

Unmasking A.I.'s bias in healthcare: The need for diverse data

David Bobrowski (BHSc)¹; Hetshee Joshi (BHSc)¹

¹Faculty of Medicine, University of Toronto, Medical Sciences Building, 1 King's College Circle, Toronto, ON, Canada, M5S 1A8.

Health Equity

The Institute for Medicine defines six elements of quality health care, as follows: safe, effective, patient-centered, timely, efficient, and equitable.¹ Health Quality Ontario (HQO) and the Canadian Medical Association envisage health equity as a system in which individuals can achieve their full health potential and receive high-quality care that is fair and appropriate to them and their needs.^{2,3} Equity in the provision of healthcare services is undermined when preventable systemic conditions, encompassing remediable social determinants of health, constrain life choices. The World Health Organization has identified the social determinants of health as the bio-psychosocial circumstances in which people are born, develop, live, and age, with special reference to marginalized communities.⁴ HQO has acknowledged that the dimension of equity has not progressed at the same rate as other sociocultural parameters, largely because health equity is complicated by multiple evolving and interdependent causes.² The Ministry of Health and Long-Term Care report – *Patients First: A Proposal to Strengthen Patient-Centered Health Care in Ontario* – reaffirms this point. Some Ontarians, particularly Indigenous peoples, Franco-Ontarians, newcomers, and people with mental health and addiction challenges, face barriers in accessing the care they need when they need it.⁵ Therefore, health equity has been recognized as a goal in addressing the challenges that continue to limit treatment opportunities for a vulnerable segment of the Canadian population.² Women's College Hospital's Health Equity Plan and Health Gap campaign exemplify efforts to close health gaps through policies, programs, and practices of the organization.⁶ Moving forward, we need to determine how best to leverage information technology such that we might address the quality and equity problems that are endemic at present.¹

Artificial Intelligence: A Double-Edged Sword

There is a growing conversation about the role of artificial intelligence (AI) and machine learning in healthcare.⁷ Recent technological advances, particularly the digitization of medical documentation, have generated massive clinical datasets that offer insights and opportunities to improve outcomes throughout the healthcare continuum, from research to patient placement. However, this volume of information is not easily amenable to manual study. Machine learning in medicine has been endorsed as a means of automating the process of data analysis. Furthermore, few medical practitioners are familiar with these technologies, and

the process of upgrading the professional knowledge base has not kept pace with the advent of novel high-tech applications.⁷

In layman's terms, machine learning is comprised of computational algorithms that independently learn from the data provided to them.^{7,8} Machine learning can be divided into two main categories: supervised and unsupervised learning. Supervised learning occurs when the user provides the solutions to the "training" dataset from which the functionality of the algorithm can be expanded. For example, in training an algorithm on how to identify a dog, a user will provide pictures of various types of animals and label them accordingly with the aim of formulating the characteristics of the animal. In unsupervised learning, the algorithm finds patterns that exist within the data rather than being instructed on the type of answer to generate.

Currently, within medicine, most machine learning applications are focused on the detection of abnormal findings and the identification of prognostic outcomes.⁸ Some published examples of these techniques include the detection of diabetic retinopathy in images of retinal fundi, and the prediction of outcomes in acute ischemic stroke after intra-arterial therapy.^{9,10} In particular, genetics and molecular medicine have recently emerged as a major area of interest in big data analytics. With the use of machine learning, successful algorithms have been developed to depict protein-protein interactions with the goal of discovering novel therapeutic targets.¹¹ While these innovations are highly promising, AI programs have demonstrated racial and gender biases. An example of the most noteworthy of these is Google's image-recognition algorithm that auto-tagged pictures of dark-skinned people as "gorillas".¹² When translated into clinical practice, such errors may have grave consequences if not controlled for quality and equity care.

