The Untapped Potential of Nursing and Allied Health Data for Improved Representation of Social Determinants of Health and Intersectionality in Artificial Intelligence Applications: A Rapid Review

IMIA Student and Emerging Professionals Group

Charlene Esteban Ronquillo (first co-author), James Mitchell (first co-author), Dari Alhuwail, Laura-Maria Peltonen, Maxim Topaz, Lorraine J. Block

1 School of Nursing, University of British Columbia Okanagan, Kelowna, Canada
2 School of Computing and Mathematics, Keele University, UK
3 Information Science Department, Kuwait University, Kuwait and Health Informatics Unit, Dasman Diabetes Institute, Kuwait
4 Department of Nursing Science, University of Turku, Finland
5 School of Nursing, Columbia University, New York, USA
6 School of Nursing University of British Columbia Vancouver, BC, Canada

Summary

Objectives: The objective of this paper is to draw attention to the consideration of social determinants of health (SDoH) and intersectionality data in electronic health records (EHRs), towards the development of equitable artificial intelligence (AI) technologies.

Methods: A rapid review of the literature on the inclusion of nursing and allied health data and the nature of health equity information representation in the development and/or use of artificial intelligence approaches alongside expert perspectives from the International Medical Informatics Association (IMIA) Student and Emerging Professionals Working Group.

Results: Consideration of social determinants of health and intersectionality data are limited in both the medical AI and nursing and allied health AI literature. As a concept being newly discussed in the context of AI, the lack of discussion of intersectionality in the literature was unsurprising. However, the limited consideration of social determinants of health was surprising, given its relatively longstanding recognition and the importance of representation of the features of diverse populations as a key requirement for equitable AI.

Conclusions: Leveraging the rich contextual data collected by nursing and allied health professions has the potential to improve the capture and representation of social determinants of health and intersectionality. This will require addressing issues related to valuing AI goals (e.g., diagnostics versus supporting care delivery) and improved EHR infrastructure to facilitate documentation of data beyond medicine. Leveraging nursing and allied health data to support equitable AI development represents a current open question for further exploration and research.

Keywords
Artificial intelligence, health equity, social determinants of health, health personnel, informatics

Yearb Med Inform 2022
http://dx.doi.org/10.1055/s-0042-1742504

1 Introduction

Inadequate representation of diverse populations in health system datasets represents an important challenge in the development of fair and equitable artificial intelligence (AI) technologies for health. Adverse impacts of AI technologies such as the development of racially biased AI algorithms that result from the lack of diversity of datasets have been spotlighted in recent famous examples, such as an AI algorithm that generated clear pictures from pixelated images which reconstructed the faces of President Obama and other people of colour into white men and women [1]. Improved representation of the features of underrepresented population subgroups in datasets used for AI development is one component towards AI fairness. It is commonly suggested that improved and more complete capture of sensitive features of underrepresented groups (e.g., demographic characteristics) and social determinants of health (SDoH) in AI datasets are necessary towards the development of less biased algorithms [2, 3]. Beyond providing a foundation for less biased algorithm development at the outset, diverse datasets are also necessary to enable computational approaches to mitigation of bias in AI; these approaches rely, to a large extent, on data quality. For instance, approaches to de-biasing algorithms include reducing the difference in model performance between predefined privileged and under-privileged groups relative to a chosen set of sensitive features (e.g., age or sex) [4-6]. Notably, many de-biasing approaches assume a sufficient volume and type of data to be already present in the dataset to enable computational bias mitigation approaches. We argue that this assumption is often false in the context of health systems. Data captured in electronic health records (EHRs) in health
systems are notoriously difficult to access, messy, incomplete, and exist in formats not easily amenable for use in AI development and refinement (e.g., narrative clinical notes). As AI algorithms rely on historical data, it is also important to note the ways that the historical contexts of institutions and broader societal values shape the diversity of datasets [7]. For example, in many nations, it is only relatively recently that there are efforts towards the improved and routine collection of sensitive data such as race, ethnicity, gender identity, and sexual orientation (among others) [8].

As the aim of health equity is to reduce and eliminate disparities in health and mitigate determinants of health that adversely affect excluded or marginalized groups [9], representation of health equity in the context of EHRs is conceptually and operationally complex. Calls for better inclusion and routine capture of SDoH data towards improving AI development are encouraging [2] and signal a recognition of the necessity of these data for developing more equitable AI. However, the inclusion of SDoH is only a beginning step. To meaningfully represent the complexity of health equity in datasets, we argue that there is a need to capture sensitive features of populations beyond SDoH and extend to the consideration of intersectionality. Briefly, intersectionality theory orients towards the consideration of social power and the compounding impacts of multiple and concurrent experiences of advantage/disadvantage, power/oppression, and privilege/marginalization that may result from one’s intersecting social locations [10, 11]. Given the complexity of health equity and the known influence of intersectionality on health and wellbeing [12-16], we propose that improved representation of intersectional data in datasets in addition to SDoH is a necessary development towards developing equitable AI.

