DUKE University School of Medicine | DUKE Institute for Health Innovation

Ensuring the Safe, Effective, and Equitable Translation of AI/ML Into Clinical Practice

Mark Sendak, MD, MPP
Population Health & Data Science Lead, Duke Institute for Health Innovation
Co-Lead, Health AI Partnership

Suresh Balu, MBA, MS
Director, Duke Institute for Health Innovation
Associate Dean for Innovation and Partnerships, Duke School of Medicine
Co-Lead, Health AI Partnership

January 2024

Duke Institute for Health Innovation	**2 mins**
Health AI Partnership	**2 mins**
Safe, Effective, and Equitable AI Translation	**20 mins**
Health Equity Across the AI Lifecycle (HEAAL)	**8 mins**

Duke Institute for Health Innovation	**2 mins**
Health AI Partnership	**2 mins**
Safe, Effective, and Equitable AI Translation	**20 mins**
Health Equity Across the AI Lifecycle (HEAAL)	**8 mins**

Duke Institute for Health Innovation

Our Mission: **Catalyze innovations at Duke**

Catalyze transformative innovation in health and healthcare through high-impact research, leadership development and workforce training and the cultivation of a community of entrepreneurship

Our Approach: **Innovation by design**

Understand user workflow, desired outcomes and problems (needs) and then collaboratively develop concepts and prototypes, and iterate through to finalize solution

DIHI domains of innovation

DIHI

Duke Institute for Health Innovation

Implementation and Health Delivery Science

- Catalyze multidisciplinary teamwork
- New care models
- Structured interface to Duke Health
- Living laboratory to incubate, refine, validate, and scale new ideas

Health Technology Innovation

- Leverage a growing health data infrastructure
- Create a connected digital health ecosystem
- Collaboration and co-development of technology
- Responsible development of datascience solutions

Leadership and Workforce Development

- Train current and future leaders across health care :
  - Leadership
  - Management
  - Innovation
  - Quantitative health sciences
- Contribute to developing the workforce of the future

Best Practices Development and Dissemination

- Disseminate best practices derived through internal R&D to elevate health innovation across ecosystem
- Convene stakeholders across settings to address common challenges in health innovation

Industry best-practice approach in catalyzing innovation

RFA

Structured

DIHI RFA approach

"Top-down + Bottom-Up" approach to sourcing innovations

- Duke Health leadership develops mission-aligned strategic themes for innovation
- Front-line faculty and staff propose "problems" aligned with strategic themes and novel solutions
- Systematic review and due diligence: Assessments on team, feasibility, resource needs, impact and value to patients
- Operational Lead engaged right from the proposal stage
- 8-12 innovations funded each year; Duration: 12-15 months
- DIHI members embedded within project innovation teams to rapidly catalyze the innovations
- Pivots as needed to support rapid evolution to create value
- Metrics: clinical utility, economic utility, cultural impact, IP and academic outputs

11 Years Catalyzing Innovations

90+ Innovation Projects

740+ Proposals

IJ

Unstructured

DIHI Innovation Jam

A Health focused Shark Tank at Duke

- Solicits and identifies high-potential healthcare and health **innovations ready for commercialization**
- **Duke Leadership as Sharks:**
  - DUHS leaders, Department Chairs, Deans of School of Medicine, Nursing, Engineering, OLV, I&E, MedBlue, Center and Institute Directors
- Innovation proposals from students, faculty, trainees and staff across campus
- **Funding** to support entrepreneurship / **formation of company** and also **develop the product/service** etc.
- Inventors offer portion of their share of Duke internal returns for investment from the sharks
- Internal syndicated investment agreements documented through MOUs.

6 Years of Jamming

30+ Pitches

12 Companies Incubated

Duke Institute
for Health Innovation

RFA 2024

DIHI

We invite you to submit your novel ideas supporting

Generative AI & Large Language Models: AI solutions to improve staff and clinician efficiency, patient journey and outcomes

Visit: dihi.org/events/rfa

email: dihi-rfa@duke.edu

@dukeinnovate

Proposals due: NOVEMBER 3, 2023

DIHI Spectrum of Value Creation

Inpatient Innovations	Transition Setting	Outpatient/ Gaps in Care	Patient & Community
Hospital at Home		HIV Pre-Exposure Prophylaxis Identification	Community COVID-19 Support
Mortality Models (inpatient / 30-day)	High-utilizer dashboard	CKD Patient Education Dissemination	Patient Reported Outcomes for Cancer Patients
Operational Enhancement	Complex Care Plans	Community-Based Palliative Care	Outpatient Procedure Concierge Program
Procedure Safety	Index Admissions with MSSP	High Value Analyte Ordering	Cancer Distress Coach
Medication Safety	Readmissions (Social Drivers for HF)	NAFLD population health rounding	Autism and Beyond
Early Detection of Deterioration	SNF transition	CKD population health rounding	Voices of Duke

Immersion in innovation and data science

- **Medical Students Scholarship**
- **Data Science in Health** masters course in BME
- **Summer Fellowship** in Data Science
- **Case Studies** and Data Camp
- **Journal Club**

Technology Infrastructure | Research and Dissemination | Education and Training

Duke Institute for Health Innovation [DIHI] – Spectrum of value creation across the ecosystem

Duke Institute for Health Innovation	**2 mins**
Health AI Partnership	**2 mins**
Safe, Effective, and Equitable AI Translation	**20 mins**
Health Equity Across the AI Lifecycle (HEAAL)	**8 mins**

Corps Sites

health ai partnership

AI INITIATIVE

Franciscan Missionaries of Our Lady HEALTH SYSTEM

KAISER PERMANENTE®

OCHIN

Berkeley UNIVERSITY OF CALIFORNIA

AMA

GORDON AND BETTY MOORE FOUNDATION

MAYO CLINIC

ONC Office of the National Coordinator for Health Information Technology
*Participating as a federal observer

UCSF University of California San Francisco

DLA PIPER

Hackensack Meridian Health

MICHIGAN MEDICINE UNIVERSITY OF MICHIGAN

PCCI

UWHealth

DukeHealth

Jefferson Thomas Jefferson University

NewYork-Presbyterian

scan HEALTH PLAN

VA

Website healthaipartnership.org

Our Mission: Empowering healthcare professionals to use AI effectively, safely, and equitably through **community-informed up-to-date standards**

