Algorithm-Based Clinical Decision Support (ABCDS) Oversight

Duke Health's Approach to Operationalizing & Governing Health AI

Michael Pencina, PhD
Michael Cary, PhD, RN
Nicoleta J. Economou, PhD

September 26, 2023

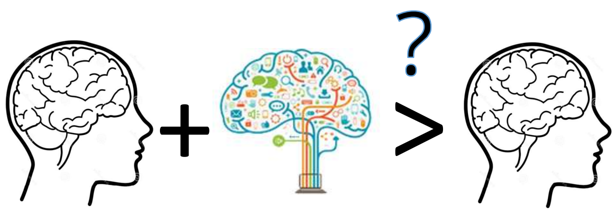
1

Agenda

- Promise of AI in Healthcare and the Current Landscape
- Algorithm-Based Clinical Decision Support (ABCDS) Oversight
- Bias Mitigation Strategies
- Benefits and Learnings from the Implementation of an Algorithmic Oversight Framework

2

Promise of Artificial Intelligence/Machine Learning in Health Care



Duke AI HEALTH

3

"Wild West" of Algorithms

"We have a Wild West of algorithms," said Michael Pencina, coalition [CHAI] co-founder and director of Duke AI Health. There's so much focus on development and technological progress and not enough attention to its value, quality, ethical principles or health equity implications."

Politico, April 4, 2023

Duke AI HEALTH

4

AI/ML Risks

Research

JAMA Internal Medicine | Original Investigation

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

Andrew Wong, MD; Erkin Otles, MEng; John P. Donnelly, PhD; Andrew Krumm, PhD; Jeffrey McCullough, PhD; Olivia DeTroyer-Cooley, BSE; Justin Pestrue, MEcon; Marie Phillips, BA; Judy Konye, MSN, RN; Carleen Penoza, MHSA, RN; Muhammad Ghous, MBBS; Karandeep Singh, MD, MMSc

IMPORTANCE The Epic Sepsis Model (ESM), a proprietary sepsis prediction model, is implemented at hundreds of US hospitals. The ESM's ability to identify patients with sepsis has not been adequately evaluated despite widespread use.

OBJECTIVE To externally validate the ESM in the prediction of sepsis and evaluate its potential clinical value compared with usual care.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study was conducted among 27 697 patients aged 18 years or older admitted to Michigan Medicine, the academic health system of the University of Michigan, Ann Arbor, with 38 455 hospitalizations between December 6, 2018, and October 20, 2019.

EXPOSURE The ESM score, calculated every 15 minutes.

MAIN OUTCOMES AND MEASURES Sepsis, as defined by a composite of (1) the Centers for Disease Control and Prevention surveillance criteria and (2) *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* diagnostic codes accompanied by 2 systemic inflammatory response syndrome criteria and 1 organ

Editorial page 1040
Multimedia
Supplemental content
CME Quiz at jamacmelookup.com and CME Questions page 1148

RESEARCH

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]*†

that rely on past data to build a predictor of future health care needs.

Our dataset describes one such typical algorithm. It contains both the algorithm's predictions as well as the data needed to understand its inner workings: that is, the underlying ingredients used to form the algorithm (data, objective function, etc.) and links to a rich set of outcome data. Because we have the inputs, outputs, and eventual outcomes, our data allow us a rare opportunity to quantify

"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…"*

Duke AI HEALTH

5

We need to do better

Prediction Models — Development, Evaluation, and Clinical Application

Michael J. Pencina, Ph.D., Benjamin A. Goldstein, Ph.D., and Ralph B. D'Agostino, Ph.D.

"Given the number of emerging prediction models and their diverse applications, no single regulatory agency can review them all. This limitation, however, does not absolve models' developers and users from applying the utmost scrutiny in demonstrating effectiveness and safety."

Duke AI HEALTH

Pencina MJ, Goldstein BA, D'Agostino RB. *N Engl J Med*. 2020 Apr 23;382(17):1583-1586. doi: 10.1056/NEJMp2000589.

