API: application programming interface
- AQI: Air Quality Index
- AWS: Amazon Web Services
- COTS: commercial-off-the-shelf
- DASE: data analysis and storage engine
- DCS: data collection system
Proposal

Tracking the Presence of Software as a Medical Device in US Food and Drug Administration Databases: Retrospective Data Analysis

Aaron Ceross¹, LLB, MSc; Jeroen Bergmann¹, BSc, MSc, PhD
Natural Interaction Lab, Department of Engineering Science, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Aaron Ceross, LLB, MSc
Natural Interaction Lab
Department of Engineering Science
University of Oxford
Parks Road
Oxford, OX1 3PJ
United Kingdom
Phone: 44 01865 273000
Email: aaron.ceross@eng.ox.ac.uk

Abstract

Background: Software as a medical device (SaMD) has gained the attention of medical device regulatory bodies as the prospects of standalone software for use in diagnostic and therapeutic settings have increased. However, to date, figures related to SaMD have not been made available by regulators, which limits the understanding of how prevalent these devices are and what actions should be taken to regulate them.

Objective: The aim of this study is to empirically evaluate the market approvals and clearances related to SaMD and identify adverse incidents related to these devices.

Methods: Using databases managed by the US medical device regulator, the US Food and Drug Administration (FDA), we identified the counts of SaMD registered with the FDA since 2016 through the use of product codes, mapped the path SaMD takes toward classification, and recorded adverse events.

Results: SaMD does not seem to be registered at a rate dissimilar to that of other medical devices; thus, adverse events for SaMD only comprise a small portion of the total reported number.

Conclusions: Although SaMD has been identified in the literature as an area of development, our analysis suggests that this growth has been modest. These devices are overwhelmingly classified as moderate to high risk, and they take a very particular path to that classification. The digital revolution in health care is less pronounced when evidence related to SaMD is considered. In general, the addition of SaMD to the medical device market seems to mimic that of other medical devices.

(JMIR Biomed Eng 2021;6(4):e20652) doi:10.2196/20652

KEYWORDS
regulation; software; medical device

Introduction

Background
Appropriate application of new digital technologies for health care is dependent on ever-evolving ethical and regulatory frameworks [1]. The US Food and Drug Administration (FDA) defines and oversees this framework for all medical devices in the United States, including software as a medical device (SaMD). Software is an integral part of many health care solutions, and in recent years, it has been acknowledged as a medical device on its own. The International Medical Device Regulators Forum (IMDRF) characterizes software as a medical device when the software itself is considered a medical device without the need for accompanying hardware. The development of more standalone software for clinical applications has led to the recognition of SaMD within medical device regulation. The FDA itself has acknowledged the strong growth potential and development of SaMD [2].
In 2011, the FDA undertook a study to predict trends for medical devices for the next 10 years. The study involved technical managers from the FDA and 15 non-FDA participants with a range of backgrounds, including clinical, policy making, and technological. It was found that reliance on software was a concept that crossed all six identified areas of growth. It was readily accepted that software would not just be an area of growth but also constituted a fundamental component of other trends. Within the last decade, standalone software has increasingly automated and facilitated a range of processes within the medical profession. It has been suggested that industry initiatives around this domain are continuously growing, which seems to be in line with popular opinion. However, there has been little empirical work describing the how the available data represent real growth and impact. It is useful to know how quickly this revolution is entering health care and potentially understand any documented issues. In this work, we explore data provided by the FDA to (1) discern new medical device product additions related to software, particularly in relation to SaMD, and (2) identify adverse incidents related to these registered devices. We aim to uncover any patterns in the data that may suggest the nature of any growth of SaMD.

### The FDA Classification Process

Although it may be possible to use an arbitrary device within a clinical setting or to address a medical issue, not all such devices may be categorized as “medical devices” for the purposes of regulation. Section 201(h) of the Food, Drug, and Cosmetic Act provides a definition of a “medical device,” which may be:

> any instrument, apparatus, implement, machine, appliance, implant, reagent for in vitro use, software, material or other similar or related article, intended by the manufacturer to be used alone or in combination, for human beings, for one or more [\ldots] specified medical purpose(s) [\ldots]

The “specified medical purposes” covers a wide range of activities, including (1) the diagnosis, prevention, monitoring, treatment, or alleviation of disease or injury; (2) the investigation, replacement, modification, or support of the anatomy or of a physiological process; (3) supporting or sustaining life; (4) the control of conception; or (5) providing information by means of in vitro examination of specimens derived from the human body.