Our Values

advance health equity

prioritize solutions that advance health equity and eliminate the AI digital divide

improve patient care

ensure that AI adoption is driven by patient care needs, not technical novelty

improve the workplace

surface socio-technical challenges in AI use and foster a positive work environment

build community

create safe spaces to share learnings and consult peers

Phase One (Apr 22 – Aug 23) Milestones

Standard AI Solution Procurement Milestones

- Community-informed best practices sourced from across the network of organizations
- Multiple co-design workshops with IDEO.org
- Focused on AI solutions used for:
  - Diagnosis or treatment decisions for individual patients
  - Prioritization of patients for healthcare services (e.g., surgery scheduling, care management prioritization, ED triaging)

8 Key Decision Points

85+ Interviews

31 Topic Guides

Health Equity Across the AI Lifecycle (HEAAL) Framework

- Developed to answer the question: "our health system is considering adopting a new solution that uses AI; how do we assess the potential future impact on health inequities?"
- Convened multi-stakeholder workshop featuring case studies, expert discussants, and framework developers
- Developed detailed procedures for healthcare organizations to follow for AI procurement

3 Case Studies

75+ Participants

37 Procedures

Duke Institute for Health Innovation	**2 mins**
Health AI Partnership	**2 mins**
Safe, Effective, and Equitable AI Translation	**20 mins**
Health Equity Across the AI Lifecycle (HEAAL)	**8 mins**

8 Key Decision Points in AI Adoption Process

Procurement

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution

Development & Adaptation

3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product

Clinical Integration

6. Execute change management, workflow integration, and scaling strategy

Lifecycle Management

7. Monitor and maintain the AI product
8. Update or decommission the AI product

8 Key Decision Points in AI Adoption Process

Procurement

1. **Identify and prioritize a problem**
2. Evaluate AI as a viable component of the solution

Development & Adaptation

3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product

Clinical Integration

6. Execute change management, workflow integration, and scaling strategy

Lifecycle Management

7. Monitor and maintain the AI product
8. Update or decommission the AI product

Align Front-Line Staff and Organizational Leaders

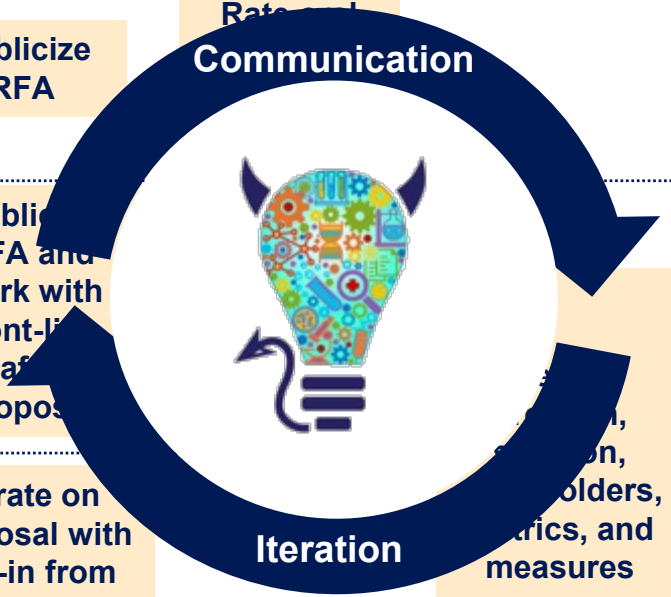
Create Alignment Throughout Project Selection

	Strategic Priorities	Proposal Solicitation	Project Selection	Scoping	Solution Development	Implementation & Evaluation
C-Suite	Specify 4-5 organizational innovation priorities	Publicize RFA	Rate oral pitches to select ~10 projects			Evaluate and determine impact and sustainability plans
Clinical and Operational Leaders (control resources)		Publicize RFA and work with front-line staff on proposal	Provide written reviews to select ~20 projects for oral pitches	Define problem, solution, stakeholders, metrics, and measures	Design and develop solution with DIHI	Define operationalization and dissemination, including communication and training
Front-Line Staff		Iterate on proposal with buy-in from clinical and operational leaders				

Align Front-Line Staff and Organizational Leaders

Create Alignment Throughout Project Selection

	Strategic Priorities	Proposal Solicitation	Project Selection	Scoping	Solution Development	Implementation & Evaluation
C-Suite	Specify 4-5 organizational innovation priorities	Publicize RFA	Rate and			Evaluate and determine impact and sustainability plans
Clinical and Operational Leaders (control resources)		Publi RFA and work with front-li staf propos				
Front-Line Staff		Iterate on proposal with buy-in from clinical and operational leaders		, n, olders, rics, and measures	Design and develop solution with DIHI	Define operationalization and dissemination, including communication and training



