2024 ANNUAL MEETING

Don’t Miss the ASTP Annual Meeting!

Washington, DC | December 4-5

www.ASTPAnnualMeeting.com

ASTP Assistant Secretary for Technology Policy

ASTP Assistant Secretary for Technology Policy

Advancing the Science and Practice of Local AI Evaluation

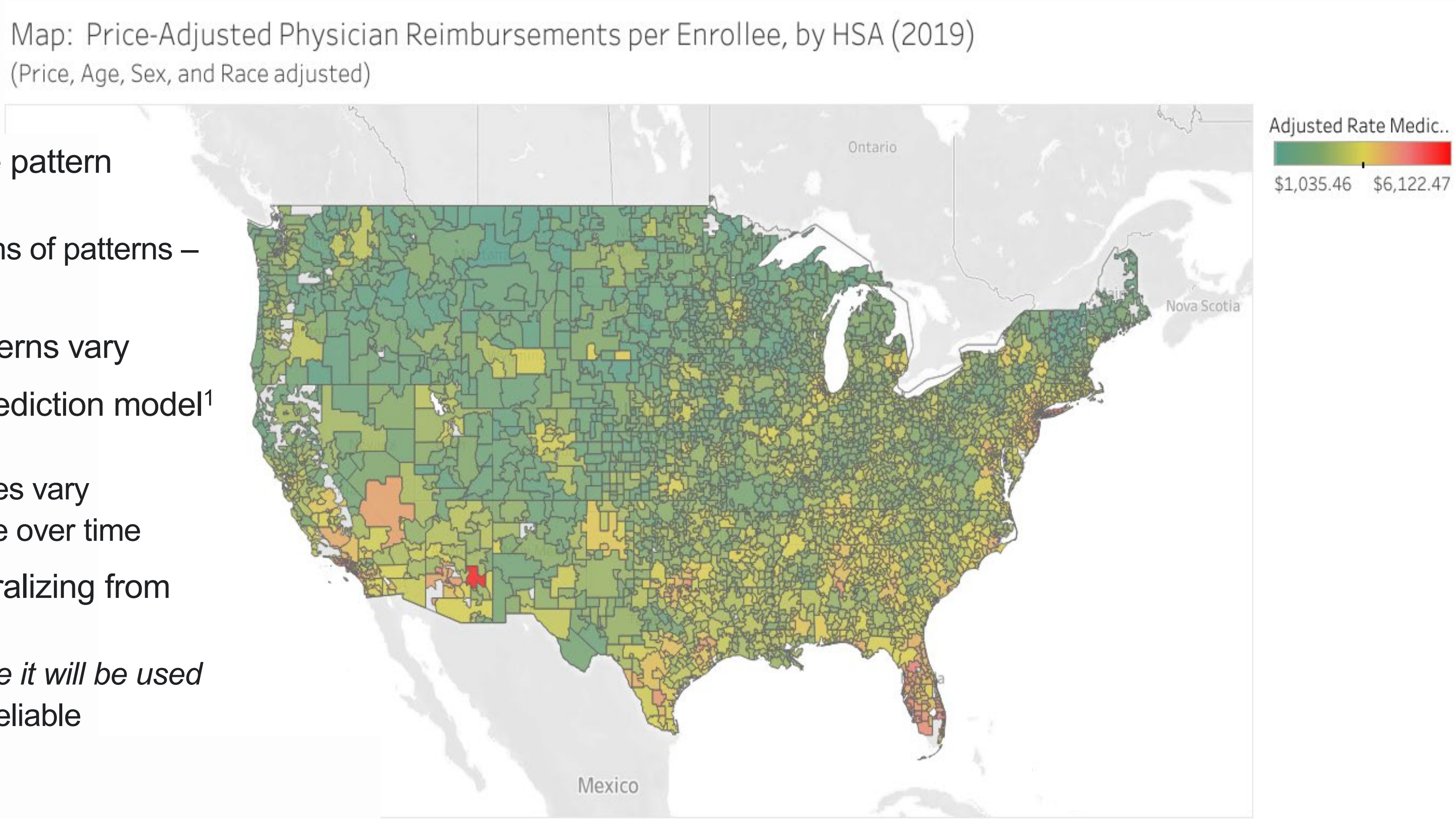
Moderator:
Jordan Everson, PhD, MPP

Panelists:
Peter Embi, MD MS
Corey Miller
Sara Murray, MD, MAS

December 4, 2024

The importance of local evaluation

- Artificial and augmented intelligence are pattern recognizers
  - Make predictions based on simplifications of patterns – including content generation
- Particularly likely to be wrong when patterns vary
- There is no such thing as a validated prediction model[1]
  - Patient populations vary
  - Measurements of predictors or outcomes vary
  - Populations and measurements change over time
- Local evaluation does not rely on generalizing from other sites[2]
  - Allows for pre-deployment testing *where it will be used*
  - Localization and monitoring to ensure reliable performance.



Map: Price-Adjusted Physician Reimbursements per Enrollee, by HSA (2019)

(Price, Age, Sex, and Race adjusted)

---

[1] Van Calster, Ben, et al. "There is no such thing as a validated prediction model." *BMC medicine* 21.1 (2023): 70.

[2] Youssef, Alexey, et al. "External validation of AI models in health should be replaced with recurring local validation." *Nature Medicine* 29.11 (2023): 2686-2687.

HTI-1 Predictive DSI Source Attributes

1 General Description and Outputs

1) Name and contact information for the intervention developer;
2) Funding source of the technical implementation for the intervention(s) development;
3) Description of value that the intervention produces as an output; and
4) Whether the intervention output is a prediction, classification, recommendation, evaluation, analysis, or other type of output.

2 Purpose

5) Intended use of the intervention;
6) Intended patient population(s) for the intervention's use;
7) Intended user(s); and
8) Intended decision-making role for which the intervention was designed to be used/for.

3 Cautioned Out-of-Scope Use

9) Description of tasks, situations, or populations where a user is cautioned against applying the intervention; and
10) Known risks, inappropriate settings, inappropriate uses, or known limitations.

4 Development and Input Features

11) Exclusion and inclusion criteria that influenced the data set;
12) Use of variables in paragraph (b)(11)(iv)(A)(5)-(13) as input features;
13) Description of demographic representativeness including, at a minimum, those used as input features in the intervention;
14) Description of relevance of training data to intended deployed setting;

5 Process used to ensure fairness

15) Description of the approach the intervention developer has taken to ensure that the intervention's output is fair; and
16) Description of approaches to manage, reduce, or eliminate bias.

