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

Marshall Chin, MD, MPH

Richard Parrillo Family Distinguished Service Professor of Healthcare Ethics in the Department of Medicine

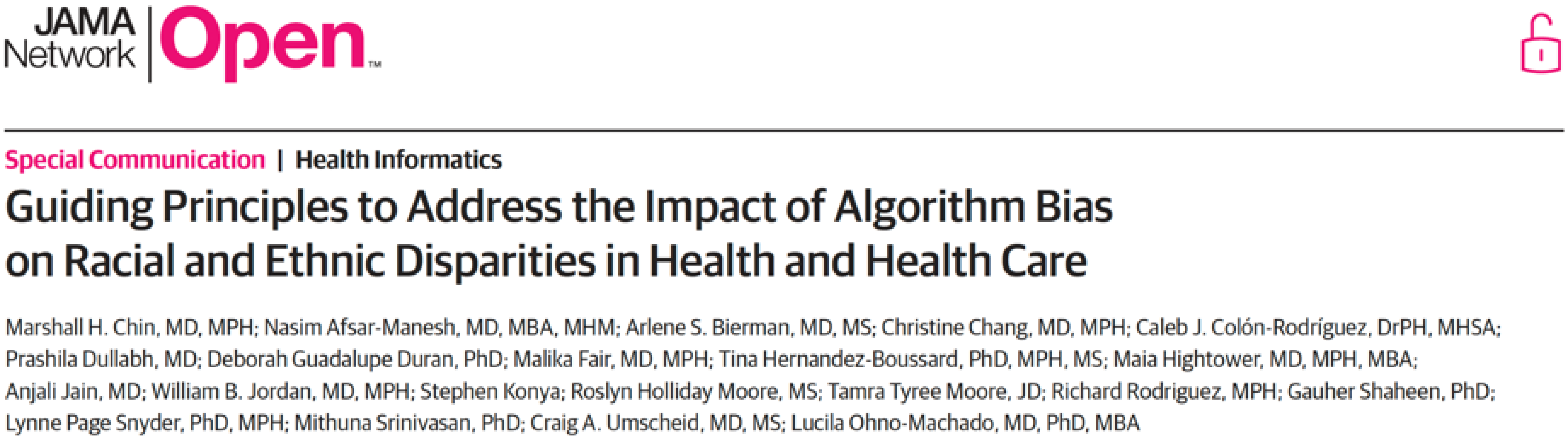
University of Chicago

@MarshallChinMD

Disclosures

Board Member/Advisory Panel:
CMS HCP LAN Health Equity Advisory Team
BCBS Health Equity Advisory Panel
Bristol-Myers Squibb Co. Health Equity Advisory Board

Research Support:
NIDDK P30 DK092949
NIDDK R25 DK130849
Robert Wood Johnson Foundation
Kaiser Foundation Health Plan, Inc.
AHRQ 1T32HS029581