8 Key Decision Points in AI Adoption Process

Procurement

1. Identify and prioritize a problem
2. **Evaluate AI as a viable component of the solution**

Development & Adaptation

3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product

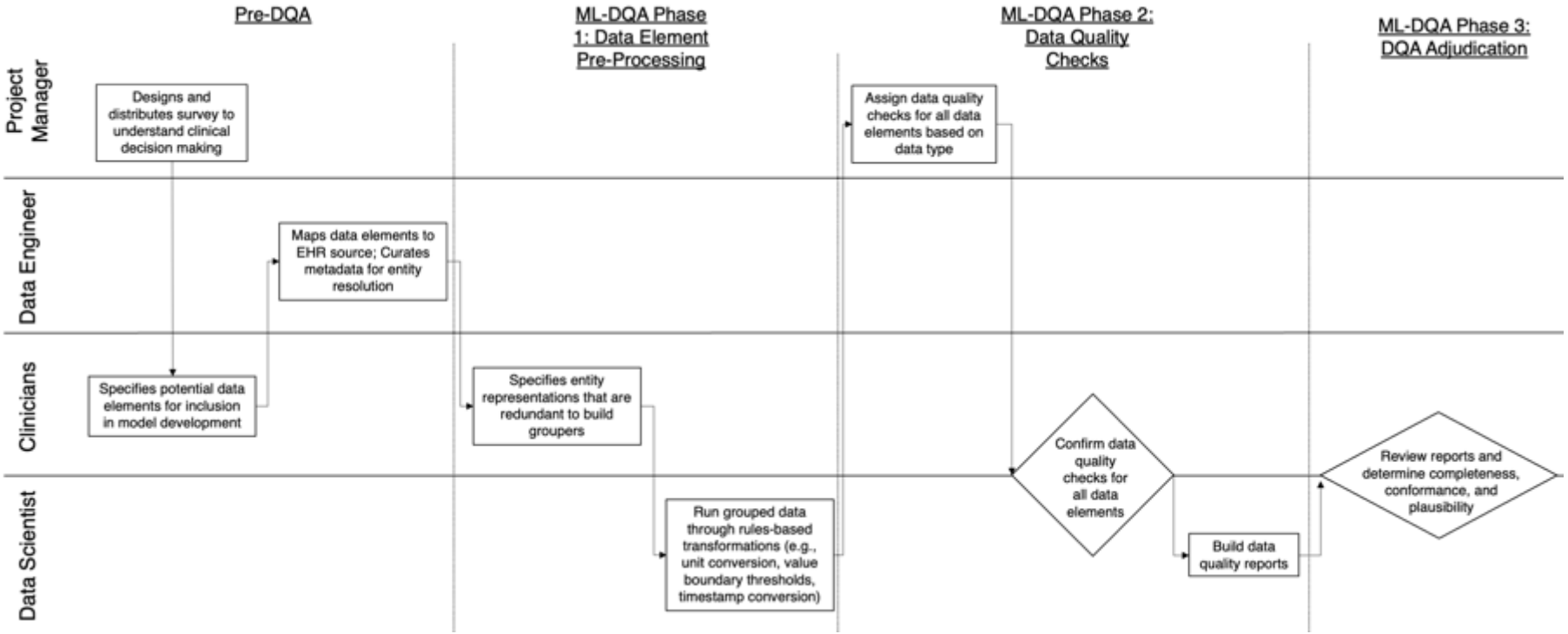
Clinical Integration

6. Execute change management, workflow integration, and scaling strategy

Lifecycle Management

7. Monitor and maintain the AI product
8. Update or decommission the AI product

ML Data Quality Assurance for Healthcare



https://proceedings.mlr.press/v182/sendak22a.html

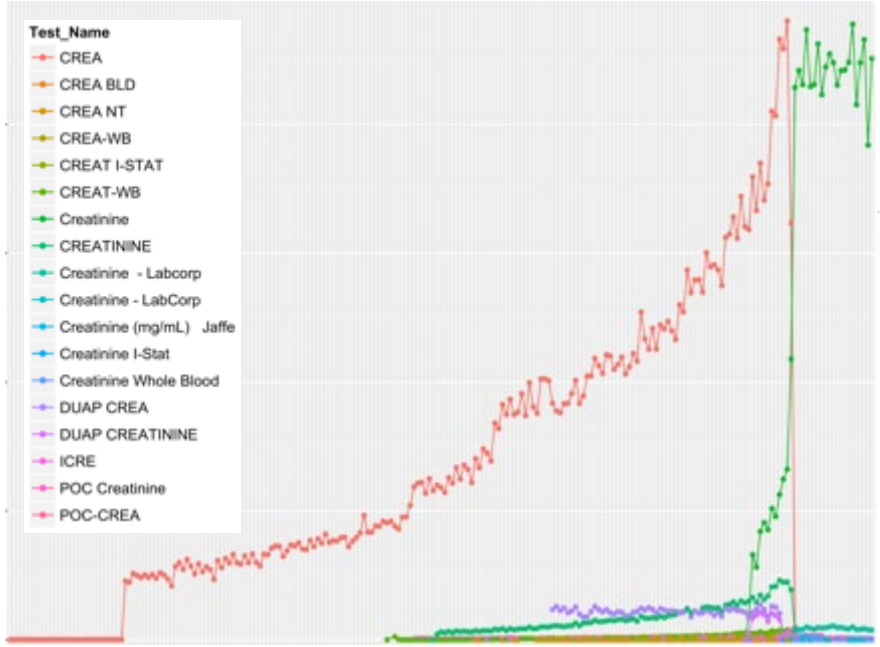
Development and Validation of ML-DQA

	Pediatric Sepsis Prediction	Lung Transplant Complication Prediction	Sepsis Prediction at Jefferson Health	Immune-Related Adverse Event Prediction	Maternal Morbidity and Mortality Prediction
Phase I: Data Element Pre-Processing					
Pre-existing groupers	108	109	30	39	310
Project-specific groupers	73	35	59	41	12
Phase II: ML-DQA Checks					
Completeness checks	144	144	70	508	404
Conformance checks	122	144	132	225	69
Plausability checks	123	144	61	301	404
Total quality checks	389	432	267	1,034	877

https://proceedings.mlr.press/v182/sendak22a.html

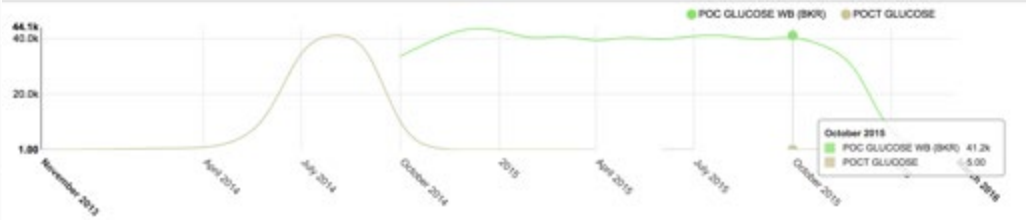
Grouper Maintenance to Address Meta Data Instability

Which Creatinine?



1996 2015

Which Glucose?



8 Key Decision Points in AI Adoption Process

Procurement

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution

Development & Adaptation

3. **Develop measures of outcomes and success of the AI product**
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product

