

AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

AHRQ

Supporting Safe Use of Artificial Intelligence in Healthcare

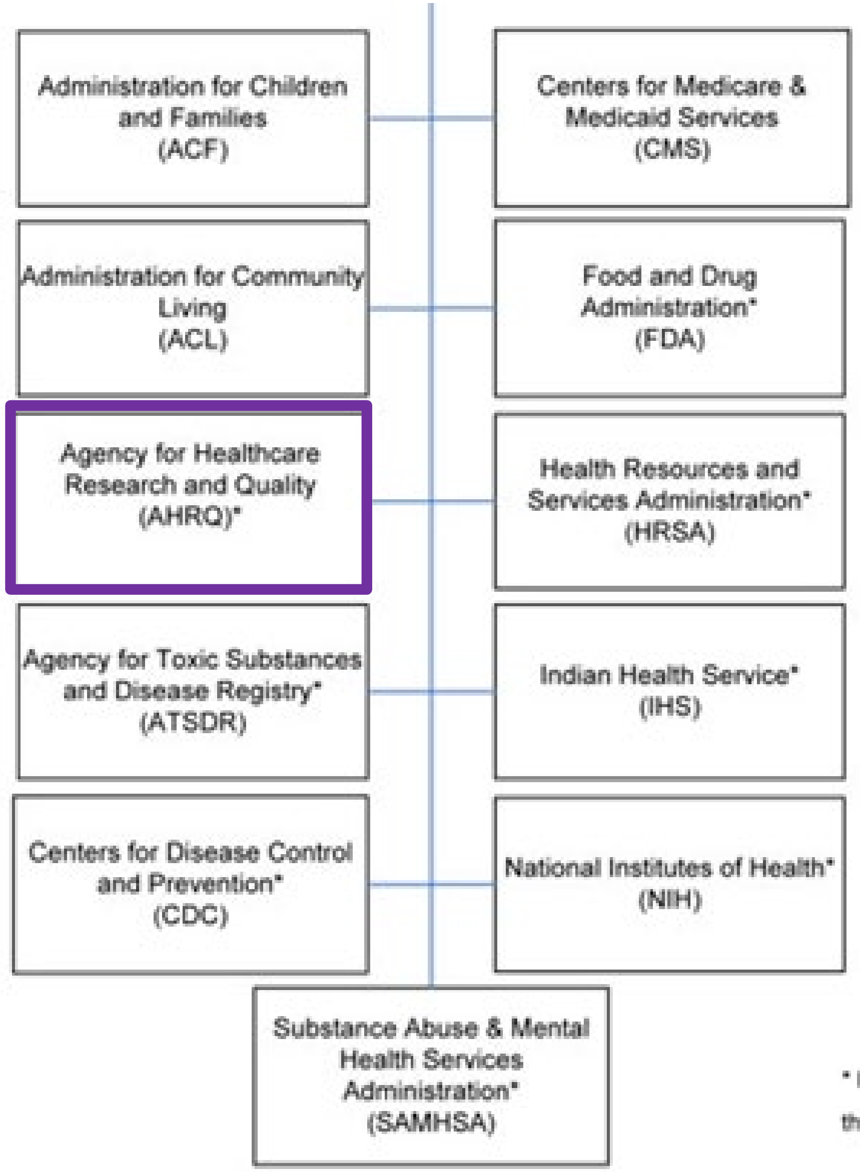
Craig Umscheid, MD, MS
Director, Center for Quality Improvement and Patient Safety
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services

DCHA
October 30, 2024

Objectives

- Describe AHRQ and AHRQ programs, and how we differ from other agencies in the Department of Health and Human Services
- Describe potential risks of using AI in healthcare delivery
- Describe select federal activities designed to mitigate safety threats associated with AI in healthcare

Eleven Divisions in Department of HHS



* Designates a component of the U.S. Public Health Service.

AHRQ in Context: Research for Healthcare Delivery

Agency for Healthcare Research and Quality (AHRQ) Mission Statement

www.ahrq.gov

- To produce evidence to make healthcare safer, higher quality, more accessible, equitable and affordable

- To work with partners to make sure that evidence is understood and used

Three Centers at AHRQ

Agency for Healthcare Research and Quality

Center for Quality Improvement and Patient Safety (CQuIPS)

Center for Evidence and Practice Improvement (CEPI)

Center for Financing, Access and Cost Trends (CFACT)

Five Aims of the National Action Alliance for Patient and Workforce Safety

Safety Self-Assessment offered by 2020 National Action Plan for Safety

- **Culture, Leadership, Governance**
  - Regular safety culture surveys
- **Workforce Safety**
  - Explicit worker safety strategy
- **Patient and Family Engagement**
  - Actively engaged PFACs
- **Learning Healthcare Systems**
  - Defined safety competencies for all workers

https://www.ihi.org/initiatives/national-steering-committee-patient-safety/national-action-plan-advance-patient-safety

Engaging the National Action Alliance Engine to Power “Safe Care Everywhere, Zero Preventable Harm for All”

By 2026, Reduce Harms by 50% from Peak of Pandemic

Outcomes | Outcomes | Outcomes

Learning Health Organizations

- Best Practice Implementation
- Learning from Network
- Research Support
- Aligned with VBP

Opportunity | Opportunity | Opportunity

National Action Alliance Engine

- National Webinars
- Implementation Initiatives
- Tools
- Funding Opportunities
- Measures and Safety Dashboard

Identify needs through baseline safety self-assessments

Examples of Tools, Funding Opportunities, and Implementation Initiatives from AHRQ

Culture, Leadership, and Governance	Patient and Family Engagement	Workforce Safety	Learning System
• **Surveys on Patient Safety Culture** (tool and pilot)	• **TeamSTEPPS 3.0** (tool & training)	• **New AHRQ grant:** Systems-Based Approaches to Improve Patient Safety by Improving Healthcare Worker Safety and Well-Being (up to $2M in funding)	• **Diagnostic Safety Measurement** (implementation initiative)

Practice Improvement Opportunity: Implementing Diagnostic Safety Tools

Work with AHRQ to pilot the below resources at your healthcare organization:

- Calibrate Dx: A Resource to Improve Diagnostic Decisions
  - ► Helps clinicians self-assess their diagnostic performance and use feedback to better calibrate their diagnostic performance
- Measure Dx: A Resource to Identify, Analyze, and Learn from Diagnostic Safety Events
  - ► Provides step-by-step instructions for developing, implementing, and sustaining diagnostic safety measurement strategies
- Toolkit for Engaging Patients To Improve Diagnostic Safety
  - ► Features strategies to address communication errors during the clinician-patient encounter that can lead to diagnostic errors

Email: IDEASproject@rand.org

Visit: https://www.ahrq.gov/diagnostic-safety/ideas-project/index.html

Proposed CMS Patient Safety Structural Measure (PSSM)

