## Contents

### Reviews

**Roles of Health Literacy in Relation to Social Determinants of Health and Recommendations for Informatics-Based Interventions: Systematic Review** ([e50898](#))  
Shwetha Bindhu, Anunita Nattam, Catherine Xu, Tripura Vithala, Tiffany Grant, Jacinda Dariotis, Hexuan Liu, Danny Wu. ............................................. 3

**e-Cigarette Tobacco Flavors, Public Health, and Toxicity: Narrative Review** ([e51991](#))  
Yehao Sun, Prital Prabhu, Ryan Rahman, Dongmei Li, Scott McIntosh, Irfan Rahman. ........................................................................................................ 17

### Corrigenda and Addenda

**Correction: Machine Learning Model for Predicting Mortality Risk in Patients With Complex Chronic Conditions: Retrospective Analysis** ([e58453](#))  
Guillem Hernández Guillamet, Ariadna Morancho Pallaruelo, Laura Miró Mezquita, Ramón Miralles, Miquel Mas, Maria Ulledemolins Papaseit, Oriol Estrada Cuxart, Francesc López Seguí. .......................................................... 27

### Original Papers

**Intention to Use Mobile-Based Partograph and Its Predictors Among Obstetric Health Care Providers Working at Public Referral Hospitals in the Oromia Region of Ethiopia in 2022: Cross-Sectional Questionnaire Study** ([e51601](#))  
Kefyalew Tilahun, Jibril Adem, Wabi Atinafu, Agmasie Walle, Nebyu Mengestie, Abraham Birhanu. .......................................................... 29

**Discussions With End Users to Inform the Vision for a Shared Care Record in Ontario: Qualitative Interview Study** ([e51231](#))  
Marta Chmielewski, Matthew Meyer. ........................................................................................................ 56

**Deriving Treatment Decision Support From Dutch Electronic Health Records by Exploring the Applicability of a Precision Cohort–Based Procedure for Patients With Type 2 Diabetes Mellitus: Precision Cohort Study** ([e51092](#))  
Xavier Pinho, Willemijn Meijer, Albert de Graaf. ........................................................................................................ 69
Characterization of Post–COVID-19 Definitions and Clinical Coding Practices: Longitudinal Study (e53445)
Monika Maripuri, Andrew Dey, Jacqueline Honerlaw, Chuan Hong, Yuk-Lam Ho, Vidisha Tanukonda, Alicia Chen, Vidul Panickan, Xuan Wang, Harrison Zhang, Doris Yang, Malarkodi Samayamuthu, Michele Morris, Shyam Visweswaran, Brendin Beaulieu-Jones, Rachel Ramoni, Sumitra Muralidhar, J Gaziano, Katherine Liao, Zongqi Xia, Gabriel Brat, Tianxi Cai, Kelly Cho

Patient Characteristics Associated With Phone and Video Visits at a Tele-Urgent Care Center During the Initial COVID-19 Response: Cross-Sectional Study (e50962)
Saif Khairat, Roshan John, Malvika Pillai, Philip McDaniel, Barbara Edson

Trends in the Ophthalmic Workforce and Eye Care Infrastructure in South India: Cross-Sectional Questionnaire Study (e50921)
Srinivasa Pallerla, Madhurima Pallerla, Sannapaneni Krishnaiah

Electronic Health Records for Population Health Management: Comparison of Electronic Health Record–Derived Hypertension Prevalence Measures Against Established Survey Data (e48300)
Katie Allen, Nimish Valvi, P Gibson, Timothy McFarlane, Brian Dixon

Effect of Long-Distance Domestic Travel Ban Policies in Japan on COVID-19 Outbreak Dynamics During Dominance of the Ancestral Strain: Ex Post Facto Retrospective Observation Study (e44931)
Junko Kurita, Yoshitaro Iwasaki

A Semantic Approach to Describe Social and Economic Characteristics That Impact Health Outcomes (Social Determinants of Health): Ontology Development Study (e52845)
Daniela Dally, Muhammad Amith, Rebecca Mauldin, Latisha Thomas, Yifang Dang, Cui Tao

Machine Learning for Prediction of Tuberculosis Detection: Case Study of Trained African Giant Pouched Rats (e50771)
Joan Jonathan, Alcardo Barakabitze, Cynthia Fast, Christophe Cox

Geospatial Imprecision With Constraints for Precision Public Health: Algorithm Development and Validation (e54958)
Daniel Harris, Chris Delcher

Viewpoints

Applying Machine Learning Techniques to Implementation Science (e50201)
Nathalie Huguet, Jinying Chen, Ravi Parikh, Miguel Marino, Susan Flocke, Sonja Likumahuwa-Ackman, Justin Bekelman, Jennifer DeVoe

Vaping: Public Health, Social Media, and Toxicity (e53245)
Yehao Sun, Prital Prabhu, Dongmei Li, Scott McIntosh, Irfan Rahman
Roles of Health Literacy in Relation to Social Determinants of Health and Recommendations for Informatics-Based Interventions: Systematic Review

Abstract

**Background:** Health literacy (HL) is the ability to make informed decisions using health information. As health data and information availability increase due to online clinic notes and patient portals, it is important to understand how HL relates to social determinants of health (SDoH) and the place of informatics in mitigating disparities.

**Objective:** This systematic literature review aims to examine the role of HL in interactions with SDoH and to identify feasible HL-based interventions that address low patient understanding of health information to improve clinic note-sharing efficacy.

**Methods:** The review examined 2 databases, Scopus and PubMed, for English-language articles relating to HL and SDoH. We conducted a quantitative analysis of study characteristics and qualitative synthesis to determine the roles of HL and interventions.

**Results:** The results (n=43) were analyzed quantitatively and qualitatively for study characteristics, the role of HL, and interventions. Most articles (n=23) noted that HL was a result of SDoH, but other articles noted that it could also be a mediator for SDoH (n=6) or a modifiable SDoH (n=14) itself.

**Conclusions:** The multivariable nature of HL indicates that it could form the basis for many interventions to combat low patient understandability, including 4 interventions using informatics-based solutions. HL is a crucial, multidimensional skill in supporting patient understanding of health materials. Designing interventions aimed at improving HL or addressing poor HL in patients can help increase comprehension of health information, including the information contained in clinic notes shared with patients.

*(Online J Public Health Inform 2024;16:e50898)*  doi: [10.2196/50898](https://doi.org/10.2196/50898)

**KEYWORDS**

health literacy; social determinants of health; SDoH; social determinants; systematic review; patient education; health education; health information; information needs; information comprehension; patient counseling; barriers to care; language proficiency
Introduction

Overview

In recent decades, medical providers, health systems, and legislators have prioritized increasing patient access to health information. For example, the 21st Century Cures Act mandates that patients must have access to their electronic health records, including clinic notes, in a rapid and convenient manner [1]. However, clinic notes and other health information can contain jargon that is difficult for patients to comprehend, reducing the utility of health information sharing. The Healthy People 2030 initiative, sponsored by the US Department of Health and Human Services, aims to address this issue by increasing patient comprehension of health information received from providers and web-based sources, such as their electronic health records [2].

A key part of health information comprehension is health literacy (HL), the ability to understand, contextualize, and make well-informed decisions based on health information [3]. Reducing HL gaps is crucial to meeting the goals set forth by Healthy People 2030 and maximizing the benefits of the 21st Century Cures Act.

Health Literacy

Having high HL correlates with greater shared decision-making between patients and physicians and promotes positive health outcomes because patients can better comprehend and act on the health information they receive [4]. Healthy People 2030 distinguishes between two dimensions of HL: personal, as previously described, and organizational [2]. Organizational HL holds health care systems and providers accountable for providing their patients with comprehensible health information to make informed decisions. This newer understanding of HL raises questions about how HL fits into the public health framework addressing disparities in health comprehension.

Social Determinants of Health and Health Literacy

Social determinants of health (SDoH) are nonmedical social and economic factors that fall into the following 5 domains: economic stability, education, health care and access quality, neighborhood and built environment, as well as social and community context [5,6]. SDoH affects health status and outcomes, and it can generate health disparities between population groups by influencing patient behavior and organizational responses. These determinants are also distinct from social factors or needs that exist at the individual level and instead exist as community- or population-level barriers [7-9].

HL has been categorized in different sources as an SDoH itself and as a midstream consequence of SDoH that can impede or improve patient interactions with health care institutions and health outcomes (ie, vaccination status and screening utilization) [10-12]. For example, a study by Schillinger et al [13] proposes that a higher education level improves HL, which was associated with better glycemic control among patients with diabetes. This is a unidirectional characterization of the relationship between SDoH, HL, and health outcomes, depicted in Figure 1.

Figure 1. Relationships characterized by the influence of social determinants of health on health literacy (personal and organizational), subsequently impacting health outcomes.

However, this may be an oversimplification. HL can evolve through continued exposure to health environments and interventions at the personal and organizational levels [14]. Moreover, even patients with high HL can struggle with comprehension in different contexts. Therefore, this relationship warrants further investigation, as there is a lack of systematic
literature analyses that macroscopically evaluate how SDoH and HL are related across different SDoH domains [15]. Understanding the nature and role of HL in interactions with SDoH can also indicate the most effective approach to designing HL-targeting interventions for patients who struggle to understand health information.

Objectives
Due to the literature gap in examining the complex relationship between HL and SDoH, we aimed to conduct a systematic literature review to (1) understand this relationship and (2) recommend informatics-based interventions to address low HL among patients.

Methods

Search Strategy
We systematically reviewed literature in PubMed and Scopus, two major biomedical and social science literature repositories. The initial database searches were conducted on June 22, 2020. The review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2009 guidelines to understand the relationship between HL and SDoH [16,17].

The search terms used were “health literacy” AND “social determinants of health.” After filtering for non-English articles and articles without abstracts, the remaining 281 articles were compiled in a Microsoft Excel sheet with their title, author, publication year, DOI or PMID, and abstract.

Screening Process
Two researchers (SB and CX) independently screened 281 papers by title and abstract and used the following exclusion criteria: (1) HL is a minor factor in the article; (2) the article is not an empirical study; (3) the article focuses on HL measurement tool development or evaluation; (4) the paper does not examine HL in relation to SDoH; (5) no abstract is available; and (6) the paper is not written in English.

Each researcher independently gave the article a score of “1” for inclusion or “0” for exclusion. The scores were summed; articles scoring “2” were automatically included, and those scoring “0” were excluded from the full article eligibility review. Disagreements (any papers with a total score of “1”) were resolved by the authors after the initial screening. The process was repeated for the full-article eligibility review and subsequent reference screening from the included full articles. Reference screening was a precautionary step to ensure the inclusion of articles that may not have been included in the initial database search. Original exclusion criteria were consistently used.

Quality Assessment
Before the information extraction, all included articles were assessed by 2 researchers (SB and TG) for study quality. Using the Agency for Healthcare Research and Quality (AHRQ) guidelines, separate quality assessments were developed for each type of study included in the review—observational studies and randomized clinical trials (RCTs) [18]. Domains included in both study types were study questions, population, interventions, outcome measurement, statistical methods, results, discussion, and disclosure of funding or sponsorship. Domains evaluated in the RCT assessment also included blinding and randomization.

The reviewers created a 3-point scoring system for the quality assessment. Articles were rated by 2 team members (SB and TG) with scores of “good,” “fair,” and “poor” for each domain and assigned numerical values of 2, 1, and 0, respectively, as per the AHRQ guidelines [18]. Values were averaged and translated back to a rating of “good” (1.50 or higher), “fair” (1-1.49), and “poor” (0-0.99).

Information Extraction
Based on quality assessment results, 43 papers were included for information extraction. Four researchers (CB, CX, AN, and TV) extracted data for the following PRISMA-based criteria: title, author, article ID, year published, location, study design, sample demographics, results, and limitations [16]. To answer the research questions, information specific to SDoH focus, HL measurement, and health outcomes was collected. The information extraction sheet is attached as Multimedia Appendix 1.

Quantitative and Qualitative Data Analysis
Extracted data were analyzed both quantitatively and qualitatively. Location, year of publication, and study design were statistically summarized. AHRQ guidelines were used to categorize studies as RCT, cross-sectional, and qualitative designs, with the last two being types of observational studies [19].

Qualitative analysis was conducted in 2 steps. First, a narrative synthesis of the chosen articles summarized the relationships between SDoH and HL. Narrative synthesis involves analyzing the data from systematic reviews to create textual explanations of observed patterns or trends rather than relying solely on statistical data. This involves developing textual descriptions of the data by extracting key information pertinent to the research question (ie, methods used or results) and exploring commonalities and differences between and within studies (ie, through visually mapping relationships) [20]. These methods were also used in a systematic review previously published by the authors [21]. The included articles were classified by SDoH domains they addressed, per the 5 domains defined by Healthy People 2030: economic stability, education, health care access and quality, neighborhood and built environment, as well as social and community context. Then, information extraction data from article results and discussion sections were used to define roles for HL. Finally, a theme visualization was conducted that plotted HL roles against publication year to understand how HL perception has evolved.

In the second step, lessons learned were summarized regarding HL roles, again using the results and discussion sections. From these same sections, the authors then extrapolated possible interventions that use HL to improve patient comprehension of health information.
Results

Literature Search Results
The PubMed and Scopus searches yielded 389 articles, resulting in 281 unique articles (Figure 2). After screening titles and abstracts, 43 articles remained for full-text eligibility assessment. Not discussing HL and SDoH together (n=95) was the largest cause for exclusion. Other papers were excluded because HL was not a substantial focus of the paper (n=62). A total of 19 articles were excluded from the full-text eligibility, once again for a minor focus on HL. References of the remaining 24 articles were screened for inclusion, yielding 20 additional articles. After the quality assessment, 1 low-quality article was excluded. Information extraction and narrative synthesis were conducted on a final sample of 43 articles.

Figure 2. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram of articles from the PubMed and Scopus search. A total of 24 articles (underlined) were included from the first database search.
Quantitative Analysis

The final 43 articles were analyzed for study location, year of publication, role of HL, and study design (Table 1). A total of 14 (32.6%) studies took place in North America (12 in the United States and 18 in Europe). All articles from South America originated in Brazil (n=4). Publication year trends revealed an increased focus and discussion of the topic in recent years, with 90.7% (n=39) of the articles being published after 2010. Most articles (n=40, 93%) had a cross-sectional design and used surveys, while 4.7% (n=2) used a qualitative design with semistructured interviews and focus questions to assess SDoH and HL.

Table 1. Summary of articles included in the literature review (N=43).

<table>
<thead>
<tr>
<th>Category</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study location</strong></td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>18 (41.9)</td>
</tr>
<tr>
<td>North America</td>
<td>14 (32.6)</td>
</tr>
<tr>
<td>Asia</td>
<td>6 (14)</td>
</tr>
<tr>
<td>South America</td>
<td>4 (9.3)</td>
</tr>
<tr>
<td>Australia</td>
<td>1 (2.3)</td>
</tr>
<tr>
<td><strong>Year published</strong></td>
<td></td>
</tr>
<tr>
<td>2006-2009</td>
<td>4 (9.3)</td>
</tr>
<tr>
<td>2010-2013</td>
<td>9 (20.9)</td>
</tr>
<tr>
<td>2014-2017</td>
<td>15 (34.9)</td>
</tr>
<tr>
<td>2018-2021</td>
<td>15 (34.9)</td>
</tr>
<tr>
<td><strong>Study design</strong></td>
<td></td>
</tr>
<tr>
<td>Cross-sectional</td>
<td>40 (93)</td>
</tr>
<tr>
<td>Qualitative</td>
<td>2 (4.7)</td>
</tr>
<tr>
<td>Randomized controlled trial</td>
<td>1 (2.3)</td>
</tr>
</tbody>
</table>

The search was not limited to 2006 for publication year; this was the earliest date among the 43 articles.

Qualitative Analysis

Narrative Synthesis

The narrative synthesis generated 4 roles for HL in relation to SDoH (Table 2). Most of the articles discussed multiple SDoH domains, but all 43 articles discussed education access and quality [5].

The most common categorization of the HL role was as a “result of SDoH” (n=23), followed by “modifiable SDoH” (n=14), and finally, as a “mediator of SDoH” (n=6). HL can be a “result of SDoH” (n=23), which suggests that SDoH domains contribute to HL levels and that it is a downstream variable [22-44]. As mentioned, 14 studies identified HL as a “modifiable SDoH,” where they identified HL as an SDoH, often citing the World Health Organization’s categorization of it; these studies suggested that HL can be improved through interventions and is actionable at multiple levels [14,45-57]. Finally, the articles that categorized HL as a “mediator of SDoH” (n=6) discussed how HL is an intermediary between other SDoH domains, such as educational attainment or economic stability, and that high HL levels can compensate for lower domain levels that compromise positive health outcomes [58-63]. Occasionally, the same paper would suggest multiple roles for HL (eg, an article’s Results and Discussion sections would inform both modifiable and mediatory roles for HL), but the most prominent relationship that appeared was used to categorize each article.

These 3 roles were plotted against the years of publication in Figure 3. In the few HL-focused articles published before 2010, HL was recognized as having a variety of roles, but only 1 article identified it as a modifiable SDoH. In the next 5-year period, being a result of SDoH was the most common role assigned to HL. In 2013, a total of 5 out of 7 articles identified HL as being a result of SDoH. As the number of published HL-focused articles increased in subsequent years, being a result of SDoH remained the most consistent and most prominent role assigned to HL to appear across all articles. Nevertheless, there has been increasing recognition of HL as a modifiable SDoH in the years 2015, 2018, and 2020, further cementing HL’s multidimensional nature.
### Table 2. Summary of narrative synthesis themes.

<table>
<thead>
<tr>
<th>Category</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SDoH</strong> domain(^b)</td>
<td></td>
</tr>
<tr>
<td>Education access and quality</td>
<td>43 (100)</td>
</tr>
<tr>
<td>Economic stability</td>
<td>38 (88)</td>
</tr>
<tr>
<td>Health care and access quality</td>
<td>11 (26)</td>
</tr>
<tr>
<td>Social and community context</td>
<td>11 (26)</td>
</tr>
<tr>
<td>Neighborhood and built environment</td>
<td>9 (21)</td>
</tr>
<tr>
<td><strong>HL</strong> role</td>
<td></td>
</tr>
<tr>
<td>Result of SDoH</td>
<td>23 (53)</td>
</tr>
<tr>
<td>Modifiable SDoH</td>
<td>14 (33)</td>
</tr>
<tr>
<td>Mediator of SDoH</td>
<td>6 (14)</td>
</tr>
</tbody>
</table>

\(^a\)SDoH: social determinant of health.
\(^b\)Most of the articles included more than 1 SDoH domain they studied.
\(^c\)HL: health literacy.

**Figure 3.** Theme visualization of the evolution of health literacy roles over time. SDoH: social determinants of health.

**Lessons Learned**

In addition to analyzing articles for the HL role, each article was further examined to determine details about the nature of the relationship between HL and SDoH. These were titled “lessons learned.” Figure 4 [14,22-63] shows an idea map that organizes articles by the role of HL and lessons learned. Although most of the articles are cross-sectional and do not always draw a causal relationship between HL and SDoH, the authors of the articles nevertheless offer hypotheses on factors influencing HL or how it interacts with SDoH and health outcomes.
**HL as a Result of SDoH (n=23)**

Being a result of SDoH was the most frequent role identified for HL. These articles characterized HL as being associated with, influenced by, or resulting from other SDoH. All articles addressed that a higher level of education, such as high school graduation, had implications on HL levels [22-44]. Bazaz et al [22], Berens et al [24], and Rocha et al [39] have suggested that HL is developed through interactions with health care due to age and disease condition, and more interaction with health care over time leads to an improvement in HL. Hou et al [31], Jovic-Vranes et al [32], Kamberi et al [34], and Todorovic et al [41] also note that lived environments have an important role in HL development. Kamberi et al [34] argue that rural versus urban environments influence SDoH, such as health care access and quality, thereby, impacting HL development [34]. Beauchamp et al [23], Berens et al [24], Cudjoe et al [26], and Sentell et al [40] observed that HL is also influenced by the patient’s primary language, especially if the patient’s primary language is different from the language of the health system. Bo et al [25] and Pop et al [37] elaborated on the relationship between education and HL; they found that lower levels of language proficiency and self-perceived health can indicate lower HL. Heizomi et al [30] and Dashhti et al [27] notice gender disparities in HL among students in Iran, with the latter observing that cultural differences encouraged technology access for men at a younger age, leading to higher HL levels among men compared to women [27,30].

**HL as a Mediator of SDoH (n=6)**

A total of 6 articles established that, as a mediatory variable, HL can both compensate for and contribute to disparities in SDoH. Some articles define HL as an SDoH itself but further classify it as a mediator for other determinants. All 6 articles included education and income as SDoH for which HL could serve as compensation [58-63]. Bennett et al [59] also suggest that having high HL can compensate for racial or ethnic disparities in health outcomes. Zanchetta et al [63] describe that HL mediates between disparities in health care access and quality as well as social cohesion and context. To address poor HL among patients, van der Heide et al [62] recommend simplifying medical jargon. When designing these interventions, Bennett et al [59] emphasize considering complex patient perspectives and unique demographic needs, such as those of research or by defining determinants as factors that impact or predict health outcomes. Aaby et al [45] classify HL as an SDoH because it is a combination of “personal competencies and situational resources” that affects individuals’ interaction with health care institutions. Some authors, despite describing HL as an SDoH, still note that it is related to other SDoH as well. Cheuhuen et al [47] identify that HL is associated with economic stability and education, and Lee et al [55] and Sentell et al [46] both associate it with the social context. Articles also identified various health outcomes that HL may impact. Cabellos-Garcia et al [48] and Zhou et al [57] identified that poor HL could lead to reduced understanding of disease conditions and engagement with providers. Nevertheless, all 14 articles emphasize that HL is a modifiable SDoH that can change over time through interventions [14,45-57].

**HL as a Modifiable SDoH (n=14)**

The 14 articles that classified HL as an SDoH did so following the World Health Organization’s classification and previous
the geriatric population, which differ from those of younger patients.

**Informatics Interventions**

The secondary objective of the study was to identify informatics-based interventions to improve HL. Articles rarely provided specific intervention recommendations but instead listed several potential problems, such as complicated medical jargon or low health awareness, that complicate patient understanding of health information. Therefore, 4 informatics-based solutions were proposed based on the research team’s knowledge and experience for the identified problems, as follows: (1) language or text simplification, (2) population-focused (or policy-based) interventions, (3) health education efforts, and (4) patient identification. Since the included articles were largely not interventional in nature, the following sections extrapolate on the recommendations with references to ongoing studies that have implemented these strategies.

**Discussion**

**Principal Findings**

This systematic review included 43 papers and reported the results following the PRISMA guidelines. Most studies were conducted in Europe in the past 5 to 10 years. The studies examined HL in relation to the two themes of SD&H—health-focused and demographic—and generated 3 roles for HL, as follows: a mediator of SD&H, a result of SD&H, and modifiable SD&H. More than half of the studies had a cross-sectional design. However, HL is a complex, actionable variable that may be targeted by various strategies.

**Proposed Interventions**

As clinical note sharing becomes more popular, generating interventions that address low HL becomes even more crucial. In this vein, we generated 4 recommendations for focused HL interventions based on the key findings of this systematic review.

Interventions with an informatics focus could play a particularly vital role in improving patient comprehension of health information as the health care field becomes increasingly mobile and technology dependent. It is important to consider experimental methods to measure the efficacy of implementing these strategies. Including control groups and validated HL measuring tools can help monitor how different interventions influence patient HL levels. Validated measuring tools include the Rapid Estimate of Adult Literacy in Medicine (REALM), REALM-Short Form, Short Assessment of Health Literacy-Spanish and English (SAHL-S&E), Brief Health Literacy Screen (BHLS), and Test of Functional Health Literacy in Adults (TOFHLA) [64-66]. The REALM, REALM-SF, and SAHL-S&E have all been validated and recommended by the AHRQ. The REALM and SAHL-S&E are recommended for research purposes to assess participant HL, while the REALM-SF, BHLS, and TOFHLA have been validated for use in screenings in clinical settings [66,67]. The REALM-SF is particularly designed to identify limited literacy levels [67]. Therefore, the clinically usable metrics may be more relevant for interventions that take place in health care settings, such as patient identification.

**Language and Text Simplification**

Text simplification addresses the tendency of clinic notes and health information in general to include medical jargon that exceeds the comprehension levels of most patients [42,44,62]. Even patients with highly educated backgrounds have shown low scores on HL surveys. Therefore, text simplification can benefit patients across all HL competencies by reducing jargon and making health information more easily understandable and usable [62]. Text simplification does not replace the existing clinic note shared between providers; it provides a simplified version for patients in addition to the original note. Current research indicates that the most effective manner of text simplification relies on manual editing techniques using human oversight of a text simplification process, combined with information visualization [68]. Although simplification improves patient comprehension, manual editing could strain health care professionals’ workload. Therefore, developing informatics interventions that automate text simplification while retaining the grammatical and logical integrity of the clinical text is important. Current automated simplification methods scored poorly due to grammatical errors, repetition, and inconsistencies in the autogenerated documents [68]. Artificial intelligence–derived text simplification methods may overcome these barriers by matching a document’s reading level to the readers’ needs, as shown in a study where ChatGPT was able to modify answers to men’s health condition questions to accommodate lower reading levels [69,70]. However, popularly used AI tools, such as ChatGPT, need considerable evaluation to minimize inaccurate information delivery and improve comprehensibility. Current studies indicate that these tools lack citations for the information they provide and cannot differentiate between low-quality and high-quality information [70,71].

**Population-Based Visualization and Cross-Cultural Communications**

HL needs are different across populations and cultural contexts, and interventions should account for these differences. For example, non–English-speaking individuals are overlooked in many HL studies, and interventions targeting English speakers will not always suit those with a limited or nonnative grasp of English [23,72]. Realizing this limitation, the OPHELIA (OPtimising HEalth LIterAc) [73] project is a multisite study that assesses HL strengths and weaknesses in their patient population at each study site and uses these responses to determine appropriate intervention methods. Equally important is including representatives from the community in intervention design. A systematic review looking at interventions that address HL among Aboriginal and Torres Strait Islander community members noted that many failed to include these patients in the design process and consequently had limited participant retention [74]. Another facet is implementing policy-level changes that increase access to HL support. This is particularly relevant for patients who face health inequity. However, implementing these changes has been slow. In the European Union, challenges such as funding constraints and obstacles to
initiatives have prevented effective execution beyond a few countries [75]. The population-based and policy-level interventions should consider visual analytics to explore meaningful patterns in a large data set and use recent advances in natural language understanding and translation to promote cross-cultural communication [76].

**Patient Identification**

Although population-focused and policy-level interventions address low HL at the macro level, such methods may overlook the individual HL needs of a patient. Therefore, screening HL levels as a part of standard practices in health care settings can help identify patients who need additional support at the clinic visit and can expedite provider response [36]. For example, Vanderbilt University Medical Center and the University of Arkansas Medical Sciences incorporate HL screening as part of their educational health assessment and have done so since 2010 and 2016, respectively [77]. Screening may also involve various informatics tools. For example, patients can be actively screened using electronic data capture tools (eg, REDCap) [78]. These informatics tools should be integrated into clinical workflow to ensure the quality of data. On the other hand, patient cohorts can be identified by reports or dashboards of electronic health records or medical text search engines (eg, Electronic Medical Record Search Engine [EMERSE]) [79]. Once the patient group is targeted, inclusive HL interventions can be designed and executed. However, implementing screening practices should be done with caution to avoid perpetuating stigma or embarrassment. Integrating screening questions within the clinical workflow and training health professionals on screening administration can help address these concerns [77].

**Health Education and Online Community Building**

Given the relevance of socialization and environment on HL development, it is important to consider interventions that cultivate HL through health education. Health care providers, such as nurses and community health workers, have important roles in providing education and reinforcing patient understanding of their health conditions [63,80]. However, the burden on health education cannot be placed on providers alone. Health education programs implemented by health care organizations and community health centers can actively and effectively improve HL [81]. It is important to adapt these programs for cultural and demographic sensitivity and patient-provider communications. For example, a recent study targeting older adult needs emphasized the need to include the patient’s caregivers and to accommodate barriers in comprehension, especially cognitive ones [82]. Health education intervention should consider developing an online community, such as ImproveCareNow, to promote collaborative care and build repositories of patient education materials with well-designed education programs to help patients improve their HL [83]. Including the input of individuals who are well-integrated into and familiar with the needs of a patient population, such as community health workers, can also be helpful in this process [80].

**Limitations**

There are a few limitations in the methodology and generalizability of our research. First, we conducted a database search of only PubMed and Scopus, limiting the scope of the article search. However, PubMed and Scopus are two of the most popular and largest databases in biomedical and social science research. During the analysis, it was clear that the results were concise and supported one another. For example, several articles noted multiple roles for HL but tended to focus on one. Second, very few articles included noncorrelated results because of their cross-sectional designs. This prevented researchers from drawing a causative relationship between HL and SDoH, but they nevertheless had hypotheses for relationships that informed our classification. Third, the PRISMA guidelines were updated in 2020 with new standards and recommendations for systematic reviews. As we had already made considerable progress in this project before the revision was published in 2021, we completed the data analysis using the 2015 reporting standards that originally informed our methods. However, in cross-referencing our methods with the 2020 revisions, our research largely adheres to the new guidelines [84].

**Conclusions**

The articles included in this literature review indicate that HL can adopt various roles in conjunction with SDoH. This flexibility makes HL an appropriate topic for intervention to accommodate poor health outcomes and improve patient autonomy. However, the complex nature of HL means that it warrants further research to understand how HL-targeted interventions impact this process.

**Acknowledgments**

The authors would like to acknowledge and thank Ms Elyse Ku at the corresponding author’s lab for her contribution in collecting and reviewing data for the literature review and Ms Somya Pandey for her copyediting effort.

This research was funded by the following intramural programs and awards at the University of Cincinnati: Office of Research Art Humanities Social Science Research Achievement Award, the University Honors Program’s (UHP) Discover program, and the 2021 Summer Undergraduate Research Fellowship Medical Exportation Program (SURF-MedEx) supported by the Medical Sciences Baccalaureate Program (MSBP).

**Conflicts of Interest**

None declared.
References


75. Rademakers J, Hofstede J, van DHI, Devill...


Abbreviations

AHRQ: Agency for Healthcare Research and Quality  
BHLS: Brief Health Literacy Screen  
EMERSE: Electronic Medical Record Search Engine  
HL: health literacy  
OPHELIA: Optimising HEalth LIterAcy  
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses  
RCT: randomized clinical trial  
REALM: Rapid Estimate of Adult Literacy in Medicine  
SAHL-S&E: Short Assessment of Health Literacy-Spanish and English  
SDoh: social determinants of health  
TOFHLA: Test of Functional Health Literacy in Adults
e-Cigarette Tobacco Flavors, Public Health, and Toxicity: Narrative Review

Yehao Sun; Prital Prabhu; Ryan Rahman; Dongmei Li, PhD; Scott McIntosh, PhD; Irfan Rahman, PhD

1Department of Environmental Medicine, University of Rochester Medical Center, Rochester, NY, United States
2Department of Clinical & Translational Research, University of Rochester Medical Center, Rochester, NY, United States
3Department of Public Health Sciences, University of Rochester Medical Center, Rochester, NY, United States

Corresponding Author:
Irfan Rahman, PhD
Department of Environmental Medicine
University of Rochester Medical Center
Box 850
601 Elmwood Avenue
Rochester, NY, 14642
United States
Phone: 1 585 275 6911
Email: Irfan_Rahman@urmc.rochester.edu

Abstract

Background: Recently, the US Food and Drug Administration implemented enforcement priorities against all flavored, cartridge-based e-cigarettes other than menthol and tobacco flavors. This ban undermined the products’ appeal to vapers, so e-cigarette manufacturers added flavorants of other attractive flavors into tobacco-flavored e-cigarettes and reestablished appeal. Objective: This review aims to analyze the impact of the addition of other flavorants in tobacco-flavored e-cigarettes on both human and public health issues and to propose further research as well as potential interventions. Methods: Searches for relevant literature published between 2018 and 2023 were performed. Cited articles about the toxicity of e-cigarette chemicals included those published before 2018, and governmental websites and documents were also included for crucial information. Results: Both the sales of e-cigarettes and posts on social media suggested that the manufacturers’ strategy was successful. The reestablished appeal causes not only a public health issue but also threats to the health of individual vapers. Research has shown an increase in toxicity associated with the flavorants commonly used in flavored e-cigarettes, which are likely added to tobacco-flavored e-cigarettes based on tobacco-derived and synthetic tobacco-free nicotine, and these other flavors are associated with higher clinical symptoms not often induced solely by natural, traditional tobacco flavors. Conclusions: The additional health risks posed by the flavorants are pronounced even without considering the toxicological interactions of the different tobacco flavorants, and more research should be done to understand the health risks thoroughly and to take proper actions accordingly for the regulation of these emerging products.

Introduction

Background

Tobacco flavoring is added to e-cigarettes to make them appealing to vapers, specifically by mimicking the taste of traditional cigarettes. Tobacco-flavored e-cigarettes are often advertised as a safer alternative to traditional cigarettes, which allow smokers to enjoy the taste they are familiar with more conveniently and smoothly for harm reduction. Tobacco-flavored e-cigarettes are very popular among various subpopulations of adults in the United States, with around 30% of vapers using these products [1]. However, the prevalence seems to be lower in dual users (vapers who also use traditional cigarettes) and vapers who used e-cigarettes as an attempt to...
quit smoking, with the percentages being 28.5% and 20.5%, respectively [2,3].

Although the taste of tobacco-flavored e-cigarettes mimics that of traditional cigarettes, the type of nicotine they contain may differ from traditional cigarettes. Recently, e-cigarette products have begun to contain synthetic nicotine or tobacco-free nicotine (TFN), a racemic mixture of both R- and S-nicotine isomers, which is different from the traditionally used tobacco-derived nicotine that is composed of pure S-nicotine [4]. Initially, e-cigarette products began to use TFN since it was not regulated by the US Food and Drug Administration (FDA), and products were able to be brought to the market since these products did not need to go through the premarket tobacco product application for e-cigarettes [4]. Although initially brought to the market without government regulation, in 2022, new legislation expanded the authority of the FDA to regulate TFN products as well [5]. Currently, limited data are available regarding the health effects of TFN, but studies have found that messaging by e-cigarette companies leads to the belief by e-cigarette users that TFN has a lower health risk compared to tobacco-derived nicotine and a higher intention to use TFN products [6]. Young adults between 18 and 25 years of age who were interested in trying TFN believed it to be less addictive than those who were uninterested, and those who have tried TFN reported that TFN products have flavors that taste better and smoother [7]. Similarly, young adults (aged 18-25 years) who were likely to purchase TFN pouches believed that TFN pouches were less harmful to a person’s health; less addictive; and tastes smoother, cleaner, and better compared to young adults who would not purchase TFN pouches [8]. Due to the perception in young adults that TFN is less harmful and addictive, there is a need for more research on the health effects of exposure to TFN to aid government regulation and properly educate the public about any potential risks of using these compounds. More recently, Nixodine (ie, (S)-6-methyl nicotine), which is a structural analog of nicotine, and synthetic-free nicotine or tobacco-free nicotine have been introduced into the market as well without much being known about their biological or toxicological effects.

Besides the use of TFN in tobacco-flavored e-cigarettes, another important modification to these products is the addition of flavorants commonly used in other flavors. On February 6, 2020, the FDA implemented enforcement priorities against all flavored, cartridge-based e-cigarettes other than tobacco- and menthol-flavored products [9]. According to Rostron et al [10] in 2020, as much as 93.2% of youth vapers started vaping with a flavored e-cigarette, and among those who are still vaping, 71% reported the flavors of e-cigarettes as a reason for use. It was also indicated that youth vapers preferred fruit and mint flavors to tobacco or menthol flavors [11]. The tobacco flavors of e-cigarettes are made to mimic the flavor of traditional tobacco cigarettes with some variation. There are many different tobacco flavors made from hundreds of brands that can provide the user with different types of tobacco flavors, including “Classic Tobacco,” “Smooth/Bold Tobacco,” and “Virginia Tobacco.” Demographically, tobacco flavors are more popular among adults and less popular among youth [12]. The lack of appeal of tobacco-flavored e-cigarettes to youth allows for fewer regulations. Therefore, we perceived that the ban on flavors other than tobacco and menthol undermined the e-cigarette products’ appeal to youth vapers, as their favorite flavors were removed from access, thus largely decreasing the manufacturers’ profit. To reverse the impacts brought by the difference in regulations, e-cigarette manufacturers started to blend other flavors into tobacco-flavored e-cigarettes, recreating the appeal for youth vapers [13,14]. For example, we found that an e-cigarette manufacturer has a fourth-generation e-cigarette product with a “Smooth Tobacco” flavor, which contains a combination of tobacco and cream flavors. The same entity also sells an e-liquid of “Tobacco Salt Rich” flavor, which is a mixture of tobacco, smokey vanilla, and creamy caramel flavors. Studies have also extracted flavorants that represent sweets and caramel-like flavors in an e-liquid marked “Smooth & Mild Tobacco” and multiple flavorants that do not belong to tobacco flavors in another tobacco-flavored e-liquid that was deidentified [13,15,16]. Such compounds include ethyl maltol, vanillin, corylone, and ethyl vanillin, which can lead to adverse health effects [14]. Additionally, the volatile organic compounds (VOCs), reactive oxygen species, and other compounds present in the tobacco-flavored e-liquids can pose further health risks.

**Objective**

The emergence of these new tobacco flavors may serve as a source for public health issues, and information related to them is critical for the establishment of regulations and interventions. Therefore, by analyzing the toxicity, characteristics, sales, social media perception, and public health aspects of tobacco-flavored e-cigarettes, this review aims to inform authorities about this issue and provide information for potential interventions (Figure 1).
Methods

To collect data, searches were conducted on Google Scholar and PubMed for papers published between 2018 and 2023 related to e-cigarette use patterns, toxicity of e-cigarette chemicals, social media posts about e-cigarettes, and public health interventions regarding e-cigarettes. Toxicity information was also included from articles published before 2018, and e-cigarette sales data and related policies were extracted from government websites and documents. The keywords for searching these sources of information included “tobacco-flavored e-cigarettes,” “e-cigarette use,” “synthetic nicotine,” “flavorants,” “e-cigarette policy,” “social media and vaping,” “vaping cessation,” and chemical names mentioned in this review.

The extracted information was discussed to identify the appeal of tobacco-flavored e-cigarettes based on sales data, to document the toxicity complications from toxicology studies, and to confirm the impacts of the addition of other flavorants in tobacco-flavored e-cigarettes by analyzing studies on related social media posts (Figure 1).

Results

e-Liquid Constituents Inhaled During Vaping

Tobacco-flavored e-cigarettes have a wide range of chemicals in the e-liquid, and different tobacco flavors have different flavoring agents. However, in general, tobacco-flavored e-cigarettes contain propylene glycol, glycerol, and 0-50 mg/mL of nicotine (in the form of freebase nicotine or nicotine salts), similar to most other e-cigarettes. Tobacco-flavored e-cigarettes have also been shown to have cinnamaldehyde [17]. Additionally, for the popular brands JUUL and Puff Bar, many other chemicals were frequently found to be in their tobacco-flavored e-cigarettes in greater than 1 mg/mL concentrations, including ethyl maltol, corylone, vanillin, and ethyl vanillin [14]. Another study found caffeine, isophorone, tributyl O-acetylcitrate, tributylphosphine oxide, triethyl citrate, and vanillin in tobacco-flavored e-liquids from popular brands such as JUUL, Blu, Smok, and Vuse Alto [18]. There are also many VOCs present in tobacco flavors such as ethanol, toluene, ethylbenzene, and styrene [17]. Moreover, tobacco flavors would also produce reactive oxygen species that cause oxidative stress when used. Overall, there are many different carbonyls, citrates, phenols, VOCs, and other organic compounds present in tobacco-flavored e-liquids and their combustion and degradation products that are inhaled during vaping.

Cellular Toxicities of Tobacco-Flavored e-Cigarette Aerosols

Existing studies have established some knowledge of the toxicities of tobacco-flavored e-cigarettes [16,19-24]. The compounds present in the e-liquid and aerosol of tobacco-flavored e-cigarettes have many toxic effects on cells. For instance, nicotine in tobacco flavors can induce mucus hypersecretion by goblet cells and decrease mucociliary clearance in the lung by suppressing α7 nicotinic acetylcholine receptor activity and cystic fibrosis transmembrane conductance regulators, resulting in a greater risk for chronic lung diseases [19]. It was revealed that tobacco flavorants can induce oxidative stress, inflammation, DNA damage, and higher levels of cell death in lung epithelial cells and inflammatory responses in different types of cells including fibroblasts [20,21]. Overall, reported in either in vivo or in vitro studies, increased reactive oxygen species or oxidative stress and the release of inflammatory cytokines were associated with tobacco flavors, and the conclusions included increased cell death, decreased cell viability, and increased inflammatory responses [22].

Mechanisms of Disease Pathogenesis Related to Toxicities of Tobacco-Flavored e-Cigarette Aerosol

Beyond cells, tobacco-flavored e-cigarettes are harmful to the user’s overall health. Inhaling nicotine from tobacco-flavored e-cigarettes can result in hypertension, chronic obstructive pulmonary disease (COPD), increased myocardial infarction...
risk, and asthma [23]. The propylene glycol found in tobacco-flavored e-cigarettes can also pose health risks when inhaled, where cough, difficulty breathing, and increased asthma risk are linked to the inhalation of propylene glycol [23]. Moreover, the heating of glycerol found in tobacco-flavored e-liquids can produce formaldehyde, which can act as a carcinogen when inhaled [23]. In another study, it is also shown that tobacco flavor accompanied by the presence of nicotine can induce an allergic inflammatory response, characterized by elevated levels of eotaxin, interleukin-6, and chemokine (C-C motif) ligand 5 (also known as RANTES) [16]. The combination can also increase the level of plasminogen activator inhibitor-1, with a higher level being a risk factor for thrombosis and atherosclerosis [16,24]. Additionally, the reactive oxygen species and VOCs present in tobacco-flavored e-cigarettes can increase exposure to free radicals, resulting in oxidative stress and lung inflammation [19]. Overall, the inhalation of compounds present in tobacco-flavored e-cigarettes poses a serious health risk and can increase lung toxicity and the likelihood of various chronic lung diseases ranging from COPD to cardiovascular disease (Figure 1).

**Tobacco-Flavored e-Cigarette Products**

Although the flavors are limited to tobacco flavors, there is still a variety of e-cigarette devices with distinct characteristics associated with tobacco flavors [25,26]. Generally, e-cigarette devices are divided into 4 generations, all of which can support tobacco flavors [26].

First-generation e-cigarettes are designed to mimic the appearance of traditional cigarettes and thus are also known as cig-a-likes [25,26]. The major components are a battery, an atomizing unit, and a fluid reservoir (cartridge) [26]. Although outdated, tobacco-flavored e-cigarettes of the first generation can still be found in some web-based and physical vape shops.

In second-generation e-cigarettes, the cartridge is replaced by a “clearomizer” installed in a pen-shaped device, so second-generation e-cigarettes are also called “vape pens” [25,26]. Third-generation e-cigarettes, on the other hand, are highly customizable and contain sub-ohm tanks, which allow even higher wattage due to decreased resistance [25,26]. Both second- and third-generation e-cigarettes use e-liquids for aerosol generation, and tobacco-flavored e-liquids can be easily found in web-based vape shops and are sold in large amounts.

Fourth-generation e-cigarettes are called “Pod-Mods,” indicating a modifiable pod cartridge that can be in various shapes [25,26]. Fourth-generation e-cigarettes use nicotine salts instead of the freebase nicotine used in previous generations, allowing a higher concentration of nicotine to be present [25]. A popular variation named “vape bars” is the most popular product in web-based vape shops.

Tobacco-flavored products associated with all the generations discussed above are widely available web-based vape shops for vapers, and the products are sold in large amounts. In web-based vape shops, the best-selling tobacco-flavored e-cigarette products are mostly vape bars (fourth-generation devices), followed by tobacco-flavored e-liquids (used by second- and third-generation devices). First-generation products and prefilled cartridges or pods (second-generation products) can also be found in another vape shop, where it claims that the first-generation product is “the new #1 selling e-cigarette on the market.” The vape shop selling primarily fourth-generation e-cigarettes has a better website design with different fonts that may attract young vapers, whereas the vape shop website that sells first- and second-generation e-cigarettes looks relatively old.

**Public Perceptions of Tobacco-Flavored e-Cigarettes on Social Media**

An examination of the public perceptions of different e-liquid flavors on over 2 million e-cigarette–related Twitter (subsequently rebranded as X) posts from May 31 to August 22, 2019, showed the public had a more negative attitude toward the tobacco flavor (sentiment score=−0.134) using sentiment analysis [27]. Meanwhile, it was also found that the public was positive toward fruit (sentiment score=0.074) and sweets flavors (sentiment score=0.156), and most of the discussions were about these 2 flavors (58.15% and 14.67%, respectively) [27]. Immediately after the flavor ban, only menthol and tobacco flavors were allowed on the market, and an increase in discussion about menthol flavors (from 16.4% to 37%) was observed [9,28]. However, there was no significant increase in discussion about tobacco flavors, indicating that vapers likely did not choose to shift to tobacco flavors immediately after the ban of their favorite flavors [28]. In contrast, the discussion of fruit and sweets flavors remained high after the ban and even increased around 5 months later (from 41% and 22.3% before the ban to 57% and 28% five months after the ban, respectively), signaling that the vapers might have sought other sources for their favorite flavors after they were banned, which indicates continued interest in these flavors [28].

Through applying generalized estimating equation (GEE) models on over 3000 Reddit posts from January 2013 to April 2019 that mentioned e-cigarette use and health symptoms in the same Reddit post, it was found that tobacco flavor was more likely to be mentioned with respiratory and throat symptoms than other symptoms [29]. A specific examination of the JUUL pod tobacco flavor with health symptoms, using similar GEE models and Reddit posts from September 2016 to April 2019, showed a high probability of the mention of the JUUL tobacco flavor with throat, respiratory, and cardiovascular symptoms [30].

**e-Cigarette Sales After Flavor Ban Regulations and Flavorants’ Appeal to Vapers**

The vast variety of e-cigarette flavorings, such as banana, mango, and cotton candy, are extremely appealing to the younger generation, which supports the nicotine addiction epidemic among today’s youth. However, the February 2020 FDA ban on flavored prefilled e-cigarette cartridges, while having the intention of curbing flavored e-cigarette use, also opened new doors for the vaping industry to continue making profits [9]. This was due to 2 loopholes in the FDA policy: the ban did not cover the sale of tobacco- and menthol-flavored prefilled cartridges or the sale of flavored disposable e-cigarettes [31]. For these reasons, e-cigarette users were able to find...
alternatives to flavored prefilled cartridges, such as the tobacco-flavored e-cigarettes outlined in this paper.

Centers for Disease Control and Prevention data show that after the FDA policy enactment, the unit share of disposable e-cigarettes went from 29.9% to 49.6%, whereas the respective unit share of prefilled cartridges lowered from 70% to 50.3% between February 2020 and July 2022 [31]. This data show the popularity of flavored e-cigarettes in the vaping population, with them quickly switching to disposable e-cigarettes once flavored prefilled cartridges became unavailable. Additionally, although the FDA ban was supposed to limit prefilled cartridge manufacturers such as JUUL from profiting off of nicotine addiction, it allowed disposable vaping brands, such as Puff Bar, Elf Bar, and Blu, to achieve a massive increase in sales by developing products that filled the “flavoring hole” left by the prefilled cartridge ban. Data showed that in response to these holes, e-cigarette users largely switched to disposable devices rather than continuing to buy the tobacco- and menthol-flavored cartridges still on the market [31]. After the 2020 ban up until July 2022, tobacco-flavored cartridge sales only increased by 11.9%, whereas all other flavor sales increased by 75.6% [31], showing the preference of the vaping population for nontobacco flavorings, which indicates that vapers are likely to be attracted to the new tobacco flavors that contain flavorants from other flavors.

Public Health Interventions Associated With Tobacco Flavors

Flavors have been cited as a key factor for the initiation of vaping by adolescents and young persons and facilitate the ongoing use of vaping products by those of all ages. Flavored vaping products are alluring to both new and established tobacco product users, and a wide variety of flavors are available. This wide variety and the ability to combine different flavors, in this case, the addition of other flavorants into tobacco flavors, could contribute to the ongoing vaping behavior among both youth and adults [12,32].

Per the FDA “Deeming” regulations, the FDA can now regulate the presence and amount of “characterizing flavors” in vaping products [33]. According to former FDA Commissioner Dr. Scott Gottlieb, e-cigarette use among youth can be characterized as an epidemic [34]. Users must be at least 18 years of age to buy vaping products in most states, but those younger than 18 years old are still able to purchase from a variety of retailers and web-based vendors [12,33].

To address the vaping epidemic, especially among youth, in 2021, the FDA implemented a flavor enforcement policy to restrict the sales of all cartridge-based, unauthorized, flavored e-cigarettes other than tobacco and menthol flavors [35,36]. Evaluation of the impact of the FDA flavor enforcement policy on e-cigarette use behavior is in progress. One study assessed the potential impact of the flavor enforcement policy on a specific vaping-related behavior change—quitting vaping—using natural language processing strategies with data collected from the Twitter platform [35]. The proportion of tweets (and Twitter users’ mentions) concerning quitting vaping was compared before and after the implementation of the FDA flavor policy [35]. Compared to before the FDA flavor policy, the proportion of tweets and Twitter user mentions after the implementation of the policy was higher [35]. They also reported that after the policy implementation (compared to before), there was an increasing trend in the proportion of female individuals and young adults (18-35 years old) mentioning quitting vaping [35]. They concluded that, as observed on Twitter, the FDA policy did have a positive effect on vaping cessation and therefore a potential influence on broader definitions of vaping behavior [35].

Another public health intervention for vaping cessation is the use of free vaping cessation apps, which have various content, features, and adherence to evidence-based approaches. In 2020, researchers conducted a systematic search of existing smartphone apps for vaping cessation [37]. A total of 8 apps were included in a quality assessment and content analysis. They concluded that the few existing vaping cessation apps use similar approaches to smoking cessation apps but are potentially valuable tools [37].

Discussion

Toxicological Complexities Brought by the Addition of Other Flavorants

Besides the toxicities of tobacco-flavored e-cigarette constituents, the introduction of other previously irrelevant chemicals may inevitably complicate the toxicity of these products. The most commonly used flavorant (in 35% of e-liquids), vanillin, is responsible for vanilla flavors in e-liquids; is likely to be present in the “Tobacco Salt Rich” e-cigarette; and was extracted from the deidentified tobacco-flavored e-cigarette introduced earlier in this paper [16,38,39]. As shown, the presence of vanillin has a positive correlation with the toxicity of e-liquids (R²=0.62) [38]. The vanillin in tobacco flavors is inflammatory and can irritate airways [19]. Another popular flavorant (in 32% of e-liquids) present in caramel flavors, ethyl maltol, was also present in the deidentified tobacco-flavored e-cigarette and has been shown to be a contributing factor for incidences of kidney lesions in rats and mild hemolytic anemia in dogs [15,16,39,40]. Furthermore, the inhalation of cinnamaldehyde and ethyl maltol, compounds found in tobacco flavors, causes oxidative stress and can lead to inflammation and epithelial barrier dysfunction, increasing the risk of diseases such as COPD [19]. These are only 2 of the flavorants used in e-liquids, and the typical number of different flavorants in a single e-liquid product would be higher than 10 [38]. It was found that the more chemicals there are in the e-liquid, the higher the toxicity that the e-liquid is likely to possess [38]. Therefore, it is predicted that the additional flavorants in tobacco-flavored e-cigarettes, which already contained many kinds of flavorants, would increase the overall toxicity of the product, and it would be hard to figure out the interactions of the toxicity mechanisms related to flavorants that originally belonged to completely unrelated species. More studies are required to fully understand this complexity and take appropriate actions regarding the regulation.
Youth Vapers’ Preferences

According to scientific studies, e-cigarette users’ preferences for e-cigarette devices were shifting toward newer-generation devices: Fourth-generation devices (prefilled pod cartridges) are the most used devices, although third-generation devices still take up a considerable proportion of use [41]. It was also observed that the shift toward newer generations is faster in youth users than in young adults or older adults [41]. Another study conducted by Lin et al [42] also agrees with this finding, as they found that adolescent and young adults’ preference is responsive to advancements in e-cigarette technology. They generally avoid using earlier-generation devices (the percentage of users who usually used disposable or large-size rechargeable e-cigarettes dropped from 88.2% to 33.1% during the study) and prefer more innovative products (the percentage of users who usually used pod-based e-cigarettes, which were only introduced into the market when the study was halfway through, was 22.3% by the end of the study) [42]. The trend found by those studies is likely applicable to tobacco-flavored e-cigarettes, as the characteristics of web-based vape shops discussed above match the trend [41,42]. The fact that youth vapers shifted to pod-based e-cigarettes quickly also made the addition of other flavorants in these products a more significant public health issue.

Implications From Social Media Studies

According to results from the social media studies mentioned earlier, vapers demonstrated continued interest in fruit and sweets flavors immediately after the flavor ban while remaining uninterested in original tobacco flavors [27,28]. Our web-based survey study also showed that most vapers continued using flavored e-cigarettes even after the flavor ban, as disposable e-cigarettes were not covered by the FDA flavor ban [43]. Therefore, when their favorite flavors are integrated back into tobacco flavors, it is expected that they would prefer the mixed flavor. Since the availability of flavors was among the top reasons for vaping and its initiation, especially in adolescents and young adults, the addition of these flavors in tobacco flavors would likely resuscitate the motivation for vapers to continue to vape [44,45].

Social media research that focused on health issues comentioned with flavors discovered that tobacco flavors were generally more likely to be comentioned with respiratory and throat symptoms, and cardiovascular symptoms were also frequently comentioned if the tobacco-flavored e-cigarettes were from JUUL [29,30]. These results are associated with traditional, tobacco-flavored cigarettes prior to the addition of new flavors, and the addition might be associated with more complicated symptoms. In the web-based vape shop, we found that for new tobacco-flavored e-cigarettes, the best-selling ones often contained new flavors categorized as “sweets” flavors or the “creme” flavor in JUUL products [29,30]. According to the same GEE models, “sweets” flavors were comentioned with throat and digestive symptoms, whereas JUUL’s “creme” flavor was comentioned with neurological, digestive, and “other” symptoms, which were not observed in the corresponding tobacco flavors [29,30]. However, the comention of flavors with health symptoms does not indicate that vaping will cause these symptoms, as it is also possible that vaping could reduce the health symptoms. Previous study showed that the toxicological effects of the flavorants may interact with each other, and the effects of such interactions are unknown [46]. Therefore, more research should be done to further understand the symptoms associated with the addition of other flavors into tobacco-flavored e-cigarettes.

Overall, as we observed more varieties of tobacco-flavored e-cigarettes sold in vape shops, the public perceptions of tobacco-flavored e-cigarettes and their associations with health symptoms mentioned on social media need to be revisited.

Vaping Communities and Flavor Addition to Tobacco Flavors

Since vapers can belong to a variety of different communities, the addition of other flavors into tobacco-flavored e-cigarettes may have different effects in these communities, and we need to focus on the differences. For example, the vaping behaviors of dual users of both traditional cigarettes and e-cigarettes are different from vapers who only use e-cigarettes [2]. Dual users usually only use e-cigarettes when they are engaging in activities or in places that encourage e-cigarette use, or when they use e-cigarettes as substitutes for traditional cigarettes [47]. This type of difference becomes exceedingly important when there is a relatively high prevalence of vaping in the community (including minority youth) or when the community is our major target of protection (including age groups such as adolescents) [48]. For instance, among young adult e-cigarette users, bisexual women were the most susceptible to e-cigarette use habits with a high level of cigarette use [49]. Such disparities are of importance when public health interventions are tailored, so knowing how specific communities respond to the addition of flavorants in tobacco-flavored e-cigarettes is critical. However, despite this importance, there is minimal data on this issue, and the differential effects remain unknown to us. Further studies should be done on these specific communities for us to comprehensively understand how new tobacco-flavored e-cigarettes impact the entire vaping population and establish regulations accordingly.

Conclusion

After the FDA implemented enforcement priorities against all flavored, cartridge-based e-cigarettes other than tobacco- and menthol-flavored products on February 6, 2020, most e-cigarette products became regulated, leaving only menthol and tobacco flavors to be widely and legally available for vapers [9]. This ban on other flavors impaired e-cigarettes’ appeal to vapers, so e-cigarette manufacturers decided to recreate similar flavors by blending the corresponding flavorants into tobacco-flavored e-cigarettes to form variant tobacco flavors, including “Smooth Tobacco” [13,16]. These mixed tobacco flavors are now widely available in web-based vape shops, and the products either come as or can be used in any generation of e-cigarettes to accommodate the preference of vapers in different age groups (it is inferred that younger vapers’ preferences switch to more innovative products more easily and they generally use newer-generation devices) [41,42].
Evidence from both the vaping market share and social media posts indicate that the manufacturers’ strategy was successful [28,31]. After the FDA regulation, the unit share of prefilled cartridges decreased, and the sales of disposable e-cigarettes of flavors other than tobacco flavors increased dramatically, indicating a strong preference for flavorants in other flavors that motivated the vapers to switch to disposable e-cigarettes [31]. Therefore, the addition of these flavorants into tobacco flavors may establish appeal to the new tobacco flavors. On the other hand, similar trends were found in social media posts, showing that fruit and sweets flavors were still often discussed after the flavor ban policies [28]. The heated discussions indicate the vapers’ strong craving for these flavors, so this further confirms that the addition of other flavorants into tobacco flavors may successfully attract vapers.

This strategy by the manufacturers can not only lead to new public health issues but also new health risks and symptoms in individual users, and it even raises issues on harm reduction approaches. The additional flavorants mixed in the new tobacco-flavored e-cigarettes may have unique toxicology mechanisms that are not observed in flavorants used in traditional tobacco flavors. For example, vanillin and ethyl maltol are likely found in a product with the flavor “Tobacco Salt Rich” and another deidentified tobacco-flavored e-cigarette, and these flavorants have been shown to increase the toxicity of e-liquids and induce incidences of kidney lesions in rats and mild hemolytic anemia in dogs, respectively [15,16,38-40]. Other flavorants may also be integrated into the recipe of tobacco-flavored e-cigarettes, and it has been shown that the toxicity of the e-liquids increases as the number of chemicals in its recipe increases [38]. Meanwhile, in the analysis of Reddit posts using GEE models, the “sweets” flavors in e-cigarettes were associated with higher comention of digestive and throat symptoms, which are not demonstrated in traditional tobacco flavors [29]. Therefore, the symptoms associated with e-cigarette use are likely to be more complicated when using the new tobacco-flavored e-cigarettes. However, our predictions of toxicology and symptoms are based on the simple addition of effects, where the interactions between the flavorants were not taken into consideration. More research needs to be done to fully understand the interactions and the overall effects.

Besides the public health issues and personal health risks associated with the addition of flavorants in tobacco-flavored e-cigarettes, the FDA flavor ban policies did have an overall positive effect in helping vapers quit vaping [36]. The use of the new vaping cessation apps is also a potentially important aspect of public health interventions [37]. To further extend the positive effects, more research should be done to analyze the effects brought by the manufacturers’ efforts to bypass the regulations and premarking approval, and emphasis should be placed on vulnerable communities regarding the vaping public health effects.

Acknowledgments

Mr. Thomas Lamb (University of Rochester Medical Center) provided help in discussions on tobacco flavors. Dr. Gagandeep Kaur provided help and suggestions to reviewers’ comments. We also thank Alannah Dalton, Chad Newton, and DJ Robinson for insightful discussions. A part of the manuscript was deposited in the preprint form on Preprints.org [50]. This study was supported by the National Cancer Institute (NCI) and the (FDA; #U54CA228110).

Authors’ Contributions

YS, DL, PP, RR, SM, and IR contributed to writing—original draft preparation. YS and IR contributed to writing—review and editing and the preparation of schematics and conceptual diagrams. IR contributed to supervision, project administration, and funding acquisition and compilation. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

4. Jordt SE. Synthetic nicotine has arrived. Tob Control 2023 Apr 07;32(e1):e113-e117 [FREE Full text] [doi: 10.1136/tobaccocontrol-2021-056626] [Medline: 34493630]


36. FDA finalizes enforcement policy on unauthorized flavored cartridge-based e-cigarettes that appeal to children, including fruit and mint. U.S. Food and Drug Administration. 2020 Jan 2. URL: https://tinyurl.com/ye2y5t2h [accessed 2023-05-23]


43. Li D, Ossip DJ, Bansal-Travers M, Xie Z. Impact of the FDA flavour enforcement policy on flavoured electronic cigarette use behaviour changes. Tob Control 2022 Nov;31(Suppl 3):s176-s183 [FREE Full text] [doi: 10.1136/tc-2022-057492] [Medline: 36328457]


Abbreviations

COPD: chronic obstructive pulmonary disease
FDA: Food and Drug Administration
GEE: generalized estimating equation
TFN: tobacco-free nicotine
VOC: volatile organic compound
Corrigenda and Addenda

Correction: Machine Learning Model for Predicting Mortality Risk in Patients With Complex Chronic Conditions: Retrospective Analysis

Guillem Hernández Guillamet1,2, BSc, MSc; Ariadna Ning Morancho Pallaruelo2, BSc; Laura Miró Mezquita1,2, BSc, MSc; Ramón Miralles3,4, MD, PhD; Miquel Àngel Mas3,4, MD, PhD; María José Ulldemolins Papaseit5,6, MD; Oriol Estrada Cuxart1,2, MD, PhD; Francesc López Seguí2,7, MSc, PhD

1Research Group on Innovation, Health Economics and Digital Transformation, Institut Germans Trias i Pujol, Badalona, Spain
2Hospital Germans Trias i Pujol, Institut Català de la Salut, Badalona, Spain
3Direcció Clínica Territorial de Cronicitat Metropolitana Nord, Institut Català de la Salut, Badalona, Spain
4Department of Geriatrics, Hospital Germans Trias i Pujol, Badalona, Spain
5Direcció d’Atenció Primària Metropolitana Nord, Institut Català de la Salut, Badalona, Spain
6Servei d’Atenció Primària Metropolitana Nord, Institut Català de la Salut, Barcelona, Spain
7Chair in ICT and Health, Centre for Health and Social Care Research (CESS), University of Vic - Central University of Catalonia (UVic-UCC), Vic, Spain

Corresponding Author:
Francesc López Seguí, MSc, PhD
Chair in ICT and Health, Centre for Health and Social Care Research (CESS)
University of Vic - Central University of Catalonia (UVic-UCC)
Carrer Miquel Martí i Pol, 1
Vic, 08500
Spain
Phone: 34 938863342
Email: francesc.lopez.segui@gmail.com

Related Article:
Correction of: https://ojphi.jmir.org/2023/1/e52782
doi:10.2196/58453

In “Machine Learning Model for Predicting Mortality Risk in Patients With Complex Chronic Conditions: Retrospective Analysis” (Online J Public Health Inform 2023;15:e52782) the authors noted one error.

The Corresponding Author’s address previously appeared as:

Chair in Palliative Care, Centre for Health and Social Care Research
University of Vic - Central University of Catalonia
Carrer Miquel Martí i Pol, 1
Vic, 08500
Spain

This has been changed to:

Chair in ICT and Health, Centre for Health and Social Care Research (CESS)
University of Vic - Central University of Catalonia (UVic-UCC)
Carrer Miquel Martí i Pol, 1
Vic, 08500
Spain

The correction will appear in the online version of the paper on the JMIR Publications website on March 21, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.
Original Paper

Intention to Use Mobile-Based Partograph and Its Predictors Among Obstetric Health Care Providers Working at Public Referral Hospitals in the Oromia Region of Ethiopia in 2022: Cross-Sectional Questionnaire Study

Kefyalew Naniye Tilahun¹, MPH; Jibril Bashir Adem², MPH; Wabi Temesgen Atinafu¹, MPH; Agmasie Damtew Walle³, MPH; Nebyu Demeke Mengestie⁴, AST; Abraham Yeneneh Birhanu⁴, AST

¹College of Medicine and Health Sciences, Ambo University, Ambo, Ethiopia
²Department of Public Health, Arsi University, Asella, Ethiopia
³Department of Health Informatics, Mettu University, Mettu, Ethiopia
⁴Department of Health Informatics, Institute of Public Health, University of Gondar, Gondar, Ethiopia

Corresponding Author:
Kefyalew Naniye Tilahun, MPH
College of Medicine and Health Sciences
Ambo University
Addis Ababa
Ambo
Ethiopia
Phone: 251 922298660
Email: kefyalewnaniye64@gmail.com

Abstract

Background: A partograph is a pictorial representation of the relationship between cervical dilatation and the time used to diagnose prolonged and obstructed labor. However, the utilization of paper-based partograph is low and it is prone to documentation errors, which can be avoided with the use of electronic partographs. There is only limited information on the proportion of intention to use mobile-based partographs and its predictors.

Objective: The objective of this study was to determine the proportion of obstetric health care providers at public referral hospitals in Oromia, Ethiopia, in 2022 who had the intention to use mobile-based partographs and to determine the predictors of their intention to use mobile-based partographs.

