Duke AI HEALTH

Intentional Governance as a Cornerstone of Trustworthy AI

Michael J. Pencina, PhD

Chief Data Scientist, Duke Health

Professor of Biostatistics and Bioinformatics, Duke University School of Medicine

"Wild West" of Algorithms

"We have a Wild West of algorithms," said Michael Pencina, coalition 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

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

Science

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]*†

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 acce for White patients. Thus by some measures of pr convenient, seemingly e bias in many contexts.

"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…"

There is growing c may reproduce racial and gender disparities via the people building them or through the data used to train them (1–3). Empirical work is increasingly lending support to these concerns. For example, job search ads for highly paid positions are less likely to be presented to women (4), searches for distinctively Black-sounding names are more likely to trigger ads for arrest records (5), and image searches for professions such as CEO produce fewer images of women (6). Facial recognition systems increasingly used

out an algorithm's training data, objective function, and prediction methodology, we can only guess as to the actual mechanisms for the important algorithmic disparities that arise.

In this study, we exploit a rich dataset that provides insight into a live, scaled algorithm deployed nationwide today. It is one of the largest and most typical examples of a class of commercial risk-prediction tools that, by industry estimates, are applied to roughly 200 million people in the United States each year. Large health systems and payers rely on

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 racial disparities in algorithms and isolate the mechanisms by which they arise. It should be emphasized that this algorithm is not unique. Rather, it is emblematic of a generalized approach to risk prediction in the health sector, widely adopted by a range of for- and

get policy interventions under the assumption that the treatment effect is monotonic in that risk, and the methods used to build the algorithm are standard. Mechanisms of bias uncovered in this study likely operate elsewhere. Second, even beyond our particular finding, we hope that this exercise illustrates the importance, and the large opportunity, of studying algorithmic bias in health care, not just as a model system but also in its own right. By any standard—e.g., number of lives affected, life-and-death consequences of the decision—

Duke AI HEALTH

We need to do better

Prediction Models — Development, Evaluation, and Clinical Application

The NEW ENGLAND JOURNAL of MEDICINE

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

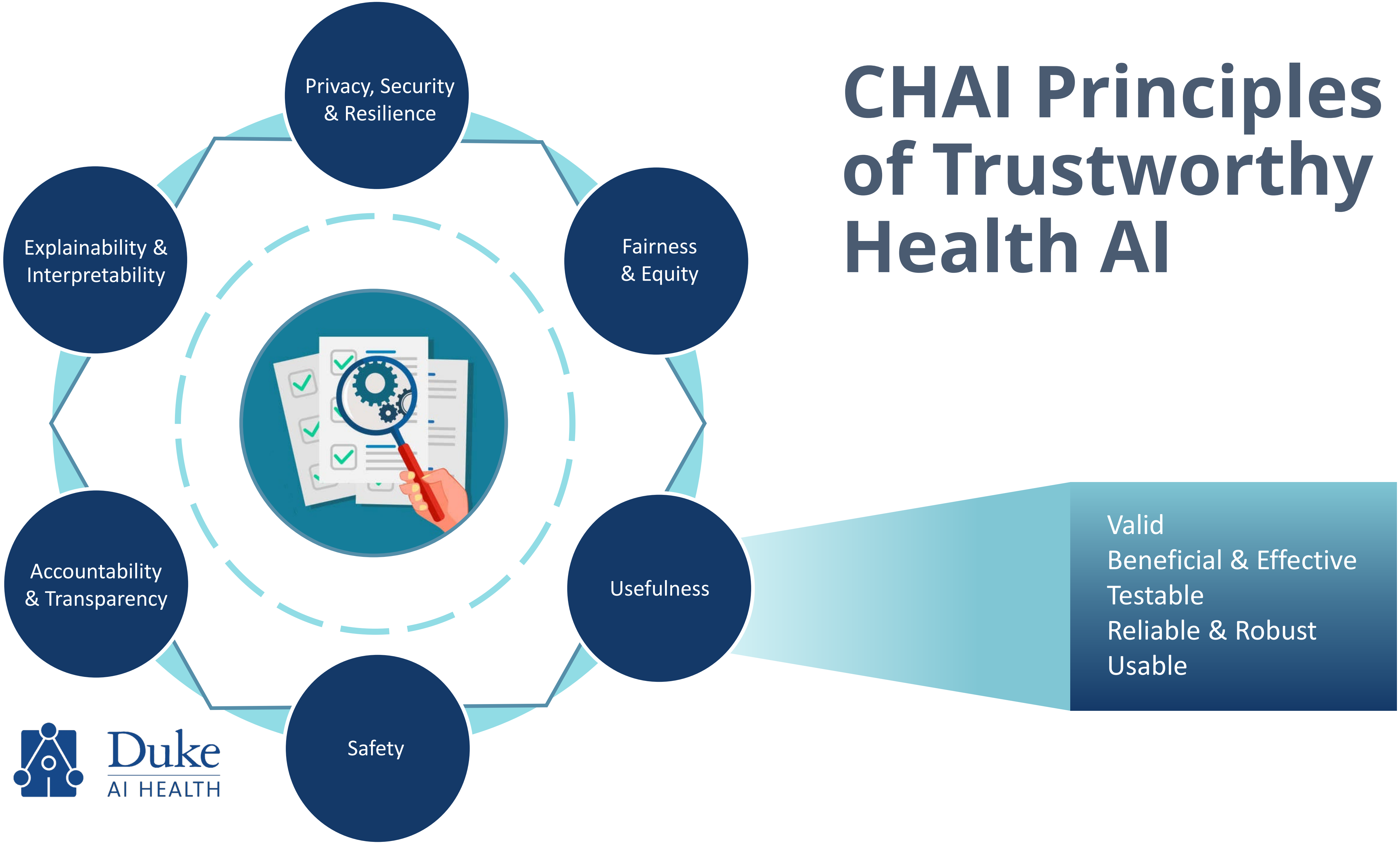
Pencina MJ, Goldstein BA, D'Agostino RB. *N Engl J Med*. 2020 Apr 23;382(17):1583-1586. doi: 10.1056/NEJMp2000589.