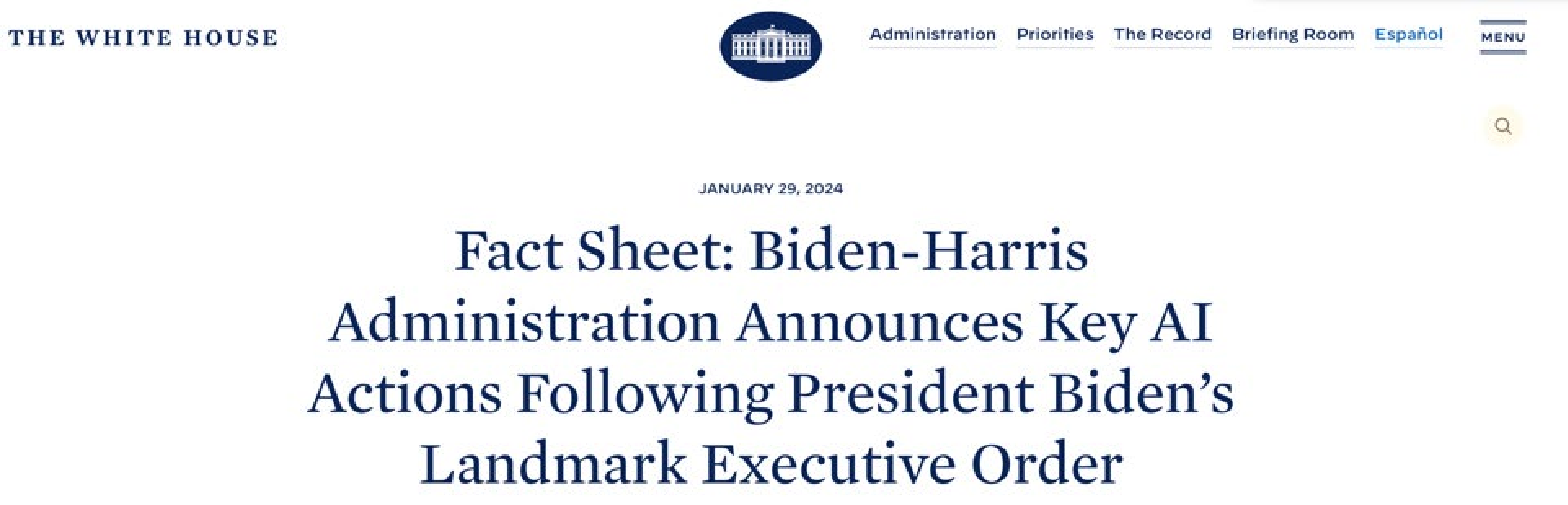


JAMA Network | Open

Special Communication | Health Informatics

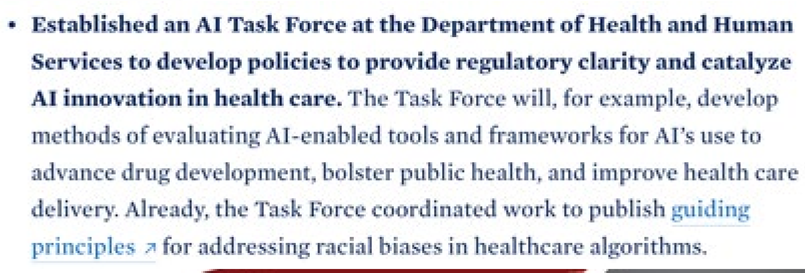
Guiding Principles to Address the Impact of Algorithm Bias on Racial and Ethnic Disparities in Health and Health Care

Marshall H. Chin, MD, MPH; Nasim Afsar-Manesh, MD, MBA, MHM; Arlene S. Bierman, MD, MS; Christine Chang, MD, MPH; Caleb J. Colón-Rodríguez, 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



JANUARY 29, 2024

Fact Sheet: Biden-Harris Administration Announces Key AI Actions Following President Biden's Landmark Executive Order



