

# Algorithmic Transparency:

## Being Clear with the Public

This letter was written by an interdisciplinary group at the University of Michigan and includes co-signatories from other institutions. It was developed and informed by research conducted by those involved in building predictive decision support interventions (DSIs), those studying public attitudes about data and artificial intelligence in clinical decision support systems, and those in current practice at academic medical centers. The views expressed in this letter represent those of the signatories and not of their institutions.

On April 18, 2023, the Office of the National Coordinator (ONC) issued a notice of proposed rulemaking (NPRM) with the goal of increased transparency about the use of algorithms being used in health information systems and in care settings. As currently drafted, the proposed rule would also require certified health information technology (IT) developers to make information about intervention risk management practices, such as risk analysis, mitigation, and governance, available to the general public. However, the proposal limits the availability of “source attributes” for a given predictive decision support intervention, or DSI, to end users only. The current source attributes include bibliographic citations relevant to the clinical decision support (CDS) intervention (e.g., citation of guideline), the developer of the CDS intervention, the funding source of CDS intervention development, and release and revision dates. In the proposed rule, the ONC proposes to expand these attributes to include the use of three additional specific types of data in a DSI in § 170.315(b)(11)(vi)(A) – demographic data elements in § 170.315(b)(11)(vi)(A)(5), SDOH data elements in § 170.315(b)(11)(vi)(A)(6), and Health Status Assessment data elements in § 170.315(b)(11)(vi)(A)(7).

We support the notion of making information about risk management practices available to the general public. We also believe that this same level of transparency should be extended to include source attributes and variables identified in § 170.315(b)(11)(vi). Additionally, we recommend that *both* the datasets used to train and evaluate predictive DSIs, as well as *all* variables used as predictors and outcomes in predictive DSIs, should *also* be considered as new source attributes. This is because recent backlash and harm resulting from the use of predictive DSIs in healthcare can largely be attributed to lack of public awareness around which *datasets* and *variables* were used by the predictive DSIs to make recommendations about people’s healthcare, as discussed below.

Predictive DSIs capture biases through the inclusion of biased training and evaluation *datasets*. Health datasets can be biased because they are commonly created through the digitization of routine clinical practices, which are known to be inequitable in both the access to and quality of care.<sup>1</sup> Avoidance of biased predictive DSIs is a key component of every responsible artificial intelligence (AI) framework, including the White House Blueprint for an AI Bill of Rights,<sup>2</sup> the Coalition of Health AI,<sup>3</sup> and the Health AI Partnership.<sup>4</sup> The use of biased datasets produces biased DSIs,<sup>5,6</sup> and algorithms produced from datasets in one healthcare setting often do not generalize to other healthcare settings.<sup>7</sup> Given the important role that datasets play in the resulting DSIs, we recommend that *information* about which

datasets were used to train and evaluate a predictive DSI should be in the public domain. We also recommend that this information should be added to the set of source attributes and variables identified in § 170.315(b)(11)(vi).

Biased datasets can harm patients through the propagation of this bias into algorithm-based decision-making. For example, gender imbalance in both clinical and medical imaging datasets has been shown to produce predictive DSIs that perform worse in women.<sup>5,6</sup> Information about which datasets were used to train and evaluate a predictive DSI is critical when determining whether an algorithm created from it is fit for purpose or whether it could be biased.

Predictive DSIs also capture biases through the inclusion of biased *variables*. For example, a widely used commercial model to predict whether a patient would “no-show” to a clinical appointment considered ethnicity, financial class, religion, and body mass index in its predictions.<sup>8</sup> There is no scientific or ethical basis for which health systems should rely on *any* of these variables in determining the accessibility of patient care. Additionally, use of these variables could lead to the use of clinical workflows that limit access to care for patients on the basis of characteristics like religion. While this information was available to end users through a login process, it was marked as “Confidential” by the vendor. The vendor eventually moved to a new version of this model that removed these variables after this information became public. Had this information been publicly available *at the start*, its use would have been flagged much earlier and likely prior to implementation. We believe that public availability would have led the vendor to remove these variables sooner, thereby minimizing the harm to patients from their inclusion.

