Impact of Healthcare Algorithms on Racial and Ethnic Disparities in Health and Healthcare

March 3, 2023
10:00 a.m.- 2:00 p.m. ET
Welcome

Prashila Dullabh, MD
NORC

Racial Bias and Healthcare Algorithms
March 3, 2023
10:00–10:05 a.m. ET
Please use the Q&A feature to send comments or questions to the panelists.

Today's webinar is being recorded for notetaking purposes.

Live captions are available. Click the “more” icon on your screen and click “show subtitle”.

Update your Zoom name with your first & last name and organization (e.g., Jane Smith – NORC).

If you experience issues with Zoom, please contact Andrew Chiao OR Tyler Taylor from NORC, via the chat or email for assistance: chiao-andrew@norc.org, taylor-tyler@norc.org
Disclaimer

- Presentations do not necessarily represent the views of AHRQ or the U.S. Department of Health and Human Services (DHHS); therefore, please do not interpret any statement in this presentation as an official position of AHRQ or of DHHS.

- Additionally, presentations and presenters were selected to include diverse perspectives and do not necessarily represent the views of the consensus panel.
Consensus Panel

Maia Hightower, MD, MPH, MBA
University of Chicago Medicine

Bill Jordan, MD, MPH
American Medical Association

Tamra Moore, JD
King & Spalding
Welcome

Deborah Guadalupe Duran, PhD
NIMHD

Racial Bias and Healthcare Algorithms
March 3, 2023
10:00–10:05 a.m. ET
Welcome Remarks

Steven Posnack, MS, MHS serves as the Deputy National Coordinator for Health Information Technology. In this role, he advises the national coordinator, leads the execution of ONC’s mission, and represents ONC’s interests at a national and international level. In conjunction with the national coordinator, Steve oversees ONC’s federal coordination, regulatory policy, public-private initiatives, and the overall implementation of statutory authorities and requirements, such as those from the 21st Century Cures Act and HITECH Act.

Dr. LaShawn McIver, MD, MPH joined CMS as the Director of the Office of Minority Health in August 2020. She is a proven public health leader with experience in driving successful health initiatives and public policy efforts aimed at promoting health equity, improving health outcomes, increasing access to care, and promoting health system reform.
ONC Remarks

Stephen Posnack, MS, MHS
Deputy National Coordinator for Health IT
Office of the National Coordinator for Health Information Technology (ONC)

Racial Bias and Healthcare Algorithms
March 3, 2023
10:05–10:10 a.m. ET
CMS Remarks

LaShawn McIver, MD, MPH
Director, Office of Minority Health
Centers for Medicare & Medicaid Services (CMS)

Racial Bias and Healthcare Algorithms
March 3, 2023
10:10–10:15 a.m. ET
Evidence Review

Key Question 2

Brian Leas, MS, MA
University of Pennsylvania School of Medicine

Racial Bias and Healthcare Algorithms
March 3, 2023
10:15-10:35 a.m. ET
Key Question 2

What is the effect of interventions, models of interventions, or other approaches to mitigate racial and ethnic bias in the development, validation, dissemination, and implementation of healthcare algorithms?
KQ 2: Clinical Topics

warfarin dosing
opioid misuse
postpartum depression
lung function
lung cancer screening
healthcare costs and utilization
kidney function
stroke risk
cardiovascular risk
organ donation
intensive care needs
KQ 2: Studies

Study Design

- 1 RCT
- 17 cohort or pre-post
- 15 simulations

Risk of Bias

- 5 Low ROB
- 23 Moderate ROB
- 5 High ROB
KQ 2: Mitigation Strategies

- Removing Input Variables
- Replacing Input or Outcome Variables
- Adding Input Variables
- Changing the Patient Mix Used for Development and Validation
- Developing Separate Algorithms by Race
- Refining Statistical and Analytical Techniques
# KQ 2: Results

<table>
<thead>
<tr>
<th>Mitigation Strategy</th>
<th>N, Studies</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Removed race</td>
<td>15</td>
<td>eGFR for kidney function</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GLI spirometry equation for lung function</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Novel risk prediction algorithm for postpartum depression</td>
</tr>
<tr>
<td>Replaced race with biological indicators</td>
<td>4</td>
<td>eGFR for kidney function</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kidney Donor Risk Index for kidney transplant suitability</td>
</tr>
<tr>
<td>Replaced biased healthcare outcome variable with unbiased variables</td>
<td>1</td>
<td>Novel risk prediction algorithm for complex healthcare needs</td>
</tr>
<tr>
<td>Added race</td>
<td>3</td>
<td>FRS for cardiovascular disease risk</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHA2DS2-VASc for stroke risk</td>
</tr>
<tr>
<td>Added biological input variables</td>
<td>3</td>
<td>ASCVD for cardiovascular disease risk</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Novel risk prediction algorithm for cardiovascular disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COAG for warfarin dosing</td>
</tr>
<tr>
<td>Added health outcome variables</td>
<td>1</td>
<td>USPSTF-2020 for lung cancer risk</td>
</tr>
<tr>
<td>Added measures of SDOH</td>
<td>1</td>
<td>Novel risk prediction algorithm for complex healthcare needs</td>
</tr>
<tr>
<td>Recalibrated after improving population representation</td>
<td>4</td>
<td>ASCVD for cardiovascular disease risk</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Novel risk prediction algorithm for postpartum depression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Donor Risk Index for liver transplant suitability</td>
</tr>
<tr>
<td>Stratified algorithms for Black and White patients</td>
<td>2</td>
<td>COAG for warfarin dosing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Novel risk prediction algorithm for opioid misuse</td>
</tr>
<tr>
<td>Statistical techniques</td>
<td>2</td>
<td>Novel risk prediction algorithm for postpartum depression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ASCVD for cardiovascular disease risk</td>
</tr>
</tbody>
</table>
KQ 2: Findings

• Aside from eGFR, substantial heterogeneity: patient populations, clinical conditions, healthcare settings, primary outcomes

• Traditional algorithms, traditional solutions

• Mitigation strategies improve algorithmic accuracy, but inference and simulation used to estimate effect on disparities

• Modeling may not fully reflect potential biases in algorithm translation, dissemination, and implementation
KQ 2: Conclusions

• We don’t know what we don’t know: might be unpublished studies that found no effect of mitigation interventions

• Further research is needed to quantify the real-world effects of modifying algorithms

• Mitigation effectiveness is largely context-specific and may depend on algorithm, clinical condition, population, setting, outcomes
Mitigating Bias in Algorithms

Christina Silcox, PhD
Duke-Margolis Center for Health Policy

Racial Bias and Healthcare Algorithms
March 3, 2023
10:35 – 10:43 a.m. ET
Bias in Different Types of Algorithms

**Millions of black people affected by racial bias in health-care algorithms**

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

**Is a racially-biased algorithm delaying health care for one million Black people?**

Sweeping calculation suggests it could be – but how to fix the problem is unclear.
Overlap in Rules-Based and Machine Learning Algorithms

Rules based on human-derived rules and physics

Machine-derived patterns and weights

Human-derived statistical relationships, programmed as rules
Sources of Bias and Inequity

How Bias and Inequities Can Arise in Health AI

- Inequitable framing of challenge or users' next steps
- Unrepresentative Data
- Biased Training Data
- Choices in Data Curation and Model Development

Preventing Bias and Inequities in AI-enabled Health Tools
https://healthpolicy.duke.edu/publications/preventing-bias-and-inequities-ai-enabled-health-tools
Implementing Fairness & Equity Principles

Duke Health’s algorithmic oversight process translates ethical & quality principles into concrete evaluation criteria and submission requirements appropriate for each checkpoint throughout the lifecycle.