Bridging or Widening Health Disparities

AI is a critical factor in the provision of more equitable care by providing an etiologic analysis of health disparities and guiding the development of personalized therapies.^{1,13} For instance, the outcomes and adverse effects of different treatment strategies can be optimized by a longitudinal analysis, wherein AI is applied to large cohorts of diverse patients in order to capture rare conditions that are infrequently observed in traditional clinical trials.¹³ The adoption of electronic medical records (EMRs) across Canada serves as the fulcrum for big data analytics by permitting the study of health disparities in populations whose data was previously not available for rigorous analysis.¹³ Canada Health Infoway reported that EMR uptake among Canadian primary care physicians was at 85% in 2017 compared to 77% in 2014.¹⁴ However, analyzing this data is complicated by differences in EHR encoding systems

Corresponding Author:
David Bobrowski
david.bobrowski@mail.utoronto.ca

throughout North America, and data fragmentation across practices and institutions requires a common language.¹³

Beyond the ongoing logistical limitations of data collection, there exist more insidious ethical concerns based on the potential for algorithms or datasets to reflect human bias or pre-existing inequities when input into an AI program.^{15,16} These human biases can be identified throughout all levels of biomedical research – from the bench to the bedside. From this perspective, while increasing attention has been focused on the importance of including women in clinical research trials, there has been little attention paid to the inclusion of both sexes in basic science and translational research involving animal and cell models.¹⁷ Since biomedical research serves as the foundation for clinical research, the lack of sex-specific variation may serve to bias results, and conclusions derived from these studies may only be applicable to one sex. The need for sex-specific reporting of data has received international support from research funders and regulators alike, with this problem in mind.^{18,19} Despite these calls to action, sex-specific reporting in published research trials remains inadequate to date.²⁰⁻²² The potentially deleterious effect of neglecting sex-specific differences is exhibited by the finding of the U.S. Food and Drug Administration in 2013 that recommended a belated 50% reduction for women in the dosage of the sleep aid, zolpidem, based on male-to-female metabolic differences.²³ Higher serum concentrations of zolpidem were found to comparatively impair function and decrease alertness in women, contributing to an increased risk of automobile accidents.

Racial minorities have been equally marginalized within the existing database. In 1997, 92% of the participants in clinical trials were Caucasian, whereas in 2014, 86% of participants were Caucasian.²⁴ The paucity of racial and ethnic diversity in health data has life-threatening implications in clinical practice. For example, genetic polymorphisms of the CYP2C19 gene can produce poorly functioning CYP2C19 enzymes that are used in the hepatic biotransformation of clopidogrel, an antiplatelet prodrug.²⁴⁻²⁶ That is to say, CYP2C19 loss-of-function alleles can impair the formation of active drug metabolites and consequently reduce platelet inhibition by clopidogrel. Since clopidogrel does not function well in individuals with this genetic variation, affected patients prescribed clopidogrel are at an increased risk of developing myocardial infarction or cerebrovascular accident, and may require alternative antiplatelet therapy.²⁴⁻²⁶ There exist marked ethnic differences in the incidence of this polymorphic deficiency, and people of Asian heritage are three times more likely to be poor metabolizers compared to people of Caucasian heritage.^{24,26}

The importance of racial differences is amplified when considering that only an estimated 20% of participants in genome-wide association studies (GWAS) are not of European descent.²⁷ This omission obscures the association between genetic variants and disease traits, and between genetic variants and responses to drugs observed in populations of European heritage.²⁷ In addition, under-representation of populations whose ancestry is not European limits the discovery of novel prognostic associations between genetics and health outcomes.²⁷ In GWAS performed on European-Americans, 25% of variants found to be associated with body mass index, type 2 diabetes and lipid levels have significantly different effect sizes in at least one non-European-American population—most frequently in African-Americans.²⁸ Likewise,

attempts to apply data from the Framingham Heart Study to predict the risk of cardiovascular events in non-Caucasian populations biased results, with the magnitude of associations between risk factors and presence of atherosclerotic disease varying between different ethnic and racial groups.^{15,29}

A number of medical specialties have already begun the process of testing AI for clinical use. Of particular interest, the field of dermatology is utilizing AI for the early detection of skin cancer to improve treatment outcomes. A study published in the *Annals of Oncology* pitted man against machine and found that AI was more sensitive and specific than consultation with a dermatologist in the diagnosis of skin cancer.³⁰ To achieve this result, AI researchers introduced more than 100,000 images of malignant and benign skin cancers and moles from an open-source repository of skin images, the International Skin Imaging Collaboration (ISIC), and input the diagnosis for each image to the deep learning convolutional neural network (CNN).³¹ However, the algorithm used by the CNN based most of its knowledge on the appearance of skin lesions on fair skinned individuals, potentially leading to the misdiagnosis of lesions in patients of colour. While fair-skinned patients are at the highest risk for developing skin cancer, the mortality rate of skin cancer in the African-American population is significantly higher (90% 5-year survival rate for Caucasians compared to 73% for African-Americans).³² This issue has been compounded by decades of clinical research focused primarily on people with light skin, excluding marginalized communities whose symptoms might present in a dissimilar way.³¹ The ISIC is currently working to expand its archive of skin lesion images to include as many skin types as possible in order to resolve this historic limitation. Once different skin tones are appropriately represented, this technology coupled with telemedicine may be able to provide remote diagnosis of skin lesions and thus also eliminate geographical barriers to care.