Clinical Integration

6. Execute change management, workflow integration, and scaling strategy

Lifecycle Management

7. Monitor and maintain the AI product
8. Update or decommission the AI product

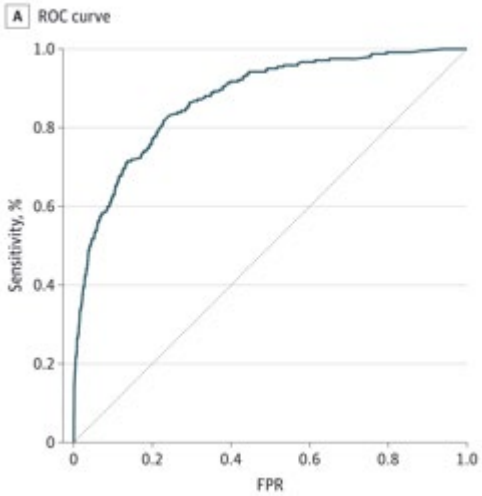
Example Categories of Measures

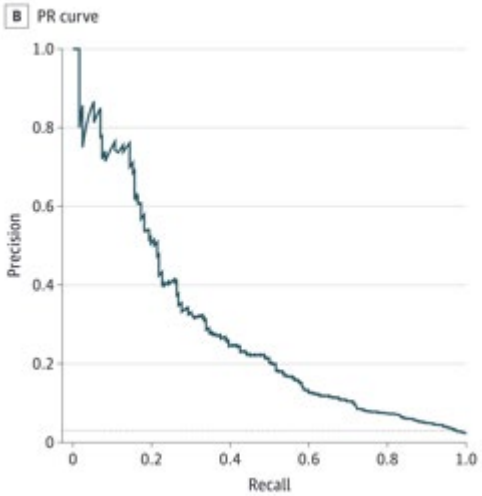
Category	Definition	Example Metrics
Model performance	**Effectiveness, accuracy, and reliability of the AI model** or algorithm in fulfilling its intended tasks within the clinical or healthcare context.	Sensitivity (recall, true positive rate), Specificity (true negative rate), Area Under the ROC Curve (AUC-ROC), F1 Score, Precision (positive predictive value).
Software performance	**Efficiency and responsiveness of processing tasks**, delivering results, and overall performance of the software components and its interactions.	Inference time, throughput, model latency, response time, resource utilization, scalability.
Clinical effectiveness	Assessment of impact of product use on healthcare outcomes.	**Mortality rate, intensive care unit requirement, complication rate**
Usability	**Quality of users' interactions with the AI-based medical software.**	Clinician satisfaction, user error rates, ease of use.
Safety and security	Safe and secure operating software, evaluating harm to patients and protection against unauthorized access, data breaches, and cyber threats.	Number of identified safety risks and mitigations, adherence to cybersecurity standards, detection of adversarial attacks, incident response time.
Business	Business objectives and outcomes	**Reduction in diagnostic time, cost savings.**

Mortality Model Performance Measures

Table 2. Prediction Accuracy by Evaluation Method, Location, and Time

Evaluation Method	Location	Time	AUROC (95% CI)	AUPRC (95% CI)
Retrospective	Hospital A	2014-2015	0.87 (0.83-0.89)	0.29 (0.25-0.37)
Retrospective	Hospital A	2018	0.85 (0.83-0.87)	0.17 (0.13-0.22)
Retrospective	Hospital B	2018	0.89 (0.86-0.92)	0.22 (0.14-0.31)
Retrospective	Hospital C	2018	0.84 (0.80-0.89)	0.13 (0.08-0.21)
Prospective	Hospital A	2019	0.86 (0.83-0.90)	0.14 (0.09-0.21)





Mortality Model Performance Measures

Threshold	Sensitivity	Specificity	PPV	Alerts, No./d		
				Total	False	True
0.01	0.88	0.66	0.05	39.9	37.8	2.1
0.02	0.76	0.81	0.08	23.3	21.5	1.8
0.03	0.68	0.88	0.11	15.3	13.6	1.7
0.04	0.61	0.91	0.12	11.9	10.4	1.5
0.05	0.57	0.93	0.15	9.1	7.7	1.4
0.06	0.54	0.95	0.18	7.4	6.1	1.3
0.07	0.52	0.95	0.19	6.5	5.3	1.3
0.08	0.50	0.96	0.21	5.8	4.5	1.2
0.09	0.48	0.96	0.22	5.2	4.1	1.2
0.10	0.44	0.97	0.22	4.8	3.7	1.1
0.11	0.43	0.97	0.24	4.4	3.4	1.0
0.12	0.41	0.97	0.24	4.1	3.1	1.0
0.13	0.39	0.98	0.26	3.7	2.7	1.0
0.14	0.39	0.98	0.27	3.5	2.6	0.9
0.15	0.36	0.98	0.27	3.2	2.3	0.9
0.16	0.35	0.98	0.28	3.1	2.2	0.9
0.17	0.34	0.98	0.30	2.8	2.0	0.8
0.18	0.33	0.98	0.32	2.6	1.7	0.8
0.19	0.31	0.99	0.32	2.4	1.6	0.8
0.20	0.29	0.99	0.33	2.2	1.5	0.7
0.21	0.28	0.99	0.33	2.0	1.4	0.7
0.22	0.28	0.99	0.35	1.9	1.3	0.7
0.23	0.27	0.99	0.36	1.8	1.2	0.7
0.24	0.26	0.99	0.38	1.7	1.1	0.6
0.25	0.26	0.99	0.41	1.5	0.9	0.6

Abbreviation: PPV, positive predictive value.