Quality & Ethical Principles
- Policies, Regulations etc.
- Fairness & Equity

Evaluation Criteria
- Committee Approval
- Evidence & Future Plans

Submission Material
- Development Teams
  - Analytical problem framing and sampling (societal, representation, measurement bias)
  - Justification re: sensitive variables as inputs, subgroup analysis on performance and impact metrics (algorithmic bias)
  - Intended and unintended uses, workflow integration (human user bias)

Educating Our Community

https://aihealth.duke.edu/algorithm-based-clinical-decision-support-abcds/
End-to-End Bias Evaluation Checklist for Predictive Models

Suchi Saria, PhD
John C. Malone Associate Professor
Computer Science, Stats, and Health Policy, Johns Hopkins University
Founder, Bayesian Health

Racial Bias and Healthcare Algorithms
March 3, 2023
10:43 – 10:51 a.m. ET
Algorithms are Sensitive to Healthcare Disparities

If not developed and implemented carefully, algorithms can propagate and create healthcare disparities.
Disparate Performance

• Many types of biases affect algorithm performance across subgroups.
  ► Exist at all stages of model development and deployment

• This leads to disparities created by the algorithm’s use in the real world.
## Sources of Bias

<table>
<thead>
<tr>
<th>Source of Bias</th>
<th>How the Bias Can Arise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model definition and</td>
<td>Use of a biased proxy target variable in place of the ideal prediction target.</td>
</tr>
<tr>
<td>design</td>
<td></td>
</tr>
<tr>
<td>Label bias</td>
<td>Use of a model, that due to its technical design, leads to inequitable outcomes.</td>
</tr>
<tr>
<td>Modeling bias</td>
<td></td>
</tr>
<tr>
<td>Data Collection</td>
<td>Poor performance in subsets of the deployment population due to non-representative training data.</td>
</tr>
<tr>
<td>Population bias</td>
<td>Bias due to differences in how features are measured across subgroups.</td>
</tr>
<tr>
<td>Measurement bias</td>
<td>Absence of validation studies that explicitly measure performance across subgroups.</td>
</tr>
<tr>
<td>Validation</td>
<td>Inconsistent user response to algorithm outputs for different subgroups.</td>
</tr>
<tr>
<td>Deployment</td>
<td></td>
</tr>
</tbody>
</table>

### Checklist for Bias Evaluation

**Case study: evaluating potential for bias in 30-day readmission risk models**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Source of bias</th>
<th>LACE</th>
<th>HOSPITAL</th>
<th>ACG</th>
<th>HATRIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Model definition and design</td>
<td>Label bias</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
</tr>
<tr>
<td>Modeling bias - general</td>
<td>RED</td>
<td>GREEN</td>
<td>RED</td>
<td>RED</td>
<td></td>
</tr>
<tr>
<td>Modeling bias – key feature missing</td>
<td>RED</td>
<td>RED</td>
<td>GREEN</td>
<td>GREEN</td>
<td></td>
</tr>
<tr>
<td>Modeling bias – accounting for bias</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td></td>
</tr>
<tr>
<td>2. Data collection and acquisition</td>
<td>Population bias</td>
<td>GREEN</td>
<td>GREEN</td>
<td>YELLOW</td>
<td>GREEN</td>
</tr>
<tr>
<td>Measurement bias - inputs</td>
<td>GREEN</td>
<td>GREEN</td>
<td>YELLOW</td>
<td>GREEN</td>
<td></td>
</tr>
<tr>
<td>Measurement bias – prediction target</td>
<td>RED</td>
<td>RED</td>
<td>GREEN</td>
<td>GREEN</td>
<td></td>
</tr>
<tr>
<td>Measurement bias - incompleteness</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td></td>
</tr>
<tr>
<td>3. Validation</td>
<td>Missing validation bias</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
<td>RED</td>
</tr>
<tr>
<td>4. Deployment and model use</td>
<td>Human use bias – different interpretation</td>
<td>RED</td>
<td>RED</td>
<td>YELLOW</td>
<td>RED</td>
</tr>
<tr>
<td>Human use bias – model use</td>
<td>YELLOW</td>
<td>YELLOW</td>
<td>YELLOW</td>
<td>YELLOW</td>
<td></td>
</tr>
<tr>
<td>Human use bias – reduce uncertain</td>
<td>GREEN</td>
<td>GREEN</td>
<td>GREEN</td>
<td>GREEN</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Model assessment heat map. An overall rating was given for each bias type based on the qualitative assessment of the checklist questions (details in Appendix 1.) RED indicates there is potential for concern. GREEN indicates there is limited potential for concern. YELLOW indicates the potential for concern is unclear or there is not enough information with which to draw a conclusion.

AI Code of Conduct (AICC) Project

Laura Adams
Senior Advisor, National Academy of Medicine

Racial Bias and Healthcare Algorithms
March 3, 2023
10:51-10:59 a.m. ET
NAM Leadership Consortium’s AI Initiatives

- **Newest AI initiative:** Align the field around a “current best practice” healthcare AI Code of Conduct to be:
  - implemented
  - tested
  - validated
  - continually improved

- Identify each stakeholder’s role in applying the Code at each stage of AI lifecycle

- Focus on equity, inclusion, and implementation vigilance
The Context

- Algorithms in healthcare are proliferating rapidly
- Evidence indicates the need for guidance, policy, and learning
- Aligning around a current best practice provides clarity, supports innovation, and promotes learning
- The NAM is in a unique position to convene a broad array of stakeholders to advance this work

*Transparency and inclusion are key project design principles*

- People support what they help create
- All of us are better than any one of us
The NAM AI Code of Conduct (AICCC) Project

• Build upon the existing body of work on AI principles/frameworks

• Then:
  ► Identify areas of convergence and gaps in current frameworks
  ► Seek expert stakeholder and public input
  ► Publish NAM Commentary paper that includes a draft “best practice” code of conduct for public comment