Methods: We performed an institution-based cross-sectional study from June 1 to July 1, 2022. Census was conducted on 649 participants. A self-administered structured English questionnaire was used, and a 5% pretest was performed. Data were entered into EpiData version 4.6 and exported to SPSS version 25 for descriptive analysis and AMOS (analysis of moment structure; version 23) for structural and measurement model assessment. Descriptive and structural equation modeling analyses were performed. The hypotheses developed based on a modified Technology Acceptance Model were tested using path coefficients and P values <.05.

Results: About 65.7% (414/630; 95% CI 61.9%-69.4%) of the participants intended to use mobile-based electronic partographs, with a 97% (630/649) response rate. Perceived usefulness had a positive influence on intention to use (β=.184; P=.02) and attitude (β=.521; P=.002). Perceived ease of use had a positive influence on attitude (β=.382; P=.003), perceived usefulness (β=.503; P=.002), and intention to use (β=.369; P=.001). Job relevance had a positive influence on perceived usefulness (β=.408; P=.001) and intention to use (β=.185; P=.008). Attitude positively influenced intention to use (β=.309; P=.002). Subjective norms did not have a significant influence on perceived usefulness (β=.020; P=.61) and intention to use (β=−.066; P=.07).

Conclusions: Two-thirds of the obstetric health care providers in our study intended to use mobile-based partographs. Perceived usefulness, perceived ease of use, job relevance, and attitude positively and significantly influenced their intention to use mobile-based electronic partographs. The development of a user-friendly mobile-based partograph that meets job and user expectations can enhance the intention to use.

DOI: 10.2196/51601
mobile-based partograph; mHealth; mobile health; cross-sectional; questionnaire; questionnaires; survey; surveys; modified TAM; technology acceptance model; intention to use; obstetric health care providers; Ethiopia; intent; intention; TAM; experience; experiences; attitude; attitudes; opinion; opinions; perception; perceptions; perspective; perspectives; acceptance; adoption; partograph; digital health; health technology; birth; women's health; obstetrics; obstetric; obstetric health care; labor monitoring

**Introduction**

**Background**

Globally, maternal mortality remains a persistent and potentially preventable issue of great concern. In 2020, every 2 minutes, a woman died due to pregnancy-related preventable causes, indicating that about 800 women died every day, resulting in a maternal mortality ratio of 223 deaths per 100,000 live births [1]. By lessening maternal mortality to roughly 70 per 100,000 live births between 2016 and 2030, the Sustainable Development Goal initiatives hope to avert the maternal mortality rates [2]. Contrary to goals set according to the Sustainable Development Goals 3.1, the global maternal mortality rate increased from 151 in 2019 to 152 per 100,000 live births in 2020 [3]. In Ethiopia, maternal death rate was 412 per 100,000 live births in 2016 [4]. It stayed high, accounting for 412 per 100,000 live births in 2019 [5]. In Ethiopia, prolonged and obstructed labor account for 22% of all maternal deaths [6]. Although prolonged and obstructed labors are among the leading causes of death in resource-poor settings, they can be diagnosed and averted with correct partograph use [7,8]. A partograph is a graphic representation of the labor’s progress that includes pertinent information about the mother and the fetus [9]. In this regard, one of the most important ways in assuring high-quality care for both the mother and the newborn during labor is to use the partograph [10].

Despite its significance, partograph use by obstetric health care providers is still low in Ethiopia [6,7,11-15]. In addition to this, the paper-based approach is prone to recording errors due to health care providers’ overburdened and retrospective data entry [16]. Paper-based partographs are also exposed to incomplete reporting of parameters. In Uganda, only 24.6% of the partograph parameters demonstrated complete details [17]. In a study in Jigjiga and Degehabur, about 64% of the partograph characteristics were only partially recorded [11]. A study in the West Shoa zone revealed that only 3% of the partographs examined was recorded according to the standard [7]. The rapid progress of technology is one of the many drivers now impacting health care systems [18], and offering health care services via mobile devices is now seen as a promising technological advancement [19]. The widespread availability of smartphones and tablets provides an opportunity for the use of a well-designed electronic partograph [16]. Digitizing the partograph improves adherence and overcomes the limitation of paper-based partographs [10,16,20,22]. The electronic partograph is a contemporary instrument for capturing labor data in real time, which can improve mother and infant outcomes [23]. Electronic partographs make it possible to improve the labor management system’s record of labor progress statistics and care given to mother and fetus, especially in low-income nations [10,20,24,25]. Electronic partographs also result in a significant reduction in the rate of prolonged labor from 42% to 29% and have a far greater usage rate than paper-based partographs [21]. In addition, electronic partographs are preferred over paper-based partographs by clinicians due to their ease of use and less time, improved performance, decreased referral rates, assured prompt referral when necessary, facilitation of reporting obligations, and enhancement of service quality [26,27].

Studies show that using mobile-based health services in the health sector have the potential to increment health service access, quality, adherence, and efficiency [28-35]. However, the technological benefit obtained depends on the rate of use and adherence of users. Human activities mainly depend on their behavioral intention, and the intention to use digital tools is a determinant factor of actual user behavior. Therefore, determining the behavioral intention to use and its predictors before the adoption of technology is important and prevents implementation failure [36]. Behavioral intention is the degree to which a person has made intentional plans to engage or abstain from engaging in a specific future conduct [37]. To facilitate future implementation, it is vital to ascertain the degree of intention to employ any digital tools in the health sector [36]. A variety of Technology Acceptance Models (TAMs) have been applied to identify and predict end user behavioral intention to utilize technology. Among these, Davis’s TAM is significant and effective at predicting users’ intent to use [38].

To increase understanding of the factors affecting behavioral intention, different scholars have modified the original TAM [39-45]. Therefore, we modified the TAM to increase the understanding of the predictors that influence the behavioral intention of obstetric health care providers to use mobile-based partographs because information on the proportion of intention to use mobile-based partographs and its predictors is limited. This study was intended to fill this gap.

**Theoretical Model and Hypothesis Development**

TAM is the most well-known methodology for establishing and evaluating each person’s intent to embrace new technology [46]. It is a commonly used model that is used to anticipate possible users’ behavioral intentions to use a technological innovation [47]. TAM had been altered, nevertheless, to boost its capacity for predicting variations in usage intention. Studies on doctors’ approval of digital personal aids in Turkey [48], smart health care services among medical practitioners in China [49], health information system acceptance by hospital personnel in Greece [50], intention to use technology to attend to clients by health professionals in Ghana [51], and sustainable adoption of eHealth systems by health care professionals in Ethiopia [39] by using modified TAM report about 71%, 71.5%, 87%, 97%, and 56% of variance in intention to use, respectively.

TAM was initially established with 2 key dimensions termed as perceived utility and perceived ease of use to identify the potential drivers of intention. The extent to which a person...
thinks using a given system would improve his or her performance at work is known as perceived usefulness [38]. According to studies [18,38,39,48,52] in health care settings, perceived usefulness had a significant and positive influence on intention to use. Perceived ease of use is another factor that establishes the end users’ behavioral intentions. The degree to which someone perceives a system to be simple to use is known as perceived ease of use [38]. The greater the user’s propensity to use something, the friendlier is the user experience [53]. Perceived ease of use significantly and positively influences intention to use [39,48,50,54] and end users’ attitudes toward using a particular approach or technology [55]. The attitude is described as an individual’s impression of the positive or negative implications of embracing technology [56]. Attitude toward using positively and significantly influences intention to use [52,57,58] and is influenced by perceived usefulness [58,59] and ease of use [39].

According to the core construct of TAM, it is possible to add antecedents to improve the predictive power and understand the potential factors influencing behavioral intention. To this end, subjective norms and job relevance are important predictors added to the original TAM. Subjective norms refer to “a person’s view that the majority of influential individuals in his life believe he should or should not engage in the behavior in question” [60]. Studies have shown that subjective norms influence perceived usefulness [61] and intention to use [49,62]. The perception of whether a technology is suitable for their jobs can also have an impact on whether they intend to use it. Job relevance examines how users view using the system for work and is found to influence perceived usefulness and intention to use [44,45,63]. Whether the specified concept will mediate the link between constructs is another key factor to consider. The mediators act as a channel for latent concept effects to reach the dependent variables [64]. Perceived usefulness [65,66] and attitude [53,67,68] have been reported to act as mediators in TAM studies. The following hypotheses are formed in light of the information presented above (Figure 1).

**Figure 1.** Our proposed modified Technology Acceptance Model based on the original Technology Acceptance Model of Davis et al [69].

Hypothesis 1: Perceived usefulness will have a positive effect on the intention to use mobile-based partographs.
Hypothesis 2: Perceived usefulness will have a positive effect on obstetric health care providers’ attitudes toward mobile-based partographs.
Hypothesis 3: Perceived ease of use will have a positive effect on obstetric health care providers’ attitudes toward mobile-based partographs.
Hypothesis 4: Perceived ease of use will have a positive effect on the perceived usefulness of mobile-based partographs.
Hypothesis 5: Perceived ease of use will have a positive influence on the intention to use mobile-based partographs.
Hypothesis 6: Obstetric health care providers’ attitudes toward using mobile-based partographs will positively influence intention to use.
Hypothesis 7: Job relevance will have a positive effect on the intention to use mobile-based partographs.
Hypothesis 8: Job relevance will have a positive effect on the perceived usefulness of mobile-based partographs.
Hypothesis 9: Subjective norms will have a positive effect on the perceived usefulness of mobile-based partographs.

Hypothesis 10: Subjective norms will have a positive effect on the intention to use mobile-based partographs.
Hypothesis 11: Perceived usefulness mediates the relationship between job relevance and intention to use.
Hypothesis 12: Perceived usefulness mediates the relationship between subjective norms and intention to use.
Hypothesis 13: Perceived usefulness mediates the relationship between perceived ease of use and intention to use.
Hypothesis 14: Attitude mediates the relationship between perceived usefulness and intention to use.
Hypothesis 15: Attitude mediates the relationship between perceived ease of use and intention to use.

**Methods**

**Study Design and Setting**

An institution-based cross-sectional study was conducted from June 1 to July 1, 2022, at public referral hospitals in the Oromia region of Ethiopia. There are 30 municipal administrations and 23 zones in the Oromia region. Thirteen public referral hospitals, 33 general hospitals, 1383 health centers, and 6797 health posts are located in the Oromia region. In the 13 public referral hospitals, there were about 649 obstetric health care professionals employed. The midwives, nurses, integrated
emergency obstetrics and surgery professionals, general practitioners, obstetricians, and gynecologists are trained obstetric health care professionals.

Study Participants and Sample Size Determination
All obstetric health care providers who were working at Oromia region public referral hospitals and were available at the time of data collection were the source and study population of this study. The sample size for this study was calculated based on the rule of thumb of structural equation modeling. The most widely used rule of thumb is that for 1 free parameter, 10 observations are required [70,71]. The proposed model of this study had 55 free parameters. Therefore, n (number of free parameters × 10; 55 × 10) = 550, where n represents the sample size. Considering a 10% nonresponse rate, the final minimum sample size was 605.

Sampling Procedure
In the 13 public referral hospitals of the Oromia region, there were 649 obstetric health care providers. From the onset, this study required a large sample size to test the developed hypothesis using the maximum likelihood estimator of the structural equation model. However, the number of study participants involved in this study was not adequate for sampling. For this reason, a census was conducted.

Study Variables
The outcome variable of this study was the intention to use mobile-based partograph, whereas the mediator variables were attitude toward using and perceived usefulness. The independent variables of this study included technology acceptance–related exogenous latent variables (perceived ease of use, subjective norm, and job relevance), sociodemographic and other related factors of obstetric health care providers (age, sex, marital status, religion, profession, qualification, and years of experience), access to mobile devices, partograph learning, in-service training, and computer courses.

Operational Definition
Behavioral intention is the extent to which an individual has made intentional plans to engage in or refrain from engaging in a specific future conduct [37]. The intention to use, in this case, refers to the likelihood of obstetric health care providers whether they intended to use mobile-based partograph if they will be offered. The construct had 4 items, and each was measured with a 5-point Likert scale response. The median score was used as a cutoff point. The obstetric health care provider who scored median and above on intention to use construct was considered as intended to use a mobile-based partograph otherwise unintended.

Data Collection and Procedure
A self-administered structured English questionnaire was used to collect data. Regarding the latent variable, the questionnaire was adapted from different literatures [44,53,63,72-77]. The adapted questionnaire was modified to fit the context of this study. The structured questionnaire had 4 parts: the first part included the sociodemographic characteristics of the obstetric health care providers, the second part related to access to mobile devices, the third part related to partograph and computer courses, and the fourth part included technology acceptance–related parameters (perceived usefulness, perceived ease of use, intention to use, attitude, job relevance, and subjective norm). A total of 21 items were used in this study to test the proposed hypothesis. A Likert scale ranging from strongly disagree (1) to strongly agree (5) was used to rate the level of participant agreement toward the prepared close-ended questions.

Data Quality Assurance
A pretest was done on 5% of the sample size among obstetric health care providers who were working out of the study area. One day of training was given to 4 data collectors and 4 supervisors on the objective of the study, data collection procedures, data confidentiality, and respondents’ rights. The data collectors were BSc graduates: 3 midwives, 2 health officers, and 3 nurses. Supervision was continuous and made by the supervisors and principal investigator throughout the data collection. After data collection, completeness was checked.

Data Processing and Analysis
Before data analysis, the coded data were entered into EpiData (version 4.6) and finally exported into SPSS (version 25; IBM Corp) for descriptive analysis and AMOS (analysis of moment structure; version 23) for structural and measurement model assessment. Structural equation modeling is a multivariate statistical analysis technique that is used to analyze structural relationships. The data set was checked for missing values, and there were no missing data. The sociodemographic data were analyzed descriptively using SPSS, and the results were presented using a frequency table. Descriptive statistics was used to compute the proportion of intention to use mobile-based partographs, and the result was presented using the bar graph. The maximum likelihood estimation method was considered, and the assumption was checked. One assumption is the presence of multiple measurements for a construct. Perceived ease of use, perceived usefulness, and intention to use each have 4 items. However, attitude, subjective norms, and job relevance each have 3 items. Multicollinearity among the independent variables was assessed using the variance inflation factor. The result obtained (variance inflation factor ranged from 1.6 to 2.027) proved that there was no multicollinearity among independent variables.

Another assumption was univariate normality, which was assessed using kurtosis and skewness values, and the result shows there was univariate normality. The kurtosis value of less than 5, a critical ratio between –1.96 and +1.96, was used to declare the presence of multivariate normality. Unfortunately, the assumption of multivariate normality was not fulfilled. Therefore, bootstrapping technique was used to manage multivariate nonnormality.

Confirmatory factor analysis was used to perform measurement model assessment. Construct reliability was tested using Cronbach α and composite reliability. A cutoff point greater than .7 was used to declare the presence of internal consistency of the item that measured construct [78]. The recommended Cronbach α value should be .7 and above [79,80]. Furthermore, the composite reliability should be greater than .7 [36].
average variance extracted and factor loading was used to measure convergent validity. In the measurement model assessment, the average variance extracted value greater than 0.50 [45, 64] and factor loading of at least 0.6 [75] should be used to establish convergent validity. Discriminant validity assesses the distinctness of construct when measured by their respective items. The discriminant validity was determined using the square root of average variance extracted, and the value should be greater than the interconstruct correlations to declare whether the discriminant validity of the construct was achieved [64]. The degree to which one construct differs from every other construct in the instrument is indicated by its discriminant validity [81].

Model goodness of fit was checked both for measurement and structural model assessment. A model fit index of the ratio of chi-square to degrees of freedom \( \leq 3 \) [36], comparative fit index >0.90, goodness-of-fit index >0.90, adjusted goodness-of-fit index >0.85, normalized fit index >0.90, standardized root mean square residual <0.08, and root means square error of approximation <0.008 index value were used to measure and declare the model’s goodness of fit [39, 50]. We planned to perform model modification to improve the model fitness if the initial model did not fit by deleting the factor loading value with <0.5 covariate error terms [82]. In this regard, even if the measurement model fitness was achieved initially, the model modification was performed since the chi-square to the degree of freedom for the structural model was 3.128. Therefore, to increase the model fitness, we covariate the error term 15 and 16 on the intention to use the latent variable. Finally, the overall model fitted the data well.

After measurement model assessment, structural model fitness was checked and the model fit the data well. Then, the structural model assessment was performed. Based on AMOS output, the standardized path coefficient and the level of significance were used to test the developed hypothesis and determine the association between the latent variables of the study. The standardized regression weights showed the strength of association between latent variables [83], and \( P \) values less than 0.05 showed the level of significance considered. The square multiple correlations were used to report the proportion of variances in endogenous latent variables explained by exogenous variables. The bootstrap method was used to test the mediation effect.

**Ethics Approval**
Ethical clearance was obtained from the University of Gondar, and this study was approved by its ethics review board (Ref/IPH/2129/2014). A letter of support was obtained from the Department of Health Informatics, and written consent was taken from each study participant.

**Results**

**Sociodemographic Characteristics of the Participants**
A total of 649 participants were planned to be included in this study from all public referral hospitals in the Oromia region for the assessment of their intention to use mobile-based partographs and the predictors for their intention to use. Among them, 97% (630/649) gave their consent and completed the questionnaire. The results of this study show that almost more than half (344/630, 54.6%) of the study participants were males, while 45.4% (286/630) of the participants were females. The median age of the study participants was 32 (IQR 9) years. The majority of the participants were in the age group of 30-39 years. About 34.4% (217/630) of the respondents were Orthodox Christians in religion. Among the study participants, 371 (58.9%) respondents were married and 218 (36.4%) were single. Regarding their profession, more than half (351/630, 55.7%) were midwives, 96 (15.2%) were nurses, and 95 (15.1%) were general practitioners. Of the total study participants, 497 (78.9%) were bachelor’s degree holders and 78 (12.5%) were a specialist in their qualifications. Almost half (328/630, 52.1%) of the study participants had 4 years of working experience with the qualification they had. All the sociodemographic characteristics of the study participants are shown in Table 1.
Table 1. Sociodemographic characteristics of the obstetric health care providers who were working at public referral hospitals in the Oromia region of Ethiopia in 2022 (N=630).

<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>344 (54.6)</td>
</tr>
<tr>
<td>Female</td>
<td>286 (45.4)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>232 (36.8)</td>
</tr>
<tr>
<td>30-39</td>
<td>299 (47.5)</td>
</tr>
<tr>
<td>&gt;40</td>
<td>99 (15.7)</td>
</tr>
<tr>
<td>Religion</td>
<td></td>
</tr>
<tr>
<td>Orthodox Christian</td>
<td>217 (34.4)</td>
</tr>
<tr>
<td>Muslim</td>
<td>167 (26.5)</td>
</tr>
<tr>
<td>Protestant</td>
<td>208 (33)</td>
</tr>
<tr>
<td>Others^a</td>
<td>38 (6)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>371 (58.9)</td>
</tr>
<tr>
<td>Single</td>
<td>218 (34.6)</td>
</tr>
<tr>
<td>Divorced</td>
<td>25 (4)</td>
</tr>
<tr>
<td>Separated</td>
<td>10 (1.6)</td>
</tr>
<tr>
<td>Widowed</td>
<td>2 (0.3)</td>
</tr>
<tr>
<td>Others^b</td>
<td>4 (0.6)</td>
</tr>
<tr>
<td>Profession</td>
<td></td>
</tr>
<tr>
<td>Midwives</td>
<td>351 (55.7)</td>
</tr>
<tr>
<td>Nurses</td>
<td>96 (15.2)</td>
</tr>
<tr>
<td>General Practitioner</td>
<td>95 (15.1)</td>
</tr>
<tr>
<td>Obstetrician and Gynecologist</td>
<td>78 (12.4)</td>
</tr>
<tr>
<td>Others^c</td>
<td>10 (1.6)</td>
</tr>
<tr>
<td>Level of qualification</td>
<td></td>
</tr>
<tr>
<td>Bachelors</td>
<td>497 (78.9)</td>
</tr>
<tr>
<td>Masters</td>
<td>48 (7.6)</td>
</tr>
<tr>
<td>Specialists</td>
<td>79 (12.5)</td>
</tr>
<tr>
<td>Others^d</td>
<td>6 (1)</td>
</tr>
<tr>
<td>Years of working experience</td>
<td></td>
</tr>
<tr>
<td>≤4</td>
<td>328 (52)</td>
</tr>
<tr>
<td>5-9</td>
<td>195 (31)</td>
</tr>
<tr>
<td>≥10</td>
<td>107 (17)</td>
</tr>
</tbody>
</table>

*aWAAqeffanna, Adventist, and Catholic.
*bIn a relationship.
*cIntegrated Emergency Obstetric Surgery professionals.
*dDiploma.

Access to Mobile Devices and Partographs
In this study, of the 630 participants, 625 (99.2%) had access to mobile devices (Table 2). Approximately 93.1% (582/630) of the obstetric health care providers had access to a smartphone. About 77.8% (490/630) of the study participants had a partograph. However, only 326 (51.7%) study participants had taken in-service training for using paper-based partographs.
Regarding the computer course, 310 (80.7%) took basic computer training. Table 2 shows the frequency of access to mobile devices, partograph learning, partograph in-service training, and computer courses.

Table 2. Access to mobile devices, partograph training, and computer courses by obstetric health care providers who were working at the public referral hospitals in the Oromia region of Ethiopia in 2022 (N=630).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access to mobile devices</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>625 (99.2)</td>
</tr>
<tr>
<td>No</td>
<td>5 (0.8)</td>
</tr>
<tr>
<td>Mobile device type</td>
<td></td>
</tr>
<tr>
<td>Smartphone</td>
<td>582 (93.1)</td>
</tr>
<tr>
<td>Tablet</td>
<td>33 (5.3)</td>
</tr>
<tr>
<td>Others&lt;sup&gt;a&lt;/sup&gt;</td>
<td>10 (1.6)</td>
</tr>
<tr>
<td>Study partograph</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>490 (77.8)</td>
</tr>
<tr>
<td>No</td>
<td>140 (22.2)</td>
</tr>
<tr>
<td>In-service training on paper-based partograph</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>326 (51.7)</td>
</tr>
<tr>
<td>No</td>
<td>304 (48.3)</td>
</tr>
<tr>
<td>Computer courses</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>384 (61)</td>
</tr>
<tr>
<td>No</td>
<td>246 (39)</td>
</tr>
<tr>
<td>Computer course level</td>
<td></td>
</tr>
<tr>
<td>Basic course</td>
<td>310 (80.7)</td>
</tr>
<tr>
<td>Advanced training</td>
<td>74 (19.3)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Basic phones or feature phone.

Intention to Use Mobile-Based Partographs
Among the 630 obstetric health care providers, 414 had intention to use mobile-based partographs. Thus, about 65.7% (414/630; 95% CI 61.9%-69.4%) or two-thirds of the study participants scored median and above of intention to use. The median score of intention to use mobile-based partograph was 16 (IQR 1.5). The minimum and maximum scores for intention to use were 4 and 20, respectively. Figure S1 of Multimedia Appendix 1 shows the proportion of intended and unintended use of mobile-based electronic partographs among obstetric health care providers who were working at public referral hospitals in the Oromia region in 2022. In this study, Cronbach α and composite reliability values were greater than .9. All the factor loading values were found in the range of 0.843 to 0.946, and the average variance extracted value was found in the range of 0.761 to 0.840. Hence, construct reliability and convergent validity of the measurement model were achieved.

Discriminant Validity
The finding of this study indicates that the square root of the average variance extracted value was greater than the value of the interconstruct correlations. Therefore, the discriminant validity of the measurement model was achieved. Table S2 of Multimedia Appendix 1 demonstrates the discriminant validity of the model. The bold values in the table represent the square root of the average variance extracted.

Measurement Indices of the Goodness of Fit of Model
In this study, all the obtained values of the measurement model fit indices were in an acceptable range. Hence, the result of this study indicated that the modified proposed model fitted the data well. Table 3 shows the result of model fit indices [36,39,50].

https://ojphi.jmir.org/2024/1/e51601
Table 3. Results of the indices of goodness of fit of the measurement model assessment.

<table>
<thead>
<tr>
<th>Model fit indices, citation</th>
<th>Cutoff point</th>
<th>Result obtained$^a$</th>
<th>Fit decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square to degree of freedom [36]</td>
<td>≤3</td>
<td>2.245</td>
<td>Accepted</td>
</tr>
<tr>
<td>Goodness-of-fit index [50]</td>
<td>&gt;0.9</td>
<td>0.945</td>
<td>Accepted</td>
</tr>
<tr>
<td>Adjusted goodness-of-fit index [50]</td>
<td>&gt;0.85</td>
<td>0.927</td>
<td>Accepted</td>
</tr>
<tr>
<td>Normed fit index [50]</td>
<td>&gt;0.9</td>
<td>0.974</td>
<td>Accepted</td>
</tr>
<tr>
<td>Comparative fit index [39,50]</td>
<td>&gt;0.9</td>
<td>0.985</td>
<td>Accepted</td>
</tr>
<tr>
<td>Root mean square residuals [39,50]</td>
<td>&lt;0.08</td>
<td>0.022</td>
<td>Accepted</td>
</tr>
<tr>
<td>Root mean square error of approximation [50]</td>
<td>&lt;0.08</td>
<td>0.044</td>
<td>Accepted</td>
</tr>
</tbody>
</table>

$^a$The result is obtained from the measurement model assessment using AMOS (analysis of moment structure; version 23) software to check model fitness.

Structural Model Assessment

Among the 10 proposed hypotheses of the direct relationship, 8 were supported by the collected data. However, hypothesis 9 and hypothesis 10 failed to support the proposed hypothesis. Our study shows that perceived usefulness positively influences intention to use ($β=.184; P=.02$). In addition, perceived usefulness had a positive influence on attitude ($β=.512; P=.002$). According to our finding, perceived ease of use positively influenced attitude ($β=.382; P=.003$), perceived usefulness ($β=.503; P=.002$), and intention to use ($β=.369; P=.001$). Perceived ease of use had the strongest path coefficient in influencing intention to use mobile-based partographs. Job relevance in the structural model had a positive and significant influence ($β=.408; P=.002$) on perceived usefulness. In addition, job relevance ($β=.185; P=.008$) positively and significantly influenced the intention to use. The results of this study also indicated that attitude ($β=.309; P=.002$) positively and significantly influenced intention to use.

Figure 2. AMOS (analysis of moment structure; version 23) output of the standardized estimate. AT: attitude; ITU: intention to use; JR: job relevance; PEOU: perceived ease of use; PU: perceived usefulness; SN: subjective norms.

Subjective norms had insignificant influences on perceived usefulness ($β=.02; P=.61$) and intention to use ($β=−.066; P=.07$). Therefore, hypotheses 9 and 10 failed to support the developed hypothesis. Perceived ease of use, attitude, and job relevance had a 0.369, 0.309, and 0.185 path coefficient in association with intention to use, respectively. Results from the AMOS output of the proposed model showed that perceived usefulness was influenced by both job relevance and perceived ease of use (Figure 2). However, perceived ease of use ($β=.503$) had higher path coefficient than job relevance ($β=.408$) in influencing perceived usefulness. Perceived usefulness and perceived ease of use influenced the attitude toward using electronic partographs. Perceived usefulness ($β=.521$) had a higher path coefficient than perceived ease of use ($β=.382$) in influencing attitude toward using electronic partographs. Table 4 shows the results of the hypothesis testing of the direct path analysis of the proposed model.
Among the 15 developed hypotheses, 5 of them deal with the mediation effect of perceived usefulness and attitude between the exogenous latent variables and outcome variable in this study. Table S3 of Multimedia Appendix 1 shows the result of the mediation analysis using the bootstrapping method. The 95% biased confidence interval and P value level were used to test the presence of a mediation. Our results showed that perceived usefulness partially mediates between job relevance and intention to use. Hence, hypothesis 11 was supported. However, perceived usefulness did not mediate the relationship between subjective norms and intention to use. Therefore, hypothesis 12 was not supported. Furthermore, attitude partially mediated the relationship between perceived usefulness and intention to use as well as perceived ease of use and intention to use. Therefore, hypothesis 13 was supported. Consequently, hypothesis 14 and hypothesis 15 were supported. Table S3 of Multimedia Appendix 1 illustrates the result of the mediation analysis.

The squared multiple correlations indicate the predictive power of the model. In this study, the proposed model explained 85% of the variance in intention to use a mobile-based partograph. Perceived ease of use and job relevance aggregatedly explained 77% of the variance in perceived usefulness. However, perceived ease of use and perceived usefulness explained about 75% of the variance in attitude toward use. Table S4 of Multimedia Appendix 1 shows the result of the predictive power of the proposed model.

### Discussion

#### Principal Findings

This study examines the proportion of obstetric health care providers in the Oromia region of Ethiopia who had the intention to use mobile-based partographs and the predictors for their intention to use. The result of our study revealed that about two-thirds (414/630, 65.7%) of the obstetric health care providers had the intention to use mobile-based partographs. Even if no similar study was conducted on the intention to use a mobile-based partograph, ascertaining the end user’s level of acceptance to use technology before execution serves as a prerequisite to judging the accomplishment of the execution [39]. In this regard, a study in north Gondar reported that 44% of the obstetric health care providers were willing to use mobile-based partographs [84]. The findings of [84] assist our study in promoting the execution of mobile-based partographs in clinical settings for enhancing the care given during the management of labor.

The results of our study regarding TAM were in line with those of the original TAM [69]. In this study, perceived ease of use (β=.503; P=.002) significantly and positively influenced the perceived usefulness of mobile-based partographs by obstetric health care providers. This study’s finding is in line with a study in Omaha on telemedicine acceptance (β=.56) [18] and that in Ethiopia on eHealth acceptance (β=.385) [39]. Further, this study finding is supported by a study conducted in Tanzania in which skilled birth attendants found that electronic partograph was useful, easy to use, and improved the quality of care [24]. This indicates that obstetric health care providers who perceive mobile-based partograph is easy to use, easy to interact with, and easy to learn are more likely to perceive it as useful, and consequently, this will lead to a high intention to use it [39].

Another association in this study was that perceived usefulness (β=.184; P=.02) positively influenced intention to use. This positive and significant relationship of the construct is in agreement with a study in Turkey on personal digital assistant acceptance (β=.41) [48], a study in Uganda on mobile phone acceptance (β=.540; P=.02) positively influenced intention to use. This indicates that mobile-based partographs is easy to use, easy to interact with, and easy to learn are more likely to perceive it as useful, and consequently, this will lead to a high intention to use it [39].

Among the 15 developed hypotheses, 5 of them deal with the mediation effect of perceived usefulness and attitude between the exogenous latent variables and outcome variable in this study. Table S3 of Multimedia Appendix 1 shows the result of the mediation analysis using the bootstrapping method. The 95% biased confidence interval and P value level were used to test the presence of a mediation. Our results showed that perceived usefulness partially mediates between job relevance and intention to use. Hence, hypothesis 11 was supported. However, perceived usefulness did not mediate the relationship between subjective norms and intention to use. Therefore, hypothesis 12 was not supported. Furthermore, attitude partially mediated the relationship between perceived usefulness and intention to use as well as perceived ease of use and intention to use. Therefore, hypothesis 13 was supported. Consequently, hypothesis 14 and hypothesis 15 were supported. Table S3 of Multimedia Appendix 1 illustrates the result of the mediation analysis.

The squared multiple correlations indicate the predictive power of the model. In this study, the proposed model explained 85% of the variance in intention to use a mobile-based partograph. Perceived ease of use and job relevance aggregatedly explained 77% of the variance in perceived usefulness. However, perceived ease of use and perceived usefulness explained about 75% of the variance in attitude toward use. Table S4 of Multimedia Appendix 1 shows the result of the predictive power of the proposed model.

### Table 4. The results of the hypothesis testing of the proposed model.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Causal path</th>
<th>β (path coefficient)</th>
<th>Critical ratio</th>
<th>P value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ITU&lt;sup&gt;a&lt;/sup&gt;→PU&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.184</td>
<td>3.294</td>
<td>.02</td>
<td>Supported</td>
</tr>
<tr>
<td>2</td>
<td>AT&lt;sup&gt;c&lt;/sup&gt;→PU</td>
<td>.521</td>
<td>9.918</td>
<td>.002</td>
<td>Supported</td>
</tr>
<tr>
<td>3</td>
<td>AT&lt;sup&gt;d&lt;/sup&gt;→PEOU</td>
<td>.382</td>
<td>7.279</td>
<td>.003</td>
<td>Supported</td>
</tr>
<tr>
<td>4</td>
<td>PU&lt;sup&gt;e&lt;/sup&gt;→PEOU</td>
<td>.503</td>
<td>11.097</td>
<td>.002</td>
<td>Supported</td>
</tr>
<tr>
<td>5</td>
<td>ITU&lt;sup&gt;f&lt;/sup&gt;→PEOU</td>
<td>.369</td>
<td>7.160</td>
<td>.001</td>
<td>Supported</td>
</tr>
<tr>
<td>6</td>
<td>ITU&lt;sup&gt;g&lt;/sup&gt;→AT</td>
<td>.309</td>
<td>6.351</td>
<td>.002</td>
<td>Supported</td>
</tr>
<tr>
<td>7</td>
<td>ITU&lt;sup&gt;h&lt;/sup&gt;→JR&lt;sup&gt;i&lt;/sup&gt;</td>
<td>.185</td>
<td>3.969</td>
<td>.008</td>
<td>Supported</td>
</tr>
<tr>
<td>8</td>
<td>PU&lt;sup&gt;j&lt;/sup&gt;→JR</td>
<td>.408</td>
<td>8.476</td>
<td>.002</td>
<td>Supported</td>
</tr>
<tr>
<td>9</td>
<td>PU&lt;sup&gt;k&lt;/sup&gt;→SN&lt;sup&gt;l&lt;/sup&gt;</td>
<td>.020</td>
<td>0.540</td>
<td>.61</td>
<td>Not supported</td>
</tr>
<tr>
<td>10</td>
<td>ITU&lt;sup&gt;m&lt;/sup&gt;→SN</td>
<td>-0.066</td>
<td>-2.060</td>
<td>.07</td>
<td>Not supported</td>
</tr>
</tbody>
</table>

<sup>a</sup>ITU: intention to use.  
<sup>b</sup>Pu: perceived usefulness.  
<sup>c</sup>AT: attitude.  
<sup>d</sup>PEOU: perceived ease of use.  
<sup>e</sup>JR: job relevance.  
<sup>f</sup>SN: subjective norms.
adoption in maternal health care ($\beta=1.86$) [54], and a study in Ethiopia on eHealth adoption ($\beta=3.87$) [39]. Thus, obstetric health care providers’ perception of mobile-based partograph usefulness is a valuable predictor of behavioral intention to use. It is important to find out how users measure the usefulness of the technology because the more the obstetric health care providers perceive that a mobile-based partograph improves productivity, performance, and effectiveness, and decreases the duration of recording, the more likely they will intend to use it. Our study shows that perceived ease of use positively ($\beta=3.82$) and significantly ($P=0.003$) influenced attitude toward using electronic partographs. This direct effect of perceived ease of use on attitude is in line with a study in Ethiopia ($\beta=3.47$) [39]. These findings indicate that if users perceive that the use of electronic partograph is easy, they will develop a positive attitude toward using it, consequently impacting their behavioral intention to use it.

We found that perceived ease of use of mobile-based partographs significantly affected intention to use ($\beta=3.69$; $P=0.001$). This showed that the likelihood of intention to use the mobile-based partograph will increase with an increase in the impression of ease of use. This outcome is consistent with a study on the adoption of health information systems that was conducted in Greece ($\beta=29$) [50] and in Ethiopia ($\beta=3.39$) [39], demonstrating that requiring less effort will boost the system’s ability to influence people’s intentions to use eHealth systems. Perceived ease of use in this study had the highest path coefficient and a significant impact on usage intention. The more obstetric health care professionals are intentional to use the mobile-based partograph, the less effort it is thought to take to operate it on both a mental and physical level. To meet user expectations, electronic partograph developers should concentrate on the device’s user-friendliness. This might increase the uptake and ongoing use of mobile-based partographs. Additionally, we found that attitudes about adopting mobile-based partographs are influenced by perceived usefulness ($\beta=521$; $P=0.002$). This finding is consistent with studies on telemedicine acceptability done in China ($\beta=43$) [85] and Ethiopia ($\beta=26$) [39]. The obstetric health care providers’ attitude toward use was impacted by how much they believed this technology improves performance and productivity in their job. Thus, intention to use is directly influenced by attitude toward use ($\beta=309$; $P=0.002$). Our study’s conclusion is consistent with that in a study on the intention of health professionals in Ethiopia to use eHealth ($\beta=526$) and indicates that actions that improve perspectives, such as ongoing training and support and information sharing on eHealth innovations, should be prioritized heavily [39]. The more positive perception the obstetric health care providers developed and had, the higher they intended to use eHealth.

In this study, job relevance significantly and positively ($\beta=0.408$; $P=0.002$) influenced perceived usefulness. This study path relationship is in line with a study that focused on the adoption of technology using modified TAM [44, 86]. There should be adequate information provision strategies for end users about the applicability and usefulness of mobile-based partographs in labor management. A study on personal digital assistant acceptance by health care professionals supports this evidence in a way that information provision about the technology applicability by health care institutions promotes technology acceptance [80]. Additionally, job relevance significantly influenced intention to use ($\beta=1.85$; $P=0.008$). This finding is in line with a study conducted on health information technology acceptance [44]. This means that the more probable obstetric health care professionals expected to use a mobile-based partograph, the more they believed it was appropriate, relevant, and vital to their work. Because of this, it is crucial to let obstetric health care professionals know about the use of mobile-based partographs in labor management.

According to this study, subjective norms had an insignificant impact on perceived usefulness ($\beta=0.20$; $P=0.61$). This insignificant influence is in line with the findings of a study on hospital information systems ($\beta=–0.18$; $P>0.05$) [87]. This might be because obstetric health care professionals are more likely to establish their independent judgments and may therefore pay less heed to what other people think. Another factor is that regardless of what is important, others may think all obstetric health care providers will be forced to use mobile-based partographs as long as the government mandates their usage in health care facilities for labor management. Subjective norms also insignificantly influence intention to use ($\beta=–0.066$; $P=0.07$). The result of this study is inconsistent with those of other studies [34, 62, 75]. Because of the time-consuming and detailed recording in paper-based partographs, obstetric health care providers could record incorrect and incomplete data and they could perform retrograde documentation to avoid accountability. However, in case of the digital partographs, retrograde documentation is not allowed and accountability is assured. To affect the belief that most significant others consider he or she should utilize the mobile-based partograph, the health care system should strengthen the concept of teamwork.

In contrast to a study on the adoption of eHealth in the Amhara region, where attitude was the strongest predictor of intention, in this study, perceived ease of use was the strongest predictor of intention to use [39]. To maximize the likelihood of initial and ongoing usage, mobile-based partographs should be provided with an easy function. The association between job relevance and intention to use as well as the relationship between perceived ease of use and intention to use is partially mediated by perceived usefulness in the study’s proposed model, which is consistent with [88]. However, perceived utility is unable to buffer the link between subjective norms and usage intention.

Perceived ease of use and intention to use were partially mediated by attitude and consistent with the study conducted using modified TAM [83]. Attitude toward using mediates the relationship between perceived usefulness and intention to use. To maximize the benefits of the additional capabilities, makers of eHealth platforms should actively work to change how physicians feel about using them [68]. The proposed modified TAM explained about 85% variance in the obstetric health care providers’ intention to use mobile-based partographs. This shows that the model’s overall predictive ability was high, and the variance in this study was almost equivalent to a study [50] utilizing modified TAM in a health care scenario.
Limitations of This Study

First, it is difficult to discuss whether the proportion of intention to use mobile-based partographs among obstetric health care providers is high or low due to the lack of similar previous studies. This study was a one-time study. Second, it is difficult to establish a cause-effect relationship in this study due to the cross-sectional nature of this study. Third, as the respondents were only from the government and tertiary care levels, caution must be exercised when applying these findings to all obstetric health care providers in the area.

Conclusion

In our study, two-thirds of the obstetric health care providers had the intention to use mobile-based partographs. Perceived usefulness, perceived ease of use, job relevance, and attitude positively and significantly influenced their intention to use mobile-based electronic partographs. Among these, perceived ease of use was the strongest potential predictor of intention to use. The connection between exogenous latent variables and intention to use was partially mediated by perceived usefulness and attitude toward usage, except for subjective norms. The modified TAM is an effective model for forecasting the intention of an obstetric health care professional to use a mobile-based partograph.

Acknowledgments

We would like to thank the University of Gondar for granting us the ethics approval to conduct this thesis. We are very grateful to Oromia referral hospital management, data collectors, supervisors, and participants of this study. KNT is immensely grateful for the support and care of his family and colleagues during his thesis work. All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Data Availability

The data set used for the analysis in this study can be obtained on request from the corresponding author.

Authors' Contributions

All the authors of this study contributed to problem identification, proposal development, data collection and analysis, and thesis write-up to final approval for publication. They accept responsibility for this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary data.

[DOCX File, 226 KB - ojphi_v16i1e51601_app1.docx ]

References


57. Abushaar L, Ismail A. Acceptance of teleconsultation among doctors and the determinant factors in teaching hospital in Malaysia. 2017 Apr 14 Presented at: 2nd Public Health International Conference (PHICO); December; Malaysia. [doi: 10.2991/phico-17.2018.9]


Abbreviations

AMOS: analysis of moment structure
TAM: Technology Acceptance Model
Applying Machine Learning Techniques to Implementation Science

Nathalie Huguet1,2*, PhD; Jinying Chen3,4,5*, PhD; Ravi B Parikh6*, MPP, MD; Miguel Marino1,2*, PhD; Susan A Flocke1,2*, PhD; Sonja Likumahuwa-Ackman1,2*, MID, MPH; Justin Bekelman6,7*, MD; Jennifer E DeVoe1,2*, MD, DPhil

1Department of Family Medicine, Oregon Health & Science University, Portland, OR, United States
2BRIDGE-C2 Implementation Science Center for Cancer Control, Oregon Health & Science University, Portland, OR, United States
3Section of Preventive Medicine and Epidemiology, Department of Medicine, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, United States
4Data Science Core, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, United States
5iDAPT Implementation Science Center for Cancer Control, Wake Forest School of Medicine, Winston-Salem, NC, United States
6Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
7Penn Center for Cancer Care Innovation, Abramson Cancer Center, Penn Medicine, Philadelphia, PA, United States

*all authors contributed equally

Corresponding Author:
Nathalie Huguet, PhD
Department of Family Medicine
Oregon Health & Science University
3181 SW Sam Jackson Park Road
Portland, OR, 97239
United States
Phone: 1 503 494 4404
Email: huguetn@ohsu.edu

Abstract

Machine learning (ML) approaches could expand the usefulness and application of implementation science methods in clinical medicine and public health settings. The aim of this viewpoint is to introduce a roadmap for applying ML techniques to address implementation science questions, such as predicting what will work best, for whom, under what circumstances, and with what predicted level of support, and what and when adaptation or deimplementation are needed. We describe how ML approaches could be used and discuss challenges that implementation scientists and methodologists will need to consider when using ML throughout the stages of implementation.

(Online J Public Health Inform 2024;16:e50201) doi:10.2196/50201

KEYWORDS
implementation science; machine learning; implementation strategies; techniques; implementation; prediction; adaptation; acceptance; challenges; scientist

Introduction

Implementation science is a research field developing and testing methods and strategies that can improve the uptake of evidence-based interventions (EBIs) and practices into routine use in targeted settings [1]. It has important applications in both clinical and public health settings, such as health care facilities, public health departments, schools, and workplaces [2-4]. For example, the RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance) framework, which was proposed by implementation scientists to guide the planning and evaluation of programs, has been used for health care– and community-based programs promoting chronic disease prevention and management, healthy aging, mental health, and health behavior change [5]. In addition, implementation science methods have been applied in clinical settings (eg, clinic-initiated cancer screening, tobacco cessation, and mental health programs) to scale up effective interventions to improve population health [4].

Implementation strategies are the methods, actions, and activities that aim to enhance the adoption, implementation, and sustainability of EBIs in clinical and public health practice. Implementation strategies can target multiple levels (eg, communities, hospitals, health care clinics, public health
departments, clinical and public health practitioners, and individual patients and community members) and may involve multiple components (eg, information technology tools, workflow changes, and policies mandating services) and activities (eg, training and incentives) [6,7]. Numerous factors, such as target populations and targeted behavior change, varied uptake of strategies across settings, the actors that deliver the implementation strategies, and the timing of the EBI implementation, can influence the implementation processes and outcomes [6-8]. Further, there is often a need to tailor or adapt implementation strategies and the associated activities to the local, dynamic context to increase implementation success. Given the multifactorial drivers and their complex relationships, implementation science could benefit from advanced data analytics frameworks and methods for artificial intelligence and machine learning (ML).

As a subfield of artificial intelligence, ML [9,10] develops automated methods and algorithms that learn from data. With this learning, it can then perform tasks such as prediction and pattern discovery. To date, ML applications in health care settings have been focused on supplementing clinical work, predicting health-related outcomes (eg, disease severity and prognosis) [11-14], and supporting clinical decisions (eg, tailoring medications and other treatments) [15-17]. Applications of ML in public health include population health surveillance and outbreak mitigation, evaluating the effectiveness of public health strategies and campaign and disaster and emergency management [18-22]. Existing literature on the application of ML in the field of implementation science is sparse [23]. However, ML has great potential to be applied in areas such as tailoring strategies and support activities, supporting decision-making on the selection of actors or settings, and predicting and understanding the impact of implementation strategies on the adoption of EBIs across different settings and target populations. The aim of this viewpoint is to introduce a roadmap for applying ML techniques to address implementation science questions, describe a few limited real-world applications of ML related to implementation science, and discuss challenges that implementation scientists and methodologists may face along the way when using ML as a strategy to monitor EBI adoption or to inform the need for interventions.

### A Roadmap for Applying ML in Implementation Science

ML approaches can be applied across the continuum of EBI implementation. Here, we use the strategic implementation framework (SIF) [24] as a roadmap to illustrate the potential application of ML at different stages of implementation, as summarized in Table 1. The SIF depicts 3 stages of implementation (ie, setting the stage; active implementation; and monitor, support, and sustain) and the distinct types of strategies needed for practice change in each stage to ensure that improvements are supported and sustained.

<table>
<thead>
<tr>
<th>Strategic implementation framework stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategic implementation framework stages</td>
</tr>
<tr>
<td>Setting the stage</td>
</tr>
<tr>
<td>Active implementation</td>
</tr>
<tr>
<td>Monitor, support, and sustain</td>
</tr>
<tr>
<td>Implementation goals and activities</td>
</tr>
<tr>
<td>• Understand the local context to select implementation strategies and prepare for implementation</td>
</tr>
<tr>
<td>• Implement strategies and support activities to improve EBI</td>
</tr>
<tr>
<td>• Monitor the sustained adoption of strategy and EBI improvements</td>
</tr>
<tr>
<td>Implementation challenges</td>
</tr>
<tr>
<td>• Limited data available or used</td>
</tr>
<tr>
<td>• The context is not static</td>
</tr>
<tr>
<td>• End users with differing priorities</td>
</tr>
<tr>
<td>• The context is not static (new guidelines, policy, and care delivery)</td>
</tr>
<tr>
<td>• Need to adapt to setting and target-ed population</td>
</tr>
<tr>
<td>Opportunities for ML application</td>
</tr>
<tr>
<td>• Predict who will adopt an EBI</td>
</tr>
<tr>
<td>• Determine the level of support needed</td>
</tr>
<tr>
<td>• Identify the need for change</td>
</tr>
<tr>
<td>• ML as a strategy</td>
</tr>
<tr>
<td>• Monitor progress</td>
</tr>
<tr>
<td>• Inform need for deimplementation</td>
</tr>
<tr>
<td>• Inform need for support</td>
</tr>
<tr>
<td>Considerations when using ML across the 3 stages</td>
</tr>
<tr>
<td>• Setting characteristics</td>
</tr>
<tr>
<td>• Time period</td>
</tr>
<tr>
<td>• Data completeness</td>
</tr>
<tr>
<td>• Multilevel strategy</td>
</tr>
<tr>
<td>• N/A</td>
</tr>
<tr>
<td>• Risk prediction bias</td>
</tr>
<tr>
<td>• Recalibration of ML</td>
</tr>
<tr>
<td>• Adaptation of ML</td>
</tr>
<tr>
<td>• Deimplementation of ML</td>
</tr>
</tbody>
</table>

aML: machine learning.

bEBI: evidence-based intervention.

N/A: not applicable.

### Setting the Stage

Setting the stage refers to preimplementation activities such as assessing readiness to change, identifying barriers and facilitators to implementing EBIs, selecting or developing strategies to support implementation, and identifying and acquiring resources. Implementation scientists often find that an effective strategy in one setting may not work well in other settings and that some may need more or different types of support (eg, hours of training, intensity of coaching support, or remote vs in-person training). As such, one of the biggest implementation science challenges is to identify what works,
Typical approaches for selecting and tailoring implementation strategies to fit the local context (eg, process mapping, intervention mapping, and coincidence analysis) address this challenge at the organization or population levels [25]. Often, data to inform the selection of an implementation strategy are limited to surveys, qualitative interviews, and organization-level data. However, clinical or public health data (eg, electronic health records [EHRs], administrative data, claims data, patient or disease registries, immunization registries, and health surveys), data linkages (eg, EHR data linked across practice sites, water quality, and air quality), and data related to implementation processes (eg, responses of patients, community members, and practitioners to a specific implementation science strategy from prior studies) are increasingly available. Implementation scientists could use ML to analyze large-scale, individual-level data to identify or predict who (individuals or subpopulations) is most likely (or least likely) to engage or respond to the intervention [26,27]. Specifically, the application of ML in the preimplementation stage could assist with the selection of the settings or actors, refinement of implementation strategies, and decisions about support activities. ML techniques could predict which sites, practitioners, or target populations will most likely respond well to certain implementation strategies (such as a training session or a health information technology tool), are most likely to need extra support, or might respond better to different strategies. These analyses could be based on prior engagement with strategies that led to increased adoption of EBIs or known characteristics of community (eg, census and environmental health), health systems (eg, geographic location), providers (eg, years of practice), patients (eg, race and ethnicity), and other targeted users.

There are currently no studies using ML approaches to tailor implementation strategies or support needs in the preimplementation stage. A few studies have used unsupervised statistical learning methods, such as latent class analysis and latent profile analysis [28], to identify subgroups of health care providers [27] and patients [26] responding differently to implementation strategies that promote provider-patient communication on critical illness or patients’ physical activities for weight reduction. For example, one study identified 3 groups (or phenotypes) of oncologists based on demographics, practice patterns, and patient panel information [27]. These phenotypes showed different responses to an EHR-based intervention (EHR nudges) aimed at improving advance care planning (ACP) discussion. Oncologists with the lowest volume of patients and a higher rate of baseline ACP discussion showed the greatest improvement compared to those with higher volume or lowest baseline ACP and intermediate volume or baseline ACP. One study used a supervised learning model to identify areas where the implementation of HIV prevention programs should be prioritized. Using state surveillance data on substance use, sexually transmitted diseases, and community characteristics (eg, percent living in poverty), ML modeling identified high-priority areas, of which 79% did not have implemented syringe services programs [29]. Similar modeling approaches could be used to better identify who will adopt what implementation strategies with what supports and tailor resource allocation before an implementation program is launched to improve the adoption and sustainability of EBIs.

Further, ML applications during the setting the stage could also facilitate monitoring when interventions are needed. For instance, using continuously collected clinical or public health data and ML-based phenotyping methods [27], it is possible to prioritize target populations who need the EBIs most at different time points or stages of the implementation of an intervention. Modeling could also trigger notifications to local clinics and public health departments about changes in quality metrics that require improvement, the resources needed to make an improvement (eg, additional staff), or changes in an environmental context (eg, climate change) [30] that could impact disease incidences and health care needs.

**Active Implementation**

During the active implementation stage, strategies and support activities are implemented to promote the adoption of an EBI (eg, disease surveillance, prescribing shingles vaccination, and lung cancer screening). During this stage, ML techniques could be incorporated as an implementation strategy. ML-based algorithms relating to the active implementation stage are currently being used to support making accurate diagnoses, disease risk estimation and surveillance, public health campaigns, and clinical decision-making. One example is the use of an ML model to identify foodborne illness in real time (FINDER). This model was developed, implemented, and tested in 2 US cities. FINDER would provide a daily list of restaurants identified as unsafe (likely to have health code violation). Health departments would then conduct an inspection in the restaurants identified by FINDER. The model identified accurately more unsafe restaurants than the previous system or reported complaints [31]. Examples in palliative care include a deep learning model that incorporates patients’ EHR data to predict mortality (those patients most likely to die within 3-12 months). The model-generated estimates were used to inform providers’ care recommendations and decisions about referring patients to palliative care [32,33]. In the context of cancer screening, ML models based on reinforcement learning or ensemble learning are being developed to more accurately identify patients with high risk of cancer [34,35]. These models could be used for cancer screening to balance the benefits of early detection and the costs of overscreening.

Further, in clinical care, clinical decision support (CDS) tools [36,37], including EHR alerts, are common implementation strategies used to promote guideline-concordant practice. ML can be used to develop “smarter” CDS tools to reduce alert fatigue. For example, an ML model was developed to predict whether a provider would respond to shingles vaccination alerts based on the provider’s characteristics (eg, demographics and clinical roles), patient’s demographics, and history of the provider’s interaction with the alerts [38]. The ML model was shown to reduce over 45% of shingles vaccination alerts without reducing weekly shingles vaccination orders [38].
Monitor, Support, and Sustain
This stage focuses on activities that ensure the sustainability of an intervention. During the monitor, support, and sustain stage, ML can inform changes needed to ensure the adoption and sustainability of practice changes. ML-based methods can leverage vast amounts of data to inform more flexible and adaptive implementation strategies. ML can also facilitate the evaluation and adaptation of strategies and inform where deimplementation is needed. For instance, ML could be used to identify when public health campaigns have reached saturation, need to be refocused, or are missing the target population. For example, during the COVID-19 pandemic, studies use ML models to identify people at greatest risk for COVID-19 death and who should be prioritized for vaccination. Different studies using different populations showed variations in who should be prioritized in informing local public health efforts [39-42]. For example, in clinical practice, implementation scientists leveraged both EHR audit logs and innovative ML-based approaches to monitor the impact of implementing a tobacco control CDS tool in the EHR system [43-45].

According to the Health Information Portability and Accountability Act (HIPAA) [46] and the 2014 release of the Meaningful Use regulations [47], all the EHRs in the United States are required to implement audit logs to unobtrusively track users’ EHR use. In a recent study, a latent-variable statistical ML model was developed to infer EHR-use activities from EHR audit log data [44]. Specifically, the ML model identified topics from EHR log data, where each topic was represented by a probability distribution of microlevel EHR actions such as loading a flow sheet, viewing a problem list, and using a favorite phrase predefined in EHR. Domain experts (3 physicians and 1 EHR specialist) reviewed these topics (eg, the top-ranked microlevel EHR actions belonging to each topic and example EHR sessions representative for each topic) and assigned an EHR-use activity (eg, visit documentation with record review and address CDS alerts) to each topic. This domain expert–informed model was then applied to EHR logs for 3703 encounters (before CDS implementation: n=2633 and after CDS implementation: n=1070) in 4 cancer clinics to monitor changes in providers’ EHR-use between 2019 and 2020 [45]. This study found that clinicians spent more time addressing CDS (more than 32-35 seconds) during a patient visit after CDS implementation (vs before CDS implementation), with compensatory unintended reductions in time spent reviewing patient vital data (less than 61 seconds) and modifying EHR (less than 7-24 seconds) [45]. These findings pointed to potential adaptations of the CDS to improve efficacy and reduce burden [43].

Data-driven findings can inform qualitative studies that aim to understand the causes of the unintended consequences and further inform the decision on refining or deimplementing certain features of the CDS tool.

In summary, despite very few real-world applications of ML in implementation science, there are many options and opportunities to use ML at different stages of implementation; however, some factors are important to take into consideration.

What Are the Factors to Consider in Using ML for Implementation Science?
As illustrated earlier, ML applications can potentially benefit implementation science across each of the SI’F stages. However, many factors can impact the use or validity of these ML-based applications in real-world settings, including achieving equitable outcomes across multiple settings or subpopulations [48].

There are various techniques used in ML [49]. Supervised learning methods can be used to build predictive models (eg, prediction of patients’ risks in illness or poor prognosis and responses of community members, patients, or providers to EBIs and implementation science strategies). Unsupervised learning methods can be used to mine data to identify patterns (eg, identify subgroups of population, patients, and health systems who have different responses to EBIs and implementation strategies). A common practice to develop and validate supervised ML models includes two stages: (1) using a data set to develop and validate (ie, internal validation) the model and (2) using a separate data set (obtained from other similar settings or from a withheld sample) to validate (ie, external validation) the developed model [50,51]. In the first stage, the model can be trained or validated through cross-validation or using a random split of the data set (eg, training or development or validation sets). The model’s parameters and hyperparameters are tuned or set using the training and development sets. In the second stage, the model’s performance is further assessed on the external validation set. Different from supervised learning, there is no ground truth (eg, labels for clusters or subgroups identified by unsupervised learning) to validate results from unsupervised learning in a real-world setting. Consequently, the evaluation process for unsupervised learning is less standard than supervised learning, and the choice of evaluation measures often depends on the unsupervised learning algorithms that are used [52,53].

In general, the quality of clustering results can be measured in 2 aspects when no external references (ie, ground truth) are available: coherence (ie, the similarity of objects falling into the same cluster) and separation (ie, the separation between clusters). Manual chart review is also useful or even necessary for qualitatively validating the clustering results in clinical settings [54]. Both supervised and unsupervised models developed on a specific sample or data set may not be readily applicable to other samples or data sets—the issue with generalizing ML models to different settings [55,56]. This issue has important implications on the use of ML in implementation science and requires paying special attention to model design, development, and validation.

The first factor to consider is that implementation strategies can be implemented at multiple levels (eg, state, county, community, population, health systems, clinicians, and patients), which would determine at which level the ML models would be based. Models developed and validated using data from one level (eg, clinic or community) need further validation and adaptation before being used for predicting outcomes at another level (eg, patient) or an intervention implemented at multiple levels [57]. For example, within the setting the stage phase, a model could
be developed using clinician and clinic characteristics (eg, specialty, provider type, and clinic geographic location) to predict which clinicians or clinics will be most likely to adopt a CDS tool. This model, however, is unlikely to be sufficient or valid in predicting the adoption of a multilevel intervention that targets both clinicians and patients (using provider nudges via EHR and patient nudges via SMS text messages). Similarly, public health programs (eg, a tobacco control or vaccination program) often use strategies targeting various levels within a public health jurisdiction (eg, individual, city, county, and state). An ML model predicting the adoption or success of such programs needs to take into account multilevel factors.

Second, the setting (eg, type of clinic and social culture of a specific community), its geographic location, and the time period used in validating the ML model are important factors to consider. These contextual factors are important in implementation science as they impact which strategy or combination of strategies are selected to scale up or modify to ensure the adoption and sustainability of EBI. Models that predict the adoption or sustainability of an implementation strategy developed in primary care clinics are unlikely to have an adequate prediction in specialty clinics in the setting the stage phase. Similarly, an ML-based strategy to improve an EBI in a rural community setting will likely need adaptation to be valid in an urban community setting. Additionally, the time period in which the model was developed needs to be taken into account. For instance, ML-based CDS developed prior to the COVID-19 pandemic may be obsolete or invalid after the pandemic in view of the widespread adoption of telehealth.

Third, when using ML models as an implementation strategy for risk prediction, they should be designed to predict the actual targeted outcome rather than the outcome that is easiest to obtain. For example, consider a risk prediction model being used to direct palliative care interventions. It is easier to train an ML-based tool to predict mortality, as a surrogate for palliative care needs, because mortality is less susceptible to measurement error and is available in palliative care medical records [58,59]. However, training an algorithm on mortality may not identify the individuals with high symptomatic or psychosocial needs who would benefit from palliative care the most. Targeting the risk prediction to the outcome that is most likely to matter for the EBI being implemented is imperative.

Finally, it is critically important to develop and validate models with equity in mind. Many of the algorithms developed in medicine are based on trials with nonrepresentative samples [60]. A recent publication examining various race-biased algorithms used for medical risk predictions demonstrated the potentially harmful consequences of biased algorithms [61]. Within implementation science, as noted earlier, strategies may not work for all. ML models validated in a specific population (eg, pediatric patients) within a specific setting (eg, hospital) could be misused and inequitable if used in a different population (eg, Latino pediatric patients receiving care in a community health center). The learning here is that ML-based implementation strategies need to be tested, validated, and adapted to fit the context of the targeted population to ensure health equity.

**What Are the Challenges?**

**Overview**

Despite the large amount of clinical data and data from pragmatic implementation trials, there are many challenges associated with data access and data quality. Further, the tools and resources needed to extract and pre-process these data for developing ML may not be easily accessible. For example, extracting and harmonizing patient-level data from the EHRs from multiple health systems to develop a pre-implementation ML model could be particularly difficult and time-consuming if these health systems have different EHR vendors. Furthermore, the application of ML in implementation science may result in unintended consequences, and issues related to the sustainability and scalability of the model need to be addressed.

**Data: Quality, Availability, and Type**

Public health data and information systems vary with regard to data quality, completeness, collection methods by systems, sampling bias, and underreporting [62-64]. In addition, the collection and generation of public health data are often time-consuming, resulting in delays in data reporting. Similarly, clinical-related data, such as EHR or health insurance claims data, are not designed for research and as such may not be collected and recorded in a systematic standard way. For example, comprehensiveness, completeness, and availability of patient demographic information (eg, race or ethnicity), health insurance data, and clinician data vary greatly by health systems and EHR vendors [65-71]. Additionally, some information that can be critical in the accuracy of ML prediction may reside in unstructured data (eg, a scanned PDF and free text of an encounter note) and, therefore, would require additional preprocessing steps, such as natural language processing [72]. Missing clinical-related data are unlikely to be random [70]. Specifically, EHR data come from a combination of clinician notes, test orders and results, documentation of diagnoses, and patient-reported information. The accuracy and completeness of these data are dependent on the source of the information. For example, the history of a cancer diagnosis can be derived from clinician diagnosis, clinical exchange systems, and patient self-reported history. A study linked EHR data with cancer registry to assess the accuracy of cancer diagnosis in the EHR [66]. Authors found that approximately 45% of cases recorded in the registry did not have a cancer history in their EHR. This information may have been in unstructured data such as in encounter notes. Data used for training an ML model may underrepresent certain patient subgroups [71]. For example, the use of insurance claims data excludes patients without health insurance, and these patients are often socioeconomically disadvantaged individuals. Variations in data documentation and completeness impact not only predictor variables used in the ML models but also the outcome variables. For example, predictive models of emergency department admissions using claims data would miss patients who are uninsured and are more likely to rely on the emergency department for care [73]. Moreover, ML models designed to develop an intervention targeting health system, school system, or community-based...
organization change may require data on staffing, supplies, or organizational capacity, which could be challenging to obtain.

**Potential for Unintended Consequences**

ML models, whether designed for predicting disease risk or for supporting clinical care management and decision-making, are susceptible to bias. Bias can be introduced at multiple points in the development and application process of ML [61,74,75]. As noted earlier, data sources and data representativeness (eg, the population, inclusion or exclusion of diseases, comorbidities, and health risk factors) can greatly influence the ML model and consequently the actions based on the ML model. Further, because ML models can generate data for other ML models, bias can be amplified and can lead to unintended consequences [76]. Char et al [77] proposed a framework for examining ethical considerations of ML models in health care settings, which poses questions about the values and ethics at multiple steps of the model development and implementation. This framework can guide decision-making to minimize bias and can promote accountability and transparency in model development.

**Sustainability and Scalability of the Model**

Public health interventions and campaigns are moving targets. For instance, climate change is leading public health departments to adapt or develop new initiatives for disaster preparedness efforts, disease surveillance, and carbon footprint reduction [78-80]. For instance, there is growing evidence of the mental health toll of climate-related events [81], yet strategies to monitor and intervene climate-related mental health burden are scarce [78]. Analogously, health care systems are ever-changing [82] as they must adapt to new clinical care guidelines, changes in reimbursement policies, care delivery modality (ie, telemedicine), quality improvement efforts, and local, state, or federal law amendments. For example, in April 2020, the American Society of Colposcopy and Cervical Pathology released new guidelines to provide recommendations on cervical cancer screening frequency and follow-up tests for abnormal cervical cancer results [83]. These guidelines significantly differ from the previous 2012 version [84]. Any implementation strategies designed to facilitate the adoption of the 2012 guidelines became obsolete and needed to be revised. For another example, EHR-based patient portals are efficient systems for communication between patients and health care providers and platforms for health information exchange. These portals can be a platform for patient-centered implementation strategies to improve the uptake of evidence-based practice. Patient portal tools have been used to improve the uptake of ACP or lung cancer screening [85]. Patient portal adoption before the COVID-19 pandemic, however, remained relatively low and varied widely across patient subgroups (eg, by age and socioeconomic status), diminishing the effectiveness of strategies implemented within the portal [86,87]. The need for social distancing and the uptake of telemedicine during the COVID-19 pandemic led to a rise in patient portal use, which could improve the reach of such strategies [88]. The uptake in patient portal during the pandemic was also associated with a rise in “e-visits,” which were communications between patients and clinicians between in-person visits [89,90]. This led to health care systems to bill for these messages following existing federal rules [90,91], which in turn may limit the use of patient portals and impact their effectiveness as an implementation strategy. This example illustrates how the changes in the health care system can impact a specific implementation strategy. Consequently, the reach, adoption, and sustainability of the EBI it aimed to improve are also impacted. These ever-changing systems pose a significant complication when using ML models [92,93]. How frequently should an ML model be adapted or recalibrated to ensure that it has accurate predictions and is unbiased and ethical? This is a critical factor impacting the use of ML in implementation science and across the 3 stages of implementation and remains to be answered by future studies.