6

Considerations for CDS development

- Population at risk
- Outcome of interest
- Time horizon
- Predictors
- Mathematical model
- Model evaluation
- Translation to CDS
- Clinical implementation

Duke AI HEALTH

Pencina MJ, Goldsein BA, D'Agostino RB. Prediction Models – Development, Evaluation, and Clinical Application. NEJM. 2020;382:1583-1586. doi:10.1056/NEJMp2000589.

7

Principles for Responsible AI

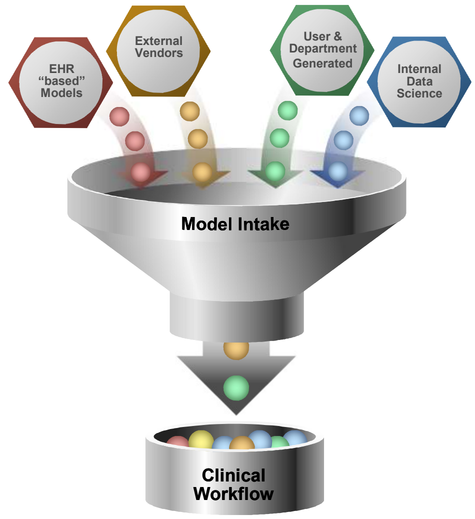
- Ensure that AI technology serves humans
- Define the task we want the AI tool to accomplish
- Describe what the successful use of the AI tool looks like
- Create transparent systems for continuously testing and monitoring AI tools

Duke AI HEALTH

8

Complex environment

Different:
- Skills
- Knowledge bases
- Resources available
- Make up of project teams



Duke AI HEALTH

9

Formation of the ABCDS Oversight Committee

In recognition of this changing landscape, the Duke Health Chancellor and the Dean of the School of Medicine charged Duke Health leadership to form an oversight framework.

Duke AI HEALTH

10

Mission Statement

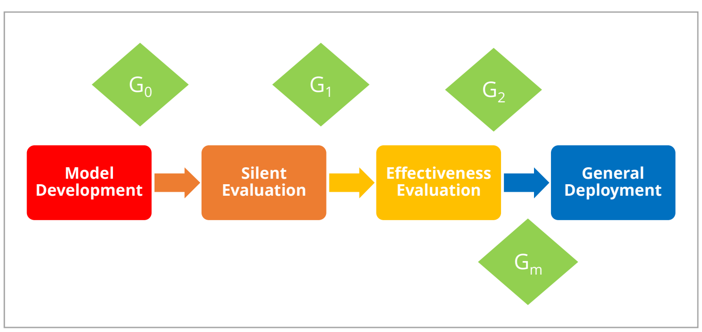
"Out of our primary focus on patient safety and high-quality care, our mission is to guide algorithm-based clinical decision support (ABCDS) tools through their lifecycle by providing governance, evaluation, and monitoring."



Duke AI HEALTH

11

ABCDS Lifecycle & Our Framework



'Just-in-time' Check-Points (**G**ates) Help Model Owners Get Ready for What's Ahead

- What are the clinical outcome and performance metrics?
- How has the model been evaluated?
- Who is the Clinical Owner?
- Who will cover maintenance costs in production?
- Will this ABCDS tool be used outside of Duke Health?
- Is this a standard of care model?
- How will the model be used in the clinic and how is it integrated with the workflow?