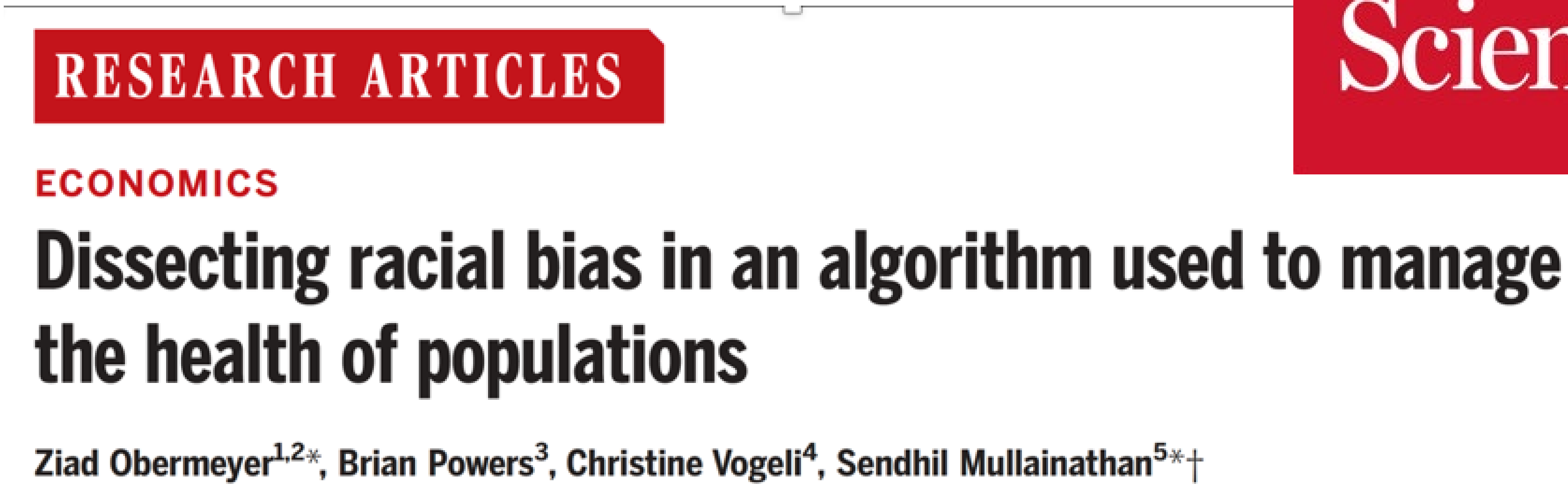
- **Established an AI Task Force at the Department of Health and Human Services to develop policies to provide regulatory clarity and catalyze AI innovation in health care.** The Task Force will, for example, develop methods of evaluating AI-enabled tools and frameworks for AI's use to advance drug development, bolster public health, and improve health care delivery. Already, the Task Force coordinated work to publish guiding principles ↗ for addressing racial biases in healthcare algorithms.

Healthcare Algorithm Definition

- Mathematical model used to inform decision-making
- Used for diagnosis, treatment, prognosis, risk stratification, triage, resource allocation
- Traditional statistical regression models – relationship between predictors and outcomes
- Artificial intelligence/machine learning – “learn” inferring relationships in large datasets - ”Black box” problem

Algorithmic Bias

- Unbiased algorithm – patients with same algorithm score or classification have same basic needs
- Bias in housing, banking, education, and health care (O’Neil. Weapons of Math Destruction 2016)
- Kidney function eGFR – inaccurately gave higher function to Black patients than white patients – led to delays in organ transplant referral (Vyas et al. NEJM 2020)