Moving Forward

Technological innovation has given rise to a sense of immediacy and opportunity among healthcare professionals who are interested in adopting these new services in the hope of improving patient outcomes. In particular, health equity remains a paramount public health concern and AI may provide an opportunity for enhanced surveillance of health disparities.^{1,13} However, potential bias inherent in healthcare-related datasets could inadvertently be built into algorithms – a challenge that must be addressed to avoid exacerbating existing structural inequalities. Fundamentally, these ethical considerations do not reflect a flaw of AI, but instead highlight the need for medical research and health care delivery models that provide a more inclusive approach to people of all ethnic backgrounds and genders. A number of initiatives are ongoing with the aim of raising awareness of population diversity in research and clinical trials, and to improve the quality of data collected on gender and race.^{6,24} At the same time, policy-makers will need to weigh regulatory safety versus added value, and the rigorous review process of evidence-based medicine versus the “fail fast” attitude of the technology industry.³³ To realize the full potential of these novel technologies and prevent widening equity gaps in the provision of healthcare services, a multi-sectoral approach must thus be taken to uphold ethical standards and build trust with patients.

References

- Health Quality Ontario. Quality Matters: Realizing Excellent Care for All [Internet]. Toronto; 2018. Available from: <http://www.hqontario.ca/Portals/0/documents/health-quality/realizing-excellent-care-for-all-en.pdf>
- Health Quality Ontario. Health Quality Ontario's Health Equity Plan [Internet]. Toronto; 2018. Available from: http://www.hqontario.ca/portals/0/documents/health-quality/health_equity_plan_report_en.pdf
- Health equity and the social determinants of health [Internet]. Canadian Medical Association. 2018 [cited 28 October 2018]. Available from: <https://www.cma.ca/En/Pages/health-equity.aspx>
- About social determinants of health [Internet]. World Health Organization. 2018 [cited 28 October 2018]. Available from: https://www.who.int/social-determinants/sdh_definition/en/
- Ministry of Health and Long-Term Care. Patients First: Action Plan for Health Care [Internet]. Toronto; 2015. Available from: http://www.health.gov.on.ca/en/news/bulletin/2015/docs/discussion_paper_20151217.pdf
- Women's College Hospital. Women's College Hospital's Health Equity Plan 2016-2020 [Internet]. Toronto; 2016 p. 1-6. Available from: https://www.womenscollegehospital.ca/assets/pdf/PatientsandVisitors/11961_WCH_Health_Equity_Plan_2016-2020.pdf
- Handelma GS, Kok HK, Chandra R V, et al. Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods. *AJR Am J Roentgenol* [Internet]. 2018 Oct 17 [cited 2018 Oct 28];1–6. Available from: <https://www.ajronline.org/doi/10.2214/AJR.18.20224>
- Deo RC. Machine Learning in Medicine. *Circulation* [Internet]. 2015 Nov 17 [cited 2018 Oct 28];132(20):1920–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26572668>
- Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* [Internet]. 2016 Dec 13 [cited 2018 Oct 28];316(22):2402. Available from: <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2016.17216>
- Asadi H, Dowling R, Yan B, et al. Machine Learning for Outcome Prediction of Acute Ischemic Stroke Post Intra-Arterial Therapy. Gómez S, editor. *PLoS One* [Internet]. 2014 Feb 10 [cited 2018 Oct 28];9(2):e88225. Available from: <http://dx.plos.org/10.1371/journal.pone.0088225>
- Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism* [Internet]. 2017 Apr [cited 2018 Oct 28];69:S36–40. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28126242>
- Hern A. Google's solution to accidental algorithmic racism: ban gorillas [Internet]. *the Guardian*. 2018 [cited 28 October 2018]. Available from: <https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people>