Attention to the lived impacts of SDoH and intersectionality, and communication of the patient story are among the key foci of many nursing and allied health professions as data sources to improve the representation of SDoH and intersectionality data in EHRs and limited representation of these data in the medical informatics AI literature. We discuss the potential for nursing and allied health data towards the broader aim of improving AI fairness and developing health equity-considerate AI. Through a review of the health professions’ informatics literature, we examine the incorporation of SDoH and intersectionality in the biomedical informatics literature and demonstrate the under exploration of nursing and allied health data in AI development in the nursing and allied health informatics literature. In this paper, we seek to develop insight into the question, “How are SDoH and intersectionality concepts considered in AI health professional research?”.

2 Objectives

The objective of this paper is to draw attention to the currently underused potential of clinical documentation by nursing and allied health professions as data sources to improve the representation of SDoH and intersectionality data in EHRs and limited representation of these data in the medical informatics AI literature. We discuss the potential for nursing and allied health data towards the broader aim of improving AI fairness and developing health equity-considerate AI.

3 Methods

This rapid literature review sought to identify published articles that were related to AI, different health care professional groups (e.g., Nursing, Pharmacy, and Medicine), intersectionality and SDoH [17]. The intent was not to perform an exhaustive review but to offer an illustration of the representation of research published in this topic area as a beginning step to better understand the current state of the evidence.

Authors CR, LB, and JM created the initial list of search terms, which were expanded and validated by the research team. Using the Medical Subject Heading (MeSH) browser hosted by the U.S. National Library for Medicine (https://meshb.nlm.nih.gov/search; Version: MeSH 2022 Preview) the selected terms were manually mapped for MeSH term representation (Table 1). Then, using PubMed as the database, we created lists of articles which used our selected MeSH terms (November 6, 2021). When searching, we included [mh:noexp] to turn off the automatic explosion of MeSH headings. This specified the search to use only the specific MeSH heading term indicated in the search strategy. This method also ensured the exclusion of any citations which had yet to be indexed, citations that were out of scope, and articles that did not have MeSH terms applied [18].

The results for each search were uploaded into EndNote X9 software, which were then exported to Covidence software for title and abstract screening. Duplicates were reviewed and removed. Each article was screened for inclusion and exclusion by one reviewer. Inclusion criteria were those articles written in English and were related to AI and the professional group. Exclusion criteria included those articles which had no abstract, editorials, commentaries, conference reports, keynotes, were not written in English, or had no relationship to AI and the professional group (table 1). Once the articles were screened, full-text data extraction was completed by one researcher, who considered the type of research and whether any intersectionality or SDoH concepts were represented. These concepts, drawn from the literature, included: geography, gender, sex, ethnicity, race, disability, class, age, discrimination, privilege, social exclusion/disadvantage, economic status, education, and citizenship [2,19-24]. Data extraction results were exported to an excel spreadsheet where descriptive statistics and syntheses of concept representation occurred.
4 Results

The searches located a cumulative result of 136 (two duplicates) nursing and allied health papers and a total of 534 (nine duplicates) medicine papers. After screening for inclusion, a total of 49 nursing and allied health papers went through to full review and data extraction. Similarly, after screening for inclusion, a total of 45 medicine papers went through to full review and data extraction. Given the almost 1:4 ratio of nursing and allied health papers to medicine search results, and the almost 1:1 ratio of nursing and allied health to medicine papers going through full review and extraction, we chose to explain our results as the representation of AI, intersectionality, and SDoH for nursing and allied health separate from medicine.

4.1 The Nature of Nursing and Allied Health Informatics and Artificial Intelligence Studies

Using the search and screenings methods described above, results were: Allied Health Personnel 9/11 relevant articles; Dentistry 21/32 relevant articles; Midwifery 0/2 relevant articles; Nurse Practitioners 1/6 relevant articles; Nursing 15/42 relevant articles; Occupational Therapy 1/2 relevant articles; Pharmacology 1/32 relevant articles; Physical Therapy 0 relevant articles; Physician Assistants 0/4 relevant articles; Social Work 0/3 relevant articles; and Speech Therapy 1/2 relevant articles. There were no MeSH terms found for Psychiatric Nursing and was therefore excluded from the article search method. Of the 49 papers evaluated, the most common types were text/opinion/discussion papers (21/49; 43%), followed by survey (10/49; 20%), case/prototype reports (7/49; 14%), literature reviews (6/49; 12%), diagnostic test accuracy (2/49; 4%), randomized control trial (1/49; 2%), descriptive (1/49; 2%), and qualitative (1/49; 2%).