Basic Principles of Trustworthy AI

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

Duke AI HEALTH

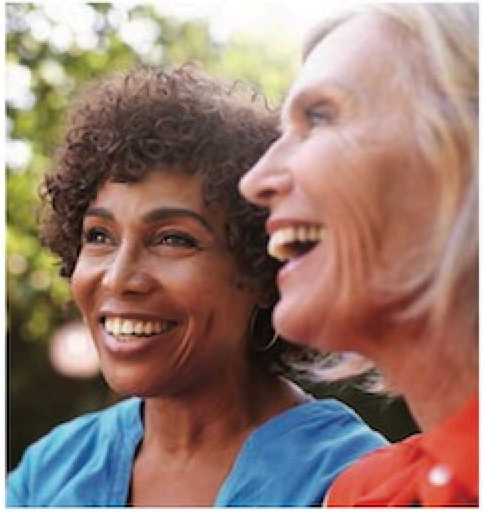
CHAI Principles of Trustworthy Health AI

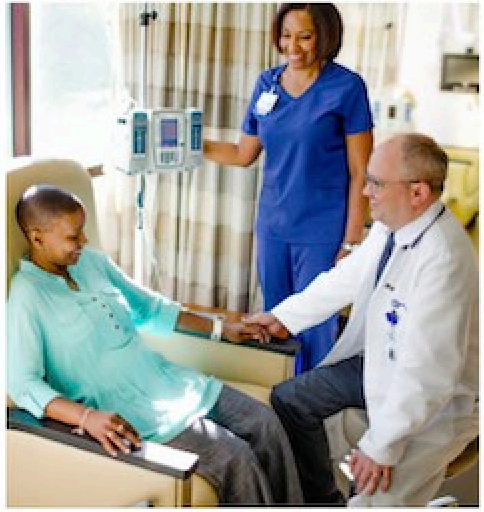


Duke AI HEALTH

Duke ABCDS 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.