PSSM Domain	Key PSSM Specifications
Domain 1: Leadership Commitment	• C-suite oversees safety self-assessment and resulting plan and metrics
Domain 2: Strategic Planning	• Strategic plan publicly shares hospital commitment to "zero preventable harm" • Hospital has action plan for workforce safety • Hospital requires implementation of a patient safety competencies for all staff
Domain 3: Culture of Safety	• Hospital conducts hospital-wide culture of safety survey • Hospital implements team communication training • Hospital participates in large-scale learning network(s) for patient safety
Domain 4: Transparency	• Hospital has a communication and resolution program, such as AHRQ's CANDOR toolkit
Domain 5: Patient Engagement	• Hospital has representative Patient and Family Advisory Council (PFAC) that provides input on safety-related activities • Patients have comprehensive access to their own medical records via patient portals

Case: Algorithm to Help Inform Management of “High Risk” Patients

- Examined differences in health status between white and black patient populations considered for a “high-risk” care management program
- Determinant of “high-risk” was commercial algorithm that’s been applied to 100s of millions of individuals in US
- Algorithm trained to predict high healthcare costs, using that as a proxy for a sicker patient population

RESEARCH ARTICLE

ECONOMICS

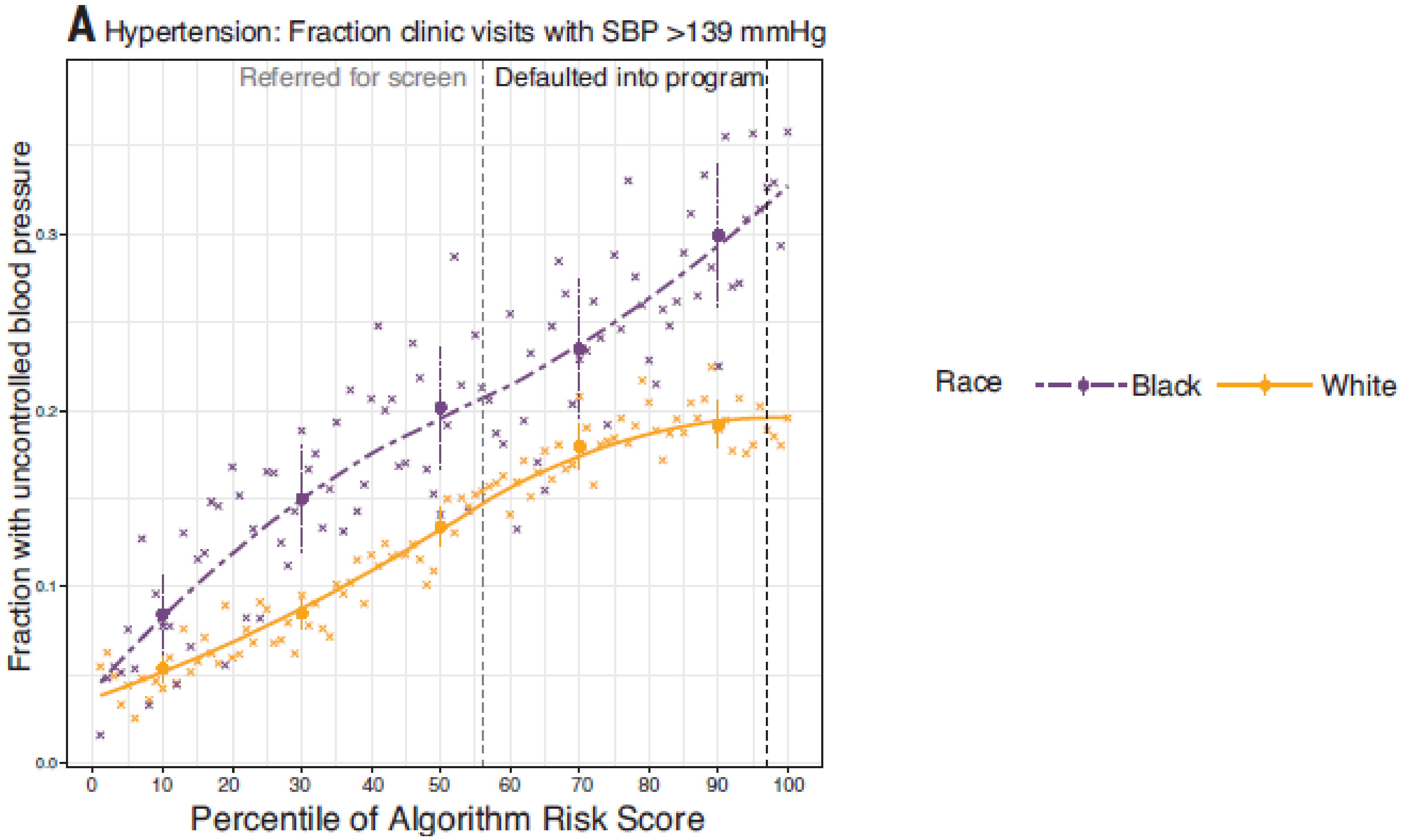
Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer[1,2]*, Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5]*†

Health systems rely on commercial prediction algorithms to identify and help patients with complex health needs. We show that a widely used algorithm, typical of this industry-wide approach and affecting millions of patients, exhibits significant racial bias: At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses. Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%. The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients. Thus, despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise. We suggest that the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.

Obermeyer, et al. *Science*. 2019; 366: 447–453.

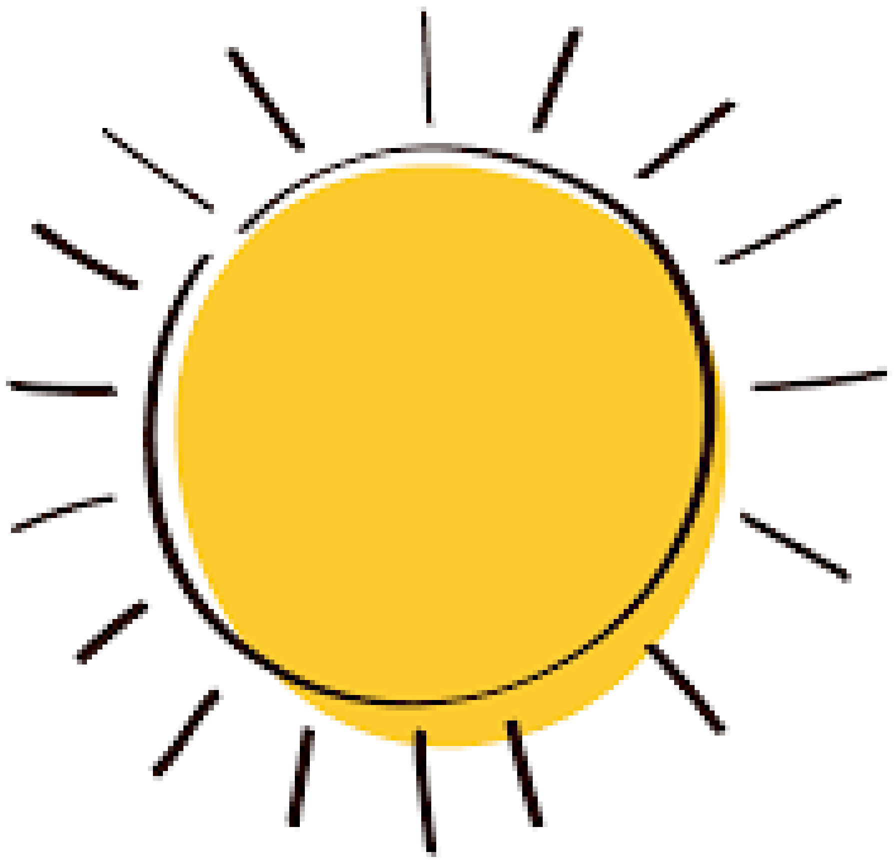
Proportion with Uncontrolled BP vs. Algorithm Predicted Risk, by Race



Obermeyer, et al. *Science*. 2019; 366: 447–453.