6 External Validation Process

17) Description of the data source, clinical setting, or environment where an intervention's validity and fairness has been assessed, other than the source of training and testing data
18) Party that conducted the external testing;
19) Description of demographic representativeness of external data including, at a minimum, those used as input features in the intervention;
20) Description of external validation process.

7 Quantitative Measures of Performance

21) Validity of intervention in test data derived from the same source as the initial training data;
22) Fairness of intervention in test data derived from the same source as the initial training data;
23) Validity of intervention in data external to or from a different source than the initial training data;
24) Fairness of intervention in data external to or from a different source than the initial training data;
25) References to evaluation of use of the intervention on outcomes, including, bibliographic citations or hyperlinks to evaluations of how well the intervention reduced morbidity, mortality, length of stay, or other outcomes;

8 Ongoing Maintenance of Intervention

26) Description of process and frequency by which the intervention's validity is monitored over time;
27) Validity of intervention in local data;
28) Description of the process and frequency by which the intervention's fairness is monitored over time.
29) Fairness of intervention in local data; and

9 Validation or Fairness Schedule

30) Description of process and frequency by which the intervention is updated; and
31) Description of frequency by which the intervention's performance is corrected when risks related to validity and fairness are identified.

Prevalence of Local Evaluation

Proportion of hospitals that reported most or all models were evaluated using data from their hospital or health system (n=1,660).		
Model Accuracy	1,009	61%
Model Bias	711	44%
Model Bias and Accuracy	709	44%

Note: 36 hospitals did not indicate whether they evaluated models for accuracy or bias and were excluded from analysis.

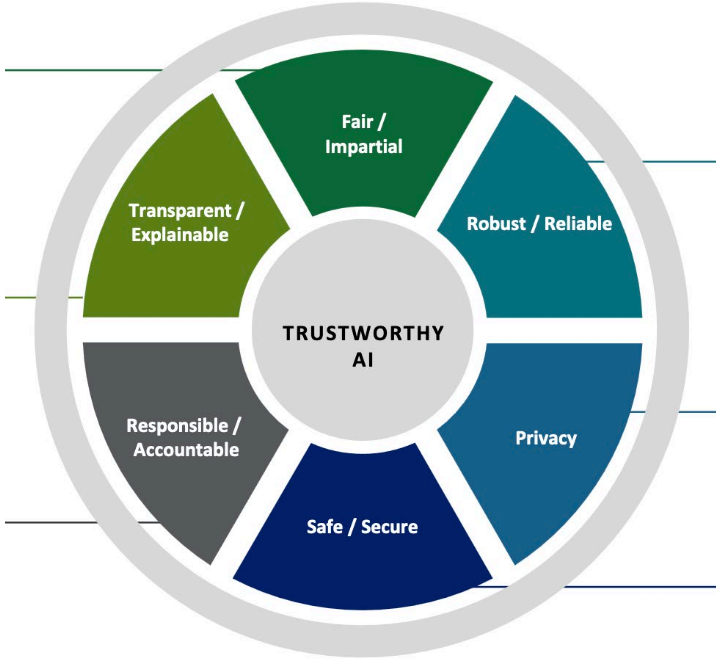
Paige Nong, Julia Adler-Milstein, Nate Apathy, A Jay Holmgren and Jordan Everson “Current Use And Evaluation Of Artificial Intelligence And Predictive Models In US Hospitals” Forthcoming, Health Affairs
Source: 2023 American Hospital Association Information Technology.

ASTP Decision Support Intervention Toolkit

- Recently awarded contract
- Assess needs for a set of tools to facilitate detection of bias in AI models
- Develop tools and share on ASTP web site
- Among other functions, tool will facilitate comparing local data to synthetic data to detect anomalies or potential unique biases in local data.

Local AI Evaluation at UCSF Health

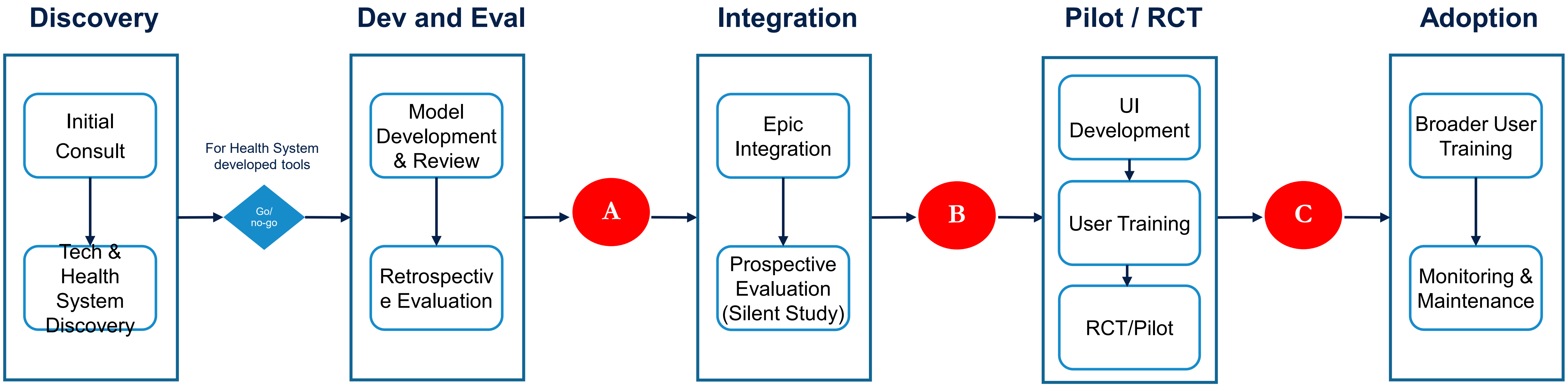
- It is very difficult for health systems to know when AI is ”trustworthy”
  - Limited regulation and standards for vendors
  - Healthcare delivery systems do not innately have AI assessment capabilities
- Early lessons in health AI use highlighted gaps across mission areas
  - Unreliable and/or biased vendor tools
  - Research tools deployed haphazardly without guardrails



https://www.hhs.gov/sites/default/files/hhs-trustworthy-ai-playbook.pdf

AI Oversight Across the AI Lifecycle

- Health AI Oversight Committee ensures all AI tools implemented in the health system are “trustworthy”
- Diverse, multidisciplinary committee with broad expertise
- Scope includes locally developed, vendor procured, and research tools