- Zhang X, Pérez-Stable E, Bourne P, et al. Big Data Science: Opportunities and Challenges to Address Minority Health and Health Disparities in the 21st Century. *Ethnicity & Disease*. 2017;27(2):95.
- Canada Infoway [Internet]. Infoway. 2018 [cited 28 October 2018]. Available from: <https://www.infoway-inforoute.ca/en/component/edocman/3362-2017-cma-workforce-survey-digital-health-results/view-document?Itemid=101>
- Char D, Shah N, Magnus D. Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *New England Journal of Medicine*. 2018;378(11):981-983.
- Rothwell P. Factors That Can Affect the External Validity of Randomised Controlled Trials. *PLoS Clinical Trials*. 2006;1(1):e9.
- Yoon D, Mansukhani N, Stubbs V, et al. Sex bias exists in basic science and translational surgical research. *Surgery*. 2014;156(3):508-516.
- Canadian Institutes of Health Research. What a Difference Sex and Gender Make: A Gender, Sex and Health Research Casebook [Internet]. Ottawa; 2012. Available from: http://www.cihr-irsc.gc.ca/c/documents/What_a_Difference_Sex_and_Gender_Make-en.pdf
- Food and Drug Administration. Guideline for the study and evaluation of gender differences in the clinical evaluation of drugs. United States: Federal Register. 1993;59:39406–39416.
- Mehta N, Rodrigues C, Lamba M, et al. Systematic Review of Sex-Specific Reporting of Data: Cholinesterase Inhibitor Example. *Journal of the American Geriatrics Society*. 2017;65(10):2213-2219.
- Phillips S, Hamberg K. Doubly blind: a systematic review of gender in randomised controlled trials. *Global Health Action*. 2016;9(1):29597.
- Geller S, Koch A, Pellettieri B, et al. Inclusion, Analysis, and Reporting of Sex and Race/Ethnicity in Clinical Trials: Have We Made Progress?. *Journal of Women's Health*. 2011;20(3):315-320.
- Questions and Answers: Risk of next-morning impairment after use of insomnia drugs; FDA requires lower recommended doses for certain drugs containing zolpidem (Ambien, Ambien CR, Edluar, and Zolpimist) [Internet]. *Fda.gov*. 2018 [cited 28 October 2018]. Available from: <https://www.fda.gov/drugs/drugsafety/ucm334041.htm>
- Knepper T, McLeod H. When will clinical trials finally reflect diversity?. *Nature*. 2018;557(7704):157-159.
- Scott S, Sangkuhl K, Stein C, et al. Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C19 Genotype and Clopidogrel Therapy: 2013 Update. *Clinical Pharmacology & Therapeutics*. 2013;94(3):317-323.
- Shu Y, Zhou H. Individual and ethnic differences in CYP2C19 activity in Chinese populations. *Acta Pharmacologica Sinica* [Internet]. 2000 [cited 28 October 2018];21(3):193-199. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/11324414?dopt=Abstract>
- Popejoy A, Fullerton S. Genomics is failing on diversity. *Nature*. 2016;538(7624):161-164.
- Carlson C, Matisse T, North K, et al. Generalization and Dilution of Association Results from European GWAS in Populations of Non-European Ancestry: The PAGE Study. *PLoS Biology*. 2013;11(9).
- Gijsberts C, Groenewegen K, Hoefler I, et al. Race/Ethnic Differences in the Associations of the Framingham Risk Factors with Carotid IMT and Cardiovascular Events. *PLOS ONE*. 2015;10(7):e0132321.
- Haenssle H, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology*. 2018;29(8):1836-1842.
- Lashbrook A. AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind [Internet]. *The Atlantic*. 2018 [cited 28 October 2018]. Available from: <https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/>
- Dermatologists provide recommendations for preventing and detecting skin cancer in people of color | American Academy of Dermatology [Internet]. *AAD*. 2018 [cited 28 October 2018]. Available from: <https://www.aad.org/media/news-releases/dermatologists-provide-recommendations-for-preventing-and-detecting-skin-cancer-in-people-of-color>
- Desveaux L, Shaw J, Wallace R, et al. Examining Tensions That Affect the Evaluation of Technology in Health Care: Considerations for System Decision Makers From the Perspective of Industry and Evaluators. *JMIR Medical Informatics*. 2017;5(4):e50.