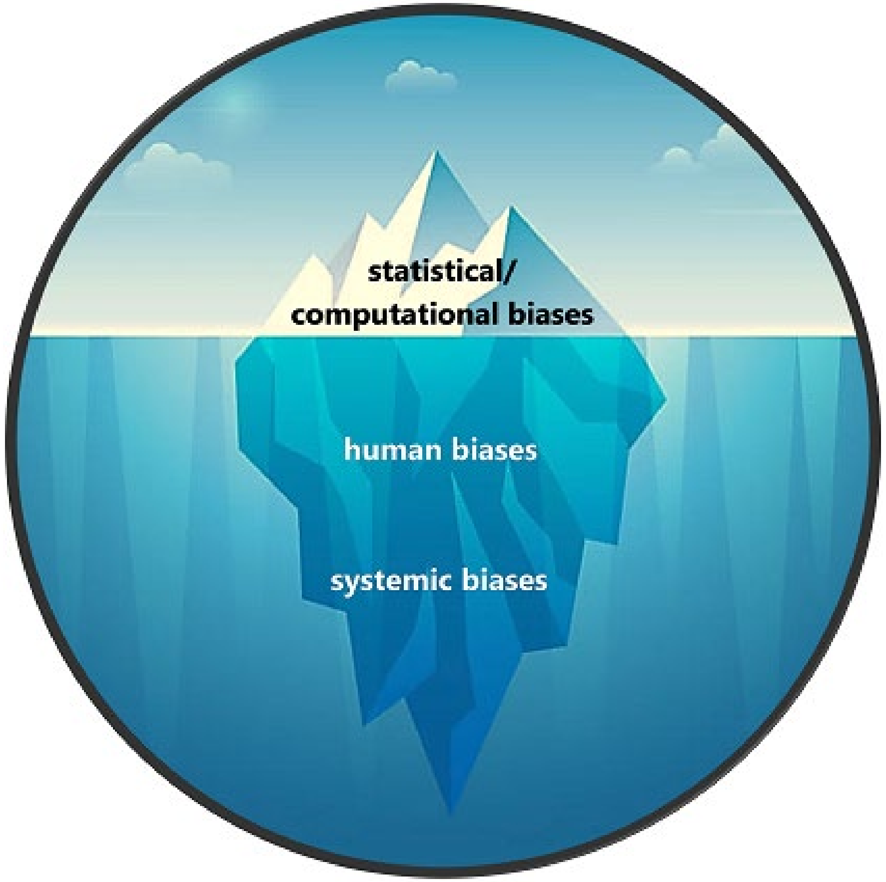
Examples of Algorithm Promise

- Making highly accurate predictions and diagnoses
- Summarizing clinical notes
- Chatbots to address patient information needs



Examples of Algorithm Risks

- Bias (ie. systematic error)
- False negatives
- False positives
- Poor generalizability (eg. overfitting)
- Poor clinical decision making
- Hallucinations (3-27% of responses)
- Privacy and confidentiality
  - ► data breach
  - ► re-identification from deidentified data



NIST Special Publication 1270

Congressional Request to AHRQ to Examine the Impact of Algorithms on Racial Bias in Healthcare

- **Congressional Outreach to AHRQ**
  - ► September 2020
  - ► Senators Warren (MA), Booker (NJ), Wyden (OR), and Representative Lee (CA)

- **AHRQ Request for Information (RFI) to Inform Planning for Evidence Review**
  - ► Posted in Federal Register March 2021
  - ► Included 11 questions to gauge awareness of algorithm use in healthcare and their potential for introducing racial bias in clinical decision making, as well as potential approaches and existing standards to mitigate bias


FEDERAL REGISTER
The Daily Journal of the United States Government
NATIONAL ARCHIVES
Notice

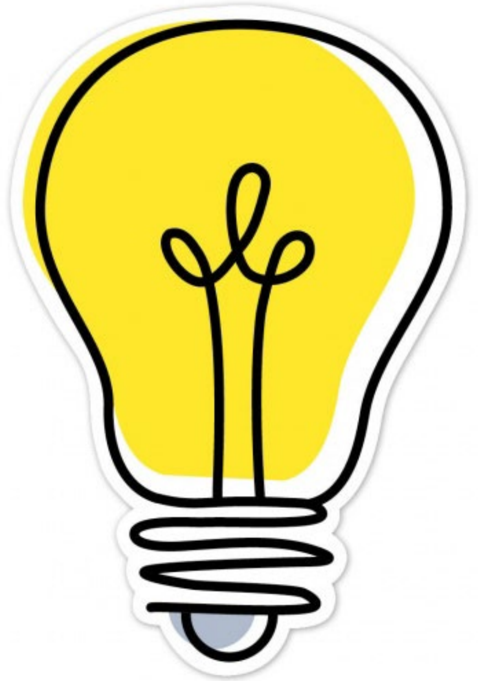
Request for Information on the Use of Clinical Algorithms That Have the Potential To Introduce Racial/Ethnic Bias Into Healthcare Delivery

AHRQ's Request For Information (RFI) to Public

- 42 responses totaling 485 pages
- Respondents included:
  - 16 clinical and professional societies
  - 9 healthcare technology organizations
  - 7 academic organizations
  - 4 federal and state agencies
  - 1 payer organization
  - 5 private citizens

Example Insights from the RFI

- Bias and disparities can result from algorithms whether or not they explicitly include race
- Great heterogeneity and lack of standardization in how race and social determinants of health are collected and defined
- Algorithms often developed using data from populations not representative of those to whom algorithm applied
- Clinicians and patients may often be unaware of algorithm use and potential for bias