Duke AI HEALTH

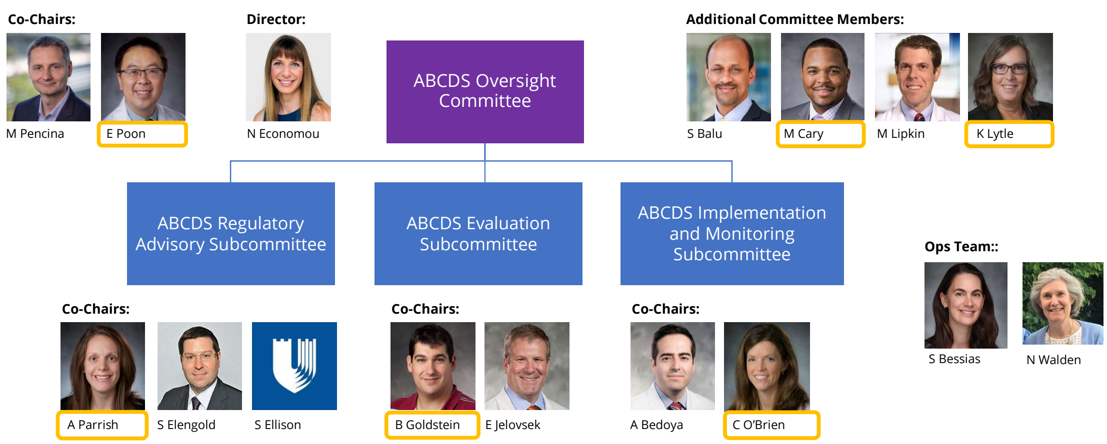
12

People

Duke AI HEALTH

13

People: ABCDS Oversight Committee



Duke AI HEALTH

14

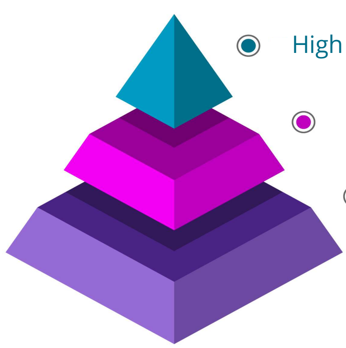
Process

15

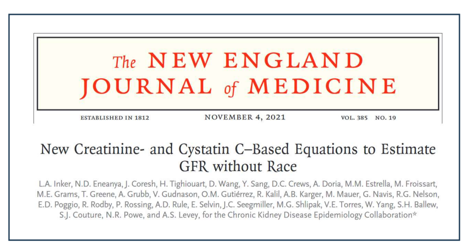
Scope of ABCDS Oversight Framework

ABCDS Tool = Algorithm(s) + Interface Algorithms Are Presented In

All electronic algorithms that could impact patient care at Duke Health fall within the scope of the ABCDS Oversight Committee and must undergo registration.



- High Risk: Data-Derived
- Medium Risk (e.g., Clinical Consensus)
- Low Risk: Standard of Care

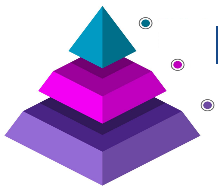

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 NOVEMBER 4, 2021 VOL. 385 NO. 19

New Creatinine- and Cystatin C–Based Equations to Estimate GFR without Race

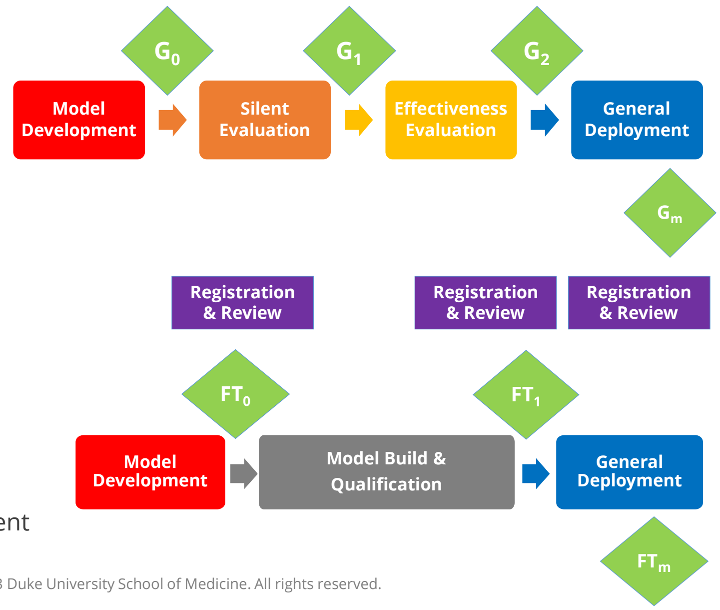
L.A. Inker, N.D. Eneanya, J. Coresh, H. Tighiouart, D. Wang, Y. Sang, D.C. Crews, A. Doria, M.M. Estrella, M. Froissart, M.E. Grams, T. Greene, A. Grubb, V. Gudnason, O.M. Gutiérrez, R. Kalil, A.B. Karger, M. Mauer, G. Navis, R.G. Nelson, E.D. Poggio, R. Rodby, P. Rossing, A.D. Rule, E. Selvin, J.C. Seegmiller, M.G. Shlipak, V.E. Torres, W. Yang, S.H. Ballew, S.J. Couture, N.R. Powe, and A.S. Levey, for the Chronic Kidney Disease Epidemiology Collaboration*

