Health care organizations are grappling with how to discover and mitigate the risks of artificial intelligence (AI) and associated algorithms worsening racial, ethnic, and socioeconomic health disparities. For example, the potential for patient harm has been shown across a variety of conditions, including inequitable access to timely care due to biased pulse oximeters and inequitable access to care management due to label choice bias.1,2 The risk of perpetuating these harms is significant, hence multiple stakeholders aim to hold health care organizations and the health care industry responsible for harm and discrimination, if it occurs.

The evidence review by Chin and colleagues3 developed a set of guiding principles to mitigate and prevent bias in health care algorithms. Their guiding principles add precision and specificity that go beyond broad mandates, such as the White House Blueprint for an AI Bill of Rights and National Institute of Standards and Technology AI Risk Management Framework. However, there may be challenges to integrate these principles with other mandates and operationalizing these guiding principles into practice for health care and technology organizations.

To help address these issues, consortia such as Health AI Partnership and Coalition for Health AI aim to curate best practices for health care professionals to use AI safely, effectively, and equitably.4 Translating AI research into policy that then guides best practices clears only 1 hurdle. The other hurdles, which may be even larger, relate to the scarce people, processes, technology infrastructure, and operational supports required for health care organizations to successfully adopt and implement best practices. To do so, we highlight opportunities to close the translation gap from principles surfaced by Chin and colleagues into practice.3

**Transparency**

There is a need to share what is in the so-called black-box of AI algorithms to improve trust and maintain trustworthiness with patients regarding use of data to develop algorithms followed by the use of algorithms in clinical care. Chin et al3 emphasize the importance of patient and community engagement and go as far as to suggest that “All patients and communities should be informed when an algorithm is used in their care, should be advised about impact of the algorithm on their treatment, and should be provided alternatives if appropriate.” Unfortunately, this aspiration is far from reality and likely would be overwhelming. Consider the current situation of AI algorithms being used in everyday life. It happens when someone does an online search, shops online, or engages with a streaming service. Given the impact algorithms may have on health care, there would need to be a stronger regulatory framework, with potential financial implications, to ensure algorithms performed as expected and safely without bias.

**Life-Cycle Management**

Given the dynamics of AI and algorithms to continuously evolve, health care organizations must adopt a total product life-cycle approach to mitigate any unintended risks or changes over time as well as cause health disparities. Thankfully, there is consensus among diverse health care settings on the key decision points of algorithm adoption. As described by Chin and colleagues,3 the process begins with problem selection and formulation and ends with algorithm updates or discontinuation.
Developing procedures for health care organizations to assess potential impacts of algorithms on health inequities similarly found important activities across the entire product life cycle. The implication of these findings is that an algorithm cannot be certified as equitable. How an algorithm is put into practice determines the impacts on equity of the solution and continuous monitoring of the overall solution is required to prevent harm.

Validation

Characterizing algorithm operating performance and validation is critical for trustworthy deployment. Notably, clinical context matters regarding how best to validate and maintain algorithm performance and whether there are any potential biases that could be integrated into the algorithm. For example, there may be situations in which we know health care delivery is biased. Developing and evaluating algorithms in that setting could perpetuate ongoing biased delivery. Consider a set of scenarios for peripheral artery disease (PAD), a condition with strong evidence of disparities in diagnosis. Algorithm performance cannot be accurately measured on historical data in a setting where health care delivery is biased. Assume that in scenario 1, health disparities prevent a subset of Black adults from completing a full diagnostic workup. The subset of Black adults affected by barriers to diagnosis are all presumed to not have PAD, regardless of their true diagnosis. Under these conditions, the observed prevalence of PAD, positive predictive value of the algorithm, and sensitivity of the algorithm all diverge from true values. The better the algorithm works at identifying PAD among Black adults who face barriers to diagnosis, the more divergent algorithm performance will appear from true performance values.

Now, assume that the same PAD algorithm is evaluated in scenario 2, an environment where health disparities are proactively addressed and all Black adults access and complete a full diagnostic workup. Under these conditions, the observed prevalence of PAD, positive predictive value of the algorithm, and sensitivity of the algorithm all align with true values. In the first scenario, which uses retrospective data to assess algorithm performance, a health care organization can determine the PAD algorithm does not work for Black patients. But if action is taken to prospectively address health disparities, as in the second scenario, a health care organization can determine the PAD algorithm performs similarly across Black and other racial subgroups. This example highlights the critical role of prospective algorithm validation to fully characterize biases and impacts on health inequities.

Governance

Federal regulators must embrace a model of collaborative governance whereby technical assistance, coordination, and support are provided to health care delivery organizations to govern algorithms. Chin and colleagues highlight efforts at our own institution, Duke Health, to internally govern algorithms. However, there is unequal distribution of algorithmic governance capabilities and resources, and most health care organizations are unable to stand up similar programs. As noted, self-forming consortia, like Health AI Partnership, aim to develop freely available content for algorithmic governance teams across care settings, but just tools and education are not enough. Technical assistance programs that provide dedicated personnel and technology infrastructure will be required to ensure that low-resource care settings can operationalize total product life-cycle disparity assessments.

Evidence Generation

Given the impact algorithms will have on health and health care delivery, it will be critical to generate evidence on the benefits and risks. Like with other medical products, the public deserves to understand the efficacy and safety through simple, embedded clinical trials in health care systems.
as well as the effectiveness through implementation science. The implication of these needs is that funding bodies, such as the National Institutes of Health, Agency for Healthcare Research and Quality, Patient-Centered Outcomes Research Institute, and others, must be willing to invest in prospective evaluation of algorithms and their impact on health, including whether there are any unintended consequences, such as worsening health disparities. Many health care organizations are unable to conduct this testing without external support and resources.

In summary, the guiding principles surfaced by Chin and colleagues\(^3\) strengthen a foundation from which to drive changes in practice. However, efforts to surface principles and curate best practices are necessary, but not sufficient. Investments in technical assistance programs will be required for health care organizations to operationalize algorithm disparity assessments, and health care organizations will need to be held accountable to implement these assessments. Finally, for everyone's benefit, we will need rigorous evidence on how safe and effective algorithms could improve health and well-being.

**REFERENCES**