None of the evaluated articles were explicitly written about intersectionality or SDoH in context to the health professions use or development of AI. However, some of the papers described using some SDoH or intersectionality concepts as data points.
in the application or development of AI. Two papers described data related to sex [25, 26], two papers described data/AI related to disability [25, 27], one paper described data/AI related to race and ethnicity [26], and one paper described data related to education and employment [28]. For example, Jain et al. [26] performed a diagnostic study, where physicians and nurse practitioners diagnosed dermatologic conditions with or without the assistance of an AI-enabled prediction model. The authors noted that the data set used to develop and test the AI included information related to age, sex, race, ethnicity, Fitzpatrick scale skin type (classification of human skin colour with scale based on reaction to ultraviolet light, [29]), and skin condition. Notably, the authors also described the groups of people who were underrepresented in the AI case training and testing were those with Fitzpatrick skin types VI (dark brown or black skin; 0; 0%), type I (pale white skin; 2; 0.2%), and type V (brown skin; 25; 2.4%); those with type III (darker white skin, 668; 63.8%) were over represented. As another example, Taylor et al. [25] piloted a case-based reasoning tool for matching people with assistive smart home technologies using occupational therapist clinical assessment data. These data included (in-part) sex, mobility, cognitive abilities, and housing. The authors noted that the piloted reasoning tool used instrumental reasoning (means-ends method). However, they reflected that for many clients needing occupational therapy interventions, there is not always a clear or defined ‘end’ (i.e., recommend technology based on imputed data). It was explained that for many, the subtlety of matching a client to a meaningful intervention still required expert clinician knowledge.

4.2 The Nature of Medical Informatics and Artificial Intelligence Studies

Using the search and screenings methods described above, relevant articles totalled 45 of the 362 initially identified as related to AI + Medicine. Of the 45 papers extracted, the most common types were discussion papers (10/45; 22%), Diagnostic test accuracy studies (9/10; 20%), Cross-sectional studies (5/45; 11%), Systematic reviews (4/45; 9%), Qualitative research studies (4/45; 9%), Case reports (3/45; 7%), Cohort studies (2/45; 4%), Clinical prediction rule (1/45; 2%), non-randomised experimental studies (1/45; 2%). A further 6 of the 45 studies (13%) were classified as other, which represented studies such as discussions or evaluations outside of the categories described.

None of the evaluated articles explicitly discuss intersectionality or SDoH in context to the health professions use or development of AI. However, several papers described using some SDoH or intersectionality concepts as data points in the application or development of AI. However, several papers described using some SDoH or intersectionality concepts as data points in the application or development of AI. Takamine et al. [30] mention the lack of socioeconomic status consideration in risk prediction models as a reason for the non-adoption of AI. No other studies discussed or mentioned economic status. Duron et al. [31] used an AI training set that had similar female to male ratios. Similar studies such as Arnold et al. [32], Anand et al. [33], and Canales [34] noted that data sets used to develop and test the AI should include information related to age, sex, race, ethnicity, etc. However, few of the studies demonstrated the use of SDoH and discussions were largely around recommendations for inclusion; none mention intersectionality. Studies also fail to discuss the impact of excluding aspects of intersectionality or SDoH nor offer discussion of bias. However, Clark et al. [35] state that aspects such as race, ethnicity, gender, and other sociodemographic factors should be introduced to ensure AI systems do not have the potential to widen inequities via algorithmic and analytic biases.

5 Discussion

This rapid review of the literature provides a first step in providing a snapshot into the current state of evidence related to the representation of SDoH and intersectionality data in the biomedical informatics literature. Our findings serve to catalyze discussion as to the challenges in conceptually and operationally incorporating health equity considerations in the development of AI. Given the complexity of intersectionality and only beginning discussions raising its relevance for equitable AI [20], it was unsurprising to find no discussion of intersectionality in our review of the medical informatics and AI literature. It is likely that there has not yet been attention paid to intersectionality in the context of AI in health, with one potential reason being that the roots of intersectionality in critical race theory, law, and the social sciences are fields that are not yet immediately linked to AI development in the context of health.