16



Fast-Track Checkpoint Review

Full Review



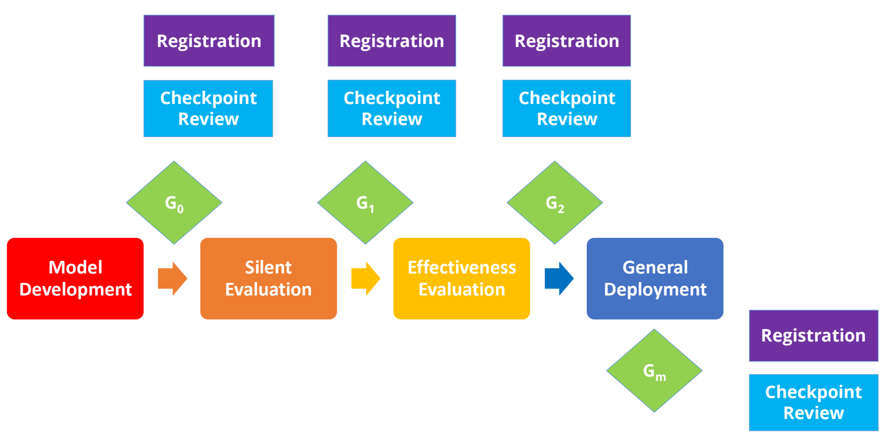
Fast-Track Review

Includes:

- Clinical Validation
- Technical Specification Document

17

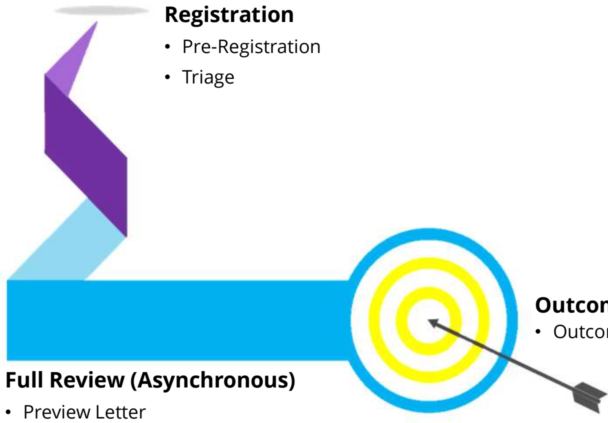
Full Checkpoint Review



Duke AI HEALTH

18

What to Expect: ABCDS Checkpoint Review



Registration
- Pre-Registration
- Triage

Full Review (Asynchronous)
- Preview Letter
- Review Meeting w/ Committee (Optional)

Outcome
- Outcome Letter

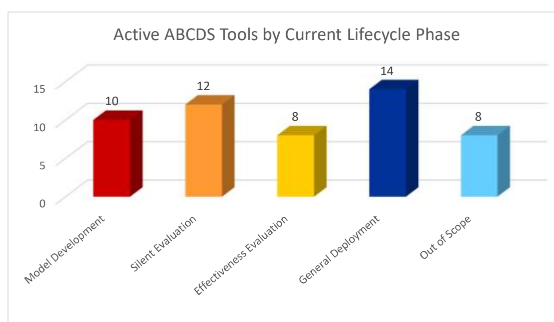
Outcomes:
- Approval
- Approval with Contingencies
- Re-review
- Denial

Duke AI HEALTH

19

Portfolio Metrics (April 2023)

ABCDS Model Registration & Review	Total
Number of registered tools	52
Number of evaluated tools	31