**Conclusions**

ML can assist with predicting what will work best, for whom, under what circumstances, and with what level of support, or what and when adaptation and deimplementation are needed. However, there are many remaining challenges with integrating ML into various stages of implementation, which require further research and investigation. Tackling these challenges has the potential to render ML as an innovative and useful tool in implementation science in years to come.

**Acknowledgments**

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health (awards P50CA244289, P50CA244690, and P50CA244693). This program is supported by funding provided through the Cancer Moonshot. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Generative artificial intelligence was not used for any portion of the paper writing.

**Authors’ Contributions**

All authors substantially contributed to the conceptualization and edits of this viewpoint and approved the final version.

**Conflicts of Interest**

JB reports personal fees from Reimagine Care, AstraZeneca, and Healthcare Foundry and grants from Lilly Loxo and Gilead. The other authors have no conflicts of interest to declare.

**References**


78. Hajat S, Gasparrini A. The excess winter deaths measure: why its use is misleading for public health understanding of cold-related health impacts. Epidemiology 2016;27(4):486-491 [FREE Full text] [doi: 10.1097/ede.0000000000000479] [Medline: 26986872]


Abbreviations

ACP: advance care planning
CDS: clinical decision support
EBI: evidence-based intervention
EHR: electronic health record
FINDER: foodborne illness in real time
HIPAA: Health Information Portability and Accountability Act
ML: machine learning
RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance
SIF: strategic implementation framework

©Nathalie Huguet, Jinying Chen, Ravi B Parikh, Miguel Marino, Susan A Flocke, Sonja Likumahuwa-Ackman, Justin Bekelman, Jennifer E DeVoe. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 22.04.2024. 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 the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.
Discussions With End Users to Inform the Vision for a Shared Care Record in Ontario: Qualitative Interview Study

Marta Chmielewski¹, BSCN, MPH; Matthew J Meyer²,³,⁴*, PhD

¹Public Health Program, Schulich School of Medicine & Dentistry, Western University, London, ON, Canada
²Middlesex London Ontario Health Team, London, ON, Canada
³Office of Health System Transformation, London Health Sciences Centre, London, ON, Canada
⁴Department of Epidemiology and Biostatistics and Interfaculty Program in Public Health at the Schulich School of Medicine and Dentistry, and the Ivey Business School, Western University, London, ON, Canada

*all authors contributed equally

Corresponding Author:
Marta Chmielewski, BSCN, MPH
Public Health Program
Schulich School of Medicine & Dentistry
Western University
1465 Richmond Street
London, ON, N6G 2M1
Canada
Phone: 1 519 685 8500 ext 56765
Email: mchmiel2@uwo.ca

Abstract

Background: Improving the health outcomes of populations of individuals through population health management requires the use of electronic health records that can exchange real-time digital information using an accurate and complete shared care record that is accessible to health care providers, services, and patients.

Objective: The aims of this study were to understand end users’ (health care providers) experiences, attitudes, and insights using current electronic health records; their expectations of what is required to establish a shared care record; and how they anticipate adapting to the use of a shared care record in daily practice. This work is the result of a quality improvement initiative deemed not to require ethics approval according to the Western research ethics board checklist.

Methods: Clinicians were contacted using voluntary response sampling and interviewed via Zoom (Zoom Video Communications) between June 2022 and July 2022. The participants were from various health care sectors and at various stages of career development.

Results: Overall, adaptation to the use of a shared care record was viewed positively by health care providers, highlighting the benefits of a centralized, shared, and accessible location for real-time data, promoting patient continuity of care. The main concerns included the privacy, confidentiality, and security of the record along with patients’ ability to interpret their own medical information found in a patient portal. The resources requested by end users included multifaceted ongoing training on the use of a shared care record.

Conclusions: This study provides practical findings that will help emphasize factors that facilitate clinicians’ practical use and process of adaptation to the use of a shared care record.

(Keywords: population health management; shared care record; health information exchange)

DOI: 10.2196/51231
**Introduction**

**Population Health Management**

Defining population health management (PHM) requires taking a step back and thinking of the bigger picture. The Population Health Alliance PHM framework explains that “a population health management program strives to address health needs at all points along the continuum of health and wellbeing through the participation of, engagement with and targeted interventions for the population” [1]. Their definition specifies that the goal of PHM is to uplift or improve the “physical and psychosocial wellbeing of individuals through cost-effective and tailored health solutions.” Breaking it down, PHM involves the use of data to proactively manage the health and well-being of an identified population of individuals while considering the diversities within that population along with their social determinants of health [2]. PHM is a constantly progressing concept that is increasing in popularity worldwide. For example, in Ontario, PHM has been characterized as a fundamental element in Ontario’s health system transformation. In the Netherlands, several PHM initiatives are working to tackle the health-related social needs of residents by building partnerships among medical care, public health, social services, and community-based organizations [3]. Managing populations of patients based on their diagnosis while maintaining their health and keeping them out of dangerous circumstances has recently become popular as it affords the ability to deliver high-quality and efficient care that is satisfying to everyone involved [4]. Examples of leaders in PHM and integrated health care delivery include those in Denmark, Spain, and the United States, such as Geisinger, Memorial Hermann, the Department of Veterans Affairs, and Kaiser Permanente. Another important mention is Epic, a widely used software company among hospitals that allows for the exchange of medical records across organizations in the United States and beyond. They are all highly regarded for high-quality and efficient health care through integrated care delivery processes [5].

PHM can be viewed as improving the health outcomes of a population using appropriately coordinated care and proper patient engagement, which is sustained through adequate economic and care models [6]. The question then becomes how to best support the entire population clinically and financially. According to the Health IT Playbook [7], examples of PHM services involve efforts to proactively help people improve their health, guarantee they obtain preventive screenings, and help them effectively manage their chronic conditions. A vital feature of this approach to care delivery is that the population whose health is being managed is a complete group of people, not only those who are pursuing health care. This population can be defined in ways such as all the employees of an employer, members of a health insurance plan, or residents of a community, but the key feature in PHM is that the health of all members of this population is considered [7].

Another crucial aspect of PHM is coordinating a diverse and progressive group of stakeholders who work together to provide programs, services, and tools for interoperable care for patients in various health care settings [8]. This is also where the integration of services occurs, such as financing and delivery of health care working together [4]. According to Jones and Smith [4], an entirely integrated care system is defined as both horizontally and vertically integrated. Vertical integration combines provider and care delivery, financing, and support services such as IT. Horizontal integration combines provider services, home health services, hospitalization (tertiary and secondary), and ambulatory care, entailing continuous and seamless care [4].

**Health Data**

The growing burden of chronic diseases challenges health care system sustainability in countries worldwide. Working toward coordinating care to prevent unnecessary hospitalizations is a crucial solution to limiting increasing health care costs. According to Burnel [9], to reach this goal, clinicians and professionals must be able to exchange information using electronic health records (EHRs). An EHR is a real-time digital form of a patient’s health care record, allowing information to be available to providers authorized to access it across different health care organizations instantly and safely [10]. Beyond providing a patient’s collected medical data, an EHR offers a comprehensive view of a patient’s care. An EHR contains information from all providers involved in a patient’s care concerning admission documents, diagnostics, ongoing assessments, and health care plans and can be shared with other health care providers; caregivers; patients; and organizations, including laboratories, medical imaging facilities, specialists, pharmacies, and clinics [10]. EHRs make it possible to reduce medical errors, increase health care provider communication, and improve care coordination [11]. The broad implementation of EHR systems in primary care has permitted the compilation of enormous amounts of clinical data that have the potential for secondary use, such as improving clinical programs, system management, and population health research [12].

**Shared Care Record**

With EHRs in mind, the concept of the shared care record is introduced. A shared care record is an enabling feature that allows PHM to be possible. According to the Patient, Family, and Caregiver Declaration of Values for Ontario [13], to enable integrated care, each resident in Ontario ought to have access to their health-related information record, which is “accurate, complete, available and accessible across the provincial health system at [their] request.” The record should be accessible to health care team members and patients as required and in a manner that encourages appropriate care and positive experiences. A complete and accurate shared care record includes up-to-date information about the person and their demographic information, the administrative services they use, their medical or clinical information, and additional health-related information involving the social determinants of health.

Shared care records permit all primary and secondary care providers to view and use a single dependable source of documentation that is up to date and provides accurate clinical information in real time about a patient [14]. It is about giving everyone access to the information they need but does not require everyone to be on one common information system.
Patient records from a variety of care providers and sources can be linked through a health information exchange (HIE) system. The vision of the shared care record would give providers, in the home database system they work in daily, access to information captured about their patients from other care providers and other systems. For example, a provider (or patient) can view medications prescribed by provider A alongside those prescribed by provider B in the same place even if providers A and B use different EHR technologies. This information could come from their primary care records, home and community care records, community mental health and addiction records, or hospital systems.

Systems worldwide are using this approach to link information on allergies, laboratory test results, procedures, appointments, and much more. The HIE simply enables information exchange between systems, for example, between hospitals and primary care [15]. Moving this information between the systems aims to help the care team locate and use the correct information to provide safe, efficient, and equitable patient-centered care. This means that a patient only needs to describe their health care history once instead of sharing it multiple times at each health care encounter. If done properly, information from this system can also provide information to public health teams to understand the health and health needs of the population [15].

The Need to Reform Service Integration in Ontario

The current health care system in Ontario is experiencing increasing strain from the aging demographic, overloaded hospitals and emergency departments, and a significant increase in chronic diseases, putting our care delivery model at risk [16,17]. Completely changing the Ontario model of health care delivery is not feasible; therefore, we must work with the existing structures. One example is health IT systems that can be better connected to improve workflows; centralize health data; and deliver information to health care providers, patients, and families where and when they need it. Ontario is not alone; fragmented care exists among health care systems worldwide involving a lack of communication between primary care physicians, other health care providers, specialists, patients, and families, leading to negative impacts on patients and gaps in continuity of care. Many systems have realized the benefits of interoperability, “defined as the ability of different health information systems to cooperatively access, integrate and exchange data to advance effective delivery of health care” [16].

Several obstacles must be kept in mind when it comes to the electronic exchange of health information, such as technical, financial, legal, and privacy barriers that can impede the implementation of interoperability. Nonetheless, as health care providers request continuous integration of information and patients stress the need for access to health data, health care organizations will be forced to share information appropriately. This may require funding for information management technology such as EHRs and IT to enable care across the continuum [4,17].

The concepts of integrated care, digital health, interoperability software, and centralized health data, exemplified by the shared care record, are crucial to exposing the benefits of a restructured and better coordinated health care system. Collectively working toward a shared care record can help reduce medical errors, health care costs, and redundant and unproductive work while improving communication among health care providers, quality of patient care, and seamless transitions of patients across health care providers and settings to create a resourceful system [17,18].

Aim of the Study

This qualitative study used semistructured interviews to improve the understanding of end users’ (health care providers) perspectives and insights regarding how they anticipate adapting to the use of a shared care record. Information gathered from the interviews will support the development of use case storyboards to inform various stakeholders across Ontario of considerations for developing a shared care record across the province. Talking with end users will help understand what a range of clinicians from different specialties believe is required to establish a shared care record and how they will adapt to its use over time.

Methods

Setting

Middlesex County is in Ontario’s Southwestern region, covering a geographical area of 2800 km² and home to >450,000 people. This region consists of a mix of urban and rural residents. London is the largest metropolitan area within Middlesex County and is home to >450,000 residents. The region also surrounds 3 sovereign First Nations: the Chippewas of the Thames, Oneida Nation of the Thames, and Munsee-Delaware Nation. For several years, legislators in Canada’s most populated province, Ontario, have endeavored to change the local health care system to create a more coordinated and financially united system [19]. This initiative resulted in the Government of Ontario Ministry of Health formation of 54 approved Ontario Health Teams (OHTs) within specific geographic areas across the province. OHTs modify how health care is financed and delivered and concentrate on collaborative partnerships in which providers and organizations such as primary care, mental health services, hospitals, and home and community care work as one synchronized team [20,21]. The Middlesex London OHT is specifically responsible for supporting the health of the population residing in the Southwestern Ontario region [22]. Using OHTs, the provincial government is assembling sustainable systems that will respond to local populations’ short- and long-term needs, support local services, and enable straightforward system navigation and transition among providers [19,23]. Another critical player providing guidance and regulation is Ontario Health, a government-formed agency working to coordinate and connect the province’s health care system [24]. This new visualization of Ontario’s health care system is aligned with the Quadruple Aim, a framework internationally understood to design and provide a system that improves patient and caregiver experiences, patient and population health outcomes, and provider experiences while reducing total costs [23].

Participants and Recruitment

Using voluntary response sampling, clinicians were contacted via email based on preexisting professional relationships. A
total of 14 health care providers were interviewed, comprising those who volunteered or agreed to participate upon request. These health care providers hold positions in various care sectors, including nursing, community care, primary care, emergency medicine, dietetics, practice specialties, occupational therapy, and physiotherapy. The list of health care provider interviewees who agreed to participate in the data collection, organized by occupation, is shown in Table 1. The providers also ranged widely across stages of career development and duration, from new graduates to experienced employees.

<table>
<thead>
<tr>
<th>Professional title</th>
<th>Care setting</th>
<th>Professionals (N=14), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registered nurse</td>
<td>Inpatient acute care</td>
<td>3 (21)</td>
</tr>
<tr>
<td>Registered nurse</td>
<td>Cardiac outpatient clinic</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Registered nurse</td>
<td>NSWOC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Clinical nurse specialist</td>
<td>Chronic diseases and clinical informatics</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Clinical dietician</td>
<td>Bariatric outpatient clinic</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Physician</td>
<td>Emergency medicine</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Physician</td>
<td>Primary care practitioner</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Registered physiotherapist</td>
<td>Outpatient clinic and inpatient acute care</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Registered practical nurse</td>
<td>Home and community care</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Clinical practice specialist</td>
<td>Occupational therapy</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Clinical practice specialist</td>
<td>Palliative care and oncology</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Occupational therapist</td>
<td>Home and community care</td>
<td>1 (7)</td>
</tr>
</tbody>
</table>

<sup>a</sup>NSWOC: Nurse Specialized in Wound, Ostomy and Continence.

**Data Collection**

Participants in the study were first introduced to the concept of a shared care record verbally and through a video demonstrating its functionality. All participant questions about a shared care record were answered before the interviews. A semi-structured question guide ensured that each interview covered essential topics and allowed participants to disclose issues and stories as they saw relevant. The use of a prepared guide also worked to decrease interviewer bias by decreasing interviewer involvement. Confidentiality and anonymity were established at the beginning of the interviews. Verbal consent was obtained from each participant to potentially use quotes from the discussions in future publications or presentation materials that result from the initiative. The semi-structured interviews averaged 15 (SD 2.56) minutes and were web-based via Zoom (Zoom Video Communications) between June 2022 and July 2022.

The interview format was chosen, as opposed to focus groups, as it allowed for direct, individual engagement with each end user. Stokes and Bergin [25] discussed the opportunity for the interviewee to truly analyze their motivations for a particular action while being given a feeling of empowerment because of the anonymity in the individual interview setting without the pressures of a group setting that may lead to a consensus view. The interviews were designed to elicit the health care providers’ understanding of and experiences with the EHRs they currently use along with their attitudes, beliefs, and expectations regarding the future use of a shared care record in their daily practice. The sequence of interview questions used and additional instructions to guide the interview are shown in Table 2. The interviews were audio recorded with permission from the participants, transcribed using web-based software, checked for accuracy, and then analyzed to develop a report. The results present the participants’ initial reactions to the concept of a shared care record and then transition to their interpretation and reflections on the use of and adaptation to a shared care record.
Table 2. Interview guide.

<table>
<thead>
<tr>
<th>Interview portion</th>
<th>To do</th>
<th>Additional notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warm-up</td>
<td>• Introduce and explain the purpose of the interview. Obtain consent to use quotes from the interview and to record the interview.</td>
<td>• Introduce the idea of the shared care record and how it works.</td>
</tr>
</tbody>
</table>
|                   | • Consider your electronic health record today (name the record), what additional health or social information regarding your patient would you like to have access to in this new shared care record or would make a difference for you, when providing care for that patient?  
  • Prompt: tell interviewee more about what a shared care record could offer them. | • Build each question off the previous one and rearrange the order as needed according to the flow of the conversation. Use the prompts to further stimulate conversation. |
|                   | • So, for you, tell us how this information would impact or change the care you provide (or can) to your patients?  
  • Prompt: what difference would it make if you had access to all of your patient’s information (the type of information you just listed) on the shared care record?  
  • Prompt: would you look at it more or make use of that information? | |
|                   | • This would be a change and change is never easy, but what do you think you and your colleagues would need to do to adapt to using a shared care record in your day-to-day practice?  
  • Prompt: what would help your colleagues benefit from this change?  
  • Prompt: what support or resources would help you with the introduction and adaptation to this system? | |
|                   | • Do you have any concerns with the concept of a shared care record?  
  • Prompt: Overall, what factors or conditions challenge or serve as barriers to your personal use of a shared care record? How might this change that? | |
|                   | • In closing, do you have any concluding thoughts or comments related to the shared care record that you would like to convey? | |
| Closing           | • Thank the interviewee for their time. Let them know not to hesitate to reach out if they have anything else they would like to discuss. | • Ask interviewee whether they give permission to be contacted in the future. |

Data Analysis

The interviewer first read and interpreted the individual transcripts to become well acquainted with the data collected. After developing the initial semantic codes based on the data, the interviewer grouped the codes into categories and themes and then reviewed, named, and discovered various connections between the themes to write the analysis. The themes were modified using an iterative process, adjusted, and grouped, with categories and subcategories added as they arose from the data analysis. The interviewer used a qualitative interpretative approach, the framework method, to analyze the data by joining thematic analysis with comparison so that the data were surveyed for known literature themes and emerging themes. The interview findings are presented in the Results section of this paper; quotes are included from the interview transcripts to illustrate the generated themes. End users are identified by their health professions in each quote.

Ethical Considerations

This publication is the result of a quality improvement initiative deemed not to require ethics approval according to the Western research ethics board checklist.

Results

Participants’ Reflections on the Shared Care Record

Theme 1: Opportunities for Using a Shared Care Record

Lack of Communication Affecting Care Delivery

A lack of adequate communication among health care providers, services, and health care facilities across the care system was a common response among participants. Discussions with end users highlighted frustrations across the care continuum, such as entering care encounters with inadequate or lacking information, more difficult care management and planning, and delays in access to information causing delays in care. Registered nurse 4 stated the following:

*Why are we doing the same assessments over and over again? Patients are forced to repeat tests because the results are not passed between the health care providers, which in the end only delays their [the patients’] treatment.*

The introduction of a shared care record could drastically improve communication among health care providers, potentially decreasing the current workload, increasing confidence in decisions, and affecting patient safety and continuity of care:
...with a client that I’m seeing, he’s a cancer patient, so he’ll go to London, and then he’ll go to Stratford for example. The two hospitals they don’t communicate very well...the communication between the two kind of gets lost in between...it’s all of these extra steps between myself and this patients’ daughter. We are trying to figure out when was his last treatment? And was the medication provided? What was that medication? And how long was he supposed to take that? So, you have Stratford asking this and trying to get through to London to get those questions answered, the whole process becomes very difficult.

[Registered practical nurse]

End-user discussions touched on the impact of poor communication on patient satisfaction, trust, and their subsequent health care journeys. Occupational therapist 1 described how enhanced communication between providers could affect the patient care experience:

From a patient perspective, it might be one less time they have to answer the same questions. Not being asked the same questions all over again seems trivial, but if you’re the patient who’s had to answer the same question twenty times you think people aren’t listening to you...we could just kind of summarize what we know, which I think also makes the patients feel like we’re all a team speaking with each other. So, we’re communicating what we know about the patient, as opposed to having to ask them the same information over and over and over again.

The retrieval of information to provide proper care was deemed exceptionally crucial among end users working in community settings, where some may not have any connections to EHRs:

I mean our nurses are walking in really with very minimal information. Again, relying on the family a lot of the time to tell us, even as far as medications that they’re on, you know, we’re going through all of their bottles and discharge lists, and lists they pull out of their wallet, and trying to reconcile it. So, it’s really pieces of information. [Clinical practice specialist]

Real-Time Information

End users collectively admired that one of the most critical advantages of the record would be the access to real-time information as it changes and becomes updated. A clinical nurse specialist described information access as “very much a game changer for clinicians,” with other participants agreeing, describing it as “taking the legwork” out of obtaining essential patient information. The shared care record real-time information feature “would not only help the patients but also the healthcare team be up to date, and they wouldn’t have to take so long to find the information they need” (registered nurse 3). Considering their daily practice allowed the end users to visualize clinical patient data being stored in one central location where the information could be accessed, analyzed, uploaded, and used, with one process going to everybody involved in the patient’s care:

I’m really intrigued by receiving information in real time, I find that especially in the community if there has been a medication that’s been added, I don’t always receive that information, unless I’m at the client’s home and I see the new bottle because they don’t typically tell me if they are on a new medication, and the doctors just prescribe it, they definitely don’t inform us. That would be information would be very helpful for me when trying to figure out why they’re having a change in their health status, or maybe a cognitive change or something along those lines. [Registered practical nurse]

Theme 2: Perceived Benefits of Using a Shared Care Record

Effective Use of Time

Access to clinical information through the shared care record was described as promoting the effective use of time and resources. End users felt that the record would provide “an accurate picture of what’s going on,” and it would be “a lot less doubling and tripling of assessments.” Registered nurse 1 recalled a common scenario occurring in their inpatient hospital unit:

...on my floor patients come up with a bag of medications or just a list of medication names and dosages that they have scribbled onto a piece of paper...the time we are spending on doing something very basic like manually inputting medication information that should be available through the pharmacy or from a physicians list would save so much time and then you would be able to spend more time doing a proper assessment on the patient, providing care or starting a treatment.

Within the complex and fast-paced acute care setting in which several of the interviewed end users worked, the ability to save time was the most significant determinant of efficiency. Effective care must be provided with often limited resources and high workloads. Occupational therapist 1 discussed the vision of the shared care record as a benefit to their work:

...we’re always looking in acute care for efficiencies. The length of stay is already very short, and the more information that we have access to when we are doing our initial triage or chart review, the earlier we can start to at least reflect on what the likely plan is.

Informed Circle of Care That Promotes Continuity of Care

The vision of the shared care record would allow the patients’ circle of care, everyone involved in the patients’ care, to be well informed, leading to better care, time savings, and less frustration. According to registered nurse 1, “it would promote continuity of care and keep everybody in the loop and informed, which is so important in healthcare in general.” End users discussed the benefit of being able to collaborate with other clinicians and share more information in general:

...even when our clients go into hospital, they [the hospital occupational therapists] have no idea what we have been working on at home. And then if they
go to discharge the client, sometimes they’ll put in new OT and PT services without realizing there’s already different things in place. There’s just poor communication, so I think this idea would make a huge difference. [Occupational therapist 2]

The consensus among the participants was that this vision becoming reality would change the way they practice, offering the ability to connect with everyone that the patients are in contact with:

...everything is just in silos right now, and I think anyone that’s had any contact with the healthcare system knows that. I think it’s very prominent in community…it is not practical how it is right now, so I think any move in that direction [access to a shared care record], would help immensely. [Registered nurse 4]

With regard to the patient-provider relationship, providers indicated that patients in the hospital setting can feel vulnerable and left out of conversations and might not know what is going on:

If we could retrieve those records...it would help the patients feel comfortable, they would be able to ask more questions during their stay and obtain answers from us as providers.

No one wants to be in the hospital as it is, so when they [the patient] come in, and they notice that their cardiologist has no idea that so and so [other healthcare providers] prescribed a certain medication, patients tend to become annoyed, and rightfully so...having that information prior to their visit would make their visit a lot easier, faster, and more efficient for them and for everyone involved. [Registered nurse 4]

Overview of Patient Health Status

The increased amount of health and health-related information accessible to health care providers would help them understand their patients’ medical requirements. A physiotherapist discussed how treatment of his patients would be enhanced “by helping me know and understand their timeline for recovery.” The record would “help understand other areas they [the patient] need help with because patients forget things and don’t always understand what other healthcare providers tell them when it comes to their injury.” Participants noted a lack of patient awareness regarding what providers are involved in their care, medication management, or even their diagnoses. Physicians discussed situations in which, unless their patients informed them, they were unaware of new allergies or changes to medication dosing made by other physicians. Registered nurse 2 discussed that, when noticing a new irregular sign or symptom, the shared care record would allow for “a quick reference, if that’s something they [the patient] have at baseline or if you need to look into it further, and just kind of base your actions on that information”:

...especially going into people’s homes, it would just make me more aware of things. People won’t always tell you the truth about things or they’ll leave out things they don’t feel are important but are impacting how they are managing at home...If you knew the information, walking into it, you kind of have a more holistic picture before moving forwards with them. [Occupational therapist 2]

Adaptation

Theme 3: Factors That Promote Use of the Record

Positive Outlook on Adaptation

Overall, the participants revealed an incredibly positive outlook when considering their own and their colleagues’ adjustment to the record. Participants made comments such as “it wouldn’t be a challenge for me and my colleague’s personally” (registered nurse 3) and “I don’t think there would be a large change-related level of concern or anxiety” (occupational therapist 1). End users acknowledged that technology is on the rise, with many individuals of all ages using digital solutions in their daily lives:

...even just logging in and seeing their bloodwork online, people are more comfortable doing that...even if they’re older or they haven’t done that they’re comfortable reaching out to their neighbor or their child to help them with that...I think it would be easier than you know even 5 years ago. [Registered nurse 4]

Health Care Provider Requests for Health Information Access

When end users were asked what other patient health or social information they would like access to when providing patient care, most made remarks regarding the difficulty of obtaining access to clinical data documented beyond the organization they worked within. The most common request among end users was access to a verified list of medications. Often, providers must rely on medication bottles, discharge lists, and family members to reconcile patient prescriptions:

...patients might know the name of the drug, or what the pill looks like, but they have no idea what it does, or why one of their physicians ordered it. For example, when they come into the clinic telling us that their nephrologist ordered something to bring their blood pressure down because their kidneys are failing, their cardiologist might just be finding out about the medication and realizing the medication could be affecting their heart. So being able to see what was ordered and when it was ordered would be such a big help. [Registered nurse 5]

The lack of integrated health IT was found to cause duplication of efforts and lack of comparison across documentation from the hospital versus the community or across organizations or different regions:

I cannot see pictures of x-rays that they [patients] get in the community. So, people get sent in with a break and I need a picture to see if I need to push on it, to put a cast on it and get it in the right place. We often repeat x-rays that probably wouldn’t need repeating if we could just see the original picture. [Physician 1]
Interoperability between organizations can reduce redundant tests, save time and costs, and result in better continuity of care. Participants regarded patient diagnostic imaging as the information they would like to access, including the actual pictures and not just the reports of x-rays, ultrasounds, computed tomography scans, echocardiograms, and other imaging modalities. Furthermore, patient exposure to radiation or contrast dye would decrease without the need to repeat tests and scans:

…it would help me if I had access to imaging without having to rely on what the patient tells me or what they can even remember. Sometimes patients don’t even know what type of imaging they had done or even what that imaging was for. [Physiotherapist]

Participants listed patient medical-related appointments as the information they wanted to see on the record. Many patients have complicated cases and multiple teams following them in the community, which can be overwhelming to manage independently. Family physician records or membership in a family health team was another common request, along with up-to-date access to all referrals sent out and specialist information:

Knowing what doctor the patient was referred to...a lot of time people will come in and will say “well my doctor sent me to a cardiologist, but I haven’t heard anything in three months” So then I’m like “well I’ll send you to a cardiologist as well.” Am I sending them to a different cardiologist? I have no idea. [Physician 1]

Access to patient social history information was highly requested, including living arrangements, home care reports, and community support or professional services that patients were using as this information significantly affects patient care planning and discharge planning. According to physician 2, “it would be nice if the patient could update things like occupation, substances, family members, consent to family members”:

In many cases, we don’t have a true understanding, at least initially on chart review, without speaking with our team about the exact specifics of what type of services or equipment or programs they are [the patient] currently involved in, in the community. A true understanding of that social and community history allows us to initially strike off maybe some options that we may not have at our disposal, or start to plan out some of the gaps that we anticipate based on what we know that they already have. [Occupational therapist 1]

End users discussed situations requiring access to patients’ medical histories and complete health records. Requests for clinical data access included a complete list of diagnoses and when the patient was assessed for them; laboratory test results; previous rehabilitation journeys; surgeries; and conversations that had taken place, which could indicate the patient’s understanding of their illness or where they are at:

Almost every single clinician that I have spoken to would say, we wish we had more, or the information is just a very brief summary...especially with acute care, the length of stay is so short, we’re trying to piece together as much as we can...so the more the more understanding of the patient’s history and journey through the healthcare system, the more efficient we can be as occupational therapists. [Occupational therapist 1]

Ongoing Training

The most common suggestion among end users was the provision of proper education and training on the new system. As one of the primary objectives of an EHR is to improve collaboration among health care providers, it only makes sense that they are offered the chance to provide feedback on the system they use daily:

From a training perspective, it’s nice to first of all be part of the process of building the system, or having some input on that system, which helps with the engagement and integration when we’re actually putting the rubber to the road...it definitely would help with the connection to the implementation. [Clinical nurse specialist]

Most participants suggested that they would benefit from getting to know the new system through proper training on the layout of the information and how to find the information that they could use. End users visualized the benefits they could obtain from an introduction to and familiarization with the system before it becomes implemented in practice:

...obviously, there needs to be training, and along with that comes the resources. Not only do the training but pay for them [the end users] to attend, which is always an issue. And then I think even support along the way, for example IT support, do we need to build that internally in our IT department...up front, it’s just really the education and making sure it is ongoing...in healthcare in general, there is a lot of turnover, so how do we sustain the education moving forward. [Clinical practice specialist]

Many participants discussed “multi-pronged approaches” as the most effective method for introducing and adapting to the record. Resources mentioned by the participants included a chat or live support option for immediate questions, videos on how to use the system, a toolkit or tip sheet developed by the system creators, and in-person and web-based computer sessions. Occupational therapist 2 described the introduction to their current EHR system as they recalled:

Clinicians felt more comfortable using the system if they had some test patients to go through trial cases of what a daily patient intervention might look like prior to the go live.

Several participants mentioned the idea of “super users”:

...our nurse colleagues on the floor, who had additional training and were more familiar with the record so that we could reach out to them if we needed help or if we had questions. [Registered nurse 1]
These super users would function to support their colleagues in the transition while helping others learn to use the record to the fullest extent.

Another resource identified was the use of clinical educators who already work to support staff with clinical updates to rules, procedures, policies, and methods of accessing information.

**Record Accessibility**

Several questions from end users concerned how to physically access the clinical data on record and the timeliness of finding information. Questions included the following: “where do we need to click?” “Under what icon?” “How do I add things to the shared care record?” “Do I have to do it manually?” “Does it just happen automatically?”

The biggest adaptation would be how to access the information, like opening the charting system will look different, so getting used to the new layout and the new system and knowing where to find things.

[Physiotherapist]

A key finding among participants included statements regarding the user interface or usability of the shared care record. Participants used terms such as “seamless,” “simple,” “accessed quickly,” “user friendly,” and “easy to follow” to describe how they envisioned the software to function:

But we need to really limit where they’re [frontline staff] finding their information, if they need to upload, that they are not having to do it to all of the people we need to report to. There are so many layers and rules, and we just need to make it as simple as possible. [Clinical practice specialist]

Participants considered the least amount of clicking and integration with their current systems to obtain data or add to their assessments as crucial features of the record. Information being uploaded automatically was considered foundational, with physician 1 commenting the following:

If we have to do an additional step at the end to get it uploaded, you’re going to get way less uptake...as long as in the back end of things, my EHR links it all up.

Regarding user-friendliness, the physicians explained that they would not appreciate retyping a password to access the charting or repeated verification of the designation upon entering the system:

I don’t want to have to log into something else, I’m already logging into so many things every day...and so there’s that information that sits somewhere but that it gets pushed to all the different places and then shared between the different places. [Physician 2]

**Theme 4: System-Dependent Considerations and Concerns**

**Clinical Data Consistency, Accuracy, and Organization**

Although participants recognized numerous potential benefits, they also discussed fundamental considerations of functional practicality, such as the consistency and accuracy of data across the record. To present clinical data across different EHR systems uniformly, health care providers must be consistent with the documentation methods and upload the documentation to the record:

I would be concerned about it being unorganized or messy, um if everyone has different styles and systems of taking notes or recording, maybe it would be difficult to find one particular piece of information that you are looking for. [Physiotherapist]

Understanding patient rostering or enrollment was discussed as a critical element of the record, understanding who is involved in a patient’s care, and participating providers can change or adjust that if needed:

I changed my practice maybe three years ago, but I am still on some people’s charts at the hospital. I have requested to have my name taken off, but unless the patient calls and changes that, nothing can happen. And so that becomes a privacy issue. I keep getting files for people who I am not actually taking care of. [Physician 2]

Beyond consistency, participants identified the importance of double-checking the information obtained from the record with the patients themselves. Updates may not be revised, data could be deleted, and mistakes can still occur:

If you have a medication record from two years ago, you would still have to do your due diligence to make sure the information you are using is accurate. [Registered nurse 1]

An auditing system of the record was suggested that could review charting to help ensure that health care providers input the required information to maximize the utility and reliability of the clinical data. According to registered nurse 5, “that way people have to take responsibility for what they’re changing or what they’re contributing towards this shared documentation.”

**Change in Workflows**

Uncertainty regarding daily practice workflows came up as a barrier to overcome when participants discussed adaptation to the use of the record. Visualizing the details of the change to their current EHR system interface was difficult for certain participants:

It’s hard to know what the change management strategy would be. [Clinical practice specialist]

My hope is that there would be very little that we would actually have to change...everything else I would expect to be kind of behind the scenes where I do my normal process, that it would just sort of happen in the background. [Physician 2]

Discussions held with professionals in the community setting revealed that there would be an adjustment to their current workflow, with the additional time spent reviewing history, reports, and other data accessible in the record before going in to see the patients:

Right now...I really only check [name of EHR] to look at when appointments are confirmed, phone numbers, names, and then I find out a lot more information from the patient once I get there [to their place of
residence...for home care, that’s just kind of how it’s been. If I am seeing four or five people every day and driving between these destinations, it would take time maybe at the beginning before getting used to it as part of the routine. [Occupational therapist 2]

Upon introducing the idea for the record, a statement from the Ontario Patient, Family, and Caregiver Declaration of Values [10] was presented to participants explaining the vision for transparency in patient access to their health records. Wondering how patients would interpret seeing physicians no longer in the patient’s circle of care re-entering their medical information and whether this would be concerning, physician 1 stated the following:

...Most of us [physicians] require understanding of, did the treatment I gave actually have a good effect? And what did the follow up doctor think? So, we will access records a few weeks later to see what happened so that we can learn...that’s how I learn and how I can change how I practice, which is super important...we are expected to do continuing professional development and take courses, but then it’s always read and learn around your cases. How am I supposed to learn around my cases when I can’t find out what the specialists thought of this unique situation that I can’t just open a textbook and read about.

Privacy, Confidentiality, and Security of the Record
The most common concern among participants regarded privacy and confidentiality. The extensive personal health information or personal information accessible in a central location, this being the record, increases the risk of a privacy breach:

I think about it, not only as a clinician, but as a user of healthcare as well. [Occupational therapist 1]

According to registered nurse 1, with “a lot more information that is available to you as the healthcare provider, it would have to be ensured that only people who are part of the patient’s circle of care are accessing this information and that you are only accessing records that are applicable to the care that you are providing.” Nevertheless, most providers, including the nurses, physiotherapist, and occupational therapist, agreed that the benefits of the shared care record would outweigh the risks, and everyone could work together to make it as secure as possible.

Patient Portal Access
The concept of a patient portal, as described in the video shown to the participants, raised many questions and some hesitation among them. According to physician 1, “if the patients can see all of the notes that I write, that might change the way I document and I may be less comfortable putting the note in [to the record].” Understanding that patients will have access to their health data, questions revolved around the extent of access provided, whether to their entire chart and complete provider documentation or solely to scheduled appointments and their care plan:

...For instance, the other day I was printing a note for a patient knowing that her family member was going to be sitting with her reading it, she [the patient] asked me not to tell her family that she smokes weed...So, now I am changing my note, instead I copy and paste into a discharge note and then kept in my actual note all the facts so that I have that in her chart for other doctors to see because that’s important medical information. [Physician 1]

Several end users expected that there would be system controls or protocols surrounding what information the patient could see or change themselves and what information the patient would not be able to change. Further discussions covered the documentation of sensitive patient information and how patients might respond to the presentation of such information:

...there might be certain things...in the record that maybe patients themselves would not want to see...if the patient was confused after a surgery and there was an episode of violence that was documented. It may almost be triggering or upsetting to them [the patients]. [Registered nurse 2]

Regardless, health care providers considered the inclusion of these types of documentation to be vital to the patient’s record:

I have a suspicion that this might be going on and I need to share that with my colleagues, because if they have a similar suspicion and it’s a pattern that’s important. [Physician 1]

When considering patient access to records, participants emphasized the importance of patients being able to interpret the information correctly and objectively, especially regarding medical jargon. Participants suggested that clinicians may be inclined to use different terminology or a different writing style or reformulate the information in a meaningful way for patients:

...kind of helping socialize the clinicians to the new reality of the patients being able to read their notes more readily or easily. I think there would be value in having a discussion as a team about those types of changes. [Occupational therapist 1]

Discussion
Overview
Concepts such as PHM, data security, and privacy can be complex to explain to individuals; however, they will become progressively essential to the design and delivery of health care. PHM is founded on interoperability, data sharing, and integration with diverse health sectors and services. Although people tend to understand the role and significance of EHRs, they may neglect the value of inputting accurate and high-quality data into them. PHM and primary health care strive for many of the same features, including person-centeredness; continuity; accessibility; and consideration of physical, mental, cultural, and social aspects of health, among others [2]. Health care providers commonly have a good understanding of the population that they serve, often living within the community themselves, and appreciate the needs and some of the determinants of health of these populations. A PHM approach rooted in quality data quantifies this understanding and enables an even deeper level of understanding [2,17]. As the vision for a shared care record using HIE technology starts coming to life,
obtaining end users’ opinions and ideas will be imperative. End-user involvement in the record’s design; development; and, ultimately, operation will help simplify the adoption of changes and attain the goals of proactive and coordinated care that actively engages patients.

**Principal Findings**

This study provides practical findings that will help emphasize factors that facilitate clinicians’ process of adaptation to the use of a shared care record. Considering the fast pace of health care, clinicians highly commended and admired a central location for real-time information availability that could promote efficiency through the effective use of time. The benefits of accessible retrieval of information were especially highlighted among end users practicing in the community setting. Discussions with end users brought forth the importance of an informed circle of care, promoting patient continuity of care, and more effective provision of care. Health care providers requested access to additional information that would help them in their practice, from medication lists and diagnostic imaging to social community and home care support, laboratory test results, and referrals. Discussions also brought forth questions regarding the interoperability of the record, its functional usability, and changes in workflows.

Adaptation to a shared care record was viewed positively by health care providers. Several end users spoke about the benefits of getting to know the new system through proper ongoing training using multifaceted approaches. Some of the approaches considered included videos, in-person and web-based computer sessions, and live user support options. The idea termed “super users” was brought forth, whereby colleagues who would be more acquainted with the software would function as support for their coworkers in the transition and adaptation to use of the record. End users wanted to understand the functionality of the record, the impact of changes on their daily workflows, and the consistency and accuracy of data across the record to maximize the utility and reliability of the clinical data. The main concerns of participants were the privacy, confidentiality, and security of the record and patient information interpretation through the patient portal.

A growing body of literature on the topic of patient access to health care provider electronic visit notes suggests that the active involvement of patients at the point of care can foster stronger patient-provider therapeutic partnerships. A study by Wolff et al [26] suggested that most patients reported benefits of reading provider notes, such as more agreement concerning treatment care plans, increased ability to formulate questions to ask their care providers, and more productive care discussions. Walker et al [27] brought forth challenges such as patients not being registered on portals to allow for access to notes or patients being unaware of provider notes being available to access. Nevertheless, the benefits of expanded patient access to clinical notes have been established, holding the potential to better support and involve patients in care, increase communication, and provide feelings of control and preparation for health care visits [26-28].

**Limitations**

This study was limited in certain ways. The range of clinicians could have included various other providers within diverse health care settings to broaden the perspectives included. Furthermore, the application of voluntary response sampling in the recruitment of health care providers for this study is a limitation because of the possible sampling bias of respondents who volunteered, meaning that the study could have involved EHR advocates. Future research should involve a subsequent round of health care provider interviews once the record has a fully developed user interface design functioning across several systems involved with the HIE initial demonstration project. At this stage, health care provider interviews may offer further understanding of the functional usability of the shared care record once the providers can visualize and use it within the home database system they work in daily. These interviews could be geared toward comprehending how information design principles align with clinician workflows, patient information examinations, or decision-making in the medical environment. Building on this effort can help populations receive high-quality care while ensuring that it meets community needs.

**Conclusions**

This study provided insights into health care providers’ perceptions of a shared care record and presented their reflections on the practical use and adaptation to the use of a shared care record. It is essential to bring end-user perspectives into the shared care record’s development, introduction, and maintenance, along with the training necessary to permit the use of the system. There is an urgent demand for high-quality, integrated, and timely health data allowing individuals, health care providers, and communities to be involved and informed partners in the provision and attainment of health care [17].

**Acknowledgments**

The authors thank all the health care providers who gave their time to be interviewed for this study. This work was supported with funds from the Office of Health System Transformation, London Health Sciences Centre.

**Conflicts of Interest**

None declared.

**References**


Abbreviations

- **EHR**: electronic health record
- **HIE**: health information exchange
- **OHT**: Ontario Health Team
- **PHM**: population health management

Edited by E Mensah; submitted 27.07.23; peer-reviewed by M Gaynor, C Turvey; comments to author 13.11.23; revised version received 17.12.23; accepted 22.12.23; published 27.02.24.

Please cite as:

Chmielewski M, Meyer MJ

*Discussions With End Users to Inform the Vision for a Shared Care Record in Ontario: Qualitative Interview Study*

Online J Public Health Inform 2024;16:e51231

URL: https://ojphi.jmir.org/2024/1/e51231
doi:10.2196/51231
PMID:384

©Marta Chmielewski, Matthew J Meyer. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 27.02.2024. 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 the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.
Deriving Treatment Decision Support From Dutch Electronic Health Records by Exploring the Applicability of a Precision Cohort–Based Procedure for Patients With Type 2 Diabetes Mellitus: Precision Cohort Study

Xavier Pinho¹, MSc; Willemijn Meijer², PhD; Albert de Graaf³, PhD
¹Netherlands Organisation for Applied Scientific Research (TNO), Utrecht, Netherlands
²Nivel, Nederlands Instituut voor Onderzoek van de Gezondheidszorg, Utrecht, Netherlands

Corresponding Author:
Xavier Pinho, MSc
Netherlands Organisation for Applied Scientific Research (TNO)
Princetonlaan 6
Utrecht, 3584CB
Netherlands
Phone: 31 88 866 42 56
Email: xavier.sacastropinho@tno.nl

Abstract
Background: The rapidly increasing availability of medical data in electronic health records (EHRs) may contribute to the concept of learning health systems, allowing for better personalized care. Type 2 diabetes mellitus was chosen as the use case in this study.

Objective: This study aims to explore the applicability of a recently developed patient similarity–based analytics approach based on EHRs as a candidate data analytical decision support tool.

Methods: A previously published precision cohort analytics workflow was adapted for the Dutch primary care setting using EHR data from the Nivel Primary Care Database. The workflow consisted of extracting patient data from the Nivel Primary Care Database to retrospectively generate decision points for treatment change, training a similarity model, generating a precision cohort of the most similar patients, and analyzing treatment options. This analysis showed the treatment options that led to a better outcome for the precision cohort in terms of clinical readouts for glycemic control.

Results: Data from 11,490 registered patients diagnosed with type 2 diabetes mellitus were extracted from the database. Treatment-specific filter cohorts of patient groups were generated, and the effect of past treatment choices in these cohorts was assessed separately for glycated hemoglobin and fasting glucose as clinical outcome variables. Precision cohorts were generated for several individual patients from the filter cohorts. Treatment options and outcome analyses were technically well feasible but in general had a lack of statistical power to demonstrate statistical significance for treatment options with better outcomes.

Conclusions: The precision cohort analytics workflow was successfully adapted for the Dutch primary care setting, proving its potential for use as a learning health system component. Although the approach proved technically well feasible, data size limitations need to be overcome before application for clinical decision support becomes realistically possible.

KEYWORDS
personalized care; electronic health records; EHRs; machine learning; type 2 diabetes mellitus; T2DM; decision-making

Introduction
The concept of learning health systems (LHSs) is an approach to health care that emphasizes continuous learning and improvement through the use of data and analytics [1]. Realizing that the US health care system was continuing to fall far short of its potential of delivering the best care at a lower cost, LHS was introduced in 2012 by the Institute of Medicine Committee on the Learning Health Care System in America as a “vision of what is possible if the nation applies the resources and tools at hand by marshaling science, information technology, incentives, and care culture to transform the effectiveness and efficiency
of care - to produce high-quality health care that continuously learns to be better,” in their report Best Care at Lower Cost: The Path to Continuously Learning Health Care in America [2]. The underlying concept of the LHS is to harness the power of data and analytics to learn from every patient and feed the knowledge of what works best back to clinicians, public health professionals, patients, and other stakeholders to create rapid cycles of continuous improvement, which should allow to derive full benefits from leveraging data, systems, and human interconnectedness on an ever-increasing scale as is also seen in other sectors of the economy [3]. To help in deriving recommendations for implementation and evaluation criteria, LHSs have been conceptualized in frameworks from various perspectives, for example, impact on quality of health care [4], health care system evolution [5], and value creation [6]. The implementation of an LHS in practice requires the transition of a complex system of multiple stakeholders, processes, and technical (information) systems, irrespective of scale: local, regional, national, or international [7]. McDonald et al [8] recently identified the following as key enablers and actions required to enact LHSs: promotion of patient engagement, ensuring availability and access to data that are fit for purpose, keeping a focus on generating and implementing knowledge, creating organizational readiness, and stimulation of learning systems at different scales. The diversity of factors identified illustrates the broadness of scope needed in assessing progress in the relatively novel field of LHSs. To date, very few reviews on the subject exist. Somerville et al [9] in a systematic review identified key implementation strategies, potential outcome measures, and components of functioning LHSs but stressed that further research is needed to better understand the impact of LHSs on patient, provider, and population outcomes and health system costs.

As is evident from this short overview of literature on LHSs, the use of data to create knowledge is one aspect of LHSs; however, it is a central one. Focusing specifically on the impact of the use of electronic health records (EHRs) on delivery or outcomes of health care, only 5 (12%) out of 43 eligible studies in a single available review study were found to document a medium-to-high level of evidence for impact [10]. This observation underlines the need for ongoing efforts to implement and evaluate the incorporation of EHR data analytics-driven knowledge generation in LHSs.

EHRs are electronic systems used to collect and store medical information of patients longitudinally over time and to collect and store information relevant to managing clinical workflows. As such, EHR data can be of a diverse nature. These data can be used for evaluation in various ways to extract knowledge [11]. Although traditionally done via statistical analyses, recent advances in machine learning techniques and applications have allowed the development and deployment of integrative algorithms that relate health care outcomes to multiple diverse sources of information present in EHRs. These algorithms can analyze large volumes of data to identify patterns and correlations and perform predictions. Approaches based on patient similarity are a typical recent example of this. A patient similarity approach tries to derive knowledge that is relevant to a given patient of interest who is presenting to the health care professional by analyzing information that is pertinent to clinically similar patients identified by a machine learning algorithm.

This study focuses on knowledge generation from EHRs in the Dutch primary health care setting based on a patient similarity approach. From a data analytics perspective, it seems appropriate to start exploring such an approach for a disease with substantial prevalence and incidence, that is, large volumes of data present in EHRs.

In 2013, approximately half of the Dutch population reported having at least 1 chronic disorder. One of the most common chronic disorders with a high disease burden is diabetes mellitus (DM). In 2021, approximately 4.9% of the Dutch population reported to have diabetes, of whom 90% (4.2%) had type 2 DM (T2DM) and the remaining had type 1 DM [12]. Therefore, T2DM was used as an example disease to explore the feasibility of an LHS approach in the Dutch primary care setting from a data analytical viewpoint.

In T2DM, a diminished insulin response in combination with insulin resistance results in hyperglycemia. Although T2DM is more common in participants aged ≥45 years, the numbers are increasing for younger individuals owing to a rise in obesity, sedentary lifestyle, and the intake of energy-dense diets [13]. Diabetes can be diagnosed and glycemic control can be monitored by measuring the glycated hemoglobin (HbA1c) levels or measuring the plasma glucose concentration. Diagnosis thresholds for HbA1c and plasma glucose concentration are >7% (53 mmol/mol) and >126 mg/dL (7 mmol/L), respectively [14]. Testing for HbA1c is convenient, fast, and standardized, but it is more costly and comes with a lower sensitivity than testing for plasma glucose. As a result, it has become a standard practice to use more frequent glucose measurements for regular monitoring and HbA1c measurements only at longer intervals or in special cases to assess the disease state and judge the necessity for treatment change. The initial steps in treating and managing T2DM involve lifestyle modifications, such as adopting a healthy diet, engaging in regular exercise, and quitting smoking. In the second step, when lifestyle modifications fail to achieve an adequate glycemic level, an antidiabetic medication is administered following national care standards. In the Netherlands, the first line of medication is metformin for non–high-risk patients and sodium-glucose transport protein 2 (SGLT-2) inhibitors for high-risk patients. There are various follow-up therapies, such as sulfonylureas, dipeptidyl peptidase 4 inhibitors, glucagon-like peptide-1 receptor agonists, SGLT-2 inhibitors, α-glucosidase inhibitors, and insulin [15].

In the Netherlands, general practitioners (GPs) are often the first point of contact for patients diagnosed with T2DM and act as gatekeepers to secondary care. Therefore, they play an important role in the diagnosis and treatment of these patients. To make treatment choices for individual patients with diabetes, physicians consider treatment guidelines and their own knowledge and experience, also—implicitly or explicitly—on the basis of research.
explicitly—considering the patient’s perspective. Following the LHS approach, the decision-making may be supported by the on-demand availability of more objective information based on larger groups of comparable patients, such that the physician can see the actual data on improvements obtained after changing treatment in similar patients. This requires integration and sharing of data between physicians and health institutes. However, in practice, in Dutch primary health care settings, medical data sharing is still mostly limited to local or small regional settings, thus hampering the implementation of an LHS.

The Netherlands Institute for Health Services Research (Nivel) hosts the Nivel Primary Care Database (PCD), in which routinely recorded data from EHRs from primary health care providers are collected and used to monitor health and use of health services in a representative sample of the Dutch population. Therefore, the Nivel-PCD was an appropriate database to study data analytical aspects for an LHS approach for patients with T2DM in the Dutch primary care setting.

This study aims to explore the applicability of a recently developed precision cohort analytics approach based on EHRs [16] as a candidate data analytical decision support tool focusing on data analytical aspects.

**Methods**

**Overview**

Our approach resides in patient matching and uses the precision cohort analytics methods developed by Ng et al [16]. In brief, there are different studies focused on building and applying matching methods. There is also evidence that patient similarity–based modeling outperforms population-based predictive methods [17]. Methods that can learn a disease-specific similarity metric by developing a locally supervised metric learning are very valuable to identify clinically similar patients. Similarity-based modeling algorithms can make use of different sources of data formats: textual data, numerical measurements, recorded signals, images, and vital signs. The algorithms used commonly are neighborhood-based algorithms, distance-based similarity metrics, correlation-based similarity metrics, cosine-similarity metrics, and cluster-based algorithms [17]. A clinical decision support (CDS) system that is intended to improve health care by improving medical decisions with clinical knowledge, patient information, and other health information intelligently filtered or presented at appropriate times [18] can for example include the following:

1. Mapping of clinical data from a patient to a specific point in a clinical pathway to advise treatment
2. Prediction of an individual’s responsiveness to different treatments based on the respective gene expression profiles
3. Identifying a patient as a candidate for a specific treatment based on a set of clinical characteristics with an associated desired treatment
4. Generating a patient trajectory graph from clinical data, capturing conditions, outcomes, interventions, and suggestions from medical guidelines at different patient group levels

The precision cohort treatment options approach used in this work is based on the abovementioned CDS approaches with the following adjustments:

1. Identification and extraction of relevant clinical treatment decision points (DPs) from the longitudinal patient data to use as events of interest for modeling and analysis
2. Selection and generation of features to determine patient similarity from different sources of information, such as guidelines, clinical measurements, prescriptions, consultations, and comorbidities
3. For a given patient of interest, creation of a precision cohort of patient events that are similar to the given clinical state
4. Demonstration of the available treatment options and statistics based on a retrospective analysis of the generated precision cohort

When a new patient presents with a need for a treatment decision, the approach can be used to create a precision cohort of the most clinically similar participants in the database and provide statistics on the past outcomes of different treatment decisions taken for these patients. This information can be used to support the health professional in their treatment choice. In this study, the published precision cohort analytics approach was adapted to the guidelines and EHR characteristics of the Dutch care setting. Results are compared with those of Ng et al [16], and their relevance for the Dutch setting is discussed. The outcomes may provide a further stimulus to ongoing initiatives to establish primary care medical data sharing at the national level in the Netherlands.

The major steps of this workflow are (1) EHR data extraction and preprocessing, (2) DPs identification and extraction, (3) patient similarity model training, (4) precision cohort identification, and (5) treatment options and outcomes analysis.

**EHR Data Extraction and Preprocessing**

The data were extracted from the Nivel-PCD [19]. Using an algorithm developed in prior research within Nivel-PCD [20], a total of 11,490 registered patients were identified as diagnosed with T2DM between 2012 and 2014, having at least 6 months of prior history in the database. The follow-up period was limited to a period of 5 years. Patients with incomplete data in this 5-year period were also included. The identification of patients with morbidities was done using an algorithm developed by Nielien et al [21] to construct episodes of illness based on routinely recorded EHRs. For the data extraction, only GP practices that permitted to use the data for scientific research were selected. Patient ID and GP practice ID are unique for practices that permitted to use the data for scientific research.

The data were extracted from the Nivel-PCD [19]. Using an algorithm developed in prior research within Nivel-PCD [20], a total of 11,490 registered patients were identified as diagnosed with T2DM between 2012 and 2014, having at least 6 months of prior history in the database. The follow-up period was limited to a period of 5 years. Patients with incomplete data in this 5-year period were also included. The identification of patients with morbidities was done using an algorithm developed by Nielien et al [21] to construct episodes of illness based on routinely recorded EHRs. For the data extraction, only GP practices that permitted to use the data for scientific research were selected. Patient ID and GP practice ID are unique for these data and cannot be linked to other data sets to reduce the risk of tracking individual patients. For the selected 11,490 patients with T2DM, several tables were provided containing clinical information on 625,641 prescriptions (date, International Classification of Primary Care [ICPC] code, and ICPC description); 402,602 consultations (date and [Dutch] CTG code); 3,360,555 measurements (date, [Dutch GP association] Netherlands Huisartsen Genootschap code, and result or value); and 228,810 comorbidities (ICPC code and start and end date [based on the episodes of illness construct mentioned above] and type of comorbidity episode: 4 weeks, 8 weeks, 16 weeks, long-lasting, or chronic [21]).
To allow for a more explicit interpretation and analysis of the data, all clinical codes were replaced by the respective descriptions (Textbox 1), and dates were rewritten to a standard format, across all the tables.

Textbox 1. List of clinical codes and source of descriptions used.

<table>
<thead>
<tr>
<th>Clinical measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Netherlands Huisartsen Genootschap [22]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prescription</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Anatomical Therapeutic Chemical Classification System [23]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comorbidity code type</th>
</tr>
</thead>
<tbody>
<tr>
<td>• International Classification of Primary Care (International Classification of Primary Care codes were processed using a table provided by Nivel)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consultation code type</th>
</tr>
</thead>
<tbody>
<tr>
<td>• CTG [24]</td>
</tr>
</tbody>
</table>

Decision Points

Identification and Extraction

From the several tables, points in time were extracted to serve as events of interest for the analysis and the modeling processes. These points are named DPs, and they are defined as points in time from the longitudinal data, of each patient, after the diagnosis date, where the disease is considered as being not under control either because HbA1c>7% or fasting glucose>7 mmol/L. These DPs thus represent opportunities to initiate a change in the treatment plan. This point must have another matching HbA1c or fasting glucose test in the follow-up period that classifies the outcome of the treatment decision, either as not under control (an HbA1c test>7% or a fasting glucose test>7 mmol/L) or as under control (an HbA1c test<7% or a fasting glucose test<7 mmol/L). Over time during longitudinal follow-up, a patient can have multiple DPs as long as the abovementioned criteria are fulfilled. A DP is composed of the following:

1. An index date:
   - Date of an HbA1c or fasting glucose test that indicates a disease-uncontrolled situation along with the new treatment decision taken at this point.

2. A baseline period preceding the DP index date that represents the disease condition and the active treatment status, featuring the following:
   - The treatment that was in effect
   - The applicable clinical guidelines
   - Data characterizing the condition of the patient
   - All available clinical history until the DP index date. In case of multiple clinical history available for the same field, the most recent was taken.

3. An observation period that follows the DP date, featuring the following:
   - The new treatment
   - The treatment outcome: either disease controlled or disease uncontrolled, during a period of 90 to 365 days after the index date

Target Outcome Variables

As described in the Introduction section, alternatively to using HbA1c International Federation of Clinical Chemistry and Laboratory Medicine as a target outcome variable, in this study, we also used fasting glucose, venous (laboratory) as a target outcome variable following the same procedure to extract DPs, but with a threshold of 7 mmol/L (Table 1). As mentioned in the Introduction section, these 2 clinical measurements are the most widely used tests for diagnosing T2DM. Although we have no information for both these metrics for all the patients, having 2 possible outcome judgment options allowed us to make wider use of the Nivel-PCD data available for each participant.

Table 1. Percentage of patients with measurements and those without and total number of decision points (DPs) that are controlled and uncontrolled for both glycated hemoglobin (HbA1c) and fasting glucose as target outcome variables (n=11,490).

<table>
<thead>
<tr>
<th>Target outcome variable</th>
<th>HbA1c</th>
<th>Fasting glucose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients with measurements, n (%)</td>
<td>10,375 (90.3)</td>
<td>10,824 (94.2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of DPs with outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controlled</td>
</tr>
<tr>
<td>9,683</td>
</tr>
<tr>
<td>8,522</td>
</tr>
</tbody>
</table>

| Patients without any measurements, n (%) | 1115 (9.7) | 666 (5.8) |
Treatments Considered

Many different prescriptions occur in the Nivel-PCD data set. To reduce the complexity of the analysis, we considered only medication-based treatments that specifically targeted T2DM (pharmacologic treatments), and we merged 3 different forms of healthy lifestyle advice encountered in the records (ie, “follow dietary advice,” “advice healthy food given,” and “advice physical activity given”) into a single nonpharmacological treatment, henceforth called “healthy lifestyle.” To select medications specifically targeting T2DM, treatment options present in the Nivel-PCD were compared against the Pharmaceutical Compass [25], containing independent pharmaceutical information for medical professionals, published by the Dutch National Healthcare Institute. Textbox 2 shows the resulting list of individual medications targeting DM that were considered in this study. Treatment options that combine multiple medications were also considered, for example, metformin and gliclazide (denoted as metformin_gliclazide).

As the dosage information was not available in the data sample used for this study, changes in drug dose were not captured and were interpreted as “no change” treatments.

For each DP, we assigned an active treatment (baseline period) and a new treatment (observation period). For assessing the active treatment of each DP, the medical history of the patient’s measurements was queried, and we applied the following reasoning: if neither pharmacologic nor lifestyle advice was found, then the DP was given a “no treatment” type of active treatment; if both pharmacologic and nonpharmacologic treatments were available, the 2 were merged (eg, metformin+healthy lifestyle). For assessing the new treatment, the same reasoning was applied; however, if the new treatment was the same as the active treatment, it was denoted as “no change.”

Textbox 2. List of medications targeting type 2 diabetes mellitus considered in this study.

<table>
<thead>
<tr>
<th>Treatment options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metformin</td>
</tr>
<tr>
<td>Sitagliptin</td>
</tr>
<tr>
<td>Insulin aspart</td>
</tr>
<tr>
<td>Insulin degludec</td>
</tr>
<tr>
<td>Insulin detemir</td>
</tr>
<tr>
<td>Insulin glargine</td>
</tr>
<tr>
<td>Insulin (human)</td>
</tr>
<tr>
<td>Repaglinide</td>
</tr>
<tr>
<td>Glimepiride</td>
</tr>
<tr>
<td>Tolbutamide</td>
</tr>
<tr>
<td>Gliclazide</td>
</tr>
</tbody>
</table>

Guidelines

Treatment decisions were made by the physicians based on clinical guidelines together with their personal past experience and considering the history and condition of the patient. The aim of using the guidelines is to improve the appropriateness of medical practice by leading to a better patient outcome while reducing costs, to aid authorities in deciding on the approval of drugs and devices, and to identify areas that need further research [26]. The guidelines for T2DM, published by the Dutch College of General Practitioners, including recommendations for the diagnosis, treatment, and management of patients, were incorporated in this study [27]. The relevant criteria for this study were derived from the guidelines to recommend medications and confirmed in a discussion with a GP:

1. Aged >70 years
2. Disease duration >10 years
3. BMI <25 kg/m²

Patient Condition

The clinical condition of patients was assessed using clinical measurements (which included data on, eg, BMI, diastolic blood pressure, and low-density lipoprotein), comorbidities, and consultation codes (Textbox 1). In addition, 2 patient condition criteria (mobility and mental state) were added based on the discussion with a GP, who emphasized that these are crucial aspects to consider when deciding which treatment is most appropriate.

Patients with mobility limitations are unlikely to perform physical exercise; therefore, even adhering to a restrictive healthy diet may not sufficiently control diabetes, necessitating medication sooner compared to patients without impaired mobility.

To assess the mobility state of the patient, we looked at comorbidities and measurements that indicated any possible obstacle to the ability to move. The list of comorbidities and measurements considered is represented in Table 2. We discriminated between the chronic and temporary duration of each comorbidity as used in the episodes of illness construct...
It has to be emphasized that the operationalization of reduced patient mobility as evident from Table 2 is a highly subjective choice made by the authors based on their intuitive understanding. It serves only for initial exploration of the feasibility of introducing additional GP considerations beyond the standard clinical guidelines.

Similarly, the mental state of the patient is also very important when deciding the appropriate treatment option. Patients who are going through events that alter their emotional state, affecting their concentration, positiveness, or willingness to adhere to treatment, may need a stricter treatment regime.

To assess the emotional state of the patient, we made an equally subjective choice of the different comorbidities that might be affecting their emotional state (Table 3).

Table 2. List of comorbidity and measurement types considered to assess the mobility state of patients. The upper part of the table shows the comorbidity International Classification of Primary Care (ICPC) code, description, and type of duration. At the bottom, it shows additional variables that were considered relevant to reflect mobility impairments, that is, the measurement type, Netherlands Huisartsen Genootschap (NHG) code description, and the duration.

<table>
<thead>
<tr>
<th>Code</th>
<th>Code description</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICPC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L15</td>
<td>Knee symptoms or complaints</td>
<td>Temporary</td>
</tr>
<tr>
<td>L95</td>
<td>Osteoporosis</td>
<td>Chronic</td>
</tr>
<tr>
<td>L90</td>
<td>Osteoarthritis of the knee</td>
<td>Chronic</td>
</tr>
<tr>
<td>N17</td>
<td>Vertigo or dizziness (excluding H82)</td>
<td>Chronic</td>
</tr>
<tr>
<td>L14</td>
<td>Leg or thigh symptoms or complaints</td>
<td>Chronic</td>
</tr>
<tr>
<td>K90</td>
<td>Stroke or cerebrovascular accident</td>
<td>Chronic</td>
</tr>
<tr>
<td>L73</td>
<td>Fracture: tibia or fibula</td>
<td>Chronic</td>
</tr>
<tr>
<td>L03</td>
<td>Low back symptoms or complaints without radiation (excluding L86)</td>
<td>Chronic</td>
</tr>
<tr>
<td>L89</td>
<td>Osteoarthritis of the hip</td>
<td>Chronic</td>
</tr>
<tr>
<td>L76</td>
<td>Fracture: other</td>
<td>Temporary</td>
</tr>
<tr>
<td>R96</td>
<td>Asthma</td>
<td>Chronic</td>
</tr>
<tr>
<td>L79</td>
<td>Sprain or strain of other joints</td>
<td>Temporary</td>
</tr>
<tr>
<td>L74</td>
<td>Fracture: hand or foot bone</td>
<td>Temporary</td>
</tr>
<tr>
<td>L16</td>
<td>Ankle symptoms or complaints</td>
<td>Temporary</td>
</tr>
<tr>
<td>L97</td>
<td>Chronic internal knee derangement</td>
<td>Chronic</td>
</tr>
<tr>
<td>N18</td>
<td>Paralysis or weakness (excluding A04)</td>
<td>Chronic</td>
</tr>
<tr>
<td>L84</td>
<td>Osteoarthritis of the spine</td>
<td>Chronic</td>
</tr>
<tr>
<td>L77</td>
<td>Sprain or strain of the ankle</td>
<td>Temporary</td>
</tr>
<tr>
<td>L75</td>
<td>Fracture: femur</td>
<td>Temporary</td>
</tr>
<tr>
<td>L70</td>
<td>Infections musculoskeletal system</td>
<td>Temporary</td>
</tr>
<tr>
<td>L78</td>
<td>Sprain or strain of the knee</td>
<td>Temporary</td>
</tr>
<tr>
<td>NHG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K93</td>
<td>Left foot amputation</td>
<td>Chronic</td>
</tr>
<tr>
<td>K94</td>
<td>Right foot amputation</td>
<td>Chronic</td>
</tr>
</tbody>
</table>
Table 3. List of comorbidity types considered to assess the emotional state of patients. The table shows the comorbidity International Classification of Primary Care (ICPC) code, description, and type of duration.

