

NAPCRG 52nd Annual Meeting — Abstracts of Completed Research 2024.

Submission Id: 6130

Title

Bias Mitigation in Primary Healthcare Artificial Intelligence Models: A Scoping Review

Priority 1 (Research Category)

Healthcare informatics

Presenters

Maxime Sasseville, PhD, RN, Vincent Couture, PhD, Jean-Sébastien Paquette, MD, MSc, Steven Ouellet, PhD, Caroline Rheaume, MD, PhD, CCMF, FCMF, dipABLM, dipIBLM, Marie-Pierre Gagnon, PhD, Malek Sahlia, MSc, BSc, Frederic Bergeron, MSI

Abstract

Background: Artificial intelligence (AI) predictive models in primary healthcare can potentially lead to benefits for population health. Algorithms can identify more rapidly and accurately who should receive care and health services, but they could also perpetuate or exacerbate existing biases toward diverse groups. We noticed a gap in actual knowledge about which strategies are deployed to assess and mitigate bias toward diverse groups, based on their personal or protected attributes, in primary healthcare algorithms.

Objectives: To identify and describe attempts, strategies, and methods to mitigate bias in primary healthcare artificial intelligence models, which diverse groups or protected attributes have been considered, and what are the results on bias attenuation and AI models performance.

Methods: We conducted a scoping review informed by the Joanna Briggs Institute (JBI) review recommendations and an experienced librarian developed a search strategy.

Results: After the removal of 585 duplicates, we screened 1018 titles and abstracts. Of the remaining 189 after exclusion, we excluded 172 full texts and included 17 studies. The most investigated personal or protected attributes were Race (or Ethnicity) in (12/17), and Sex, using binary “male vs female” in (10/17) of included studies. We grouped studies according to bias mitigation attempts in 1) existing AI models or datasets, 2) sourcing data such as Electronic Health Records, 3) developing tools with “human-in-the-loop” and 4) identifying ethical principles for informed decision-making. Mathematical and algorithmic preprocessing methods, such as changing data labeling and reweighing, and a natural

language processing method using data extraction from unstructured notes, showed the greatest potential. Other processing methods, such as groups recalibration and equalized odds, exacerbated predictions errors between groups or resulted in overall models miscalibrations.

Conclusions: Results suggests that biases toward diverse groups can be more easily mitigated when data are open-sourced, multiple stakeholders are involved, and at the algorithm' preprocessing stage. Further empirical studies with more diverse groups considered, such as nonbinary gender identities or Indigenous peoples in Canada, are needed to confirm and to expand this knowledge.

Downloaded from the Annals of Family Medicine website at www.AnnFamMed.org. Copyright © 2024 Annals of Family Medicine, Inc. For the private, noncommercial use of one individual user of the Web site. All other rights reserved. Contact copyrights@aafp.org for copyright questions and/or permission requests.