Duke AI HEALTH

20

Implementing Quality & Ethics with Our Framework

Transparency & Accountability

Impact & Safety

Fairness & Equity

Usability & Adoption

Regulatory Compliance

Duke AI HEALTH

21

Implementing Quality & Ethics with Our Framework

Transparency & Accountability

Impact & Safety

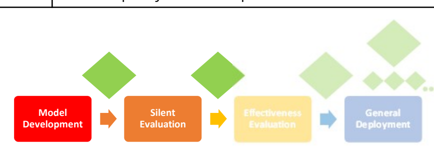
Fairness & Equity

Usability & Adoption

Regulatory Compliance

Principle	Criteria	Submission Materials
Clinical Impact & Safety	The ABCDS software, in comparison to current state, stands to improve clinical care.	✓ Evidence that the tool has potential to impact clinical outcomes or processes ✓ List of key impact metrics (clinical outcomes and/or process improvement) with definitions, following TRIPOD guidelines[5] ✓ List of core performance metrics (e.g. sensitivity, PPV, etc.) and results from development ✓ Calibration curves, threshold selections and justification if applicable
	Plans for Silent Evaluation will inform the decision to proceed with pilot implementation in clinic.	Silent Evaluation Plan, including: ✓ Summary of benefits you expect to demonstrate and criteria to proceed into Effectiveness Evaluation ✓ Study design, including in/exclusion criteria, timeframe and sample size considerations ✓ Core performance metrics with shell tables ✓ Data analysis plan ✓ Data quality evaluation plan

Sample evaluation criteria supporting the principle of clinical impact and safety at the G_0 Checkpoint evaluation between pilot implementation and general deployment



Duke AI HEALTH

(Unpublished work)

22

Implementing Quality & Ethics with Our Framework

Transparency & Accountability

Impact & Safety

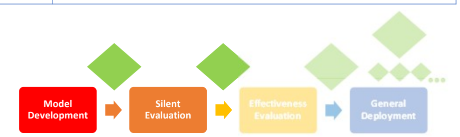
Fairness & Equity

Usability & Adoption

Regulatory Compliance

Principle	Criteria	Submission Materials
Fairness & Equity	The principles of fairness and equity are reflected in the development process.	✓ Summary of fairness & equity considerations during development ✓ Strategy for subgroup analysis with list of key stratification variables (e.g. race, age group, etc.) and rationale ✓ Subgroup analysis of chosen impact metrics ✓ Subgroup analysis of chosen model performance metrics ✓ Interpretation of findings and recommendations

Sample evaluation criteria supporting the principle of clinical impact and safety at the G_0 Checkpoint evaluation between pilot implementation and general deployment

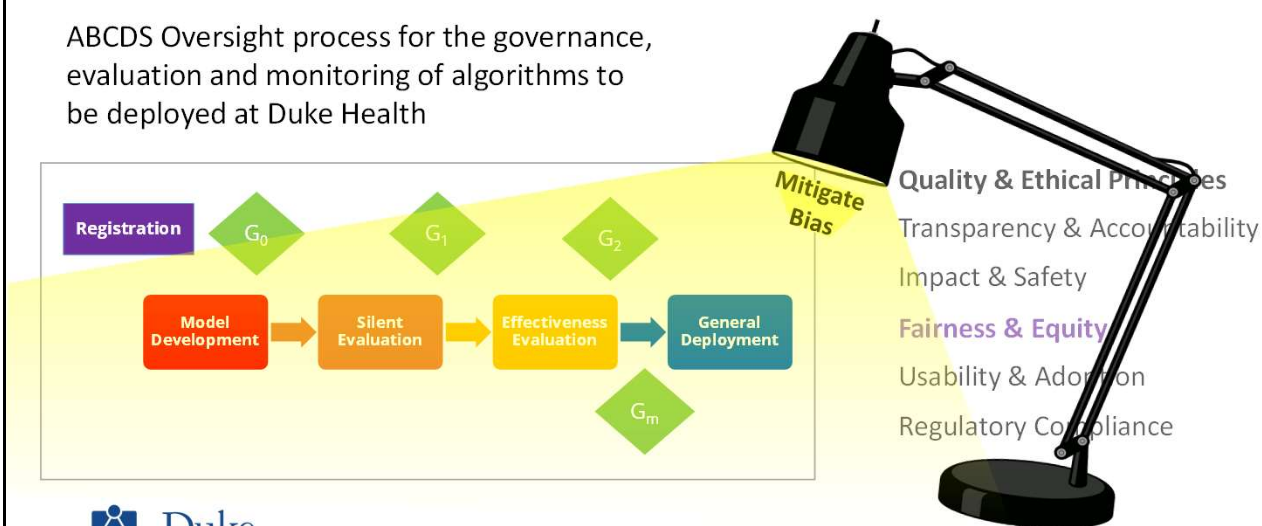


Duke AI HEALTH

23

Mitigating Bias Through Algorithmic Oversight

ABCDS Oversight process for the governance, evaluation and monitoring of algorithms to be deployed at Duke Health