RESEARCH ARTICLES

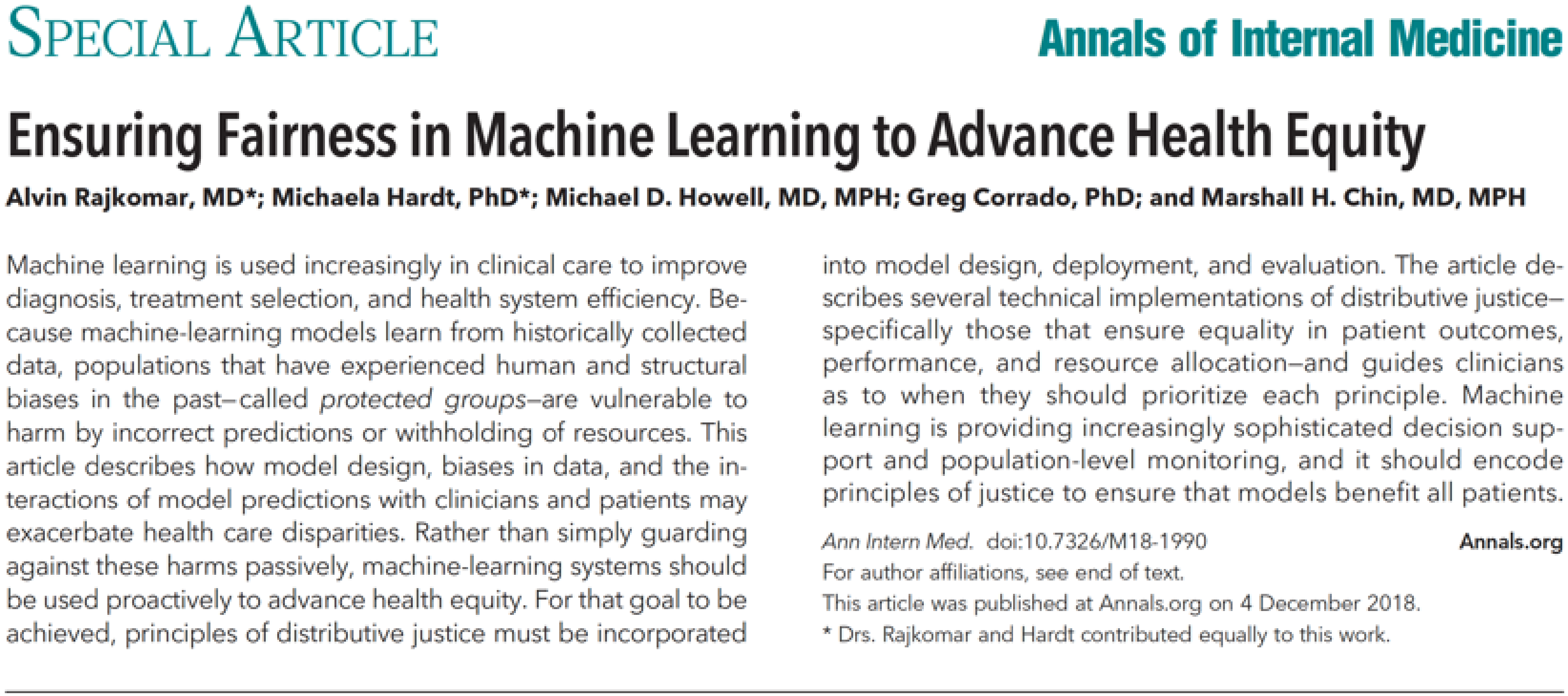
ECONOMICS

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]*†

Eligibility for chronic disease management program
Blacks had to be sicker than whites to qualify
$ bad proxy for health (Access barriers lead to less resource utilization for Blacks)

Sources of Bias



SPECIAL ARTICLE

Annals of Internal Medicine

Ensuring Fairness in Machine Learning to Advance Health Equity

Alvin Rajkomar, MD*; Michaela Hardt, PhD*; Michael D. Howell, MD, MPH; Greg Corrado, PhD; and Marshall H. Chin, MD, MPH

Machine learning is used increasingly in clinical care to improve diagnosis, treatment selection, and health system efficiency. Because machine-learning models learn from historically collected data, populations that have experienced human and structural biases in the past–called *protected groups*–are vulnerable to harm by incorrect predictions or withholding of resources. This article describes how model design, biases in data, and the interactions of model predictions with clinicians and patients may exacerbate health care disparities. Rather than simply guarding against these harms passively, machine-learning systems should be used proactively to advance health equity. For that goal to be achieved, principles of distributive justice must be incorporated into model design, deployment, and evaluation. The article describes several technical implementations of distributive justice–specifically those that ensure equality in patient outcomes, performance, and resource allocation–and guides clinicians as to when they should prioritize each principle. Machine learning is providing increasingly sophisticated decision support and population-level monitoring, and it should encode principles of justice to ensure that models benefit all patients.

Ann Intern Med. doi:10.7326/M18-1990 **Annals.org**

For author affiliations, see end of text.

This article was published at Annals.org on 4 December 2018.

* Drs. Rajkomar and Hardt contributed equally to this work.

Model Development

- Data (datasets; how data obtained [e.g. pulse oximetry overestimates oxygen saturation in Black patients])
- Design