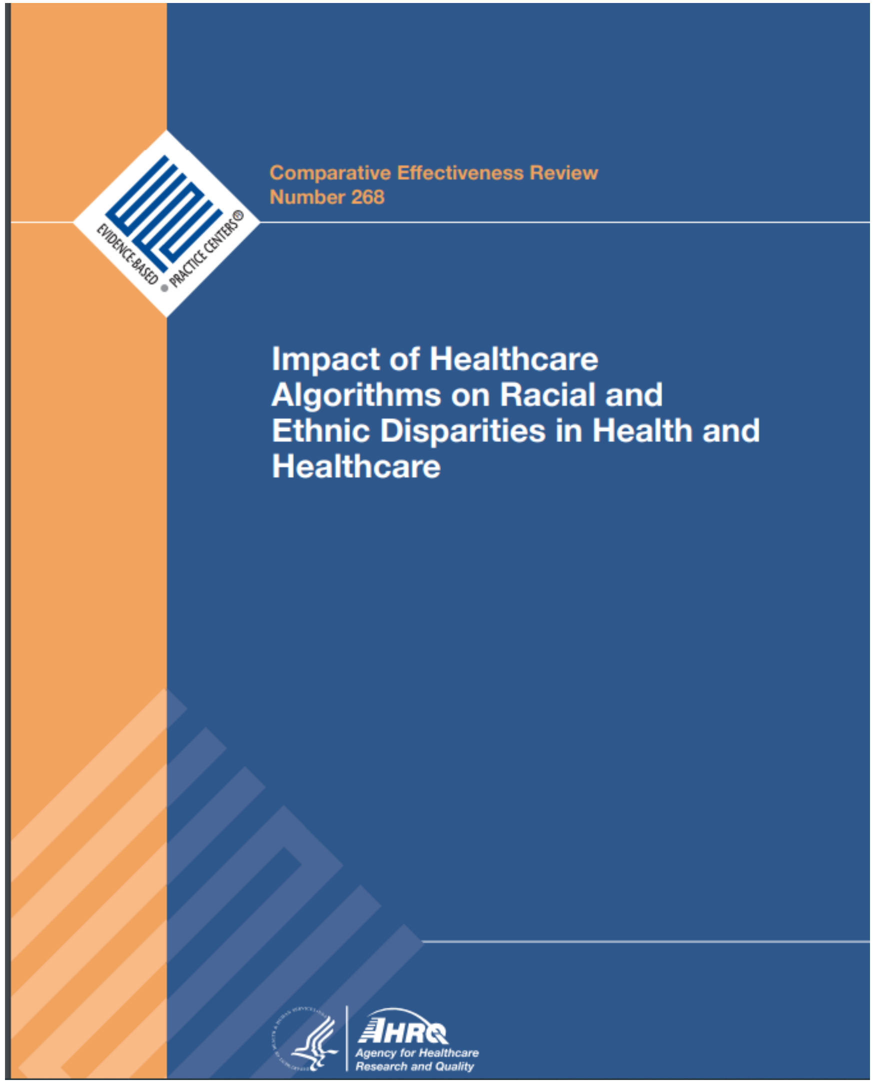


JAMA Health Forum. 2023;4(6):e231197. doi:10.1001/jamahealthforum.2023.1197

Evidence Review: Impact of Algorithms on Racial Disparities in Healthcare

- **Key Question 1:** Effect of algorithms on racial differences in health and healthcare?

- **Key Question 2:** Effect of approaches to mitigate racial bias resulting from healthcare algorithms?


Comparative Effectiveness Review
Number 268

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

https://effectivehealthcare.ahrq.gov/products/racial-disparities-health-healthcare/research

Key Question 1: Effect of algorithms on racial differences in health and healthcare?

- The effect of algorithms is complex, and some have been shown to exacerbate disparities, some reduce disparities, and others have no effect

- An algorithm may exacerbate disparities for one outcome, but reduce disparities for another outcome

- Many algorithms in clinical use exacerbate racial disparities (e.g. eGFR)

- Disparities can be reduced when they are identified and used to inform algorithm development (e.g. prostate CA screening)

https://effectivehealthcare.ahrq.gov/products/racial-disparities-health-healthcare/research

Key Question 2: Effect of approaches to mitigate racial bias resulting from healthcare algorithms?

- Mitigation strategies can improve algorithmic accuracy, but may not address biases in how algorithms are implemented in practice
- Mitigation effectiveness is highly context-specific and may depend on algorithm, clinical population, setting, outcomes examined
- Further research is needed to quantify the real-world effects of modifying algorithms on healthcare inequities

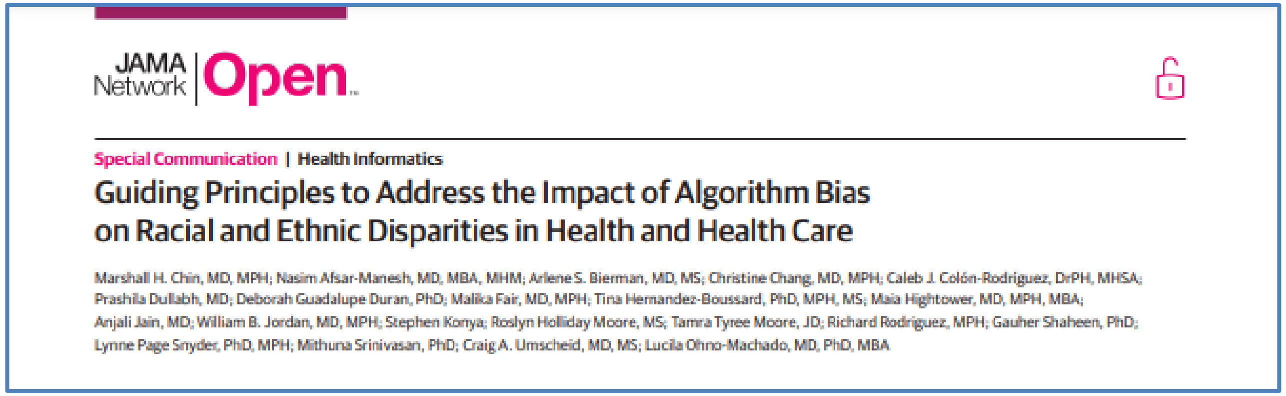
https://effectivehealthcare.ahrq.gov/products/racial-disparities-health-healthcare/research

Developing Guiding Principles to Address the Impact of Algorithms on Racial Disparities in Healthcare

- **March 2023:** AHRQ and National Institute on Minority Health and Health Disparities (NIMHD) host public meeting to share draft evidence review and receive additional perspectives from key stakeholders to inform expert panel who drafted guiding principles

- **May 2023**: Expert panel shared draft of guiding principles in public meeting, and received and responded to public comments

- **December 2023**: Final guiding principles published



JAMA Network | Open

Special Communication | Health Informatics

Guiding Principles to Address the Impact of Algorithm Bias on Racial and Ethnic Disparities in Health and Health Care

Marshall H. Chin, MD, MPH; Nasim Afsar-Manesh, MD, MBA, MHM; Arlene S. Bierman, MD, MS; Christine Chang, MD, MPH; Caleb J. Colón-Rodriguez, DrPH, MHSA; Prashila Dullabh, MD; Deborah Guadalupe Duran, PhD; Malika Fair, MD, MPH; Tina Hernandez-Boussard, PhD, MPH, MS; Maia Hightower, MD, MPH, MBA; Anjali Jain, MD; William B. Jordan, MD, MPH; Stephen Konya; Roslyn Holliday Moore, MS; Tamra Tyree Moore, JD; Richard Rodriguez, MPH; Gauher Shaheen, PhD; Lynne Page Snyder, PhD, MPH; Mithuna Srinivasan, PhD; Craig A. Umscheid, MD, MS; Lucila Ohno-Machado, MD, PhD, MBA