<table>
<thead>
<tr>
<th>ICPC code</th>
<th>Description</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>P01</td>
<td>Feeling anxious, nervous, tense, or inadequate</td>
<td>Temporary</td>
</tr>
<tr>
<td>P02</td>
<td>Acute stress or transient situational disturbance</td>
<td>Temporary</td>
</tr>
<tr>
<td>P20</td>
<td>Disturbances of memory, concentration, or orientation</td>
<td>Chronic</td>
</tr>
<tr>
<td>P74</td>
<td>Anxiety disorder or anxiety state</td>
<td>Chronic</td>
</tr>
<tr>
<td>Z18</td>
<td>Illness problem with a child</td>
<td>Temporary</td>
</tr>
<tr>
<td>Z19</td>
<td>Loss or death of a child</td>
<td>Chronic</td>
</tr>
<tr>
<td>P06</td>
<td>Disturbances of sleep or insomnia</td>
<td>Temporary</td>
</tr>
<tr>
<td>Z15</td>
<td>Loss or death of a partner</td>
<td>Temporary</td>
</tr>
<tr>
<td>A80</td>
<td>Accident or injury NOS(^a)</td>
<td>Temporary</td>
</tr>
<tr>
<td>P03</td>
<td>Feeling depressed</td>
<td>Temporary</td>
</tr>
<tr>
<td>P99</td>
<td>Other mental or psychological disorder</td>
<td>Chronic</td>
</tr>
<tr>
<td>P72</td>
<td>Schizophrenia</td>
<td>Chronic</td>
</tr>
<tr>
<td>P76</td>
<td>Depressive disorder</td>
<td>Chronic</td>
</tr>
<tr>
<td>P73</td>
<td>Affective psychosis</td>
<td>Chronic</td>
</tr>
<tr>
<td>Z25</td>
<td>Problems resulting from assaults or harmful events</td>
<td>Temporary</td>
</tr>
<tr>
<td>P04</td>
<td>Feeling or behaving irritable or angry</td>
<td>Temporary</td>
</tr>
<tr>
<td>P77</td>
<td>Suicide attempt</td>
<td>Chronic</td>
</tr>
<tr>
<td>P98</td>
<td>Other or unspecified psychoses</td>
<td>Chronic</td>
</tr>
<tr>
<td>P70</td>
<td>Dementia (including senile and Alzheimer)</td>
<td>Chronic</td>
</tr>
</tbody>
</table>

\(^a\)NOS: not otherwise specified.

Patient Similarity Modeling

Feature Selection

Patient similarity was evaluated firstly based on the characteristics captured with the DP, that is, active treatment, applicable guidelines, and patient condition (including clinical history). In addition, to enrich the patient’s clinical information, more features were engineered. For instance, a Boolean feature named has_chronic_comorbidity was created to represent if the patient had or did not have a chronic comorbidity recorded at the time of interest in the medical history. In addition, the number of prescriptions, number of clinical measurements, number of consultations, and number of comorbidities from the disease diagnosis date until the DP date were added as candidates to the similarity features set.

As some algorithms cannot work directly with categorical data, one-hot encoding was applied to the variable new_treatment, converting this single column into N new columns, with N being the total number of different treatments (including combinations) observed in the data set. The same was done for the active_treatment variable.

The final data frame of DPs extracted and processed from the Nivel data set was composed of multiple feature types: identifiers, dates, and categorical and numerical data. The resulting DPs data frame had a large amount of missing data. As a first selection step, only features that had at least 80% nonmissing values were retained [28]. Thereafter, the remaining missing values were imputed using the k-nearest neighbors (KNN)-based KNNImputer method from the scikit-learn library [29], with the following parameters: 2 nearest neighbors and a uniform weighting of all points in the neighborhood. The result of this step was a data frame of DPs without any missing values.

A further selection of the most salient features associated with disease control was done using the approach by Ng et al [16]. A total of 200 different L1-regularized logistic regression models (also known as the least absolute shrinkage and selection operator [LASSO] models) for predicting disease outcomes were created. LASSO is a linear regression technique that incorporates a penalty term to the sum of squared errors to shrink the coefficients toward 0 and perform feature selection. The penalty term is determined by the \( \alpha \) parameter. Each model used a randomly selected subset (75%) of the data. The features selected by at least 150 of the 200 models were considered stable, and the remaining features were discarded. To find the best \( \alpha \) parameter for the LASSO model, a grid search approach was applied. Different values of \( \alpha \) were tested (100, 20, 10, 2, 1.67, 1.43, 1.25, 1.11, and 1), and the value corresponding to the highest \( F_1 \)-score was selected.
As the 2 target outcome variables were unbalanced, we used the $F_1$ score [30], rather than accuracy, as a metric of prediction performance.

It is important to note that 2 different data frames were built, and separate similarity models were built, for the 2 outcome variables HbA1c and fasting glucose. Each of the data frames was split into 2 sets: a training set that was used to train the patient similarity model and a scoring set used as a repository of clinical events and to create multiple precision cohorts.

**Similarity Model Training**

The similarity model used the set of stable features obtained from the previous steps to calculate patient similarity. This model learns a T2DM distance measure that is a modified version of the Mahalanobis distance (MD). The MD measures the distance relative to a centroid or central point, in which all means from all variables intersect; the larger the MD, the further away from the centroid the data point is. The MD can also be used to calculate the distance between 2 points, or in this case, 2 patients ($x_i$ and $x_j$) using the covariance matrix. Here, a modified version of the MD formula was used, with the covariance matrix replaced by a weight matrix W, which is learned from the training data.

$$
(1)
$$

The similarity model sets the weight matrix to maximize the target class discriminability (disease control state) by adjusting the weights of every feature. This was done by locally separating points from different classes while keeping together points that belong to the same class, using a large margin nearest neighbor (LMNN) [31] machine learning algorithm.

The number of points to consider for the calculations is defined by the number of neighbours parameter (K). Different values of K were tested: 2, 3, 4, 5, 6, and 7.

To assess the effect of the LMNN algorithm, we compared the performance of a KNN classification model [29] on the raw data and the data transformed by the similarity model [32]. For the KNN algorithm, a grid search approach was used to find the best set of parameters:

1. Number of neighbors (N): 3, 4, 5, ..., or 30
2. Weight function: uniform (all points in each neighborhood weighted equally) or distance (closer neighbors of a query point will have a greater influence than neighbors that are further away).

The learned similarity weights for each variable for T2DM separately for HbA1c and fasting glucose as target outcome variables are shown in Multimedia Appendices 1 and 2.

**Precision Cohort Construction**

Precision cohort construction consists of selecting the most clinically similar patient DPs based on the characteristics of the patient of interest at the time of the consultation. The process of generating precision cohorts needs to ensure that the baseline confounders are adjusted so that the effect analysis is valid (ie, a good covariate balance is achieved). The selection was done in a 2-step procedure: a filtering step and a similarity rating.

To filter the patients who are more similar to the patient of interest, filter variables were generated. These filter variables were composed of guidelines plus the active treatment. For instance, if the patient was aged 80 years and was currently taking metformin, the filter variable was age_above_70y+metformin. By using this filter variable, it was ensured that only patients aged >70 years and who were taking metformin were selected for the precision cohort. It is important to mention that only data from the “baseline period” of the DP were used for the filtering process, which means that the treatment decision at the time of the DP (index date) was not considered, as we want to analyze the entire pool of new treatments in the precision cohort. Combining the set of guidelines with the set of active treatments resulted in a very large number of filter cohorts of widely varying sizes. Considering that it is difficult to recommend treatment options for patients with clinically odd profiles as this method requires a large patient pool to recommend statistically significant treatment options, we focused only on the most representative cohorts, that is, those with >200 DPs.

For the second step (similarity rating), the similarity model explained in the Similarity Modeling section was used to calculate the similarity scores for the filtered DPs. The similarity score is a distance metric; thus, smaller scores indicate a higher similarity between a patient’s DP and the DP of the patient of interest. The similarity score was converted to a normalized distance using a minimum-maximum normalization method to allow an easier interpretation of these scores.

Thereafter, the final precision cohort was generated by retaining only the “most similar” patients. However, reducing the cohort size may compromise the cofactor balance. Therefore, the covariate balance of the precision cohort with varying sizes was calculated to assess bias and matching validity by comparing the “no treatment change” (new treatment is the same as the active treatment) with the “treatment change” (new treatment different from the active treatment) groups. Covariate balance was calculated as the difference in the means of each covariate between the 2 groups divided by the SD of the treated group [33]. The closer this value was to 0, the better balance we had between the groups.

To have a trade-off between the covariate balance value and the number of DPs in the precision cohort, a normalized distance value of 2 was defined as a cutoff. This value was chosen after visual inspection of several covariate balance plots for different precision cohorts. Some studies agreed that covariate balance values <0.1 were satisfactory [34], although another study suggested that a value of 0.25 was good enough [35]. With a normalized distance cutoff value of 2, a covariate balance of 0.1 was achieved for most of the precision cohorts.

In summary, the precision cohort construction process involved the following:

1. Filtering DPs with the same filter variables as the patient of interest
2. Calculating the similarity scores for the filtered DPs
3. Ranking DPs based on similarity scores (normalized distance)
4. Retaining only the DPs with a normalized distance <2

**Treatment Options and Outcomes Analysis**

Treatment outcomes analyses were performed both from a global perspective (ie, across all filter cohorts) and a personalized perspective (ie, in the precision cohorts). The latter can be used as a retrospective analysis to generate personalized treatment options for a given patient of interest.

The set of DPs in the precision cohort was grouped by treatment decision. For each of these treatment groups, we computed the following:

1. The number of DPs
2. The percentage of DPs that have a controlled outcome
3. The difference in the respective outcome compared with the “no treatment” change option
4. A statistical significance assessment using a Bonferroni corrected $\chi^2$ $P$ value of 0.5 to adjust for multiple comparisons

To better visualize the difference between the different treatments for each precision cohort, the results were presented in a Sankey diagram [36].

To reduce the number of treatment options with a low number of DPs, we decided to only include those with at least 1% of the total DPs in the precision cohort, with a minimum of 10 DPs. For instance, if the precision cohort had 2000 DPs, we only included treatment options with at least 20 DPs.

**Ethical Considerations**

This study was approved according to the governance code of Nivel-PCD (NZR-00320.048). The use of EHRs for research purposes is allowed under certain conditions. When these conditions are fulfilled, neither obtaining informed consent from patients nor approval by a medical ethics committee are obligatory for this type of observational study containing no directly identifiable data (Art. 24 General Data Protection Regulation Implementation Act jo art. 9.2 sub j General Data Protection Regulation).

**Results**

Separate analyses were conducted for HbA1c and fasting glucose as target outcome variables.

**HbA1c Outcome Scenario**

**Decision Points**

For the HbA1c scenario, we found 17,328 DPs across the available longitudinal data from the 11,490 patients with T2DM. Although seemingly large, this was still 10-fold lower compared with previous work by Ng et al [16], who retrieved >171,000 DPs for 24,373 patients with T2DM from their data set, for HbA1c as an outcome.

The processed HbA1c set was split into 5199 DPs for the training set and 12,129 DPs for the scoring set. As the class (disease control) was unbalanced, we had to ensure that we had the same proportion of each class for both training and scoring sets. For this case, of the 17,328 DPs, we had 9704 (56%) uncontrolled DPs and 7624 (44%) controlled DPs. This same proportion was maintained for the training set: there were 2905 uncontrolled DPs and 2294 controlled DPs in the training set and 6793 uncontrolled DPs and 5336 controlled DPs in the scoring set.

**Patient Similarity Modeling**

The methods of feature generation, missing data imputation, and feature selection explained in the Methods section were applied to this subset. The LASSO model $\alpha$ value used was 1.43, with an $F_1$-score of 0.613.

Next, the tuned version of the KNN algorithm was used to evaluate the LMNN algorithm performance (refer to the Methods section). The $F_1$-score was 0.606 for raw data versus 0.613 for transformed data. Thus, indeed, the LMNN algorithm resulted in improved classification performance.

The learned similarity weights for each variable for HbA1c as the target outcome variable are shown in Multimedia Appendix 1. These weights are disease specific for T2DM. A total of 26 features were retained for the similarity model. Interestingly, there was rather limited variation in size: similarity weights were all of comparable value (typically ranging between 0.4 and 0.6), except for features “#comorbidities” and “Systolic BP.”

**Precision Cohort Construction**

The largest observed filter cohorts for HbA1c are shown in Figure 1. These 25 cohorts covered approximately 75% of all the DPs in the scoring set. Of these 25 cohorts, 20 (80%) had >100 DPs. Only 10 cohorts had >200 DPs, considered potentially useful for constructing precision cohorts. As expected, the cohorts containing metformin were the largest ones as it is the first line of medication for the treatment of T2DM. The cohort “healthy lifestyle_metformin” was the largest cohort for the HbA1c outcome scenario. The cohort “all_guidelines_variables_false” also contained a large number of DPs. In this cohort of patients, who were aged ≤70 years, were not mobility or mentally impaired, and had a BMI of ≥25 kg/m², no antidiabetic treatment (medication or healthy lifestyle advice) was administered or the information was not registered. Furthermore, of the 25 largest cohorts, 17 (68%) had “healthy lifestyle,” 16 (64%) had “metformin,” 11 (44%) had “mobility impaired,” 7 (28%) had “age_above_70y,” and 3 (12%) had “mental impaired.” No other pharmacological treatments than metformin and gliclazide (6 occurrences) were represented in the 25 largest cohorts.

Following the procedures explained in the Precision Cohort Construction section, precision cohorts were generated for a...
number of randomly chosen patients from various filter cohorts. Figure 2 shows a covariate balance plot for 1 randomly chosen patient in the metformin filter cohort as an example. As can be appreciated from the figure, the best cofactor balance was achieved for normalized distance 2.0.

Figure 1. The 25 largest cohorts for glycated hemoglobin (HbA1c) as a target outcome variable based on the filter variables. The blue bars represent the number of decision points (DPs) on a logarithmic scale (left vertical axis), and the orange line shows the cumulative coverage on a linear percentage scale (right vertical axis).

Figure 2. Illustration of the metformin precision cohort generation for the glycated hemoglobin (HbA1c) outcome target. The decision points (DPs) in the subset were grouped by normalized distance (similarity score) to the patient of interest. The blue bars represent the number of DPs per bar or grouped DPs (left vertical axis), and the red dotted line shows the cumulative covariate balance values for the different bars or grouped DPs (right vertical axis).

Treatment Options and Outcomes Analysis

Table 4 shows the overall percent controlled of the 20 most representative treatment options on a global scale, that is, across the cohorts based on the filter variables. In >47% of cases, the GP decided to continue the current treatment despite the HbA1c levels being classified as uncontrolled. Despite no change, >44% of patients had their HbA1c levels subsequently “controlled” during the follow-up. Some alternatives, however, had much better outcomes. For instance, the option of stopping the “no treatment” option and starting taking metformin, that is, the first line of medication treatment according to the guidelines, resulted in 57.6% (213/370) of DPs with HbA1c controlled, whereas starting treatment with healthy lifestyle advice resulted in >64%
controlled outcomes. Despite this difference, starting treatment with metformin was chosen almost 4 times more often than starting with healthy lifestyle treatment (370 vs 98 cases). The combination of metformin and lifestyle advice was beneficial over continuing each of these treatments as a single treatment. This list is relevant for analyzing the global picture of the available treatment options and the corresponding outcomes. However, for individual patients, the treatment option analysis might differ from the overall picture depending on the precision cohort that more closely reflects the clinical scenario for the particular patient. As an example of such a case-specific analysis, Figure 3 uses a Sankey diagram to represent the different treatment options for a given patient that belongs to the cohort “Metformin.”

In Table 4, it can be noticed that adding healthy lifestyle advice to the metformin prescription was a treatment option with a statistically significant better-associated outcome (210/324, 64.8%), whereas changing from metformin to only healthy lifestyle advice alone led to a worse disease outcome in the precision cohort for this particular patient. Changing the treatment to gliclazide or tolbutamide, or adding gliclazide to metformin, also resulted in better outcomes; however, the differences were not statistically significant because of the low number of cases involved. We can also see that from the most similar patients who kept taking only metformin, 45.69% (605/1324) of the patients improved their disease condition. For the remaining 54.31% (719/1324) of the patients, the outcome was “uncontrolled” in the follow-up.

Table 4. The 20 largest observed global treatment option groups for glycated hemoglobin as the target outcome variable. The list shows the size of each group and the associated percentage controlled during follow-up (n=12,129).

<table>
<thead>
<tr>
<th>Treatment option</th>
<th>Frequency, n (%)</th>
<th>DPs&lt;sup&gt;a&lt;/sup&gt; controlled, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No change</td>
<td>5747 (47.38)</td>
<td>2541 (44.21)</td>
</tr>
<tr>
<td>Metformin_new+healthy lifestyle_metformin_stop&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1324 (10.92)</td>
<td>605 (45.69)</td>
</tr>
<tr>
<td>Healthy lifestyle_new+healthy lifestyle_metformin_stop&lt;sup&gt;c&lt;/sup&gt;</td>
<td>402 (3.31)</td>
<td>170 (42.29)</td>
</tr>
<tr>
<td>Metformin_new+no treatment_stop&lt;sup&gt;b&lt;/sup&gt;</td>
<td>370 (3.05)</td>
<td>213 (57.57)</td>
</tr>
<tr>
<td>Healthy lifestyle_metformin_new+metformin_stop&lt;sup&gt;b&lt;/sup&gt;</td>
<td>324 (2.67)</td>
<td>210 (64.81)</td>
</tr>
<tr>
<td>Healthy lifestyle_metformin_new+healthy lifestyle_stop&lt;sup&gt;b&lt;/sup&gt;</td>
<td>246 (2.03)</td>
<td>165 (67.07)</td>
</tr>
<tr>
<td>Healthy lifestyle_metformin_new+no treatment_stop&lt;sup&gt;b&lt;/sup&gt;</td>
<td>222 (1.83)</td>
<td>138 (62.16)</td>
</tr>
<tr>
<td>Gliclazide_new+healthy lifestyle_metformin_STOP&lt;sup&gt;c&lt;/sup&gt;</td>
<td>217 (1.79)</td>
<td>86 (39.63)</td>
</tr>
<tr>
<td>Metformin_new+healthy lifestyle_STOP&lt;sup&gt;b&lt;/sup&gt;</td>
<td>169 (1.39)</td>
<td>91 (53.85)</td>
</tr>
<tr>
<td>Gliclazide_new+healthy lifestyle_metformin_stop&lt;sup&gt;c&lt;/sup&gt;</td>
<td>151 (1.24)</td>
<td>65 (43.05)</td>
</tr>
<tr>
<td>Gliclazide_new+metformin_stop&lt;sup&gt;b&lt;/sup&gt;</td>
<td>148 (1.22)</td>
<td>80 (54.05)</td>
</tr>
<tr>
<td>Gliclazide_metformin_new+gliclazide_healthy lifestyle_metformin_stop&lt;sup&gt;c&lt;/sup&gt;</td>
<td>147 (1.21)</td>
<td>38 (25.85)</td>
</tr>
<tr>
<td>Gliclazide_new+gliclazide_healthy lifestyle_stop&lt;sup&gt;c&lt;/sup&gt;</td>
<td>123 (1.01)</td>
<td>41 (33.33)</td>
</tr>
<tr>
<td>Healthy lifestyle_new+no treatment_stop&lt;sup&gt;b&lt;/sup&gt;</td>
<td>98 (0.81)</td>
<td>63 (64.29)</td>
</tr>
<tr>
<td>Gliclazide_healthy lifestyle_metformin_new+healthy lifestyle_metformin_stop&lt;sup&gt;b&lt;/sup&gt;</td>
<td>68 (0.56)</td>
<td>32 (47.06)</td>
</tr>
<tr>
<td>Gliclazide_metformin_new+metformin_stop&lt;sup&gt;c&lt;/sup&gt;</td>
<td>59 (0.49)</td>
<td>30 (50.85)</td>
</tr>
<tr>
<td>Metformin_new+gliclazide_healthy lifestyle_stop&lt;sup&gt;b&lt;/sup&gt;</td>
<td>59 (0.49)</td>
<td>16 (27.12)</td>
</tr>
<tr>
<td>Gliclazide_healthy lifestyle_new+gliclazide_healthy lifestyle_metformin_stop&lt;sup&gt;c&lt;/sup&gt;</td>
<td>58 (0.48)</td>
<td>17 (29.31)</td>
</tr>
<tr>
<td>Healthy lifestyle_metformin_new+gliclazide_healthy lifestyle_stop&lt;sup&gt;c&lt;/sup&gt;</td>
<td>58 (0.48)</td>
<td>12 (20.69)</td>
</tr>
<tr>
<td>Healthy lifestyle_metformin_new+gliclazide_healthy lifestyle_metformin_stop&lt;sup&gt;c&lt;/sup&gt;</td>
<td>57 (0.47)</td>
<td>21 (36.84)</td>
</tr>
</tbody>
</table>

<sup>a</sup>DP: decision point.

<sup>b</sup>Treatment options with a controlled percentage higher than the “no change” treatment option.

<sup>c</sup>Treatment options with a lower percentage than the “no change” treatment option.
Fasting Glucose Scenario

Decision Points
For the fasting glucose target outcome variable, we found 41,014 DPs across the available longitudinal data, that is, approximately 2.5 times more than for the HbA1c scenario.

The processed data set was split into 12,304 DPs for the training set and 28,710 for the scoring set. Again, the disease control target outcome variable was not balanced; thus, the proportions were kept the same for both training and scoring sets. In this case, we had 79% (32,401/41,014) uncontrolled DPs and 21% (8613/41,014) controlled DPs. Accordingly, the same proportion was kept for both the training and the testing sets.

Patient Similarity Modeling
Similarly to the HbA1c scenario, the methods of feature generation, missing data imputation, and feature selection explained in the Methods section were applied for the fasting glucose scenario. The LASSO model α value used was 1.67, with an $F_1$-score of 0.691. The training set was used to train the similarity model. The optimal k value for the LMNN algorithm was 5. The KNN algorithm was applied to the raw data and the transformed data. The best combination of parameters was N=6 and weight function=distance, that is, the same as that for the HbA1c scenario (data not shown).

Next, the tuned version of the KNN algorithm was used to evaluate the LMNN algorithm performance. The $F_1$-score was 0.849 for raw data and 0.856 for transformed data. Thus, only a small improvement in classification performance was achieved by the LMNN algorithm.

The learned similarity weights for each variable for fasting glucose as target outcome variable are shown in Multimedia Appendix 2. These weights are disease specific for T2DM. A total of 48 features were retained for the similarity model, 22 more than that for HbA1c. In total, 25 features were of the “active _treatment_”-type, whereas for HbA1c, only 5 features were of this type. This indicates that the active treatment was much more predictive of disease outcome than in the HbA1c scenario. With approximately half of weight values between 0.4 and 0.65, and half between 0.7 and 1.15, there was much more variation in the size of weights compared with the HbA1c scenario. Similar to the HbA1c scenario, the feature “Systolic BP” had a low weight, probably because it had little discriminating power.

Precision Cohort Construction
In Figure 4, the 25 largest observed filter cohorts for fasting glucose are represented. Together, these covered approximately 80% of DPs in the data set. In line with the much larger number of DPs compared with the HbA1c scenario, 20 of the largest filter cohorts had >200 DPs. Contrary to the HbA1c scenario, here the largest cohort was the “all_guidelines_variables_false,” followed by the “healthy lifestyle_metformin cohort,” which is in accordance with what we expected as the fasting glucose values are used more frequently to diagnose T2DM and most patients do not start medication immediately. Furthermore, of the 25 largest cohorts, 17 (68%) had “healthy lifestyle,” 16 (64%) had “metformin,” 11 (44%) had “mobility impaired,” 7 (28%) had “age_above_70y,” and 3 (12%) had “mental impaired,” all identical to the HbA1c scenario. Moreover, 23 (92%) of the 25 largest cohorts for the fasting glucose scenario were also in the list of the 25 largest cohorts for the HbA1c scenario, and the ranking in size was similar (≤3 positions difference). No other pharmacological treatments than metformin and gliclazide (4 occurrences) were represented in the 25 largest cohorts.

Various precision cohorts for randomly selected patients were constructed to verify that the chosen normalized distance threshold of 2.0 indeed resulted in optimal cofactor balance overall (data not shown) for the fasting glucose outcome scenario as well.

https://ojphi.jmir.org/2024/1/e51092

Online J Public Health Inform 2024 | vol. 16 | e51092 | p.80
(page number not for citation purposes)
Figure 4. The 25 largest cohorts for fasting glucose as target outcome variable based on the filter variables. The blue bars represent the number of decision points (DPs) on a logarithmic scale (left vertical axis), and the orange line shows the cumulative coverage on a linear percentage scale (right vertical axis).

Treatment Options and Outcomes Analysis

Table 5 shows the overall controlled percentage of the 20 most representative treatment options, that is, across all filter cohorts. In >57.4% (16,480/28,709) of the cases, the existing treatment was continued, and in only 23.29% (3839/16,480) of the cases, this led to patients becoming “controlled” as judged by the fasting glucose value. Interestingly, in this scenario, the option of starting to take metformin and stopping the “no treatment” resulted in a lower success percentage when compared with the “no change” option. More differences are apparent when comparing with the HbA1c scenario; overall, the percentages of a “controlled” outcome seem more than 2-fold lower and never >37%, thereby seemingly indicating a much more pessimistic perspective.

Figure 5 represents the different treatment options observed in a precision cohort for a given patient that belongs to the cohort “Metformin with mobility impairments” using a Sankey diagram.

Analyzing the Sankey diagram in Figure 5, we can notice that for patients with mobility impairments who were taking metformin, only a small proportion of those who were kept on the same prescription met with an improved disease control outcome. The ones who changed from metformin to gliclazide showed an improvement, although the number of DPs for these 2 options was not very large, and the result was not statistically significant as a consequence. These results showed that the set of patients, similar to the patient of interest, had difficulties in having the fasting glucose controlled and might need more attention from the health care professionals. Many other cohorts could be included; we chose 2 examples that had a decent number of patients to show how this approach works in practice.
Table 5. The 20 largest observed global treatment option groups for fasting glucose as the target outcome variable. The list shows the size of each group and the associated percentage controlled during follow-up (n=28,710).

<table>
<thead>
<tr>
<th>Treatment option</th>
<th>Frequency, n (%)</th>
<th>DPs(^a ) controlled, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No change</td>
<td>16,480 (57.4%)</td>
<td>3839 (23.29%)</td>
</tr>
<tr>
<td>Metformin(<em>{-})new+healthy lifestyle(</em>{-})metformin(_{-})stop(^b)</td>
<td>3130 (10.9%)</td>
<td>586 (18.72%)</td>
</tr>
<tr>
<td>Healthy lifestyle(<em>{-})new+healthy lifestyle(</em>{-})metformin(_{-})stop(^b)</td>
<td>1011 (3.52%)</td>
<td>201 (19.88%)</td>
</tr>
<tr>
<td>Metformin(<em>{-})new+no treatment(</em>{-})stop(^b)</td>
<td>639 (2.23%)</td>
<td>110 (17.21%)</td>
</tr>
<tr>
<td>Healthy lifestyle(<em>{-})new+no treatment(</em>{-})stop(^c)</td>
<td>638 (2.22%)</td>
<td>203 (31.82%)</td>
</tr>
<tr>
<td>Healthy lifestyle(<em>{-})metformin(</em>{-})new+metformin(_{-})stop(^b)</td>
<td>612 (2.13%)</td>
<td>59 (9.64%)</td>
</tr>
<tr>
<td>Healthy lifestyle(<em>{-})metformin(</em>{-})new+healthy lifestyle(_{-})stop(^c)</td>
<td>429 (1.49%)</td>
<td>157 (36.6%)</td>
</tr>
<tr>
<td>Healthy lifestyle(<em>{-})metformin(</em>{-})new+no treatment(_{-})stop(^b)</td>
<td>399 (1.39%)</td>
<td>69 (17.29%)</td>
</tr>
<tr>
<td>Gliclazide(<em>{-})metformin(</em>{-})new+gliclazide(<em>{-})healthy lifestyle(</em>{-})metformin(_{-})stop(^b)</td>
<td>284 (0.99%)</td>
<td>28 (9.86%)</td>
</tr>
<tr>
<td>Gliclazide(<em>{-})new+healthy lifestyle(</em>{-})stop(^c)</td>
<td>265 (0.92%)</td>
<td>97 (36.6%)</td>
</tr>
<tr>
<td>Gliclazide(<em>{-})new+healthy lifestyle(</em>{-})metformin(_{-})stop(^b)</td>
<td>259 (0.9%)</td>
<td>32 (12.36%)</td>
</tr>
<tr>
<td>Gliclazide(<em>{-})new+gliclazide(</em>{-})healthy lifestyle(_{-})stop(^b)</td>
<td>223 (0.78%)</td>
<td>39 (17.49%)</td>
</tr>
<tr>
<td>Gliclazide(<em>{-})new+healthy lifestyle(</em>{-})metformin(_{-})stop(^c)</td>
<td>204 (0.71%)</td>
<td>54 (26.47%)</td>
</tr>
<tr>
<td>Gliclazide(<em>{-})new+metformin(</em>{-})stop(^b)</td>
<td>165 (0.57%)</td>
<td>22 (13.33%)</td>
</tr>
<tr>
<td>Gliclazide(<em>{-})new+healthy lifestyle(</em>{-})metformin(<em>{-})new+healthy lifestyle(</em>{-})metformin(_{-})stop(^b)</td>
<td>105 (0.37%)</td>
<td>5 (4.76%)</td>
</tr>
<tr>
<td>Healthy lifestyle(<em>{-})new+metformin(</em>{-})stop(^b)</td>
<td>102 (0.36%)</td>
<td>15 (14.71%)</td>
</tr>
<tr>
<td>Metformin(<em>{-})new+gliclazide(</em>{-})healthy lifestyle(_{-})stop(^b)</td>
<td>102 (0.36%)</td>
<td>13 (12.75%)</td>
</tr>
<tr>
<td>Gliclazide(<em>{-})metformin(</em>{-})new+gliclazide(<em>{-})healthy lifestyle(</em>{-})stop(^b)</td>
<td>98 (0.34%)</td>
<td>8 (8.16%)</td>
</tr>
<tr>
<td>Healthy lifestyle(<em>{-})metformin(</em>{-})new+gliclazide(<em>{-})healthy lifestyle(</em>{-})stop(^b)</td>
<td>93 (0.32%)</td>
<td>6 (6.45%)</td>
</tr>
<tr>
<td>Metformin(<em>{-})tolbutamide(</em>{-})new+healthy lifestyle(<em>{-})metformin(</em>{-})tolbutamide(_{-})stop(^b)</td>
<td>91 (0.32%)</td>
<td>2 (2.2%)</td>
</tr>
</tbody>
</table>

\(^a\)DP: decision point.
\(^b\)Treatment options with a controlled percentage lower than the “no change” treatment option.
\(^c\)Treatment options with a controlled percentage higher than the “no change” treatment option.
**Discussion**

**Principal Findings**

This contribution evaluated the feasibility of using a precision cohort treatment options approach to generate personalized treatment options for patients with T2DM in the Dutch primary care setting. The approach involves the identification of relevant clinical treatment DPs from the longitudinal patient data to use as events of interest for modeling and analysis, patient similarity modeling for precision cohort construction, and treatment options and outcomes analysis as main elements. All procedures functioned well from a technical viewpoint; however, data size limitations proved challenging for reaching statistical significance for differences in outcomes of multiple treatment options.

**Decision Points**

Key information considered with the DPs included outcomes, clinical information (treatments and guidelines), and patient condition (measurements and comorbidities).

**Outcomes**

Two target outcome variables for T2DM were considered in this study, independently, to make the best use of the available data, as both HbA1c and fasting glucose are clinical measurements that can provide information about the T2DM state of the patient. The use of the 2 different outcome variables in separate scenarios enables a comparison between them and, in our perspective, offers the physician an opportunity to choose the one more relevant for a given clinical situation. In comparing the results for both scenarios, it is important to consider their significance for differences in outcomes of multiple treatment options.

To mimic the situation in practice where measurements of HbA1c and fasting glucose are intermittently made in the same patient to follow disease status, we also tried to analyze a “mixed case” scenario in which both HbA1c and fasting glucose measurements were used to build the DPs (eg, use HbA1c to identify a noncontrolled situation, then use fasting glucose to evaluate the treatment follow-up). This approach, however, did not result in the expected increase in the number of DPs extracted from the data, so it was not pursued further. The observed discrepancy in results for the different target variables makes clear that for the implementation of the precision cohort analytics approach in practice in the future, a unified definition of when the disease is to be considered “controlled,” and a common decision on what metric is to be used to assess it, needs to be made.

**Treatments and Guidelines**

The data extraction retrieved all the usual medications used to treat T2DM, except for SGLT-2 inhibitors. Since 2021, the recommendations for high-risk patients include SGLT-2
inhibitors as medication. The absence of this treatment in the data set seemingly might indicate that the current Dutch guidelines for patients with T2DM were not followed in the study population. However, this can be explained not only by the fact that the data used here are from before 2021 but also by the fact that patients considered high risk are more likely to be referred to a specialist in secondary care, whereas the Nivel-PCD is concerned with primary care.

In the Netherlands, T2DM treatment over time has focused increasingly on lifestyle adjustments where possible, especially in the early stage of treatment where lifestyle improvement is the first treatment of choice except for high-risk cases [37]. This is, for example, reflected in a recent analysis of patients with T2DM diagnosed between 2015 and 2019, which showed that half of these patients did not receive antidiabetic medication prescriptions within 1 year of the diagnosis data [38]. Thus, lifestyle advices are probably often given; however, the fact that the “all_guidelines_variables_false” cohort was the fourth largest cohort for the HbA1c scenario seems to indicate that lifestyle adjustment is not integrally registered in the data fields within the Nivel-PCD (it is hardly imaginable that patients diagnosed with T2DM would receive no treatment at all). Incomplete registration of lifestyle treatment might have occurred because it is a default standard choice, or possibly because it is registered in the text fields of the patient dossiers that GPs use to make notes and are not collected in the Nivel-PCD. This will likely have led to an underestimation of the “healthy lifestyle advice” treatment group and likely to some misclassification, especially of patients in cohorts that involved “no treatment_stop” in this analysis.

Moreover, the merging of the nonpharmacologic treatments such as diet and exercise (registered as separate treatments in the Nivel-PCD) into a single “healthy lifestyle advice” as done in this study did not allow us to assess the individual contribution of each of the separate lifestyle interventions. However, this analysis still offers a way to evaluate the importance of lifestyle interventions for the patient’s health. Future studies could explore the efficacy of individual lifestyle adjustments as a first line of therapy for patients with T2DM, provided they are adequately registered.

The absence of medication dosage information in the available data set means that the current approach cannot be used to inform decisions regarding dosage changes. Similarly, the influence of medications that are not directly targeting diabetes, for example, the ones targeting blood pressure, was not considered in this retrospective analysis. Nevertheless, information on prescriptions of these other medications is available in the data set, and as drugs may have interactions, future studies might explore the possibility of including a larger spectrum of medications for a deeper analysis, leading to more refined models and decision-making tools.

**Patient Condition**

The idea to use both mobility and emotional states complementary to standard clinical measurements such as blood pressure, blood lipids, and creatinine came as a suggestion from a GP consulted for the study. We took care to also include the aspect that such impairment may be temporal instead of chronic, for example, in the case of a broken leg. As mentioned, for this exploratory analysis, we made subjective, intuitive choices to operationalize these conditions, which were not validated by independent experts. Therefore, they remain subject to debate. However, the fact that 11 (44%) of the 25 largest cohorts included “mobility impaired” suggests that it is indeed relevant to include a mobility assessment as a filter variable for defining the precision cohorts. The same holds for emotional state; however, with inclusion in only 3 (12%) of the 25 largest cohorts, the importance seems lower than for mobility impairment.

The patient condition characteristics used in this analysis are necessarily limited to the information available in the Nivel-PCD. Information about social behavior, ethnicity, socioeconomic status, medication adherence, and many other factors was not included, which might have a major impact on the disease outcome. This fact points to a world beyond what was analyzed in this study.

**Patient Similarity Modeling and Precision Cohort Construction**

The selection of precision cohorts was done in a 2-step procedure, first a filtering step and then a similarity rating. The filtering step was based on guidelines, mobility, and mental state. The vast majority of the largest filter cohorts involved metformin and lifestyle advice as treatments, reflecting the importance of these as the first line of treatments according to the clinical guidelines. Patient similarity modeling and precision cohort construction were technically well feasible. However, data size limitations became apparent for the HbA1c scenario, where only 10 filter cohorts were considered sufficiently large (>200 DPs) to allow the construction of precision cohorts. This contrasts with the analysis of Ng et al [16] of the US EHR data, where the 75 largest T2DM cohorts based on filter variables had >200 DPs.

**Treatment Options and Outcomes Analysis**

The technical feasibility of treatment options and outcome analysis was also well established. For the HbA1c target outcome, approximately half of the global treatment options resulted in better outcomes than the “no change” option. For fasting glucose as the target outcome, this was the case for only very few treatment options, suggesting that the use of fasting glucose as the target outcome needs careful consideration. The few examples of precision cohort–based treatment options and outcomes analysis clearly illustrated the lack of statistical power available with the current data set. This further underlines that the method has high data availability requirements. Because a correction for multiple testing has to be applied, reaching statistical significance for differences in outcomes of multiple treatment options proved challenging, even for filter cohorts with 1000 DPs, as demonstrated in Figures 3 and 5. It is estimated that to overcome this limitation, the data set size should be at least an order of magnitude larger.

**Future Perspectives**

The method developed and applied for the Dutch primary care situation in this study aims to create precision cohorts that include a set of patients who are more similar to the patient of
interest in different attributes, from classic clinical measurements to assessments of mobility state and even mental state. This would allow the physician to selectively consider a group of patients that were in a very similar situation to the patient of interest and view statistics on past outcomes of available treatment options. Over time, with the increase in the available patient information and developments in computational methods, the ability to thus incorporate past clinical experience to generate more personalized treatment options for individual patients is enhanced, thereby potentially contributing to better treatment outcomes.

In this study, we took a highly reductionist approach to defining treatment outcome, that is, an HbA1c or fasting glucose clinical test result that falls below a predetermined threshold. As such it is of a highly reductionist nature.

Although we recognize that a patient’s perspective and experience are crucial factors both in decision-making and in evaluating treatment efficacy, our data-driven approach was as yet unable to account for those aspects. Still, the technique can support shared decision-making by practitioner and patient because the information on expected treatment outcomes is more personalized toward the individual patient and therefore more relevant in the discussion when balancing risks and expected outcomes with patient preferences and values. Although this study did not explore the actual use of the precision cohort approach for shared decision-making in practice, it offers valuable insight into the potential use from a data availability perspective.

It is important to mention that this approach does not aim to replace or lessen the actions of the physicians but to provide refined tools to support them in the medical decision-making process.

The workflow elaborated in this study was applied to the T2DM case but can be applied to any other disease or health disorder for which rich data and guidelines are available. Indeed, Ng et al [16] applied their approach to hypertension and hyperlipidemia as well as T2DM. However, applying the approach to different diseases requires rerunning all the steps of the workflow to adapt for the different diseases, including patient selection, choice of target outcome variables, incorporation of applicable clinical guidelines, selection of salient features, and tuning of the similarity model.

In this study, data availability was identified to be a principally limiting factor for feasibility. Considering that further personalization will lead to yet smaller cohorts, it is evident that increasing the pool of data for the precision cohort approach is essential to achieve a more meaningful and more robust analysis. Given that approximately 1 million patients have T2DM in the Netherlands, there is a realistic perspective for this; however, it will require combining health data from different EHR sources nationwide, which is a challenge in itself. This problem is aggravated for diseases for which the amount of data (ie, patients) is much smaller or the spectrum of treatment choices is larger.

Although methods to reduce the presence of bias in the data were applied, having more data available offers a possibility to improve the workflow, especially with respect to better selection of the confounder variables. This may lead to a better generalization, improving the performance of the workflow as a whole.

Conclusions
This study explored the feasibility of applying a patient similarity–based precision cohort approach to derive personalized treatment options for patients with T2DM treated in primary health care in the Netherlands using the Nivel-PCD. A previously published data analysis and modeling workflow for US EHR data was successfully adapted for this Dutch primary care setting, proving its potential for use in an LHS context. Although the approach proved technically well feasible, data size limitations need to be overcome before application for CDS purposes becomes realistically possible.

Acknowledgments
This research was partly funded by the Netherlands Organisation for Health Research and Development (project 104006001). The funders in the study had no role in the study design, data analysis, interpretation of the results, or writing of the report.

The authors are indebted to Mark Nielen for helpful discussions during the conceptualization of the study.

Data Availability
The data sets generated during and analyzed during this study are not publicly available. The Data from the Nivel Primary Care Database are available for research by other organizations upon approved data use request.

Authors’ Contributions
XP was involved in co-designing and executing the study and writing the manuscript. AdG was involved in conceptualizing, obtaining funding, co-designing the study, and writing the manuscript. WM was involved in conceptualizing the study, accomplishing study data access, and reviewing the manuscript. XP performed all data analytics work. All authors read and approved the final manuscript.

Conflicts of Interest
None declared.
Multimedia Appendix 1
Features used for the similarity model for glycated hemoglobin. For each variable, the name, the use (filter or similarity), the type of feature (numerical, Boolean, etc), and the similarity weight (influence on the model) are shown.

Multimedia Appendix 2
Features used for the similarity model for fasting glucose. For each variable, the name, the use (filter or similarity), the type of feature (numerical, Boolean, etc), and the similarity weight (influence on the model) are shown.

References


25. Farmacotherapeutisch Kompas homepage. Farmacotherapeutisch Kompas. URL: https://www.farmacotherapeutischkompas.nl/ [accessed 2024-04-10]


**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDS</td>
<td>clinical decision support</td>
</tr>
<tr>
<td>DM</td>
<td>diabetes mellitus</td>
</tr>
<tr>
<td>DP</td>
<td>decision point</td>
</tr>
<tr>
<td>EHR</td>
<td>electronic health record</td>
</tr>
<tr>
<td>GP</td>
<td>general practitioner</td>
</tr>
<tr>
<td>HbA1c</td>
<td>glycated hemoglobin</td>
</tr>
<tr>
<td>ICPC</td>
<td>International Classification of Primary Care</td>
</tr>
<tr>
<td>KNN</td>
<td>k-nearest neighbors</td>
</tr>
<tr>
<td>LASSO</td>
<td>least absolute shrinkage and selection operator</td>
</tr>
<tr>
<td>LHS</td>
<td>learning health system</td>
</tr>
<tr>
<td>LMNN</td>
<td>large margin nearest neighbor</td>
</tr>
<tr>
<td>MD</td>
<td>Mahalanobis distance</td>
</tr>
<tr>
<td>Nivel</td>
<td>The Netherlands Institute for Health Services Research</td>
</tr>
</tbody>
</table>
PCD: Primary Care Database
SGLT-2: sodium-glucose transport protein 2
T2DM: type 2 diabetes mellitus

Please cite as:
Pinho X, Meijer W, de Graaf A
Deriving Treatment Decision Support From Dutch Electronic Health Records by Exploring the Applicability of a Precision Cohort–Based Procedure for Patients With Type 2 Diabetes Mellitus: Precision Cohort Study
Online J Public Health Inform 2024;16:e51092
URL: https://ojphi.jmir.org/2024/1/e51092
doi: 10.2196/51092
PMID: 38691393
Characterization of Post–COVID-19 Definitions and Clinical Coding Practices: Longitudinal Study

Monika Maripuri1, MBBS, MPH; Andrew Dey1, MPH; Jacqueline Honerlaw1, RN, MPH; Chuan Hong2, PhD; Yuk-Lam Ho1, MPH; Vidisha Tanukonda1, MD; Alicia W Chen1, MS; Vidul Ayakulangara Panickan2, PhD; Xuan Wang3, PhD; Harrison G Zhang2, BA; Doris Yang2; Malarkodi Jebathilagam Samayamuthu4, MD; Michele Morris4, BA; Shyam Visweswaran5, MD, PhD; Brendin Beaulieu-Jones2, MBA, MBI, MD; Rachel Ramoni5, DMD, ScD; Sumitra Muralidhar5, PhD; J Michael Gaziano1,6, MPH, MD; Katherine Liao5, MPH, MD; Zongqi Xia8, MD, PhD; Gabriel A Brat2,9, MPH, MD; Tianxi Cai2, ScD; Kelly Cho1,6, MPH, PhD

1Veterans Affairs Boston Healthcare System, Boston, MA, United States
2Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States
3Department of Population Health Sciences, University of Utah, Salt Lake City, UT, United States
4Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, United States
5Office of Research and Development, US Department of Veterans Affairs, Washington, DC, United States
6Division of Aging, Department of Medicine, Mass General Brigham, Harvard Medical School, Boston, MA, United States
7Division of Rheumatology, Inflammation, and Immunity, Brigham and Women’s Hospital, Boston, MA, United States
8Department of Neurology, University of Pittsburgh, Pittsburgh, PA, United States
9Department of Surgery, Beth Israel Deaconess Medical Center, Boston, MA, United States

Corresponding Author:
Monika Maripuri, MBBS, MPH
Veterans Affairs Boston Healthcare System
2 Avenue de Lafayette
Boston, MA, 02111
United States
Phone: 1 617 390 6976
Email: monika.maripuri@va.gov

Abstract

Background: Post–COVID-19 condition (colloquially known as “long COVID-19”) characterized as postacute sequelae of SARS-CoV-2 has no universal clinical case definition. Recent efforts have focused on understanding long COVID-19 symptoms, and electronic health record (EHR) data provide a unique resource for understanding this condition. The introduction of the International Classification of Diseases, Tenth Revision (ICD-10) code U09.9 for “Post COVID-19 condition, unspecified” to identify patients with long COVID-19 has provided a method of evaluating this condition in EHRs; however, the accuracy of this code is unclear.

Objective: This study aimed to characterize the utility and accuracy of the U09.9 code across 3 health care systems—the Veterans Health Administration, the Beth Israel Deaconess Medical Center, and the University of Pittsburgh Medical Center—against patients identified with long COVID-19 via a chart review by operationalizing the World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC) definitions.

Methods: Patients who were COVID-19 positive with either a U07.1 ICD-10 code or positive polymerase chain reaction test within these health care systems were identified for chart review. Among this cohort, we sampled patients based on two approaches: (1) with a U09.9 code and (2) without a U09.9 code but with a new onset long COVID-19–related ICD-10 code, which allows us to assess the sensitivity of the U09.9 code. To operationalize the long COVID-19 definition based on health agency guidelines, symptoms were grouped into a “core” cluster of 11 commonly reported symptoms among patients with long COVID-19 and an extended cluster that captured all other symptoms by disease domain. Patients having ≥2 symptoms persisting for ≥60 days that were new onset after their COVID-19 infection, with ≥1 symptom in the core cluster, were labeled as having long COVID-19 per chart review. The code’s performance was compared across 3 health care systems and across different time periods of the pandemic.
Results: Overall, 900 patient charts were reviewed across 3 health care systems. The prevalence of long COVID-19 among the cohort with the U09.9 ICD-10 code based on the operationalized WHO definition was between 23.2% and 62.4% across these health care systems. We also evaluated a less stringent version of the WHO definition and the CDC definition and observed an increase in the prevalence of long COVID-19 at all 3 health care systems.

Conclusions: This is one of the first studies to evaluate the U09.9 code against a clinical case definition for long COVID-19, as well as the first to apply this definition to EHR data using a chart review approach on a nationwide cohort across multiple health care systems. This chart review approach can be implemented at other EHR systems to further evaluate the utility and performance of the U09.9 code.

(Keywords: veterans; long COVID-19; postacute sequelae of SARS-CoV-2; PASC; International Classification of Diseases; U09.9 ICD-10 code; algorithm validation; chart review; electronic health records; COVID-19)

Introduction

Characterizing the public health burden of postacute sequelae of SARS-CoV-2, also known as post–COVID-19 condition or colloquially as long COVID-19, has been difficult, given that multiple clinical case definitions have been proposed by various international health agencies [1-3]. While the exact components of these definitions vary, they share some common underlying features such as the development of symptoms that are new onset after a COVID-19 infection and the persistence of new onset symptoms for a duration of time after acute infection period. Electronic health records (EHRs) provide a uniquely rich resource for studying this condition at scale, and there have been multiple efforts to describe long COVID-19 symptoms and estimate prevalence in various EHR systems [4-9].

Specifically, the introduction of the International Classification of Diseases, Tenth Revision (ICD-10) code U09.9 for “Post COVID-19 condition, unspecified” has provided an alternative method of evaluating this condition in EHRs, and its use has been described in various health care systems [10-13]. However, there is no universal diagnostic guideline for defining long COVID-19, and thus, there is no standard guideline for assigning U09.9. The use of the U09.9 code and its accuracy in identifying long COVID-19 have not yet been evaluated against any existing clinical case definitions in a multicenter setting. Clinical coding of long COVID-19 has the potential for misclassification, given the heterogeneity and ambiguity around the definition of long COVID-19 [14].

This study aims (1) to characterize the use of ICD-10 code U09.9 across 3 health care systems and (2) to evaluate the accuracy of the U09.9 code against patients identified with long COVID-19 via chart review.

Methods

Data Sources and Study Cohort

The Consortium for Clinical Characterization of COVID-19 by EHR (4CE) is an international consortium for data-driven studies on the COVID-19 pandemic [15]. Three health care systems from the 4CE contributed chart review results for this study, namely, the national Veterans Health Administration (VHA), the Beth Israel Deaconess Medical Center (BIDMC), and the University of Pittsburgh Medical Center (UPMC). Over 15 million patients are collectively provided care across all 3 health care systems [16-18]. The VHA is the largest integrated health care system in the United States with 171 medical centers throughout the country [16]. The BIDMC is an academic medical center that is part of Beth Israel Lahey health care system located in Boston, and the UPMC is a Pittsburgh-based health care system with 40 hospitals across Pennsylvania [17,18].

EHR data from the 3 health care systems were used to identify patients who were COVID-19 positive, define patient characteristics, and obtain clinical notes for chart review. The sampling strategy for our chart review is described in Figure 1. Patients who had their first incidence of COVID-19 diagnosis reported within the participating health care systems’ EHR with either a U07.1 ICD-10 code for “COVID-19” or a positive polymerase chain reaction test performed between March 1, 2020, and December 31, 2021, were identified for chart review. From this COVID-19–positive cohort, we then sampled patients based on two approaches: (1) the presence of the U09.9 ICD-10 code, which was first introduced in the United States in October 2021, or (2) the presence of at least 1 new onset long COVID-19–related ICD-10 code if the patient did not have a U09.9 code. These long COVID-19–related ICD-10 codes were selected to enrich the chart review sample for patients who may potentially have long COVID-19. At the VHA, we further sampled patients from 2 time periods: those who were COVID-19 positive before September 1, 2021 (pre-U09.9 period), and those who were COVID-19 positive after this date (post-U09.9 period).

The presence of long COVID-19–related ICD-10 codes was identified via a data-driven process using EHR data from 10 health care systems at the 4CE [19]. Initial steps consisted of extracting longitudinal codified features such as ICD-10 codes and mapping these codified features to phecodes for new onset of conditions after COVID-19 infection. Phecodes are a curated grouping of ICD-10 codes used to analyze EHR data characterizing specific clinical symptoms or diagnoses [20]. New onset conditions were defined as those that were not present before the initial COVID-19 infection. The conditions were selected such that patients with COVID-19 are associated with a higher risk of a new onset of the condition after adjusting for baseline confounders such as age, sex, self-reported race,
and health care use. Marginal testing using a logistic regression framework was then performed to identify associated new onset conditions emerging 3 months after the initial infection. Conditions that passed marginal testing were then subject to conditional randomization analyses via distillation to robustly test whether a condition’s new onset is conditionally dependent on prior COVID-19 infection [19]. The Benjamini-Hochberg procedure was used to adjust for multiple comparisons [21]. Multimedia Appendix 1 [19,20] presents a list of the phecodes identified.

Figure 1. Patient sampling strategy for chart review. BIDMC: Beth Israel Deaconess Medical Center; EHR: electronic health record; ICD-10: International Classification of Diseases, Tenth Revision; PASC: postacute sequelae SARS-CoV-2; PCR: polymerase chain reaction; UPMC: University of Pittsburgh Medical Center; VHA: Veterans Health Administration.

Ethical Considerations
Exempt approval for this study was received by the Central Institutional Review Board at Veterans Affairs Boston Healthcare System (MVP000), BIDMC (2020P000565), and UPMC (STUDY20070095).

Chart Review Approach
The primary aim of our chart review was to operationalize the clinical case definition for long COVID-19 by the World Health Organization (WHO), with secondary aims to compare against a less stringent WHO definition and the Centers for Disease Control and Prevention (CDC) definition [3,22]. The chart review protocol (Multimedia Appendix 2) was developed at the VHA with guidance from 4CE subject matter experts to operationalize the WHO and CDC clinical case definitions. Long COVID-19 symptoms were identified from the WHO definition as well as through a literature review, and 11 commonly occurring symptoms among patients with long COVID-19 were classified into a “core” symptom cluster [23-26]. All other symptoms were classified into an “extended” symptom cluster based on their disease domain, which included cardiovascular, neurological, dermatological, musculoskeletal, digestive, and respiratory domains. For patients to be labeled as having long COVID-19 per the WHO definition during the chart review (reported here as “WHO-2”), at least 2 new onset symptoms after their COVID-19 infection were required (Figure 2). These could be either (1) two “core” symptoms or (2) one “core” and 1 “extended” cluster symptom, each of which must have persisted for 60 days or longer. All sampled patient charts had at least 6 months of clinical notes for review after the incident COVID-19 infection to allow appropriate assessment of symptoms. During chart review, all symptoms were collected based on their onset and duration of persistence for either 30 or 60 days (Multimedia Appendix 2) to allow evaluation against multiple long COVID-19 definitions. The less stringent WHO definition (reported here as “WHO-1”) was defined as a patient having just 1 core symptom persisting for at least 60 days or longer, and the CDC definition was defined as a patient having just 1 core symptom persisting for at least 30 days.

Reviewers had access to all clinical notes 1 year prior to the incident COVID-19 infection to determine baseline symptoms and conditions. Any symptoms present at the time of the incident COVID-19 infection or exacerbations of existing conditions were not considered new onset and thus not captured in the review. Chart reviewers were instructed to look for consistent mention or documentation of COVID-19-related symptoms in the notes and mark the duration of persistence (30 or 60 days). However, symptoms that waxed and waned over time were captured. At the VHA, a total of 500 patient charts were reviewed, and 200 patient charts were reviewed at each of the other 2 sites—BIDMC and UPMC.
Characterizing the U09.9 ICD-10 Code

To characterize the use of the U09.9 ICD-10 code in clinical practice, the following three metrics were investigated: (1) the frequency of the U09.9 code used over time from October 2021 to September 2022, (2) the frequency of the U09.9 code used across Veterans Integrated Service Networks (VISNs; which are regional systems of care at the VHA), and (3) the time elapsed between COVID-19 diagnosis and U09.9 code assignment.

Results

Characteristics of Study Cohort

Demographics and cohort sizes across the health care systems varied; notably, patients in VHA who had a COVID-19 diagnosis between March 1, 2020, and December 31, 2021, were generally White and male veterans. Among those who were assigned a U09.9 code, the demographics were generally similar with a few notable exceptions (Table 1). At the BIDMC and UPMC, however, the demographics were different, with a higher proportion of female patients who were assigned the U09.9 code. We also observed across all 3 health care systems that a higher proportion of those assigned a U09.9 code had received at least 1 dose of a COVID-19 vaccine compared to the general population of patients who were COVID-19 positive.

We observed a substantial variation in the use of the U09.9 code to diagnose long COVID-19 over time and region. Figures 3 and 4 show the results of our characterizations of the U09.9 code use 12 months following its introduction in the United States on October 1, 2021. Figure 3 shows the frequency of the U09.9 code used to diagnose long COVID-19 was highest from January to March 2022 at health care system 1, from February to March 2022 at health care system 2, and from December 2021 to January 2022 at health care system 3.

There were also large regional differences in the use of the U09.9 code across VHA health care system VISNs (Figure 4). VISN17 assigned the U09.9 code to 28.4% (6304/22,196) of all patients who received this code at the VHA, while VISN1 assigned the U09.9 code to just 2.4% (533/22,196) of all patients who received the code at the VHA.
<table>
<thead>
<tr>
<th>Demographics</th>
<th>Patient cohort and health care system</th>
<th>All patients who were COVID-19 positive(^\text{a}) from March 1, 2020, to December 31, 2021</th>
<th>All patients who were COVID-19 positive(^\text{a}) with a U09.9 ICD-10(^\text{b}) code</th>
<th>All patients who were COVID-19 positive(^\text{a}) with a new onset long COVID-19 feature</th>
<th>All patients who were COVID-19 positive(^\text{a}) from March 1, 2020, to December 31, 2021</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VHA(^\text{c}) (n=307,909)</td>
<td>BIDMC(^\text{d}) (n=30,294)</td>
<td>UPMC(^\text{e}) (n=147,653)</td>
<td>VHA (n=22,196)</td>
<td>BIDMC (n=164)</td>
</tr>
<tr>
<td>Age at incident COVID-19 diagnosis (years), mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>59.2 (16.1)</td>
<td>47.5 (20.5)</td>
<td>45.7 (23.8)</td>
<td>61.7 (15.1)</td>
<td>54.7 (13.9)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>272,957 (88.7)</td>
<td>16,579 (54.7)</td>
<td>63,421 (43)</td>
<td>19,275 (86.8)</td>
<td>58 (35.4)</td>
</tr>
<tr>
<td>Female</td>
<td>34,898 (11.3)</td>
<td>13,715 (45.3)</td>
<td>84,232 (57)</td>
<td>2921 (13.2)</td>
<td>106 (64.6)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>3521 (1.1)</td>
<td>31 (0.1)</td>
<td>704 (0.5)</td>
<td>273 (1.2)</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Asian</td>
<td>3326 (1.1)</td>
<td>1200 (4)</td>
<td>1303 (0.9)</td>
<td>249 (1.1)</td>
<td>4 (2.4)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>69,067 (22.4)</td>
<td>5074 (16.7)</td>
<td>14,966 (10.1)</td>
<td>3621 (16.3)</td>
<td>29 (17.7)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>3189 (1)</td>
<td>21 (0.07)</td>
<td>35 (0.02)</td>
<td>239 (1.1)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>White</td>
<td>208,457 (67.7)</td>
<td>8983 (29.7)</td>
<td>125,503 (85)</td>
<td>16,102 (72.5)</td>
<td>105 (64)</td>
</tr>
<tr>
<td>Not reported</td>
<td>0 (0)</td>
<td>14,985 (49.5)</td>
<td>5141 (3.5)</td>
<td>0 (0)</td>
<td>25 (15.2)</td>
</tr>
<tr>
<td>Vaccination status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Received at least 1 dose of COVID-19 vaccine</td>
<td>199,235 (64.7)</td>
<td>9649 (31.9)</td>
<td>58,244 (39.4)</td>
<td>15,295 (68.9)</td>
<td>112 (68.3)</td>
</tr>
<tr>
<td>Did not receive at least 1 dose of COVID-19 vaccine</td>
<td>104,907 (34.1)</td>
<td>20,645 (68.1)</td>
<td>89,409 (60.6)</td>
<td>6684 (30.9)</td>
<td>52 (31.7)</td>
</tr>
</tbody>
</table>

\(^\text{a}\)COVID-19 positive is defined as a patient having either a U07.1 ICD-10 code or a documented positive polymerase chain reaction test.

\(^\text{b}\)ICD-10: International Classification of Diseases, Tenth Revision.

\(^\text{c}\)VHA: Veterans Health Administration.

\(^\text{d}\)BIDMC: Beth Israel Deaconess Medical Center.

\(^\text{e}\)UPMC: University of Pittsburgh Medical Center.
Chart Review Results

Chart review at the VHA was conducted by 2 clinical reviewers (M Maripuri and JH) with a 1% overlap and an interrater reliability of 80%. At the BIDMC and UPMC, a chart review was conducted by 1 clinical reviewer (BB-J and MJS, respectively). The most common symptoms identified during chart review among patients with long COVID-19 as per the WHO-2 definition were shortness of breath, fatigue, cough, and loss of smell or taste from the core symptom cluster. Among the extended symptom clusters, we most commonly saw symptoms across the cardiovascular, gastrointestinal, neurological, and respiratory disease domains.

Chart review was performed on a total of 900 patients infected with COVID-19 across 3 health care systems. These 3 institutions were anonymized to provide an unbiased interpretation of the results. The performance of the U09.9 code was evaluated by calculating the positive predictive value (PPV) or the probability of having long COVID-19 (as defined by chart review classification) among the patients who had the U09.9 code. When using the WHO-2 definition to define long COVID-19, the PPV was 29.8% at health care system 1 and 62.4% and 23.2% at health care systems 2 and 3, respectively (Figure 5). However, when we consider the WHO-1 and the CDC definitions, the PPV of long COVID-19 was higher at all 3 health care systems, but at health care system 2, the PPV was slightly higher for the WHO-1 definition and remained the same for the CDC definition (Figure 5).
We also evaluated PPV in the sample of patients with new onset long COVID-19 features. Using the WHO-2 definition, the PPV was 7% at health care system 1 and 6.7% and 3% at health care systems 2 and 3, respectively.

Weighted sensitivity calculations were used due to the nonrandom nature of our labeled set in relation to the overall cohort. Our approach involved a random sampling of patients with the U09.9 code, while those without this code were intentionally downsampled. Applying standard sensitivity calculations to such a biased labeled set could lead to misleading results. To counteract this, weighted sensitivity was used, which assigns a specific sampling weight to each case in our labeled data set. This weight is calculated to be inversely proportional to the likelihood of a case’s inclusion in the sample. By integrating this weighting system, we ensure that each case’s contribution to the sensitivity analysis accurately mirrors its representation in the full cohort, thereby yielding more representative and reliable results. The overall performance of the U09.9 code based on the WHO-2 definition resulted in a weighted sensitivity of 15% at health care system 1 and 4.9% and 19.1% at health care systems 2 and 3, respectively.

Additionally, at the VHA, we looked at the prevalence of long COVID-19 among patients with the U09.9 ICD-10 code across different time periods based on their first COVID-19 infection date. From the chart reviewed cohort, 44.6% (50/112) had long COVID-19 from the pre-U09.9 period and 22.7% (53/233) from the post-U09.9 period at the VHA. The PPV of patients with long COVID-19 in the pre-U09.9 period is higher, as many of these patients were backcoded.

Through our chart review, we observed that patients were given the U09.9 code over a wide range of time from fewer than 29 days to over 365 days following their initial COVID-19 infection. Most patients did not have persisting symptoms after acute infection, and waxing and waning of symptoms were frequently observed.

**Figure 5.** Comparison of results among all sites. (A) Predictive values of ICD-10 U09.9 code for various clinical definitions of long COVID-19. (B) Capture of PASC symptoms using diagnosis codes and natural language processing data. (C) Average number of new onset PASC symptoms in long COVID-19 clinic patients. CDC: Centers for Disease Control; HS: health care system; ICD-10: International Classification of Diseases, Tenth Revision; PASC: postacute sequelae of SARS-CoV-2; SOB: shortness of breath; WHO: World Health Organization.
Discussion

Principal Results
This study provides a comprehensive, multicenter evaluation of the U09.9 code against proposed clinical case definitions for long COVID-19. It is also one of the first studies to apply a clinical case definition to EHR data using a chart review approach. The use of EHR data allowed evaluation of the U09.9 code across multiple health care systems nationwide. The availability of ample clinical notes enabled reviewers to ascertain whether observed symptoms after COVID-19 infection were truly new onset and to evaluate the duration of new symptoms, which are critical components of case definitions for long COVID-19. Another strength of this study was that we evaluated both the WHO and CDC clinical case definitions for long COVID-19 since one universal definition is not currently available. Our symptom collection approach (Multimedia Appendix 2) captured discrete symptoms by the duration of 30 or 60 days, which allowed for multiple case definitions to be applied.