• Incorporate feedback and dive deeper into such issues as:
  ► Bias elimination and promotion of equity
  ► Involvement of those affected by AI in the development of AI
  ► Algorithmic vigilance post-implementation
  ► Data linkage and sharing
The NAM AICC Project (cont’d)

- Incorporate input into a NAM Special Publication which will include:
  - A current best practice healthcare AI Code of Conduct
  - A proposed methodology for Code implementation, testing, validation and continuous learning:
    - Collaborations with others are essential (e.g., Coalition for Health AI (CHAI) work in establishing algorithmic assurance/validation labs)
  - Role of each stakeholder at each AI lifecycle phase vis-à-vis the Code
  - Priorities for accelerating progress

Contact: ladams@nas.edu
Resources to Explore AI Bias Mitigation Strategies

Dr. Luca Calzoni, MD MS PhD Cand.
NIMHD ScHARE

Racial Bias and Healthcare Algorithms
March 3, 2023
10:59 – 11:07 a.m. ET
AI Bias Can Lead to Health Disparities

AI can be a transformative tool for improving care and population health, when developed with Health Equity

Many algorithms are biased by design, or trained on biased data
- Example: an algorithm to distinguish between malignant and benign moles trained on light-skinned patients

Biased algorithms can exacerbate existing inequities in socioeconomic status, race, ethnic background, disability, religion, gender or sexual orientation, and lead to health disparities

Harm to populations can also come:
- when Big Data is used for AI without consent
- when algorithms are not applied fairly across populations, or the socio-cultural context is missed due to lack of diversity in the AI workforce

If bias is in the world, it will always be present in the data and learned by AI

The public sector can:
1. address society biases introduced or uncovered by AI
2. establish fairness standards
3. regulate AI deployed in health systems
   but regulation is not enough, and clinicians cannot bear all liability risks/ burdens of identifying biases

To prevent biases from resulting in health disparities, we need:
- implementation strategies across the entire AI development cycle
- critical thinking and an ethical inquiry approach in AI users

Panch et al., 2019; FDA, 2017; Shachar, 2023
Look Deeper with More Eyes

“For the first time in history, we have a technology (AI) that is opening our eyes to who we are, and could allow us to play a conscious role in who we want to become.

Jennifer Aue - IBM Director for AI Transformation

<table>
<thead>
<tr>
<th>Who we are</th>
<th>Who we want to become</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Human biases are perpetuated or amplified in AI applications</td>
<td>▪ Develop AI with Equity to Prevent Health Disparities:</td>
</tr>
<tr>
<td></td>
<td>▪ Use models in <strong>context</strong></td>
</tr>
<tr>
<td></td>
<td>▪ Ensure R’s in AI:</td>
</tr>
<tr>
<td></td>
<td>▪ <strong>Repeatability</strong></td>
</tr>
<tr>
<td></td>
<td>▪ <strong>Replicability</strong></td>
</tr>
<tr>
<td></td>
<td>▪ <strong>Reproducibility</strong></td>
</tr>
<tr>
<td></td>
<td>▪ Increase AI workforce <strong>diversity</strong> to <strong>look deeper with more eyes</strong></td>
</tr>
</tbody>
</table>

Join our effort to implement mitigation strategies in:
- Project design
- Data
- Algorithm development and training
- Implementation

**AI Bias Mitigation Collaboratives**

ScHARE
ScHARE is a cloud-based population science data platform that offers researchers at all experience levels and disciplines:

- The ability to collaboratively use AI tools in a secure setting
- Access to SDoH and other population science large datasets

ScHARE fills three critical gaps:

1. Collaboratively advance **AI bias mitigation strategies and ethical inquiry** by increasing the use of diverse eyes and skills
2. Promote participation of **women and populations with health disparities** in data science
3. Leverage health disparities and healthcare outcomes **research opportunities** afforded by Big Data, AI and cloud computing

Register for ScHARE:

bit.ly/join-schare

Join our email list:

bit.ly/schare-news
Join the **SCHARE** Think-a-Thons

Think-a-Thons are **virtual meetings** for people conducting health disparities and healthcare outcomes research

- **Monthly sessions** (2 hours)
- Designed for new and experienced users
- Two types:
  1. **Instructional**
  2. **Research-focused**: teams collaborate around health disparities and healthcare outcomes research projects leading to publications
- Networking, mentoring and coaching opportunities

**Join our Think-a-Thons:**

bit.ly/think-a-thons

**Contact us:**

schare@mail.nih.gov

**AI bias mitigation:**

let’s keep the dialogue open
Impact of Healthcare Algorithms on Racial and Ethnic Disparities in Health and Healthcare

Consensus Panel Discussion/Q & A

March 3, 2023
11:07 – 11:45 a.m. ET
Discussion Questions

• What works, what’s missing, additional experiences and insights on bias mitigation strategies, including approaches to implementation?

• What guidance is needed to mitigate bias/what are the next steps, for different parts of AI lifecycle?
  ▶ Guidance for algorithm development, testing, and updating?
  ▶ Guidance for algorithm implementation?
Impact of Healthcare Algorithms on Racial and Ethnic Disparities in Health and Healthcare

Break
March 3, 2023
11:45 a.m.– 12:15 p.m. ET

Please take thirty minutes for lunch
Evidence Review
Contextual Question 2

Gary Weissman, MD, MSHP
University of Pennsylvania School of Medicine

Racial Bias and Healthcare Algorithms
March 3, 2023
12:15 – 12:35 p.m. ET
Contextual Question 2

What are existing or emerging national or international standards or guidance for how algorithms should be developed, validated, implemented, and updated to avoid introducing bias that could lead to health and healthcare disparities?
CQ 2: Methodology

• Primary literature searches
• Supplemental searches for guidelines, frameworks, white papers
• AHRQ’s Request for Information
• Technical Expert Panel and Key Informants
CQ 2: Who Develops Standards?

**Academic researchers**
- US
- UK
- Australia

**Regulatory Agencies**
- FDA
- NIST
- Canadian regulators

**Think tanks**

**Advocacy Groups**

**Corporations**
CQ 2: Examples

Good Machine Learning Practice for Medical Device Development:
Guiding Principles
October 2021

The U.S. Food and Drug Administration (FDA), Health Canada, and the United Kingdom’s Medicines and Healthcare products Regulatory Agency (MHRA) have jointly identified 10 guiding principles that can inform the development of Good Machine Learning Practice (GMLP). These guiding principles will help promote safe, effective, and high-quality medical devices that use artificial intelligence and machine learning (AI/ML).