Chin MH, Afsar-Manesh N, Bierman AS, et al. JAMA Network Open. 2023;6(12):e2345050. DOI: https://doi.org/10.1001/jamanetworkopen.2023.45050

https://effectivehealthcare.ahrq.gov/products/collections/healthcare-algorithms-meeting-agenda

Expert Panel for Guiding Principles

Expert Panel
- Marshall Chin, MD, MPH, University of Chicago (co-Chair)
- Lucila Ohno-Machado, MD, PhD, MBA, Yale (co-Chair)
- Tina Hernandez-Boussard, PhD, MPH, MS, Stanford
- Nasim Afsar-Manesh, MD, MBA, MHM, Oracle Health
- Malika Fair MD, MPH, AAMC
- William Jordan, MD, MPH, AMA
- Gauher Shaheen, PhD, Elevance Health
- Tamra Tyree Moore, JD, Prudential Financial
- Maia Hightower, MD, MPH, MBA, Equality AI

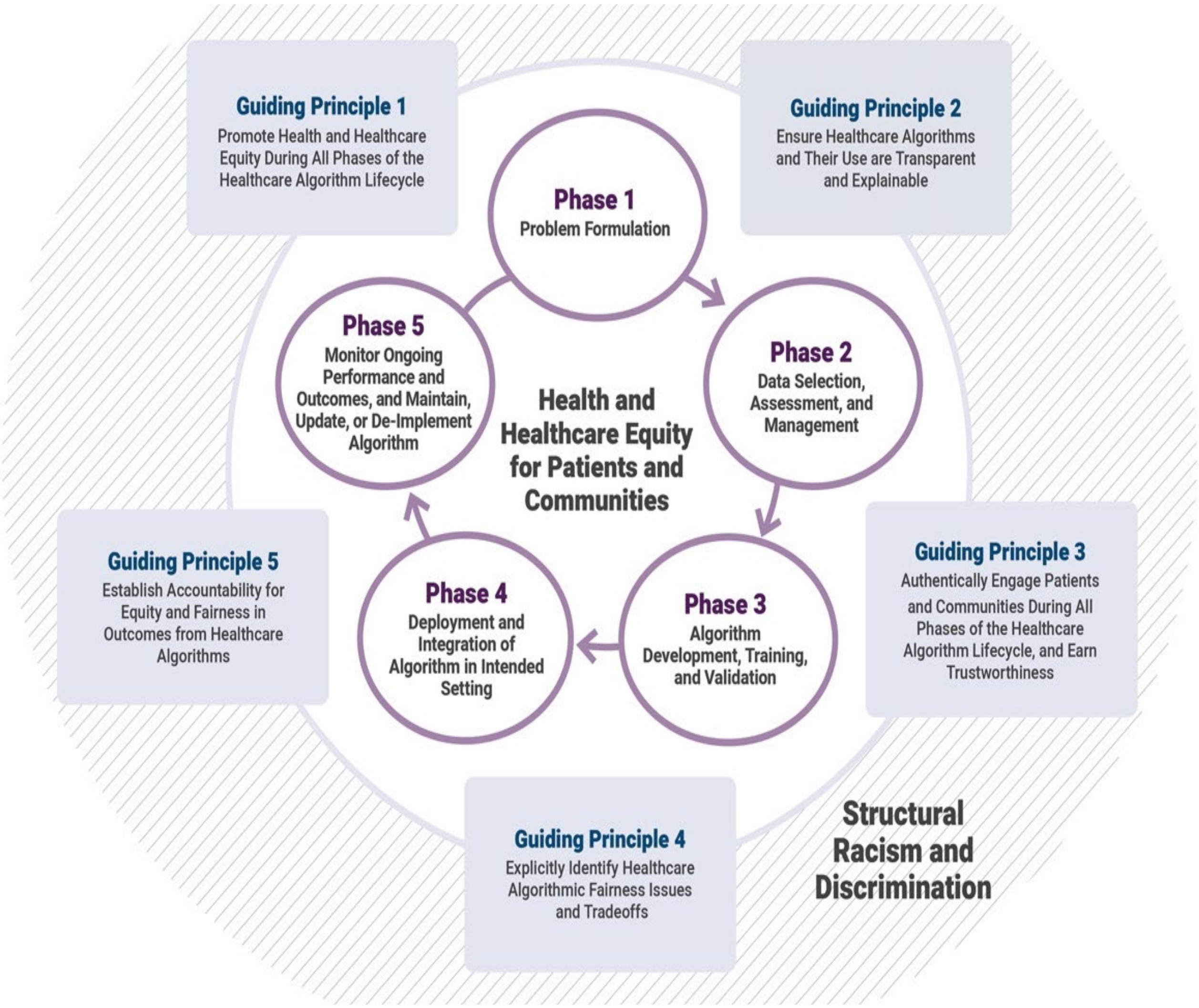
Federal Panel
- Craig Umscheid, MD, MS, AHRQ
- Anjali Jain, MD, AHRQ
- Christine Chang, MD, MPH, AHRQ
- Arlene Bierman, MD, MS, AHRQ
- Deborah Guadalupe Duran, PhD, NIMHD/NIH
- Stephen Konya, ONC
- Caleb Colón-Rodríguez, DrPH, MHSA, Office of Minority Health, HHS
- Roslyn Moore, MS, Office of Minority Health, HHS

NORC (contracted support)
- Prashila Dullabh, MD
- Mithuna Srinivasan, PhD
- Richard Rodriguez, MPH
- Lynne Page Snyder, PhD, MPH

Guiding Principles

Five principles should guide efforts:

1) promote healthcare equity in all phases of the algorithm lifecycle
2) ensure healthcare algorithms and their use are transparent and explainable
3) engage the community in authentic partnerships
4) explicitly explore tradeoffs
5) establish accountability for equity and fairness in outcomes



Chin MH et al. *JAMA Network Open*. 2023.