A registered medical device receives a classification according to the risk it poses to the individual. The device’s intended use and purpose provide an indication of the risk level and, thus, the classification. The FDA describes three risk levels that set the types of controls and assessments that need to be considered before the device can be placed on the market. The classification and associated risks and controls are provided in Table 1.

<table>
<thead>
<tr>
<th>Class</th>
<th>Risk level</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td>Low to moderate</td>
<td>General controls</td>
</tr>
<tr>
<td>Class II</td>
<td>Moderate to high</td>
<td>General controls and special controls</td>
</tr>
<tr>
<td>Class III</td>
<td>High</td>
<td>General controls, special controls, premarket clearance</td>
</tr>
</tbody>
</table>

The classification of a device also plays a role in determining its pathway to final approval as a medical device. Class I devices are exempt from premarket submission. All devices are subject to general controls, which include notification, basic safety measures, and registration requirements. Special controls are those that consist of more stringent risk management processes, such as mandatory reporting of adverse events. Class III devices also require these measures, as well as a premarket approval (PMA) application, due to their high-risk nature (eg, supporting or sustaining human life) or novelty. A PMA involves more rigorous testing of the product before it can be released to market. Figure 1 provides a simplified view of the FDA paths for approval across different classifications.

The FDA uses a system of product codes to facilitate the classification of medical devices. A product code is a three-letter combination that designates the technological type of a device and its class. The Center for Devices and Radiological Health provides the names and attributes of each product code. The classification codes allow for rapid identification of the device type and the types of regulation that would apply, with the aim of making the path to market more efficient and rapid. The 510(k) is a submission made to the FDA to demonstrate that the device to be marketed is equivalent in safety and effectiveness to a legally marketed device, which is not subject to premarket approval.
Figure 1. A simplified model of the FDA classification pathways for new medical devices. The 510(k) is a submission made to the FDA to demonstrate that the device to be marketed is equivalent in safety and effectiveness to a legally marketed device, which is not subject to premarket approval. FDA: US Food and Drug Administration.

SaMD Regulation

In 2013, there was regulatory recognition that standalone software may constitute a medical device, given the proliferation of developed systems. The software was previously classified according to an affiliated hardware device before the explicit inclusion of software in the regulations [8]. The IMDRF is a voluntary group of medical device regulators from different jurisdictions, working together under the World Health Organization’s Global Harmonization Task Force with a focus on the harmonization of medical device regulation. They drafted a guidance document [9] for SaMD to harmonize their definitions. In the document, software was recognized as a device without the need for affiliated hardware. This widened the scope of medical devices to include analytical software as well as mobile apps.

Methods

Addition of New Devices

In this work, we used the Global Unique Device Identification Database (GUDID), which is maintained by the FDA and made freely available to the public [10]. It contains records of devices that have a unique identification number. The device companies submit the relevant information concerning their product to the database. The data are made available either through an API or as downloadable text files. GUDID divides its files into 9 separate files [11]. The devices are identified in each of the data sets through a “Primary Device Identifier” number, which is unique to each device. In this work, we used the text files available from the website (full release dated August 21, 2020) and analyzed the data in the R language. The files include the unique ID of the device, description of the device, manufacturer, date of addition to the database, product code, and device classification. The device publish date is the date that the device record was created in the database.

To identify SaMD, we used the Global Medical Device Nomenclature (GMDN) terms [12] and the FDA product codes [13]. The GMDN is an internationally accepted scheme that identifies medical devices through a 5-digit numeric value and generic terms associated with this unique value [9]. In a similar manner, the FDA has developed product codes for medical devices that associate a device with a generic description and...
We extracted SaMDs by subsetting those devices with the string “software” as a term.

We analyzed the data between February 2014 and August 2020, comprising 2,628,409 devices. A total of 32 product codes contain the term “software” in their name for these years. Devices that are software but were not assigned one of the relevant codes previously mentioned were not considered in this analysis. In this work, we used a Sankey diagram to explore and visualize the pathway to market approval. The diagram shows the transitions and relations within the data, allowing for a more immediate understanding of the relationship of the variables within the data.