Algorithmic Bias Playbook
Ziad Obermeyer
Rebecca Nissan
Michael Stern
Stephanie Eareff
Emily Joy Bembeneck
Sendhil Mullainathan
June, 2021

Microsoft Responsible AI Standard, v2
GENERAL REQUIREMENTS
FOR EXTERNAL RELEASE
June 2022
Table 14. Guidance, Standards, and Recommendations

<table>
<thead>
<tr>
<th>Resources</th>
<th>Stakeholder</th>
<th>Summary of Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preventing bias and inequities in AI-enabled health tools</td>
<td>Authors identified 4 areas of algorithmic bias. 1) Ineptitude framing of the healthcare challenge 2) Unrepresentative training data 3) Insufficient training data 4) Inefficient care with choices in data selection, curation, preparation, and model development. They also offer recommendations for key stakeholders. Developers should recognize the potential for harm, follow good machine learning practices, work with diverse teams, and develop an understanding of the problems being addressed, the data being used, potential differences across subgroups, and how the algorithm is likely to be used. Purchasers and users should test algorithms in their populations immediately and over time, focusing on patient outcomes. Health systems/payers/other owners of large health datasets should prioritize standardization reduce bias in subjective measures, and note where their data may differ across groups. FDA and other agencies should ensure that devices that use AI perform well for all subgroups, require clear, accessible labeling, and build systems to monitor for biased outcomes.</td>
<td></td>
</tr>
</tbody>
</table>

Who audits the algorithms? Recommendations from a field scan of the algorithmic auditing ecosystem | The medical algorithmic audit | Authors identified 4 areas of algorithmic bias. 1) Ineptitude framing of the healthcare challenge 2) Unrepresentative training data 3) Insufficient training data 4) Inefficient care with choices in data selection, curation, preparation, and model development. They also offer recommendations for key stakeholders. Developers should recognize the potential for harm, follow good machine learning practices, work with diverse teams, and develop an understanding of the problems being addressed, the data being used, potential differences across subgroups, and how the algorithm is likely to be used. Purchasers and users should test algorithms in their populations immediately and over time, focusing on patient outcomes. Health systems/payers/other owners of large health datasets should prioritize standardization reduce bias in subjective measures, and note where their data may differ across groups. FDA and other agencies should ensure that devices that use AI perform well for all subgroups, require clear, accessible labeling, and build systems to monitor for biased outcomes. | |

The medical algorithmic audit | Academic researchers based primarily in United Kingdom | Presents rationale (based on fairness and justice) and describes components of an algorithmic audit tailored to medicine. Expands on work of Raj by emphasizing intended use, intended impact, explanatory error analysis, subgroup testing, and adversarial testing in the context of healthcare. | |

Who audits the algorithms? Recommendations from a field scan of the algorithmic auditing ecosystem | Algorithmic Justice League | Not specific to healthcare, focuses on AI. Presents 6 recommendations for policymakers: 1) Require the owners and operators of AI systems to engage in independent algorithmic audits against clearly defined standards 2) Notify individuals when they are subject to algorithmic decision-making systems 3) Mandate disclosure of key components of audit findings for peer review 4) Consider real-world harm in the audit process, including through standardized harm incident reporting and response mechanisms 5) Directly involve the stakeholders most likely to be harmed by AI systems in the algorithmic audit process 6) Formalize evaluation and, potentially, accreditation of algorithmic auditors. | |

Microsoft responsible AI standard, v2: general requirements | Microsoft | Microsoft's detailed standards for AI algorithms. Shaped around 6 core goals: accountability, transparency, fairness, reliability and safety, privacy and security, and inclusiveness. Numerous principles relevant to healthcare disparities, including: F3.1) Identify and prioritize demographic groups, including marginalized groups, that may be at risk of being differentially affected by the system based on intended uses and geographic areas where the system will be deployed. Include: 1) groups defined by a single factor and 2) groups defined by a combination of factors. F2.2) Evaluate all data sets to assess inclusiveness of identified demographic groups and collect data to close any gaps. F2.1) Recess the system design, including the choice of training data, features, objective function, and training algorithm, to pursue the goals of minimizing differences between the rates at which resources and opportunities are allocated to identified demographic groups, paying particular attention to those that exceed the target maximum difference, while recognizing that doing so may appear to affect system performance and it is essential to do the steps to make such tradeoff explicitly. | |

Microsoft responsible AI standard, v2: general requirements | National Institute of Standards and Technology (NIST) | NIST has been developing the groundwork for consensus standards on bias in AI. The proposal is organized around 3 key stages: Pre-Design, Design and Development, and Deployment. Within each stage, the authors discuss challenges that can introduce bias and suggest multiple potential solutions. | |

A proposal for identifying and managing bias in artificial intelligence | Academic researchers at the University of Chicago Booth School of Medicine and the University of California Berkeley School of Public Health | Developers should recognize the potential for harm, follow good machine learning practices, work with diverse teams, and develop an understanding of the problems being addressed, the data being used, potential differences across subgroups, and how the algorithm is likely to be used. | |

A proposal for identifying and managing bias in artificial intelligence | National Institute of Standards and Technology (NIST) | Presents rationale (based on fairness and justice) and describes components of an algorithmic audit tailored to medicine. Expands on work of Raj by emphasizing intended use, intended impact, explanatory error analysis, subgroup testing, and adversarial testing in the context of healthcare. | |

A proposal for identifying and managing bias in artificial intelligence | Microsoft | Microsoft's detailed standards for AI algorithms. Shaped around 6 core goals: accountability, transparency, fairness, reliability and safety, privacy and security, and inclusiveness. Numerous principles relevant to healthcare disparities, including: F3.1) Identify and prioritize demographic groups, including marginalized groups, that may be at risk of being differentially affected by the system based on intended uses and geographic areas where the system will be deployed. Include: 1) groups defined by a single factor and 2) groups defined by a combination of factors. F2.2) Evaluate all data sets to assess inclusiveness of identified demographic groups and collect data to close any gaps. F2.1) Recess the system design, including the choice of training data, features, objective function, and training algorithm, to pursue the goals of minimizing differences between the rates at which resources and opportunities are allocated to identified demographic groups, paying particular attention to those that exceed the target maximum difference, while recognizing that doing so may appear to affect system performance and it is essential to do the steps to make such tradeoff explicitly. | |

A proposal for identifying and managing bias in artificial intelligence | University of California Berkeley School of Public Health | Developers should recognize the potential for harm, follow good machine learning practices, work with diverse teams, and develop an understanding of the problems being addressed, the data being used, potential differences across subgroups, and how the algorithm is likely to be used. | |

A proposal for identifying and managing bias in artificial intelligence | University of Chicago Booth School of Medicine | Developers should recognize the potential for harm, follow good machine learning practices, work with diverse teams, and develop an understanding of the problems being addressed, the data being used, potential differences across subgroups, and how the algorithm is likely to be used. | |