While the limited representation of intersectionality is unsurprising, we expected to find that representation and discussion of SDoH in the medical AI literature would be more common; our findings did not align with this expectation. Compared to intersectionality, recognition of the importance of SDoH across the health literature is substantial, as noted by the World Health Organization [36] and numerous national and international organizations. We also found that the term ‘Social Determinants of Health’ exists in MeSH, facilitating the searchability and coding of related papers, while the concept of Intersectionality could not be identified as a MeSH term. Nevertheless, very few articles in the medical informatics and AI literature mention any aspects of SDoH in the context of AI fairness and bias, suggesting that SDoH considerations are not yet routine in the development of AI for health. A potential contributor to the lack of SDoH representation in datasets are limitations posed by EHR designs, poor data quality, and similar to intersectionality, challenges in operationalizing the complex concept of SDoH into a format that is amenable to documentation in EHRs [3, 37]. Fluid definitions of SDoH and the lack of standards for the capture and representation of SDoH stand as substantial challenges [37]. There are important arguments for and against the need for some reductionism to facilitate data re-use versus the risk of losing meaningful, nuanced, and complete representation of the patient story. As established standards and consensus around the capture of SDoH and initial mention of intersectionality to support equitable AI development remains to be seen, looking to alternative data sources to improve representation of SDoH and intersectionality data may be a way forward in the interim.
The scope and nature of practice of nurses [38] and allied health professionals, particularly those who work in the community and public health, require rich clinical documentation of patient contexts and environments, as these inform care plans, goals of care, and guide decision making for care delivery and service referrals. This documentation that aims to capture patient context and social needs as it relates to their health and well-being and likely captures details of SDoH [39], and potentially, intersectionality. To leverage rich data collected by nursing and allied health, however, a number of issues must be addressed. One is a question of prioritization with regard to AI development targets, with a greater focus of AI towards supporting diagnostics (in the sphere of medicine) versus AI for the purpose of supporting care management and delivery (in the sphere of nursing and allied health). These trends suggest that there may be implicit values being attributed to some AI goals as compared to others, fuelling current foci of AI developments targeting medical goals. A related issue is the question of data availability and quality between medical data and allied health data. In many settings, EHRs are geared towards capture and reporting of medical data, manifested as the dominant and sometimes exclusive use of the International Classification of Diseases codes (for medicine). As such, the EHR infrastructure in the majority of health systems lends itself to producing the largest volumes of structured medical data, providing the datasets necessary for AI development. Meanwhile, there is evidence of the insufficiency of EHR designs for nursing [40], which likely relates to the tendency to capture rich nursing and allied health data (that can include details of SDoH and intersectionality) within narrative free-text clinical notes, a data source that is only beginning to be explored for use in AI developments [39, 41, 42]. While the results of our review highlight limited discussion of SDoH and intersectionality in the nursing and allied health AI literature similar to what was found in medicine, it is possible that this limitation is linked to the aforementioned issues of insufficiency of EHRs to support clinical documentation outside of medicine as well as the likelihood of narrative free-text documentation within EHRs that is completed by these health professions. As such, we offer that the potential to leverage nursing and allied health data to improve the capture and representation of SDoH and intersectionality in health systems datasets remains an outstanding question that requires further research and exploration. There is a particular opportunity to explore synergies in the form of increased engagement with AI developments in nursing [43-45] and allied health [46-49].

Limitations

While we did not intend to fully examine the depth of literature on this topic, our findings are limited by the method of rapid literature review completed. We excluded papers based on the abstracts’ lack of mention of the application of AI for use by a specific health care professional group. This decision could have missed papers whose authors may have had assumptions of applicability/use, rather than an explicit mention in the abstract. Bias may have been introduced through our decision to use only one reviewer to screen the literature and perform data extraction. A deeper review is necessary to further advance our understanding of how health care professional AI incorporates and impacts intersectionality and SDoH.

6 Conclusions

This paper sought to explore the topic of health equity and AI for different health care professional groups. Using a rapid literature review approach, we evaluated related papers for any representation of intersectionality and SDoH concepts. We found that none were explicitly written on the subject, and very few mentioned using or applying these concepts in AI design, development, or implementation. This suggests that equity-related concepts in AI are lacking, but opportunities to address this limitation could be found through better inclusion nursing and allied health professional data. Overall, the improved capture and representation of intersectionality data in health datasets and AI bias mitigation using these data will require interdisciplinary efforts and perspectives.

References