UCSF IMPACC
(Impact Monitoring Platform for AI in Clinical Care)

IMPACC = AI Monitoring Infrastructure + Robust Adjudication Process

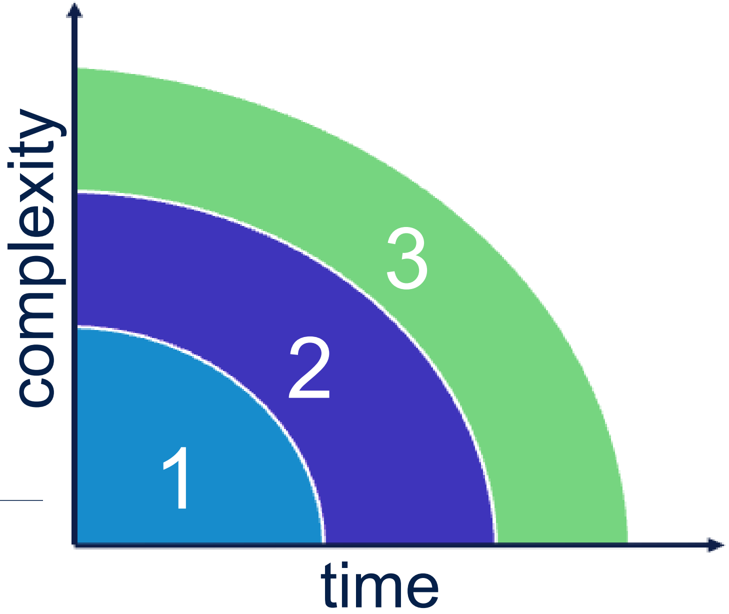
- Unique partnership bridging health system and campus/academic expertise
- Generalizable monitoring infrastructure for all enterprise AI tools deployed at UCSF
  - Implementation and use, algorithmic vigilance, KPIs and outcomes

3 Horizons of IMPACC AI Monitoring

Horizon 1: Basic Metrics

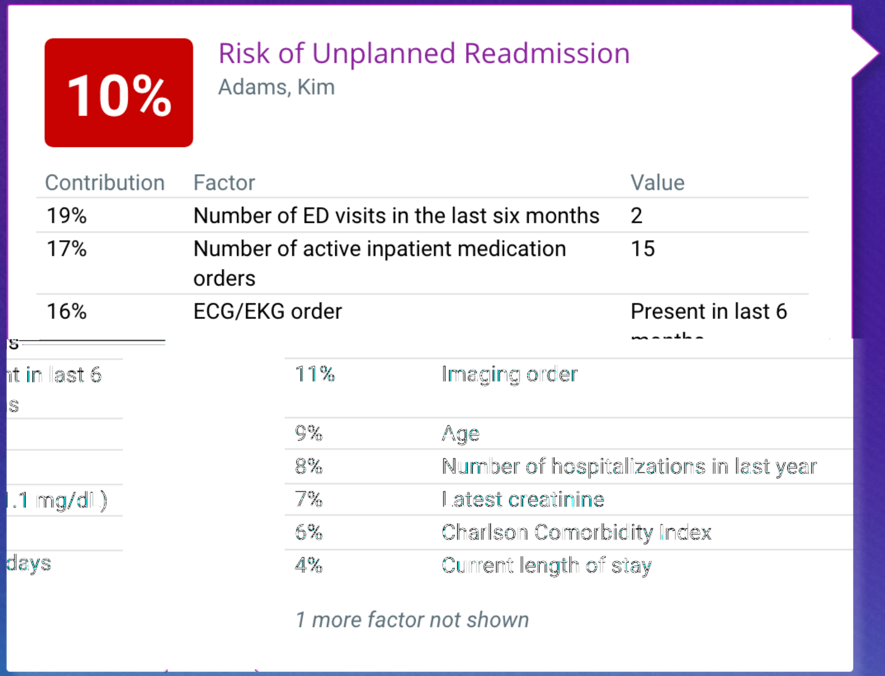
Horizon 2: Advanced Insights

Horizon 3: Open Research Questions



AI @ Epic

Predictive AI

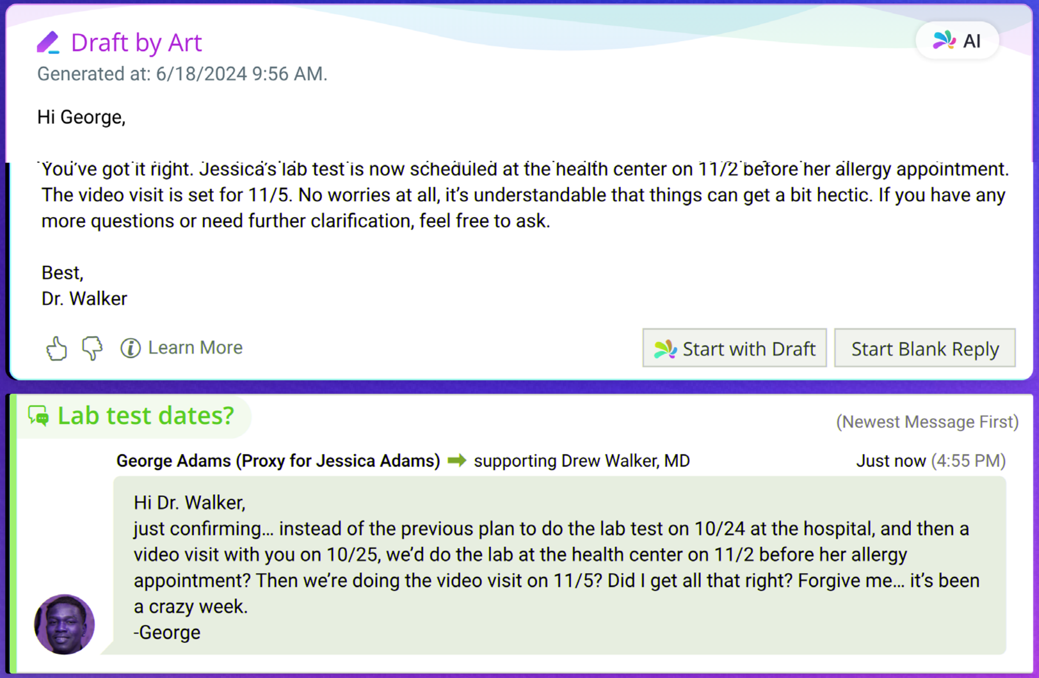


10% Risk of Unplanned Readmission
Adams, Kim

Contribution	Factor	Value
19%	Number of ED visits in the last six months	2
17%	Number of active inpatient medication orders	15
16%	ECG/EKG order	Present in last 6
11%	Imaging order	t in last 6 s
9%	Age	
8%	Number of hospitalizations in last year	
7%	Latest creatinine	.1 mg/dl)
6%	Charlson Comorbidity Index	
4%	Current length of stay	days

1 more factor not shown

Generative AI



Draft by Art
Generated at: 6/18/2024 9:56 AM.