There were large variations in the accuracy of the use of the U09.9 code for long COVID-19. We observed that 1 center had a much higher predictive value for patients with long COVID-19 among the U09.9 cohort across all definitions than the other 2 health care systems. This health care system also had the highest average number of new onset symptoms among patients seen in long COVID-19 clinics. The U09.9 code assignment at this health care system could have been more accurate due to a higher proportion of patients being seen at long COVID-19 clinics.

In a recent publication on this work, we also evaluated the capture of long COVID-19 symptoms using ICD-10 codes and natural language processing (NLP) data [14]. The Narrative Information Linear Extraction NLP tool was used to extract relevant concept unique identifiers that were manually mapped to each of the long COVID-19 symptoms we studied [27]. Among the chart-reviewed patients with the U09.9 code, we then identified the ones who had the various symptoms through the chart review. We then assessed what proportion of them with the symptom (ie, brain fog) had a corresponding ICD-10 or concept unique identifier mention of the concept and calculated the proportion of patients with the NLP or codified data capture. We shared the findings from this evaluation in Figure 5B. The performance of NLP was significantly better for all the commonly occurring long COVID-19 symptoms. The accuracy of using either a diagnosis code or NLP had the best results with 97% accuracy for loss of smell or taste and 87% accuracy for chest pain and cough.

Limitations
There were several limitations to this study. The cohort at the VHA had a higher proportion of male patients who were generally older and predominantly White. Incident COVID-19 infection was required for inclusion in the chart review, and it is possible that patients had an infection outside of the health care systems that was not recorded in the EHR. Patients may have also had symptoms that were not reported at health care system visits. The variation in the number of long COVID-19 clinics across regions may have led to a differential capture of symptoms for those patients who were seen at long COVID-19 clinics versus those who were seen by their primary care providers. We observed that symptoms among these patients were well documented as most long COVID-19 clinics have a specific template for evaluating and capturing COVID-19 symptoms [28]. In some instances, it was difficult to assess whether a symptom was truly new onset due to COVID-19 infection or a result of underlying health conditions noted at baseline. While the WHO definition has been in use since 2021, long COVID-19 is still an evolving disease, and the case definition may change over time as the condition is further characterized. We also faced some challenges in optimizing a heterogenous and sparse data capture within the EHR systems.

Comparison With Prior Work
The use of the U09.9 code has been described in several cohorts. The National Institutes of Health’s National COVID Cohort Collaborative (N3C) reported on the growing use of U09.9 from October 2021 through January 2022 in a nationwide cohort of 21,072 patients with the code [10]. However, the N3C did not require patients in the cohort to have a positive COVID-19 test to evaluate the use of the U09.9 code, and 37.2% (n=12,550) of patients did not have a COVID-19 index date. McGrath et al [12] also reported increasing use of the U09.9 code in the months following its release in the nationwide HealthVerity cohort of 56,143 patients with a U09.9 code that included children younger than 18 years of age. Similar to the N3C, this cohort did not require a COVID-19 positive test for evaluation of the U09.9 code, and only 70.4% (n=8879) had a documented COVID-19 infection. Of the patients who were COVID-19 positive with U09.9, the median time from infection to U09.9 diagnosis was 56 (21-200) days. A study in Sweden by Bygdell et al [11] reported 10,196 patients with the U09.9 code. They also found that 2% of the population who were COVID-19 positive in the 2 largest regions of Sweden had U09.9 at least 28 days after infection.

Conclusions
Our findings suggest that the U09.9 code should be used judiciously in EHR-based studies of long COVID-19. Given the low PPV of the U09.9 code, its use as a proxy for long COVID-19 is not recommended. However, the sensitivity of the code makes it useful for identifying patients who may have long COVID-19 and thus require further clinical evaluation. This was one of the initial efforts toward validating long COVID-19 against a clinical case definition and the U09.9 code through a chart review on a nationwide cohort. The chart review approach developed at the VHA can be implemented at other EHR systems to further evaluate the utility and performance of the U09.9 code. Further efforts to develop a more refined and reproducible phenotyping algorithm for long COVID-19 using NLP are underway, using the chart review labels from our study for algorithm training and development.
Acknowledgments
This study was supported by the Office of Research and Development, Veterans Health Administration through the Centralized Interactive Phenomics Resource program. The authors thank the Consortium for Clinical Characterization of COVID-19 by Electronic Health Record for their collaboration. This publication does not represent the views of the Department of Veteran Affairs or the US government.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Selected postacute sequelae of SARS-CoV-2 feature phecodes.
[DOCX File, 25 KB - ojphi_v16i1e53445_app1.docx]

Multimedia Appendix 2
Chart review protocol.
[DOCX File, 27 KB - ojphi_v16i1e53445_app2.docx]

References


15. 4CE: consortium for clinical characterization of COVID-19 by EHR. Covidclinical.net. URL: https://covidclinical.net/ [accessed 2022-11-05]
16. Veterans Health Administration. US Department of Veterans Affairs. URL: https://www.va.gov/health [accessed 2023-03-22]
17. By the numbers: UPMC facts and figures. UPMC. URL: https://www.upmc.com/about/facts/numbers [accessed 2023-03-22]
18. About Beth Israel Deaconess Medical Center. Beth Israel Deaconess Medical Center. URL: https://www.bidmc.org/about-bidmc [accessed 2023-03-22]

Abbreviations

4CE: Consortium for Clinical Characterization of COVID-19 by Electronic Health Record
BIDMC: Beth Israel Deaconess Medical Center
CDC: Centers for Disease Control and Prevention
EHR: electronic health record
ICD-10: International Classification of Diseases, Tenth Revision
NC3: National COVID Cohort Collaborative
NLP: natural language processing
PPV: positive predictive value
UPMC: University of Pittsburgh Medical Center
VHA: Veterans Health Administration
VISN: Veterans Integrated Service Networks
WHO: World Health Organization

Edited by E Mensah; submitted 11.10.23; peer-reviewed by D Drew, L Zhang; comments to author 28.12.23; revised version received 19.01.24; accepted 19.03.24; published 03.05.24.

Please cite as:
Patient Characteristics Associated With Phone and Video Visits at a Tele-Urgent Care Center During the Initial COVID-19 Response: Cross-Sectional Study

Saif Khairat1,2,3, PhD, MPH; Roshan John1, PSM; Malvika Pillai1, PhD; Philip McDaniel4, MA; Barbara Edson5, RN, MBA, MHA

1Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
2Cecil G Sheps Center for Health Services Research, Chapel Hill, NC, United States
3School of Nursing, University North Carolina at Chapel Hill, Chapel Hill, NC, United States
4Digital Research Services Department, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
5UNC Health, Morrisville, NC, United States

Corresponding Author:
Saif Khairat, PhD, MPH
Carolina Health Informatics Program
University of North Carolina at Chapel Hill
428 Carrington Hall
Chapel Hill, NC, 27514
United States
Phone: 1 9198435413
Email: Saif@unc.edu

Abstract

Background: Health systems rapidly adopted telemedicine as an alternative health care delivery modality in response to the COVID-19 pandemic. Demographic factors, such as age and gender, may play a role in patients’ choice of a phone or video visit. However, it is unknown whether there are differences in utilization between phone and video visits.

Objective: This study aimed to investigate patients’ characteristics, patient utilization, and service characteristics of a tele-urgent care clinic during the initial response to the pandemic.

Methods: We conducted a cross-sectional study of urgent care patients using a statewide, on-demand telemedicine clinic with board-certified physicians during the initial phases of the pandemic. The study data were collected from March 3, 2020, through May 3, 2020.

Results: Of 1803 telemedicine visits, 1278 (70.9%) patients were women, 730 (40.5%) were aged 18 to 34 years, and 1423 (78.9%) were uninsured. There were significant differences between telemedicine modalities and gender (P<.001), age (P<.001), insurance status (P<.001), prescriptions given (P<.001), and wait times (P<.001). Phone visits provided significantly more access to rural areas than video visits (P<.001).

Conclusions: Our findings suggest that offering patients a combination of phone and video options provided additional flexibility for various patient subgroups, particularly patients living in rural regions with limited internet bandwidth. Differences in utilization were significant based on patient gender, age, and insurance status. We also found differences in prescription administration between phone and video visits that require additional investigation.

DOI: 10.2196/50962

Keywords

telehealth; telemedicine; tele-urgent care; virtual urgent care; nonemergency care; televist; phone visit; video visit; urgent care; health services research; COVID-19; health disparities; insurance status; cross-sectional study
Introduction

Health systems rapidly adopted telemedicine as an alternative health care delivery modality in response to the COVID-19 pandemic. Demographic factors, such as age and gender, may play a role in patients’ choice of a phone or video visit [1-3]. However, it is unknown whether there are utilization differences between phone and video visits.

The pandemic led to a rise in phone and video consultations, providing an opportunity to study their usage across demographics and outcomes, such as medication prescriptions. Telemedicine can help improve health access and reduce disparities for vulnerable populations [4-7]. Although we know that medication prescription differs between in-person and video visits [8], there is a gap in the knowledge regarding differences in prescription administration, whether medication was prescribed or not, between telephone and video visits. Driven by prior differences in prescription administration among providers based on gender and specialty [7], we hypothesized that prescription administration, a service outcome of telemedicine, may differ between phone and video visits.

Phone-based treatment has been found feasible, acceptable, and effective compared to face-to-face visits. It is a promising alternative in telemedicine, offering tailored interventions [9]. Phone visits have taken less time and have been used more frequently, but there have not been significant differences in patient perceptions or other clinical outcomes [10].

Telemedicine’s growth during the pandemic has led to a need for understanding the limitations of telephone-based versus video-based consultations for clinical care [11,12]. Patients reported that video consultations were more favorable compared to phone consultations, claiming that video visits led to improved outcomes, better diagnostic accuracy, and patient satisfaction [13-18].

Previous studies have looked at the impact of phone or video visits on vulnerable patients [19-23], but there is a lack of research on the differences in patient characteristics between the 2 modes of telemedicine-based care. Understanding these differences can help health organizations and policy makers tailor telehealth options to better suit patients.

Telemedicine use during the pandemic has been examined in various clinical environments, such as primary care, geriatrics, and subspecialties [5,22,23]. It is unclear how phone and video health care delivery in urgent care clinics was affected during the initial phases of the COVID-19 pandemic, especially regarding wait times and visit duration. The demand for urgent care clinics increased due to emergency department overcrowding, cost increase, and long wait times [24,25]. Therefore, it is important to understand the changes in urgent care practices considering telemedicine deployment postpandemic.

In this exploratory study, we examined patient and service characteristics of on-demand telehealth utilization and whether they differed by modality during the initial phase of the pandemic when the health care system suspended all in-clinic visits. We used the Donabedian framework of structure-process-outcome to inform this study design [26].

Methods

Study Overview

We conducted a descriptive analysis on a cross-sectional study of patients using a statewide, on-demand tele-urgent care clinic in the southeastern United States region. The Virtual Urgent Clinic (VUC) is an on-demand clinic open for nonemergency concerns 24 hours a day and 7 days a week. Regardless of whether they are new or existing patients, any individuals can register and access the virtual clinic through the web-based portal. To use the telemedicine service, individuals must create an account, input their medical history, and request a virtual care visit. Individuals can choose their telemedicine modality—telephone or video—through a computer, tablet, or phone.

The cost of the visit was the same for phone and video visits. The clinic provides on-demand service such that individuals can log on to the web-based portal and choose to have a visit immediately or schedule a visit for a later date. Board-certified physicians are available 24 hours every day of the week to provide care for patients. If an individual is an existing patient, documentation of the virtual visit is integrated into the electronic medical record after the visit is completed.

Data Collection

VUC monthly data were collected from March 3, 2020, through May 3, 2020, using the institutional data warehouse. The data set included patient information, such as age, gender, insurance status, and residential address, and service characteristics, such as telemedicine modality, wait time, visit duration, and medication prescription outcomes. To avoid double counting of patients or visits, each patient and each visit received a unique identifier. Incomplete encounters were recorded in the data set as incomplete if the call was not completed for any reason. The rate of incomplete encounters was only 7.9% (142/1803) of the total visit volume in this study and was included to better understand the characteristics of patients who sought care via telehealth.

Outcomes

Our primary endpoints were the characterization of telemedicine modalities (phone vs video) on patient characteristics measured by demographics and insurance status, utilization measured by the volume of visits; and service characteristics measured by medication prescriptions and visit wait times. The secondary endpoint was utilization, which was measured by the number of visits from rural and urban neighborhoods.

Statistical Analysis

The study data included patient age, gender, health insurance status, address, number of medication prescriptions, number of visits, and choice of telemedicine modality. For each of these variables, we calculated descriptive statistics for each demographic category stratified by modality (phone or video) and the total of both groups. A $\chi^2$ test was calculated to check for significant differences between telemedicine visits and these
variables. Additionally, we calculated the average wait time and visit duration for phone and video visits. A 2-sample $t$ test assuming unequal variances (Welch $t$ test) was also conducted to determine if there was a statistically significant difference in the average wait times and visit duration lengths between phone and video telemedicine visits.

To examine the predictors of prescription administration, we constructed a logistic regression model with a dichotomous dependent variable of prescription administration ($0=$ no prescription and $1=$ at least 1 prescription given) as a dependent outcome variable and patient age, gender, insurance status, location, and telemedicine modality as independent variables in the model predictors. We used a $P$ value level of .05 to indicate statistical significance.

**Geospatial Analysis**

Geographical locations for patients with VUC visits over the phone or video were examined to assess the urban-rural spread of the patients in this data set. Using the US Census definition, cities with populations of 50,000 people or more were designated as urban, and those with less than 50,000 people were designated as rural. In the telemedicine data set, 198 places in North Carolina were found, of which 179 were classified as rural and 19 were classified as urban, which was used to develop the health access map. A $\chi^2$ analysis was used to determine the significance between an encounter from an individual in an urban or rural area and the encounter modality.

To understand the association between telehealth modality and location, we used ArcGIS (Esri) to map zip code–level populations, as reported in the 2010 US Census Bureau data, with VUC visits based on Zip Code Tabulation Areas (ZCTAs). We used the 2016 American Community Service (ACS) to calculate the percentage of households with internet access by ZCTA. We then mapped the ACS data and visit counts from the VUC by modality on the North Carolina (NC) map to better understand the preference of patients for modalities based on internet availability.

We used natural breakdowns to quantify the percentage of households with internet in each NC zip code to determine the threshold for low, medium, and high categories based on the 2016 ACS data set. The colors along the bottom row (gray to light blue to teal) represent ZCTAs with a low percentage (0%-71%) of households with internet access and an increasing number of phone (or video) visits. The colors in the middle row (light pink to light purple to blue) represent ZCTAs with a medium percentage (72%-82%) of households with internet access and an increasing number of phone (or video) visits. The colors along the top row (pink to purple to dark purple) represent ZCTAs with a high percentage (83%-100%) of households with internet access and an increasing number of phone (or video) visits. The colors along the diagonal (gray to light purple to dark purple) represent ZCTAs with low internet access and low telemedicine visits, medium internet access and medium telemedicine visits, and high internet access and high telemedicine visits. For phone visits, the breaks were 1-2 (low), 3-6 (medium), and 7-37 (high). For video visits, the breaks were 1 (low), 2-3 (medium), and 4-15 (high). We used quantiles to determine the threshold for low, medium, and high categories based on the ACS 5-year estimates from 2015-2019.

**Ethical Considerations**

University of North Carolina at Chapel Hill institutional review board approval was obtained prior to conducting this study (18-1628).

**Results**

**Telemedicine Visit Overview**

Table 1 shows a series of visit counts of the patients who used the telemedicine service during the observed period categorized by the patient characteristics captured in this study. It also indicates the $\chi^2$ and $P$ values for significance tests for the differences between these observed characteristics.
Table 1. Percentage statistics and \( \chi^2 \) values for phone and video telemedicine visits.

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Phone visits (n=1414)</th>
<th>Video visits (n=389)</th>
<th>Total visits (N=1803)</th>
<th>Chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visits per day, mean (SD)</td>
<td>22.8 (9)</td>
<td>6.3 (3.1)</td>
<td>29.1 (10.7)</td>
<td>N/A(^a)</td>
<td>N/A</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>16.79 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Women</td>
<td>1033 (73)</td>
<td>245 (63)</td>
<td>1278 (70.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>377 (26.7)</td>
<td>144 (37)</td>
<td>521 (28.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonbinary</td>
<td>4 (0.3)</td>
<td>0 (0)</td>
<td>4 (0.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), n (%)</td>
<td></td>
<td></td>
<td></td>
<td>24.99 (4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&lt;18</td>
<td>96 (6.8)</td>
<td>57 (14.7)</td>
<td>153 (8.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>579 (40.9)</td>
<td>151 (38.8)</td>
<td>730 (40.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-50</td>
<td>486 (34.4)</td>
<td>123 (31.6)</td>
<td>609 (33.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51-64</td>
<td>185 (13.1)</td>
<td>44 (11.3)</td>
<td>229 (12.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥65</td>
<td>68 (4.8)</td>
<td>14 (3.6)</td>
<td>82 (4.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health insurance status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>18.91 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Insured</td>
<td>329 (23.3)</td>
<td>51 (13.1)</td>
<td>380 (21.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uninsured</td>
<td>1085 (76.7)</td>
<td>338 (86.9)</td>
<td>1423 (78.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residence, n/N (%)</td>
<td></td>
<td></td>
<td></td>
<td>6.74 (1)</td>
<td>.009</td>
</tr>
<tr>
<td>Rural</td>
<td>782/1370 (57.1)</td>
<td>189/381 (49.6)</td>
<td>971/1751 (55.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>588/1370 (42.9)</td>
<td>192/381 (50.4)</td>
<td>780/1751 (44.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prescriptions per visit, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>24.07 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Received</td>
<td>980 (69.3)</td>
<td>218 (56)</td>
<td>1198 (66.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did not receive</td>
<td>434 (30.7)</td>
<td>171 (44)</td>
<td>605 (33.6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)N/A: not applicable.

Patient Characteristics

Phone visits constituted most of the 1803 total visits (n=1414, 78.4%), with an average of 22.8 (SD 9) daily visits, while video visits accounted for the remaining visits (n=389), with a daily average of 6.3 (SD 3.1) visits. Most of the patients were women across both phone and video modalities (phone visits: n=1033, 73%; video visits: n=245, 63%). Among age groups, patients aged 18 to 34 years had the most visits (phone visits: n=579, 40.9%; video visits: n=151, 38.8%), with patients aged 35 to 50 years being the next most represented age group (phone visits: n=486, 34.4%; video visits: n=123, 31.6%). Across both modalities, the least present age group included patients older than 65 years (phone visits: n=68, 4.8%; video visits: n=14, 3.6%). Most patients across both modalities were uninsured (phone visits: n=1085, 76.7%; video visits: n=338, 86.9%).

Significant differences between telemedicine modalities and gender (\( P < .001 \)), age (\( P < .001 \)), insurance status (\( P < .001 \)), health access (\( P = .009 \)), and prescriptions given (\( P < .001 \)). This suggests that men, patients younger than 18 years, uninsured patients, and patients residing in urban areas preferred the video modality for telemedicine visits, and video visits were more associated with not getting prescriptions.

Telemedicine Service Characteristics

Prescription Administration

More patients received at least 1 prescription (phone visits: n=980, 69.3%; video visits: n=218, 56%) from a telemedicine visit rather than no prescription. Video visits were more associated with no prescriptions than phone visits (\( P < .001 \)). Significant differences were found in medication prescription administration between phone and video visits (\( P < .001 \); Table 1).

For phone visits, of a total of 1414 phone visits, 980 (69.3%) resulted in at least 1 prescription given, while the other 434 did not receive any prescriptions. On average (SD), patients received 1 (1.02) prescription per encounter. Of all phone visits, 434 (30.7%) patients did not receive a prescription, 944 (66.8%) patients received 1-3 prescriptions in an encounter, and 36 (2.5%) patients received 4-7 prescriptions in an encounter.

For video visits, from a total of 389 video visits, 218 (56%) resulted in at least 1 prescription given, while the other 171 did not receive any. The average (SD) number of prescriptions per encounter was 0.84 (1.00). Of all video visits, 171 (43.9%) patients did not receive a prescription, 210 (54%) patients received 1-3 prescriptions in an encounter, and 8 (2.1%) patients received 4-7 prescriptions in an encounter.

We found that 5 patient characteristics were strong predictors of telemedicine prescription administration (Table 2). Predictors...
that were positively associated with prescription administration were patients aged 18 to 34 years ($\beta=.62, P<.001$), 35 to 50 years ($\beta=.81, P<.001$), and older than 65 years ($\beta=.94, P=.002$). Predictors that were negatively associated with prescription administrations were video visits ($\beta=-.47, P<.001$) and male patients ($\beta=-.38, P<.001$). There was no significant relationship between patients’ insurance status and prescription rates.

Table 2. Logistic regression model showing patient demographic associations with telemedicine prescription administration. The independent variables were modality, age, gender, and insurance status. The depended variable was prescriptions given.

|                        | Estimate | SE  | z score | Pr(>|z|)$^a$ | $R^2$ |
|------------------------|----------|-----|---------|--------------|-------|
| Model intercept        | 0.2711   | 0.1853 | 1.463   | .14          | 0.027269 |
| Modality               |          |      |         |              |       |
| Video                  | -0.4724  | 0.1204 | -3.922  | $.001        | N/A$^b$ |
| Gender                 |          |      |         |              |       |
| Men                    | -0.3878  | 0.1108 | -3.5    | $.001        | N/A    |
| Nonbinary              | 0.4606   | 1.1715 | 0.393   | .69          | N/A    |
| Health insurance status|          |      |         |              |       |
| Insured                | 0.1629   | 0.1308 | 1.245   | .21          | N/A    |
| Age (years)            |          |      |         |              |       |
| 18-34                  | 0.6227   | 0.1858 | 3.351   | $.001        | N/A    |
| 35-50                  | 0.8057   | 0.19  | 4.241   | $.001        | N/A    |
| 51-64                  | 0.3573   | 0.2173 | 1.644   | .10          | N/A    |
| ≥65                    | 0.9421   | 0.3045 | 3.094   | 0.002        | N/A    |

$^a$Pr(>|z|): $P$ value associated with the value in the z score column.

$^b$N/A: not applicable.

**Wait Times and Visit Duration**

The average wait time for patients to start their phone visits was 64.1 (SD 129.9) minutes, while the average wait time for patients with video visits was 24.6 (SD 45.6) minutes. The average visit duration for phone visits was 7.3 (SD 4.4) minutes, while the average visit duration for patients in video visits was 9.0 (SD 5.9) minutes. Significant differences existed between the average wait times and durations for phone and video visits (Welch $t$ test $P<.001$ for both wait times and duration). For phone and video visits in this data set, the daily wait times for patients to see a physician across each modality are indicated in Figures 1A and 1B, respectively. The number of physicians working daily shown in these figures peaked at a maximum of 33 physicians on March 21 and 22. The number of phone sessions facilitated was also at its peak on these days at 47 phone visits. Phone users experienced the longest wait times in the second half of March, but both phone and video users experienced extended wait times in this same period compared to April.
Figure 1. Comparison of (A) phone visit and (B) video visit wait times with a count of daily visits and physicians working.

A) Phone visits

B) Video visits

Telemedicine Utilization in Rural and Urban Areas

Of the 1080 NC zip codes, 262 (24.3%) had a low percentage of households with internet access, 277 (25.6%) had a medium percentage of households with internet access, and 269 (24.9%) had a high percentage of households with internet access. There were 272 (25.3%) zip codes with no internet access.

The overall utilization of video visits was higher in areas with high percentages of households having internet access (Figure 2). Among the individuals from zip codes with low internet access there were 127 (83.5%) phone visits and 25 (16.5%) video visits. Zip codes with medium internet access had 367 (80.8%) phone visits and 87 (19.2%) video visits, and those with high internet access were 879 (76.2%) phone visits and 274 (23.8%) video visits.
Visits to the telemedicine-based clinic came from 431 (40%) unique NC zip codes. Of these, 251 (58.2%) were rural zip codes and 180 (41.8%) were urban zip codes (Figure 2). The density of the visits, shown in larger icons in Figure 2, originated mostly from major metropolitan areas like the state capitol or the Research Triangle Park. Phone visits provided further reach into areas with low internet access, while video visits mainly occurred in urban settings with high access to internet services.

Phone visits originated from 290 (26.9%) unique NC zip codes, of which 170 (58.6%) were from rural areas, 80 (27.6%) were from urban areas, and 50 (16.8%) were from out of state. Video visits occurred in 141 (32.4%) unique NC zip codes, of which 80 (56.7%) were from rural areas, 56 (39.7%) were from urban areas, and 5 (3.6%) were from out of state. Phone visits provided better reach into rural areas; however, video visits had widespread coverage, demonstrating the potential to complement phone visits in rural areas. Both phone and video visits within urban areas provided comparable coverage as expected.

**Discussion**

**Principal Findings**

We conducted a cross-sectional study of telemedicine urgent care visits completed through phone or video using a statewide,
on-demand urgent care telemedicine clinic, focusing on demographics, utilization, and service characteristics. We observed significant differences in service characteristics between phone and video visits. The rate of medication prescription was much higher among phone visits compared to video visits. Patients had a higher probability of receiving a prescription during a phone visit, while the probability of receiving a prescription was lower during a video visit. Differences in gender, age, and telemedicine modality were associated with significant variations in prescription administration.

Similarly, significant differences in wait time and visit duration were observed between phone and video visits, where phone visits had higher wait times and longer visit durations. The high volume of requests for phone visits can justify the long wait. It was unclear if providers compensated for the long wait times by providing more visit time or if patients who waited longer had more questions based on the differences in visit durations.

Utilization of phone and video visits differed significantly. Women, insured patients, and those residing in rural areas preferred phone visits, while men, uninsured patients, and those residing in urban areas preferred video visits. Patients older than 65 years were equally split. The increase in video visits was due to pandemic-related cancellations of in-person appointments. Video visits were more common for children due to the need for clinical examination. Phone visits were more common in rural areas with no internet access for video visits. Rural patients preferred phone visits while urban patients preferred video visits. The reason for this preference is unclear. We suspect that a combination of privacy concerns, lack of confidence in their internet connection, and a lack of awareness may drive patients’ decisions; however, more investigation is needed [27,28].

Tying our findings to similar studies in the literature was a challenge because of a gap in studying the differences between telephone and video visits on the same outcomes [29]. Comparative studies have indicated that there has not been a meaningful difference between these modalities, having similar consultation session lengths, content, and perceived quality [30-32]. One study reported that older, rural, and ethnic minority patients were associated with lower utilization rates of video visits compared to phone visits [23]. A previous study reported that patients who had telephone visits had longer visit durations than those who had video visits [32], which contradicts our finding where video visits were longer in duration. A few studies have indicated increased utilization of telemedicine to trend toward women, with women being more likely to attend telephone-based interventions and to benefit from such interventions in the context of addiction treatment [33,34]. Moreover, another study showed that no major differences in utilization were found between video and telephone visits [31], which contradicts our findings demonstrating higher utilization of telephone visits compared to video visits.

Other studies explored telemedicine modalities separately demonstrating limitations due to selection bias in patient populations, such as including patients from a single hospital or clinic setting [13,14,16]. There is also concern that these studies often cater to specialized medical concerns or treatment options, which limits the demographic diversity of the patients recruited regarding factors such as age or gender, making the findings less generalizable [17,30]. Little was known regarding the patient characteristics of telephone or video telemedicine modalities across the rural-urban divide, patient insurance statuses, and prescriptions provided to the patients.

The COVID-19 policy waivers by the Center of Medicaid and Medicare and private insurers to include phone and video visits appear to be an effective decision that increased access and reduced disparities [35,36]. Additionally, this study shows that internet access is still limited in rural areas, which may limit the ability to conduct patient video visits, resulting in more phone visits. We recommend policymakers to continue to support video and phone visits equally, and we highlight the importance of building internet capacity within rural and vulnerable communities to expand the effective use of telemedicine.

**Limitations**

This study had several limitations. We conducted a cross-sectional study as we could not randomize patients to a telemedicine modality due to the complexity of the process and given the sensitivity of COVID-19. In addition, the study was conducted over 2 months (March 3, 2020, to May 3, 2020) at the height of the pandemic with a limited amount of data; however, this reflected the initial response to the pandemic when telemedicine was the primary option for care. A large proportion of patients in this study were uninsured. Uninsured patients preferred telehealth during the initial phase of the pandemic due to the suspension of in-person visits and the shutdown of health care systems and primary care clinics, which are more expensive for uninsured patients compared to emergency departments [37]. This study did not include a comparison to in-person consultations because the health care system suspended all nonessential visits during the observed study period, starting on March 20, 2020. There were no data collected on race, ethnicity, or type of insurance used or covered, which could have added value to the findings of this study. The diagnosis type may confound the difference in prescription administration of phone and video visits. We could not merge the telemedicine data with the electronic health record data to assess the difference in documentation quality between phone and video visits. No information was available to determine if the visit wait times in the data set included those seeking a telemedicine visit immediately as opposed to at a later date. Wait times could be separated for those seeking immediate appointments to improve our findings. Physician-level data was not accessible, limiting our assessment of factors such as clinician preparedness. Finally, the study findings were limited to 1 site, and so the generalizability to other settings is limited.

**Conclusion**

Our study analyzed the use of phone and video visits at a telemedicine clinic during the COVID-19 pandemic. We discovered that providing patients with a variety of phone and video options was beneficial for many patient groups, especially those in rural or low-bandwidth areas. Gender, age, and insurance status were also factors affecting usage. Moreover,
we observed differences in prescription administration between the 2 modalities that require further investigation. Our findings indicate that phone visits were more prevalent in rural regions compared to urban areas. To promote telemedicine adoption and quality, we must work toward improving internet infrastructure in rural areas, educating patients on selecting the appropriate modality, and establishing equitable service policies for phone and video visits.

Acknowledgments
This study was funded by the National Institutes of Health National Library of Medicine (grant 1T15LM012500-01). Saif Khairat was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (grant RC2TR004380). The authors would like to acknowledge the assistance of Dr. Robert Gianforcaro for his invaluable guidance throughout this study.

Authors’ Contributions
SK contributed to the data collection, study design, analysis, and manuscript writing. RJ contributed to the data analysis and manuscript writing. MP contributed to data analysis and manuscript writing. PM contributed to map creation and manuscript revision. BE contributed to the study design and data collection. All authors have reviewed and agreed to submitting the manuscript.

Conflicts of Interest
None declared.

References


Abbreviations

ACS: American Community Service  
NC: North Carolina  
VUC: Virtual Urgent Clinic  
ZCTA: Zip Code Tabulation Area

©Saif Khairat, Roshan John, Malvika Pillai, Philip McDaniel, Barbara Edson. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 19.01.2024. 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 the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.
Trends in the Ophthalmic Workforce and Eye Care Infrastructure in South India: Cross-Sectional Questionnaire Study

Srinivasa R Pallerla1, MBBS, MD, MCEH; Madhurima R Pallerla2, BOptom, MHA; Sannapaneni Krishnaiah3, MSc, PhD

1Andhra Pradesh Right to Sight Society, Hyderabad, India
2Sri Jyothi Eye Clinic, Hyderabad, India
3LV Prasad Eye Institute, Hyderabad, India

Corresponding Author:
Srinivasa R Pallerla, MBBS, MD, MCEH
Andhra Pradesh Right to Sight Society
Plot No 12 BN Reddy Colony Road
No 14 Banjara Hills
Hyderabad, 500034
India
Phone: 91 9849078882
Email: srinivasar107@hotmail.com

Abstract

Background: This study is part of broad-based research to determine the impact of blindness control activities in general and with special reference to the Andhra Pradesh Right to Sight Society (APRTSS) activities in the southern Indian states of Andhra Pradesh and Telangana. As part of the global “VISION 2020: The Right to Sight” initiative, the APRTSS was established in the undivided state of Andhra Pradesh in 2002. Since then, the APRTSS has been actively implementing the strategies of VISION 2020 to reduce visual impairment and blindness in the state.

Objective: The availability and distribution of the eye care workforce are essential to reach the goals of VISION 2020: The Right to Sight, the global initiative to eliminate avoidable blindness. This study assessed the trends in the availability and distribution of eye health professionals and eye care infrastructure in 2 southern Indian states: Andhra Pradesh and Telangana.

Methods: This cross-sectional study used a pretested questionnaire to gather data for the year from 2012 to 2013. Data for 2002 to 2003 were collected from available historical records. The questionnaires were pretested in a pilot study conducted before the main survey. Pretested questionnaires were administered to all eye care professionals—ophthalmologists (n=1712) and midlevel ophthalmic personnel (MLOP; n=1250)—eye care facilities with ≥10 inpatient beds or performing ≥100 cataract surgeries per annum (n=640), local nongovernmental eye care organizations (n=182), and international eye care organizations (n=10). Data were collected for 2 different time periods: the baseline year of 2002 to 2003 and the target year of 2012 to 2013. Data analysis was conducted using SPSS version 19.0.

Results: The response rates were 81.1% (519/640) for eye care facilities, 96.1% (1645/1712) for ophthalmologists, and 67.6% (845/1250) for MLOP. From 2002-2003 to 2012-2013, there has been an increase in eye care facilities, from 234 to 519 (121.8%); ophthalmologists, from 935 to 1712 (83.1%); and MLOP, from 767 to 1250 (63%). The ophthalmologist:population ratio improved from 1:88,260 in 2002-2003 to 1:51,468 in 2012-2013. The MLOP:population ratio improved from 1:168,283 in 2002-2003 to 1:138,117 in 2012-2013 but still falls short of the ideal number.

Conclusions: Both southern Indian states are able to meet the requirements for ophthalmologists and eye care infrastructure as per the goals of VISION 2020. However, the number of MLOP falls short of the ideal ratio for the population. This study has some limitations. For example, most of the data collected through questionnaires were based on self-report, which might introduce bias due to memory recall or over or under-reporting of certain information. However, this was addressed by cross-checking the collected data with information from supplementary sources.

DOI: 10.2196/50921

Keywords
trends; human resources; infrastructure; eye care; South India
Introduction

Blindness and visual impairment represent a major public health problem in India [1–4]. The major causes of blindness and visual impairment in Andhra Pradesh and Telangana include cataract, refractive errors, retinal diseases, glaucoma, and corneal opacities, as reported in the Andhra Pradesh Eye Diseases study [5]. To tackle the problem of blindness and visual impairment, we need adequate human resources and sufficient infrastructure in eye care. Since the global “VISION 2020: the Right to Sight” initiative was launched in 1999, there has been a lot of progress in not only lessening the burden of blindness and visual impairment but also increasing the number of skilled eye care professionals and eye care infrastructure [6,7].

In line with the global Vision 2020 initiative, the undivided Andhra Pradesh state (the state was divided into Andhra Pradesh and Telangana states in 2014) established the Andhra Pradesh Right to Sight Society (APRTSS) in 2002 to work toward the VISION 2020 goals. Since its formation, the APRTSS has coordinated closely with major stakeholders in eye care such as those in the government, nongovernmental organization (NGO), and private sectors. Its activities include human resource development, infrastructural strengthening, disease control, and advocacy. To determine the impact of APRTSS VISION 2020 activities, we carried out a research project collecting information about the APRTSS activities from the baseline year of 2002 to 2003—the year in which the APRTSS was established—and the target year of 2012 to 2013—after a period of 10 years.

As part of the aforementioned research project, we carried out a survey about the ophthalmic workforce and infrastructure to identify the trends over a period of 10 years. An evidence base is essential to understand trends in human resources for health [8]. However, no regular mechanism exists in India to collect data on human resource trends in the provision of eye care services [9]. This study fills that gap by identifying trends in eye care. The results of the survey will be helpful to identify gaps, strengthen the eye care facilities, and overcome the maldistribution of human resources and infrastructure, in order to achieve the goals of VISION 2020. This study assessed trends in the availability and distribution of eye health professionals and eye care infrastructure in 2 southern Indian states: Andhra Pradesh and Telangana.

Methods

Study Design

This cross-sectional study used a pretested questionnaire for the year 2012 to 2013. The data for the 2002-2003 period were collected from available historical records.

We used questionnaires in both electronic and hard copy formats to collect the data. The questionnaires were developed based on the 6 building blocks of the universal health care system [10].

Ethical Considerations

This study was conducted as part of the research project on the “Impact of implementation of blindness control activities in the state of Andhra Pradesh,” which was approved by the ethics committee of the LV Prasad Eye Institute (reference number: LEC 09-13-097) and conducted in accordance with the tenets of the Declaration of Helsinki.

Definitions

For the purpose of this study, an eye care facility was defined as any health care facility where ophthalmologist services are available. The eye care facilities were identified as secondary or tertiary eye care facilities. For the purpose of this study, secondary eye care was defined as any eye care facility having an ophthalmologist conducting cataract and basic minor surgical procedures. Tertiary eye care was defined as any eye care facility with secondary eye care services as well as at least one subspecialty such as cornea, glaucoma, retina, or oculoplasty. Eye care facilities were categorized as government eye care facilities if they were established and funded by the government or other public sources such as universities and public sector organizations. NGO eye care facilities functioned on a no-profit, no-loss basis. Eye care facilities with a profit motive, irrespective of whether owned by an individual or a group of people or agencies, were categorized as private eye care facilities.

Inclusion Criteria

All eye care facilities with ≥10 inpatient beds or performing ≥100 cataract surgeries per annum were eligible.

Questionnaire

The questionnaire had 4 sections. Each section was distributed to concerned eye care professionals both in electronic form and hard copy to obtain the data.

Section 1: Questionnaire for Eye Care Facilities

The questionnaire for eye care facilities (Multimedia Appendix 1) was distributed to the director, superintendent, administrator, or manager in charge of the care facility. It was completed to obtain information for both the baseline and target years. It contained questions ranging from the services available, human resources, infrastructure, training facilities for eye care professionals, and any other relevant data.

Section 2: Questionnaire for Ophthalmologists

The questionnaire for ophthalmologists (Multimedia Appendix 2) was sent to all ophthalmologists working in government, NGO, and private eye care facilities. It was intended to be completed both by email and in hard copies by surface mail. It contained questions about demographic details; whether the ophthalmologist performs surgeries; whether the ophthalmologist practices in any subspecialties such as anterior segment surgeries, glaucoma, or retina; the average number of cataract surgeries per month; the principal method followed during cataract surgeries; professional experience; academic activity; and any training undergone.
Section 3: Questionnaire for MLOP

The questionnaire for MLOP (Multimedia Appendix 3) was distributed to all optometrists, ophthalmic assistants, and nurses working in all government, NGO, and private eye care facilities. It contained questions to elicit information on knowledge, skills, experience, and special training undergone such as in contact lens practice, refresher training in retinoscopy methods, and biomedical training for equipment maintenance. We also collected information on how many refractions were conducted per month, how many pairs of spectacles were prescribed per month, any administrative work, and any research activities.

Section 4: Questionnaire for District Blindness Control Societies and NGOs in Eye Care

The questionnaire for district blindness control societies (DBCSs) and NGOs in eye care (Multimedia Appendix 4) was distributed to program managers to obtain information on the impact of the implementation of blindness control activities in the district. It contained 3 subsections: section A for program managers of DBCS, section B for NGOs in eye care, and section C for international NGOs in eye care who were active in the state.

Follow-Up

Follow-up mechanisms were instituted every 2 weeks after mailing the questionnaire to the various stakeholders, and reminders were sent at the 3rd month and again at the 6th month.

Table 1. Response rates for eye care facilities, eye care professionals, and eye care organizations.

<table>
<thead>
<tr>
<th>Questionnaire recipient</th>
<th>Questionnaires distributed, n</th>
<th>Response rate, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye care facilities</td>
<td>640</td>
<td>519 (81.1)</td>
</tr>
<tr>
<td>Ophthalmologists</td>
<td>1712</td>
<td>1645 (96.1)</td>
</tr>
<tr>
<td>Midlevel ophthalmic personnel</td>
<td>1250</td>
<td>845 (67.6)</td>
</tr>
<tr>
<td>Local NGOs(^a)</td>
<td>182</td>
<td>165 (90.7)</td>
</tr>
<tr>
<td>International NGOs</td>
<td>10</td>
<td>9 (90)</td>
</tr>
<tr>
<td>DBCSs(^b)</td>
<td>23</td>
<td>23 (100)</td>
</tr>
</tbody>
</table>

\(^a\) NGOs: nongovernmental organizations.
\(^b\) DBCSs: district blindness control societies.

Eye Care Facilities and Service Delivery

The number of eye care facilities in the undivided state increased from 234 in 2002-2003 to 519 in 2012-2013 (121.8% increase). From 2002-2003 to 2012-2013, there was a marginal increase in the number of eye care facilities in the government sector (44 to 58, 31.8%), there was a substantial increase in the NGO sector (105 to 165, 57.1%), and the highest increase was seen in the private sector (85 to 296, 248.2%; Table 2).

Table 2. Number of eye care facilities in the combined state of Andhra Pradesh in 2002-2003 and 2012-2013.

<table>
<thead>
<tr>
<th>Type of facility</th>
<th>Facilities in 2002-2003, n</th>
<th>Facilities in 2012-2013, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government</td>
<td>44</td>
<td>58</td>
</tr>
<tr>
<td>NGO(^a)</td>
<td>105</td>
<td>165</td>
</tr>
<tr>
<td>Private</td>
<td>85</td>
<td>296</td>
</tr>
</tbody>
</table>

\(^a\) NGO: nongovernmental organization.

Additional Data Sources

In addition to the data collected through questionnaires, we gathered information from the following sources: (1) member directory for the All India Ophthalmological Society and its website, (2) directory of the Andhra Pradesh Ophthalmological Society and its website, (3) directory of the Telangana Ophthalmological Society and its website, (4) directory of the Andhra Pradesh Paramedical Board, and (5) websites of leading eye care institutions.

The information obtained from these sources helped us cross-check the data received through the questionnaires from eye care facilities, ophthalmologists, MLOP, and DBCSs. The data collected were entered in Excel sheets by 2 different data operators and cross-checked for any typographical errors. The data were analyzed using SPSS version 19.0 (IBM Corp) for Windows.

Results

Participants

As per the inclusion criteria, a total of 640 eye care facilities were identified, and a questionnaire was sent to the directors or those in charge of the facilities. Of the 640 facilities, responses were received from 519. Table 1 shows the number of questionnaires distributed to the various participants and the response rates. All the DBCSs responded to the questionnaire, whereas the lowest response rate was from MLOP.
to 79 in 2012-2013 (119.4% increase). The secondary and tertiary eye care facilities experienced a large jump in number from 2002-2003 to 2012-2013, whereas there was no increase in the number of tertiary eye care facilities in the government sector for the same period (Table 3).

Of 519 eye care facilities, 455 facilities (87.7%) were offering patient care services exclusively. Only 17% (88/519) of eye care facilities offered training facilities for eye care professionals and eye bank services in addition to patient care.

Regarding the eye care workforce, there was a substantial increase in the number of ophthalmologists in both southern Indian states. There was an insufficient increase in MLOP to meet the need. There was a large jump in the number of eye care managers, mostly in NGO and private eye care facilities (Table 4).

Table 3. Increase in secondary and tertiary eye care facilities from 2002 to 2012 by sector.

<table>
<thead>
<tr>
<th>Eye care facility sector</th>
<th>Facilities in 2002-2003, n</th>
<th>Facilities in 2012-2013, n</th>
<th>Increase, %</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Secondary</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Government</td>
<td>34</td>
<td>48</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>NGO</td>
<td>88</td>
<td>139</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Private</td>
<td>76</td>
<td>253</td>
<td>233</td>
<td></td>
</tr>
<tr>
<td>All secondary</td>
<td>198</td>
<td>440</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td><strong>Tertiary</strong></td>
<td></td>
<td></td>
<td></td>
<td>.009</td>
</tr>
<tr>
<td>Government</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>NGO</td>
<td>17</td>
<td>26</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Private</td>
<td>9</td>
<td>43</td>
<td>378</td>
<td></td>
</tr>
<tr>
<td>All tertiary</td>
<td>36</td>
<td>79</td>
<td>119</td>
<td></td>
</tr>
</tbody>
</table>

*aNGO: nongovernmental organization.

Table 4. Eye care workforce in the 2002-2012 period.

<table>
<thead>
<tr>
<th>Job role</th>
<th>Andhra Pradesh, n</th>
<th>Telangana, n</th>
<th>Both states, n</th>
<th>Increase, %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ophthalmologists</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor or senior consultant&lt;sup&gt;a&lt;/sup&gt;</td>
<td>132</td>
<td>288</td>
<td>146</td>
<td>338</td>
</tr>
<tr>
<td>Assistant professor or junior consultant&lt;sup&gt;b&lt;/sup&gt;</td>
<td>257</td>
<td>364</td>
<td>248</td>
<td>467</td>
</tr>
<tr>
<td>Ophthalmologists acting as superintendents or directors</td>
<td>69</td>
<td>148</td>
<td>83</td>
<td>107</td>
</tr>
<tr>
<td>All ophthalmologists</td>
<td>458</td>
<td>800</td>
<td>477</td>
<td>912</td>
</tr>
<tr>
<td><strong>Midlevel ophthalmic personnel (MLOP)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optometrists, refractionists, ophthalmic assistants, vision technicians</td>
<td>272</td>
<td>410</td>
<td>238</td>
<td>472</td>
</tr>
<tr>
<td>Ophthalmic nurses and general nurses working in eye care facilities</td>
<td>58</td>
<td>111</td>
<td>72</td>
<td>130</td>
</tr>
<tr>
<td>All MLOP</td>
<td>330</td>
<td>521</td>
<td>310</td>
<td>602</td>
</tr>
<tr>
<td>Eye care managers</td>
<td>69</td>
<td>163</td>
<td>83</td>
<td>244</td>
</tr>
</tbody>
</table>

<sup>a</sup>Ophthalmologists with ≥10 years of experience.
<sup>b</sup>Ophthalmologists with ≤10 years of experience.

The ophthalmologist:population ratio ranged from 1:6309 in Hyderabad district, which is the capital area, to 1:193,822 in Nalgonda district (Table 5). This shows there was a maldistribution of ophthalmologists among the districts in the state. The ratio of optometrists and allied personnel to the population ranged from 1:66,209 in Ranga Reddy district to 1:34,206 in Guntur district. Overall, the ophthalmologist:population ratio in the state was 1:49,404, which appears to be optimal as per the VISION 2020 guidelines.

We looked at the number of eye care beds available for the population, and this improved from an average of 1:17,457 in 2002-2003 to an average of 1:13,877 in 2012-2013 (Table 6). There was also a lot of variation in the availability of eye care beds among the districts; for example, in Hyderabad district, 1 eye care bed was available for 3805 persons, compared with 1 eye care bed for 30,014 persons in Karimnagar. The total number of eye care beds increased from 4339 in 2002-2003 to 6103 in 2012-2013 (40.6% increase). On average, 1
ophthalmologist was available per 100,000 people/6 eye care beds in 2002-2003, which increased to an average of 2 ophthalmologists per 100,000 people/7 eye care beds in 2012-2013. A greater number of ophthalmologists per 100,000 population will improve the accessibility and availability of ophthalmologists to the public.

Table 5. Human resources in eye care in the districts of undivided Andhra Pradesh.

<table>
<thead>
<tr>
<th>District name</th>
<th>Population, n</th>
<th>Ophthalmologists, n</th>
<th>MLOP(^a), n</th>
<th>MLOP: population ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adilabad</td>
<td>2,479,347</td>
<td>2,741,239</td>
<td>N/A(^d)</td>
<td>22</td>
</tr>
<tr>
<td>Hyderabad</td>
<td>3,686,460</td>
<td>3,943,323</td>
<td>N/A(^d)</td>
<td>625</td>
</tr>
<tr>
<td>Karim Nagar</td>
<td>3,477,079</td>
<td>3,776,269</td>
<td>N/A(^d)</td>
<td>42</td>
</tr>
<tr>
<td>Khammam</td>
<td>2,565,412</td>
<td>2,797,370</td>
<td>N/A(^d)</td>
<td>30</td>
</tr>
<tr>
<td>Mahbub Nagar</td>
<td>3,506,876</td>
<td>4,053,028</td>
<td>N/A(^d)</td>
<td>18</td>
</tr>
<tr>
<td>Medak</td>
<td>2,662,296</td>
<td>3,033,288</td>
<td>N/A(^d)</td>
<td>14</td>
</tr>
<tr>
<td>Nalgonda</td>
<td>3,238,449</td>
<td>3,488,809</td>
<td>N/A(^d)</td>
<td>18</td>
</tr>
<tr>
<td>Nizamabad</td>
<td>2,342,803</td>
<td>2,551,335</td>
<td>N/A(^d)</td>
<td>29</td>
</tr>
<tr>
<td>Ranga Reddy</td>
<td>3,506,670</td>
<td>5,296,741</td>
<td>N/A(^d)</td>
<td>99</td>
</tr>
<tr>
<td>Warangal</td>
<td>3,231,174</td>
<td>3,512,576</td>
<td>N/A(^d)</td>
<td>55</td>
</tr>
<tr>
<td>Anantapur</td>
<td>3,639,304</td>
<td>4,081,148</td>
<td>N/A(^d)</td>
<td>45</td>
</tr>
<tr>
<td>Chittoor</td>
<td>3,735,202</td>
<td>4,174,064</td>
<td>N/A(^d)</td>
<td>44</td>
</tr>
<tr>
<td>East Godavari</td>
<td>4,872,622</td>
<td>5,154,296</td>
<td>N/A(^d)</td>
<td>93</td>
</tr>
<tr>
<td>Guntur</td>
<td>4,405,521</td>
<td>4,887,813</td>
<td>N/A(^d)</td>
<td>41</td>
</tr>
<tr>
<td>Kadapa</td>
<td>2,573,481</td>
<td>2,882,469</td>
<td>N/A(^d)</td>
<td>23</td>
</tr>
<tr>
<td>Krishna</td>
<td>4,218,416</td>
<td>4,517,398</td>
<td>N/A(^d)</td>
<td>108</td>
</tr>
<tr>
<td>Kurnool</td>
<td>3,512,266</td>
<td>4,053,463</td>
<td>N/A(^d)</td>
<td>57</td>
</tr>
<tr>
<td>Nellore</td>
<td>2,659,661</td>
<td>2,963,557</td>
<td>N/A(^d)</td>
<td>58</td>
</tr>
<tr>
<td>Prakasam</td>
<td>3,054,941</td>
<td>3,397,448</td>
<td>N/A(^d)</td>
<td>47</td>
</tr>
<tr>
<td>Srikakulam</td>
<td>2,528,491</td>
<td>2,703,114</td>
<td>N/A(^d)</td>
<td>11</td>
</tr>
<tr>
<td>Visakhapatnam</td>
<td>3,789,823</td>
<td>4,290,589</td>
<td>N/A(^d)</td>
<td>171</td>
</tr>
<tr>
<td>Vizianagaram</td>
<td>3,789,823</td>
<td>2,344,474</td>
<td>N/A(^d)</td>
<td>15</td>
</tr>
<tr>
<td>West Godavari</td>
<td>3,796,144</td>
<td>3,936,966</td>
<td>N/A(^d)</td>
<td>47</td>
</tr>
<tr>
<td>All districts</td>
<td>7,572,754</td>
<td>8,458,077</td>
<td>858(^e)</td>
<td>1712</td>
</tr>
</tbody>
</table>

\(^a\)MLOP: midlevel ophthalmic personnel.
\(^c\)Census 2011 [12].
\(^d\)N/A: not available.
\(^e\)Approximate number from supplementary records.
### Table 6. Population and number of eye care beds by district.

<table>
<thead>
<tr>
<th>District name</th>
<th>Population, n</th>
<th>Eye care beds, n</th>
<th>Eye care bed:population ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adilabad</td>
<td>2,479,347</td>
<td>2,737,738</td>
<td>207</td>
</tr>
<tr>
<td>Hyderabad</td>
<td>3,686,460</td>
<td>4,010,238</td>
<td>855</td>
</tr>
<tr>
<td>Karimnagar</td>
<td>3,477,079</td>
<td>3,811,738</td>
<td>102</td>
</tr>
<tr>
<td>Khammam</td>
<td>2,565,412</td>
<td>2,798,214</td>
<td>97</td>
</tr>
<tr>
<td>Mahbub Nagar</td>
<td>3,506,876</td>
<td>4,042,191</td>
<td>144</td>
</tr>
<tr>
<td>Medak</td>
<td>2,662,296</td>
<td>3,031,877</td>
<td>87</td>
</tr>
<tr>
<td>Nalgonda</td>
<td>3,238,449</td>
<td>3,483,648</td>
<td>198</td>
</tr>
<tr>
<td>Nizamabad</td>
<td>2,342,803</td>
<td>2,552,073</td>
<td>142</td>
</tr>
<tr>
<td>Ranga Reddy</td>
<td>3,506,670</td>
<td>5,296,396</td>
<td>127</td>
</tr>
<tr>
<td>Warangal</td>
<td>3,231,174</td>
<td>3,934,842</td>
<td>242</td>
</tr>
<tr>
<td>Anantapur</td>
<td>3,639,304</td>
<td>4,083,315</td>
<td>182</td>
</tr>
<tr>
<td>Chittoor</td>
<td>3,735,202</td>
<td>4,170,468</td>
<td>123</td>
</tr>
<tr>
<td>East Godavari</td>
<td>4,872,622</td>
<td>5,151,549</td>
<td>192</td>
</tr>
<tr>
<td>Guntur</td>
<td>4,405,521</td>
<td>4,889,320</td>
<td>207</td>
</tr>
<tr>
<td>Kadapa</td>
<td>2,573,481</td>
<td>2,884,524</td>
<td>152</td>
</tr>
<tr>
<td>Krishna</td>
<td>4,218,416</td>
<td>4,529,009</td>
<td>102</td>
</tr>
<tr>
<td>Kurnool</td>
<td>3,512,266</td>
<td>4,046,601</td>
<td>107</td>
</tr>
<tr>
<td>Nellore</td>
<td>2,659,661</td>
<td>2,966,082</td>
<td>93</td>
</tr>
<tr>
<td>Prakasam</td>
<td>3,054,941</td>
<td>3,392,764</td>
<td>220</td>
</tr>
<tr>
<td>Srikakulam</td>
<td>2,528,491</td>
<td>2,699,471</td>
<td>147</td>
</tr>
<tr>
<td>Visakhapatnam</td>
<td>3,789,823</td>
<td>4,288,113</td>
<td>205</td>
</tr>
<tr>
<td>Vizianagaram</td>
<td>3,789,823</td>
<td>2,342,868</td>
<td>132</td>
</tr>
<tr>
<td>West Godavari</td>
<td>3,796,144</td>
<td>3,934,782</td>
<td>276</td>
</tr>
<tr>
<td>All districts</td>
<td>75,727,541</td>
<td>84,665,533</td>
<td>4339</td>
</tr>
</tbody>
</table>

### Discussion

#### Principal Findings

Estimates indicate there are 4.95 million people who are blind (0.36% of the total population), 35 million people who are visually impaired (2.55%), and 0.24 million children who are blind in India [13]. Cataract and refractive errors remain the major causes of blindness and visual impairment, respectively, in India [13-16]. Cataract is responsible for nearly two-thirds of the blindness load in the older population in India [1-4], and one-fifth of blindness is due to uncorrected refractive errors [1-3]. There have been significant improvements in the field of blindness prevention, management, and control since the “VISION 2020: The Right to Sight” initiative [17]. In view of this background, India needs a pool of well-qualified, skilled, and optimal eye care professionals and sufficient infrastructure to eliminate avoidable and needless blindness and visual impairment.

The global advisory committee for VISION 2020 recommended a set of criteria for human resources and infrastructure based on expert consensus of the number of cataract procedures that could be performed by a surgeon per year under optimal conditions and the number of beds required for the same per 1 million population [9]. It was assumed that at least 50 procedures per bed per year could be optimally performed. Based on these assumptions, the following norms were recommended: 1 ophthalmologist per 50,000 population, 1 MLOP per 50,000 population, and 1 eye care bed per 20,000 population.

In this study, the ophthalmologist:population ratio in 2002-2003 was 1:88,822, and in 2012-2013, it reached 1:51,416. The state had almost reached the optimal ophthalmologist:population ratio. Previous data show that the national average ophthalmologist:population ratio is 1:107,000, ranging from 1:9000 in some regions to 1:608,000 in some areas [9]. There was a decrease in the percentage of ophthalmologists in the government sector and virtually no change in the percentage of ophthalmologists in the NGO sector. In addition, there was a substantial increase in the number of ophthalmologists in the private sector from 2002-2003 to 2012-2013. Some of the ophthalmologists, who were mainly working in the private sector, offered their services for a few hours a day or 1 to 2 days a week to NGO eye care facilities, either free or for a fee.
per our study definition, these ophthalmologists who were providing their services part-time for the NGO eye care facilities were treated as working in the private sector only. Hence, the number of ophthalmologists working in the NGO sector appears to be under-reported when compared with that of other sectors.

As per VISION 2020, there should be 20 ophthalmologists and 50 beds per 1 million population [18]. The importance of the ophthalmologist:population ratio is that it can serve as a guide to forecast ophthalmic manpower requirements [19]. As per the norm, the number of available eye care beds is sufficient, and there is no need to increase the number of eye care beds; in addition, there is a shift toward day surgeries for cataract [8].

The distribution of ophthalmologists was skewed toward urban areas. Due to the lack of educational facilities for their children and other lifestyle-related infrastructure in underdeveloped areas, ophthalmologists and private eye care facilities tend to be established in developed urban areas. In the Telangana region, the majority of the ophthalmologists were practicing in Hyderabad City, whereas in coastal Andhra, many of the ophthalmologists were practicing in the urban areas of Visakhapatnam and Vijayawada. Compared with the coastal Andhra region, this phenomenon of ophthalmologists working in urban areas was more pronounced in the Telangana region. As urban areas became more crowded with ophthalmologists, there was a trend that some ophthalmologists started their practices in smaller towns in 2012-2013. In 2002-2003, ophthalmologists were mainly present in the district headquarters and major population areas. This trend changed in 2012-2013 when more eye care facilities were opened in less populated areas.

Murthy et al [20] reported that 69% of ophthalmologists worked in the private and NGO sectors, while 31% were working in the government sector. In this study, 88% of ophthalmologists were working in the private and NGO sectors, and the remaining 12% were working in the government sector. In this study, the majority of the ophthalmologists in the government sector were working in teaching institutions rather than in district and subdistrict hospitals similar to that reported by Murthy et al [20]. In this study, we found the average number of surgeries performed by surgeons in the NGO sector was significantly higher than that in other sectors in both the baseline and target years. After the ophthalmologists in the NGO sector, ophthalmologists in the government sector were performing more surgeries than those in the private sector.

Ophthalmologists with less than 10 years of experience were performing more cataract surgeries than those with more than 10 years of experience (P=.001). This may be because some of the senior ophthalmologists were involved in teaching and research. This finding corroborates the fact that nonteaching ophthalmologists were performing more cataract surgeries than their teaching counterparts.

The state should ideally have 1693 MLOP for its population of 84.6 million. The state needs 1080 more MLOP to reach this number. The majority of the MLOP either were not trained in streak retinoscopy or did not have access to streak retinoscopes. There is a need for a strategy to ensure that all MLOP can perform streak retinoscopy.

There were many reasons for the increase in the number of both secondary and tertiary eye care facilities in all 3 sectors—government, NGO, and private—from 2002-2003 to 2012-2013. The number of eye care facilities as well as the number of eye care professionals increased during this period. The highest increase in eye care facilities (248%) was seen in the private sector due to the establishment of many institutions for eye care professionals in both government and NGO sectors. People trained at these institutes either were absorbed into the private sector or started their own practice, because there was no recruitment in the government sector or minimal opportunities in the NGO sector. This is the reason why the number of secondary eye care facilities increased more than tertiary eye care facilities. Another reason was, compared with other fields in medical practice, it is easier to start a solo practice in eye care, as it does not depend on cooperation from other medical streams. For example, to start a general surgery or orthopedics practice, one requires the services of an anesthetist. To start a pediatric practice, good laboratory services are required. Of the 519 eye care facilities functioning in 2012-2013, 253 (48.7%) were from the private sector. This was similar to the findings reported by Murthy et al [1], in which more than one-half of the eye care facilities belonged to the private sector.

Limitations
This study has some limitations. Most of the data collected through questionnaires were based on self-report, which might introduce bias due to memory recall or over or under-reporting of certain information. However, this was addressed by cross-checking the collected data with information from the supplementary sources mentioned in the Methods section.

Conclusion
Regarding human resources, there was a substantial increase in the number of ophthalmologists, particularly in the private sector. In fact, the percentage of ophthalmologists in the government sector decreased from the baseline year to the target year, whereas in the NGO sector, it remained the same.

Though all 3 sectors—government, NGO, and private—showed an increase in the number of eye care facilities from the baseline year to the target year, substantial increases were seen in the private sector and, to some extent, in the NGO sector. Most of the eye care facilities offered patient care services only. The outpatient services and inpatient services were also higher in 2012-2013 in all 3 sectors, but the NGO sector contributed a major share, followed by the private sector. Regarding outreach activities, the NGO sector dominated the services, to the extent of 80%-97%. One NGO facility collected the majority of eyes for corneal transplantation, and the remaining eye care facilities in the government, NGO, or private sector showed very little improvement in their collection of eyes.

Regarding eye care infrastructure, there was a 41% increase in the number of beds available for eye care, and this increase was mainly due to the NGO sector, followed by the private sector. The average number of surgeries per surgeon per annum was highest in the NGO sector, followed by the government sector. There was a major shortage of MLOP in the state to attain the ideal ratio of 1 MLOP per 50,000 population. To attain the ideal
number of MLOP, there is an urgent need to increase the number of training facilities for MLOP. Overall, the functioning of the DBCSs for planning and supervising district eye care programs was satisfactory.

Acknowledgments
The authors would like to show appreciation to all the eye care professionals and eye care administrators who readily provided the information required for this mammoth study. We would also like to acknowledge the support received from the Government of Andhra Pradesh Health Department in giving access to the records necessary for our study.

This research study received financial support from the Government of Andhra Pradesh, India.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Questionnaire for eye care facilities.
[XLSX File (Microsoft Excel File), 658 KB, - ojphi_v16i1e50921_app1.xlsx]

Multimedia Appendix 2
Questionnaire for ophthalmologists.
[XLSX File (Microsoft Excel File), 485 KB, - ojphi_v16i1e50921_app2.xlsx]

Multimedia Appendix 3
Questionnaire for midlevel ophthalmic personnel.
[XLSX File (Microsoft Excel File), 494 KB, - ojphi_v16i1e50921_app3.xlsx]

Multimedia Appendix 4
Questionnaire for district blindness control societies (DBCSs) and nongovernmental organizations (NGOs) in eye care.
[XLSX File (Microsoft Excel File), 463 KB, - ojphi_v16i1e50921_app4.xlsx]

References
Electronic Health Records for Population Health Management: Comparison of Electronic Health Record–Derived Hypertension Prevalence Measures Against Established Survey Data

Katie S Allen1, BS; Nimish Valvi1, PhD; P Joseph Gibson2, PhD; Timothy McFarlane2, PhD; Brian E Dixon1,2, PhD

1Regenstrief Institute, Inc, Indianapolis, IN, United States
2Richard M Fairbanks School of Public Health, Indiana University–Purdue University Indianapolis, Indianapolis, IN, United States

Corresponding Author:
Katie S Allen, BS
Regenstrief Institute, Inc
1101 West 10th Street
Indianapolis, IN, 46202
United States
Phone: 1 3172749024
Email: allenkat@regenstrief.org

Abstract

Background: Hypertension is the most prevalent risk factor for mortality globally. Uncontrolled hypertension is associated with excess morbidity and mortality, and nearly one-half of individuals with hypertension do not have the condition under control. Data from electronic health record (EHR) systems may be useful for community hypertension surveillance, filling a gap in local public health departments’ community health assessments and supporting the public health data modernization initiatives currently underway. To identify patients with hypertension, computable phenotypes are required. These phenotypes leverage available data elements—such as vitals measurements and medications—to identify patients diagnosed with hypertension. However, there are multiple methodologies for creating a phenotype, and the identification of which method most accurately reflects real-world prevalence rates is needed to support data modernization initiatives.

Objective: This study sought to assess the comparability of 6 different EHR-based hypertension prevalence estimates with estimates from a national survey. Each of the prevalence estimates was created using a different computable phenotype. The overarching goal is to identify which phenotypes most closely align with nationally accepted estimations.

Methods: Using the 6 different EHR-based computable phenotypes, we calculated hypertension prevalence estimates for Marion County, Indiana, for the period from 2014 to 2015. We extracted hypertension rates from the Behavioral Risk Factor Surveillance System (BRFSS) for the same period. We used the two 1-sided t test (TOST) to test equivalence between BRFSS- and EHR-based prevalence estimates. The TOST was performed at the overall level as well as stratified by age, gender, and race.

Results: Using both 80% and 90% CIs, the TOST analysis resulted in 2 computable phenotypes demonstrating rough equivalence to BRFSS estimates. Variation in performance was noted across phenotypes as well as demographics. TOST with 80% CIs demonstrated that the phenotypes had less variance compared to BRFSS estimates within subpopulations, particularly those related to racial categories. Overall, less variance occurred on phenotypes that included vitals measurements.