Built upon the NAM algorithm lifecycle framework adapted by Roski

Operationalizing Guiding Principles to Address Bias Across an Algorithm's Lifecycle (I)

Guiding Principle	Considerations for Operationalizing Guiding Principle
1. Promote Health and Healthcare Equity During All Phases of the Healthcare Algorithm Lifecycle	▪ Researchers and research sponsors (e.g., funders, scientific journals) should routinely assess the impact of healthcare algorithms on health equity. ▪ Validate algorithms for the specific purpose for which they are being deployed and across populations. Evaluate training datasets for representativeness of specific populations. Document any lack of representativeness. If appropriate, take mitigation measures before training the algorithm. ▪ Continually monitor algorithm performance for equitable impact across populations.
2. Ensure Healthcare Algorithms and their Use are Transparent and Explainable	▪ Develop profiles of algorithm training data with the distributions of key features of the population (e.g., race, gender, socioeconomic status, age), and make the distributions available for independent review. ▪ Design regulations to ensure transparency, explainability, and interpretability. For example, require algorithm information labels to clearly communicate design features and the intent of the algorithm to stakeholders. Enough information should be provided to assess validity and bias. ▪ Develop reporting guidelines for publications examining algorithms that are explicit about bias, similar to proposals to address equity in observational studies and randomized controlled trials. Explain algorithm biases and mitigation measures to the stakeholder community. ▪ Make information available for patients and communities when an algorithm is used in their care, what aspects of their personal data were used in the algorithm, what impact the algorithm had on their care such as diagnosis, prognosis, or treatment, and how the algorithm performs for their sociodemographic group.
3. Authentically Engage Patients and Communities During All Phases of the Healthcare Algorithm Lifecycle, and Earn Trustworthiness	▪ Engage patients and communities in decisions about those problems best addressed by algorithm solutions. ▪ Algorithm development teams should include a diverse group of people who are involved in decision-making. ▪ Put safeguards in place to protect patient autonomy and privacy in healthcare algorithm development, deployment, and monitoring. ▪ Speak to those most impacted by algorithmic bias to acknowledge potential or demonstrated harms and agree on methods of redress and closure.

Chin MH et al. *JAMA Network Open*. 2023.

Model Fact Sheet

Model Facts

Model name: Deep Sepsis

Locale: Duke University Hospi

Approval Date: 09/22/2019

Last Update: 01/13/2020

Version: 1.0

Summary

This model uses EHR input data collected from a patient's current inpatient encounter to estimate the probability that the patient will meet sepsis criteria within the next 4 hours. It was developed in 2016-2019 by the Duke Institute for Health Innovation. The model was licensed to Cohere Med in July 2019.

Mechanism

- **Outcome** ....sepsis within the next 4 hours, see outcome definition in "Other Information"
- **Output** ....0% - 100% probability of sepsis occurring in the next 4 hours
- **Target population** ....all adult patients >18 y.o. presenting to DUH ED
- **Time of prediction** ....every hour of a patient's encounter
- **Input data source**........electronic health record (EHR)
- **Input data type** ....demographics, analytes, vitals, medication administrations
- **Training data location and time-period** ....DUH, diagnostic cohort, 10/2014 – 12/2015
- **Model type**........ Recurrent Neural Network

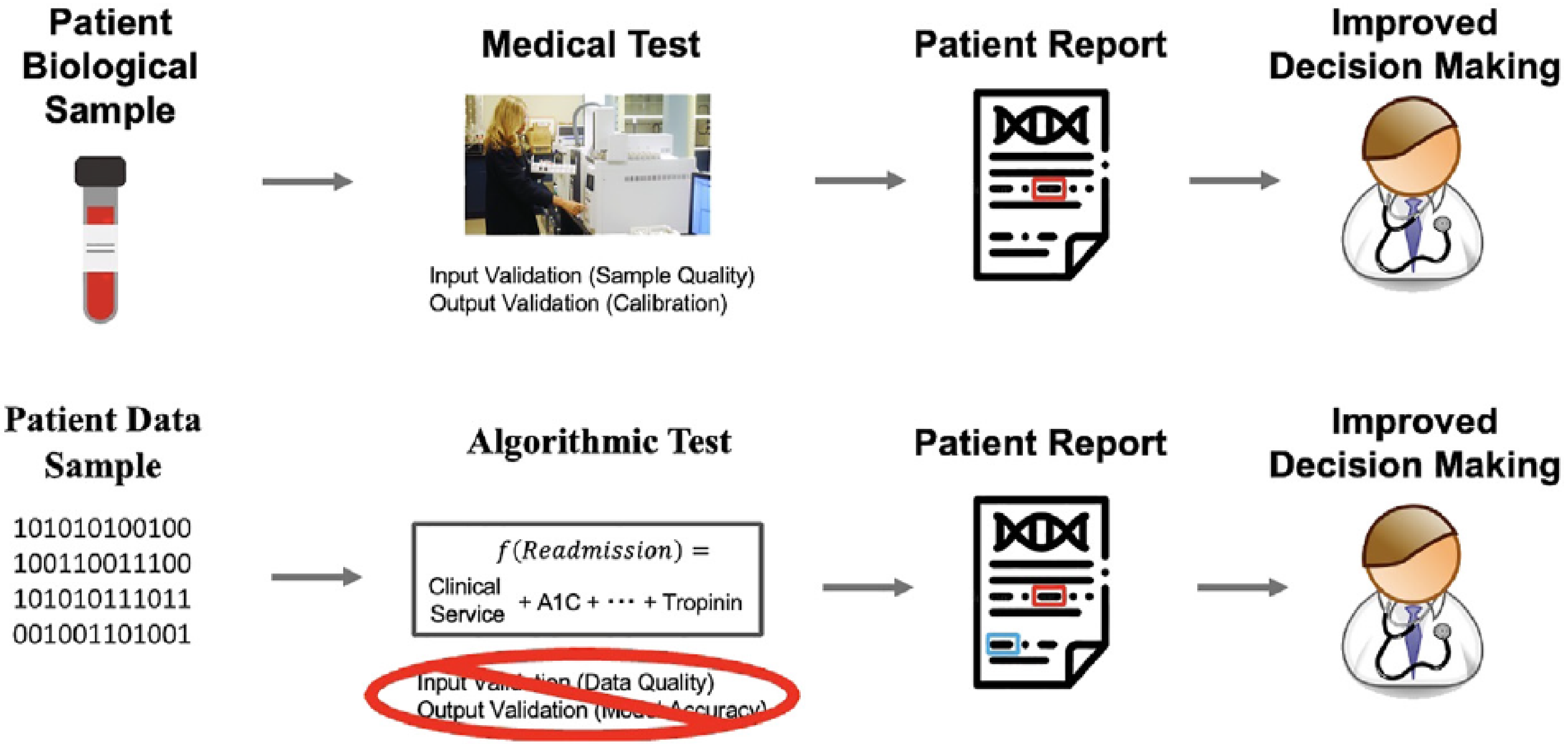
Sendak et al. *NPJ Dig Med.* 2020.