Hi George,

You've got it right. Jessica's lab test is now scheduled at the health center on 11/2 before her allergy appointment. The video visit is set for 11/5. No worries at all, it's understandable that things can get a bit hectic. If you have any more questions or need further clarification, feel free to ask.

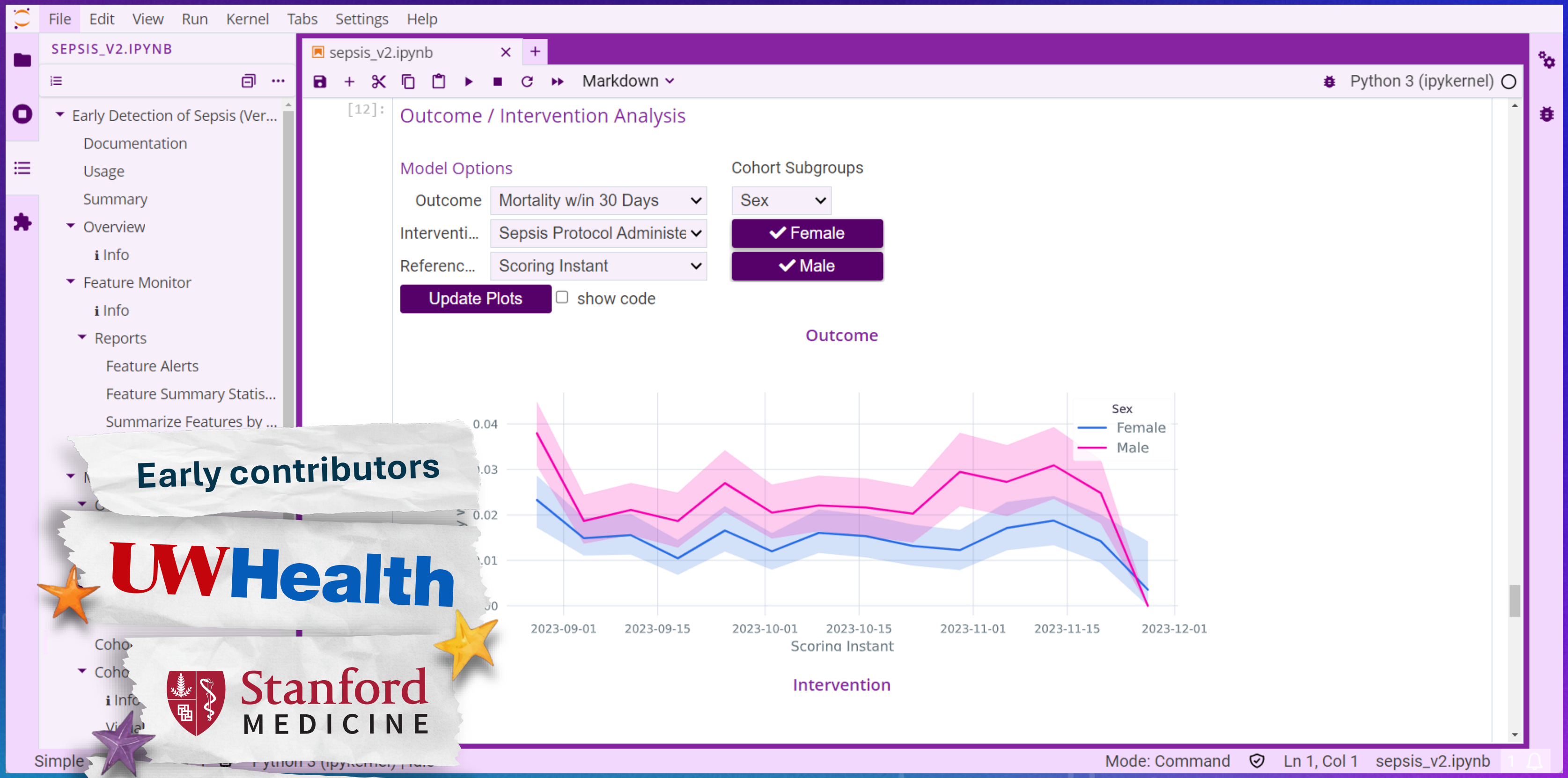
Best,
Dr. Walker

Lab test dates?
(Newest Message First)

George Adams (Proxy for Jessica Adams) ➡ supporting Drew Walker, MD — Just now (4:55 PM)

Hi Dr. Walker,
just confirming... instead of the previous plan to do the lab test on 10/24 at the hospital, and then a video visit with you on 10/25, we'd do the lab at the health center on 11/2 before her allergy appointment? Then we're doing the video visit on 11/5? Did I get all that right? Forgive me... it's been a crazy week.
-George

AI Trust & Assurance Suite



Outcome / Intervention Analysis

Model Options

Outcome: Mortality w/in 30 Days

Interventi...: Sepsis Protocol Administe

Referenc...: Scoring Instant

Update Plots ☐ show code

Cohort Subgroups

Sex

✓ Female

✓ Male

Early contributors

UWHealth

Stanford MEDICINE

UC San Diego

- Implements major 3rd-party standards
- Performance on local patient mix
- Real-time, ongoing monitoring
- Open-source template & schema

Epic

Enabling Algorithmovigilance for Safe, Effective, & Equitable Health AI

Peter J. Embí, MD, MS, FACP, FACMI, FIAHSI

Professor of Biomedical Informatics and Medicine
Chair, Department of Biomedical Informatics
Endowed Directorship in Biomedical Informatics
Co-Director, ADVANCE AI Center
Senior Vice-President for Research & Innovation