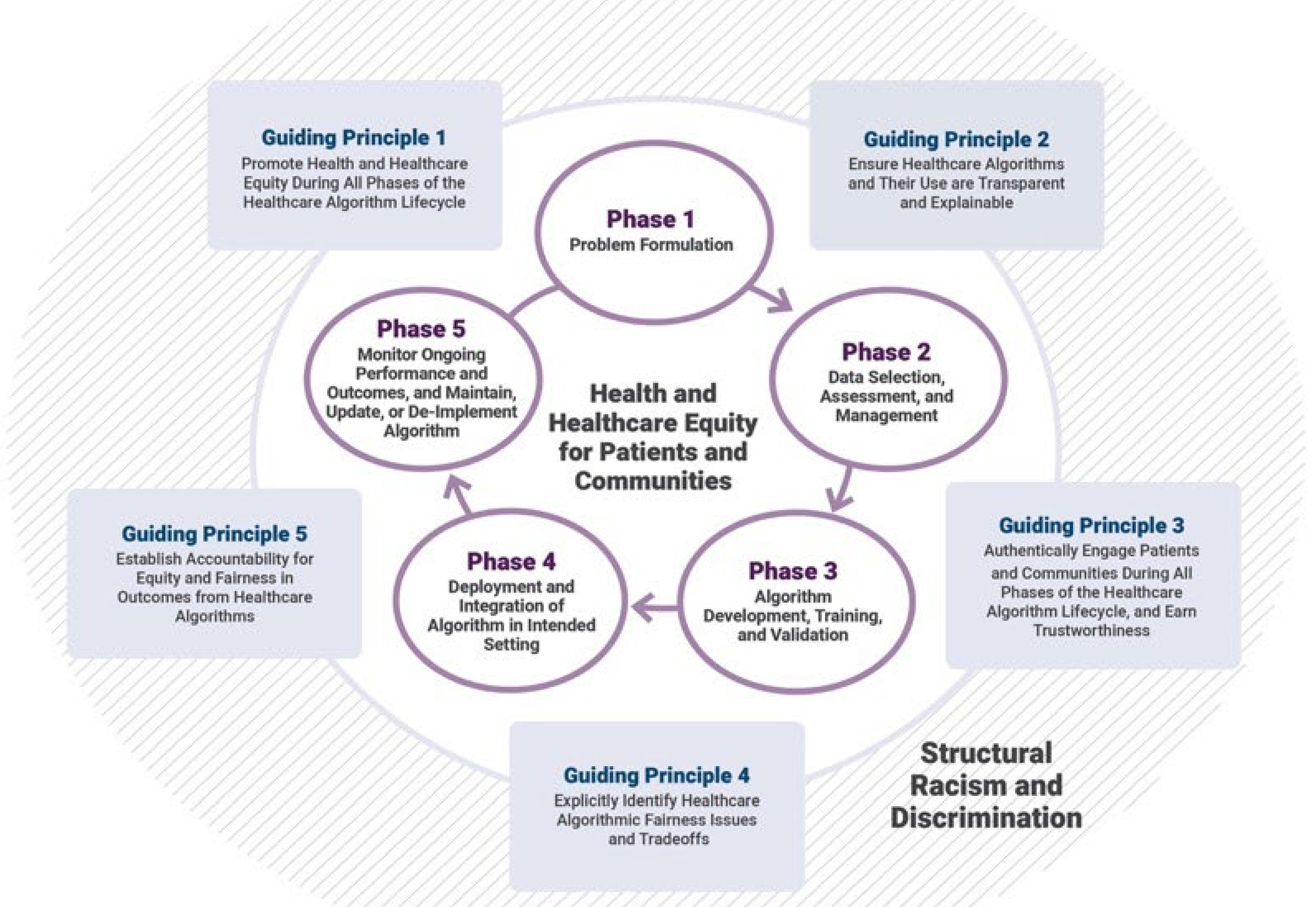
Model Deployment

Guiding Principles to Address the Impact of Healthcare Algorithms on Racial and Ethnic Disparities in Health and Healthcare

- Congressional request
- AHRQ, NIMHD, ONC, HHS OASH
- 9 person diverse panel
- Process
  - Evidence review
  - 2 days of stakeholder panels
  - Public virtual meeting – feedback, comments

JAMA Network Open™

Guiding Principles and Algorithm Lifecyle



Guiding Principle 1
Promote Health and Healthcare Equity During All Phases of the Healthcare Algorithm Lifecycle

Guiding Principle 2
Ensure Healthcare Algorithms and Their Use are Transparent and Explainable

Phase 1
Problem Formulation

Phase 2
Data Selection, Assessment, and Management

Phase 3
Algorithm Development, Training, and Validation

Phase 4
Deployment and Integration of Algorithm in Intended Setting

Phase 5
Monitor Ongoing Performance and Outcomes, and Maintain, Update, or De-Implement Algorithm