Operationalizing Guiding Principles to Mitigate and Prevent Bias Across an Algorithm's Lifecycle (II)

Guiding Principle	Considerations for Operationalizing Guiding Principle
4. Explicitly Identify Healthcare Algorithmic Fairness Issues and Tradeoffs	▪ Model performance across patient cohorts should be measured using multiple objective measures (e.g., accuracy measures such as sensitivity, specificity, and area under the receiver operating curve, predictive values, calibration, residuals) that are appropriate for the intended use of the algorithm. ▪ Fairness of the model output across patient cohorts should be measured using metrics such as demographic parity (same proportion of groups assigned to positive or negative class) and equalized odds (groups have same false positive rate and same false negative rate). ▪ Model fairness should be optimized for equity in clinical outcomes or resource allocation using bias mitigation methods (e.g., disparate impact remover, label choice experiments, reweighing) and human judgment.
5. Ensure Accountability for Equity and Fairness in Outcomes from Healthcare Algorithms	▪ Governmental agencies, accreditation organizations, and professional associations should implement regulatory processes, policies, and standards to mandate transparency and regular monitoring and validation of healthcare algorithms for equity and fairness. Incentives for fairness in healthcare algorithms should be created. Equity and fairness checks should be built into each phase of the algorithm life cycle for both technical bias and human bias that lead to inequities in model performance, clinical outcomes, and resource allocation. Unfair algorithms should be deactivated, removed, or discontinued. A structured reporting process could identify signals of emerging problems both locally and nationally and facilitate addressing the problems systematically. ▪ A legal and administrative framework and culture should be created to redress harm caused by algorithms. The framework should encourage quality improvement, collaboration, and transparency, as is recommended in the patient safety field. ▪ Algorithm developers, implementers, and users (including but not limited to healthcare delivery organizations) should adopt policies, procedures, and processes to monitor for equity and fairness at each stage of the algorithm lifecycle. They should implement effective and transparent data collection mechanisms to support monitoring. They should identify clear algorithm stewardship and bias mitigation roles for each involved stakeholder group. ▪ Healthcare delivery organizations and algorithm vendors should invest in infrastructure, governance, and teams with diverse skills and experiences to support equity and fairness in algorithm development and use. ▪ Algorithms should not be deployed before validation on the impacted population. Under-resourced institutions with limited technical capability should be supported in validation. ▪ Journals, funders, and research professional associations should identify standards for ensuring accountability for equity and fairness in outcomes from healthcare algorithms, for the algorithms to be published, funded, and rated as high-quality.

Chin MH et al. *JAMA Network Open*. 2023.

The Importance of Local Governance of AI



Rojas JC, Teran M, Umscheid CA. *Crit Care Clin.* 2023;39(4):769-782.

Predictive Analytics Programs at Large Healthcare Systems in the USA: a National Survey

- Surveyed healthcare leaders at all non-academic healthcare system member sites of The Scottsdale Institute, organization of 60 non-profit healthcare systems committed to sharing best practices, with focus on IT and innovation
- Responses occurred between 4/13/2021 and 5/17/2021
- Response rate was 60% (25/42)
- Majority (16/25, 64%) reported having an individual or team focused on clinical applications of predictive algorithms
- Most programs existed for five or fewer years (11/16, 69%), and implemented six or fewer algorithms (11/16, 69%)
- Only a minority (6/25, 24%) had dedicated budget for predictive analytics

Rojas JC, Rohweder G, Guptill J, Arora VM, Umscheid CA. *J Gen Intern Med.* 2022;37(15):4015-4017.

Current State of Healthcare System Governance

Do you tend to buy or internally build your predictive analytic algorithms?****	
Buy all	2 (8)
Mostly buy	11 (44)
Even between build or buy	3 (12)
Mostly build	8 (32)
Build all	1 (4)
What are the current focus areas of the models deployed in your healthcare system?***	
Sepsis risk/identification	14 (88)
Hospital readmission risk	14 (88)
Inpatient length of stay prediction	10 (67)
Ambulatory no-show prediction	10 (63)
Acute care utilization prediction	6 (38)
Cardiac arrest risk	5 (31)
ICU transfer risk	4 (25)

Rojas JC, Rohweder G, Guptill J, Arora VM, Umscheid CA. *J Gen Intern Med.* 2022;37(15):4015-4017.

Federal Approaches to Mitigate Risks (Facilitate Promise) of AI

- Office of National Coordinator for Health IT (ONC)
- White House
- AHRQ

AHRQ, 5600 Fishers Lane, Rockville, MD

Federal Agencies: ONC

- **ONC EHR Certification Program**
  - ► Existing scope includes ensuring transparency of predictive clinical decision support (CDS)

- **Health Data, Technology, and Interoperability (HTI-1) Proposed Rule:**
  - ► Goes into effect 2025
  - ► Include new requirements for Health IT used to support decision-making based on predictive CDS
  - ► Revised criteria require:
    - − availability of model cards for predictive CDS used in EHRs (to include info on outcome of interest, data used in development, predictive performance)
    - − annual public disclosures by organizations that they have competencies to manage risks of predictive CDS

Micky Tripathi, PhD, MPP

White House Executive Order on AI

https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

In summary, the Secretary of HHS shall:

- Establish an AI Task Force to develop a strategic plan on responsible use of AI in healthcare
- Prioritize grantmaking to support responsible AI development and use
- Establish an AI safety program in partnership with Patient Safety Organizations that:
  - ► creates a common framework to identify clinical errors resulting from AI in healthcare
  - ► analyzes captured data to develop best practices aimed at avoiding these harms
  - ► disseminates best practices to appropriate stakeholders

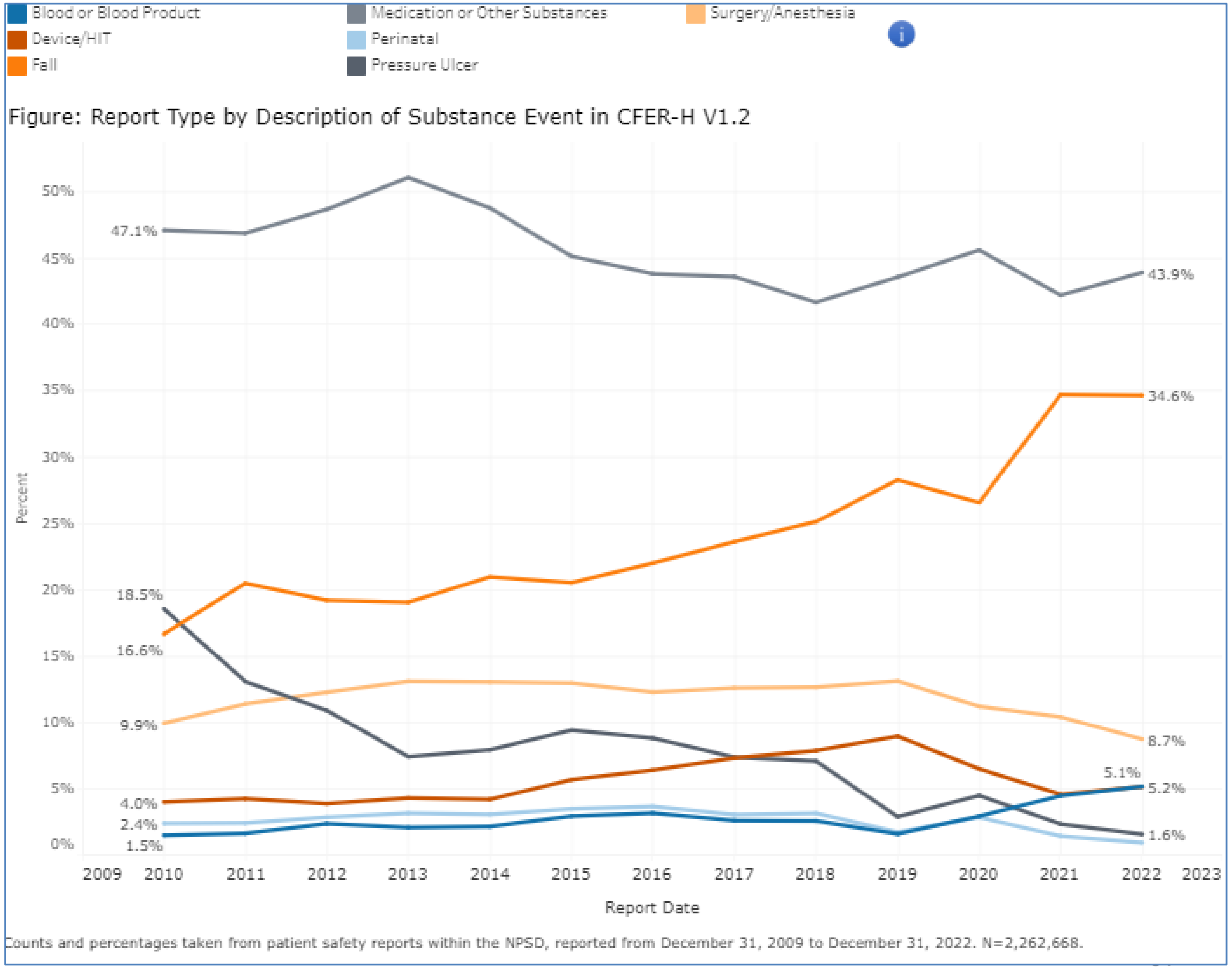
AHRQ's Patient Safety Organization (PSO) Program