Number Needed to Evaluate = 1 / PPV

8 Key Decision Points in AI Adoption Process

Procurement

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution

Development & Adaptation

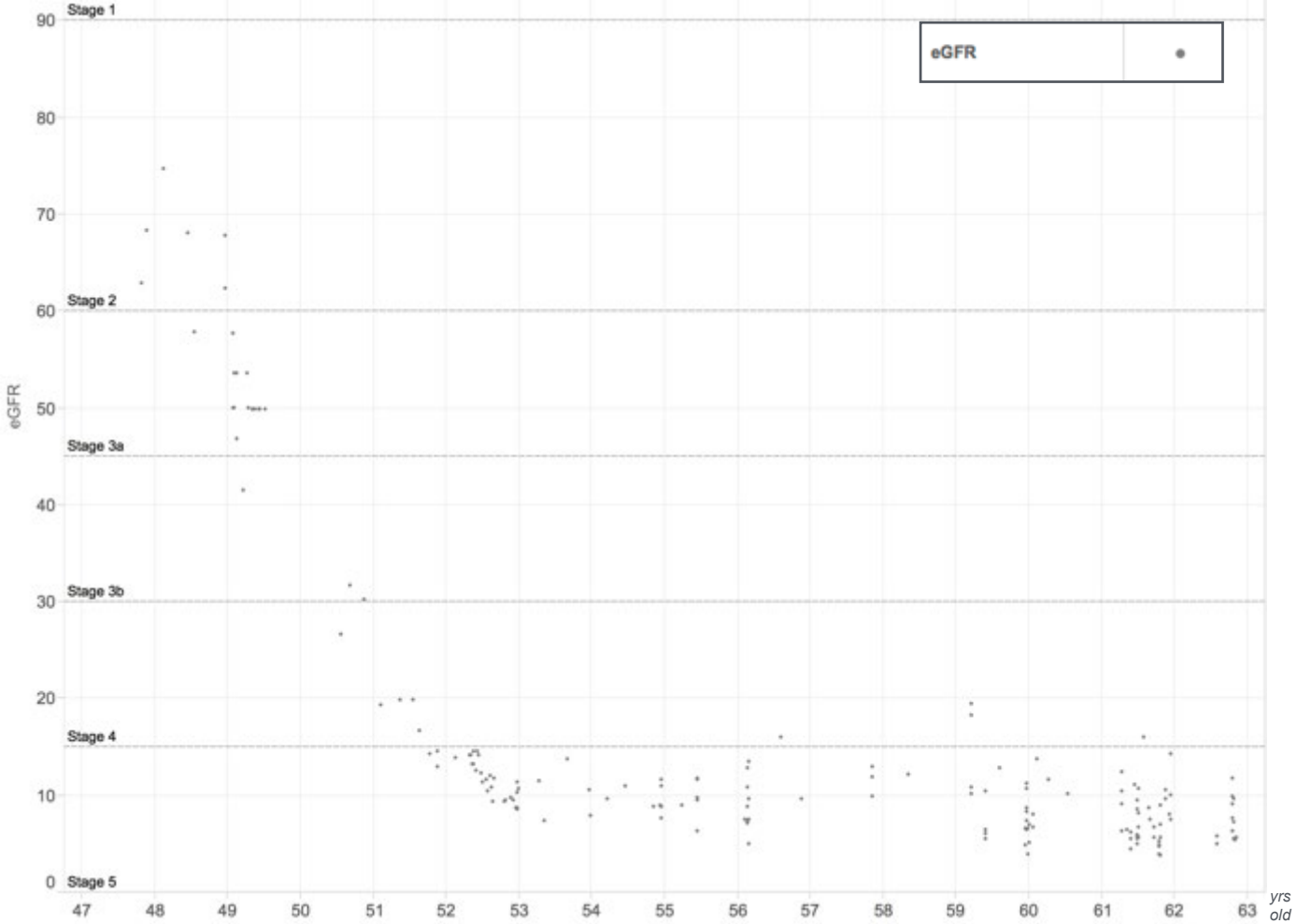
3. Develop measures of outcomes and success of the AI product
4. **Design a new optimal workflow to facilitate integration**
5. Evaluate pre-integration safety and effectiveness of the AI product

Clinical Integration

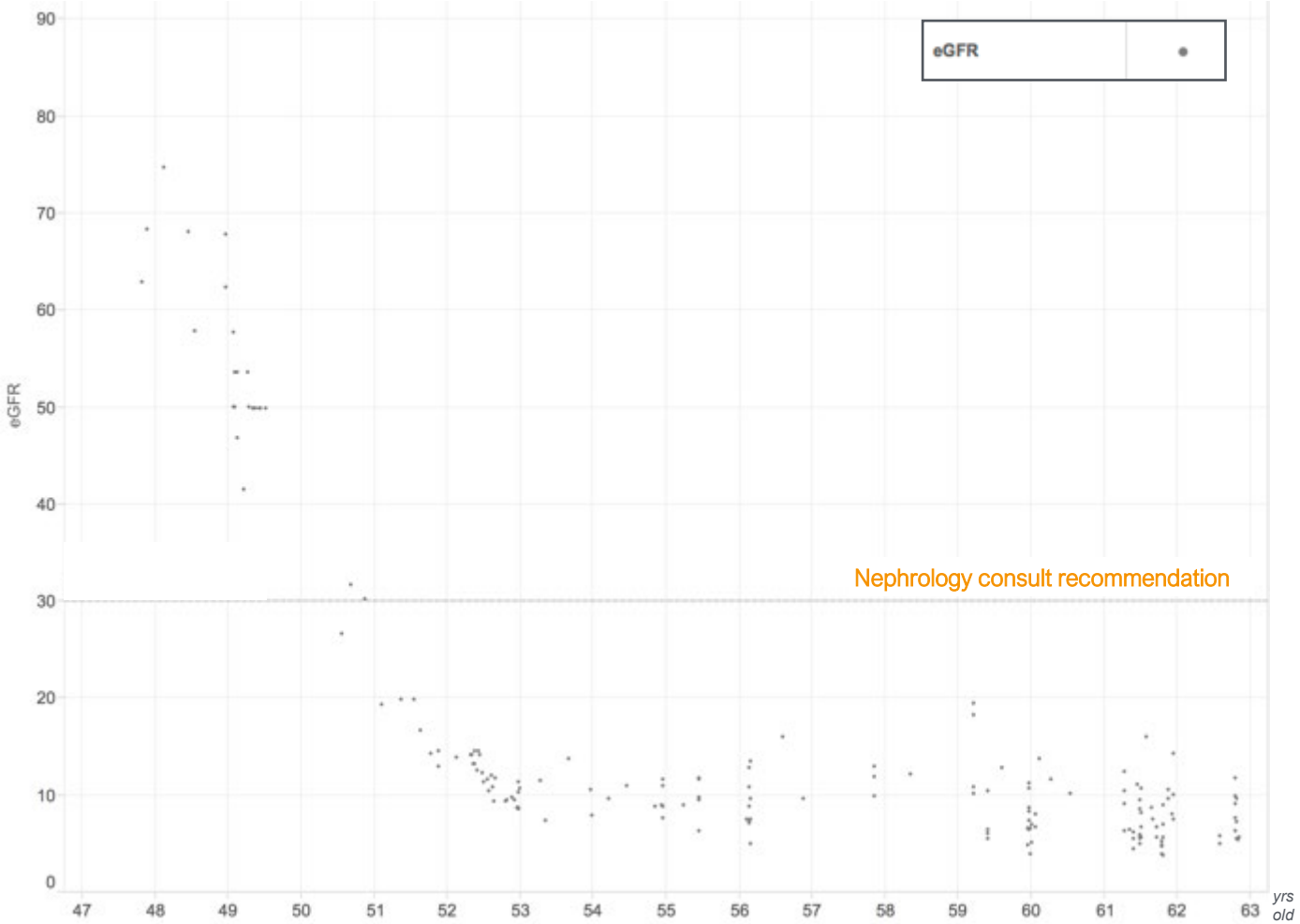
6. Execute change management, workflow integration, and scaling strategy