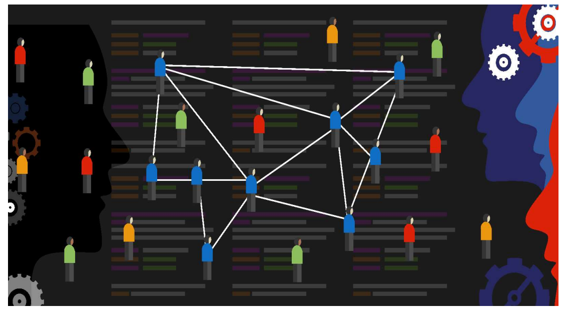


Duke AI HEALTH

24

What is Bias in Clinical Algorithms?

Bias refers to the difference in how one or more subgroups is treated, represented or perceived, resulting in unfair/unjust outcomes.



Duke AI HEALTH

Accessed on July 25, 2023, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

25

Societal Bias

Bias Type	Example	Assessment	Mitigation Strategy
Societal Bias Bias due to training data shaped by present and historical inequities and their fundamental causes	Predictive policing algorithms[1] are trained on data that reflects structural racism and criminalization of, e.g., homelessness and poverty. Groups that are more likely to interact with the police are more likely to be identified by policing algorithms as "at risk" for future offense.	*Please discuss the real-world inequities reflected in your training data and how they inform the problem formulation and intended purpose of your model.*	• *Restriction to particular settings or use cases* • *Human-in-the-loop deployment design* • *Multi-stakeholder engagement*

Label Bias
Aggregation Bias
Learning Bias
Representation Bias
Evaluation Bias
Human Use Bias

Duke AI HEALTH

[1] Angwin, J. Larson, S. Mattu, L. Kirchner, "Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks," *ProPublica*, 23 May 2016; www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

26

Label Bias

Bias Type	Example	Assessment	Mitigation Strategy
Label Bias Use of a biased proxy target variable in place of the ideal prediction target.	An algorithm[1] used to identify patients for high-risk care management services predict healthcare costs as a proxy for healthcare *need*. Despite having greater health needs, Black patients have lower average healthcare spending (due to structural barriers in access to care) and are thus less likely to be recognized by the algorithm as 'high risk.'	*Please discuss any proxies used as inputs or outputs. Provide a rationale and describe implications for use.*	• *Eliminating proxies (where possible) or choosing a proxy as close as possible to the intended idea or concept*

Label Bias
Aggregation Bias
Learning Bias
Representation Bias
Evaluation Bias
Human Use Bias

Duke AI HEALTH

[1]Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019 Oct 25;366(6464):447-453. doi: 10.1126/science.aax2342.

27

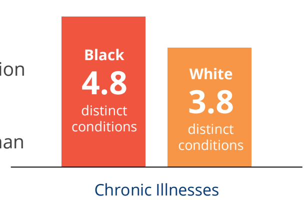
Why is it Important to Identify Racial/Ethnic Bias in Health Algorithms?

Algorithms are used to identify patients with complex health needs in order to provide more comprehensive care management. However, these algorithms can exhibit significant racial bias.

A 2019 study of one such algorithm found:

Black patients who are considerably sicker than White patients are given the same risk score

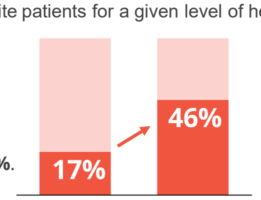
At the risk level that would result in automatic identification for the care management program, Black patients had **26%** more chronic illnesses than White patients.



Why is this?

This algorithm assigned risk scores based on past health care spending. Black patients have lower spending than White patients for a given level of health.

