

VIEWPOINT

Crystal N. Johnson-Mann, MD

Department of Surgery,
University of Florida
Health, Gainesville.

Tyler J. Loftus, MD

Department of Surgery,
University of Florida
Health, Gainesville.

Azra Bihorac, MD

Department of
Medicine, University of
Florida Health,
Gainesville.

Equity and Artificial Intelligence in Surgical Care

In 2003, the Institute of Medicine reported that Black US individuals receive fewer procedures and poorer-quality care than White individuals, independent of socioeconomic determinants of health. Seventeen years later, access to care and perioperative level of care assignments potentiate disparities in surgical care, particularly affecting Black patients.¹ These disparities are partially attributable to implicit bias that is entrenched deeply in US culture and media and are exacerbated when patient and physician demographics are mismatched. It is difficult to identify modifiable mechanisms of implicit bias because its latent mental constructs cannot be directly observed, but the weight of evidence suggests that many well-intentioned clinicians have 2 conflicting cognitive processes: one that is governed by a conscious, explicit system of beliefs and values, and one subconscious, implicit process that adapts to repeated stimuli. The former process is typically fair and equitable; the latter may drive implicit bias. Efforts to overcome implicit bias and health care disparities by building awareness and enacting structural changes to credentialing agencies and training curricula have yielded modest progress; additional strategies are needed. This Viewpoint endeavors to impart understanding of mechanisms by which artificial intelligence can either propagate or counteract disparities and suggests methods to tilt the balance toward fairness and equity in surgical care.

Artificial Intelligence Can Propagate Disparities

Artificial intelligence algorithms learn from data; when trained on biased data, algorithms produce biased results. Injustices occur when an algorithm is applied to a patient who is poorly represented by training data. In deriving a data set of inpatient surgeries at the University of Florida Health campus in Gainesville, 14% of all patients were Black; 73 miles away at the Jacksonville campus, 34% of all patients were Black. It seems unlikely that a model derived from these data sets would accurately represent patients in Detroit, Michigan, where approximately 79% of the population is Black, or Rochester, Minnesota, where approximately 8% of the population is Black.² Similarly, socioeconomic determinants of health vary substantially across geographic domains and people groups; these personalized variables (eg, residing neighborhood characteristics, education, and income) are predictive of postoperative outcomes.³ Therefore, careful model development, validation, and testing across diverse data sets is necessary to minimize the negative effects of biased source data.

Incorporation of clinician judgements in algorithm training can introduce implicit biases held by clinicians. For example, clinician notes contain information that can be learned by artificial intelligence models, affecting model accuracy and bias. In a study⁴ using psychia-

trists' notes to predict intensive care unit mortality and psychiatric readmission, there were significant differences in documented diagnoses across race, and model performance was worst when predicting outcomes for women and patients with public insurance.⁴ This source of bias can be mitigated by using input data that is "upstream" of clinician judgements, eg, objective vital sign and laboratory data collected by standardized methods.⁵ Careful management of each model input feature can minimize the risk of bias in model outputs.

Using race, ethnicity, sex, and gender as input variables for decision-support tools has potential benefits and harms. If the outcome of interest is influenced by these variables via a biologically proven or plausible mechanism, then ignoring them may degrade model accuracy; if not, then including these variables has the potential to introduce bias. This binary distinction becomes unrealistic when it is unknown whether demographic factors influence pathophysiology or reflect systemic bias, which is difficult to ascertain. One prominent decision-support tool incorporates the observation that Black patients have increased risk for mortality after coronary artery bypass.⁶ If this observation is attributable to suboptimal access to care for Black patients, then incorporating race/ethnicity as an input variable could decrease the likelihood that Black patients will garner the benefits of an indicated procedure.⁷ Therefore, it is critically important to use machine learning interpretation mechanisms, causal inference, and clinical interpretation of biologic plausibility to understand whether race, ethnicity, sex, and gender are causal, pathophysiologic mechanisms of disease or simply indicators that the outcome of interest is influenced by systemic bias. These elements should be mandated in machine learning-specific expansions of reporting guidelines for predictive models, such as the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement.⁸

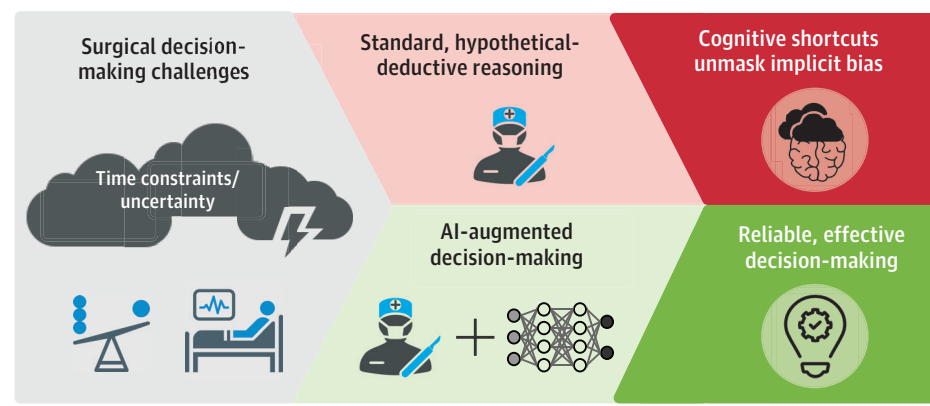
Artificial Intelligence Can Counteract Disparities