Lifecycle Management

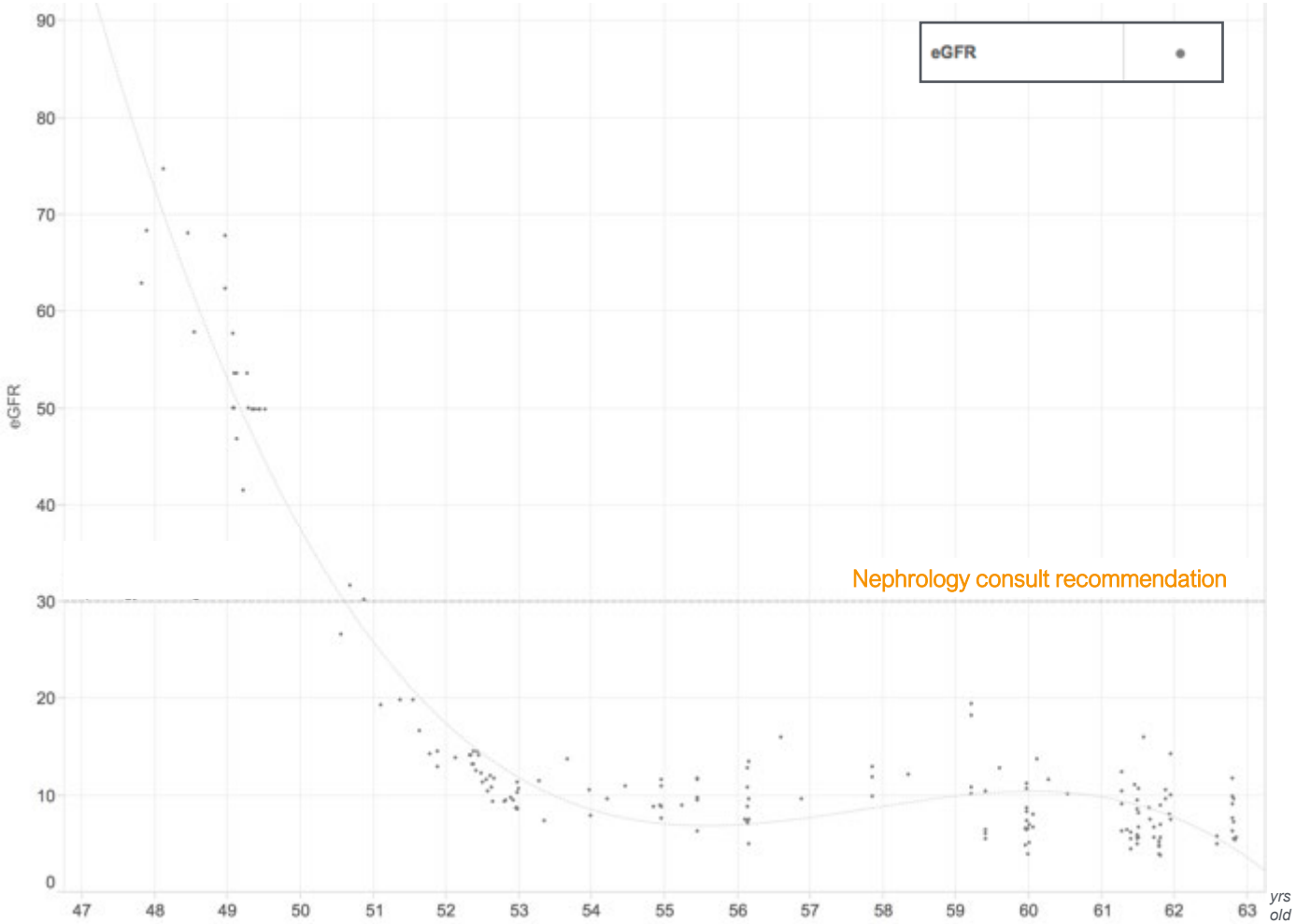
7. Monitor and maintain the AI product
8. Update or decommission the AI product