ASTP/ONC Annual Meeting

December 4, 2024

VANDERBILT UNIVERSITY
MEDICAL CENTER

AI Discovery & Vigilance to Accelerate Innovation & Clinical Excellence

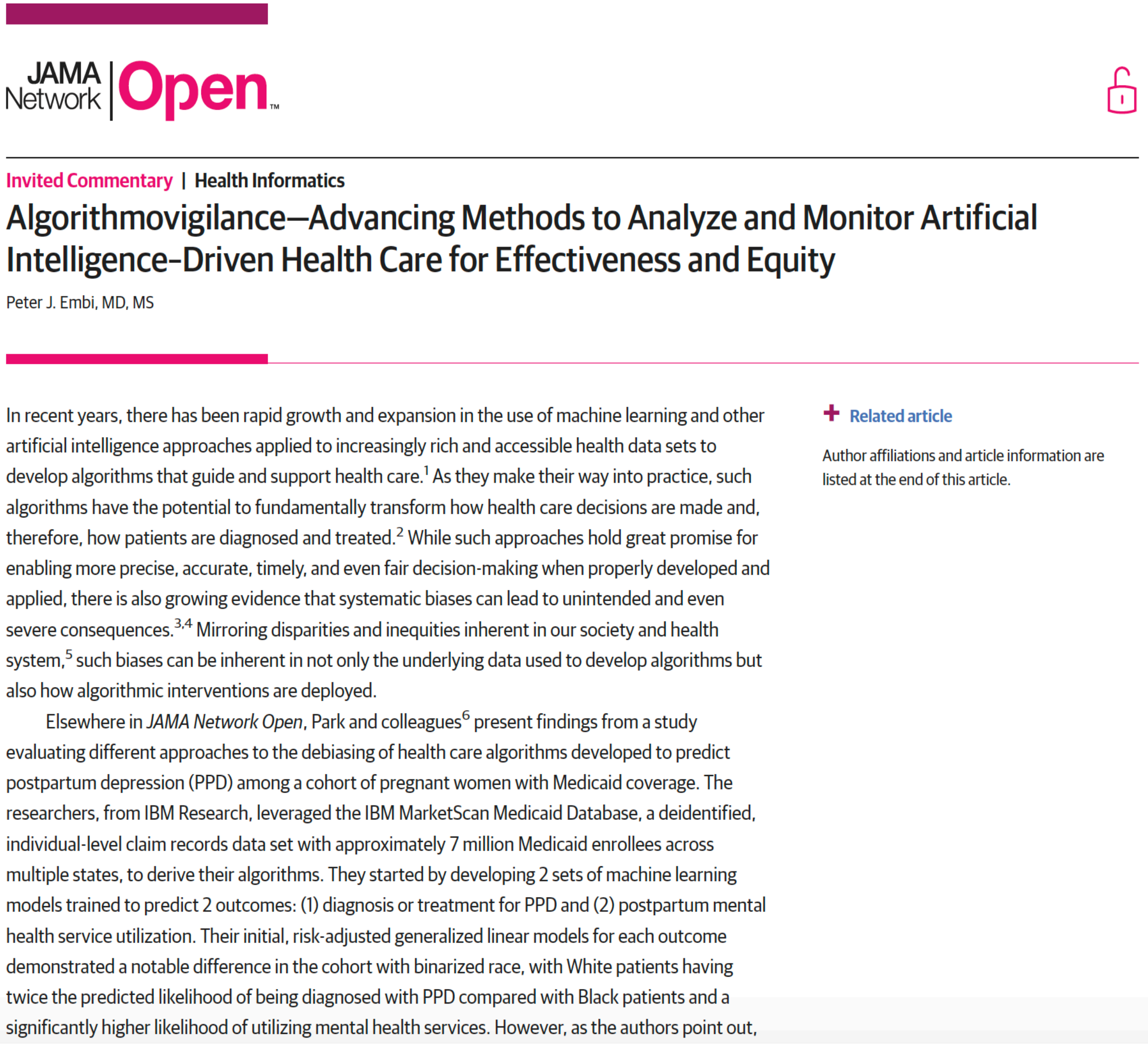
ADVANCE

”Algorithmovigilance”

“The scientific methods and activities relating to the evaluation, monitoring, understanding, and prevention of adverse effects of algorithms in health care.”

Akin to pharmacovigilance for monitoring drug effects

Increasingly important as AI/ML-derived algorithms are used



JAMA Network Open

Invited Commentary | Health Informatics

Algorithmovigilance—Advancing Methods to Analyze and Monitor Artificial Intelligence–Driven Health Care for Effectiveness and Equity

Peter J. Embi, MD, MS

In recent years, there has been rapid growth and expansion in the use of machine learning and other artificial intelligence approaches applied to increasingly rich and accessible health data sets to develop algorithms that guide and support health care.[1] As they make their way into practice, such algorithms have the potential to fundamentally transform how health care decisions are made and, therefore, how patients are diagnosed and treated.[2] While such approaches hold great promise for enabling more precise, accurate, timely, and even fair decision-making when properly developed and applied, there is also growing evidence that systematic biases can lead to unintended and even severe consequences.[3,4] Mirroring disparities and inequities inherent in our society and health system,[5] such biases can be inherent in not only the underlying data used to develop algorithms but also how algorithmic interventions are deployed.