- What is a PSO?
  - Collects and analyzes patient safety event data voluntarily reported by healthcare provider organizations
  - Provides feedback to healthcare providers to prevent future patient safety events
  - Working with a PSO makes it possible for information from healthcare providers to receive certain legal protections

- AHRQ's PSO Program:
  - Regulates PSOs (e.g. certify and list PSOs, assess compliance, monitor performance)
  - Develops Common Formats to promote standardized measurement and reporting of patient safety incidents
  - Maintains the Network of Patient Safety Databases (NPSD)

PSO Data Informs Dashboards for Network of Patient Safety Databases (NPSD)

- Most commonly reported safety events are medication-related and falls
- Role for this system to track AI-related patient safety events?



https://www.ahrq.gov/npsd/data/dashboard/generic.html

Funding Opportunities in IT and Safety

Special Emphasis Notice:

- NOT-HS-22-004: AHRQ Announces Interest in Research on Digital Healthcare Safety (nih.gov)

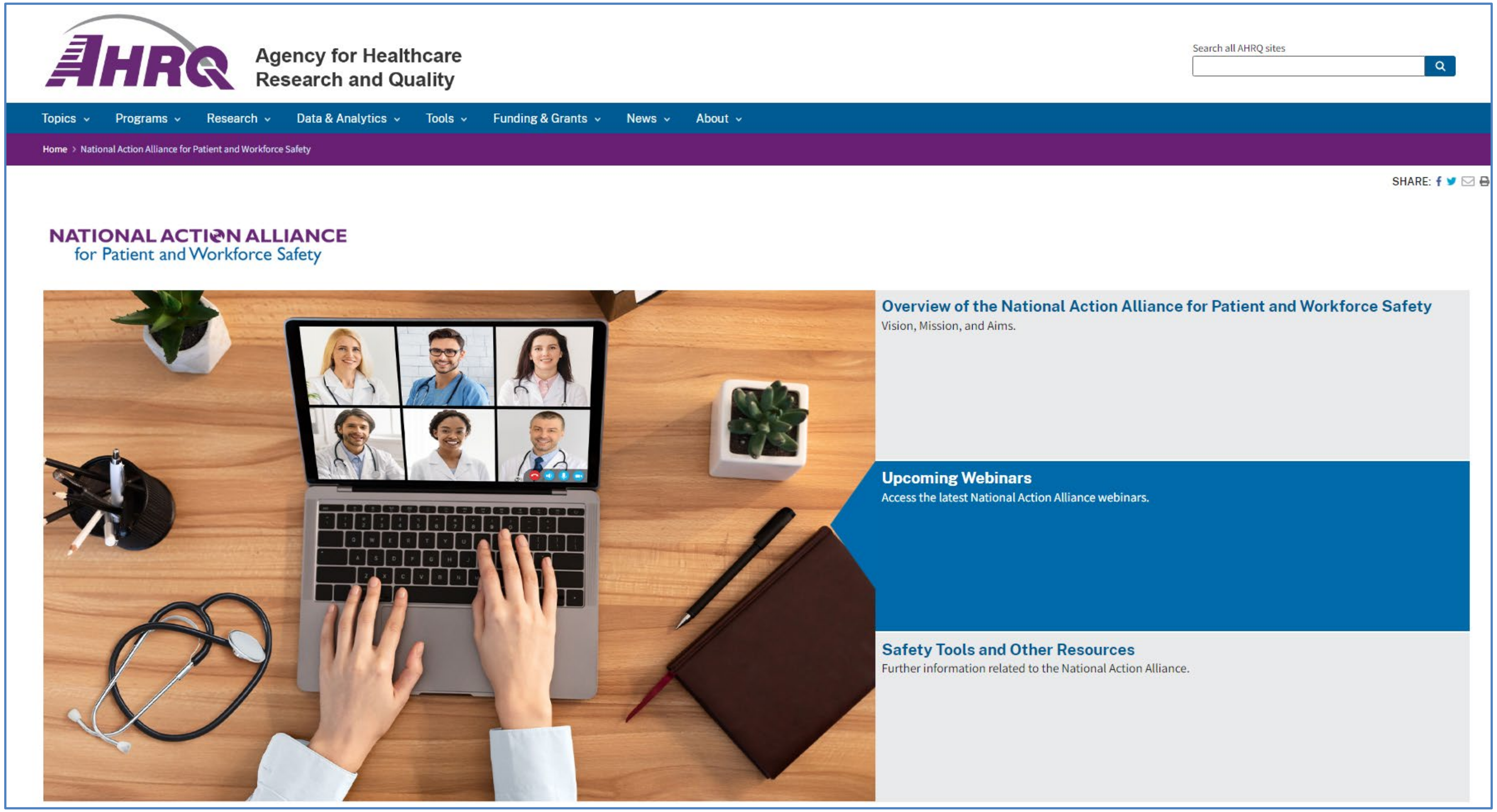
Notice of Funding Opportunity:

- PA-24-261: Examining the Impact of Artificial Intelligence (AI) on Healthcare Safety (R18)
  - Posted July 12, 2024
  - First Application Due Date September 25, 2024
  - Earliest Award Date Summer 2025

Revisiting Objectives

- Describe the mission of AHRQ, and how it differs from other agencies in the Department of Health and Human Services
- Describe potential risks of using AI in healthcare delivery
- Describe select federal activities designed to mitigate safety threats associated with AI in healthcare

The National Action Alliance Website Serves as a Hub to Engage



https://www.ahrq.gov/action-alliance/index.html,
craig.umscheid@ahrq.hhs.gov