Good machine learning practice for medical device development | US Food and Drug Administration, Health Canada, Medicines and Healthcare Products Regulatory Agency | Developers should recognize the potential for harm, follow good machine learning practices, work with diverse teams, and develop an understanding of the problems being addressed, the data being used, potential differences across subgroups, and how the algorithm is likely to be used. | |
Recent guidance focuses more on AI, less on traditional algorithms

Recommendations often focus on these issues:

• Representative datasets
• Diverse, multidisciplinary teams
• Transparency
• Accountability
• Fairness
CQ 2: Unanswered Questions

• Role of self-regulation vs. external requirements

• Third-party auditing

• Standards for who?
  ▶ Algorithm developers
  ▶ End-users

• “Off-label” use

• Scope of concern
  ▶ Traditional algorithms
  ▶ AI
  ▶ Imaging

• Beyond race and ethnicity


Managing AI Bias with the NIST AI Risk Management Framework: A Socio-Technical Approach

Reva Schwartz
Principal Investigator – Bias in AI
National Institute of Standards and Technology

Racial Bias and Healthcare Algorithms
March 3, 2023
12:35 - 12:43 p.m. ET
Current focus on computational/statistical bias obfuscates the other two categories.
What is the AI RMF?

Voluntary resource for organizations designing, developing, deploying, or using AI systems to manage AI risks and promote trustworthy and responsible AI

- Flexibly applied
- Rights-preserving
- Measurable

AI Risk Management Framework

Govern

Manage

Map

Measure

National Institute on Minority Health and Health Disparities
How can the AI RMF help organizations manage the risks from AI bias?

• Operationalizes trustworthiness and societal values
• AI Governance and organizational culture
• Socio-technical approach
• Risks and impacts focused
Example: Why a socio-technical framing matters for managing bias

- model uses proxy for narcotic use (referred to as “addictiveness”) that is inherently unobservable
- resulting in
  - people being denied medication that they genuinely needed,
  - bias - women – who are more likely to experience chronic pain – were more likely to be wrongly identified as “highly addictive”


See also: https://www.wired.com/story/opioid-drug-addiction-algorithm-chronic-pain/
THANK YOU

Contact us via email at aiframework@nist.gov

For more info on the NIST AI RMF, visit https://www.nist.gov/itl/ai-risk-management-framework
Understanding AI + Health IT

Stephen Konya
ONC

Racial Bias and Healthcare Algorithms
March 3, 2023
12:43 - 12:51 p.m. ET
ONC – Understanding AI + Health IT

ONC Activities

- Standards
- Certification
- Exchange

ONC Objectives

- Federal
- Coordination
- State & Public

Advance the development and use of health IT capabilities
Establish expectations for data sharing
The Evolution of Health IT & Digital Health (in the US)

ONC is charged with formulating the federal government’s health IT strategy to advance national goals for better and safer health care through an interoperable nationwide health IT infrastructure.

Office of the National Coordinator (ONC)

HITECH Act 2009

Laying the foundation of EHRs across the industry

Leveraging EHRs to drive value

21st Century Cures Act 2016

Founded in 2004 by executive order

NIH

National Institute on Minority Health and Health Disparities
Proposed strategy and recommendations

• Are based on the premise that risk and corresponding controls should focus on health IT functionality – not the platform(s) on which such functionality resides or the product name/description of which it is a part.
• Seek to advance a framework that is relevant to current functionalities and technologies yet sufficiently flexible to accommodate the future and rapid evolution of health IT.

Agencies’ proposed strategy identifies three categories of health IT:

1) administrative health IT functions
2) health management health IT functions, and
3) medical device health IT functions

Source: FDASIA Health IT Report – Proposed Strategy & Recommendations for a Risk-Based Framework (ONC, FDA, FCC)

- JASON supports the collection and curation of new health data sources for AI applications as well. For example:
  - Capturing smartphone data
  - Integrating social and environmental data
  - Supporting AI competitions
- The recommendations in the new report underline the importance of ONC’s efforts toward interoperable and standardized health data and AHRQ’s efforts to effectively use those data to improve the quality and safety of patient care.
- These efforts will improve capabilities to exchange and appropriately use high-quality health data – critical elements in powering AI efforts in health and healthcare.

An ONC Artificial Intelligence Showcase (2022) - “Seizing the Opportunities and Managing the Risks of Use of AI in Health IT”

30+ showcase presentations grouped into the following 3 categories;

- Advancing Responsible AI in Health IT – Guiding Principles
- Transparency and Accountability
- Evaluating Data Input Needs & Real-World Performance

- View Agenda [PDF - 246 KB]
- View Presentation Slides [PDF - 12 MB]
- Watch Event Recording

### Figure 2: Framework of all AI Use Cases in Healthcare

<table>
<thead>
<tr>
<th>Population health</th>
<th>Individual health</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Care routing</strong></td>
<td><strong>Care services</strong></td>
</tr>
<tr>
<td><strong>Prevention</strong></td>
<td><strong>Diagnosis</strong></td>
</tr>
<tr>
<td><strong>Acute treatment</strong></td>
<td><strong>Follow-up and chronic treatment</strong></td>
</tr>
</tbody>
</table>

#### Population health:
- Surveillance and prediction
- Population risk management
- Intervention selection
- Intervention targeting

#### Individual health:
- Self-referral
- Triage
- Personalized outreach

#### Framework:

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Behavior change</td>
<td>Prevention</td>
</tr>
<tr>
<td>4A Data-driven diagnosis</td>
<td>Diagnosis</td>
</tr>
<tr>
<td>5A Clinical decision support</td>
<td>Acute treatment</td>
</tr>
<tr>
<td>6A Compliance monitoring</td>
<td>Follow-up and chronic treatment</td>
</tr>
<tr>
<td>4B Image-based diagnosis</td>
<td>Prevention</td>
</tr>
<tr>
<td>5B Monitoring: Inpatient monitoring, device monitoring</td>
<td>Acute treatment</td>
</tr>
<tr>
<td>5C AI-facilitated care: Self-care guidance, psych counseling</td>
<td>Follow-up and chronic treatment</td>
</tr>
<tr>
<td>5D AI-facilitated care: Robotic surgery, robotic PT</td>
<td>immune system</td>
</tr>
</tbody>
</table>

#### Health Systems:
- Medical records
- Fraud prevention
- Capacity planning and personnel management
- Quality assurance and training
- Claims processing
- Coding and billing

#### Pharma & Medtech:
- Clinical trial support and recruitment
- Supply chain and planning optimization
- Drug discovery and medtech R&D
- Process optimization
- Drug safety and pharmacovigilance
- Real world evidence and HEOR

*Figure from USAID’s “Artificial Intelligence in Global Health: Defining a Collective Path Forward” [https://www.usaid.gov/cii/ai-in-global-health](https://www.usaid.gov/cii/ai-in-global-health)
Key Findings:

• Respondents acknowledged the unacceptable nature of findings that CDSTs produce avoidable differences for patients of color
  ‣ One-third of respondents said they are not planning to reevaluate use of race and ethnicity in clinical algorithms

• Raised the absence of a central hub of accountability as a barrier to addressing these complex issues across scientific and medical professions
  ‣ Some recommended leadership from largest and most influential organizations (e.g., the Centers for Medicare & Medicaid Services) must assemble stakeholders to develop standards, guidance, and best practices for using race in CDSTs.