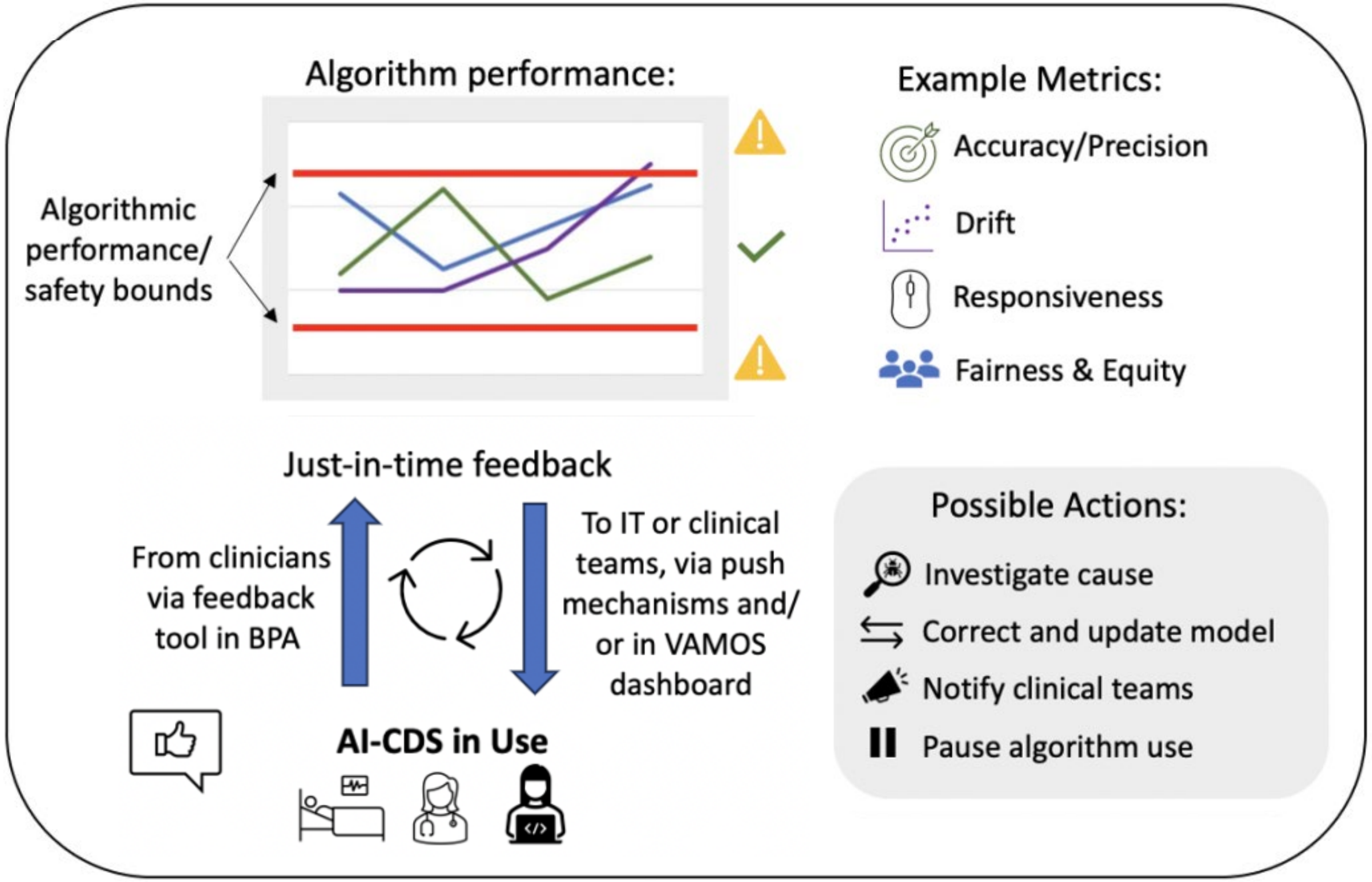
Elsewhere in *JAMA Network Open*, Park and colleagues[6] present findings from a study evaluating different approaches to the debiasing of health care algorithms developed to predict postpartum depression (PPD) among a cohort of pregnant women with Medicaid coverage. The researchers, from IBM Research, leveraged the IBM MarketScan Medicaid Database, a deidentified, individual-level claim records data set with approximately 7 million Medicaid enrollees across multiple states, to derive their algorithms. They started by developing 2 sets of machine learning models trained to predict 2 outcomes: (1) diagnosis or treatment for PPD and (2) postpartum mental health service utilization. Their initial, risk-adjusted generalized linear models for each outcome demonstrated a notable difference in the cohort with binarized race, with White patients having twice the predicted likelihood of being diagnosed with PPD compared with Black patients and a significantly higher likelihood of utilizing mental health services. However, as the authors point out,

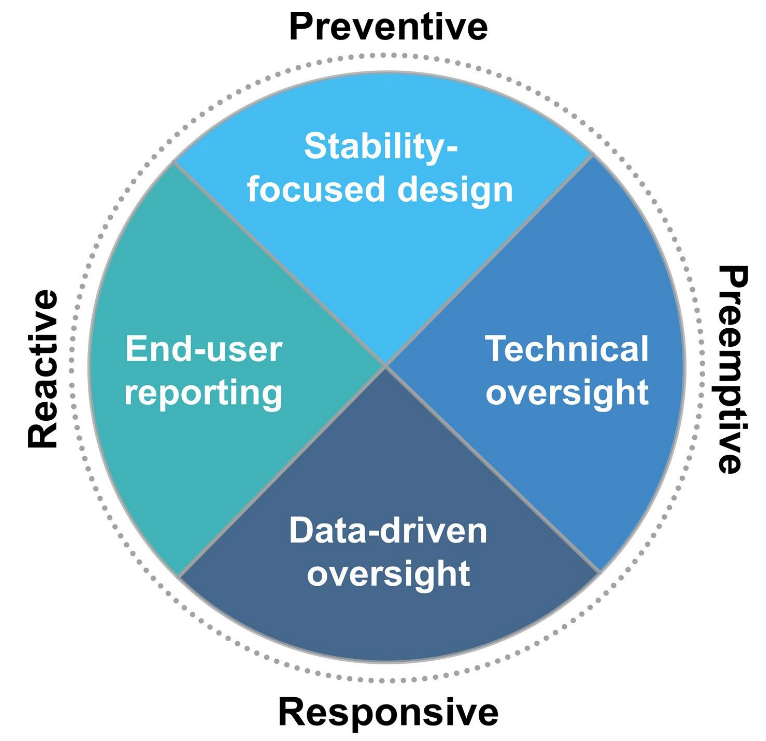
+ Related article

Author affiliations and article information are listed at the end of this article.

Embi PJ. JAMA Network Open. 2021;4(4):e214622.