Adverse Events

Adverse events related to medical devices are recorded in the Manufacturer and User Facility Device Experience (MAUDE) database [14]. The database comprises both mandatory reports (eg, those obtained from manufacturers) and voluntary reports (received from patients and clinicians). The data are made available as text files in a pipe-delimited format and contain fields related to the product, its name, its type, and the number of affected patients (no personal data are available in these data). These are split into separate files. In this work, we used the device data files. The database acknowledges that it is limited in its surveillance and contains incomplete and/or inaccurate descriptions of events. As such, it is not possible to use MAUDE to detect the prevalence of any type of event, owing to the potential of underreporting. Nevertheless, the data can be informative and provide an indication about the types of risks encountered by patients in the use of medical devices. In addition, we cross-referenced GUDID data information with FDA reports of adverse events through the use of product codes.

Results

Addition of New Devices

During the period from 2014 to 2020, 6193 devices were registered with “software” as a GMDN term. However, it is unclear from the GMDN whether the device is solely composed of software or merely incorporates it. To resolve this, we relied on product codes. Of the total devices registered with software product codes, 515 had only a single product code that was related to software. These devices were identified as SaMD.

Table 2 shows that most of these devices were Class II, and nearly all that were identified as SaMD by product code (476/515, 92.4%) fell within this classification. It should be noted that the figures for software do not add up to 100% owing to rounding as well as to removal of records listed with unknown device classes.

Table 2. Classification proportions for all the GUDID data for software (generally, as a subset by GMDN terms) and SaMD (defined as a subset by product codes).

<table>
<thead>
<tr>
<th>Class type</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total GUDID</strong> (N=2,628,409)</td>
<td></td>
</tr>
<tr>
<td>Class I</td>
<td>599,277 (22.8)</td>
</tr>
<tr>
<td>Class II</td>
<td>1,968,678 (74.9)</td>
</tr>
<tr>
<td>Class III</td>
<td>49,940 (1.9)</td>
</tr>
<tr>
<td><strong>Software GMDN</strong> (n=6193)</td>
<td></td>
</tr>
<tr>
<td>Class I</td>
<td>793 (12.8)</td>
</tr>
<tr>
<td>Class II</td>
<td>5208 (84.1)</td>
</tr>
<tr>
<td>Class III</td>
<td>155 (2.5)</td>
</tr>
<tr>
<td><strong>SaMD product code</strong> (n=515)</td>
<td></td>
</tr>
<tr>
<td>Class I</td>
<td>12 (2.3)</td>
</tr>
<tr>
<td>Class II</td>
<td>476 (92.4)</td>
</tr>
<tr>
<td>Class III</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

The pathways to classification are shown in Table 3. The table shows that the majority of devices were submitted through premarket notification, while only 5 devices, comprising <1% of the total devices, followed a 510(k) exemption path.

The pathway to classification is shown for both SaMD and non-SaMD devices (Figure 4). SaMD, almost without exception, seem to have a more singular path to classification. Their main path for market entry is through premarket notification.
Figure 2. Registration of devices in the Global Unique Device Identification Database by yearly quarter. The vertical axis is on a logarithmic scale. GMDN: Global Medical Device Nomenclature; Q: quarter; SaMD: software as a medical device.

Figure 3. Frequency of software product codes set by the US Food and Drug Administration [13], showing the dominance of index-generating electroencephalograph software (product code OLW) among SaMD registrations.
Table 3. Pathways to classification for software as a medical device (n=515).

<table>
<thead>
<tr>
<th>Submission type</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premarket notification (510(k))</td>
<td>483 (93.8)</td>
</tr>
<tr>
<td>Contact office of device evaluation</td>
<td>22 (4.3)</td>
</tr>
<tr>
<td>510(k) exempt</td>
<td>5 (0.97)</td>
</tr>
</tbody>
</table>

**Figure 4.** Paths to classification compared between SaMD (red) and non-SaMD (grey) devices. The data are shown in log scale to visualize the distinctions between the paths. The questions are given at the bottom of each column. SaMD: software as a medical device.