If this bias was eliminated, the percentage of Black patients automatically enrolled in the program would rise from **17%** to **46%**.



Duke AI HEALTH

[1]Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019 Oct 25;366(6464):447-453. doi: 10.1126/science.aax2342.

28

Aggregation Bias

Bias Type	Example	Assessment	Mitigation Strategy
Aggregation Bias Bias due to use of a one-size-fits-all model for data in which there are underlying groups or types of examples.	A natural language processing (NLP) model developed to scan clinical notes and suggest medication review is used across hospitals in a large health system in which documentation practices differ between locations, leading to poor performance in recently-acquired rural hospitals switching EHR systems.	*Please discuss the ways that the data used to train your model may be observed differently across subgroups.*	• *Use of subpopulation-specific models instead of or in addition to one-size-fits-all models* • *Use of subgroup-specific thresholds in a one-size-fits-all model* • *Imputation or other strategies to improve mapping from inputs to labels across subgroups*

Label Bias
Aggregation Bias
Learning Bias
Representation Bias
Evaluation Bias
Human Use Bias

Duke AI HEALTH

29

Learning Bias

Bias Type	Example	Assessment	Mitigation Strategy
Learning Bias Bias due to modeling choices that amplify performance disparities across subgroups.	A development team is working on prediction of asthma exacerbation and uses a variety of methods to generate candidate models. The final model is selected by ranking the candidates on a single performance metric, AUROC. The focus on a single summary metric conceals large performance differences by race leading to poor prediction in the demographic most exposed to environmental asthma triggers.	*Please describe how the model was optimized and the performance metrics used among candidate models.*	• *Penalized optimization methods* • *Subgroup analysis to inform model selection*

Label Bias
Aggregation Bias
Learning Bias
Representation Bias
Evaluation Bias
Human Use Bias

Duke AI HEALTH

30

Representation Bias

Bias Type	Example	Assessment	Mitigation Strategy
Representation Bias Bias emerging from non-representative training data which can lead to poor performance in subsets of the deployment population.	A melanoma detection model[1] achieved accuracy parity with a board-certified dermatologist; however, the model was trained primarily on light-colored skin. As such, the algorithm is likely to underperform for patients with dark skin.	*Please discuss the quality and representativeness of your training data.* *If your model is adaptive, please discuss how you will ensure representativeness of the training data on an ongoing basis.*	• *Integration with data from other sources* • *Supplementation with synthetic data* • *Up- or down-sampling approaches* • *Acknowledgement of limitations in model brief or other training materials* • *Refitting an out-of-the-box model to the local population*

Label Bias
Aggregation Bias
Learning Bias
Representation Bias
Evaluation Bias
Human Use Bias

Duke AI HEALTH

[1]Wang HE, et al. A bias evaluation checklist for predictive models and its pilot application for 30-day hospital readmission models. *J Am Med Inform Assoc*. 2022 Jul 12;29(8):1323-1333. doi: 10.1093/jamia/ocac065.

31

Evaluation Bias

Bias Type	Example	Assessment	Mitigation Strategy
Evaluation Bias Bias emerging from a validation dataset that is not reflective of the deployment population and/or the training population.	A health system implements a new vendor model to predict in-hospital deterioration after receiving a validation report showing strong performance in other health systems that share the same EHR. Once the model is connected to the local data source, it produces an unexpected number of false alerts.	*Briefly summarize plans for local validation.*	• *Local validation (required)* • *Re-fitting the model on development sample that better represents the deployment population* • *Post-deployment monitoring with chart review (required)*

Label Bias
Aggregation Bias
Learning Bias
Representation Bias
Evaluation Bias
Human Use Bias

Duke AI HEALTH

32

Human Use Bias

Bias Type	Example	Assessment	Mitigation Strategy
Human Use Bias Inconsistent user response to algorithm outputs for different subgroups.	A machine learning algorithm1 developed to help pathologists differentiate liver cancer types did not improve every pathologist's accuracy despite the model's high rate of correct classification. Instead, pathologists' accuracy was improved when the model's prediction was correct but decreased when the model's prediction was incorrect.	*Briefly describe how your algorithm fits into the clinical workflow. If it will replace an existing model or process, please include a comparison to baseline.*	• *Workflow design solutions* • *End user training* • *Post-deployment monitoring with chart review (required)* • *Collection of end user feedback and metrics of adoption*