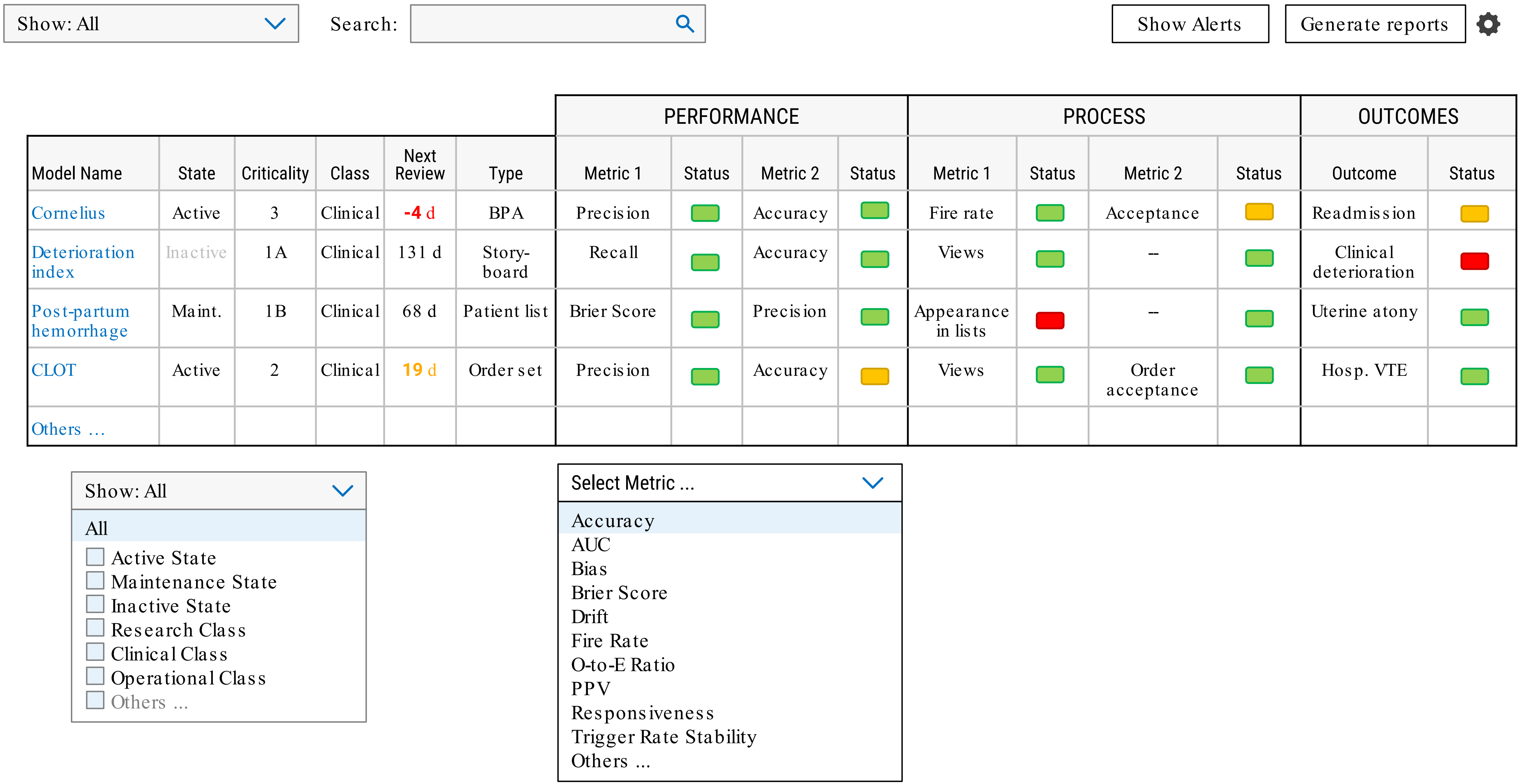
The Vanderbilt Algorithmovigilance Monitoring and Operations System (VAMOS)





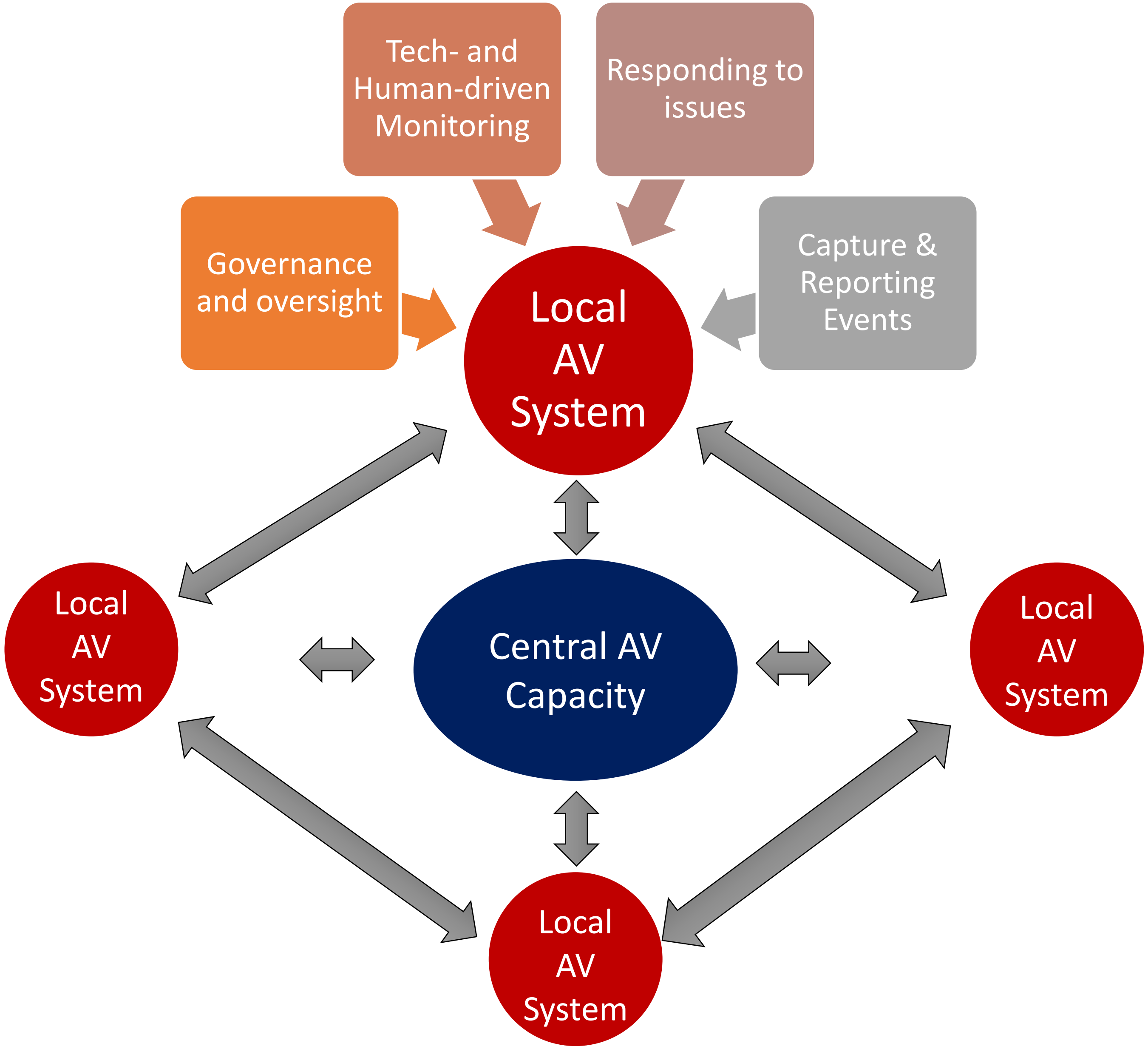
Sustainable deployment of clinical prediction tools—a 360° approach to model maintenance
Davis S, Embi PJ, Matheny M. J Am Med Inform Assoc, May 2024