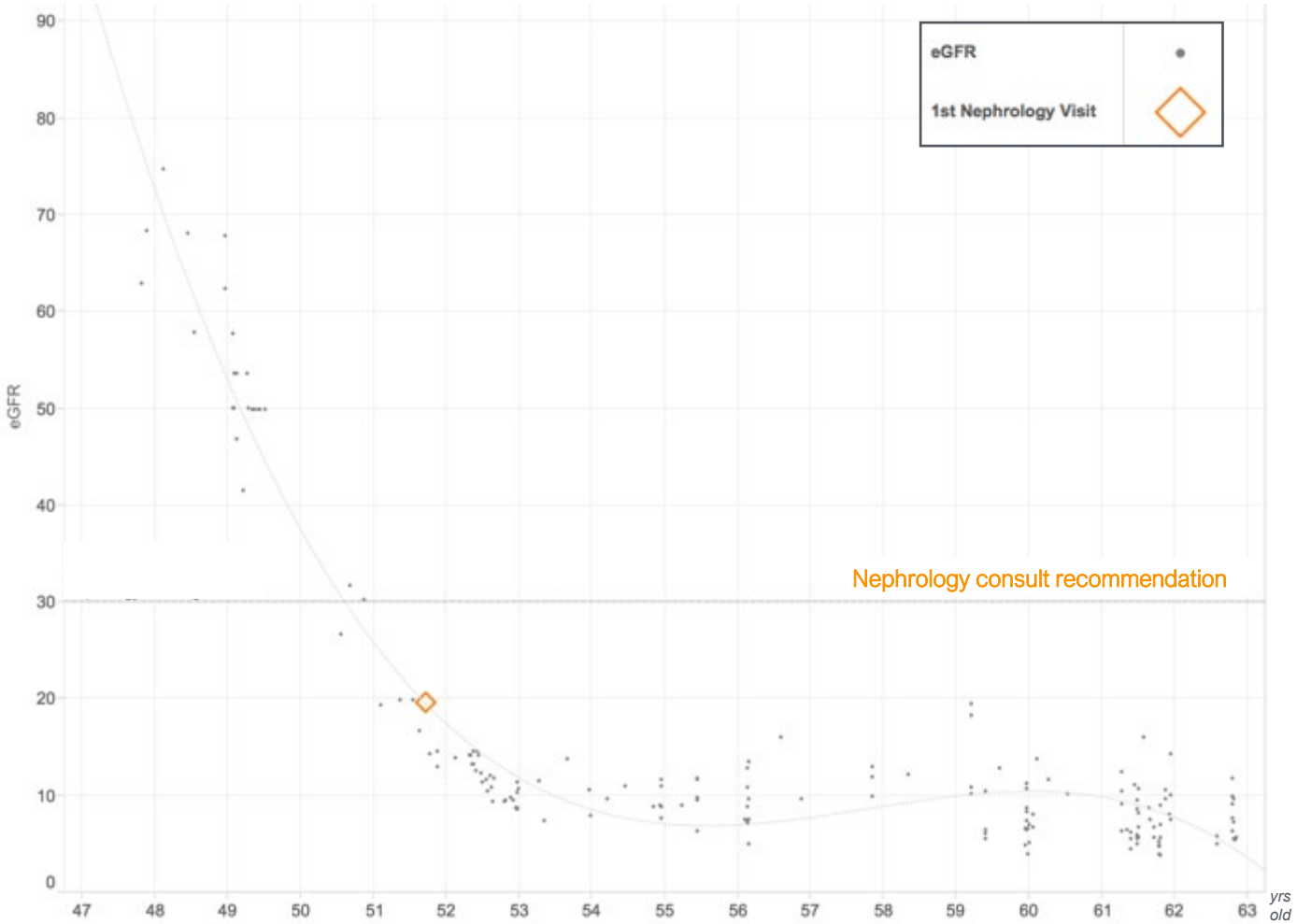
Credit: RJ Andrews



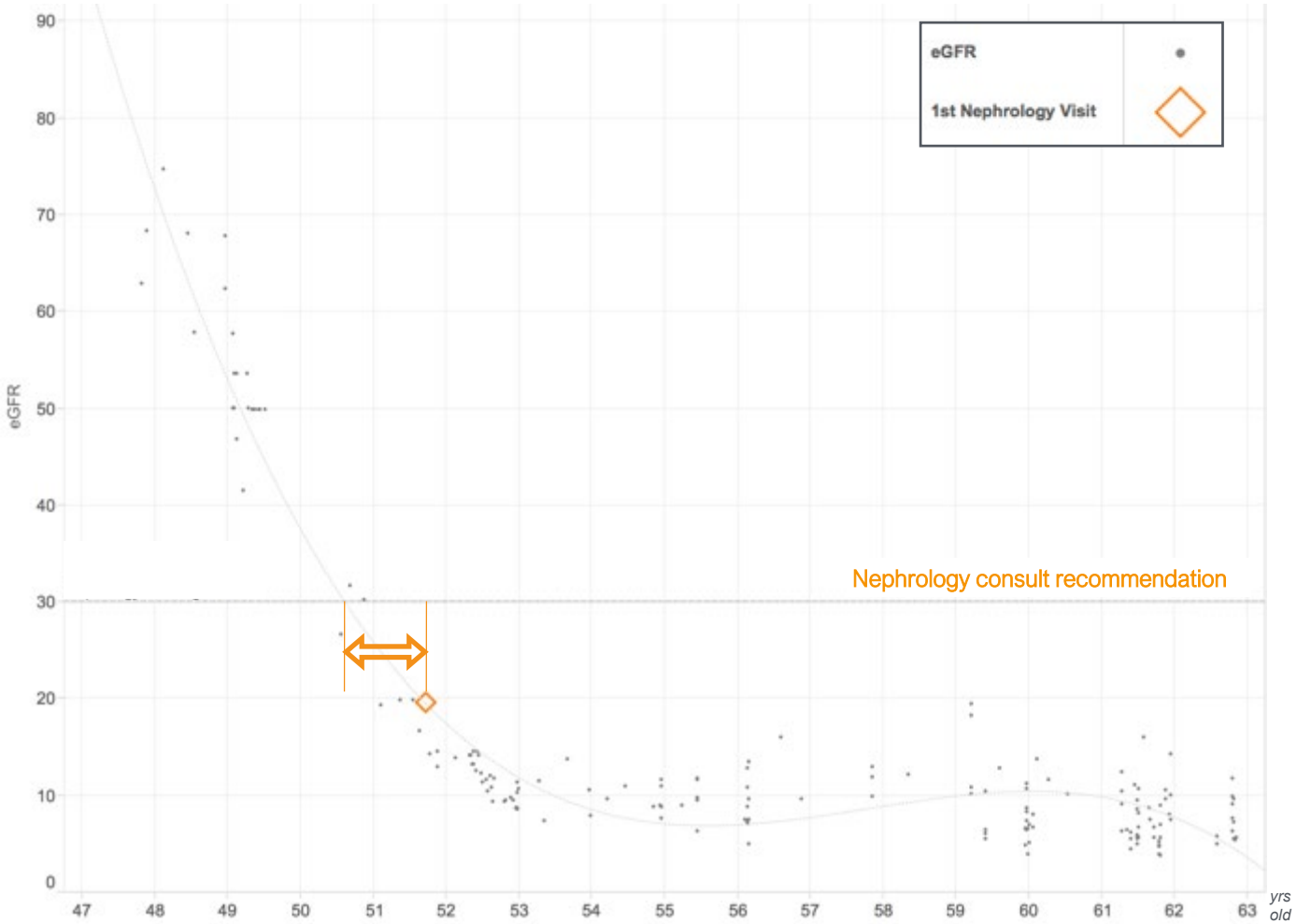
Credit: RJ Andrews