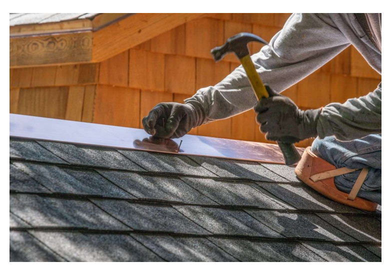
- Label Bias
- Aggregation Bias
- Learning Bias
- Representation Bias
- Evaluation Bias
- Human Use Bias

Duke AI HEALTH

Wang HE, et al. A bias evaluation checklist for predictive models and its pilot application for 30-day hospital readmission models. *J Am Med Inform Assoc*. 2022 Jul 12;29(8):1323-1333. doi: 10.1093/jamia/ocac065.

33

Impacting How We Deliver Patient Care



Duke AI HEALTH

34

Lessons Learned

- **Successful AI Governance is a Team Sport**
  - ✓ Lots of skillsets, perspectives and languages to bring together
- **Culture Shift is Hard**
  - ✓ Show Teams how to succeed by addressing gaps in their knowledge, skillsets, and/or bandwidth
  - ✓ Governance's role is Coach and Facilitator (not Punisher)
  - ✓ There is no such thing as over-communication in a complex system
- **Benefits of Centralized Governance**
  - ✓ Transparency of process expectations
  - ✓ Institutional visibility into all the 'skeletons in the closet'
- **Conscious Decision (thus far) Not to Regulate Who Gets to Build AI Models**

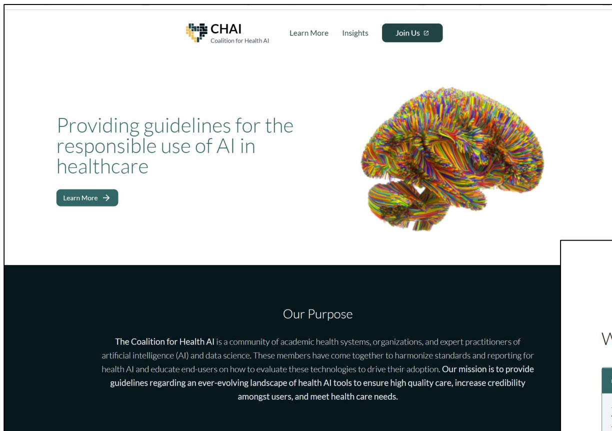
Duke AI HEALTH

35

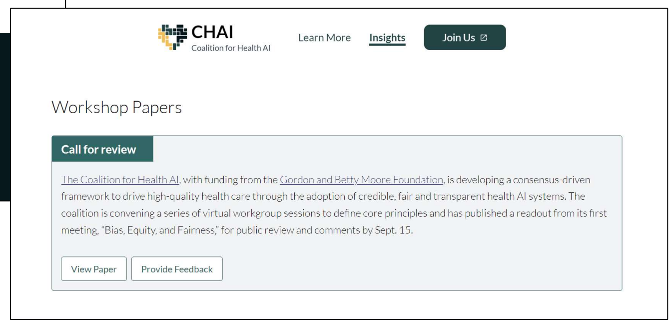
Future Directions

- QMS
  - SOPs
  - Centralized model monitoring / safety surveillance
- Incorporating the patient voice

36



https://www.coalitionforhealthai.org/



https://www.coalitionforhealthai.org/insights/

Duke AI HEALTH

37

Learn More…

https://aihealth.duke.edu/algorithm-based-clinical-decision-support-abcds/

Bedoya AD, et al. A framework for the oversight and local deployment of safe and high-quality prediction models.
J Am Med Inform Assoc. 2022 Aug 16;29(9):1631-1636. doi: 10.1093/jamia/ocac078. PMID: 35641123; PMCID: PMC9382367.

Contact us at abcds@duke.edu or nicoleta.economou@duke.edu

DukeHealth

Duke AI HEALTH

38

Thank you

Duke
AI HEALTH

39