• Emphasized role of bias in CDST development and care delivery, suggesting solution lies upstream (e.g., at the level of health technology research and development and through clinician education)

• Strategies must be enacted to proactively correct and confront the challenges of the misuse of race and ethnicity in CDSTs

(access full report here)
"Many CDSTs are proprietary, making it difficult for independent researchers to evaluate and validate these tools to ensure they function as intended and do not disadvantage certain patients."

Figure 5. CDST Vendor Market Share

Cerner 25%
EPSi (Strata) 14%
Epic 11%
Stanson Health 6%
NDSC/Change 2%
CPSI/Evident 2%
Premier 5%
Nuance 5%
Truven/IBM 4%
Elsevier 4%
Zynx Health 3%
CPSI/Evident 2%
NDSC/Change 2%
Zynx Health 3%
Elsevier 4%
Truven/IBM 4%
Premier 5%
Nuance 5%
Stanson Health 6%
NDSC/Change 2%
CPSI/Evident 2%

#2 - Relationship Between CDSTs and Health Equity

- While CDSTs have vastly improved medicine, they also remain vulnerable to implicit and explicit biases inherent to big data.
- There is significant potential for the more advanced tools to address racial inequalities if care is taken to acknowledge the underlying data’s susceptibility to bias and proactively “clean” the data.
- Because big data replicates or amplifies human biases, adding its elements to CDST architecture with questionable foundations can yield unintended, yet avoidable consequences.

Source: House Ways and Means Committee Report: “Clinical Decision Support Tools (CDST) and the (Mis)use of Race”
#3 – CDSTs# Race/Equity Related Findings:

NEJM Study (June 2020)

- Racial correction in clinical algorithms is harmful for a range of conditions, from childbirth to cancer.
- Study authors concluded that race had been misinterpreted or misused in multiple CDSTs, resulting in worse outcomes for people of color.
- Incorporating race data into clinical algorithms can entrench disparities by potentially producing different treatment approaches for individuals that are not based on precision medicine but are simply chosen because of race/ethnicity, historical differences in outcomes based on race, discrimination, racism, and biases about race.

Source: House Ways and Means Committee Report: “Clinical Decision Support Tools (CDST) and the (Mis)use of Race”
US Core Data for Interoperability (USCDI):
The Minimum Dataset of the Health Care Delivery System

- ONC standard for minimum dataset required for interoperability
  - Defines required data elements and vocabulary standards
  - Agnostic to format

- Updated on annual cycle with federal agency and industry input
  - Updates based on multiple criteria including standards maturity and public/industry priority
With the publication of Draft USCDI v4, ONC is accepting feedback on its content until April 17, 2023.

ONC plans on releasing a final USCDI v4 in July 2023.
New Releases:
SDOH Toolkit and Learning Forum Sessions for the Health IT Community

1) Health IT Buzz Blog post

2) Register for the 2023 series, covering;
• Community Level Governance
• Values, Principles, and Privacy
• Implementation, Measurement and Evaluation
• SDOH Information Exchange Learning Forum Summary
• Through its role as the National Coordinator for Health IT, and by leveraging the FHITCC, ONC conducted an internal interagency assessment surrounding the use of Artificial Intelligence (AI) in the field of health IT.
• The survey aimed to achieve the following:
  ► compile a robust list of federal activities and initiatives currently underway
  ► collect insights on key areas of interest and focus priorities for those agencies
  ► and attempt to identify specific point of contacts for AI related work located within each agency.
• This activity was designed to compliment several other federal activities that were/are also intended to gain a better understanding of the current and planned federal landscape with respect to the responsible use of AI in Health IT. (i.e., activities being led by NAIIO/OSTP, OMB, OCIO/OCAIO, HRSA, AHRQ, ASPE, OCR, FDA, NIH, FTC, NIST, VA, etc.)
# 22 Different Agencies Responded

<table>
<thead>
<tr>
<th>Agency</th>
<th>Agency</th>
<th>Agency</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACF</td>
<td>FDA</td>
<td>OCIO / OCAIO</td>
</tr>
<tr>
<td>AHRQ</td>
<td>FTC*</td>
<td>OCR</td>
</tr>
<tr>
<td>CDC</td>
<td>HRSA</td>
<td>ONC</td>
</tr>
<tr>
<td>CMS</td>
<td>IHS</td>
<td>SSA*</td>
</tr>
<tr>
<td>DHA*</td>
<td>NASA*</td>
<td>USAID*</td>
</tr>
<tr>
<td>DoE*</td>
<td>NIH</td>
<td>VA*</td>
</tr>
<tr>
<td>DoS*</td>
<td>NIST*</td>
<td></td>
</tr>
<tr>
<td>DoT*</td>
<td>OASH</td>
<td></td>
</tr>
</tbody>
</table>

*Non-HHS Agencies
Based on your knowledge, which of the following use cases for the deployment of AI in healthcare, are considered as areas of potential interest for your agency? (check all that apply)

<table>
<thead>
<tr>
<th>ANSWER CHOICES</th>
<th>RESPONSES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>70.59%</td>
</tr>
<tr>
<td>Bias / Equity</td>
<td>64.71%</td>
</tr>
<tr>
<td>Clinical Care / Clinical Decision Support Tools (CDSTs)</td>
<td>64.71%</td>
</tr>
<tr>
<td>Natural Language Processing (NLP) / Voice Tech / Conversational AI / Chatbots</td>
<td>58.82%</td>
</tr>
<tr>
<td>Predictive Analytics (clinical)</td>
<td>58.82%</td>
</tr>
<tr>
<td>Medical Device / Diagnostics</td>
<td>52.94%</td>
</tr>
<tr>
<td>Population / Public Health</td>
<td>52.94%</td>
</tr>
<tr>
<td>Human / Social Services</td>
<td>47.06%</td>
</tr>
<tr>
<td>Imaging</td>
<td>47.06%</td>
</tr>
<tr>
<td>Predictive Analytics (business)</td>
<td>47.06%</td>
</tr>
<tr>
<td>Other (please specify)</td>
<td>41.18%</td>
</tr>
<tr>
<td>Administrative / Operational Functions</td>
<td>35.29%</td>
</tr>
<tr>
<td>Genomics / Precision Medicine</td>
<td>29.41%</td>
</tr>
</tbody>
</table>
Has your agency publicly issued any official guidance, policy statements, regulatory requirements, and/or strategy relating to the design, development or use of AI within healthcare?