Conclusions: This study demonstrates that certain EHR-derived prevalence estimates may serve as rough substitutes for population-based survey estimates. These outcomes demonstrate the importance of critically assessing which data elements to include in EHR-based computer phenotypes. Using comprehensive data sources, containing complete clinical data as well as data representative of the population, are crucial to producing robust estimates of chronic disease. As public health departments look toward data modernization activities, the EHR may serve to assist in more timely, locally representative estimates for chronic disease prevalence.

DOI: 10.2196/48300

Keywords:
public health informatics; surveillance; chronic conditions; electronic health record; health management; hypertension; surveillance; public health; prevalence; population-based survey
Introduction

Hypertension is the most prevalent risk factor for mortality throughout the world [1]. The condition is characterized by elevated systolic blood pressure (>140 mm Hg) or diastolic blood pressure (>90 mm Hg) [2]. An estimated 1 out of 3 adults in the United States has been diagnosed with hypertension, which translates to almost 75 million Americans [2]. This results in substantial use of health care services and medications, as well as lost wages [3,4]. The estimated direct and indirect costs of hypertension exceed US $48 billion each year in the United States [5]. In concurrence with economic loss, uncontrolled hypertension is associated with excess morbidity and mortality, and nearly one-half of individuals with hypertension do not have the condition under control [2].

Uncontrolled hypertension is associated with an increased risk of coronary heart disease, stroke, and kidney disease, which are the 3 leading causes of death in the United States [5]. Hypertension is a comorbid condition for nearly 70% of individuals who have their first myocardial infarction and almost 80% of those who have their first stroke [6]. Additionally, hypertension is associated with an excess risk of severe COVID-19 illness with a risk of hospitalized more than double that of nonhypertensive individuals [7]. The association with increased morbidity and mortality is a critical public health concern given the high prevalence of the condition. To address this increasing public health concern, public health programs and policies aimed at reducing morbidity, mortality, and costs associated with hypertension are required. To create these policies, public health departments are reliant on timely, accurate, stable estimates of disease prevalence. This is required both for timely detection and effective evaluation.

Identifying the prevalence of hypertension as well as measuring hypertension control at the community level remains a challenge for local health departments. While clinical guidelines from the National Quality Forum and others (eg, Centers for Disease Control and Prevention and Healthcare Effectiveness Data and Information Set) exist [8], measurement happens at the level of a provider or health system as opposed to the community. Public health departments typically rely on surveys for measuring community-level estimates of hypertension. However, surveys have known limitations including cost and timeliness due to long gaps between data collection and when results are available. Additionally, the local samples are insufficiently small for precise estimates within communities and subpopulations (ie, wide CIs). Therefore, local health departments seek alternative methods for obtaining timely, complete, accurate, and precise information about the prevalence of chronic conditions such as hypertension and measures of control for individuals with chronic illness.

Since the passing of the Health Information Technology for Economic and Clinical Health Act of 2009, electronic health record (EHR) systems have become more common, representing a potential data source for chronic disease surveillance. As of 2016, over 70% of ambulatory providers use EHR systems [9]. As health care organizations increasingly capture data from routine health care visits in EHR systems, national initiatives, including the digital Learning Health System of the National Academy of Medicine [10] and the Robert Wood Johnson Foundation’s data for health [11], aim to leverage such data to improve the delivery of health care and community health outcomes. The hope is that by leveraging existing digital data sources, public health agencies may access more timely and precise information to assess and improve health in their communities.

While there exists much optimism about EHR systems’ ability to provide timely, complete, and accurate estimates for hypertension and other chronic diseases, evidence to date has been mixed. In a systematic review of the quality of data used for quality-of-care measurement, the completeness of data varied “substantially across studies,” ranging from 0.1% to 51% for blood pressure and from 10% to 38% for smoking status [12]. Missing data ranged between 24% and 38% for cholesterol; 3% and 31% for blood pressure; and 5% and 23% for blood glucose (hemoglobin A1c) [12].

Despite these challenges, EHR data may be useful for community health surveillance. More recent work by the New York City (NYC) Department of Mental Health and Hygiene shows promising results in using EHR data for measuring the prevalence and control of chronic diseases [13,14]. By querying EHR systems in primary care practices representing 15% of the city’s population, the health department found prevalence rates were in line with community-based surveys for diabetes, obesity, hypertension, and smoking even when the survey respondents were limited to those who had received primary care in the prior year (NYC Health and Nutrition Examination Survey and the NYC Community Health Survey [15]). More recent studies give hope that EHR data could be used by health departments to improve the timeliness and precision of their community health assessments [16-18].

Given limited prior evidence, we sought to validate computable phenotypes for hypertension using EHR data available through a community-based health information exchange (HIE) network. The use of HIE data was selected to examine data representing a geographic community rather than the population of a single health system. Our goal is to identify methods that can be leveraged by health departments for the surveillance of chronic illnesses and the calculation of control measures.

Accordingly, the objective of this analysis was to analyze the equivalence of EHR-based methods for deriving the prevalence of hypertension compared to an established community survey. To facilitate this analysis, 6 distinct EHR-based phenotypes for hypertension were used to establish prevalence rates in 1 county. These rates were then tested for equivalency with the prevalence calculated by a national survey. We hypothesized that at least 1 of the selected phenotypes would produce equivalent estimates.
Methods

Data Sources

Indiana Network for Patient Care

The primary data source was the Indiana Network for Patient Care (INPC), a regional HIE with data covering emergency department visits, hospital admissions, and large outpatient health care clinics from across the state. Data were supplemented with direct extracts from 1 health system to provide additional vital measurements and medication data that were not currently shared with the INPC. For this study, the focus was Marion County, Indiana, which is the county containing the largest city, Indianapolis, and we leveraged 3 of the 5 major health systems. Using the 3 health systems ensures that approximately 780,000 (80%) of the population of Marion County was captured for this study. According to the 2010 census, Marion County had a resident population of 977,203 with a racial composition of 30% Black or African American, 11.6% Hispanic, and 61.9% White.

Data were extracted for all adults (at least aged 18 years as of January 1, 2014) living in Marion County who sought care (outpatient, inpatient, or emergency department encounters) at 1 of the 3 large integrated delivery networks that connect to the INPC between January 1, 2014, and December 31, 2015. We used 2 years of data to capture a representative number of clinical encounters since individual health care use may not occur annually. This period was used due to the availability of comprehensive data from 3 of the 5 major health systems in the area. Given the period covered in this data set, the data do not establish current prevalence rates for Marion County but rather serve as an example for the surveillance methodology deployed. The algorithms to detect hypertension in the community were implemented on the data set, which contained diagnosis codes, vital measurements, and medications.

Behavioral Risk Factor Surveillance System

For the gold standard comparison, we used the Behavioral Risk Factor Surveillance System (BRFSS)—the US national survey related to health-related behaviors, chronic health conditions, and the use of preventive services. The prevalence estimates produced by the BRFSS are carefully developed, validated, and weighted to minimize biases in response or coverage [19]. The BRFSS collects data in all 50 states, the District of Columbia, and territories. However, for small geographics (eg, county) or population subgroups, the BRFSS is imprecise with large CIs. For this study, the data related to the 2015 prevalence of hypertension in Marion County, Indiana, was used.

Measures

To facilitate analysis, BRFSS prevalence measures were compared to EHR-based measures extracted from the HIE. The 2015 BRFSS results include an overall hypertension prevalence rate as well as rates by age, race, and gender for Marion County. These measures were extracted from the US Centers for Disease Control and Prevention website [20].

The computable phenotypes used for this study were previously developed and reported separately [21]. Briefly, 6 phenotypes for hypertension were developed using algorithms (or rules) executed using 1 or more types of structured EHR data. These rules were validated using chart review to calculate sensitivity, specificity, and positive predictive value [21]. Defining multiple permutations allowed for evaluating the best-performing phenotype. The phenotypes are as follows:

- P1: clinical diagnostic codes only (in which an individual has either 1 inpatient or 1 outpatient encounter documenting a hypertension diagnosis)
- P2: vital statistics only (in which an individual has at least 1 blood pressure reading above the hypertension threshold)
- P3: vital statistics only (in which an individual has at least 2 blood pressure readings above the hypertension threshold)
- P4: clinical diagnosis and vital statistics (P1 and P2)
- P5: clinical diagnosis and vital statistics (P1 and P3)
- P6: Inclusive of P1-P5 and medications (P1, P2, or the use of hypertension medication)

Using the 6 different EHR-based computable phenotypes, we calculated hypertension prevalence estimates from data for residents of Marion County, Indiana, from the years 2014 and 2015. Prevalence was calculated as the number of persons with data satisfying the given phenotype divided by the number of persons with any HIE record for a health care encounter.

Ethical Considerations

Exempt approval for this study was received by the Indiana University Institutional Review Board (1701925087).

Statistical Analysis

Demographics for the INPC-derived cohort were calculated using P6, which is the broadest and most sensitive phenotype [21]. Using the estimates for Marion County outlined above, equivalency testing was performed. Equivalence testing examines whether 2 independent statistics are similar enough to be treated as though they are equivalent. The null hypothesis is that the statistics differ by at least a specified amount. If the test results in a P value <.05, then the null hypothesis is rejected with a conclusion that the 2 statistics differ by less than the specified amount. We used the two 1-sided t test (TOST) to test equivalence between BRFSS- and INPC-based prevalence estimates. The TOST was performed at the overall level as well as stratified by age, gender, and race. The TOST was performed with 80% and 90% CI. As with other large national surveys, BRFSS estimates have wide CIs. Accordingly, widening the TOST analysis threshold was considered to account for the wide CIs within the BRFSS data set compared to the small CIs associated with the larger INPC data set. The 95% CI of the BRFSS overall hypertension estimates for Marion County is 7-7.5 percentage points wide. The stratified BRFSS hypertension rates are slightly wider. Accordingly, our specified amounts align with the CIs for the BRFSS. This study used SAS (version 9.4; SAS Institute Inc) and Excel 365 (Microsoft) for analyses.

Results

The demographics for the BRFSS and INPC cohorts are presented in Table 1. The EHR-based phenotypes were calculated from INPC data for 548,232 patients, which was the number of adult patients with at least 1 clinical encounter during
the period. Overall, the cohort was 61.2% (n=335,548) women and 27% (n=148,117) Black or African American. Of the total INPC-derived cohort, 210,764 (38.4%) patients were identified as hypertensive by phenotype P6, which is the broadest— and most sensitive— definition of hypertension according to Valvi et al [21]. The INPC-derived hypertension cohort was 57.6% (121,307/210,764) women and 33.2% (70,060/210,764) Black or African American. The BRFSS-derived hypertensive cohort was 55.2% (197/357) women and 17.6% (63/357) Black or African American. The INPC cohort was more racially diverse than the BRFSS cohort overall. The BRFSS cohort had less representation of the younger population and overrepresentation of those aged 65 years and older.

Table 1. Cohort demographics.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Overall population</th>
<th>Hypertensive population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BRFSSb (n=934), n (%)</td>
<td>INPCc (n=548,232), n (%)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>524 (56.1)</td>
<td>335,548 (61.2)</td>
</tr>
<tr>
<td>Men</td>
<td>410 (43.9)</td>
<td>212,684 (38.8)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>152 (16.7)</td>
<td>148,117 (27)</td>
</tr>
<tr>
<td>White</td>
<td>702 (75.2)</td>
<td>308,213 (56.2)</td>
</tr>
<tr>
<td>Other</td>
<td>80 (8.6)</td>
<td>91,902 (16.8)</td>
</tr>
<tr>
<td>Age group (y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-39</td>
<td>197 (21.1)</td>
<td>214,655 (39.2)</td>
</tr>
<tr>
<td>40-64</td>
<td>406 (43.5)</td>
<td>240,064 (43.8)</td>
</tr>
<tr>
<td>65+</td>
<td>331 (35.4)</td>
<td>93,513 (17)</td>
</tr>
</tbody>
</table>

aTable 1 contains gender, race, and age counts and percentages for each of the cohorts. The cohorts include the overall population for both BRFSS and INPC as well as the hypertensive population.
bBRFSS: Behavioral Risk Factor Surveillance System.
cINPC: Indiana Network for Patient Care.

The TOST analysis was undertaken at both the 90% and 80% CIs. The TOST analysis at the 90% CI resulted in 2 phenotypes (P2 and P5) having statistically significant results, indicating their equivalency to BRFSS estimates, or, more specifically, given the assumptions of this analysis, it is at least 90% likely that hypertension prevalence estimates from the BRFSS and phenotypes P2 and P5 will differ by no more than 5 percentage points. However, performance in the stratified groups was much poorer with statistical significance for women only in phenotypes P1 and P4. By the nature of TOST, the wider an estimate’s CI, the less chance that the null hypothesis will be rejected; some stratified groups have CIs so wide that their TOSTs had zero power. The analysis at the 80% CI yielded statistically significant results across multiple phenotypes. At the 80% CI, phenotypes P2, P3, and P5 showed equivalency overall, with P2 and P5 also showing equivalence in 9 of the demographic subsets and P3 showing equivalence in 7 of those subsets. Tables 2-4 depict the full 80% CI analysis for P2, P3, and P5. All remaining analyses are included in the Multimedia Appendices 1 and 2.
Table 2. Full 80% CI analysis for phenotype 2, with overall ≥1 vitals indicated. This table depicts all analytical results for P2 at the 80% CI.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>BRFSS&lt;sup&gt;a,b&lt;/sup&gt;, n/N (%)</th>
<th>INPC&lt;sup&gt;c,d&lt;/sup&gt;, n/N (%)</th>
<th>%Δ&lt;sup&gt;e&lt;/sup&gt; (Δ80% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>235/934 (28.4)</td>
<td>159,330/548,298 (29.1)</td>
<td>0.7 (–1.8 to 3.1)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>127/410 (31)</td>
<td>66,758/212,684 (31.4)</td>
<td>0.4 (–10.6 to 11.4)</td>
</tr>
<tr>
<td>Women</td>
<td>137/524 (26.1)</td>
<td>92,570/335,548 (27.6)</td>
<td>1.5 (–6.6 to 9.6)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>54/152 (35.7)</td>
<td>57,026/148,120 (38.5)</td>
<td>2.8 (–3.3 to 8.9)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>White</td>
<td>187/702 (26.6)</td>
<td>89,205/308,224 (28.9)</td>
<td>2.3 (–0.3 to 5)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Other</td>
<td>18/80 (22.6)</td>
<td>13,099/91,954 (14.2)</td>
<td>–8.4 (–15 to –1.7)</td>
</tr>
<tr>
<td>Age group (y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-39</td>
<td>21/197 (10.8)</td>
<td>49,634/214,685 (23.1)</td>
<td>12.3 (9.2 to 15.4)</td>
</tr>
<tr>
<td>40-64</td>
<td>133/406 (32.8)</td>
<td>76,795/240,084 (32)</td>
<td>–0.8 (–4.5 to 2.9)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>65+</td>
<td>204/331 (61.6)</td>
<td>31,238/88,569 (35.3)</td>
<td>–26.3 (–30 to –22.6)</td>
</tr>
<tr>
<td>Men by race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>24/60 (40.6)</td>
<td>22,226/56,004 (39.7)</td>
<td>–0.9 (–7.1 to 5.2)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>White</td>
<td>91/314 (29.1)</td>
<td>38,832/120,672 (32.2)</td>
<td>3.1 (–1 to 7.2)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Other</td>
<td>9/36 (24.1)</td>
<td>5,700/36,008 (15.8)</td>
<td>–8.3 (–18.1 to 1.6)</td>
</tr>
<tr>
<td>Women by race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>30/92 (32.2)</td>
<td>34,800/92,113 (37.8)</td>
<td>2.5 (–0.9 to 5.8)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>White</td>
<td>95/388 (24.4)</td>
<td>50,373/187,541 (26.9)</td>
<td>–7.6 (–16.3 to 1.1)</td>
</tr>
<tr>
<td>Other</td>
<td>9/44 (20.8)</td>
<td>7,379/55,894 (13.2)</td>
<td>5.6 (–2.4 to 13.5)</td>
</tr>
<tr>
<td>Men by age group (y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-39</td>
<td>18/99 (18.5)</td>
<td>20,478/77,992 (26.3)</td>
<td>7.8 (2.3 to 13.3)</td>
</tr>
<tr>
<td>40-64</td>
<td>56/178 (31.2)</td>
<td>33,928/98,778 (34.3)</td>
<td>3.1 (–2.2 to 8.5)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>65+</td>
<td>90/133 (67.4)</td>
<td>11,957/34,606 (34.6)</td>
<td>–32.8 (–38.6 to –27.1)</td>
</tr>
<tr>
<td>Women by age group (y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-39</td>
<td>3/98 (3.4)</td>
<td>29,155/136,663 (21.3)</td>
<td>11.5 (15.6 to 20.2)</td>
</tr>
<tr>
<td>40-64</td>
<td>78/228 (34.2)</td>
<td>42,866/141,286 (30.3)</td>
<td>–3.9 (–9.1 to 1.4)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>65+</td>
<td>114/198 (57.5)</td>
<td>19,281/53,954 (35.7)</td>
<td>–21.8 (–29.1 to –14.5)</td>
</tr>
</tbody>
</table>

<sup>a</sup>BRFSS: Behavioral Risk Factor Surveillance System.
<sup>b</sup>Sample size=934.
<sup>c</sup>Indiana Network for Patient Care.
<sup>d</sup>Sample size=548,298.
<sup>e</sup>Δ: mean difference.
<sup>f</sup>Behavioral Risk Factor Surveillance System and Indiana Network for Patient Care phenotypes were determined as statistically equivalent by the two 1-sided t test method.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>BRFSS(^a,b), n/N (%)</th>
<th>INPC(^c,d), n/N (%)</th>
<th>%Δ(^e) (Δ80% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>235/934 (28.4)</td>
<td>122,051/548,298 (22.3)</td>
<td>−6.1 (−8.6 to −3.7)(^f)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>127/410 (31)</td>
<td>50,997/212,684 (24)</td>
<td>−7 (−18 to 4)</td>
</tr>
<tr>
<td>Women</td>
<td>137/524 (26.1)</td>
<td>71,053/335,548 (21.2)</td>
<td>−4.9 (−13 to 3.1)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>54/152 (35.7)</td>
<td>45,513/148,120 (30.7)</td>
<td>−5 (−11.1 to 1.2)</td>
</tr>
<tr>
<td>White</td>
<td>187/702 (26.6)</td>
<td>67,594/308,224 (21.9)</td>
<td>−4.7 (−7.4 to −2)(^f)</td>
</tr>
<tr>
<td>Other</td>
<td>18/80 (22.6)</td>
<td>8,944/91,954 (9.7)</td>
<td>−12.9 (−19.5 to −6.2)</td>
</tr>
<tr>
<td><strong>Age group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-39</td>
<td>21/197 (10.8)</td>
<td>34,282/214,685 (16)</td>
<td>5.2 (2.1 to 8.2)(^f)</td>
</tr>
<tr>
<td>40-64</td>
<td>133/406 (32.8)</td>
<td>60,657/240,084 (25.3)</td>
<td>−7.5 (−11.2 to −3.8)</td>
</tr>
<tr>
<td>65+</td>
<td>204/331 (61.6)</td>
<td>25,699/88,569 (29)</td>
<td>−32.6 (−36.3 to −28.9)</td>
</tr>
<tr>
<td><strong>Men by race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>24/60 (40.6)</td>
<td>17,678/56,004 (31.6)</td>
<td>−9 (−15.2 to −2.9)</td>
</tr>
<tr>
<td>White</td>
<td>91/314 (29.1)</td>
<td>29,448/120,672 (24.4)</td>
<td>−4.7 (−8.8 to −0.6)(^f)</td>
</tr>
<tr>
<td>Other</td>
<td>9/36 (24.1)</td>
<td>3,871/36,008 (10.8)</td>
<td>−13.3 (−23.2 to −3.5)</td>
</tr>
<tr>
<td><strong>Women by race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>30/92 (32.2)</td>
<td>27,835/92,113 (20.3)</td>
<td>−4.1 (−7.4 to −0.7)(^f)</td>
</tr>
<tr>
<td>White</td>
<td>95/388 (24.4)</td>
<td>38,146/187,541 (9.1)</td>
<td>−11.7 (−20.4 to −3)</td>
</tr>
<tr>
<td>Other</td>
<td>9/44 (20.8)</td>
<td>5,072/55,894 (30.2)</td>
<td>−2 (−9.9 to 6)(^f)</td>
</tr>
<tr>
<td><strong>Men by age group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-39</td>
<td>18/99 (18.5)</td>
<td>13,875/77,992 (17.8)</td>
<td>−0.7 (−6.2 to 4.8)(^f)</td>
</tr>
<tr>
<td>40-64</td>
<td>56/178 (31.2)</td>
<td>27,100/98,778 (27.4)</td>
<td>−3.8 (−9.1 to 1.6)(^f)</td>
</tr>
<tr>
<td>65+</td>
<td>90/133 (67.4)</td>
<td>9,694/34,606 (28)</td>
<td>−39.4 (−45.1 to −33.6)</td>
</tr>
<tr>
<td><strong>Women by age group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-39</td>
<td>3/98 (3.4)</td>
<td>20,407/136,663 (14.9)</td>
<td>11.5 (9.2 to 13.8)</td>
</tr>
<tr>
<td>40-64</td>
<td>78/228 (34.2)</td>
<td>33,556/141,286 (23.8)</td>
<td>−10.4 (−15.7 to −5.2)</td>
</tr>
<tr>
<td>65+</td>
<td>114/198 (57.5)</td>
<td>16,005/53,954 (29.7)</td>
<td>−27.8 (−35.1 to −20.5)</td>
</tr>
</tbody>
</table>

\(^a\) BRFSS: Behavioral Risk Factor Surveillance System.
\(^b\) Sample size=934.
\(^c\) INPC: Indiana Network for Patient Care.
\(^d\) Sample size=548,298.
\(^e\) Δ: mean difference.
\(^f\) Behavioral Risk Factor Surveillance System and Indiana Network for Patient Care phenotypes were determined as statistically equivalent by the two 1-sided t test method.
Table 4. Phenotype 5, overall ≥1 clinical diagnosis or ≥1 vitals indicated. This table depicts the full analytical results for P5 at the 80% CI.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>BRFSS&lt;sup&gt;a,b&lt;/sup&gt;, n/N (%)</th>
<th>INPC&lt;sup&gt;c,d&lt;/sup&gt;, n/N (%)</th>
<th>%Δ&lt;sup&gt;e&lt;/sup&gt; (Δ80% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>235/934 (28.4)</td>
<td>151,645/548,298 (27.7)</td>
<td>−0.7 (−3.2 to 1.7)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>127/410 (31)</td>
<td>63,992/212,684 (30.1)</td>
<td>−0.9 (−11.9 to 10.1)</td>
</tr>
<tr>
<td>Women</td>
<td>137/524 (26.1)</td>
<td>87,652/335,548 (26.1)</td>
<td>0 (−8 to 8.1)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>54/152 (35.7)</td>
<td>71,464/148,120 (48.2)</td>
<td>12.5 (6.4 to 18.7)</td>
</tr>
<tr>
<td>White</td>
<td>187/702 (26.6)</td>
<td>137,674/308,224 (44.7)</td>
<td>18.1 (15.4 to 20.8)</td>
</tr>
<tr>
<td>Other</td>
<td>18/80 (22.6)</td>
<td>31,158/91,954 (33.9)</td>
<td>11.3 (4.6 to 17.9)</td>
</tr>
<tr>
<td>Age group (y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-39</td>
<td>21/197 (10.8)</td>
<td>36,157/214,685 (16.8)</td>
<td>6 (3 to 9.1)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>40-64</td>
<td>133/406 (32.8)</td>
<td>74,864/240,084 (31.2)</td>
<td>−1.6 (−5.3 to 2.1)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>65+</td>
<td>204/331 (61.6)</td>
<td>38,356/88,569 (43.3)</td>
<td>−18.3 (−22 to −14.6)</td>
</tr>
<tr>
<td>Men by race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>24/60 (40.6)</td>
<td>21,091/56,004 (37.7)</td>
<td>−2.9 (−9.1 to 3.2)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>White</td>
<td>91/314 (29.1)</td>
<td>37,622/120,672 (31.2)</td>
<td>2.1 (−2 to 6.2)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Other</td>
<td>9/36 (24.1)</td>
<td>5,268/36,008 (14.6)</td>
<td>−9.5 (−19.3 to 0.4)</td>
</tr>
<tr>
<td>Women by race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>30/92 (32.2)</td>
<td>30,285/88,868 (34.1)</td>
<td>1.9 (−5.1 to 1.6)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>White</td>
<td>95/388 (24.4)</td>
<td>41,094/181,412 (22.7)</td>
<td>−1.7 (−6.1 to 9.8)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Other</td>
<td>9/44 (20.8)</td>
<td>5,959/54,954 (10.8)</td>
<td>−10 (−18.7 to −1.3)</td>
</tr>
<tr>
<td>Men by age group (y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-39</td>
<td>18/99 (18.5)</td>
<td>14,819/77,992 (19)</td>
<td>0.5 (−5 to 6)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>40-64</td>
<td>56/178 (31.2)</td>
<td>33,567/98,778 (34)</td>
<td>2.8 (−2.6 to 8.2)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>65+</td>
<td>90/133 (67.4)</td>
<td>15,011/34,606 (43.4)</td>
<td>−24 (−29.8 to −18.3)</td>
</tr>
<tr>
<td>Women by age group (y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-39</td>
<td>3/98 (3.4)</td>
<td>21,331/136,663 (15.6)</td>
<td>12.2 (9.9 to 14.5)</td>
</tr>
<tr>
<td>40-64</td>
<td>78/228 (34.2)</td>
<td>41,296/141,286 (29.2)</td>
<td>−5 (−10.2 to 0.3)</td>
</tr>
<tr>
<td>65+</td>
<td>114/198 (57.5)</td>
<td>23,345/53,954 (43.3)</td>
<td>−14.2 (−21.5 to −6.9)</td>
</tr>
</tbody>
</table>

<sup>a</sup>BRFSS: Behavioral Risk Factor Surveillance System.<br><sup>b</sup>Sample size=934.<br><sup>c</sup>Indiana Network for Patient Care.<br><sup>d</sup>Sample size=548,298.<br><sup>e</sup>Δ: mean difference.<br><sup>f</sup>Behavioral Risk Factor Surveillance System and Indiana Network for Patient Care phenotypes were determined as statistically equivalent by the two 1-sided t test method.

**Discussion**

**Principal Findings**

Our study examined the prevalence estimates of 6 distinct EHR-based phenotypes to ascertain whether EHR-derived estimates are equivalent to estimates produced by survey methods. The 2 clinical phenotypes (P2 and P5) relying primarily on vital statistics data showed the closest equivalence to BRFSS hypertension prevalence estimates. This suggests that clinical variables, such as blood pressure readings, are important in classifying hypertension cases when compared to national survey data. However, clinical measurements are often missing from national surveys (eg, BRFSS). When clinical measurements are present (eg, the National Health and Nutrition Examination Survey), the survey possesses an even smaller...
sample size and is frequently more costly. Establishing robust local prevalence estimates may require local health departments to capture blood pressure measurements, which is cost prohibitive. EHR data may provide a more economical approach to the collection of clinical measurements. Additionally, EHRs can supply these measurements regularly forgoing the need for additional, specific public health data collection efforts.

Interestingly, phenotypes that relied on diagnosis code data performed less robustly. Previous studies have demonstrated the underreporting of conditions when relying on diagnostic codes alone [22-24]. Accordingly, it is possible that diagnostic codes themselves are not sensitive enough for identification of hypertension. Further, 1 possible reason for this is the type of encounter for which an individual is seen. For example, if the patient is being seen primarily in emergency or inpatient settings, a diagnosis of hypertension may not be coded, but the vital measurements would be available.

In our results, P6, which is the broadest and most sensitive definition of hypertension [21], did not align with the BRFSS at the overall population level. The hypertension BRFSS instrument item asks “has a doctor told you that you have hypertension?” [20]. This allows for variability in interpretation and may include individuals with a single elevated blood pressure incident or someone who is prehypertensive. Accordingly, it is logical that a computable phenotype using a combination of clinical data elements would be more sensitive to a diagnosis of hypertension but not to the broad question posed by the BRFSS. However, the phenotypes using a variety of clinical measurements may be a more robust measurement of hypertension for local health departments to deploy.

The results showcase the importance of the inclusion of vital statistics, which proved more sensitive for overall comparison and certain subpopulations when the CI threshold was lower. The results of P6 being associated with lower CIs were not surprising given the smaller sample sizes inherent in analyses of subpopulations. Compared to estimates from survey data, more numerous records available in the HIE or multiple EHR systems would allow for smaller CIs in estimates about subpopulations.

While not all algorithms demonstrated equivalency, 2 of the phenotypes demonstrated the potential for EHR data to provide prevalence estimates that are likely to be within 10 percentage points of BRFSS estimates. Accordingly, the use of EHR data may be a better option to estimate disease burden than costly community health surveys. EHR data have several benefits. First, EHR-derived prevalence estimates are timelier. This methodology can be implemented regularly (e.g. quarterly and semiannually) to address the needs of the community compared to national surveys. National surveys are typically conducted annually and require time for postprocessing for data. These conditions result in delayed estimates, making the data untimely for certain population health questions. For certain conditions and interventions, this may prove useful for the identification of community needs as well as the timely assessment of community-level interventions. For example, we are using these methods to estimate changes in childhood obesity in multiple urban neighborhoods that received community-level interventions to address childhood obesity [25].

Second, the EHR-derived measures can be tailored to the specific needs of local health departments. Working in coordination with health care systems or HIE networks, local health departments may arrange to receive the data most relevant to their specific question rather than using proxy constructs from national data. Additionally, the EHR-based measures were manually validated and demonstrated to be of high quality, showing strong specificity and positive predictive values [21]. As reported in the results, the computable phenotypes identified a higher prevalence for the Black or African American community. Some of this variation could be attributed to the overrepresentation of inner-city health system patients within the County. However, the demographic analysis supports the premise that the BRFSS may be underrepresentative of the Black or African American population. This argument may be bolstered by the higher prevalence of subpopulations represented within the INPC demographics, both the overall cohort and the hypertension cohort. High-quality estimates, partnered with customization to local needs, will ultimately provide more robust measures for the local health departments.

Further, 1 limitation in the broader use of this methodology is most public health agencies’ lack of legal authority to require reporting of data about chronic conditions. Currently, hospitals are not required to report clinical measurements or metrics related to chronic diseases, such as hypertension, to public health authorities beyond discharge data. Discharge data primarily consist of diagnostic codes, which may not reliably capture chronic disease burden as discussed above. Currently, the reporting of these data is voluntary and, therefore, unlikely to occur given the resources, human, and technological requirements to do so on the part of providers. However, HIE networks (such as INPC) have existing infrastructures that can be leveraged to address community surveillance needs. Data are already aggregated across health care systems and providers within the community, addressing a large amount of the work required to implement surveillance of chronic conditions. This analysis suggests support for leveraging HIE networks in the community for chronic disease surveillance.

The widening use of the Fast Healthcare Interoperability Resources standard and the Trusted Exchange Framework and Common Agreement for health data exchange may also increase public health agencies’ opportunity to access EHR data [26,27]. There are still barriers to the full adoption of HIE networks into the public health environment, such as infrastructure [28] and data quality [29]. However, the COVID-19 pandemic revealed the role HIE could play in support of public health needs [17]. This is increasingly becoming important given the burden of post–COVID-19 conditions [30] and the potential increase in chronic conditions after the pandemic. Surveillance of chronic conditions is critical to public health practice. The efforts to modernize the nation’s public health infrastructure, which are currently underway, should consider the important role HIE networks can play in support of chronic disease surveillance. Admittedly, future work will involve the implementation of HIE networks in those areas of the United States where they are not currently present.
A second limitation is the inconsistent and imprecise equivalency we have demonstrated between the HIE and BRFSS estimates. The BRFSS estimates themselves are fairly imprecise even for a population of about 1 million, as in Marion County, and so make a weak “gold standard,” especially for subpopulations. Conversely, EHR data only reflect persons with health care encounters, and persons with frequent visits are more likely to have enough EHR data to satisfy some phenotype definition. With health care use varying by health status, race, age, employment, and other factors, EHR data would need adjustment for systematic biases before being interpreted as representative of the general community or subpopulations of interest. Further research would reveal what adjustments can improve how well EHR-based estimates approximate population health statistics. This study is subject to limitations related to the quantity and type of available data. Equivalence may be improved by a more complete capture of an area’s health care providers, especially in ambulatory and primary care settings. Improved data capture would increase the EHR-based prevalence estimates. Data might be weighted according to patient characteristics, such as race, age, gender, or type of health insurance, allowing estimates to be adjusted to be more representative of the general population.

As noted above, this study is subject to limitations related to data availability, namely the period for which comprehensive data was available. There have been advancements in EHR adoption and use in the period from 2014 to now. EHR and HIE adoption will continue to be advanced by data modernization activities, which have in turn been spurred by gaps identified in the COVID-19 pandemic. The data availability of important measurements such as vitals, medications, and diagnoses will likely become routinely captured and shared as part of these activities. This suggests, and more recent literature suggests, that the accuracy of computable phenotypes may improve with these advancements [31,32].

Conclusions
This study demonstrates the feasibility of using EHR-derived prevalence estimates as rough substitutes for population-based survey estimates at the community level. It highlights the importance of critically assessing which data elements to include when deriving the EHR-based estimates. Using comprehensive data sources, containing complete clinical data as well as data representative of the population, may enhance local estimates. The number of people represented in EHR data versus survey data may allow for locally accurate EHR-based measurements of subpopulations. This is critical when considering health disparities as more robust measurements for subpopulations may enable targeted public health interventions.

Acknowledgments
This study’s team would like to thank Regenstrief Data Services for their expertise with data extraction. This work was funded, in part, by the Centers for Disease Control and Prevention under a contract with the Task Force for Global Health (CFDA P93.283) entitled “Enhancing DOH Capacity for Using EHR Data for Cardiovascular Disease.” The authors further acknowledge funding from the Public Health Informatics Program and Center for Biomedical Informatics at the Regenstrief Institute. No form of artificial intelligence was used in the creation or editing of this paper.

Data Availability
The data sets generated or analyzed during this study are not publicly available due to privacy and governance concerns but are available from the corresponding author upon reasonable request and with the completion of appropriate governance.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Two 1-sided t test analyses at 80% CI.
[DOCX File, 35 KB - ojphi_v16i1e48300_app1.docx]

Multimedia Appendix 2
Two 1-sided t test analyses at 90% CI.
[DOCX File, 30 KB - ojphi_v16i1e48300_app2.docx]

References

https://ojphi.jmir.org/2024/1/e48300


Abbreviations
- **BRFSS**: Behavioral Risk Factor Surveillance System
- **EHR**: electronic health record
- **HIE**: health information exchange
- **INPC**: Indiana Network for Patient Care
- **NYC**: New York City
- **TOST**: two 1-sided t test

©Katie S Allen, Nimish Valvi, P Joseph Gibson, Timothy McFarlane, Brian E Dixon. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 13.03.2023. 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 the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.
Effect of Long-Distance Domestic Travel Ban Policies in Japan on COVID-19 Outbreak Dynamics During Dominance of the Ancestral Strain: Ex Post Facto Retrospective Observation Study

Junko Kurita¹, MSc; Yoshitaro Iwasaki², BSc

¹Department of Nursing, Faculty of Sports & Health Science, Daitobunka University, Higashimatsuyama-shi, Japan
²Iwasaki Industrial Corporation, Kagoshima, Japan

Corresponding Author:
Junko Kurita, MSc
Department of Nursing
Faculty of Sports & Health Science
Daitobunka University
560 Iwadono
Higashimatsuyama-shi, 3558501
Japan
Phone: 81 0493 31 1503
Email: kuritaj@ic.daito.ac.jp

Abstract

Background: In Japan, long-distance domestic travel was banned while the ancestral SARS-CoV-2 strain was dominant under the first declared state of emergency from March 2020 until the end of May 2020. Subsequently, the “Go To Travel” campaign travel subsidy policy was activated, allowing long-distance domestic travel, until the second state of emergency as of January 7, 2021. The effects of this long-distance domestic travel ban on SARS-CoV-2 infectivity have not been adequately evaluated.

Objective: We evaluated the effects of the long-distance domestic travel ban in Japan on SARS-CoV-2 infectivity, considering climate conditions, mobility, and countermeasures such as the “Go To Travel” campaign and emergency status.

Methods: We calculated the effective reproduction number \( R(t) \), representing infectivity, using the epidemic curve in Kagoshima prefecture based on the empirical distribution of the incubation period and procedurally delayed reporting from an earlier study. Kagoshima prefecture, in southern Japan, has several resorts, with an airport commonly used for transportation to Tokyo or Osaka. We regressed \( R(t) \) on the number of long-distance domestic travelers (based on the number of airport limousine bus users provided by the operating company), temperature, humidity, mobility, and countermeasures such as state of emergency declarations and the “Go To Travel” campaign in Kagoshima. The study period was June 20, 2020, through February 2021, before variant strains became dominant. A second state of emergency was not declared in Kagoshima prefecture but was declared in major cities such as Tokyo and Osaka.

Results: Estimation results indicated a pattern of declining infectivity with reduced long-distance domestic travel volumes as measured by the number of airport limousine bus users. Moreover, infectivity was lower during the “Go To Travel” campaign and the second state of emergency. Regarding mobility, going to restaurants, shopping malls, and amusement venues was associated with increased infectivity. However, going to grocery stores and pharmacies was associated with decreased infectivity. Climate conditions showed no significant association with infectivity patterns.

Conclusions: The results of this retrospective analysis suggest that the volume of long-distance domestic travel might reduce SARS-CoV-2 infectivity. Infectivity was lower during the “Go To Travel” campaign period, during which long-distance domestic travel was promoted, compared to that outside this campaign period. These findings suggest that policies banning long-distance domestic travel had little legitimacy or rationale. Long-distance domestic travel with appropriate infection control measures might not increase SARS-CoV-2 infectivity in tourist areas. Even though this analysis was performed much later than the study period, if we had performed this study focusing on the period of April or May 2021, it would likely yield the same results. These findings might be helpful for government decision-making in considering restarting a “Go To Travel” campaign in light of evidence-based policy.

(Online J Public Health Inform 2024;16:e44931) doi:10.2196/44931
Introduction

Important features of countermeasures against the COVID-19 outbreak in Japan were restrictions such as staying at home, wearing a mask, holding virtual meetings at organizations, and conducting contact tracing. All these measures were implemented on a voluntary basis; that is, the government strongly recommended such measures but none was required as a matter for law enforcement. Therefore, lockdowns such as those that occurred in the United States or some European countries, entailing enforced laws, never occurred in Japan. Even though these countermeasures were recommended by the government without enforced laws, aside from laws implemented at border controls and for quarantines, most Japanese people cooperated with the recommendations voluntarily.

At the beginning of the COVID-19 outbreak in Japan, school closures and voluntary event cancellations were required from February 27 through March of 2020. Large commercial events were also cancelled. Subsequently, a state of emergency was declared from April 7 through May 25, with voluntary restrictions against leaving the home and requiring the shutting down of businesses serving customers. During this period, the first peak in the outbreak was reached on April 3, 2020. Another peak then emerged on July 29, as shown in Figure 1. The so-called “Go To Travel” campaign (GTTC) started on July 22, 2020, with 50% subsidized travel and coupons issued for shopping at tourist destinations. The policy was aimed at reinforcing sightseeing businesses, even though such a measure entailed the possibility of expanding the outbreak. Thereafter, the GTTC continued through December 2020, by which time a third wave of infections had emerged, which was larger than either of the prior two waves. Therefore, the GTTC was implicated as the main underlying reason for the third wave [1].

![Figure 1. Numbers of newly confirmed COVID-19 cases in Kagoshima prefecture, Japan, from March 26, 2020, to the end of February 2021. Bars represent the epidemic curve showing the numbers of patients by onset date. The arrow indicates the period during which the “Go To Travel” campaign was in effect.](image-url)
nonpharmaceutical interventions, including lockdowns, strongly reduced transmission in at least 11 European countries. However, another study including 131 countries found that the introduction and relaxation of lockdowns or movement restrictions had only limited the effects on infectiousness, except for public event bans, although their data were limited to the end of July 2021 [12]. Another study indicated that strict movement restrictions in Argentina imposed as of March 2020 were effective at reducing mobility, but not for mitigating the outbreak [13]. These mixed results suggest that such countermeasures might not have significantly affected the infectivity of SARS-CoV-2.

By contrast, there is abundant evidence demonstrating the effects of international travel and restrictive policies on pandemic dynamics, mainly focusing on the effects of long-distance domestic travel [14-17]. However, despite a report indicating that 80% of patients with COVID-19 on an island in Canada had acquired a travel-related infection from July 1, 2020, to May 31, 2021 [18], in Japan, only 12% of patients had a recent history of international travel in the very early phase of the pandemic from January 13 to March 31, 2020 [19].

Some studies have emphasized associations among international trade and outbreak sizes [20,21]. Although the traded goods are likely not the source of infectiousness, international trade volumes might be related to outbreak size, as this volume can reflect the movements of business representatives accompanying international trade. Unfortunately, these two studies were based on cross-sectional analyses and therefore it was not possible to isolate the effects of the number of travelers from the effects of high population densities or air pollution discharged by manufacturing industries. Therefore, these studies did not provide sufficient evidence to confirm that travelers were mainly responsible for expanding the outbreak. Moreover, immediately after the pandemic was declared, most international borders had been closed in principle and almost all planned face-to-face meetings related to international trade were held as virtual meetings [22].

Countermeasures against the spread of COVID-19 differed considerably with respect to international and domestic travel, as the former involves border controls and restrictive quarantines, which might be more effective at reducing transmission [23-26]. By contrast, domestic travel recommendations might be limited to voluntary restrictions against going out and long-distance domestic travel. Even in the case of a long-distance domestic travel ban, a chain of short-distance trips to a neighboring city, including commuting to a school, workplace, or shopping center, might ultimately act to transmit viruses over long distances, given some delay. In this sense, experience and evidence related to international travel might not be directly applicable to long-distance domestic travel.

To our knowledge, no study has examined the impact of long-distance domestic travel on outbreak situations in rural areas. One might expect that such information might be less available for epidemiological analysis. Although annual or monthly data related to travelling or sightseeing might be generally available, such records are generally not widely available. Moreover, these data would be quite aggregated, and the number of data points would likely be too small to support sufficient power for statistical analyses considering short time periods of less than 1 year. Fortunately, epidemiologists, statisticians, and companies managing resort hotels and buses to airports in rural areas can provide travel-related data collaboratively. In fact, daily data of bus users from airports and visitors to these hotels are available for many areas in Japan. Therefore, the hypothesis that sightseeing visitors and long-distance domestic travelers were largely responsible for spreading the virus and contributed to COVID-19 outbreaks in rural areas can be tested directly. This hypothesis served as the rationale for ceasing the GTTC and for banning long-distance domestic travel during the first and second states of emergency in Japan. Nevertheless, this rationale has neither been analyzed nor confirmed to date.

Therefore, the objective of this study was to directly examine the hypothesis that long-distance domestic travel was responsible for expanding the COVID-19 outbreak, supporting the rationale and legitimacy of the policy followed in Kagoshima prefecture, Japan. This area was selected given that Kagoshima, located north of Okinawa but in southern Japan (Figure 2), has one airport that is used for commuting to more urban areas of the country, such as Tokyo and Osaka. Moreover, collaborative data obtained from epidemiologists and from leading tourist industry companies were available for Kagoshima, offering a valuable resource for this analysis that can contribute to more insightful consideration and policy evaluation.
Figure 2. Map of Japan with Kagoshima prefecture indicated in the red circle and the airport routes to the main tourism cities for long-distance domestic travel marked.

Methods

Sample and Data

Data reflecting the daily numbers of Kagoshima airport limousine bus users were provided by Iwasaki Industrial Corp., Ltd., of Kagoshima. However, the information does not completely reflect the traffic to and from the airport, as some airport users commuted to or from the airport by taxi, private car, or rental car. Moreover, some tourists visited Kagoshima without using an airline, such as by train, car, bus, or ship. Nevertheless, most tourists from Tokyo, Osaka, or other urban areas typically use airlines to visit Kagoshima. Therefore, although the extent of domestic travel could not be verified completely, we infer that the available information accurately reflects the general picture of movement during this time.

The study period was defined as June 20, 2020, through February 2021. Before this period, COVID-19 cases had been confirmed only sporadically. Therefore, the effective reproduction number $R(t)$ could not be stably estimated. However, after this period, the Alpha variant strain emerged and dominated up to 35% of all cases throughout Japan by the end of March 2021 [27]. The infectiousness of the Alpha variant was estimated to be 35%-90% higher than that of the ancestral strain [28-31]. Such a large difference in virus characteristics could affect the estimations related to our study objectives. Therefore, we limited the study period to the time prior to the emergence of the Alpha variant strain.

Variables

Climate variables considered were average temperature (measured in degrees Celsius) and relative humidity data for Kagoshima during the day; these data were obtained from the Japan Meteorological Agency [32].

We also used mobility data provided by Google, which includes data for six types of locations: restaurants, shopping malls or amusement centers, grocery stores or pharmacies, parks, transition areas, workplaces, and homes [33]. These data show mobility comparisons according to a base day; a value of 100 was assigned if the number of people recorded for a given type of place was the same as that recorded on the base day.

Additionally, we considered the impact of major countermeasures against the pandemic implemented in Japan: two emergency state declarations, the GTTC, and school closure and voluntary event cancellation. The latter measure extended from February 27 through March in 2020, requiring school closure and cancellation of voluntary events, along with cancellation of private meetings. The first state of emergency was declared on April 7, 2020, which ceased at the end of May. This involved required school closures, shutting down of some businesses, and voluntary restrictions against going out. For Kagoshima prefecture, the state of emergency spanned from April 16 to May 14, 2020. To subsidize travel and shopping at tourist destinations, the GTTC started on July 22, 2020, and was halted at the end of December 2020.
The second state of emergency was declared on January 7, 2021, and continued until March 21, 2021, for the 11 most-affected prefectures in Japan. This countermeasure required restaurant closure at 8 PM, along with voluntary restrictions against going out, but it did not require school closure. During the study period, the GTTC and the second state of emergency were in effect. Although this second state of emergency was not declared for Kagoshima prefecture, it was implemented in major cities, including Tokyo and Osaka.

**Models and Data Analysis**

The numbers of newly confirmed COVID-19 cases each day were reported by the Kagoshima Prefecture Office from May 13, 2020, through February 21, 2021 [34]. The effective reproduction number \(R(t)\) was estimated according to a previous study [35]. We first estimated the onset date of patients for whom onset dates were not reported. Letting \(f(k)\) represent the empirical distribution of the incubation period and letting \(N_t\) denote the number of patients for whom onset dates were not available as published at date \(t\), the number of patients for whom the onset date was known is designated \(t-1\). The number of patients with onset date \(t-1\) for whom onset dates were not available was estimated as \(f(1)N_t\). Similarly, patients with onset date \(t-2\) and for whom onset dates were not available were estimated as \(f(2)N_t\). Therefore, the total number of patients for whom the onset date was not available, given an onset date of \(s\), was estimated as \(\sum_{k=s} f(k)N_t + k\) for the long duration extending from \(s\).

Moreover, the reporting delay for published data from the Ministry of Health, Labour and Welfare of Japan might be considerable. In other words, if \(s+k\) is larger than that in the current period \(t\), then \(s+k\) represents the future for period \(t\). Consequently, \(N_{s+k}\) is not observable. Such a reporting delay engenders underestimation of the number of patients. Therefore, the formula must be adjusted to \(\sum_{k=t-1}^{\infty} f(k)N_t + k\sum_{k=t-1}^{\infty} f(k)\). Similarly, patients for whom the onset dates were available are expected to be affected by the reporting delay. Therefore, the formula \(M_{s+k} = \sum_{k=s}^{\infty} f(k)\) was used, where \(M_{s+k}\) denotes the number of patients for whom the onset date was period \(s\) as of the current period \(t\).

We defined \(R(t)\) as the number of infected patients on day \(t\) divided by the number of patients who were presumed to be infectious. The number of infected patients was calculated from the epidemic curve by the onset date using an empirical distribution of the incubation period, which is \(\sum_{k=s} g(k)E_{t-k}\), where \(E_t\) denotes the number of patients for whom the onset date was period \(t\). The distribution of infectiousness in symptomatic and asymptomatic cases \(g(k)\) was assumed to be 30% on the onset day, 20% on the following day, and 10% for the subsequent 5 days [36]. Therefore, the number of infectious patients was calculated as \(\sum_{k=s} g(k)E_{t-k}\) and \(R(t)\) was defined as \(\sum_{k=s} f(k)E_{t-k}\sum_{k=s} g(k)E_{t-k}\). The empirical distributions of \(f\) and \(g\) based on actual data in Japan were obtained from an earlier report [35].

The bootstrapping procedure was applied to calculate the 95% CIs of \(R(t)\). We used fully replicated bootstrapping for a constant number of cases in this study period. There were \(L\) patients in the actual data of this study period, with numbering of the patients from the initial case to the last case. Initially, no patient was on the bootstrapped epidemic curve. If a random variable drawn from a uniform distribution of \((0,1)\) was included in the interval \([i/(L-1), (i+1)/(L-1)]\), then we added 1 to the onset date of the \(i\)th patient to the bootstrapped epidemic curve. We replicated this procedure \(L-1\) times. Thereby, we obtained the bootstrapped epidemic curve with \(L-1\) patients. We calculated \(R(t)\) based on the bootstrapped epidemic curve. We denote \(R(t)\) based on the \(j\)th bootstrapped epidemic curve as \(R(t)^j\). We repeated these processes 10,000 times to obtain \(R(t)^j\) \((j=1–10,000)\). We reordered superscripts in each \(t\) from \(\min_j[R(t)^j]\) to \(\max_j[R(t)^j]\) and denoted the reordered \(R(t)^j\) as \(R(t)^{jk}\) \((k=1–10,000)\). The estimated \(R(t)\) was then taken as the median of \(R(t)^{jk}\), denoted \(R(t)^{5000}\), and its 95% CI is \(R(t)^{250} – R(t)^{9750}\).

To clarify associations among \(R(t)\) and the GTTC or other variables in addition to climate, mobility, and countermeasures, we used observed least-squares regression to regress the daily \(R(t)\) on daily dummy variables for the GTTC (G\(t)\)); daily data of airport limousine bus users and visitors at the resort hotels (L\(t)\); as well as dummy variables for daily climate (H\(t)\) for temperature and H\(t)\(2\)\(t)\) for humidity), mobility (M\(i\)\(t)\) \((i=1–6)\), and the second state of emergency as follows:

\[
R(t) = \alpha + \beta G(t) + \gamma L(t) + \delta_1 H_1(t) + \delta_2 H_2(t) + \delta_3 H_3(t) + \delta_4 H_4(t) + \delta_5 H_5(t) + \delta_6 H_6(t) + \theta P(t) + \epsilon(t)
\]

The start of the study period was June 20, 2020. Therefore, school closure and voluntary event cancellation and the first state of emergency had ceased; accordingly, we are unable to estimate their effects on \(R(t)\).

We anticipated the following influence of the explanatory variables: airport limousine bus users and visitors at the resort hotels or part of the GTTC would contribute to increased infectivity if the policy banning long-distance domestic travel was rational, and countermeasures such as the emergency state or school closure and voluntary event cancellation were presumed to decrease infectivity. We adopted 5% as the significance level and performed all statistical analyses using Stata SE 17.0 software (Stata Corp).

**Ethical Considerations**

Information about the number of patients used for this study was collected under the Law of Infection Control, Japan, published by the Kagoshima Prefectural Office [34]. Iwasaki Industrial Corp Ltd provided the number of airport limousine bus user data from their business records. Both data sets only provided the number of persons, and thus did not include private or privacy information. There were therefore no ethical issues related to this study.

**Results**

Figure 1 shows the numbers of newly confirmed cases of COVID-19, including asymptomatic cases, in Kagoshima from March 26, 2020, to February 28, 2021. The initial case was
detected in Kagoshima on March 26. However, data were sporadic in the initial phase. From June 2020, cases were reported continuously.

Figure 3 presents the estimated R(t) and 95% CI for the study period. Before June 2020, the R(t) was large and volatile because very few cases were reported. After June 2020, because new cases were reported almost daily, the R(t) became smaller and exhibited less volatility. The largest peak was in November 2020 while the GTTC was in effect.

Figure 4 portrays the number of the airport limousine bus users during the study period. The main peak of airport limousine bus users occurred before the outbreak emergence. During April and May of 2020, when the first state of emergency was declared, this number decreased considerably, reaching 0 in September 2020 when the airport was closed because a typhoon struck the area.

Table 1 presents the estimation results (adjusted $R^2=0.2772$; n=273 observations) from the regression model. Climate conditions, including temperature and humidity, did not show any significant associations with R(t). In addition, there was no association of specific places with R(t) over the shorter period, except for restaurants and grocery stores. One can infer that going to a restaurant increased infectivity, whereas going to a grocery store, perhaps as a reflection of “staying home,” reduced infectivity. This result might indicate that a “stay-at-home” policy, including a lockdown or voluntary ban against going out as practiced in Japan, was legitimate. However, staying at “home” itself and going to a “workplace” were not found to be significant factors influencing R(t), even though they had negative coefficients. The first and second states of emergency and the GTTC had negative and significant effects on R(t). In particular, the estimated coefficients of these variables were quite large. The second state of emergency, which was not applied to Kagoshima, also showed an association with reduced infectivity.

Figure 3. Effective reproduction number (black line) with 95% CIs (gray lines) of COVID-19 in Kagoshima prefecture, Japan, from June 20, 2020, to the end of February, 2021.
Table 1. Estimated effects of various factors on infectivity (effective SARS-CoV-2 reproductive number) obtained using data from June 20, 2020, to the end of February 2021 in Kagoshima prefecture, Japan.

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Estimated coefficient</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airport limousine bus users</td>
<td>−0.003</td>
<td>.02</td>
</tr>
<tr>
<td>Temperature</td>
<td>0.013</td>
<td>.76</td>
</tr>
<tr>
<td>Humidity</td>
<td>0.003</td>
<td>.89</td>
</tr>
<tr>
<td>Restaurant, shopping mall, or amusement center</td>
<td>0.094</td>
<td>.03</td>
</tr>
<tr>
<td>Grocery store or pharmacy</td>
<td>−0.095</td>
<td>.03</td>
</tr>
<tr>
<td>Park</td>
<td>−0.038</td>
<td>.08</td>
</tr>
<tr>
<td>Transition</td>
<td>0.028</td>
<td>.52</td>
</tr>
<tr>
<td>Workplace</td>
<td>−0.035</td>
<td>.44</td>
</tr>
<tr>
<td>Home</td>
<td>−0.303</td>
<td>.09</td>
</tr>
<tr>
<td>Second state of emergency</td>
<td>−3.774</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>GTTC(^a)</td>
<td>−3.040</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Constant</td>
<td>7.646</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>F value (df=11, 261)(^b)</td>
<td>10.48</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\(^a\) GTTC: “Go To Travel” campaign.
\(^b\) The F test was used to evaluate the null hypothesis that all coefficients except for constant terms are 0.

Discussion

Principal Findings

Estimation results for the GTTC and the number of airport limousine bus users indicate that the promotion of long-distance domestic travel might decrease SARS-CoV-2 infectiousness. This finding thus appears to be inconsistent with a legitimate policy banning long-distance domestic travel, including cessation of the GTTC. Our findings suggest that during sightseeing or long-distance domestic travel, tourists/visitors and hosts might be much more conscientious about infection control and might therefore be less likely to infect others than when they are in their hometown. These phenomena were also indicated through a psychographical study and market research.
In other words, people in their hometown might exhibit less conscientious behaviors and might therefore be more likely to become infected. Therefore, discouraging long-distance domestic travel might actually engender worse infection rates overall within a local area.

This finding is consistent with earlier studies [35,39]. One study showed that the GTTC reduced infectiousness, whereas the other study found that events with an audience might not raise infectiousness compared to events without an audience.

Another study [40] using two types of patient data (onset date and the date of testing positive) found much higher travel-associated COVID-19 incidence during the period of July 22-26, when the GTTC was initiated, than during either an earlier period of June 22 to July 21, July 15-19, or June 22 to July 21 in terms of the incidence rate ratio (IRR). The same study also compared the period of August 8-31.

Some notable points can be identified from this previous study [40]. First, the proportion of people with a travel history during the GTTC period was comparable to that during the two prior periods. In particular, the proportion of people with a travel history among patients with COVID-19 who had an available onset date was smaller for the GTTC period than during the prior period of July 15-19. However, the authors found a significantly higher COVID-19 incidence at the beginning of the GTTC. These findings might merely reflect the fact that the total number of patents in the GTTC period was higher than that during the prior period. In other words, they did not control for the underlying outbreak situation and therefore found an incorrect association. Use of the IRR would be valid if the underlying outbreak situation other than the examining point was the same in the two considered periods. Therefore, application of the IRR might be inappropriate for addressing this issue. At the very least, controlling for the potential differences in the outbreak situation is considered to be necessary. The underlying outbreak situation, unrelated to the GTTC, was reflected in the number of patients without a travel history or any sightseeing. One potential approach to control for the underlying outbreak situation is to consider the share of patients with a travel history or sightseeing. However, that share did not increase markedly during the initial stage of the GTTC. This lack of a marked increase indicates that the authors’ results and conclusions are misleading.

Second, Anzai and Nishiura [40] referred to the period of August 8-31, 2020, when the GTTC was still in effect. The proportion of patients with a travel history was much smaller than that during the period of July 22-26 when the GTTC started or in the prior period. Although the authors did not compare the COVID-19 incidence in August with that of either the prior period or July 22-26 when the GTTC started, the rate of incidence in August 2020 was likely lower than that in other periods. In fact, some patients with active COVID-19 infections traveling under the GTTC might have been included in the study period, August 2020, as described above. Their inclusion might be inconsistent with the authors’ conclusion.

Third, we observed the peak of newly infected persons on July 23, 2020, which was the start date of the GTTC, for the entirety of Japan. Therefore, we infer that the GTTC might have reduced infectiousness. We also considered the potential effect of climate conditions on the variation in infectivity. At around the end of July, the rainy season in Japan ends and summer begins, accompanied by high temperatures. Therefore, the GTTC might have been insufficient to increase the number of COVID-19 patients and cancel out the benefits from the improved climate conditions. Taken together, these points suggest that the GTTC might not have been the main factor determining the course of the COVID-19 outbreak.

Moreover, if the GTTC did have a strong effect on the outbreak dynamics, then there would be an increase in the number of patients with no travel history. For example, one can consider a patient traveling under the GTTC on July 22 and 23, with disease onset occurring on July 24. Although this patient had a travel history in the GTTC period, they would not be included in a group of patients with a travel history whose onset date corresponds to the initial GTTC period of July 27-31. Nevertheless, presymptomatic patients are well known to become infectious during the symptomatic period [36]. In the above scenario, this patient might infect staff members of hotels or other individuals encountered in the visiting areas. However, if their onset dates were July 27 and 28, they would be included in the group of patients without a travel history in the GTTC start period of July 27-31. Therefore, the GTTC certainly increased the number of patients without a travel history but did not increase the number of patients with a travel history in this case. Consequently, when considering the effects of the GTTC, it is important to account for the total number of patients with COVID-19, irrespective of their travel history.

Finally, it is noteworthy that this study could have been performed in the middle or end of March 2021, if we had analyzed those data at that time. We found the same results as those found from this study. In fact, this analysis was performed in 2022, although similar research without the valuable data used for this study was posted to the medRxiv preprint server on January 4, 2021, and we obtained the same results for the GTTC [35]. In general, an ex ante policy evaluation is necessary, although it was very difficult to estimate its effects precisely. By contrast, an ex post evaluation performed as soon as possible could have been possible if such preparation had been arranged before policy activation. If such preparation had been done, then the policy banning long-distance domestic travel with no legitimate rationale could have been prevented in 2021 and thereafter.

Another study in Japan [41] showed that the GTTC was responsible for the introduction of an emerging sublineage of SARS-CoV-2 in October 2020 to Hokkaido, Japan’s second largest island. The ratio of the number of travelers at Hokkaido to that on the same month in the prior year was the largest in October 2020. However, this might not necessarily imply that the outbreak was accelerated by the GTTC or the number of travelers.

For this study, because daily airport user data were not available, we used the number of daily airport limousine bus users as a proxy of daily airport users, including those who did not use limousine buses. However, monthly airport user data have since been published [42]. Therefore, we further evaluated the
representativeness of airport limousine bus users for airport users on a monthly basis. The correlation coefficient between monthly airport limousine bus users and airport users during 2020 and 2021 was 0.9881 (P<.001). Therefore, we can infer that airport limousine bus users constitute a good proxy of overall airport users. Moreover, even though bullet-train or bus services were available as a means of transportation to Kagoshima from neighboring or nearby prefectures, airlines are the only means of transportation to Kagoshima from areas with large populations in Japan, such as Osaka and Tokyo. Therefore, we can infer that airport limousine bus users are a good proxy of long-distance domestic travel volumes for Kagoshima.

This study excluded some variables suggested by earlier studies, such as vaccination, contact tracing, or mass gathering events, which potentially affect infectivity. Vaccination for COVID-19 started in March 2021 in Japan for health care workers. Therefore, there was no vaccination performed during our study period [43-45]. Moreover, contact tracing had been performed with the same intensity during the study period in Japan and was continued until the Omicron variant strain emerged. The public health center could not be traced with the same intensity. Contact tracing might be effective, at least in the very early stage, when the number of cases was limited to a few patients per public health center [46]. However, even in the very early stage of the pandemic in Japan, 80% of the infection sources were unknown. Therefore, contact tracing should not be expected to be effective in most cases in Japan [47]. Nevertheless, we were not able to estimate the effects or intensity of contact tracing for this study because it did not vary during the study period.

Mass gathering events such as the Olympic and Paralympic Games were also excluded from our analyses because of the study period. Even though international visitors seeking to see the games were refused because many players and officers were already crowded in a small area, an outbreak in the players’ village had been expected [48].

**Limitations**

This study has some limitations. First, this study specifically assessed data from Kagoshima. Therefore, it remains unclear whether the same results would hold for other regions or for the entirety of Japan.

Second, we particularly examined the ancestral strain of SARS-CoV-2, which might be less infective than the Alpha variant strain [28-31] and the subsequent dominant Delta and Omicron variant strains [29,30,49,50]. Thus, the effects of a policy banning long-distance domestic travel might have been different under a scenario of the dominance of these variant strains.

Third, if complete daily information about long-distance domestic travel to Kagoshima prefecture were available, obviating the use of data particularly addressing only some travel, then the implications might differ from those obtained with this analysis. We consider that our data do reflect complete and precise travel information, although it is not possible to prove this at present.

Fourth, regression analyses such as that used for this study cannot demonstrate causality. Although we interpreted the number of airport limousine bus users as showing decreased infectivity, lower infectivity pushed up the number of airport limousine bus users. Therefore, the results need to be interpreted cautiously.

**Conclusion**

We demonstrated that the GTTC or the increase of tourists and long-distance domestic travel visitors might not contribute to increasing COVID-19 infectiousness. Therefore, the policy banning long-distance domestic travel, including cessation of the GTTC, was neither fair nor rationally justified. Even though this analysis was performed much later than the study period of focus, the same results would be obtained considering the periods of April or May 2021 if we had performed this study at that time. The findings might have been helpful at that time for more rational decision-making when the government was considering whether to restart the GTTC. If so, then evidence-based policy might be suggested and operated. This perspective is in line with that of an earlier study [51].

**Acknowledgments**

We acknowledge the providers of the unique data used in this study, as well as the great efforts of all staff at public health centers, medical institutions, and other facilities fighting against the spread and destruction of COVID-19. This study is based on the authors’ opinions and does not reflect any stance or policy of their affiliations.

**Data Availability**

The data of airport limousine buses are not publicly available because they are sales data of the private company. However, the data are available from the corresponding author on reasonable request. Other data used for this study are available from the corresponding author.

**Authors’ Contributions**

JK was responsible for the coordination of the study and analyzed the data. YI collected and prepared the data for analysis. Both authors contributed to the writing and review of the final manuscript.
Conflicts of Interest
None declared.

References
1. Imada K. Japan to suspend Go to Travel campaign in cities where coronavirus cases are surging. TimeOut. 2020 Nov 24. URL: https://timeout.com/566w4x9k [accessed 2024-04-12]

https://ojphi.jmir.org/2024/1/e44931

Online J Public Health Inform 2024 | vol. 16 | e44931 | p.140 (page number not for citation purposes)


Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTTC</td>
<td>“Go To Travel” campaign</td>
</tr>
<tr>
<td>IRR</td>
<td>incidence rate ratio</td>
</tr>
<tr>
<td>R(t)</td>
<td>effective reproduction number</td>
</tr>
</tbody>
</table>

© Junko Kurita, Yoshitaro Iwasaki. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 22.04.2024. 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 the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.
A Semantic Approach to Describe Social and Economic Characteristics That Impact Health Outcomes (Social Determinants of Health): Ontology Development Study

Daniela Dally1*, BS, MS; Muhammad Amith2,3*, PhD; Rebecca L Mauldin4, LMSW, PhD; Latisha Thomas4, MSW; Yifang Dang5, MS; Cui Tao6, PhD

1The University of Texas Health Science Center at Houston School of Public Health, The Brownsville Region, Brownsville, TX, United States
2Department of Biostatistics and Data Science, University of Texas Medical Branch, Galveston, TX, United States
3Department of Internal Medicine, University of Texas Medical Branch, Galveston, TX, United States
4School of Social Work, The University of Texas at Arlington, Arlington, TX, United States
5School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States
6Department of Artificial Intelligence and Informatics, Mayo Clinic, Jacksonville, FL, United States
*these authors contributed equally

Corresponding Author:
Cui Tao, PhD
Department of Artificial Intelligence and Informatics
Mayo Clinic
4500 San Pablo Road
Jacksonville, FL, 32224
United States
Phone: 1 904 953 0255
Email: Tao.Cui@mayo.edu

Abstract

Background: Social determinants of health (SDoH) have been described by the World Health Organization as the conditions in which individuals are born, live, work, and age. These conditions can be grouped into 3 interrelated levels known as macrolevel (societal), mesolevel (community), and microlevel (individual) determinants. The scope of SDoH expands beyond the biomedical level, and there remains a need to connect other areas such as economics, public policy, and social factors.