Health and Healthcare Equity for Patients and Communities

Guiding Principle 5
Establish Accountability for Equity and Fairness in Outcomes from Healthcare Algorithms

Guiding Principle 3
Authentically Engage Patients and Communities During All Phases of the Healthcare Algorithm Lifecycle, and Earn Trustworthiness

Guiding Principle 4
Explicitly Identify Healthcare Algorithmic Fairness Issues and Tradeoffs

Structural Racism and Discrimination

Guiding principles apply at each phase to mitigate and prevent bias in an algorithm.
Operationalization of the principles takes place at three levels - individual, institutional, and societal.

Chin et al. JNO 2023

Principle 1: Promote Equity in All Phases of Algorithm Lifecycle

“Health equity means that everyone has a **fair and just opportunity** to be as healthy as possible.”

Robert Wood Johnson Foundation 2017

“Achieving health equity requires valuing everyone equally with focused and ongoing societal efforts to **address avoidable inequalities, historical and contemporary injustices—which includes systemic racism**—and the elimination of health and healthcare disparities.”

CMS HCP LAN HEAT 2021

Equity

- Algorithms should be fair. Equitable outcomes for health and health care
- Bias in algorithms should be detected, mitigated, and prevented
- Algorithm performance monitored
- Healthcare decisions: human + algorithm
- No digital divide

Principle 2: Ensure Transparency and Explainability

"All relevant individuals should understand how their data is being used and how AI systems make decisions; algorithms, attributes, and correlations should be open to inspection."

HHS. Trustworthy AI Playbook 2021

Explainability

- Evidence and reasons for approaches, processes, outcomes
- Explanations are understandable to users
- Explanations correctly reflect the system’s process for generating the output
- Information ensures system only operates under conditions for which it was designed
- Outputs only used when the system achieves sufficient confidence in its results

Zuckermann BL et al. 2022

Principle 3: Authentically Engage Patients & Communities; Earn Trustworthiness

- Patients engaged in choosing problem, and algorithm data selection, development, deployment, and monitoring
- Patients aware how algorithm impacts care
- Data sovereignty – e.g. indigenous peoples
- Trustworthiness earned through authenticity, ethical practices, security of data, timely disclosures of algorithm use

Principle 4: Identify Fairness Issues and Tradeoffs

- Algorithmic fairness and bias issues arise from both ethical choices and technical decisions at each stage of the algorithm's lifecycle

- Distributive justice metrics
  - Societal - clinical outcomes, resource allocation
  - Technical - algorithms' performance metrics
    - Different technical definitions of algorithmic fairness are mathematically mutually incompatible, trading off maximizing accuracy of an algorithm for entire group and minimizing accuracy differences among subgroups across definitions

Fairness and Tradeoffs

- Mitigate bias
  - Social - diverse teams and stakeholder co-development
  - Technical - algorithmic fairness toolkits
- View algorithms and accompanying policies and regulations through:
  - Frames of equity of harms and risks
  - Explicit identification of trade-offs among different competing values and options
- Optimize model fairness for equity in clinical outcomes or resource allocation using bias mitigation methods and human judgment

Principle 5: Accountability

- Individuals and organizations must accept responsibility to achieve equity and fairness in outcomes from healthcare algorithms and be accountable
- Organizations should establish processes at each stage of the lifecycle of the algorithm to facilitate equity and fairness in outcomes
  - Involve model developers, end users, clinicians, administrators, and community representatives
- Organizations should have an inventory of their algorithms
  - Periodically screen for and mitigate bias
- Oversee prediction models
  - Checkpoint gates
  - Oversight governance structure
  - Investment

Accountability

- Regulations and incentives should support equity and fairness
- Algorithms should not be deployed before validation on the impacted population
- Those persons and communities who have been harmed by unfair algorithms should be redressed