Yes 23%
No 77%
ONC – Understanding AI + Health IT

Health Equity by Design / AI Bias

As part of ONC’s focus on ensuring health equity by design with regard to the adoption and use of technology, we are specifically interested in identifying work underway by federal agencies to try and mitigate the potential for systemic biases that could be exacerbated by the use of algorithms in health care and human service settings.

Please indicate if your agency is currently, or plans to pursue work in this area:

- Yes 38%
- Maybe 24%
- Unknown 21%
- No 17%
Please indicate your interest in the following areas of potential need for federal coordination when it comes to the adoption and use of AI specifically in the practice of health care.

- Standards Development / Adoption
- Surfacing & Alignment of Guiding Principles / Best Practices
- Certification Criteria and Enforcement Mechanisms
- Open and Inclusive Industry Convenings / Engagement
- Other (please specify)
- No Other Coordination is Necessary
ONC’s Health IT Buzz Blog Series: Artificial Intelligence & Machine Learning

• Back to the Future: What Predictive Decision Support Can Learn from DeLoreans and The Big Short (Dec. 2022)
• Two Sides of the AI/ML Coin in Health Care (Oct. 2022)
• Minimizing Risks and Maximizing Rewards from Machine Learning (Sep. 2022)
• Getting the Best out of Algorithms in Health Care (Jun. 2022)
Thank you!

Subscribe to our weekly eblast at healthit.gov for the latest updates!

Email: Stephen.Konya@hhs.gov

Phone: 202-690-7151

Health IT Feedback Form: https://www.healthit.gov/form/healthit-feedback-form

Twitter: @onc_healthIT

LinkedIn: Office of the National Coordinator for Health Information Technology

Youtube: https://www.youtube.com/user/HHSONC
Artificial Intelligence/Machine Learning (AI/ML)-Enabled Medical Devices

Matthew Diamond MD, MPH
Chief Medical Officer, Digital Health Center of Excellence
Center for Devices & Radiological Health (CDRH), US FDA

Racial Bias and Healthcare Algorithms
March 3, 2023
12:51- 12:59 p.m. ET
FDA’s Collaborative Patient-Centered Approach to AI/ML-Enabled Medical Devices

We’re working collaboratively with stakeholders to build a proactive, patient-centered approach to AI/ML-enabled devices that promotes health equity.
AI/ML Medical Device Software Action Plan

- Holistic, patient-centered approach to AI/ML-enabled devices
- Invitation for collaboration with broad set of stakeholders
- Five Aims encompassing regulatory policy and science

Five Aims:

- Update the proposed AI/ML framework
- Strengthen FDA’s role in harmonizing GMLP
- Foster a patient-centered approach
- Support development of regulatory science methods
- Advance real-world performance pilots
Patient-Centered Approach Incorporating Transparency to Users

AI/ML-enabled devices have unique considerations that necessitate a proactive patient-centered approach:

- that takes into account issues including usability, equity, trust, and accountability
- that promotes transparency to all users and to patients more broadly

Patient Engagement Advisory Committee (PEAC) Meeting held Oct 2020

Workshop on Transparency of AI/ML-enabled devices held Oct 2021
Regulatory Science Supporting FDA’s Strategic Priority to Promote Health Equity

- Collaborate on regulatory science efforts to develop evaluation methods for AI/ML-enabled medical software, especially related to algorithm bias and robustness.
- Ensure important performance considerations – including with respect to race, ethnicity, disease severity, gender, age, and geographical considerations – are addressed throughout the total product lifecycle.
- Facilitate more rapid and continuous improvement of AI/ML-enabled device performance across diverse populations.
- Ongoing research being conducted in collaboration with Centers for Excellence in Regulatory Science and Innovation (CERSIs) and within FDA’s OSEL.
- In Collaborative Communities we work together to achieve common objectives on medical device challenges and promote health equity.
We provide stakeholders ongoing clarity through guidance, including on:

- Collection of Race and Ethnicity Data in Clinical Trials
- Evaluation of Sex-Specific Data in Medical Device Clinical Studies
- Patient Engagement in the Design and Conduct of Medical Device Clinical Studies
- Evaluation and Reporting of Age-, Race-, and Ethnicity-Specific Data in Medical Device Clinical Studies
- Clinical Decision Support Software
- Software as a Medical Device (SAMD): Clinical Evaluation
Internationally Harmonized GMLP Guiding Principles

• **Good Machine Learning Practice (GMLP):** accepted practices in AI/ML algorithm design, development, training, and testing that can facilitate the quality development and assessment of AI/ML-enabled technologies

• Ten guiding principles issued by US FDA, MHRA (UK) and Health Canada to promote global harmonization in efforts for the identification of best practices and the creation of standards

• Intended to help inform the development of GMLP and encourage broad stakeholder engagement

<table>
<thead>
<tr>
<th>Good Machine Learning Practice for Medical Device Development: Guiding Principles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-Disciplinary Expertise are Leveraged Throughout the Total Product Life Cycle</td>
</tr>
<tr>
<td>Clinical Study Participants and Data Sets are Representative of the Intended Population</td>
</tr>
<tr>
<td>Selected Reference Datasets are Based Upon Best Available Methods</td>
</tr>
<tr>
<td>Focus is Placed on the Performance of the Human-AI Team</td>
</tr>
<tr>
<td>Users are Provided Clear, Essential Information</td>
</tr>
</tbody>
</table>
Further Questions or Feedback:

www.fda.gov/digitalhealth

DigitalHealth@fda.hhs.gov

Matthew Diamond, MD, PhD
Chief Medical Officer
CDRH Digital Health Center of Excellence
Office of Strategic Partnerships & Technology Innovation (OST)
Center for Devices and Radiological Health (CDRH), U.S. Food and Drug Administration
matthew.diamond@fda.hhs.gov
(301) 332-5126
Digital Equity - A Global Perspective

Dr. Bilal A Mateen, MBBS, MPH
Wellcome Trust

Racial Bias and Healthcare Algorithms
March 3, 2023
12:59 – 1:07 p.m. ET
Characteristics of trustworthy data science

Technical, institutional and social characteristics need to be considered at every stage of a development process.

**Technical**

The properties and effectiveness of data science and digital tools

- Tools must be effective in solving the problems they were created to address, in the intended context or environment
- Software should be easy to maintain, open by default, and can evolve to meet changing needs

**Institutional**

The policy and regulatory environment

- Policy needs to facilitate responsible and sustainable innovation
- Data standards should support efficient data exchange and analysis, including across borders
- Governance must balance the need for adequate control over data, with access for science and research

**Social**

The role of communities that develop, use and are affected by data science and digital tools

- Datasets and digital tools need to be designed for – and with – diverse populations, so they don’t only work for some groups at the expense of others
- People should have sufficient levels of visibility and control over data about their health
- Research cultures need to value people working on data science and research software
A Structural Problem Starting with the Data we Use
Requiring Regulatory, Academic & Private Sector Collaboration

6.3.1 Diversity and inclusion in development

6.3.1.1 The supplier shall document how they ensure they and their suppliers accommodate the diversity of users impacted by the product. If this is not possible, the supplier shall provide a justification as to why.