**Adverse Events**

For the years from 2015 to 2019, there were 5.1 million reported adverse events in MAUDE for all devices (Figure 5). A subset of the database was examined consisting of only those product codes related to software during the years available for the GUDID. During the same reporting period, 215 adverse events were reported for devices with product codes related to software. This represents a total of 38 manufacturers. This subset does not capture all software-related events but only those related to SaMD (Figure 6). In slightly over half the SaMD cases (21/38, 55%), the device was reported to have been evaluated by the manufacturer. This is in contrast to 37% (1,900,000/5,100,000) of the cases for all reported adverse events.
Discussion

This work represents a first empirical look at SaMD pathways to market and representation in adverse events based on publicly available data. These findings give a clearer understanding of the nature of SaMD within the regulatory environment. SaMD patterns for entry into market and in adverse events do not seem to deviate from those of medical devices in general. However, the number of new devices entering into the market and adverse events for both types of devices have been rising in the last few years. This rise, however, is rather modest, and it seems that regulations may be an (appropriate) barrier, as not all technologies developed are indeed safe or perform at a suitable level.
level. The number of adverse incidents related to SaMD has also been rising, but at a faster rate than the number of devices. This could indicate that software enters the market earlier than it should, or it may simply identify a tendency toward better reporting of adverse events. This may also explain the higher percentage of adverse event reporting that was found for SaMD manufactures. However, the number of adverse events reported for SaMD is so small compared to the overall number of reported events that any interpretation needs to be carefully considered.

Almost all SaMD requires a 510(k) premarket notification, as demonstrated in Table 3. This indicates that the majority of SaMD is not an exempt product and that manufacturers often aim to enter the market by describing the similarity of their devices to other products that are already available. It has been suggested that although the 510(k) clearing process may offer expediency in bringing devices to market, this may impact the safety of the device. It remains a question for further investigation whether the 510(k) process has a negative impact on the safety of SaMD [15].

There is a noted anomalous spike in the third quarter of 2016 across the data. In that year, the United States passed the 21st Century Cures Act [16], which was aimed at facilitating the acceleration of medical product development by fast-tracking new innovations and advances that could benefit patients. This may have had an effect on the approval of devices. The rise in new device approvals that year may be related to the requirements of this legislation and the reclassification of some devices. It should be noted that GUDID is reliant on device labelers for information and the system allows for bulk uploads, which could help explain this feature in the data. However, no causation data are available to verify this.

Several challenges still remain in developing SaMD. The modern software development pattern frequently uses a form of iterative cycles wherein problems (as well as needed features) are identified and developed within the cyclical period. However, safety-critical software requires formal verification to determine that it performs as intended and that it can manage identified risks appropriately. Although there is strong evidence that formal verification methods more readily address regulatory compliance, the associated documentation, management, and training costs may not directly contribute to the delivery of customer value [17]. As such, medical device regulations may arguably make it difficult to use modern software development approaches [18]. The FDA has embarked on approaches designed to address the particular issues of software management within health care, such as the use of precertification of SaMD. Lee and Kesselheim [19] highlight that the FDA does not have the resources to validate every single iteration of software. Therefore, if new features are added, they may have certification despite not having any clinical evidence to support claims of treatment or diagnosis. This arguably limits the surveillance that can be conducted by the regulator. This may also have an impact on risk management, which in turn has an effect on the regulatory outcomes for such devices. It is possible that software development itself is not yet optimized for medical devices. Likewise, a question remains as to whether artificial intelligence and machine learning should be distinguished from SaMD, which at the time of publication was an ongoing discussion topic within the FDA [20].

Overall, it seems that SaMD has not yet developed at a different rate from that of other medical devices. Although more research is needed to robustly explain the results reported here, this work does provide useful insight for considering the digital revolution in medicine and how it relates to the market reality.

Conflicts of Interest
None declared.

References

Abbreviations

FDA: US Food and Drug Administration
GMDN: Global Medical Device Nomenclature
GUDID: Global Unique Device Identification Database
IMDRF: International Medical Device Regulators Forum
MAUDE: Manufacturer and User Facility Device Experience
PMA: premarket approval
SaMD: software as a medical device

©Aaron Ceross, Jeroen Bergmann. Originally published in JMIJ Biomedical Engineering (http://biomedeng.jmir.org), 03.11.2021. 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 JMIJ Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on https://biomedeng.jmir.org/, as well as this copyright and license information must be included.