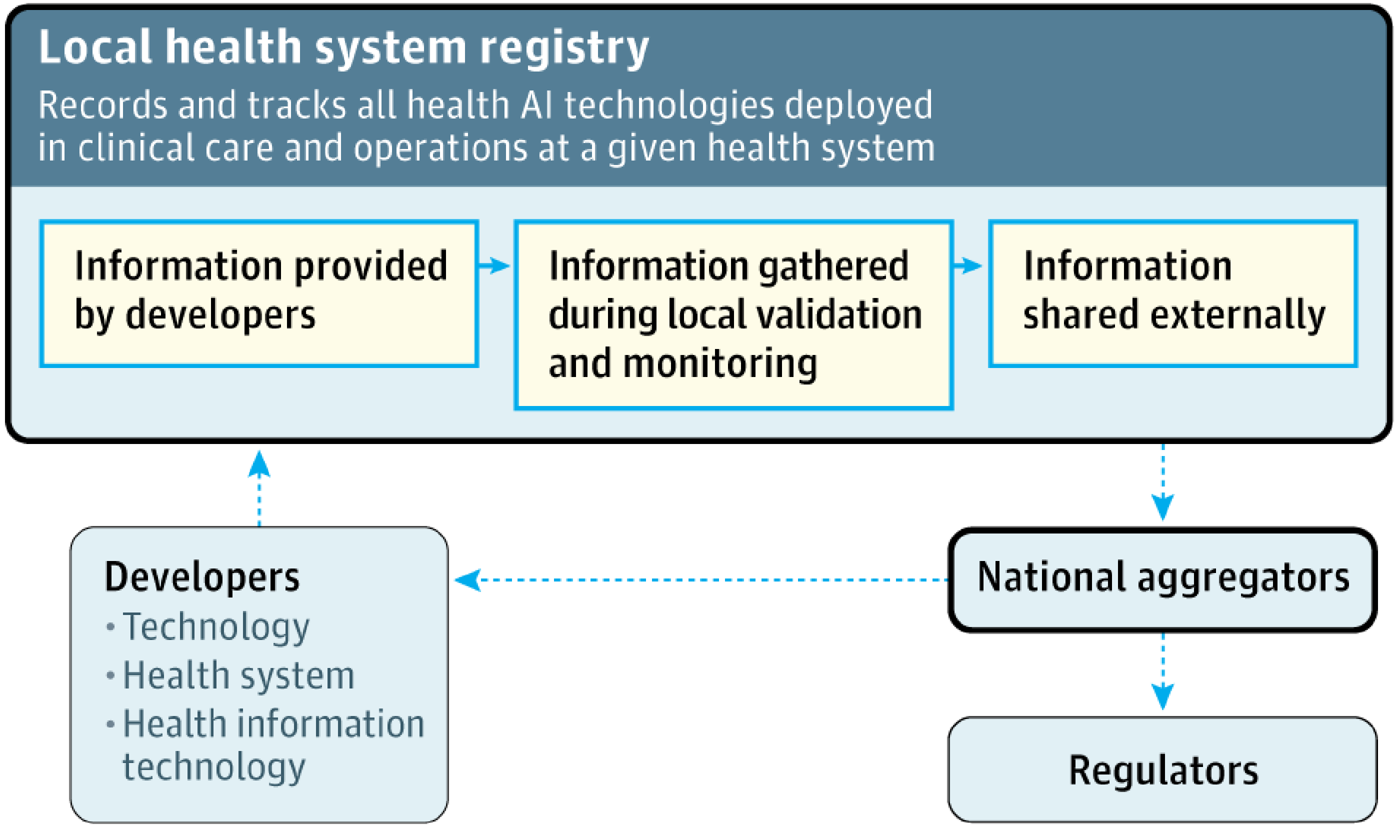




Duke AI HEALTH

From: **A Federated Registration System for Artificial Intelligence in Health**

JAMA. Published online August 12, 2024. doi:10.1001/jama.2024.14026



Federation of Local Health Artificial Intelligence (AI) Registries. Each health system maintains a local health AI registry that captures information from developers as well as from local validation and monitoring and shares it externally with national aggregator(s). Aggregated information can be fed back to health system users and developers and fed forward to regulators (dotted lines).