VAMOS DASHBOARD MOCK-UP



						PERFORMANCE				PROCESS				OUTCOMES	
Model Name	State	Criticality	Class	Next Review	Type	Metric 1	Status	Metric 2	Status	Metric 1	Status	Metric 2	Status	Outcome	Status
Cornelius	Active	3	Clinical	-4 d	BPA	Precision	🟩	Accuracy	🟩	Fire rate	🟩	Acceptance	🟨	Readmission	🟨
Deterioration index	Inactive	1A	Clinical	131 d	Storyboard	Recall	🟩	Accuracy	🟩	Views	🟩	--	🟩	Clinical deterioration	🟥
Post-partum hemorrhage	Maint.	1B	Clinical	68 d	Patient list	Brier Score	🟩	Precision	🟩	Appearance in lists	🟥	--	🟩	Uterine atony	🟩
CLOT	Active	2	Clinical	19 d	Order set	Precision	🟩	Accuracy	🟨	Views	🟩	Order acceptance	🟩	Hosp. VTE	🟩
Others …															

Creating Local and Federated Algorithmovigilance (AV) Systems



Thank You!

Business Card

Questions or Comments?

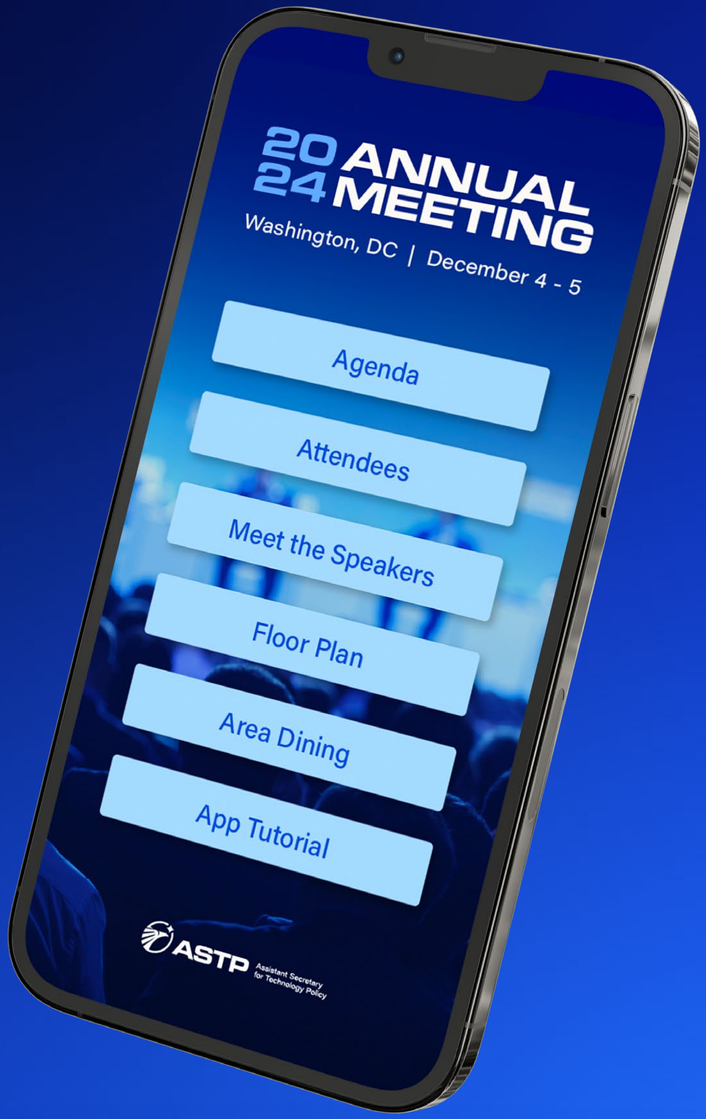
@embimd.bsky.social

peter.embi@vumc.org

2024 ANNUAL MEETING

Stay up-to-date on the agenda with the Annual Meeting app!

Download **ShoApp** in the App Store or Google Play.



2024 ANNUAL MEETING

Washington, DC | December 4 - 5

20 ANNUAL
24 MEETING

𝕏 @HHS_TechPolicy

Share your content on X and don't forget to use the hashtag **#ASTP2024**

Today’s Agenda

December 4th

9:00 am – 11:30 am	Keynote Remarks from Micky Tripathi Morning Plenary: TEFCA - Year One in the Books and Looking to the Future
11:30 am – 1:00 pm	Lunch on your own
1:00 pm – 2:00 pm	Breakout Sessions I – *View the ASTP Annual Meeting app for details*
2:00 pm – 2:15 pm	Break
2:15 pm – 3:15 pm	Afternoon Plenary: Collaboration, Harmonization, and Standardization: How USCDI+ is Raising the Floor for Interoperable Data Use
3:15 pm – 3:45 pm	Break
3:45 pm – 5:00 pm	Breakout Sessions II – *View the ASTP Annual Meeting app for details*

2024 ANNUAL MEETING

Tomorrow’s Agenda

December 5th

9:25 am – 11:30 am	Morning Plenaries: - What the GPT? Does AI Have a Place in Health Care Delivery? - Meeting the Mission with AI: How HHS is Using AI to Advance Health and Human Services
11:30 am – 1:00 pm	Lunch on your own
1:00 pm – 2:00 pm	Breakout Sessions III – *View the ASTP Annual Meeting app for details*
2:00 pm – 2:15 pm	Break
2:15 pm – 3:15 pm	Breakout Sessions IV – *View the ASTP Annual Meeting app for details*

2024 ANNUAL MEETING

Today's Agenda

December 5th

9:25 am – 11:30 am	Morning Plenaries: - What the GPT? Does AI Have a Place in Health Care Delivery? - Meeting the Mission with AI: How HHS is Using AI to Advance Health and Human Services
11:30 am – 1:00 pm	Lunch on your own
1:00 pm – 2:00 pm	Breakout Sessions III – *View the ASTP Annual Meeting app for details*
2:00 pm – 2:15 pm	Break
2:15 pm – 3:15 pm	Breakout Sessions IV – *View the ASTP Annual Meeting app for details*