Objective: Providing a computable artifact that can link health data to concepts involving the different levels of determinants may improve our understanding of the impact SDoH have on human populations. Modeling SDoH may help to reduce existing gaps in the literature through explicit links between the determinants and biological factors. This in turn can allow researchers and clinicians to make better sense of data and discover new knowledge through the use of semantic links.

Methods: An experimental ontology was developed to represent knowledge of the social and economic characteristics of SDoH. Information from 27 literature sources was analyzed to gather concepts and encoded using Web Ontology Language, version 2 (OWL2) and Protégé. Four evaluators independently reviewed the ontology axioms using natural language translation. The analyses from the evaluations and selected terminologies from the Basic Formal Ontology were used to create a revised ontology with a broad spectrum of knowledge concepts ranging from the macrolevel to the microlevel determinants.

Results: The literature search identified several topics of discussion for each determinant level. Publications for the macrolevel determinants centered around health policy, income inequality, welfare, and the environment. Articles relating to the mesolevel determinants discussed work, work conditions, psychosocial factors, socioeconomic position, outcomes, food, poverty, housing, and crime. Finally, sources found for the microlevel determinants examined gender, ethnicity, race, and behavior. Concepts were gathered from the literature and used to produce an ontology consisting of 383 classes, 109 object properties, and 748 logical axioms. A reasoning test revealed no inconsistent axioms.

Conclusions: This ontology models heterogeneous social and economic concepts to represent aspects of SDoH. The scope of SDoH is expansive, and although the ontology is broad, it is still in its early stages. To our current understanding, this ontology represents the first attempt to concentrate on knowledge concepts that are currently not covered by existing ontologies. Future
direction will include further expanding the ontology to link with other biomedical ontologies, including alignment for granular semantics.

(Keywords: social determinants of health; ontology; semantics; knowledge representation)

Introduction

Background

Ontologies are an important resource that have advanced the biomedical sciences. Originating from the philosophical domain and later incorporated into the computing and information sciences, ontologies represent and model our physical reality using semantics to describe domain entities (ie, knowledge base) [1]. These artifacts can be used to house vocabularies to generate inferences with the help of software reasoners such as HermiT [2], ELK [3], and FaCT++ [4]. Logically structured vocabularies can be used with reasoning tools to implement problem-solving software in clinical settings. In addition, biomedical researchers have advanced and wielded ontologies to be used in applications for artificial intelligence, natural language processing, information retrieval, and indexing (eg, data integration, harmonization, and exchange) [5]. Some impactful examples of ontologies include the Systematized Nomenclature of Medicine–Clinical Terms [6] and Gene Ontology [7], which are hosted on the National Center for Biomedical Ontology [8] and the OBO Foundry [9]; for example, the National Center for Biomedical Ontology BioPortal is an open repository of >700 biomedical ontologies [8], whereas the OBO Foundry hosts interoperable biomedical and health ontologies that share a common framework [9]. All the OBO Foundry–approved ontologies are built upon the Basic Formal Ontology (BFO), a common upper-level ontology, for interoperability and reuse. More than ever, there is a strong need to use ontologies for social health behavior sciences with the downstream goal of harmonizing biological and behavioral data [10].

Social Determinants of Health

Since the early 19th century, the public health community has sought to determine how social determinants are associated with behavior, health outcomes, and health inequalities [11]. Factors such as social position can influence an individual’s health status and thus lead to disease-inducing behaviors [11]. The link between social determinants and disease is a central point for public health research [11]. Over the years, public health researchers have classified these determinants as social determinants of health (SDoH). SDoH have been described by the World Health Organization as the conditions in which individuals are born, live, work, and age [12]. These nonbiological factors influence health outcomes in terms of health status, well-being, mortality, and life expectancy.

SDoH encompass many different areas, such as social and political context, governance, physical and living environment, community, safety, education, occupation, income, cultural and social values, biological and behavioral factors, wellness, food, and the health care system [12]. These categories can be represented by 3 levels of organization: macrolevel, mesolevel, and microlevel determinants [12]. Macrolevel determinants consist of socioeconomic hierarchies that govern access to resources in society through policy making [11]. Mesolevel determinants include concepts such as environment, neighborhood quality, occupation, and crime. This intermediate level is also concerned with psychosocial risk factors such as a stressful environment, the quality of social networks, and high physical or social demand [11]. Finally, microlevel determinants describe individual interactions, behaviors, lifestyle, and genetics [11]. Associated with these determinants are health inequalities, or the unfair and avoidable differences in health status among individuals [12], including inequities caused by structural or systemic factors.

Research Objective

The overarching goal of this research was to develop a biomedical ontology to model and represent knowledge on SDoH. More specifically, this work attempted to provide a broad spectrum of concepts ranging from the macrolevel to microlevel determinants focusing on social and economic characteristics as well as social-related health policies. By developing an ontology for SDoH, we can standardize the current scientific knowledge of this area based on a lightweight literature review and consensus from domain experts. Accomplishing this may help provide a computable ontology artifact that can link health data to concepts involving SDoH and advance informatics methods and tools to understand the impact each determinant has on human populations. In addition, modeling SDoH may also help to reduce existing gaps in the literature through explicit links between the determinants and biological factors. This in turn can allow researchers and clinicians to make better sense of data and discover new knowledge through the use of semantic links.

Existing relevant ontologies usually focus on biology and biomedicine; however, the scope of SDoH expands beyond the biomedical level (ie, microlevel) and relates to aspects that are not necessarily biology based, such as economics, public policy, social factors, and so on. Some of the more mature ontologies, such as the ones hosted on the OBO Foundry, have some interoperability due to a shared framework, but there remains a need to connect the heterogeneous SDoH concepts within the biomedical level and elucidate meaning from the knowledge. We therefore put forth the following research objective: using ontological methods, we can represent, formalize, and connect concepts pertaining to social, policy, and economic factors of SDoH. The output of this effort is an initial ontology artifact that models the social, policy, and economic concepts and their relationships in composing the scope of SDoH to build future work. To accomplish this, we (1) analyzed the literature on the 3 aspects and the aforementioned concepts within these aspects.

https://ojphi.jmir.org/2024/1/e52845

Online J Public Health Inform 2024 | vol. 16 | e52845 | p.144

(page number not for citation purposes)
and (2) produced an evaluated ontology artifact that reflects the intricate connections of the social and economic concepts of SDoH. This final experimental ontology artifact will be logically consistent with evaluation from domain experts and reasoning tools, grounded from a review of the literature to determine high-level concepts that stretch across SDoH, and aligned with a shared framework for biomedical ontologies to enable interoperability and reusability.

**Methods**

**Overview**

A brief yet comprehensive review was conducted to develop ontology terminology that effectively captures the concepts related to SDoH. This review served as a foundation for structuring and defining the key elements within the ontology. The literature reviewed aimed to examine how human health is affected by nonbiological factors that are associated with SDoH. The concepts were curated in concept map drafts from the review of SDoH, and the determinant of health model was used as a guide for concept development [13]. Later, we used Web Ontology Language, version 2 (OWL2) [14], the BFO [15-17], and semantic reasoners to construct and validate the ontology artifact.

**Review of Social and Economic Factors Impacting Health**

Peer-reviewed articles were searched and evaluated by the primary author on PubMed from September 17 to October 8, 2021. Boolean operators and MeSH (Medical Subject Headings) terms were used to refine literature searches conducted using the advanced search feature on PubMed. Multiple concepts and relationships were combined through Boolean expressions, that is, "Social determinants of health AND (health policy OR health care system)," to broaden the search. Certain phrases were enclosed in parentheses to isolate parts of the search query for precision and specificity. MeSH terms with regard to SDoH were provided by PubMed and used to construct the queries. A summary of each search is described in Table 1.

**Table 1. Literature search overview.** Advanced search queries for each level of the social determinants of health were searched on the PubMed database between September 17 and October 8, 2021. The table displays the query, applied filter, and number of results each search yielded (N=2833).

<table>
<thead>
<tr>
<th>Level</th>
<th>Search query</th>
<th>Results, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macrolevel</td>
<td>“Social determinants of health AND (health policy OR health care system OR health disparities)”</td>
<td>721 (25.45)</td>
</tr>
<tr>
<td>Macrolevel</td>
<td>“Income inequality AND welfare AND health policy”</td>
<td>10 (0.35)</td>
</tr>
<tr>
<td>Macrolevel</td>
<td>“Environmental determinants of health AND climate change”</td>
<td>216 (7.62)</td>
</tr>
<tr>
<td>Mesolevel</td>
<td>“Work OR socioeconomic position AND (health inequalities)”</td>
<td>300 (10.59)</td>
</tr>
<tr>
<td>Mesolevel</td>
<td>“Socioeconomic outcomes AND (housing OR food)”</td>
<td>291 (10.27)</td>
</tr>
<tr>
<td>Mesolevel</td>
<td>Food OR poverty AND (health inequalities)”</td>
<td>250 (8.82)</td>
</tr>
<tr>
<td>Mesolevel</td>
<td>“Social determinants of health AND (crime rate OR domestic violence)”</td>
<td>14 (0.49)</td>
</tr>
<tr>
<td>Microlevel</td>
<td>“Social determinants of health (gender OR age OR ethnicity OR race OR inequalities OR education)”</td>
<td>1031 (36.39)</td>
</tr>
</tbody>
</table>

Articles of interest must have met the following criteria: free full text available, publication date <10 years ago, and published in English. With accessibility in mind, free full text was included as an eligibility criterion. Older publications may have been relevant to this paper but were not considered because they may not reflect current knowledge. Thus, the publication date was set to <10 years ago. As English is the primary language of all authors of this study, it was included as an eligibility criterion for the literature search. Finally, the article type must have been a book or document, systematic review, journal article, observational study, case report, or clinical study. Collectively, the search queries yielded a total of 2833 nonduplicate citations.
Figure 1. Iterative process for gathering the articles of interest. The PubMed searches produced 2833 nonduplicate citations; by applying the inclusion and exclusion criteria described in the main text, we removed 2805 (99.01%) citations, leaving 27 (0.95%) articles for review.

Ontology Design and Development

The review helped us capture some basic salient high-level knowledge that we can encode into ontology from concept maps. The motivation is to gain a bird’s-eye view of SDoH and proceed from a top-down approach in developing the experimental ontology. We developed iterative multiple concept maps using draw.io [18] to identify concepts and relationship links among the concepts. Our analysis of the concept maps revealed 4 generalized relationships that bridged the various concepts: type of, part of, dependency, and causal. Figures 2-5 reveal the final drafted concept maps.
Figure 2. Determinants that impact health outcomes and behaviors. Dotted concept ovals indicate additional child concepts that are further described in Figures 3-5.

Figure 3. The relationship of concepts associated with macrolevel determinants. Concepts were derived from literature keywords, such as “health policy,” “income inequality,” “welfare,” and “environment.” Dotted concept ovals indicate additional child concepts.
Figure 4. The relationship of concepts associated with mesolevel determinants. This map displays the most detailed network of relationships and was formed from the following keywords: “work,” “work conditions,” “psychosocial work factors,” “socioeconomic position,” “socioeconomic outcomes,” “food,” “poverty,” “housing,” and “crime.” Dotted concept ovals indicate additional child concepts.

Figure 5. The relationship of concepts associated with microlevel determinants. Key elements of this map were gathered from keywords such as “physiology,” “gender,” “ethnicity,” “race,” and “behavior.” Dotted concept ovals indicate additional child concepts.

The part of relationship is illustrated with a forked link connection and indicates that 2 concepts were part of each other; for example, this is demonstrated in Figure 2 between the concepts “Macrolevel determinants” and “Social determinants of health,” where macrolevel determinants are one part (meronym) of the composition of SDoH (holonym). A dependency relationship was demonstrated as a dotted-line link connection and referred to concepts that were dependent on each other. This can be seen in Figure 3 between the concepts “Behavioral health” and “Mental health status,” where an individual’s behavioral health is dependent on the status of their mental health.

A causal relationship was represented as a thick line link connection and described 2 concepts that had a cause-and-effect relationship. An example of this is demonstrated with the concepts “greenhouse gas” and “extreme high temperature,” where there is a causal relationship between greenhouse gas and increased temperatures. Finally, a type of relation was illustrated as an open arrowhead similar to Unified Modeling Language (UML) notation.
Alignment With the BFO

To ensure semantic interoperability, we aligned our ontology with the BFO [15,17,21]. The BFO is an upper-level ontology that models entities using metalevel categories based on philosophical realism [16]. It is a widely regarded standard framework for creating biomedical and health reference ontologies that enable sharing, interoperability, and consistency with other ontologies by way of the metalevel categories and properties. To advance this work further, we aligned our exploratory ontology with a few of the metalevel concepts from the BFO. Currently, we have made some early attempts to align the object properties with OBO Foundry properties. Earlier, we identified 4 basic relationships that connect the concepts from our ontology model. We reviewed the BFO model and identified object property relationships that semantically correspond with our 4 relationship connections. The OBO Foundry’s part of or has part (BFO_0000050 [22]) object property was used to reflect the part of relationship [23]. The OBO Foundry’s causally related to (RO_0002410 [24]) object property reflected a causal relationship, and the OBO Foundry depends on (RO_0002502 [25]) object property was used to reflect the dependency relationship [23]. Naturally, the type of relationship was handled by OWL2’s SubClassOf axiom.

In addition to the identified property relationships, we settled on classifying the concepts using the 2 basic categories continuant (BFO_0000002 [26]) and occurrent (BFO_0000003 [27]). A continuant is defined as “an entity that persists, endures, or continues to exist through time while maintaining its identity” [26], essentially an entity or object. An occurrent is defined as “an entity that unfolds itself in time or it is the instantaneous boundary of such an entity (for example a beginning or an ending) or it is a temporal or spatiotemporal region which such an entity occupies_temporal_region or occupies_spatiotemporal_region” [27], basically an event or process. Each of the concepts in our model was classified into these 2 very basic classes from the BFO. Classifying these concepts into these BFO categories helped leverage the aforementioned property relationships because they were dependent on whether the connecting concepts were aligned with the BFO concepts.

Results

Overview

The literature search identified several topics of discussion for each determinant level. For macrolevel determinants, topics included health policy, income inequality, welfare, and the environment. For mesolevel determinants, the selected articles investigated areas such as work, work conditions, psychosocial work factors, socioeconomic position (SEP), socioeconomic outcomes, food, poverty, housing, and crime. Among all 3 levels, the highest number of articles for discussion were available for mesolevel determinants. Finally, the articles found for microlevel determinants examined gender, ethnicity, race, and behavior. In the following paragraphs, we discuss SDoH in detail.

Policy Making

Social policies and programs, fair employment and working conditions, and living environment are all likely to have the greatest impact on SDoH [30]. Social protection measures, increased coverage and quality of early years care, parental employment support, and gender equality in employment and education may improve early childhood development and even help to reduce child poverty. Affordable housing can be met through minimum housing standards and government actions [30]. Air quality legislation may have some benefits on air pollution and overall living [30]. The effects of climate change may be reduced by improving early warning systems and extreme weather preparedness.
Without action, climate change has the potential to raise agricultural prices, and this may threaten food security in low-income regions [31]. Families and individuals with low-income status are most susceptible to climate-related diseases such as malaria. Providing universal health care coupled with climate resilience measures is needed to reduce climate change impact on those with low-income status [31]. Bouzid et al [32] point out that several systematic reviews discuss diseases associated with climate change, but more focus should go toward the management of droughts, floods, air pollution, and food safety. The lack of research in these areas is likely due to the unpredictable nature of, for example, floods and government bodies that are primarily concerned with disaster response rather than research [32].

Policy Outcomes and Interventions
Health policies are fundamental for health and safety and are designed to improve quality of life. The most common types of implementation measures used to assess health policy outcomes include acceptability, feasibility, appropriateness, and compliance [33]. Well-tested quantitative measures are not used enough, and this may directly affect policy outcomes [33]. Most policy intervention tools at the school, district, state, or province level assess wellness policies from high-income countries such as the United States. Data from a systematic review showed that low- and middle-income countries lacked policy intervention initiatives [34]. Similar studies have investigated the relationship between income inequality and subjective well-being.

Evidence on the impact of social assistance on human health remains unclear [35]. Not enough articles discuss the differences between social assistance recipients and nonrecipients [35]. There is a lack of strong methods and study designs to evaluate the health effects of policies mainly in part due to insufficient data. Population-based health surveys do not provide enough information on respondent characteristics [35]. The available methods used to evaluate policy interventions require researchers to identify instances of large-scale policy change when social assistance programs are hardly ever affected by big changes. Instead, areas to be looked at are tobacco, food labeling, greater income redistribution, and labor market regulations [35].

A systematic review assessed randomized social experiments on social policy interventions for health outcomes in the United States and found that investments in early life, income support, and health insurance interventions may hold the potential to improve mental health and health in general [36]. The authors’ power analyses suggested that the models that were used were underpowered to detect health effects and outcomes. The authors noted that policy-related experiments should focus on design to accurately measure the relationship between health outcomes and policy interventions.

Income Inequality and Low SEP
According to a meta-analysis, income inequality was not influenced by measures used to assess subjective well-being or geographic region [37]. Instead, the level of country development, more specifically job opportunities, may be linked to income inequality in low- and middle-income countries. This serves as an indicator to government policy makers that reducing income inequality may lead to an improvement in subjective well-being [37]. While income inequality may have some effect on well-being, political economy may also influence population health. A systematic review revealed that there is a gap in the literature on many aspects of political economy, and it is unclear whether there is a relationship between political economy and population health [38]. Although there is no evidence, it seems that social democratic states with higher public spending tend to have better population health, but there is still no significant relationship between welfare state type and health inequalities [38].

In addition to income inequality, a low SEP may also contribute to poor health outcomes. There is consistent evidence that individuals who have a low SEP are often associated with hospital death and poor-quality end-of-life care [39]. Individuals with a poor education and who resided in impoverished neighborhoods were most likely to die in the hospital, receive acute-based care, and not receive specialized palliative care [39]. Future research on end-of-life interventions should consider SEP and its effects across the social strata [39].

Physical Environment and Health
A systematic review conducted by Lago et al [40] analyzed the relationship between health and physical environment, lifestyle, and social and economic conditions. On the basis of their evidence, the authors concluded that the main factor linking socioeconomic status and health status was income. Individuals with a higher level of income, as opposed to those with lower income, were associated with a lower chance of negative health outcomes [40]. The current association between income distribution and health is the general conclusion because individuals belonging to a lower social class have been shown to have worse average health. Different variables such as education may also play a role in determining health status because it is usually correlated to individuals’ social class [40]. Warmth and energy interventions may lead to improvements in respiratory health, mental health, and overall health for individuals with low-income status. Studies that targeted existing chronic respiratory diseases linked to inadequate warmth were most likely to see health improvement [41].

A mixed methods study demonstrated that energy performance interventions reduced energy use and helped raise indoor temperatures [42]. Despite there being a lack of evidence that suggests that energy performance investments improve health, data did show that improvements in social and economic conditions are better for overall well-being and health [42]. Economic conditions such as a low SEP are linked to poor health outcomes [42]. Individuals with a low SEP had an increased risk of cardiometabolic disorders and mortality according to Petrovic et al [43], who examined the role of health behaviors in socioeconomic equality in health. Behaviors such as smoking, alcohol consumption, physical activity, and diet were considered, as well as health outcomes such as cardiometabolic disorders and mortality. Of all behaviors examined, smoking contributed to the most social inequalities in health. The authors conclude that health behaviors may contribute to socioeconomic
inequalities, but this is dependent on population and study characteristics [43].

Impact of Food Availability on Nutrition

Individuals with low- and middle-income status are subject to food scarcity and poor nutritional health [44]. Supplementary feeding had a positive effect on weight and growth in low- and middle-income countries and was most beneficial to individuals who were poorly nourished. There were moderate positive effects on psychomotor development and mixed evidence on improved cognitive development [44]. Groups with lower income tend to select energy-dense diets that do not consist of vegetables or fruit [45]. Fats, refined grains, and added sugars are less expensive than nutrient-dense foods [45]. As a result, there may be a link between high obesity rates and low-cost calories [45]. Pregnant or postnatal women had an increased intake of fruits and vegetables after being enrolled in a food subsidy program [46]. Mean birth weight was slightly higher in 2 high-quality studies [46]. There is currently not enough evidence on the true impact of food subsidy programs for both children and adults [46].

Work Conditions and Occupational Health

Currently, no data suggest that workplace health promotion programs (WHPPs) increase socioeconomic inequalities in health, and there is not enough quantitative data on the ability of WHPPs to reduce social inequalities [47]. WHPPs seem to be the most helpful for working individuals who have a low SEP, but most of the programs were equally effective for groups from lower and higher socioeconomic backgrounds [47]. Most studies on working conditions supported the notion that adverse working conditions can mediate the association between SEP and well-being [48]. Studies that examined occupational categories or employment grades as indicators of SEP had the strongest findings in comparison to those that used education or income [48].

There is strong evidence that both physical and psychosocial factors are the cause of approximately one-third of the socioeconomic inequalities in health [49]. Despite limited longitudinal studies, cross-sectional evidence consistently showed that both physical and psychosocial work factors contributed to socioeconomic differences in self-rated health. Work factors may also play a role in inequalities, but there is not enough evidence to determine specific types of work factors [49]. In comparison to men, women experienced worse working conditions and higher job insecurity and also experienced poorer self-perceived physical and mental health [49,50]. Employed men had less emotional support, worked longer hours, and faced higher physical demands; however, they also held higher job statuses and had greater levels of effort-reward imbalance [50]. Although men were subject to more physically demanding tasks, women reported more musculoskeletal symptoms [50]. Health disparities between genders may stem from less favorable working conditions experienced by women [50]. Women are more commonly exposed to repetitive movements with low loads and awkward working positions than men [50]. Anthropometric differences in bone mass, fatty tissue, and muscle may also influence these health outcomes [50].

Socioeconomic Factors and Domestic Violence

Employment, income, social class, ethnicity, race, and living conditions all make up socioeconomic factors that may contribute to domestic violence [51]. The highest frequency of violence against women is found in a family environment, with the spouse being the most common perpetrator, and is most prevalent in low-income countries [51]. Individuals who experienced sexual dissatisfaction, unsatisfactory environmental conditions, and mental disorders tend to partake in acts of violence [51]. Certain countries have established laws to better protect women, but there needs to be an integrated approach for both national and international government organizations to achieve social change [51].

Discrimination and Poor Health Outcomes

The literature has shown a significant relationship between poor health and racism and a relationship with even higher significance between poor mental health and poor physical health [52]. Health outcomes indicated an association between racism and suicidal ideation, planning, and attempts. Depression was the most reported health outcome and had the same magnitude of association as racism [52]. Health care providers with different training, experience, and specialty backgrounds may hold implicit bias against racial and ethnic minority people [53]. A systematic review revealed that bias is associated with patient-provider interactions rather than health outcomes [53]. This indicates that patient-provider interaction can mediate the relationship between provider bias and patient health outcomes [53].

Institutionalized racism refers to the macrolevel systems, social forces, institutions, ideologies, and processes that interact with one another to cause inequities among racial or ethnic groups [54]. Although public health literature mentions the term institutionalized racism, it does not always engage with the concept or dive deep into the mechanisms through which health injustice is perpetuated [54]. To better understand racial and ethnic groups considered disadvantaged, the term should be explicitly mentioned in public health research as a central concept of health inequities [54]. Disparities in the neonatal intensive care unit exist in structure, process, and outcomes and generally disadvantage infants from racial and ethnic minority groups [55]. Hispanic and Black infants are most likely to receive care in poor-quality hospitals. In addition, hospitals serving racial and ethnic minority groups are underresourced and may lack quality improvement infrastructure. Quality improvement initiatives may have the best effect on populations considered disadvantaged who experience poor-quality care [55].

Gender Attitude and Sociocultural Norm

There may be several factors that can shape gender attitudes in early adolescence. In a study conducted in 29 countries, data demonstrated that young adolescents from varying cultures all express similar stereotypes and gender attitudes [56]. A gender study demonstrated that adolescents commonly endorsed norms that perpetuated gender inequalities such as masculinity established on toughness and skills or femininity based on physical appearance and shaming of sexuality [56].
Sociodemographic characteristics such as gender, race, immigration status, and age cause a variation in the results; however, family and peers are the central influences in building gender attitudes [56].

**Statistical Analysis**

Initial metrics from the ontology exhibited 245 classes, 47 object properties, and 346 logical axioms. Four evaluators independently reviewed 232 statements, specifically SubClassOf axioms, produced by the Hootation natural language translation software. Each statement was categorized as a 0 or a 1 to indicate expression accuracy. Statements that were not accurate were annotated as 0, and accurate statements were annotated as 1. Unsure responses were annotated as 0. The levels of agreement for each evaluator were calculated using a web-based program called ReCal3 (“Reliability Calculator for 3 or more coders”) [57]. Intercoder reliability was assessed through an average pairwise agreement and an average pairwise Cohen \( \kappa \) value [58].

Individual levels of agreement were as follows: evaluator 1=54%, evaluator 2=58%, evaluator 3=56%, and evaluator 4=76%. The average percentage agreement in terms of the average number of shared responses was 60.85% (SD 10.13%). The pairwise agreement also demonstrated that evaluator 2 and evaluator 4 had the highest similarity (74.14%) among shared responses, and the lowest percentage for shared responses was between evaluator 1 and evaluator 3 (48.71%). The pairwise agreements between evaluator 1 and 2 (56.47%), evaluator 1 and 4 (70.26%), evaluator 2 and 3 (60.35%), and evaluator 3 and 4 (55.17%) were recorded. The relationship among these results is demonstrated more accurately through the average pairwise Cohen \( \kappa \) value (0.19), which determined the interrater reliability. The results are presented in Table 2.

**Table 2.** Evaluation of subclass accuracy in percentage. Evaluators were asked to rate expression accuracy with 0 (no and unsure) and 1 (yes). The individual levels of agreement and disagreement are shown for each evaluator.

<table>
<thead>
<tr>
<th>Evaluator</th>
<th>Agreed (yes(^a); %)</th>
<th>Disagreed (no(^a) and unsure; %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluator 1</td>
<td>54</td>
<td>46</td>
</tr>
<tr>
<td>Evaluator 2</td>
<td>58</td>
<td>42</td>
</tr>
<tr>
<td>Evaluator 3</td>
<td>56</td>
<td>44</td>
</tr>
<tr>
<td>Evaluator 4</td>
<td>76</td>
<td>24</td>
</tr>
</tbody>
</table>

\( ^a \)Yes indicates the evaluator denoted a knowledge statement from the ontology was true, whereas, no indicates the evaluator assessed it to be false and unsure for if the statement was unknown to the evaluator to be true or false.

The average Cohen \( \kappa \) value was extremely low (0.19), as was the pairwise Cohen \( \kappa \) value for evaluators 1 and 3 (−0.04). The other Cohen \( \kappa \) values between evaluators 1 and 2 (0.12), evaluators 1 and 4 (0.38), evaluators 2 and 3 (0.19), evaluators 2 and 4 (0.44), evaluators 3 and 4 (0.04) were recorded. The statistical analyses helped identify concepts that required revision or omission. Statements that were classified as 0 were reviewed for analysis and possible error. Concepts with high levels of disagreement were revised, and new concepts were added to create a more logically structured ontology.

**Discussion**

**Principal Findings**

The ontology that was developed attempted to model the macrolevel and mesolevel conceptualizations of SDoH in more detail. Interpretations from the literature demonstrated that macrolevel factors are crucial determinants of health and health inequities. Individuals considered disadvantaged are almost always at risk for poor health and poor health outcomes. The main drivers of health inequalities seem to be a lack of education, affordable housing, basic housing needs, income, and access to health care. More specifically, women and racial and ethnic minority people are subject to these determinants, and this is the same for individuals living in low- and middle-income countries. Data from the articles also identified gaps in the literature for current research on low- and middle-income societies. Moreover, policy outcomes determined by SDoH can be measured in many ways; yet, there is little quantitative data on their validity. Finally, findings from the literature provided a solid foundation of knowledge and analysis that guided the design and development of the ontology.

Each of the 3 determinant levels interacts with, and dynamically influences, the other 2; therefore, delineation among the micro-, meso-, and macrolevel determinants is not always clear [59]; for example, the primary effects of discrimination are microlevel factors, such as the imposed psychological context from the individual enacting the discrimination and the individual experiencing it. However, the act of discrimination also has effects on the meso- and macrolevel determinants. The willingness of providers to live and work in underserved communities is considered a mesolevel factor, while the ability of the health care system to create recruitment and retention policies is a macrolevel factor. For an adequate transformation of these complex systems to occur, there will need to be an emphasis on the interactions among the levels and their interdependence [60]. Our work is imperative to the understanding of the ontology of SDoH because it will further the scholarly understanding of public health, lead to the development of necessary policy and interventional changes, and reduce the gap in health care literacy.

The statistical analyses from the evaluations were used to create a revised version of the ontology with a broad spectrum of knowledge concepts ranging from the macrolevel to microlevel determinants. Interpretation of the statements varied, and this may have posed a potential challenge for proper ontology evaluation; for example, the average Cohen \( \kappa \) values indicated that there was no effective agreement, implying that statements from the ontology were not accurate. The low levels of
agreement were mostly attributed to poor labeling and poor association between class and subclass axioms. Poor labeling referred to items that were not specific enough (eg, burnout→job strain, translated to “every burnout is a job strain”). Poor associations among expressions were found to be untrue or mislabeled (eg, poor→income inequality, translated to “every poor is an income inequality”). Personal opinions on statement evaluations were considered but not always incorporated for revision; for example, the concept poor energy performance was not understood by the first 3 evaluators, but it was cited in the literature and described poor energy efficiency in homes, such as poor heating or poor insulation [42].

After some iterative revisions of the ontology, we imported the minimal BFO concepts and property relationships discussed earlier into the Protégé environment and encoded the concept alignment with the BFO terms. We used the FaCT++ reasoner to perform a check of the logical consistency of our final aligned ontology model, and it revealed no inconsistent axioms. At the time of this writing, the core ontology exhibited 383 classes, 109 object properties, and 748 logical axioms, and we included an import of the Simple Knowledge Organization System ontology for additional annotation properties [61]. This preliminary ontology is currently hosted on GitHub [62]. Figure 6 shows a screenshot of the ontology in the Protégé tool with all essential concepts aligned (by assertion and inference) with the BFO categories and properties.

Figure 6. Screenshot of the experimental ontology in Protégé with alignment with Basic Formal Ontology concepts and properties.

Determining the accuracy of ontology concepts may help to produce a well-structured ontology. Moreover, appropriately addressing SDoH is fundamental for improving health and reducing long-standing inequalities. Modeling concepts transform metadata into a knowledge domain, which facilitates new knowledge discovery. By linking this ontology of SDoH
with other biomedical ontologies, researchers can make use of shared data for data exchange and information integration for biomedical tools such as computer-aided reasoning or decision support applications, enhance existing ontology knowledge bases, produce precise definitions of SDoH concepts in natural language, and provide a better understanding of the terminology associated with SDoH to reduce gaps in the literature.

Several concepts exist beyond the macro-, meso-, and microlevel determinants, which are included in the final version of the ontology. Concepts that impact or contribute to SDoH include academic degree, access to food, access to health care, behavioral health, discrimination, distance to medical resources, economic growth, economy, employment status, environmental determinants, exposure to crime, disease, food security, gross domestic product, gender attitude, gender identity, health inequities, health literacy, health outcomes, health services, health care coverage, history of incarceration, income, individual behavior, media culture, medical conditions, military service, national employment rates, nutritious diet, patient engagement, patient safety, personal health management, quality health care interventions, quality of life, sexual activity, sexual orientation, social class, social constructs, and vaccine. Each of these items contains additional subclasses (Figures 2-5).

**Macrolevel Determinants**

Class axioms for the macrolevel determinants included government programs, health care system, income inequality, macroeconomic conditions, macroeconomic policies, multisectoral approach, public policy, social security benefits, and social welfare. Each of these classes has been broken down further, as illustrated in Figure 3. Government programs, social security benefits, and social welfare were created to assist individuals who belong to a low social class, have a secondary-level education or less, and who are unemployed or work minimum wage jobs [30]. Both national and local governments intend to improve overall health by formulating macroeconomic policies and implementing multisectoral action initiatives to develop comprehensive strategies for addressing SDoH, promote inclusion and transparency in decision-making, and adopt equity-focused approaches in planning and resource allocation [30].

Currently, the US federal government mandates several public policies to improve the quality of life through the drug policy, agricultural policy, water policy, and energy policy [12]. Macroeconomic conditions such as employment and inflation can help regulate the economy, but these are highly dependent on current national employment rates [38]. Likewise, fiscal policies may help to reduce government spending, control debt, and regulate taxation, which in turn controls the economy [38]. Findings from the literature are reported on adult populations and rarely focused on children.

**Mesolevel Determinants**

The focal point for the mesolevel determinants is the physical environment. It is the level that contained the highest number of classes and subclass axioms. In addition to physical environment, classes included access to recreational activity, affordable housing, crime level, geographic location, psychosocial factor, psychosocial work factor, residence quality, residential location, residential safety, transportation, transportation quality, and walkability. The concepts that warrant the most discussion are physical environment, residence quality, and psychosocial work factor. The environment in which an individual lives and works affects their ability to function and socialize. The quality of housing has major implications on health outcomes [41]. Evidence suggests that poverty and low income affect housing circumstances.

Poor residence quality, such as insufficient heating or insufficient ventilation, may lead to illness [42]. Likewise, poor housing conditions such as mold presence, overcrowding, and unrepaired damage to property may also affect healthy living. Negative health outcomes are associated not only with residence quality but also with work environment. Exposure to psychosocial work factors was linked to poor mental health status [49]. Working long hours and being subject to high physical demands can result in depression, burnout, or work exhaustion [49]. Undesirable working conditions may affect job performance and ultimately employment status [48]. Occupations differ in both psychosocial work factors and work conditions; therefore, these concepts could be elaborated further. Mesolevel factors are presented in Figure 4.

**Microlevel Determinants**

Microlevel determinant class axioms were identified as biological factor, bodily function, human physiology, individual factor, individual lifestyle, nutrition, participation in physical activity, and physical fitness. Each of these concepts has subclasses that are illustrated in Figure 5. The relationship between individual factors such as education and health is complex. Low educational attainment may result in poor health. Cognitive disabilities and health conditions may affect educational outcomes, which in turn affect health literacy [63]. Low health literacy is associated with poor health outcomes and mortality. Individuals who do not understand the severity of their health conditions are less likely to seek medical care [63]. Poor operation of bodily functions may also result in undesired health outcomes. Likewise, poor management of diet and nutrition can affect physical fitness [45]. Individuals with a low SEP are subject to food insecurity and often malnourished [44]. Their inability to purchase food or healthy food options reflects their diet and nutritional status, resulting in illness [44].

Another microlevel factor that disrupts healthy living is discrimination. Individuals who are discriminated against for their race, gender, sexual orientation, disability, or age may experience depression and suicidal ideation. Discrimination that occurs in a hospital setting is prominent against African American and Hispanic individuals and results in poor or delayed treatment [53]. Negative gender attitudes may elicit aggressive behavior and lead to domestic or physical violence [56]. Attitudes toward gender may be attributed to sociocultural norms or individual beliefs; for example, individuals living in low- and middle-income countries with a high poverty rate often express toxic masculinity [56].

Individuals who identify as lesbian, gay, bisexual, transgender, or queer are targets for discrimination, bullying, isolation, and violence [64]. This is true even in the health care system, where
transgender women are commonly admitted as men, despite them expressing their gender [64]. Similarly, the normalized societal attitude toward individuals with disabilities is often exclusionary [64]. As health systems are often not designed with the needs of individuals with disabilities in mind, these individuals frequently face challenges, requiring them to navigate and challenge established norms [64]. Overall, findings from the literature emphasized that microlevel factors play a large role in human behavior and health outcomes.

Conclusions
In this paper, we examined the range of social and economic factors covering SDoH and modeled these aspects using ontology-based methods and tools to create a representational artifact. With this artifact, data and resources can be linked and aggregated to address clinical research that could analyze the link between the aforementioned factors and possible biological factors sourced in published bioinformatics ontologies. To our knowledge, this is the first ontology to focus on knowledge concepts that are not addressed by current biomedical ontologies for SDoH. The latest version of this ontology is available on GitHub [62] for public early release and future updates. Overall, this preliminary work is a demonstration of the possibility to model these heterogeneous social and economic concepts that can be aligned with the greater body of biomedical ontologies. However, the social and economic scope of SDoH is expansive, and although the ontology is broad, it is still in its early stages and could be expanded further with more granular social and economic concepts. Future consideration will be given to developing specific subdomains that can act as federated modules that can integrate with this ontology. Finally, we will include further aligning of this work with the BFO, using more precise semantic properties to accurately reflect the relationships among the concepts, which will provide further alignment with the existing validated biomedical ontologies.

Acknowledgments
This work was supported by the National Institute on Aging (RF1AG072799 and R56AG074604), and the Cancer Prevention Research Institute of Texas (RP220244).

Data Availability
The data sets generated during and analyzed during this study are available in the 3M Ontology repository [29].

Conflicts of Interest
None declared.

References
5. Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. Yeb Med Inform 2008:67-79 [FREE Full text] [Medline: 18660879]
17. Home. Basic Formal Ontology. URL: http://basic-formal-ontology.org/ [accessed 2024-03-04]
18. Untitled diagram. draw.io. URL: https://app.diagrams.net/ [accessed 2024-03-05]
22. part of. OLS Ontology Search. URL: http://purl.obolibrary.org/obo/BFO_0000050 [accessed 2024-02-27]
25. depends on. OLS Ontology Search. URL: http://purl.obolibrary.org/obo/BFO_0000009 [accessed 2024-02-27]
29. 3M ontology. GitHub. URL: https://github.com/UTHealth-Ontology/3M [accessed 2024-03-05]
30. Key policies for addressing the social determinants of health and health inequities. World Health Organization. 2017 Aug 03. URL: https://www.who.int/publications/i/item/9789289052658 [accessed 2024-03-04]


62. 3M Ontology. GitHub. URL: https://github.com/UTHealth-Ontology/3M [accessed 2023-09-17]


Abbreviations

BFO: Basic Formal Ontology
MeSH: Medical Subject Headings
OWL2: Web Ontology Language, version 2
SDoH: social determinants of health
SEP: socioeconomic position
STAR: situation, task, action, and result
WHPP: workplace health promotion program

©Daniela Dally, Muhammad Amith, Rebecca L Mauldin, Latisha Thomas, Yifang Dang, Cui Tao. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 13.03.2024. 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 the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.
Vaping: Public Health, Social Media, and Toxicity

Yehao Sun¹; Prital Prabhu¹; Dongmei Li²; Scott McIntosh³, PhD; Irfan Rahman¹, PhD

¹Department of Environmental Medicine, University of Rochester Medical Center, Rochester, NY, United States
²Department of Clinical & Translational Research, University of Rochester Medical Center, Rochester, NY, United States
³Department of Public Health Sciences, University of Rochester Medical Center, Rochester, NY, United States

Corresponding Author:
Irfan Rahman, PhD
Department of Environmental Medicine, University of Rochester Medical Center
Box 850
601 Elmwood Avenue
Rochester, NY, 14642
United States
Phone: 1 5852756911
Email: Irfan_Rahman@urmc.rochester.edu

Abstract

This viewpoint aims to provide a comprehensive understanding of vaping from various perspectives that contribute to the invention, development, spread, and consequences of e-cigarette products and vaping. Our analysis showed that the specific characteristics of e-cigarette products as well as marketing strategies, especially social media marketing, fostered the spread of vaping and the subsequent effects on human health and toxicity. We analyzed the components of e-cigarette devices and e-liquids, including the latest variants whose impacts were often overlooked. The different forms of nicotine, including salts and freebase nicotine, tobacco-derived nicotine, tobacco-free nicotine, and cooling agents (WS3 and WS23), have brought more choices for vapers along with more ways for e-cigarette manufacturers to advertise false understandings and present a greater threat to vapers’ health. Our work emphasized the products of brands that have gained significant influence recently, which are contributing to severe public health issues. On the other hand, we also discussed in detail the toxicity of e-liquid components and proposed a toxicity mechanism. We also noticed that nicotine and other chemicals in e-liquids promote each other’s negative effects through the oxidative stress and inflammatory nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) pathway, a mechanism leading to pulmonary symptoms and addiction. The impact of government regulations on the products themselves, including flavor bans or regulations, has been limited. Therefore, we proposed further interventions or harm reduction strategies from a public health perspective.

(Keywords: addiction; behavior; behaviors; device; devices; e-cigarette; e-cigarettes; effect; effects; e-liquid; flavors; nicotine; public health; social media; human health; toxicity; vape user; vape; vapor; vapers; vaping)

Vaping: A Public Health Issue

Recently, e-cigarette use among the adolescent population has become a concerning public health challenge in the United States. The popularity of the relatively new invention of e-cigarettes reversed the nation’s effort to reduce tobacco use among youths in the past 30 years [1]. According to data collected in 2021 in the United States, 11% of young adults aged 18-24 years currently use e-cigarettes, and 24% of 12th-grade students reported having engaged in vaping activities in the past 30 days [1,2]. This public health issue is not limited to the United States, however, as the prevalence of past 30-day vaping among adolescents aged 12-16 years in 68 different countries worldwide was 9.2% (2021) [3]. This is especially concerning, as e-cigarette use among early-adolescent smokers is a predictor of any smoking and more frequent cigarette smoking in late adolescence [4]. Besides the impact on smoking behaviors, adolescents who vape can develop severe respiratory symptoms that may even require ventilation, and other symptoms, including gastrointestinal tract symptoms, coughing, and hypoxemia, can also arise throughout the body [5].

e-Cigarette use has become a worldwide public health issue that demands our attention. To gather information and present it to the public, this paper has taken literature from PubMed and Google Scholar into consideration and introduced the advantages and disadvantages of vaping based on current literature by...
including conflicting reports and potential biases on social media and public health.

**e-Cigarette Brands and Characteristics**

Due to the public impact of e-cigarette use, it is important to understand the characteristics of popular e-cigarette products. According to the 2022 National Youth Tobacco Survey, Puff Bar (29.7%), Vuse (23.6%), JUUL (22%), SMOK (13.5%), NJOY (8.3%), Hyde (7.3%), and Blu (6.5%) were found to be the most popular brands among youth [6]. Another study by Ali et al [7] showed that during the 4-week period that ended December 25, 2022, the top 5 brands with the highest e-cigarette unit sales were Vuse, JUUL, Elf Bars (Funky Vape or EB Design), NJOY, and Breeze Smoke. Among the products from all the brands mentioned above, only products from SMOK were third-generation e-cigarettes (also known as mods, a type of highly customizable aerosol-generating devices that use e-liquids and have subohm tanks that allow for higher wattage), while all other products were fourth-generation e-cigarettes (also known as pod mods, a type of modifiable pod cartridges including vape bars that use nicotine salts and can come in different shapes). Most fourth-generation e-cigarettes introduced above contain salt-form tobacco-derived nicotine (TDN) with a concentration of ≤5% and a volume below 2 mL. The products from Hyde are unique in that some of them contain larger volumes (up to 10 mL) that support a higher number of puffs, and these products may contain tobacco-free nicotine (TFN). The characteristics of TDN and TFN are discussed in a subsequent section.

In 2023, however, several new brands became more popular, including Elf Bars (EB Design, Lost Mary, Funky Vape, BC5000, and BC 7000), all from iMiracle Associates (originated from Shenzhen, China), FLUM float bars, ESCO bars, and Tyson. All of the products within these brands are fourth-generation e-cigarettes with nicotine salts as the source of nicotine. All the products contain a larger volume (up to 13 mL) and thus more puffs (≥2500 puffs). Only some products from iMiracle contained TFN, while the other brands only used TDN.

Among all the popular brands, Elf Bars (Funky Vape or EB Design, now rechargeable with pods) from iMiracle requires the most attention due to worries of it becoming the next “JUUL” that produces massive public health issues (e-cigarettes by JUUL were extremely popular from 2017 to 2020 and caused massive public health issues) [8]. In the youth population in England, around 50% of past 30-day vapers reported the use of Elf Bars (EB Design) [9]. Researchers have also suggested that the increase in vaping frequency and the shifted interest toward disposable e-cigarettes among English vapers could have been driven by the popularity of Elf Bars (EB Design) [9]. Moreover, an analysis of data from the National Poison Data System showed that 60.8% of e-cigarette–associated cases in poison centers with reported brand information were related to Elf Bars (EB Design) [10]. Therefore, its popularity, its ability to change vapers’ vaping behaviors that were demonstrated by the shift in vapers’ interests, and its potential for causing severe health issues all make Elf Bars (EB Design) a dangerous brand from a public health perspective, and public health interventions should be applied to prevent another “JUUL” from emerging and causing severe issues.

In addition to the most popular brands on the market, another set of brands also requires our attention, despite not being as popular. Both freebase nicotine and nicotine salts are used in brands including EC Blend (uses TDN), Halo (TDN), Coastal Clouds (TDN), Bad Drips (TDN and TFN), Naked (TDN and TFN), Cloud Nurdz (TFN), and Primus (TDN); however, these compositions change regularly. e-Cigarettes originally used freebase nicotine, while more recent e-cigarettes started using nicotine salts, which have a lower pH and therefore are less irritating to the throat and allow for higher doses [11].

The difference between freebase nicotine and nicotine salts is not restricted to user experience, as different physiological and toxic properties were observed. Research in pharmacokinetics has shown that nicotine salts are absorbed more rapidly and can reach higher concentrations, while freebase nicotine is metabolized more slowly, causing it to remain at higher concentrations in male rats, which may be generalizable to humans [12]. As advertised, these characteristics allow vapers to experience an instant nicotine “hit” or a prolonged sensation, or they can even achieve both by using both types of products. However, the availability of both types of products in these brands allows vapers to encounter the different adverse effects of either type. Generally, nicotine salts produce changes in levels of more types of proinflammatory cytokines and have more complicated effects on human nasal epithelial cells, but freebase nicotine can also cause unique changes, including increased secretion of interleukin-7 [13]. The results above were yielded with the same nicotine concentration, while nicotine salt products often have double the nicotine concentration than freebase nicotine products [14]. Therefore, the toxicity of nicotine salt products should be more serious than that of freebase nicotine products, but the threat of the availability of freebase nicotine along with nicotine salt is also not neglectable.

**e-Liquids Composition, Toxicology, and Associated Pulmonary Symptoms or Responses**

Besides the toxicities specific to nicotine salts, there are various toxic effects of almost all the components in the e-liquids [15-24]. The most abundant chemicals, propylene glycol, and vegetable glycerin, can negatively affect cell viability as they can cause decreases in cell growth to a similar degree to dimethyl sulfoxide [15]. Propylene glycol is known to cause respiratory arrest in rats after administration of 25 mg/kg/day for 3 days [16]. Chronic propylene glycol exposures are associated with reported symptoms of chronic wheezing, chest tightness, and weaker lung function, while acute exposures are associated with coughs and other upper airway symptoms as well as ocular irritations [17,18]. Heating vegetable glycerin in the presence of other acids may cause it to undergo pyrolysis reactions, releasing acrolein, which can cause nasal cavity irritations, lung lining damage, and even contribute to the onset of cardiovascular diseases [19-21]. Furthermore, the chemical
that vapers are addicted to (i.e., nicotine) can induce health issues by increasing reactive oxygen species, causing lipid peroxidation, and damaging human DNA [22] (Figure 1).

The most commonly used flavorants in e-liquids mentioned above also have pronounced toxicities [15,23,24]. Vanillin, the flavorant used in 35% of e-liquids, is positively correlated with the toxicity of the e-liquids ($R^2=0.62$) [15]. Meanwhile, ethyl maltol leads to incidences of kidney lesions in rats and mild hemolytic anemia in dogs [23]. Ethyl butyrate, a type of ethyl ester flavor additive, can be broken down under high temperatures (may be achieved by atomizers) into carboxylic acids [24]. These carboxylic acids can then decompose into ketene, a chemical known to be a strong respiratory poison that can cause severe lung damage even at lower concentrations [24]. Generally, it is shown that the more chemicals there are in an e-liquid and vaporized (aerosols), the higher toxicity that e-liquid is likely to have [15].

**Figure 1.** Biological pathways to pulmonary responses and nicotine addiction. NF-κB: nuclear factor kappa-light-chain-enhancer of activated B cells; TLR: toll-like receptors; TRP: transient receptor potential-like channels.

Another group of deleterious chemicals that are present in e-liquids and aerosols are volatile organic chemicals (VOCs), including benzene, aldehydes, toluene, and ethylbenzene [25]. VOCs are significant risk factors for asthma among children [26]. In fact, the risk for asthma doubles for every 10 μg/m$^3$ increase in the concentration of toluene, and it even triples for benzene [26]. Meanwhile, VOCs are also associated with reduced lung function in all age groups [27]. Besides the chemicals in the e-liquids themselves, vapers may inhale other toxic chemicals, including heavy metals (cadmium, lead, nickel, copper, arsenic, and chromium), while vaping and experience health risks, including increased oxidative stress, DNA damage, and decreased cellular viability in tissues [28,29]. It is observed that later puffs contain higher concentrations of toxic metals, and these higher concentrations can lead to a 30% increase in DNA damage after a 7-day exposure [28]. The commonly used flavorant, ethyl maltol, can also enhance the absorption and toxicity of copper in lung epithelial cells, as apoptosis, DNA damage, and oxidative stress occurred after coexposure to ethyl maltol and copper at low concentrations, which cannot initiate toxic effects individually [30]. Such toxicities can be applicable to most vapers, as higher cadmium, lead, silver, vanadium, nickel, and chromium concentrations were found in the urine or serum of e-cigarette users [31-33].

Due to the toxicity of these chemicals, although e-cigarettes are potentially less harmful than traditional cigarettes (less nitric oxide, oxidants, aldehydes, and no carbon monoxide inhaled), they are not as safe as expected by vapers [34]. Moreover, studies that demonstrated less harmful effects for e-cigarettes mostly focused on short-term effects, so whether the long-term health effects of e-cigarettes are also lower remains unclear [34]. More longitudinal studies should be conducted to confirm the long-term, chronic effects of e-cigarette use.

**Tobacco-Free and Tobacco-Derived Nicotine**

In addition to the toxic effects brought by the presence of certain chemicals, the source of such chemicals may have an effect as well. The source of nicotine in e-liquids can be divided into 2 categories: TDN and TFN (nicotine that is not derived from tobacco plants). The type of nicotine originally used in e-cigarettes was TDN, but TFN started to be used in e-cigarettes since products containing TFN were not defined as tobacco products by the Federal Food, Drug, and Cosmetic Act and thus could evade regulations. e-Cigarette manufacturers have also advertised TFN as safer and with smoother flavors, causing vapers to believe that TFN has lower risks and to be willing to use and even pay more for TFN products [35]. Similarly, young adults have an overall positive perception of TFN products’ flavors, and those who are willing to try TFN products believe that TFN products are less addictive [36]. Fortunately, on April 14, 2022, new legislation took TFN into consideration and
granting the Food and Drug Administration the authority to regulate TFN products, and TFN products are no longer a viable way to evade regulations [37].

Considering the chemical composition of TDN and TFN, the fundamental difference between them is chirality. The nicotine in TDN products is mostly S-nicotine (only 0.1%-1.2% is R-nicotine), while the nicotine in TFN products is a racemic mixture of R- and S-nicotine (50% R-nicotine and 50% S-nicotine) [38,39]. There is limited information on how the chirality of nicotine may impact its health effects, but it is stated that TFN is pharmaceutically pure and does not contain impurities found in TDN, including tobacco-specific nitrosamines, which may contribute to negative health effects [40,41]. Research has shown that TFN products of beverages or minty or iced (containing cooling agents, such as WS3 and WS23) flavors generate significantly less reactive oxygen species (ROS) than their TDN counterparts, which concurs with the statement earlier [42]. However, no significant difference was detected with fruity or tobacco flavors, indicating that flavorants also play a role in the process of generating ROS [42]. Therefore, more research should be done to confirm the difference posed by TFN itself, and cellular experiments are also crucial to understanding the inflammatory response of humans to the different forms of nicotine.

Mechanism of Toxicity

Overall, the inhalation of e-liquid aerosols can lead to lung injuries, and Kaur et al [43] proposed mechanisms for this process. In this mechanism, the chemicals in the aerosol produce ROS in the lungs and induce oxidative stress, DNA damage, and epithelial barrier dysfunction. These cellular toxicities then trigger the transient receptor potential-like channels or toll-like receptors, activate the NF-xB complex, and stimulate the release of inflammatory cytokines, including interleukin-1 beta, interleukin-6, and tumor necrosis factor-alpha, which are all parts of inflammatory responses [43]. These inflammatory responses then manifest as symptoms by inducing endoplasmic reticulum stress, mitochondrial dysfunction, airway hyperactivity, epigenetic modifications, and other disease-developing mechanisms [43].

Besides the inflammatory aspects of the activation of the NF-xB complex, such activation can also contribute to the addiction to e-cigarettes [43-45]. The NF-xB complex is known to facilitate the positive reinforcing effect of drugs through reward sensitization in the nucleus accumbens, thus exacerbating the development and maintenance of nicotine addiction in vapers [44]. The major addiction development pathway, on the other hand, originates from the binding of nicotinic acetylcholine receptors with nicotine [45]. Signals are then sent to the reward system in the central nervous system, including the nucleus accumbens, to reinforce the behavior and eventually lead to addiction [45]. Figure 1 describes the known biological pathways of the toxicity of and addiction to e-cigarettes.

As a result of these pathways, the addiction potential of e-cigarettes is not lower than that of traditional cigarettes [46]. A study that used the Fagerström test for nicotine dependence showed that exclusive e-cigarette users (mean 3.5) have a nicotine dependence level more than 2 times higher than that of traditional cigarette smokers (mean 1.6) [46]. Dual users also demonstrated higher dependence when using e-cigarettes (mean 4.7) than when using traditional cigarettes (mean 3.2) [46]. The high addiction potential of e-cigarettes may make them unsuitable for being used as smoke cessation tools, especially in young adults who have a higher risk of addiction to e-cigarettes, as demonstrated in the study [46].

Social Media and Vaping

In 2022, the number of social media users in the United States reached over 302 million. Around 90% of the US population used social media as of 2023. Social media platforms, such as Twitter, Reddit, Instagram, TikTok, YouTube, and Facebook, have become increasingly popular, especially among youth and young adults. Instagram and YouTube are the most broadly used social media platforms among youth in the United States, with 80% of youth using YouTube and 72% of youth using Instagram [47].

With the increasing popularity of social media platforms in the United States, e-cigarette companies and vape shops have aggressively marketed vaping products on social media [48-52]. Vaping marketing and promotion posts dominate vaping-related social media posts with more user engagement [47,50,52]. Social media accounts of e-cigarette companies or vape shops usually post well-designed pictures with vaping products, images linking vaping with luxury lifestyles, price promotions, discounts, and product giveaways [53-55]. Examples of extensive use of such strategies include the “Doit4juul” campaign on social media platforms, including Instagram and YouTube, initiated by the Juul manufacturer, and the campaign largely contributed to the rapid growth of the brand [56]. They also sponsor influencers with many followers to help them market vaping products [57,58]. Around US $75 million (inflation-adjusted 2021) was spent on marketing vaping products by the vaping industry in the third quarter of 2019, including marketing and promotion on various social media platforms that reach most of the US population, including youth under 18 years [57,59,60]. The massive marketing and promotions of vaping products on social media resulted in the misperception of vaping as a harmless activity [58]. They also increased the risk of vaping initiation, especially among youth and young adults [61-63]. Research also found that the initiation of vaping is associated with subsequent cigarette smoking [64,65]. However, a recent epidemiological longitudinal survey study using Population Assessment of Tobacco and Health Wave 1-5 data (2013-2019) implicates that baseline vaping is not associated with subsequent cigarette smoking initiation in youth who have never smoked before after adjusting for behavior risk factors, such as alcohol, marijuana, and other tobacco product use [66]. Fortunately, the advertisements for combustible cigarettes are more restricted by the government, and combustible cigarette companies can only focus on marketing at the point of sale and product packaging [67,68].

Besides the massive marketing and promotion activities of vaping industries, the public also uses social media platforms to share their opinions and user experiences on vaping products.
Our longitudinal examination of the vaping flavors mentioned in 2.8 million Reddit posts from January 2013 to April 2019 showed that the top 2 flavors were fruit and sweet, consistent with previous survey results during a similar period [50]. A further examination of the association of vaping with health symptoms mentioned in the same Reddit posts showed a significant comentioning of fruit-flavored vaping products and cardiovascular symptoms [48]. A sentiment analysis of over 2.7 million vaping-related Twitter posts (tweets) from May 31 to August 22, 2019, found the fruit, mint, and sweet flavors were more positively perceived, and the beverage and tobacco flavors were more negatively perceived by the public [50]. A systematic review of the vaping-related social media studies from 2007 to 2017 found the major topics related to vaping on social media included the health effects of vaping, vapor testimony, benefits and risks associated with vaping, regulations of vaping products, and vaping as smoking cessation aids [69]. Being exposed to antivaping content on social media was associated with reduced vaping activities among young people [70]. Vaping-related social media posts could also be used to examine the impact of vaping product regulation policies on public attitudes toward vaping [71–73]. With the high prevalence of vaping in youth and young adults and the e-cigarette or vaping use-associated lung injury (EVALI) outbreak in 2019, many state governments and the FDA started to ban flavored vaping products.

Overall, social media plays an important role in vaping product marketing, promotion, and communication with the public about the potential harms of vaping. Regulations on social media marketing of vaping products can help reduce vaping initiation in youth and young adults. Meanwhile, social media platforms could also be used to provide education to the public about the potential harms of vaping and deliver vaping cessation interventions or harm reductions to reduce the uptake of vaping in youth and young adults. An example of such an application is the Real Cost e-Cigarette Prevention Campaign [74]. However, the low frequency of educational posts and the lack of appeal to youth social media users may impact the efficacy of the intervention, and more research should be done to address these issues and improve its efficacy [74].

Health Effects of Vaping

Due to the recent and fast e-cigarette popularization with the contributions of social media, there is limited scientific evidence regarding its exact long-term effects. However, e-cigarette use has caused the onset of respiratory symptoms among consumers, labeled EVALI, the etiology of which involves the inhalation of vitamin E acetate (VEA) along with other chemicals emitted from e-cigarettes [75]. Patients with EVALI present to the hospital with sterile exogenous pneumonitis and the symptoms of cough, chest pain, dyspnea, nausea, gastrointestinal tract symptoms, fatigue, and fever [76]. Although the identification of VEA in bronchoalveolar lavage fluid is commonly associated with EVALI, sampling and identifying VEA from bronchoalveolar lavage fluid are not widely available, so the diagnosis of EVALI remains to be done through exclusion [77]. Nevertheless, as of 2020, almost 3000 cases of lung injury hospitalizations related to vaping have been reported in the United States [78]. Lung biopsies taken from 8 of these patients in various centers revealed acute lung injuries, including organizing pneumonia, diffuse alveolar damage, or interstitial inflammation [79]. In addition to respiratory damage, e-cigarette use can have effects elsewhere. In a recent study, researchers noticed an association between e-cigarette use and seizures in youth, perhaps due to the high levels of nicotine or flavoring chemicals inhaled when vaping—this finding also raises concerns about the impacts of e-cigarette use on brain development and other neurological complications in the youth population [80]. e-Cigarettes have been shown to produce an increase in blood pressure and aortic stiffness, lending themselves to further cardiovascular stress [81], and can also be potentially carcinogenic due to formaldehyde-releasing agent formation during the vaporization process [82] and high levels of nitrosamines present in e-cigarette flavorings [83].

If used as a cessation method for traditional smoking, e-cigarettes have the potential to limit traditional smoking. However, e-cigarette corporations’ decision to expand their consumer base past quitting smokers ushered in a new generation plagued by nicotine addiction. Since 1999, combustible nicotine intake has been steadily decreasing among high schoolers, down from an average of 5 days per month to only 1; however, after the popularization of e-cigarettes among school-aged children in 2015–2017 (coinciding with the release and rapid growth of the JUUL corporation), these numbers have begun to rise, once again approaching 5 days [84]. While nicotine addiction in youth may be lucrative for e-cigarette corporations, it perpetuates a cycle of toxicant inhalation, leading to the dangerous symptoms described above.

Public Health Interventions Against Toxicity

The serious health effects and the wide spread of e-cigarettes led to the urgent need to address the public health issue. However, the sole motivation of the tobacco industry is to maximize profit, so it is unlikely it will ever be motivated to abandon profit and help address the public health issue of tobacco product use [85]. Historically and to this day, this profit has clearly been at the expense of public health. Such expenses include the use of aggressive tactics, including lobbying for industry-favorable laws and regulations [86], marketing directly and indirectly to youth [87], targeting many channels and subpopulations with evolving marketing strategies [88], and advanced public relations approaches to undermine and misrepresent evidence-based science by inflating scientific uncertainty to undercut public health initiatives and regulatory actions [85,89]. Many industries have subsequently adopted this game plan of disrupting normative science, which leads to an assertion of personal accountability for what are actually industrially generated health hazards [86,90].

Despite overall drops in rates of smoking (including mentholated products), menthol product use has increased, especially among young adults, female individuals, and Black users. Bans on menthol, a flavoring that has historically been disproportionately
marketed to African American communities and associated with a lower likelihood of cessation among persons from these communities, were thought to have a high potential for increasing cessation in the communities [91]. Governments have taken steps to outlaw the retail sale of flavored tobacco products, including menthol cigarettes. The regulations primarily include the requirement of premarket tobacco product applications for e-cigarettes (as in traditional cigarettes), raising the age restriction of e-cigarette purchases to 21 years, broadening smoke-free policies and high taxation (originally only against traditional cigarettes) to e-cigarettes, and flavor bans [1]. These regulations target different aspects of e-cigarette use, but they all have imperfections in that the implementation of such regulations has a long processing time and not all states and jurisdictions would choose to implement such policies [1]. Therefore, e-cigarette sales are still largely underregulated.

On the other hand, the tobacco industry has once again aggressively fueled erroneous information about these policies, including the claim that such flavor bans target African American smokers’ freedom of choice, in service of their goals to protect profit [92]. The example demonstrated that the industries would exert all their effort and resources to oppose such prohibitions [92].

Telephone quitlines have demonstrated effectiveness for smoking cessation for more than 20 years [88]. They can produce both short-term and long-term intervention effects [88]. It has also been shown that promotions through television and radio have tripled call rates [88]. It is reasonable to infer that such promotions and best practices can be effective for vaping cessation, especially when promoted on social media. Since young adults and adolescents are the most vulnerable to nicotine addiction and the other negative effects of vaping, it is important to analyze the quitlines’ effectiveness in these understudied populations [89]. More research is needed to analyze the effectiveness of quitlines on adolescent and young adult vape users, including analysis of the effects of quitline-related treatment modalities such as websites, user chat rooms, and SMS text messaging. Additionally, it is needed to analyze the reach and efficacy of social media promotion of such eHealth intervention platforms.

Summary and Conclusion

Overall, we have discussed the various aspects of vaping and its related toxicity. Social media marketing is an essential part of the companies’ strategies due to its high influence among youth. Their strategies are successful for their companies, as they gained substantial growth in their sales, while causing almost 3000 hospitalizations associated with vaping in the United States [78]. The most popular and most problem-causing brands discussed in the study should be monitored, especially Elf Bars (EB Design or Funky Vape or Funky Land), which is responsible for 60.8% of e-cigarette–associated cases in poison centers with reported brand information [10]. The toxicity mechanism of and addiction to e-liquids was also proposed in this study, including the chemicals’ contributions to nicotine addiction in vapers (Figure 1). More research should be done on the newer variants of e-liquid components to address the lack of understanding of these components and their potential to cause public health issues (ie, chronic effects on toxicities and health effects). In addition, a lack of knowledge about the long-term effects of vaping may undermine the determination of e-cigarette regulations, so longitudinal studies should be conducted to acquire the crucial information [34]. Investigations on the potential effectiveness of vaping cessation on the mitigation of both short- and long-term health effects would also be crucial in combating the vaping epidemic.

Considering the complexity of the toxicity that stemmed from the new e-liquid formula (bars and e-liquids with devices) and subsequent health issues, vaping cessation interventions or harm reduction should also address all these concerns. From a public health perspective, social media marketing of vaping products should be prohibited, and social media should instead be leveraged to provide education and cessation resources to the public. Government or federal regulations alone on flavored tobacco products may have a limited impact since corporations can bypass such regulations with their resources [87]. Viable intervention strategies should include telephone quitlines and related eHealth interventions, such as websites and SMS text messaging, which should be promoted to and tailored for adolescent and young adult populations.