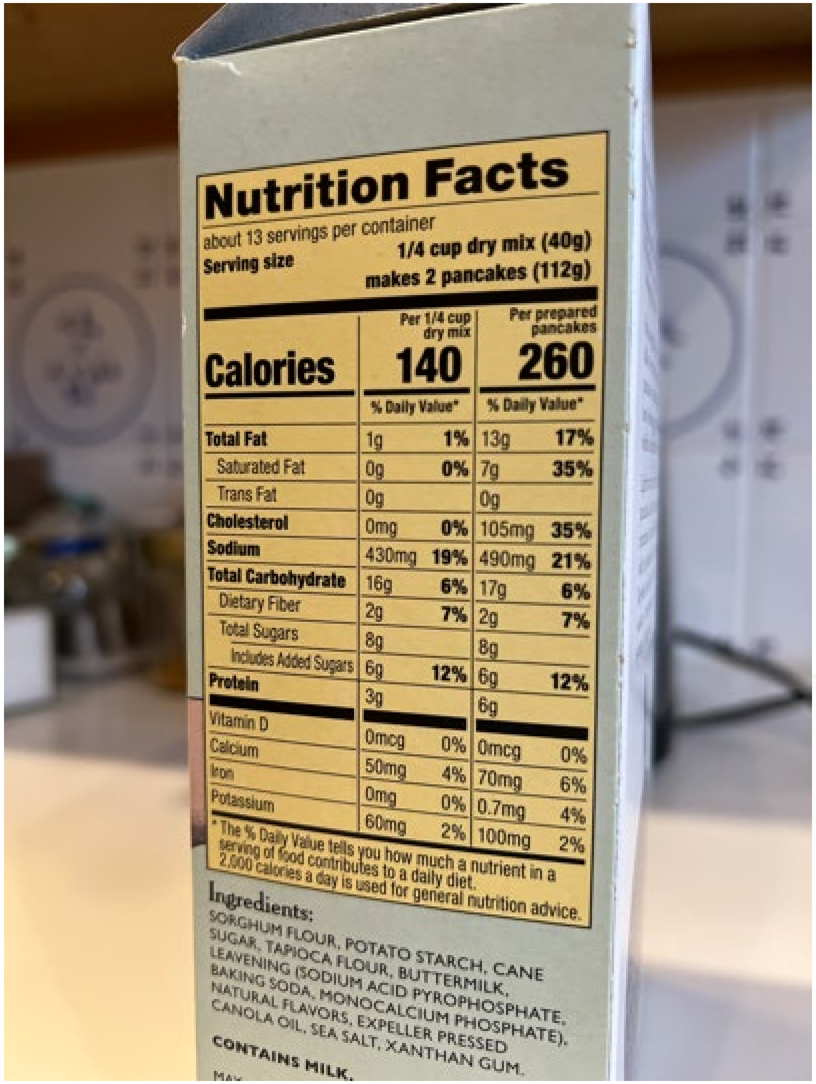
Overarching Issues and Challenges

- Technical definitions and metrics of fairness often do not translate clearly or intuitively to ethical, legal, social, and economic conceptions of fairness – 2 different worlds
- Trade-offs among competing fairness metrics and values are common
  - Clinical outcomes / Resource allocation vs. Technical performance metrics
    - Different technical definitions of algorithmic fairness are mathematically mutually incompatible, trading off maximizing accuracy of an algorithm for entire group and minimizing accuracy differences among subgroups across definitions

Ethics and Fairness

- No cookie cutter solution – Individualize to "Use case"
- Problem formulation
  - Health or Profit?
  - Equity/Justice or Efficiency/Save $?
- Distributive justice
  - Health outcomes
  - Resource allocation

Communications: “Nutrition Label”



Nutrition Facts

about 13 servings per container

Serving size **1/4 cup dry mix (40g)**
makes 2 pancakes (112g)

	Per 1/4 cup dry mix		Per prepared pancakes	
Calories	**140**		**260**	
	% Daily Value*		% Daily Value*	
Total Fat	1g	**1%**	13g	**17%**
Saturated Fat	0g	**0%**	7g	**35%**
Trans Fat	0g		0g	
Cholesterol	0mg	**0%**	105mg	**35%**
Sodium	430mg	**19%**	490mg	**21%**
Total Carbohydrate	16g	**6%**	17g	**6%**
Dietary Fiber	2g	**7%**	2g	**7%**
Total Sugars	8g		8g	
Includes Added Sugars	6g	**12%**	6g	**12%**
Protein	3g		6g	
Vitamin D	0mcg	0%	0mcg	0%
Calcium	50mg	4%	70mg	6%
Iron	0mg	0%	0.7mg	4%
Potassium	60mg	2%	100mg	2%

*The % Daily Value tells you how much a nutrient in a serving of food contributes to a daily diet. 2,000 calories a day is used for general nutrition advice.