Problematic variables do not necessarily have to be linked to personal patient characteristics, such as race, to be problematic.<sup>9,10</sup> For example, a recent evaluation of a proprietary sepsis model noted that the use of antibiotics was used as a *predictive* variable of developing sepsis. This is problematic because patients are given antibiotics once they have already developed sepsis. A model that relies on antibiotic use to predict whether or not a patient will develop sepsis in the future will be less effective and result in false negatives. Without transparency into the variables used to generate this model, researchers never would have been able to isolate the problematic variable and fix the model. The vendor has since removed antibiotics from the model,<sup>11</sup> although information about the model remains confidential and not in the public domain.

These problems are not limited to a single use case or vendor. Over a dozen widely used models have been shown to inappropriately consider race when generating predictions, with varying levels of transparency about the role of race in the models.<sup>12</sup> The consideration of race is inappropriate because using race to make medical decisions can perpetuate and amplify race-based health inequities already present in our health system. Even when race is not present in a model, proxy variables may unintentionally capture race and thus worsen race-based health inequities.<sup>13</sup> This issue may not necessarily be recognized by clinician end users due to lack of expertise. Placing information about model variables in the public domain will allow for a wider lens of scrutiny.

In light of the need for a broader consideration of transparency than is currently provided in the NPRM, we recommend the following:

- Place the source attributes identified in § 170.315(b)(11)(vi) for predictive DSIs in the public domain. It is not sufficient to have the attributes *only* available for “user review”

(<https://www.federalregister.gov/d/2023-07229/p-621>). The source attributes should be readily accessible by the public and in the public domain. This recommendation is aligned with Executive Order 13960 Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, which states that “agencies shall be transparent in disclosing relevant information regarding their use of AI to appropriate stakeholders, including the Congress and the public.”<sup>14</sup> Having source attributes in the public domain would help to achieve this goal.

- Add information about which data sources were used for training and evaluation of the predictive DSIs to the set of source attributes and variables identified in § 170.315(b)(11)(vi). Additionally, place this information in the public domain.
- Add information about *all* variables (predictors and outcomes) used in predictive DSIs to the set of source attributes and variables identified in § 170.315(b)(11)(vi). Additionally, place this information in the public domain. All variables should be included because demographic and proxy variables may be important for effective evaluation of bias, discrimination, utility, and quality. In cases where listing all variables may not be possible (such as models using x-ray imaging data), a clear statement to that effect should be made.
- In response to the question as to whether having this information publicly available would improve public confidence in predictive DSIs by enabling research on source attribute information (<https://www.federalregister.gov/d/2023-07229/p-709>), we believe that making information about models readily available and easy to find is critical to improving public confidence in predictive DSIs and in enabling research on source attribute information. Research should also include studies that inform evidence-based best practices for communications strategies, such as model cards, product labels, and dissemination, and studies that evaluate the impact of governance on transparency and public trust.

**Signed,**

Karandeep Singh, MD, MMSc  
Assistant Professor  
University of Michigan

Jodyn Platt, MPH, PhD  
Associate Professor  
University of Michigan

**University of Michigan (Ann Arbor, MI)**

Ranjit Aiyagari, MD  
Sardar Ansari, PhD  
Emily Balczewski, BA  
Michael Burns, MD, PhD  
Jie Cao, MPH  
Gloria Carmona, MPH  
Astrid Fishstrom, LMSW  
Allen J. Flynn, PharmD, PhD  
Charles P. Friedman, PhD

Professor, Chief Medical Information Officer  
Research Assistant Professor  
MD/PhD Student  
Assistant Professor  
PhD Student  
Senior Research Specialist  
Senior Research Area Specialist  
Assistant Professor  
Josiah Macy, Jr. Professor, Chair of Learning Health Sci.