NOTE 1: Assessment of potential biased processes and impacts could be limited when development teams are homogeneous, i.e., non-diverse. It is assumed that the diversity of the team reflects current good practice for inclusion criteria.

6.3.1.2 Suppliers shall undertake and document a process to ensure that the users identified and engaged in 5.1 are representative of the intended users

NOTE 2: Within the intended groups of users stakeholders should be diverse (e.g., across ethnicity, gender, socioeconomic status, age, and geography). For example, for a product which treats conditions found in older populations, engagement might not require stakeholders with diverse ages. It is still necessary to engage stakeholders across other social categories (e.g., different genders and ethnicities) within older populations to meet the requirements for b). Suppliers might need to consider and account for the range of digital skills and engagement among different groups of people when seeking a wide range of participants.

6.3.1.3 Suppliers shall document both its justification for the chosen approach and any specific framework of methodology followed and the outcomes.

6.3.2 Assessment of model bias risks

6.3.2.1 Developers shall undertake a risk assessment to identify risks of bias in model development and deployment that might result in inequitable outcomes.

NOTE: Possible approaches to complete the algorithmic risk and impact assessment include:

a) an extension of standard project risk management processes;
b) Assessment List for Trustworthy AI [12] (specifies a Human Rights Impact Assessment);
c) the Ada Lovelace Algorithmic Impact Assessment (AIA) [13] or the Black Box report [14] (which divides the audit for bias and assessment of impact);
d) the Canadian government AIA [15] (outlines an additional approach for a quantitative method); and
e) the NHS England Equality and Health Inequality Assessment (EHIA) form [16].

6.3.2.2 The risk assessment (see 6.3.2.1) shall document the degree of potential impact and harm for sub-groups and outline mitigations.

6.3.2.3 The risk assessment shall include:

a) explanations for the mechanism referred to in 8.1.4 for monitoring of real-world impact on sub-groups post-deployment; and
b) details of when the assessment is to be reviewed, e.g., at key stages of model maintenance, new use-cases, updates, and decommissioning.

Correspondence

https://doi.org/10.3386/1599-022-0567-w

Tackling bias in AI health datasets through the STANDING Together initiative

T o the Editor—As of June 2022, a wide range of artificial intelligence (AI as a Medical Device (MDSmM)) has received regulatory approval internationally, with at least 340 devices cleared by the US Food and Drug Administration (FDA). Despite the enormous potential of AI, there’s rapid growth in healthcare has been accompanied by concerns that AI models may learn biases, expressed in medical practice and patient outcomes. This has been exemplified several AI systems that have shown the ability of algorithms to systematically misrepresent and exacerbate health problems in minority groups. This raises concerns that, without appropriate safeguards, AI models prioritize sample size. There are concerns that many health datasets do not adequately represent minority groups; however, the extent of this problem is unknown. Many datasets do not provide demographic information, such as an ethnicity and race. Publicly available datasets for skin cancer and rare imaging have shown inconsistent and incomplete demographic reporting, and are disproportionately created from smaller number of high-income countries. For skin cancer datasets, the reporting of key demographic information, such as ethnicity and skin tone, even when clinically relevant, was only present in 25% of datasets.

Under-representation in datasets can affect the fairness of AI systems by two observations and labels were constructed. These concerns have motivated calls for better documentation practices and the creation of benchmarks for fairness for datasets and Benchmarks. The aforementioned problems are becoming increasingly recognized by regulators of medical devices, December 2021. The US FDA, Health Canada and the UK Medicines and Healthcare products Regulatory Agency (MHRA) jointly published 10 guiding principles for good machine-learning practice. This specifically states that data should be representative of the intended population to order “manage any bias, procure appropriate and generalizable performance across the intended population, assessed usability.

Equity in medical devices independent review: terms of reference

April 2022

Purpose

The purpose of the review is to establish the extent and impact of potential ethnic and other unfair biases in the design and use of medical devices and to make recommendations for more equitable solutions.
And Careful Consideration of Technical Solutions

Figure 1. **Pareto curve.** We depict the typical trade-off assumed by most fairness studies in computer vision. By adjusting fairness, the accuracies of a classifier on its best- and worst-performing groups form a *Pareto curve* (dotted gray line). Points A and B are maximally efficient configurations because they lie on the curve; B is fairer since it yields a lower accuracy difference between the groups (see bar plot on right). Point C is as fair as B, but is inefficient because it reduces the accuracies of both groups. In our experiments, we find that accuracy-based fairness constraints applied to deep neural networks tend to achieve inefficient configurations like C.

<table>
<thead>
<tr>
<th>Privacy Level</th>
<th>Average White Influence</th>
<th>Average Black Influence</th>
<th>Most Helpful Ethnicity</th>
<th>Most Harmful Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.29 ± 2.40</td>
<td>0.71 ± 1.40</td>
<td>White</td>
<td>White</td>
</tr>
<tr>
<td>Low</td>
<td>−0.22 ± 0.70</td>
<td>−0.03 ± 0.17</td>
<td>White</td>
<td>White</td>
</tr>
<tr>
<td>High</td>
<td>−0.11 ± 3.94</td>
<td>0.03 ± 1.35</td>
<td>White</td>
<td>White</td>
</tr>
</tbody>
</table>

**Table 4:** Group influence summary statistics across all privacy levels for white (majority) and Black (minority) training patients on both white and Black test patients in MIMIC-III. Privacy changes the most helpful group from Black patients to the majority white patients and minimizes their helpful influence. This needs careful consideration as the use of ethnicity is still being investigated in medical practice.
Impact of Healthcare Algorithms on Racial and Ethnic Disparities in Health and Healthcare

Consensus Panel Discussion/Q & A

March 3, 2023
1:07– 1:40 p.m. ET
Discussion Questions

• What’s missing: gaps in experience and insights related to algorithmic standards/stewardship, including pitfalls?

• Guidance needed: which standards are needed and possible and by whom? How to develop standards and gain adoption?
Impact of Healthcare Algorithms on Racial and Ethnic Disparities in Health and Healthcare

Closing Remarks

Christine Chang, MD
Associate Director, EPC Division, AHRQ
Racial Bias and Healthcare Algorithms
March 3, 2023
1:40-1:45 p.m. ET
What’s Next?

• Panel deliberations
• Webinar presenting panel recommendations: May 15, 2023
• For inquiries and to be added to the distribution list
  AIDisparities@norc.org
Thank you!