Ingredients:
SORGHUM FLOUR, POTATO STARCH, CANE SUGAR, TAPIOCA FLOUR, BUTTERMILK, LEAVENING (SODIUM ACID PYROPHOSPHATE, BAKING SODA, MONOCALCIUM PHOSPHATE), NATURAL FLAVORS, EXPELLER PRESSED CANOLA OIL, SEA SALT, XANTHAN GUM.

CONTAINS MILK.

MAY

Communications Challenges

- Explaining probabilities and distributions
- Assessing and explaining real data and synthetic data
- Assessing and explaining the validity of applying a specific algorithm to a specific individual person
- Analogy: Legal informed consent versus patients truly understanding and providing informed consent

Regulations, Incentives, & Culture

- Regulations and incentives should support equity and fairness while also promoting innovation – dilemma European AI guidelines
- Ethical, legal, social, and administrative framework and culture should be created that redresses harm while encouraging quality improvement, collaboration, and transparency, similar to recommendations for patient safety

Call to Action

“ChatGPT and other artificial intelligence language models have spurred widespread public interest in the potential value and dangers of algorithms. Multiple stakeholders must partner to create systems, processes, regulations, incentives, standards, and policies to mitigate and prevent algorithm bias in health care. Dedicated resources and the support of leaders and the public are critical for successful reform. It is our obligation to avoid repeating errors that tainted use of algorithms in other fields.”

Chin et al. JNO 2023

"… the only solution is to apply to artificial intelligence algorithms the very thing they are designed to supersede—human intelligence."

Goodman, Goel, Cullen. Ann Intern Med 2018

References 1

- Advancing Health Equity Through APMs: Guidance for Equity-Centered Design and Implementation. Centers for Medicare and Medicaid Services Health Care Payment Learning and Action Network Health Equity Advisory Team. December 15, 2021. http://hcp-lan.org/workproducts/APM-Guidance/Advancing-Health-Equity-Through-APMs.pdf https://hcp-lan.org/advancing-health-equity-through-apms/
- Braveman P, Arkin E, Orleans T, Proctor D, Plough A. What is Health Equity? Robert Wood Johnson Foundation. May 1, 2017. https://www.rwjf.org/en/insights/our-research/2017/05/what-is-health-equity-.html
- Chin MH, et al. Guiding principles to address the impact of algorithm bias on racial and ethnic disparities in health and health care. JAMA Network Open. 2023 Dec 15;6(12):e2345050.
- Goodman SN, Goel S, Cullen MR. Machine learning, health disparities, and causal reasoning. Ann Intern Med 2018;169(12):883-884. doi: 10.7326/M18-3297.Epub 2018 Dec 4.

References 2

- Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. Oct 25 2019;366(6464):447-453. doi:10.1126/science.aax2342
- O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown; 2016.
- Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018 Dec 18;169(12):866-872. doi: 10.7326/M18-1990.
- US Department of Health and Human Services. Trustworthy AI (TAI) Playbook; 2021. https://www.hhs.gov/sites/default/files/hhs-trustworthy-aiplaybook.pdf

References 3

- Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. N Engl J Med. 2020;383(9):874-882. doi:10.1056/NEJMms2004740
- Zuckerman BL, et al. Options and Opportunities to Address and Mitigate the Existing and Potential Risks, as well as Promote Benefits, Associated With AI and Other Advanced Analytic Methods. OPRE Report 2022-253. US Department of Health and Human Services Office of Planning, Research, and Evaluation, Administration for Children and Families; 2022. https://www.acf.hhs.gov/opre/report/options-opportunities-address-mitigate-existing-potential-risks-promote-benefits

Thank You!

Marshall Chin, MD, MPH

@MarshallChinMD