Paul Grant, MD	Professor, Associate Chief Medical Information Officer
Reema Hamasha, MHI	Research Associate Specialist
Peter Higgins, MD, PhD, MSc	Professor
Ariella Hoffman-Peterson, MSL	PhD Student
Cornelius A. James, MD	Assistant Professor
Erin O'Brien Kaleba, MPH	Director of the Michigan Medicine Data Office
Sharon Kardia, PhD	Professor
Andrew Krumm, PhD	Assistant Professor
Rich Medlin, MD, MSIS	Assistant Professor
Charles Mayo, PhD	Professor
Sean R. Meyer, MBA, PhD	Lead Data Scientist
Bhramar Mukherjee, MS, PhD	Professor and Chair of Biostatistics
Brahmajee Nallamothu, MD, MPH	Professor
Erkin Ötleş, MS	MD/PhD Student
Carleen Penozza, MHSA, BSN, RN-BC	Chief Nursing Informatics Officer
W. Nicholson Price II, JD, PhD	Professor
Michael Sheppard, MSIS	Director of Data Foundations
Kayte Spector-Bagdady, JD, MBe	Co-Director, Ctr. for Bioethics & Social Sciences in Med.
Jeremy Sussman, MD, MS	Associate Professor
Jessica Virzi, MSN	Clinical Informaticist
Akbar Waljee, MD, MSc	Professor
Andrew Wong, MD	Resident Physician

#### **Brigham and Women's Hospital (Boston, MA)**

Emily Alsentzer, PhD	Research Fellow
----------------------	-----------------

#### **Carnegie Mellon University (Pittsburgh, PA)**

Alex John London, PhD	Clara L West Professor of Ethics and Philosophy
-----------------------	---

#### **Cedars-Sinai Medical Center (Los Angeles, CA)**

David Ouyang, MD	Assistant Professor
------------------	---------------------

#### **Columbia University (New York, NY)**

Shalmali Joshi, PhD	Assistant Professor
---------------------	---------------------

#### **Duke University (Durham, NC)**

Mark Sendak, MD, MPP	Population Health and Data Science Lead
----------------------	---

#### **Global Health Research, Management and Solutions (Troy, MI)**

Adnan Hammad, PhD	Adjunct Professor
-------------------	-------------------

#### **Harvard Medical School (Boston, MA)**

Isaac Kohane, MD, PhD	Professor
Lisa S. Lehmann, MD, PhD, MSc	Associate Professor

#### **Icahn School of Medicine at Mount Sinai (New York, NY)**

Girish N. Nadkarni, MD, MPH	Irene and Dr. Arthur M. Fishberg Professor
-----------------------------	--