Acknowledgments

We would like to thank Chad Newton, Alannah Dalton, Ryan Rahman, and Desjamebra Robinson for their insights and useful discussions. We would also like to thank the Western New York-based Center for Research on Flavored Tobacco Products (CRoFT #U54CA228110).

Disclaimer

A part of the manuscript was deposited in the preprint form on Preprints.org [93].

Authors’ Contributions

YS, PP, DL, SM, and IR contributed to the writing and original draft preparation. YS, PP, and IR were responsible for writing, reviewing, and editing. YS and IR prepared schematics and conceptual diagrams. IR contributed to supervision, editing, project administration, acquisition, and compilation. All authors have read and agreed to the published version of the manuscript.
Conflicts of Interest
None declared.

References


9. Hammond D, Reid JL. Trends in vaping and nicotine product use among youth in Canada, England and the USA between 2017 and 2022: evidence to inform policy. Tob Control 2023;tc-2023-058241 [FREE Full text] [doi: 10.1136/tc-2023-058241] [Medline: 37940402]


14. Pennings JLA, Havermans A, Pauwels CGGM, Krüsemann EJZ, Visser WF, Talhout R. Comprehensive Dutch market data analysis shows that e-liquids with nicotine salts have both higher nicotine and flavour concentrations than those with free-base nicotine. Tob Control 2023;32(e1):e78-e82 [FREE Full text] [doi: 10.1136/tobaccocontrol-2021-056952] [Medline: 34987081]


67. Advertising and promotion. US Food and Drug Administration. 2020. URL: https://tinyurl.com/35dxn3r7 [accessed 2023-12-22]


71. Gao Y, Xie Z, Li D. Investigating the impact of the New York State flavor ban on e-cigarette-related discussions on Twitter: observational study. JMIR Public Health Surveill 2022;8(7):e34114 [FREE Full text] [doi: 10.2196/34114] [Medline: 35802417]


Abbreviations

EVALI: e-cigarette or vaping use-associated lung injury
NF-κB: nuclear factor kappa-light-chain-enhancer of activated B cells
ROS: reactive oxygen species
TDN: tobacco-derived nicotine
TFN: tobacco-free nicotine
VEA: vitamin E acetate
VOC: volatile organic chemical
Machine Learning for Prediction of Tuberculosis Detection: Case Study of Trained African Giant Pouched Rats

Joan Jonathan1*, MSc; Alcardo Alex Barakabitze1*, PhD; Cynthia D Fast2,3,4*, PhD; Christophe Cox2*, MSc

1Department of Informatics and Information Technology, Sokoine University of Agriculture, Morogoro, United Republic of Tanzania
2APOPO Rodent Project, Sokoine University of Agriculture, Morogoro, United Republic of Tanzania
3Evolutionary Ecology Group, Department of Biology, University of Antwerp, Antwerp, Belgium
4Rutgers Center for Cognitive Science, Piscataway, NJ, United States

*all authors contributed equally

Corresponding Author:
Joan Jonathan, MSc
Department of Informatics and Information Technology
Sokoine University of Agriculture
PO Box 3038
Morogoro
United Republic of Tanzania
Phone: 255 763 630 054
Email: joanjonathan@sua.ac.tz

Abstract

Background: Technological advancement has led to the growth and rapid increase of tuberculosis (TB) medical data generated from different health care areas, including diagnosis. Prioritizing better adoption and acceptance of innovative diagnostic technology to reduce the spread of TB significantly benefits developing countries. Trained TB-detection rats are used in Tanzania and Ethiopia for operational research to complement other TB diagnostic tools. This technology has increased new TB case detection owing to its speed, cost-effectiveness, and sensitivity.

Objective: During the TB detection process, rats produce vast amounts of data, providing an opportunity to identify interesting patterns that influence TB detection performance. This study aimed to develop models that predict if the rat will hit (indicate the presence of TB within) the sample or not using machine learning (ML) techniques. The goal was to improve the diagnostic accuracy and performance of TB detection involving rats.

Methods: APOPO (Anti-Persoonsmijnen Ontmijnende Product Ontwikkeling) Center in Morogoro provided data for this study from 2012 to 2019, and 366,441 observations were used to build predictive models using ML techniques, including decision tree, random forest, naïve Bayes, support vector machine, and k-nearest neighbor, by incorporating a variety of variables, such as the diagnostic results from partner health clinics using methods endorsed by the World Health Organization (WHO).

Results: The support vector machine technique yielded the highest accuracy of 83.39% for prediction compared to other ML techniques used. Furthermore, this study found that the inclusion of variables related to whether the sample contained TB or not increased the performance accuracy of the predictive model.

Conclusions: The inclusion of variables related to the diagnostic results of TB samples may improve the detection performance of the trained rats. The study results may be of importance to TB-detection rat trainers and TB decision-makers as the results may prompt them to take action to maintain the usefulness of the technology and increase the TB detection performance of trained rats.

KEYWORDS
machine learning; African giant pouched rat; diagnosis; tuberculosis; health care
**Introduction**

**Background**

African giant pouched rats (*Cricetomys ansorgei*) are native to sub-Saharan Africa, making them resistant to local parasites and diseases [1]. The term “pouched rat” refers to their large cheek pouches that are used for carrying food back to their burrows, where the food is either eaten or stored. These rats are nocturnal and omnivorous, eating various insects, fruits, and vegetables. They are large (adult males and females weigh about 1.3 kg and 1.2 kg, respectively) and are long-lived, averaging 8 years in captivity. Moreover, they have a highly developed olfactory capacity, enabling them to do specific detection tasks with training [2]. As such, in 1997, APOPO (Anti-Persoonsmijnen Ontmijnende Product Ontwikkeling or “Anti-Personnel Landmines Detection Product Development” in English) started researching how to train these rats for scent detection. APOPO is a Belgian nongovernmental organization whose mission is to protect people and the planet using scent detection animals [3]. Rat pups born at APOPO’s breeding facility are weaned from their mother at 10 weeks old. Rats begin training in a custom-engineered line cage immediately after they are weaned. Training for tuberculosis (TB) detection takes place in this apparatus, which requires upwards of 9 months to master. Each rat’s home cage is outfitted with a clay nest pot to simulate the rat’s natural underground burrow, a wood shaving substrate, and unlimited access to water that is routinely infused with a multivitamin and electrolyte supplement. The majority of the diet of the rats is provided during training sessions in the form of crushed commercial rodent chow pellets mixed with mashed bananas and avocados, which serves as appetitive reinforcement for the operant conditioning procedures. This diet is supplemented with a variety of fresh fruits, vegetables, and grains [3].

While APOPO began with training rats to detect landmines in former conflict zones, the demonstrated success influenced the 2001 idea to also train the rats to detect the presence of *Mycobacterium tuberculosis* in human sputum samples [4]. Data reported annually to the World Health Organization (WHO) by countries show that TB is one of the major causes of ill health and death worldwide. TB is a life-threatening infectious disease that attacks the lungs and can also harm other parts of the body. The transmission occurs from one person to another when a person with TB talks, sneezes, or coughs. The development of novel, accurate, robust, and rapid diagnostic capabilities will result in improved case detection, disease surveillance, health care delivery, and quality of future research [5]. In 2004, APOPO and Sokoine University of Agriculture (SUA) partnered with the Tanzanian National Institute of Medical Research (NIMR) and the Tanzanian National Tuberculosis and Leprosy Program (NTLP) to develop a scent-detection technology for diagnosing human TB in resource-poor areas [6]. While microscopy is the most commonly used method to detect TB in developing countries, its effectiveness remains a problem [3]. In Tanzania, the Ministry of Health, Community Development, Gender, Elders, and Children (MOHCDGEC) permitted APOPO to conduct research using rats to detect TB bacteria in sputum samples [7].

**Conceptual Framework**

The theoretical concepts and empirical framework of this study are based on Signal Detection Theory (SDT). SDT describes how features of the stimulus and detector factors affect performance on stimulus detection tasks [8]. SDT helps to distinguish between the sensitivity of a detector and the underlying signal. In medical diagnosis, this translates to the efficacy of a diagnostic tool to accurately detect the presence of a pathogen or other signal with medical significance [9], that is, the diagnostic “sensitivity.” However, in rats, determining diagnostic accuracy depends on the rat’s training and the diagnostic results from partner health clinics using WHO-endorsed methods. During training, the behavior of each...

---

**Figure 1.** Tuberculosis (TB) testing and detection using trained rats. The rats test and detect TB-negative and TB-positive samples.
rat is recorded, including indication responses committed in response to samples known to either contain or not contain TB (TB positive or TB negative). These data allow trainers to accurately track each rat’s discrimination learning [4]. There are numerous independent variables related to each rat evaluation session, including the rat’s identity (name), age, sex, and bodyweight, as well as the characteristics of the sample itself, including DOTS clinic diagnostic results (ID_BL_DOTS) and results of any applicable confirmatory diagnosis within APOPO’s laboratory (ID_BL_APOPO), which are combined to form another independent variable called TB_Status.

In this study, one of the primary dependent variables was captured as hit, which refers to whether or not (true or false) the rat provided an indication (continuously sniffed the sample for at least 3 seconds, as estimated by the rat handler). Combining the hit variable with WHO-endorsed diagnostic results (ID_BL_DOTS and ID_BL_APOPO) provided 4 possible outcomes termed rat performance for each sample evaluated (Figure 2), including correct hit, miss, false alarm, and correct reject, which are used in determining the diagnostic accuracy of each rat. Correct hit refers to samples that the rat indicated and were confirmed to contain TB; false alarm (or suspect) refers to samples that the rat indicated but which could not be confirmed to contain TB. Additionally, miss (sample confirmed to be TB positive) and correct reject (no TB mycobacterium confirmed) refer to samples that the rat failed to indicate (sniff for 3 seconds) [3]. In other words, the rat’s sensitivity represents the percentage of correct hits out of the sum of total correct hits and total misses (all confirmed TB-positive samples evaluated by the rat). Similarly, the rat’s specificity represents the percentage of correct rejects out of the sum of correct rejects and false alarms (all samples found to be TB negative) [10]. By this logic, sensitivity refers to the rat’s ability to accurately find true positive (TP) cases, while specificity measures its ability to accurately reject negative cases. Hence, sensitivity (correct hit) and specificity (correct reject) together comprise overall diagnostic accuracy.

![Figure 2. Relationship among the status of tuberculosis (TB), hit, and the performance of the rat. Hit refers to whether or not (true or false) the rat provided an indication.](https://ojphi.jmir.org/2024/1/e50771)

From Figure 2, if the TB status was already known to be positive at the time of the rat evaluation and hit was true, the rat’s behavior was categorized as “correct hit.” Conversely, if the TB status was positive and hit was false, the rat’s behavior was categorized as “miss.” On the other hand, if the TB status was determined to be negative at the time of the rat evaluation and hit was true, the rat’s behavior was categorized as “false alarm” or suspect. Finally, if the TB status was negative and hit was false, the rat’s behavior was categorized as “correct reject.”

Hence, contrary to the study by Jonathan et al [10], this study considered the status of TB in the sample the rat was evaluating. In that study, the modeling methods only used the dichotomous variable of hit as true or false (ie, did the rat sniff the sample for ≥3 seconds) without regard for what the rat was sniffing. Within the data set analyzed, about 78.8% of samples were not hit (hit=false), somewhat reflecting the estimated underlying prevalence of TB across the samples. However, assuming this distribution reflected that the most common outcome (hit=false) served as the desired or correct outcome in all instances when modeling rat performance, the models predicted when a trained rat would fail to detect TB (ie, miss a TB-positive sample or correctly reject a TB-negative sample) rather than detect it. Furthermore, the predictive power of the models did not take into account what the rats were smelling, since the rats were trained to perform differently (hit true or false) depending on the presence of TB within the sample. Therefore, the aim of this study was to replicate the procedures of the study by Jonathan et al [10] but with the inclusion of variables related to the detection of TB and with expansion of modeling to include 2 additional machine learning (ML) algorithms.

**Objectives of the Study**

This study applied the same data set from APOPO’s TB-detection rat training and research center in Morogoro, Tanzania, as used by Jonathan et al [10] but with the inclusion of WHO-endorsed diagnostic results, including those provided by partner DOTS clinics (smear microscopy, ID_BL_DOTS)
and, where applicable, those performed by APOPO (either concentrated smear microscopy, ID_BL_FM, or fluorescent microscopy, ID_BL_APOPO) to confirm samples flagged suspect by the rats. As with Jonathan et al [10], this study used the decision tree, random forest, and naïve Bayes algorithms and included support vector machine (SVM) and k-nearest neighbor (kNN) ML techniques to improve the accuracy of the predictive models. Furthermore, it provides extensive simulations using real data to determine if ML techniques can accurately predict the performance of rat TB detection. Additionally, this paper compares the classification accuracy performance of the 5 ML predictive models. The rest of this paper is organized as follows: the Related Work subsection provides details of related literature focusing on African giant pouched rat TB detection, including the current status and its implications, along with the application of ML in diagnosing and detecting TB; the Methods section presents the methodology of this study; the Results section provides a description of the performance results and performance measurements of the predictive models; and the Discussion section discusses the findings, provides conclusions, and mentions the scope for future work.

Related Work

**Diagnosis of TB by African Giant Pouched Rats: Current Status and its Implications**

African giant pouched rats cost-efficiently complement other TB diagnostic tools through second-line screening via scent detection to increase TB case detection. Patient samples are provided by partner DOTS clinics that perform initial screening. The rats can test up to 100 samples in 20 minutes or less, while a laboratory technician requires about 4 days to accomplish the same task using microscopy [11]. Samples that the clinic deems TB negative but which the rats indicate are TB positive are then retested using WHO-endorsed methods, such as concentrated smear microscopy or GeneXpert. Samples that are confirmed positive are communicated to the respective DOTS clinic, effectively providing 24-hour result turnaround and improved linkage to care [6]. Applying this method since 2007 has enabled TB-detection rats to identify more than 29,000 patients who had a missed diagnosis during initial screening [4]. Thus, rat scent detection technology is of great importance to the community and public health hospitals because it increases case detection, enables treatment, and curbs the spread of the disease [3].

**Application of ML and Big Data Analytics in Diagnosing and Detecting TB**

Technology advancement has allowed access to data from multidimensional sources with high throughput velocity. The term used to describe this kind of data is “big data,” which is difficult to analyze for interesting patterns or inefficiencies without ML technologies [12]. The application of ML in health care is important to improve human health, and ML and big data analytic technologies have brought advancements in TB health care services owing to the increase of health care data and the availability of analytics to solve health problems [13]. ML is a technology that enables a machine to learn from past data and predict the outcome. Thus, in health care, ML contains sophisticated algorithms that help to learn features from a large volume of health care data and then use the obtained insights to assist clinical practices [14]. Big data analytics is the use of advanced analytic techniques on vast amounts of data in different formats, such as structured, semistructured, and unstructured data, from different sources. Big data analytics can help to discover useful information that facilitates decision-making and health care outcome prediction. Therefore, ML and big data analytics can assist physicians by providing up-to-date medical information from clinical practices for proper patient care. As such, the application of ML and big data analytics can help to reduce diagnostic and human errors in the outcomes of clinical practices [15].

ML in health care depends on different techniques, which include classification, clustering, and association, for its operation. These techniques help to learn past data and detect knowledge patterns [16]. Classification techniques are used to develop models that predict future events from the manipulated data and offer solutions to real-world health problems such as diagnosis and treatment of diseases [16]. Classification is the ML technique that operates by building predictive models that categorize and assign labels to manipulated and newly encountered instances [16]. These predictive models help solve multiclassification problems through prediction and analysis. Moreover, the models are used as decision-support tools that help medical professionals interpret diagnosis results [17]. For example, Abdar et al [18] used the boosted C5.0 and CHAID classification algorithms to build a decision tree model for the early diagnosis and prediction of liver disease. In addition, ML technologies were used in the diagnosis of TB to categorize and find relationships among the manipulated variables [19]. This study developed an efficient and reliable framework for automatic TB bacilli detection based on deep learning and ML algorithms. The study also suggested that a classification model can be used to discriminate between positive and negative samples [19].

The classification algorithms recently used in the diagnosis of TB include decision tree, random forest, naïve Bayes, SVM, and kNN [20]. These algorithms are suggested as an alternative for health care professionals to improve the diagnosis of TB. The decision tree algorithm C4.5 was used to build a model to predict the presence of TB bacteria. The results showed that the decision tree had a prediction accuracy of 99% [21]. The decision tree generates rules that are simple and easy to understand and interpret for a decision maker [16].

Moreover, a random forest classification algorithm was used to discriminate the TB bacilli with a sensitivity and specificity of above 89.34% and 62.89%, respectively. Furthermore, it is proposed that the naïve Bayes algorithm can be used for the diagnosis of TB [22]. Additionally, SVM is known as a useful model to identify abnormalities in the lungs for the diagnosis of TB [23]. Following this, algorithm comparison is of great importance to find a reliable algorithm in the given data [24].
Methods

ML Algorithms
In this study, the ML algorithms used are decision tree, random forest, naïve Bayes, SVM, and kNN to build predictive models that categorize data and assign a label to manipulated and newly encountered data. The purpose of involving different algorithms is to compare and improve the prediction accuracy of rats for TB detection.

Real Data Sets
This paper used 2 data sets provided by APOPO: detection rats data set and RAT_WEIGHT data set, which were combined to form the final data set, as shown in Table 1. The detection rats data set contained 471,133 observations from 2011 to 2019 and involved 18 variables (17 independent and 1 dependent). The RAT_WEIGHT data set contained 1438 records collected from 2012 to 2019 and involved 4 independent variables. Moreover, these data contained 5 female rats with IDs 56, 72, 80, 85, and 96. However, the fifth rat with ID 96 from the RAT_WEIGHT data set was eliminated in the analysis because it lacked the necessary detection performance variables in the detection rats data set. Therefore, 4 female rats were used in this study. The 2 data sets and corresponding variables are displayed in Table 1.

Table 1. Rats data set description.

<table>
<thead>
<tr>
<th>Data set and number</th>
<th>Variable name</th>
<th>Data type</th>
<th>Description</th>
<th>Variable type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection rats data set</td>
<td>DOTS_NAME</td>
<td>String</td>
<td>Name of the DOTS&lt;sup&gt;a&lt;/sup&gt; center</td>
<td>Independent</td>
</tr>
<tr>
<td>1</td>
<td>DOTS_PATIENTS_NUMBER</td>
<td>Integer</td>
<td>Number of patients from the DOTS center</td>
<td>Independent</td>
</tr>
<tr>
<td>2</td>
<td>ENTRY_YEAR</td>
<td>Integer</td>
<td>Year when the patient attended the DOTS center</td>
<td>Independent</td>
</tr>
<tr>
<td>3</td>
<td>ID_SAMPLE</td>
<td>Integer</td>
<td>Identification of the sample</td>
<td>Independent</td>
</tr>
<tr>
<td>4</td>
<td>ID_BL_DOTS</td>
<td>Integer</td>
<td>Identification of the bacteria level from the DOTS center</td>
<td>Independent</td>
</tr>
<tr>
<td>5</td>
<td>HIT</td>
<td>Boolean</td>
<td>TB&lt;sup&gt;b&lt;/sup&gt; detection rat performance (categorical variable)</td>
<td>Dependent</td>
</tr>
<tr>
<td>6</td>
<td>ID_BL_APOPO</td>
<td>Integer</td>
<td>Identification of the bacteria level from the APOPO&lt;sup&gt;c&lt;/sup&gt; center</td>
<td>Independent</td>
</tr>
<tr>
<td>7</td>
<td>ID_CONFIGURATION</td>
<td>Integer</td>
<td>Identification of the cage during training</td>
<td>Independent</td>
</tr>
<tr>
<td>8</td>
<td>ID_BL_FM</td>
<td>Integer</td>
<td>Identification of the bacteria level by fluorescence microscopy</td>
<td>Independent</td>
</tr>
<tr>
<td>9</td>
<td>ID_EVALUATION_SESSION</td>
<td>Integer</td>
<td>Identification of the evaluation session</td>
<td>Independent</td>
</tr>
<tr>
<td>10</td>
<td>SESSION_DATE</td>
<td>Date</td>
<td>Date when a session was performed</td>
<td>Independent</td>
</tr>
<tr>
<td>11</td>
<td>ID_RAT</td>
<td>Integer</td>
<td>Identification of the rat</td>
<td>Independent</td>
</tr>
<tr>
<td>12</td>
<td>RAT_NAME</td>
<td>String</td>
<td>Name of the rat</td>
<td>Independent</td>
</tr>
<tr>
<td>13</td>
<td>GENDER</td>
<td>String</td>
<td>Sex of the rat</td>
<td>Independent</td>
</tr>
<tr>
<td>14</td>
<td>AGE</td>
<td>Integer</td>
<td>Age of the rat</td>
<td>Independent</td>
</tr>
<tr>
<td>15</td>
<td>START_TIME</td>
<td>Date/Time</td>
<td>Date and time when the detection task started</td>
<td>Independent</td>
</tr>
<tr>
<td>16</td>
<td>END_TIME</td>
<td>Date/Time</td>
<td>Date and time when the detection task ended</td>
<td>Independent</td>
</tr>
<tr>
<td>17</td>
<td>DOB</td>
<td>Date</td>
<td>Date when the rat was born</td>
<td>Independent</td>
</tr>
<tr>
<td>RAT_WEIGHT data set</td>
<td>ID_RAT</td>
<td>Integer</td>
<td>Identification of the rat</td>
<td>Independent</td>
</tr>
<tr>
<td>1</td>
<td>RAT_NAME</td>
<td>String</td>
<td>Name of the rat</td>
<td>Independent</td>
</tr>
<tr>
<td>2</td>
<td>WEIGHT_DATE</td>
<td>Date</td>
<td>Date when the weight of the rat was measured</td>
<td>Independent</td>
</tr>
<tr>
<td>3</td>
<td>WEIGHT</td>
<td>Integer</td>
<td>Weight of the rat</td>
<td>Independent</td>
</tr>
</tbody>
</table>

<sup>a</sup>DOTS: directly-observed treatment, short-course.

<sup>b</sup>TB: tuberculosis.

<sup>c</sup>APPO: Anti-Persoonsmijnen Ontmijnende Product Ontwikkeling.
Applied Variables

The data underwent initial preprocessing to obtain the required variables for developing the predictive models. All data preparation was implemented by Python owing to its large number of libraries for scientific computing and the development of ML predictive models [24]. The sample (either TB negative or the bacterial concentration of TB positivity provided by the partner DOTS clinic, ID_BL_DOTS) was compared to APOPO’s confirmatory diagnosis (where applicable) using concentrated smear microscopy (ID_BL_APOPO) to create a variable termed Definitive_Status. This variable reflected the APOPO result when one was provided; otherwise, it indicated the DOTS clinic result. The Definitive_Status was then transformed into the dichotomous variables of TB_Status to reflect the final status of the sample as either positive or negative for TB (collapsing across bacterial concentrations for positive samples). Then, TB_Status was compared to hit to compute the dependent variable of Rat_Performance, which consists of 4 categories: correct hit, miss, false alarm, and correct reject (Figure 2).

After the data preparation, 4 variables for the detection performance of the rats, including TB_Status, age, weight, and hit, as shown in Table 2, were used to build the predictive model. Moreover, this study used 366,441 observations for analysis after removing the null rows from the rats data set to prevent noises, outliers, and inconsistencies in the data. The sklearn model selection library through a train-test split class was used to partition the data (366,441 observations) into 256,508 observations (70%) in the training data and 109,933 observations (30%) in the test data. It is important to mention that, due to the binary nature of many variables and the underlying prevalence of TB infections, the data used in this study lack a normal distribution, as shown in Multimedia Appendix 1.

Categorical variables were used to build predictive models, and 256,508 observations (70%) were used for training the models. The TB_Status variable consisted of 10.90% (27,950/256,508) positive samples and 89.10% (228,558/256,508) negative samples. The hit variable consisted of 21.33% (54,719/256,508) true values and 78.67% (201,789/256,508) false values.

Table 3 shows a statistical summary of the distribution of continuous variable data before and after the random data split. Despite most of the distributions being the same, the mean of age and weight variables showed a difference of 0.01. Moreover, the SD of ID_RAT and weight differed by 0.01.

### Table 2. Description of the dependent and independent variables used to build predictive models.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Data type</th>
<th>Variable type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TB_Status</td>
<td>Final diagnosis of the sample as either TB(^a) positive or TB negative. Combines the diagnostic results of both DOTS(^b) and APOPO(^c) (lab confirmation, when applicable) wherein APOPO status (results) overrides DOTS.</td>
<td>Object</td>
<td>Independent/cat-egorical</td>
<td>True or false</td>
</tr>
<tr>
<td>Age</td>
<td>Age of the rat in years at the time when the rat evaluated the patient sample in question</td>
<td>Object</td>
<td>Independent</td>
<td>Age ranges from 0.79 to 7.95 years</td>
</tr>
<tr>
<td>Weight</td>
<td>Average rat body weight (in grams) per year because most of DetectionRats-Data describes the daily detection tasks and misses their corresponding weights since the weight of the rats from the RAT_WEIGHT data set was measured every week.</td>
<td>Object</td>
<td>Independent</td>
<td>Average rat body weight ranges from 843.67 to 1054.83 grams</td>
</tr>
<tr>
<td>Hit(^d)</td>
<td>Defined as a continuous sniff (nose insertion into the cage hole) for ≥3 seconds. True means the rat “indicated” that the sample contained TB (held its nose in the hole for at least 3 seconds). False means the rat rejected the sample (did not hold its nose for at least 3 seconds).</td>
<td>Object</td>
<td>Dependent/cat-egorical</td>
<td>True or false</td>
</tr>
</tbody>
</table>

\(^a\)TB: tuberculosis.
\(^b\)DOTS: directly-observed treatment, short-course.
\(^c\)APOPO: Anti-Persoonsmijnen Ontmijnende Product Ontwikkeling.
\(^d\)Hit refers to whether or not (true or false) the rat provided an indication.
Table 3. Descriptive statistics of the continuous variables used to build predictive models before and after random data split.

<table>
<thead>
<tr>
<th>Data split status and variable</th>
<th>Age (years)</th>
<th>Weight (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before random data split (n=366,441, 100%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.83</td>
<td>899.40</td>
</tr>
<tr>
<td>SD</td>
<td>1.72</td>
<td>84.37</td>
</tr>
<tr>
<td>IQR</td>
<td>3.71</td>
<td>866.80</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.79</td>
<td>843.67</td>
</tr>
<tr>
<td>Maximum</td>
<td>7.95</td>
<td>1054.83</td>
</tr>
<tr>
<td><strong>After random data split (n=256,508, 70%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.84</td>
<td>899.41</td>
</tr>
<tr>
<td>SD</td>
<td>1.72</td>
<td>84.36</td>
</tr>
<tr>
<td>IQR</td>
<td>3.71</td>
<td>866.80</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.79</td>
<td>843.67</td>
</tr>
<tr>
<td>Maximum</td>
<td>7.95</td>
<td>1054.83</td>
</tr>
</tbody>
</table>

**Model Building**

The predictive model in this study was developed using 5 different ML techniques: decision tree, random forest, naïve Bayes, SVM, and kNN. This study used Python libraries for data preprocessing, matrix processing, mathematical functions, visualization, and classification. These are Pandas, Numpy, Matplotlib, and Scikit-learn [25]. The repetitive approach was used to generate a decision tree by dividing the training data. The data were divided recursively until the same class of variables, depending on conditions, using roughly 15,000 samples per leaf, were distributed among each division to create the decision tree. After that, each node in the decision tree used a split point to test the altered variables and choose how to divide the data. The split decision was concerned with the information gain and entropy of a computed variable. The variable that had the greatest information gain and the least entropy was therefore divided and put to the test. The choice regarding the data split and decision tree building was made based on information gain and entropy [16]. This study used pruning to maintain control over the parameters being used to remedy expansion.

During the training procedure, many decision trees were randomly constructed using the random forest technique. Based on the provided manipulated variables, the algorithm’s ultimate decision was based on the selection of the majority of the trees. There was a connection between the outcome and the number of trees in the forest. The outcome was therefore more accurate with an increase in the number of trees. As a result, the technique handled 500 trees in the ensemble, and it calculated the error rate using the training set of information. In the random forest approach, the training data were used to generate random splits for the root node and variable node. Since there was no parameter control during training, the connection between trees remained strong. Additionally, the frequency and values of the adjusted variables from the provided data were counted to generate the classification model using the naïve Bayes method. This method determined the dependent variable’s a priori probabilities as well as the conditional probabilities for each independent variable based on the altered data. The naïve Bayes technique has been specifically utilized to contrast its prediction performance with the outcomes produced by other ML techniques. It does not display the weights of each variable included in the classification.

SVM is one of the most common supervised ML algorithms owing to its greater predictive power. SVM analyzes data, recognizes patterns, and produces input-output functions from a set of labeled training data. It works by classifying a response variable by drawing a decision boundary line or hyperplane to separate 2 classes. Then, the maximum margin hyperplanes are constructed to optimally separate the output classes from each other in the training data. The goal is to find the optimal separating hyperplane where the separating margin is maximized. The linear kernel was used to allow flexibility and loss functions. The kNN algorithm is a supervised ML algorithm that works by identifying a set of k-nearest observations to the test point and calculating mainly the Euclidean distance between an observation and its kNN in training data. The k in kNN refers to the number of nearest neighbors the classifier will retrieve and use to make its prediction. The chosen k in kNN was 1, as it is suggested to provide the best test prediction.

**Performance Measurements**

This study used accuracy, specificity, sensitivity, and F1 score as metrics to evaluate the performance of the generated predictive models and compare classification performances. These measurements were supported in the scikit-learn library through the classification report class.

**Accuracy**

The classification accuracy was calculated based on the confusion matrix, which accurately categorized the actual class labels of the test data and the class labels of the predicted models. It was also obtained by dividing the number of truly classified instances by the number of instances in the test phase. Accuracy considers TP, true negative (TN), false positive (FP), and false negative (FN). The classification accuracy for the data set was measured according to the following formula:
Accuracy = \( \frac{TP + TN}{TP + FP + TN + FN} \) \hspace{1cm} (1)

**Sensitivity**

Sensitivity is defined as the number of TP cases over the number of TP cases plus the number of FN cases. Sensitivity identifies the correct positive predictions relative to the total actual positive cases. It is sometimes called a recall metric. The formula of sensitivity is as follows:

Sensitivity = \( \frac{TP}{TP + FN} \) \hspace{1cm} (2)

**Specificity**

Specificity is the ratio between TN cases and all negative cases. In this study, the precision measure identified the correct positive predictions relative to total positive predictions. For diagnostic tools, this could be termed positive predictive value (PPV) or precision. It essentially provides confidence that any given positive response reflects a truly positive condition [25]. The formula of specificity is as follows:

Specificity = \( \frac{TN}{TN + FP} \) \hspace{1cm} (3)

**F1 Score**

The F1 score is the harmonic mean of specificity and sensitivity. Basically, it is the weighted average of specificity and sensitivity. The F1 score was calculated from the specificity and sensitivity of the test data set [25]. The formula of the F1 score is as follows:

F1 score = \( \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \) \hspace{1cm} (4)

It is important to mention that specificity and sensitivity are similar to precision and recall, respectively.

**Restrictions of the Study**

This study ran the predictive models on a computer with a Core i5-5300U CPU at 2.30 GHz (2301 MHz, 2 cores, 4 logical processors) and 8 GB of RAM. The sample size, on the other hand, was small, with only 4 rats and a gender imbalance. Moreover, the hit variable consisted of fewer true values (21.26%) than false values (78.74%).

**Ethical Considerations**

The study was approved by the SUA (DPRTC/R/142/vol.01/104) and Medical Research Coordinating Committee of Tanzania (NIMR/HQ/R.8a/Vol.1X/3905). The use of African giant pouched rats as a potential tool for TB diagnosis has received ethics clearance from the Tanzanian Medical Research Coordinating Committee [26]. The Office of Laboratory Animal Welfare has approved APOPO’s Animal Welfare Assurance (OLAW; Assurance Identification Number A5720-01).

**Results**

Comparing Classification Performance Measurements of the Predictive Models

This study used different ML techniques to build the predictive models following the methodology presented in Figure 3. Moreover, this study employed several metrics, including accuracy, sensitivity, specificity, and F1 score, to measure the classification performance of the predictive models based on test data. Figure 4 shows the confusion matrices of the SVM and random forest classifiers, while Table 4 summarizes the performance of all 5 ML techniques used to build the predictive models. The accuracy classification performance of the kNN technique was low at about 81.25%, while the best performing algorithm was SVM. As it can be seen from Table 4, validation showed that the SVM classifier based on the 4 variables shown in Table 2 achieved an accuracy of 83.39%, but it also reported that SVM had better ability to recognize the status of TB as either positive or negative in a given sample.

**Figure 3.** Process flow of machine learning–based prediction models of rat tuberculosis detection performance. The rectangle symbols represent data, while the histogram entails model evaluation metrics. DT: decision tree; kNN: k-nearest neighbor; NB: naive Bayes; RF: random forest; SVM: support vector machine.
Figure 4. Confusion matrices of the predictive models. (A) Support vector machine classifier; (B) Random forest classifier.
Table 4. Comparing the classification performance of classifiers of rat tuberculosis detection.

<table>
<thead>
<tr>
<th>Classification performance measurement</th>
<th>Predictive model</th>
<th>Decision tree</th>
<th>Random forest</th>
<th>Naïve Bayes</th>
<th>Support vector machine</th>
<th>K-nearest neighbor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy, %</td>
<td></td>
<td>83.32</td>
<td>83.38</td>
<td>82.56</td>
<td>83.39</td>
<td>81.25</td>
</tr>
<tr>
<td>Sensitivity, %</td>
<td></td>
<td>65.00</td>
<td>65.00</td>
<td>63.00</td>
<td>66.00</td>
<td>64.05</td>
</tr>
<tr>
<td>Specificity, %</td>
<td></td>
<td>79.00</td>
<td>79.00</td>
<td>77.00</td>
<td>78.00</td>
<td>72.05</td>
</tr>
<tr>
<td>F1 score, %</td>
<td></td>
<td>67.00</td>
<td>67.00</td>
<td>66.00</td>
<td>69.00</td>
<td>66.05</td>
</tr>
<tr>
<td>Correctly classified observations (true positive), n</td>
<td></td>
<td>91,602</td>
<td>91,602</td>
<td>90,370</td>
<td>91,602</td>
<td>89,326</td>
</tr>
<tr>
<td>Incorrectly classified observations (false negative), n</td>
<td></td>
<td>18,331</td>
<td>18,331</td>
<td>19,163</td>
<td>18,331</td>
<td>20,607</td>
</tr>
</tbody>
</table>

Important Variables Influencing the TB Detection Performance of the Rats

This study used the random forest variable importance function to output the predictor variables based on the mean decrease in Gini (impurity). Random forest showed high performance in the feature ranking. The mean decrease in the Gini value is the average (mean) of a variable’s total decrease in the likelihood of incorrect classification of a new instance of a random variable from the data set. Multimedia Appendix 2 shows the predicted variable importance based on the mean decrease in the Gini value using the random forest algorithm.

Table 5. Random forest variable importance based on the mean decrease in the Gini value.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable name</th>
<th>Mean decrease in the Gini value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>TB_Status</td>
<td>0.817152</td>
</tr>
<tr>
<td>1</td>
<td>Age</td>
<td>0.156190</td>
</tr>
<tr>
<td>2</td>
<td>Weight</td>
<td>0.026657</td>
</tr>
</tbody>
</table>

Algorithm for the Prediction of Rat TB Detection Performance

The study also employed a prediction algorithm for TB detection as illustrated in Textbox 1.

Textbox 1 shows the algorithm that predicts if the rat will hit the sample or not. First, data were imported and normalized to acquire the required data format. Then, the statistical summary of the independent variables used to build predictive models was described. Considering Figure 3, the train_test_split library was used to divide the data set into training data (70%) for developing the models and test data (30%) for validating the models. The predictive models were trained based on the decision tree, random forest, naïve Bayes, SVM, and kNN classifiers, using the train data. Meanwhile, the validation of the models was performed using the test data. Then, accuracy, sensitivity, specificity, and F1 score were used to measure the classification performance of each classifier, as reported in Table 4. Furthermore, the input variables TB_Status, age, and weight were entered for prediction. Following the prediction, models were validated using the test data. Hence, data visualization was performed using the Matplotlib library for proper interpretation of the results. On the other hand, if the constraints were not met, the algorithm could be terminated.

In addition to the above algorithm for the prediction of rat TB detection performance, Figure 3 indicates the process flow of ML models and their predictions using Python libraries. The TB input data set was imported as a .csv file. After preprocessing the data, the sklearn model selection library was used to partition the data into training data (70%) and test data (30%) by using a simple random split method. The training data were used to build a predictive model using decision tree, random forest, naïve Bayes, SVM, and kNN classifiers. After building the predictive model, the inputs, including TB_Status, age, and weight, were computed to predict if the rat would hit the sample or not. Thereafter, the predictive models were evaluated for their prediction performance using accuracy, sensitivity, specificity, and recall metrics.
**Textbox 1. Algorithm for the prediction of rat tuberculosis detection performance.**

I. Import and normalize the dataset (.csv)

II. Calculate IQR, mean, SD, minimum, and maximum

III. Perform splitting of the data set

1. if splitting is successful and not any constraints then
   - train the model

2. Perform machine learning (ML) modeling based on decision tree, random forest, naïve Bayes, support vector machine, and k-nearest neighbor

3. Perform validation of the ML modeling

4. Perform ML model prediction

5. Validate the prediction model by calculating accuracy, sensitivity, specificity, and F1 score

6. if accuracy and other parameters are good then
   - input: TB_Status, age, weight

7. Perform ML model prediction if the rat would hit the sample or not

8. Update the predicted value of new data for reporting

9. Make data visualization in Python

10. else

11. Perform termination check

12. else

13. End

---

**Discussion**

**Principal Findings**

The aim of this study was to build on the prior work of Jonathan et al [10] to develop models that predict if a trained TB-detection rat would hit (indicate the presence of TB within) a patient sample or not using ML techniques by incorporating variables related to the diagnostic results of the TB samples. This study used decision tree, random forest, naïve Bayes, SVM, and kNN ML techniques to build predictive models. The ML techniques successfully categorized the data by assigning a label to each computed data point. The results revealed that for the 5 different algorithms used, the classification accuracy was the greatest for SVM, suggesting its superiority to the decision tree, random forest, naïve Bayes, and kNN classifiers. The SVM classifier outperformed by yielding a classification accuracy of about 83.39% for predicting if the rat would hit the sample or not. This level of accuracy surpasses the 78.82% accuracy found with decision tree and naïve Bayes by Jonathan et al [10], suggesting that the inclusion of sample information serves as a valuable variable that influences the performance of TB-detection rats and improves the accuracy of the prediction models. Moreover, Jonathan et al [10] employed a small amount of data compared to the data used in this study. In fact, TB_Status was found to be the most significant variable in predicting rat TB detection performance. However, there was an insignificant accuracy difference between the constructed models and those created by Jonathan et al [10], which could be due to the characteristics of the data [16]. Therefore, the additional variables are likely to influence rat behavior, and the true status of patient samples can only be determined by available diagnostics.

**Conclusion**

This study has shown the usefulness of ML techniques to identify factors that influence TB detection performance of rats. The techniques used were decision tree, random forest, naïve Bayes, SVM, and kNN to develop models that predict if the rat would hit the sample or not by incorporating valuable variables related to TB detection performance of rats. The performance of the predictive models was measured by accuracy, sensitivity, specificity, and F1 score metrics. The results showed that the SVM predictive model outperformed in the classification and prediction of the performance of rats in TB detection by yielding the highest accuracy of 83.39%. Furthermore, the obtained results suggest that the inclusion of variables related to the diagnostic results of TB samples improves the performance of the predictive models. Therefore, the results might benefit TB-detection rat trainers and TB decision-makers in improving the diagnostic accuracy of rats by predicting if a trained TB-detection rat would hit a patient sample or not. They can adopt several measures, including ensuring that all hit samples are confirmed within APOPO’s laboratory (ID_BL_APOPO). Furthermore, taking into consideration that the age of the rat at hit and clinic diagnostic results are predictors of detection performance.
Acknowledgments

We thank all reviewers who provided their insightful comments and suggestions for the improvement of the study and thank the APOPO (Anti-Persoonsmijnen Ontmijnende Product Ontwikkeling) TB Training and Research Center in Morogoro that provided data for this study. This work was supported (grant number CC003) in part by the Government of Tanzania through the Research and Innovation Grants of the Sokoine University of Agriculture (SUA). All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Distribution of continuous independent variables.

Multimedia Appendix 2

Random forest variable importance plot.

References


Abbreviations

APPOPO: Anti-Persoonsmijnen Ontmijnende Product Ontwikkeling

DOTS: directly-observed treatment, short-course

FN: false negative

FP: false positive

kNN: k-nearest neighbor

ML: machine learning

SDT: Signal Detection Theory

SUA: Sokoine University of Agriculture

SVM: support vector machine

TB: tuberculosis

TN: true negative

TP: true positive

WHO: World Health Organization
Geospatial Imprecision With Constraints for Precision Public Health: Algorithm Development and Validation

Daniel Harris¹, PhD; Chris Delcher², PhD

¹Institute for Biomedical Informatics, University of Kentucky, Lexington, KY, United States
²Department of Pharmacy Practice and Science, College of Pharmacy, University of Kentucky, Lexington, KY, United States

Corresponding Author:
Daniel Harris, PhD
Institute for Biomedical Informatics
University of Kentucky
Healthy Kentucky Research Building, Room 361
760 Press Avenue
Lexington, KY, 40508
United States
Phone: 1 8593237100
Email: daniel.harris@uky.edu

Abstract

Background: Location and environmental social determinants of health are increasingly important factors in both an individual’s health and the monitoring of community-level public health issues.

Objective: We aimed to measure the extent to which location obfuscation techniques, designed to protect an individual’s privacy, can unintentionally shift geographical coordinates into neighborhoods with significantly different socioeconomic demographics, which limits the precision of findings for public health stakeholders.

Methods: Point obfuscation techniques intentionally blur geographic coordinates to conceal the original location. The pinwheel obfuscation method is an existing technique in which a point is moved along a pinwheel-like path given a randomly chosen angle and a maximum radius; we evaluate the impact of this technique using 2 data sets by comparing the demographics of the original point and the resulting shifted point by cross-referencing data from the United States Census Bureau.

Results: Using poverty measures showed that points from regions of low poverty may be shifted to regions of high poverty; similarly, points in regions with high poverty may be shifted into regions of low poverty. We varied the maximum allowable obfuscation radius; the mean difference in poverty rate before and after obfuscation ranged from 6.5% to 11.7%. Additionally, obfuscation inadvertently caused false hot spots for deaths by suicide in Cook County, Illinois.

Conclusions: Privacy concerns require patient locations to be imprecise to protect against risk of identification; precision public health requires accuracy. We propose a modified obfuscation technique that is constrained to generate a new point within a specified census-designated region to preserve both privacy and analytical accuracy by avoiding demographic shifts.

Introduction

Geographic information systems (GISs) are increasingly important for public health research and policy makers and are instrumental in measuring socioeconomic equity in health care [1,2]. Social determinants of health are the conditions in which individuals are born, live, work, and age; mesolevel determinants are from the physical environment and encompass items such as geographic location and access to resources [3]. Location-based exposures tied to geographic location are a pivotal element to one’s health [4-6]; ongoing research suggests that zip code is on par with genetic code in influencing individual health [7-10]. Even greater research utility lies in more precisely geolocating patients beyond the zip code level, yet privacy regulations often prevent high-resolution patient residential address data from being shared for research purposes.
Privacy is paramount when working with health care data and access is regulated at different levels through both institutional policies and government-mandated legal protections [12]. The Health Insurance Portability and Accountability Act (HIPAA) mandates privacy protections of personal health information in the United States; it outlines which data elements are considered private, including patient addresses needed for geospatial analysis for place-based epidemiology.

Although institutional review boards may grant researchers and other agencies access to identified data that pose minimal risk to the patient, there may exist institutional hesitancy to disclose this data due to the inherent privacy and sensitivity of residential addresses. As a current example, this tension is apparent between state and federal public health and safety agencies using the Overdose Detection Mapping Application Program (ODMAP), which maps in real time the exact locations of suspected drug overdoses, often occurring in residential locations [13,14]. Geocoding a patient’s address (ie, converting to geographic coordinates) is often an intermediate step in secondary data analyses; it is either used to link the patient to external geographic units (eg, census tracts to obtain neighborhood socioeconomic status from the United States Census Bureau) or to calculate distance from other entities, such as health care providers and facilities. For example, accessibility of buprenorphine, a medication for opioid use disorder, may be determined using addresses of health care providers that are authorized to prescribe the medication [15]. In these examples, imprecise locations may be sufficient for confident linkage to administrative units or approximate distance measures and preferred for research to preserve privacy. In these scenarios, thoughtful and controlled techniques designed to generate inexact data are needed to reduce precision [12,16].

To this end, geospatial or location-based privacy methods seek to maintain an appropriate level of confidentiality for a given task, service, or application while balancing the utility that these offer [17-19]. For example, users of location-based services on a cellular phone expect some level of privacy when sending personal data, and different strategies exist that anonymize pools of people by anonymizing data at point of collection. Location-based $k$-anonymity provides a method where one’s data and location are indistinguishable from $k$–1 other people [20]. Other methods, such as geographic masking, alter coordinates systematically to limit the risk of reidentification when releasing data [21]; no universally accepted method exists for protecting geospatial privacy [16].

Point obfuscation refers to the deliberate degrading of the resolution of coordinate information with the goal of protecting the privacy of the individual represented by the point [22]; this may be referred to as geographic masking, geomasking, jittering, or dithering and relies upon transformations or perturbations using randomness or artificial noise [22]. The $N$-RAND algorithm generates $N$ candidate points in a given area and selects the furthest point [23]; $\theta$-RAND limits candidate points to a specific area defined by a chosen angle [24]. We introduce modifications to the pinwheel obfuscation method which shifts points along pinwheel-like paths for a randomly chosen angle [25]; examples are shown in Figure 1. The noise added by this method is asymmetrical and highly variable, making it less open to privacy attacks designed to eliminate uniform and predictable noise [25]. However, the limitation of any point obfuscation technique is that coordinate shifts may change real-world locations and distort the linked health-related metrics. For example, a study defining a participant’s rurality based on administrative units may be impacted if obfuscated coordinates move the participant across boundaries into an urban area.

Figure 1. Pinwheel obfuscation at theta 45 degrees (A) and 15 degrees (B).

Address correction positively impacts the accuracy of assigning a patient to a geographic area [26]. On the other hand, the goal of point obfuscation is to intentionally generate an incorrect address to preserve patient privacy without compromising analytical conclusions. Our paper demonstrates that indiscriminate point obfuscation impacts studies linking points to neighborhood-level socioeconomic demographics and subsequently provides new methods needed to constrain...
pinwheel obfuscation to yield results confined in specific census-designated regions, such as blocks, block groups, or tracts. The constraint reduces concerns that neighborhood-level measures are inappropriately assigned at the patient level, leading to misclassification bias. We use poverty status as an example of data recorded by the United States Census Bureau to explore the potential impact of unintentional administrative boundary shift. Also, we demonstrate how indiscriminate point obfuscation impacts hot spot analysis at the census tract level. The role of this work is to provide evidence that point obfuscation techniques may substantially alter neighborhood-level socioeconomic demographics and that the intentional imprecision in these techniques must be constrained to support precision public health.

Methods

Overview

We implemented our methods using PostGIS, an open-source project that adds geospatial objects and procedures to the PostgreSQL database. We previously demonstrated PostGIS as a capable environment for geospatial privacy research [27,28]. We make our custom PostGIS functions available as open-source software [29]. The pinwheel technique was originally designed because other point obfuscation methods could be reversed by methods designed to filter uniform noise; the randomness of the pinwheel has been shown to maintain high variability, making it less susceptible to privacy attack [25]. Our geographically constrained pinwheel algorithm leverages the same concept as the original pinwheel algorithm and improves its research utility by adding constraint checking that controls how the new obfuscated point is selected. There is a function, PINWHEEL, that obfuscates a single point given a specific theta, maximum radius r, and a calculated random degree a; these are used to calculate a projection distance that can leverage PostGIS’s projection function, ST_PROJECT, to calculate the resulting obfuscated point.

The left-hand side of Figure 1 shows 1000 candidate points for a given seed point using a 45-degree angle; we sequentially varied the maximum radius while keeping theta constant at 45 degrees. Similarly, the right-hand side of Figure 1 shows the same simulation but with 15-degree angles; this demonstrates that theta controls the width of the pinwheel layers and that small angles naturally yield closer layers. In practice, a random degree may be used to further obfuscate the results.

Census Bureau geometries are input as reference data. The smallest census-based geographic boundary, the census block, is contained in the block group; a census block group typically represents between 600 and 3000 people. Block groups are organized into census tracts, which typically contain between 1200 and 8000 people and have an optimum size of 4000 people [30]. The United States Census Bureau publishes and updates files containing the geometries for these regions; these geometries are used in point intersection calculations to assign a region to a given point.

To obfuscate, points from the original list, P1, are fed into the PINWHEEL function and saved into P2. To constrain obfuscation, our method recalculates the pinwheel candidate for any generated candidates falling outside a specific region associated with the original point; we can constrain to standard administrative units: state, county, tract, block group, and block geometry. Furthermore, we can constrain obfuscation to custom geometries, such as buffer zones or other areas that may be relevant for research projects. The region of the points can be calculated with PostGIS’s ST_CONTAINS function, which tests the intersection of the points with the Census Bureau geometries. The regions are compared for every matching pair of existing and new points in P1 and P2. If the regions are dissimilar, PINWHEEL is rerun on the existing point. This continues until all points in P1 have a matching obfuscated point in P2 where P1 regions align with P2 regions.

We tested our methods with 2 different data sets, with geographic coverage ranging from multiple states to a single large urban area. Our first data set contained 1,000,000 records formatted in the Observational Medical Outcomes Partnership (OMOP) common data model [31]; we previously leveraged this data for geospatial research on open data and privacy [27]. Our second data set contains 58,102 case records from the Medical Examiner Case Archive from Cook County, Illinois, which contains the city of Chicago; we previously used this open data for research on geographic clustering of fatal overdoses [32] and to create an open data pipeline for spatial analyses on substance use disorders [33]. This open data set was released by the Cook County Medical Examiner’s Office (CCMEO) and offers details on all deaths recorded by the CCMEO from August 2014 to April 2022, including the address where the incident occurred and the address of the death. Deaths by suicide recorded in the CCMEO data set were also used for our hot spot analysis example.

Ethical Considerations

This study was exempt from ethical review since no private health data were used and no human subjects were involved.

Results

As expected, the pinwheel obfuscation method initially resulted in points shifting into different geographic regions; the frequency of these region shifts is summarized in Table 1 for our OMOP and CCMEO data. We categorized these shifts by census-designated regions by increasing population size (block, block group, tract, county, and state). Shifts were proportional to the maximum radius allowed; a radius of 1000 meters was the largest distance tested and naturally generated the most region shifts. Blocks are the smallest geographic unit and experienced the most change: 747,934 of 1,000,000 (74.8%) of the OMOP points and 53,415 of 58,102 (91.9%) of the CCMEO points were moved to different census blocks after obfuscation up to 1000 meters away. This empirically indicates that smaller geographical regions are more likely to shift when using any significant distance in the obfuscation method. Blocks are the smallest of the census-designated administrative boundaries. There is no maximum size for a census block; minimum block size is 30,000 square feet (2787.1 m²) for polygons bounded by roads or 40,000 square feet (3716.1 m²) otherwise, which is
smaller than the largest obfuscation distance selected for our testing [34]. Additionally, the results demonstrate that even with very small distances unintentional consequences may occur; even moving the point using a radius of 1 meter resulted in misclassification. Although relatively rare, point obfuscation 1 meter away could change a point’s county in 0.0007% of the OMOP data (7/1,000,000) or a point’s census tract in 0.09% of the CCMEO data (58/58,102). There are no universally accepted best practices for point obfuscation, and the most effective allowed distance may vary with study area and application [16]. Due to the inclusion of the city of Chicago in the Cook County data, the census-designated regions in the CCMEO data are geographically smaller than those in our OMOP data, which cover multiple states; census tracts generally contain between 1200 and 8000 residents, meaning urban census tracts are geographically smaller than rural census tracts. This can be seen in our results, where a radius of 100 meters or larger yielded a higher percentage of shifts in our CCMEO data than our OMOP data.

Our results demonstrate that indiscriminate point obfuscation can shift a point into different census-designated geographical regions; this is a natural and expected consequence of moving a point. However, we now discuss and quantify the potential impact of shifting to linked administrative units (i.e., neighborhoods) by comparing census demographics before and after obfuscation. The United States Census Bureau conducts large-scale surveys, such as the decennial census and the American Community Survey (ACS). The yearly ACS samples approximately 250,000 household units monthly. From the ACS, we picked the estimated number of “individuals with income in the past 12 months below poverty level” as an example demographic; these data are publicly available at the census tract level. We chose poverty status due to its saliency in health outcomes research [35,36].

We give a high-level overview of the obfuscation impact on poverty status measurement in Table 2 to justify the need for a geographically constrained obfuscation technique when assigning points to a population-based rate of individuals living under the poverty line (as a percentage of the total population). In our OMOP data, a pinwheel distance of 1000 meters resulted in 22.4% (n=224,065) of records with a different poverty rate after obfuscation where those changes were, on average, a mean 7.3% (SD 7.4%) away from the original rate (median 5%, range 91.9% to –78.6%). We also show the magnitude of the difference between the original and obfuscated address by showing minimum and maximum differences of rates. Negative differences imply the obfuscated record had a lower assigned poverty rate while positive differences imply the obfuscated record had a higher assigned poverty rate. For completeness and to complement Table 1, we include small distances of 10 and 1 meters in Table 2, although we did not anticipate such small distances would impact poverty rate assignments.

Table 1. The number of records shifted into a different region varied per obfuscation radius (in meters) and size of region.

<table>
<thead>
<tr>
<th>Data set and radius (meters)</th>
<th>Records, n×1000 (%)</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>Block group</td>
<td>Tract</td>
</tr>
<tr>
<td>Observational Medical Outcomes Partnership (n=1,000,000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>747 (75)</td>
<td>367 (37)</td>
</tr>
<tr>
<td>500</td>
<td>616 (62)</td>
<td>210 (21)</td>
</tr>
<tr>
<td>100</td>
<td>236 (24)</td>
<td>35 (3.5)</td>
</tr>
<tr>
<td>1</td>
<td>3.2 (0.34)</td>
<td>0.359 (0.03)</td>
</tr>
<tr>
<td>1</td>
<td>0.304 (0.03)</td>
<td>0.044 (0.004)</td>
</tr>
<tr>
<td>Cook County Medical Examiner’s Office (n=58,102)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>53 (91.9)</td>
<td>40 (70.4)</td>
</tr>
<tr>
<td>500</td>
<td>49 (85.4)</td>
<td>29 (50.7)</td>
</tr>
<tr>
<td>100</td>
<td>27 (47.7)</td>
<td>7.8 (13.4)</td>
</tr>
<tr>
<td>10</td>
<td>1.2 (2.22)</td>
<td>0.495 (0.85)</td>
</tr>
<tr>
<td>1</td>
<td>0.187 (.032)</td>
<td>0.082 (0.14)</td>
</tr>
</tbody>
</table>
Table 2. Poverty rates are substantially different before and after obfuscation.

<table>
<thead>
<tr>
<th>Data set and distance (meters)</th>
<th>Records with changed poverty rate, n (%)</th>
<th>Difference in rate (%), mean (SD)</th>
<th>Difference in rate (%), median (maximum to minimum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observational Medical Outcomes Partnership (n=1,000,000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>224,065 (22.4)</td>
<td>7.3 (7.4)</td>
<td>5 (91.9 to –78.6)</td>
</tr>
<tr>
<td>500</td>
<td>113,682 (11.3)</td>
<td>7.3 (7.5)</td>
<td>5 (91.9 to –78.6)</td>
</tr>
<tr>
<td>100</td>
<td>16,121 (1.6)</td>
<td>7.4 (7.7)</td>
<td>5.1 (78.6 to –78.6)</td>
</tr>
<tr>
<td>10</td>
<td>193 (0.01)</td>
<td>6.5 (6.4)</td>
<td>4.2 (35.9 to –32.6)</td>
</tr>
<tr>
<td>1</td>
<td>18 (0)</td>
<td>11.7 (8.1)</td>
<td>10.5 (38.8 to –12)</td>
</tr>
<tr>
<td>Cook County Medical Examiner’s Office (n=58,102)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>30,843 (53.1)</td>
<td>8.3 (7.7)</td>
<td>6 (61.7 to –69)</td>
</tr>
<tr>
<td>500</td>
<td>19,139 (32.9)</td>
<td>8.2 (7.6)</td>
<td>5.9 (58.4 to –61.7)</td>
</tr>
<tr>
<td>100</td>
<td>4506 (7.8)</td>
<td>8.2 (7.5)</td>
<td>6 (58.4 to –58.4)</td>
</tr>
<tr>
<td>10</td>
<td>343 (0.6)</td>
<td>8.1 (7.3)</td>
<td>6.2 (33.2 to –55.4)</td>
</tr>
<tr>
<td>1</td>
<td>58 (0.1)</td>
<td>6.8 (6.5)</td>
<td>4.4 (31 to –18.2)</td>
</tr>
</tbody>
</table>

A larger percentage of records in the CCMEO data experienced rate changes in comparison to our OMOP data. With a distance of 1000 meters, 53.1% (n=30,843) of records were assigned into a region with a different poverty rate where those changes were a mean 8.3% (SD 7.7%) away from the original rate (median 6%, range 61.7% to –69%). The magnitude of changes was not substantially different from our OMOP data (averages of 7.3% vs 8.3% and medians of 5% vs 6%, respectively, for 1000 meters); yet the frequency of these changes was notably higher (22.4% vs 53.1%, respectively, for 1000 meters).

Figure 2 shows an example census tract (17031031100) in Cook County, Illinois; 33 deaths were recorded in this area. Figure 2A shows an example simulation using pinwheel obfuscation with a 1000-meter radius; Figure 2B shows the results of our geographically constrained pinwheel obfuscation. The original point is orange, and the obfuscated point is blue; the census tracts are colored according to quintile of our poverty measure, where lightly colored areas have the lowest poverty rates. For this example, pinwheel obfuscation resulted in 22 of 33 (66%) of the points shifting census tracts; 12 of 33 (36%) of these were shifted into areas of higher poverty, while 10 of 33 (33%) were shifted into areas of lower poverty. Of those positive shifts, 4 of 12 were pushed to the highest category (33%) while the other 8 were moved into the second-highest poverty quintile. This example shows how obfuscation may move points from one extreme to another.

Hot spot analysis is known to be an effective tool for understanding how health outcomes and social determinants of health concentrate and cluster together [37-39]. We explored the impact of indiscriminate point obfuscation on a hot spot analysis of deaths by suicide from our CCMEO data; suicides were identified by the manner of death field in the CCMEO data and span the time period August 2014 to April 2022. Figure 3 shows the results of hot spot analyses using ArcGIS Pro [40] on the original data (Figure 3A) and data obfuscated with the pinwheel method using an unconstrained 1000-meter radius (Figure 3B). The obfuscation naturally blurred the correct hot spots, but (unexpectedly) new hot spots emerged, as identified by the pink boxes in Figure 3B. Most notably, a hot spot (95% confidence) spills into the neighboring and uninhabited Lake Michigan (census tract 17031990000). Highlighted in green are regions of interest with substantial change; the upper green box demonstrates the disappearance of a hot spot (99% confidence), and the lower green box demonstrates how indiscriminate obfuscation can bridge 2 hot spots together and weaken the signal that distinct clusters exist. By definition, the hot spots corresponding to the geographically constrained pinwheel method are identical to the true clusters in Figure 3A because of the confinement to the point’s original census tract. When linked to an administrative boundary such as census tract, results are consistent before and after obfuscation when the pinwheel method is constrained; only distance-based results would be impacted by moving the original point. By constraining the pinwheel process to a specific geographic region, the results of any method depending upon aggregation within those regions will not be impacted by our method.
Figure 2. Poverty and point obfuscation of original points (orange) to obfuscated points (blue) using a pinwheel (A) and a geographically constrained pinwheel (B).
Discussion

Principal Findings

We demonstrated that imprecise point obfuscation results in shifts across geographic regions and showed that these shifts do result in points geolocating in regions with vastly different socioeconomic contexts. This justifies the need for more precise point obfuscation techniques; our method constrains the candidate points into a specific region, which guarantees identical regional demographics after privacy protection is applied. The official poverty rate for the United States was 11.5% in 2022, and in the state of Kentucky, it was 16.5%, which places it 46th in a ranking of poverty rates in the United States [41]. For comparison, New Hampshire was ranked first and has the lowest poverty rate of 7.2% [41]. These example rates for the United States indicate areas experiencing differences of 5% to 7% in poverty rate, such as those reported in Table 2, represent vastly different socioeconomic dynamics. The extreme of this is illustrated in Table 2, where the largest difference between rates before and after obfuscation was 91.9% when records having relatively low poverty rates were assigned into areas having extreme poverty rates of 100% after obfuscation. The frequency of rate differences was substantially higher in the CCMEO data, which represents only Cook County, Illinois, and is home to Chicago, the third-largest urban area in the United States. In the 1990s, there were notable declines in the concentration of poverty in Chicago [42]. This decline, mixed with concerns regarding how gentrification has impacted the socioeconomic dynamics of Chicago, may explain why changes in poverty rate would occur at a higher frequency than in our larger OMOP data [43]. The impact of these shifts is important to understand when working with sensitive, protected health information and social determinants of health to correctly identify associations between place and health. As an example of poverty and health, women in high-poverty places are at greatest risk of being diagnosed with late-stage breast cancer [44]. Furthermore, different research studies may require different definitions of “neighborhood” when accessing socioeconomic statuses; for example, a person with multiple economic disadvantages may have a much narrower spatial range and limited social mobility.

We included small distances in our analysis to show that shifts occur even at very small distances. In obfuscation practice, small distances, such as 1 meter, would likely not be used due to the shifted point being too close to the original point and therefore not providing privacy; the balance between protecting privacy and protecting utility is context sensitive [21].

Limitations

Different analytical applications may be variably sensitive to shifts in demographics; our method eliminates analytical concerns by avoiding shifts altogether. The caveat to our method is that constraining inherently limits the maximum distance a point can travel, which may not be suitable for all applications in terms of privacy requirements. For example, in Figure 2, pinwheel obfuscation moved points on average 482 meters away, while our geographically constrained pinwheel algorithm moved points on average 217 meters away. Some applications may not be suitable for point obfuscation; for example, research studies requiring distance to be preserved between subjects and waypoints such as hospitals or clinics may not tolerate any shifting of geographic coordinates.

Public Health Impact

A previous evaluation of the pinwheel obfuscation method indicated that it had no major impact on geospatial analyses such as heat maps and hot spots [45]. However, we demonstrated that erroneous hot spots may be generated when analyzing deaths by suicide in Cook County, Illinois, including hot spots in uninhabitable areas. The issue presented in this paper is not a deficiency of the pinwheel technique, but a deficiency of data linkage using obfuscated points generated from any technique; if the variable that data linkage depends upon changes during obfuscation, then utility is harmed. We contend that any point
obfuscation technique may be constrained to specific geographies.

An alternative solution could be that socioeconomic demographics are calculated using the real address data before data release, but data owners, especially state and local health agencies, have varying degrees of technical sophistication and may not be able to compute demographics. Our research group geocodes electronic health records on behalf of our local health care enterprise on campus, and we make the data available to any university researcher using our local data warehouse. Mobile applications, networking, and research with Internet of Things technology have explored privacy at different levels; our work is a step closer to context-aware point obfuscation within the epidemiology domain.

Conclusions
A growing number of publicly available data sets are including precision geographic data for analysis. Our own work has explored decedent data published from open data portals for use in precision public health [14,33]. Point obfuscation can naturally shift a point into a different census-designated region; the regional differences before and after shifting highlight significantly different socioeconomic demographics. This is a natural consequence of moving a point and is not a weakness of the techniques themselves. We chose poverty as an example demographic due to its popularity in public health research; we also wish to explore the results of linking other census-level demographics. As future work, we will evaluate other techniques of point obfuscation and explore how these techniques may differ from those presented here. We show that it is possible to enhance point obfuscation by constraining where the new point may be placed; this ensures that the original point is obfuscated in a way that will not impact analyses depending upon the linkage to external region-based data.

Acknowledgments
This project is fully supported by the US Centers for Disease Control and Prevention of the US Department of Health and Human Services as part of grant 1R01CE003360-01-00. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement by, the US Centers for Disease Control and Prevention, the US Department of Health and Human Services, or the US government.

Conflicts of Interest
None declared.

References


Abbreviations

ACS: American Community Survey
CCMEO: Cook County Medical Examiner’s Office
GIS: geographic information system
HIPAA: Health Insurance Portability and Accountability Act
OMOP: Overdose Detection Mapping Application Program
ODMAP: Observational Medical Outcomes Partnership

©Daniel Harris, Chris Delcher. Originally published in the Online Journal of Public Health Informatics (https://ojphi.jmir.org/), 21.05.2024. 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 the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://ojphi.jmir.org/, as well as this copyright and license information must be included.