Risk-Based Process

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

The NEW ENGLAND JOURNAL of MEDICINE

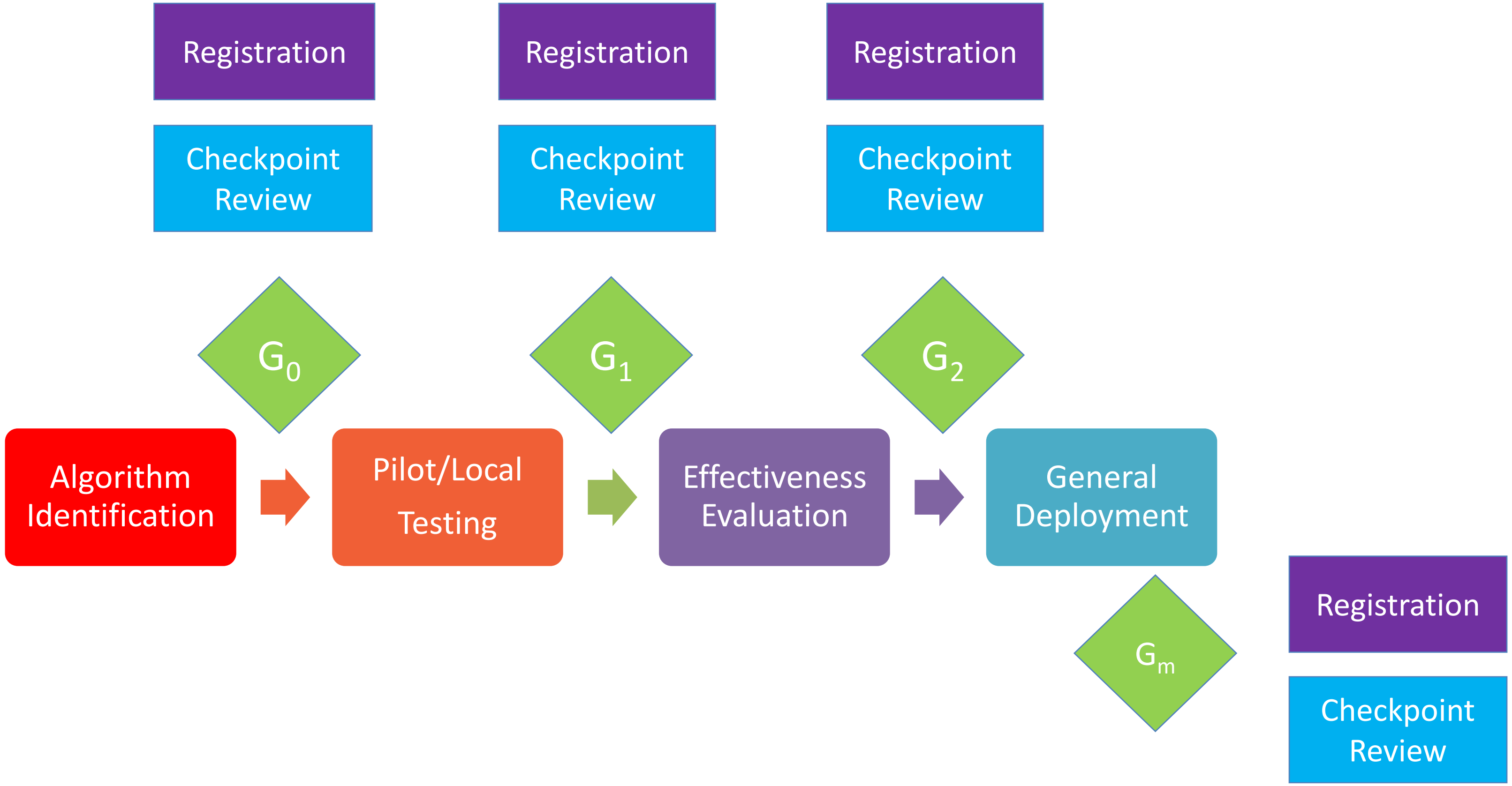
ESTABLISHED IN 1812 NOVEMBER 4, 2021 VOL. 385 NO. 19

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*

Duke AI HEALTH

Lifecycle Management



Duke AI HEALTH