**KU Leuven (Leuven, Belgium)**

Ben Van Calster, M.Psych, M.Stat Associate Professor

**Leiden University Medical Center (Leiden, the Netherlands)**

Ewout Steyerberg, MSc, PhD Professor

**Maastricht University (Maastricht, The Netherlands)**

Laure Wynants, PhD Assistant Professor

**Mayo Clinic (Rochester, MN)**

Kianoush Kashani, MD, MS Professor

**Memorial Sloan Kettering Cancer Center (New York, NY)**

Andrew Vickers, PhD Attending

**Mozilla Foundation**

Inioluwa Deborah Raji Fellow

**New York University Langone Health (New York, NY)**

Narges Razavian, PhD Assistant Professor

**Oregon Health & Science University (Portland, OR)**

Patrick G Lyons, MD, MSc Assistant Professor

**RTI International (Berkeley, CA)**

Joshua E. Richardson, PhD, MS, MLIS Research Health IT Scientist

**Stanford University (Stanford, CA)**

Nigam Shah, MBBS, PhD Professor

**University of California, San Diego (San Diego, CA)**

Gabriel Wardi, MD, MPH Associate Professor

**University of California, San Francisco (San Francisco, CA)**

Julia Adler-Milstein, PhD Professor

**University College Cork (Cork, Ireland)**

Darren Dahly, PhD Principal Statistician at the Clinical Research Facility

**University College London (London, United Kingdom)**

Pearse Keane, MD, FRCOphth Professor of Artificial Medical Intelligence

**University Medical Center Utrecht (Utrecht, the Netherlands)**

Maarten van Smeden, PhD Associate Professor

**University of Minnesota (Minneapolis, MN)**

Paige Nong, PhD Assistant Professor

**University of Oxford (Oxford, England)**

Gary Collins, PhD Professor

**University of Toronto (Toronto, Canada)**

Melissa McCracken, PhD Assistant Professor

**University of Washington (Seattle, WA)**

Andrew Lees, MD Primary Care Doctor, Clinical Informatics Fellow  
Angad Singh, MD Clinical Assistant Professor, Associate CMIO  
Andrew White, MD Professor, Assistant Chief Medical Information Officer

**Vanderbilt University (Nashville, TN)**

Frank E Harrell Jr, PhD Professor  
Blackford Middleton, MD, MPH, MSc Professor (retired)

**Washington University School of Medicine (St. Louis, MO)**

Philip R.O. Payne, PhD Associate Dean and Chief Data Scientist

**Yale University (New Haven, CT)**

Rohan Khera, MD, MS Assistant Professor

## References

1. Suresh H, Gutttag J. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. Equity and Access in Algorithms, Mechanisms, and Optimization. New York, NY, USA: Association for Computing Machinery; 2021. p. 1–9.
2. Blueprint for an AI bill of rights [Internet]. The White House. 2022 [cited 2023 May 17]. Available from: <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>
3. CHAI [Internet]. [cited 2023 May 17]. Available from: <https://coalitionforhealthai.org>
4. Health AI Partnership [Internet]. Duke Institute for Health Innovation (DIHI); 2019 [cited 2023 May 17]. Available from: <https://healthaipartnership.org>
5. Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proc Natl Acad Sci U S A*. 2020 Jun 9;117(23):12592–12594. PMID: PMC7293650
6. Cao J, Zhang X, Shahinian V, Yin H, Steffick D, Saran R, Crowley S, Mathis M, Nadkarni GN, Heung M, Singh K. Generalizability of an acute kidney injury prediction model across health systems. *Nat Mach Intell*. Springer Science and Business Media LLC; 2022 Dec 1;4(12):1121–1129.
7. Lyons PG, Hofford MR, Yu SC, Michelson AP, Payne PRO, Hough CL, Singh K. Factors associated with variability in the performance of a proprietary sepsis prediction model across 9 networked hospitals in the US. *JAMA Intern Med* [Internet]. 2023 Apr 3; Available from: <http://dx.doi.org/10.1001/jamainternmed.2022.7182> PMID: PMC10071393
8. Discrimination by artificial intelligence in A commercial electronic health record—A case study [Internet]. Forefront Group. Health Affairs (Project Hope); 2020. Available from: <http://dx.doi.org/10.1377/forefront.20200128.626576>
9. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, Pestrue J, Phillips M, Konye J, Penozza C, Ghous M, Singh K. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med*. 2021 Aug 1;181(8):1065–1070. PMID: PMC8218233
10. Ross C. Epic’s sepsis algorithm is going off the rails in the real world. The use of these variables may explain why [Internet]. STAT. 2021 [cited 2023 May 17]. Available from: <https://www.statnews.com/2021/09/27/epic-sepsis-algorithm-antibiotics-model/>
11. Ross C. Epic overhauls popular sepsis algorithm criticized for faulty alarms [Internet]. STAT. 2022 [cited 2023 May 17]. Available from: <https://www.statnews.com/2022/10/03/epic-sepsis-algorithm-revamp-training/>
12. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. *N Engl J Med*. 2020 Aug 27;383(9):874–882. PMID: 32853499
13. 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. PMID: 31649194
14. Executive Office of the President. Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government [Internet]. Federal Register. 2020. p. 78939–78943. Available from: <https://www.federalregister.gov/d/2020-27065>