In surgical decision-making, clinicians make high-stakes decisions with incomplete information while facing time constraints. Under these conditions, humans tend to rely on heuristics, or cognitive shortcuts. Heuristics facilitate efficient decision-making and offer powerful advantages if they are guided by intuition that is honed by extensive experience with similar decision-making scenarios; unfortunately, heuristics are also influenced by implicit bias. To mitigate bias from heuristic-based decision-making within time constraints and uncertainty, decision-support tools must keep pace with clinical workflow. The emerging availability of live-streaming electronic health record data offers opportunities to generate and deliver artificial intelligence model outputs to health care professionals in real time. These models could identify

Corresponding

Author: Azra Bihorac, MD, Division of Nephrology, Hypertension, and Renal Transplantation, Department of Medicine, University of Florida Health, Precision and Intelligent Systems in Medicine (Prisma[®]), PO Box 100224, Gainesville, FL 32610-0224 (abihorac@ufl.edu).

Figure. Surgical Decision-Making Augmented by Artificial Intelligence (AI)



AI can augment surgical decision-making by minimizing the effect of cognitive shortcuts and implicit bias.

scenarios in which the decision made by a clinician is significantly different than the decision that yields the highest probability of achieving the goal of care, as illustrated in the Figure.

Artificial intelligence-augmented decision-making is founded on the hypothesis that a human with efficient access to accurate, computer-enabled information will make better decisions than a human who does not have access to such information. In this paradigm, the human is the most important element. If the computer does not provide the right information at the right time in the right format, then clinicians will ignore the information. Further, computer-enabled information must align with patient-centered outcomes. A model that recommends treatments yielding the highest probability of 30-day survival may not align with patients' values and preferences for resource intensity and quality of life after discharge. Therefore, input from patients and clinicians is essential to the design and implementation of optimal artificial intelligence-augmented decision-support tools.

Conclusions

Implicit bias is a major threat to surgical care. Surgery referral patterns and individual surgeon judgement are inherently affected by implicit bias, resulting in disparities; surgical decision-support tools propagate these disparities when source data and model input features are biased. Each input feature must be managed carefully to minimize the risk of bias in model outputs, including assumptions about whether associations between outcomes and race, ethnicity, sex, and gender are attributable to pathophysiologic mechanisms or systemic bias; this should be mandated in model reporting guidelines. Artificial intelligence models can also mitigate disparities by augmenting the decision-making process when time constraints and uncertainty unmask implicit bias. To humanize this process, the computer must conform to clinician needs and patient values. Understanding and implementing these principles is necessary to shift the influence of artificial intelligence toward equity and optimal surgical care in the next 17 years and beyond.

ARTICLE INFORMATION

Published Online: February 24, 2021.
doi:10.1001/jamasurg.2020.7208

Conflict of Interest Disclosures: Dr Loftus reported grants from the National Institutes of Health and was supported by the National Institute of General Medical Sciences of the National Institutes of Health (grant K23 GM140268). Dr Bihorac reported grants from the National Institutes of Health (GM110240) during the conduct of the study and grants from Atox Bio, La Jolla, Mallinckrodt, and Canadian Institutes of Health Research outside the submitted work. Dr Johnson-Mann reported grant support from the University of Florida. No other disclosures were reported.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Additional Contributions: The authors thank Chair of Surgery Gilbert Upchurch Jr, MD, for supporting

diversity, equity, and inclusion as primary missions of the University of Florida Department of Surgery.

REFERENCES

- Haider AH, Scott VK, Rehman KA, et al. Racial disparities in surgical care and outcomes in the United States: a comprehensive review of patient, provider, and systemic factors. *J Am Coll Surg*. 2013; 216(3):482-92.e12. doi:10.1001/jamainternmed.2018.3763
- Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med*. 2018;178(11):1544-1547. doi:10.1001/jamainternmed.2018.3763
- Bihorac A, Ozrazgat-Baslanti T, Ebadi A, et al. MySurgeryRisk: development and validation of a machine-learning risk algorithm for major complications and death after surgery. *Ann Surg*. 2019;269(4):652-662. doi:10.1097/SLA.0000000000002706
- Chen IY, Szolovits P, Ghassemi M. Can AI help reduce disparities in general medical and mental health care? *AMA J Ethics*. 2019;21(2):E167-E179. doi:10.1001/amajethics.2019.167
- Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. *JAMA*. 2019. doi:10.1001/jama.2019.18058
- Shahian DM, Jacobs JP, Badhwar V, et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: part 1: background, design considerations, and model development. *Ann Thorac Surg*. 2018;105(5):1411-1418. doi:10.1016/j.athoracsur.2018.03.002
- Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight: reconsidering the use of race correction in clinical algorithms. *N Engl J Med*. 2020;383(9):874-882. doi:10.1056/NEJMe2004740
- Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63. doi:10.7326/M14-0697