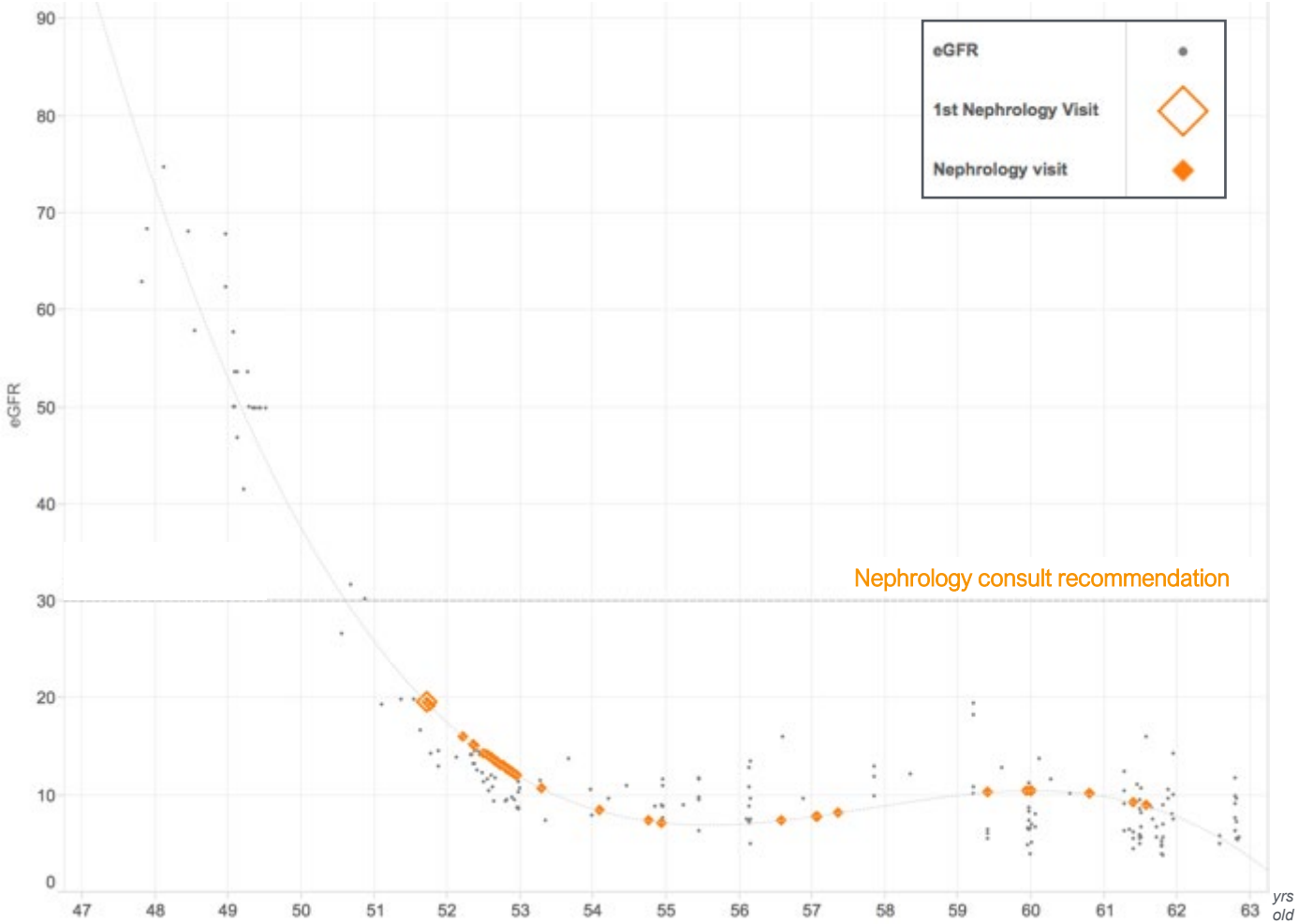
Credit: RJ Andrews



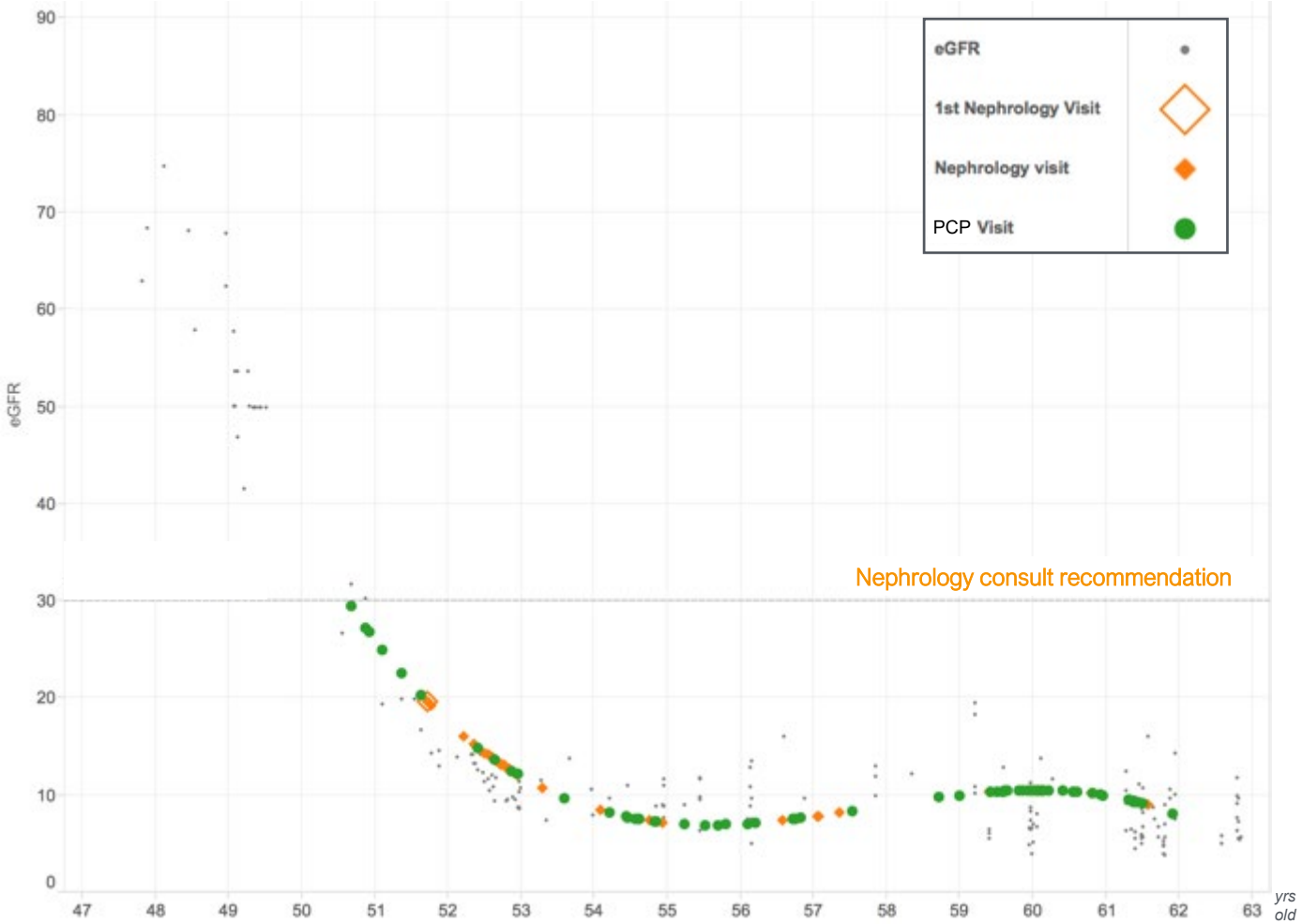
Credit: RJ Andrews



Credit: RJ Andrews

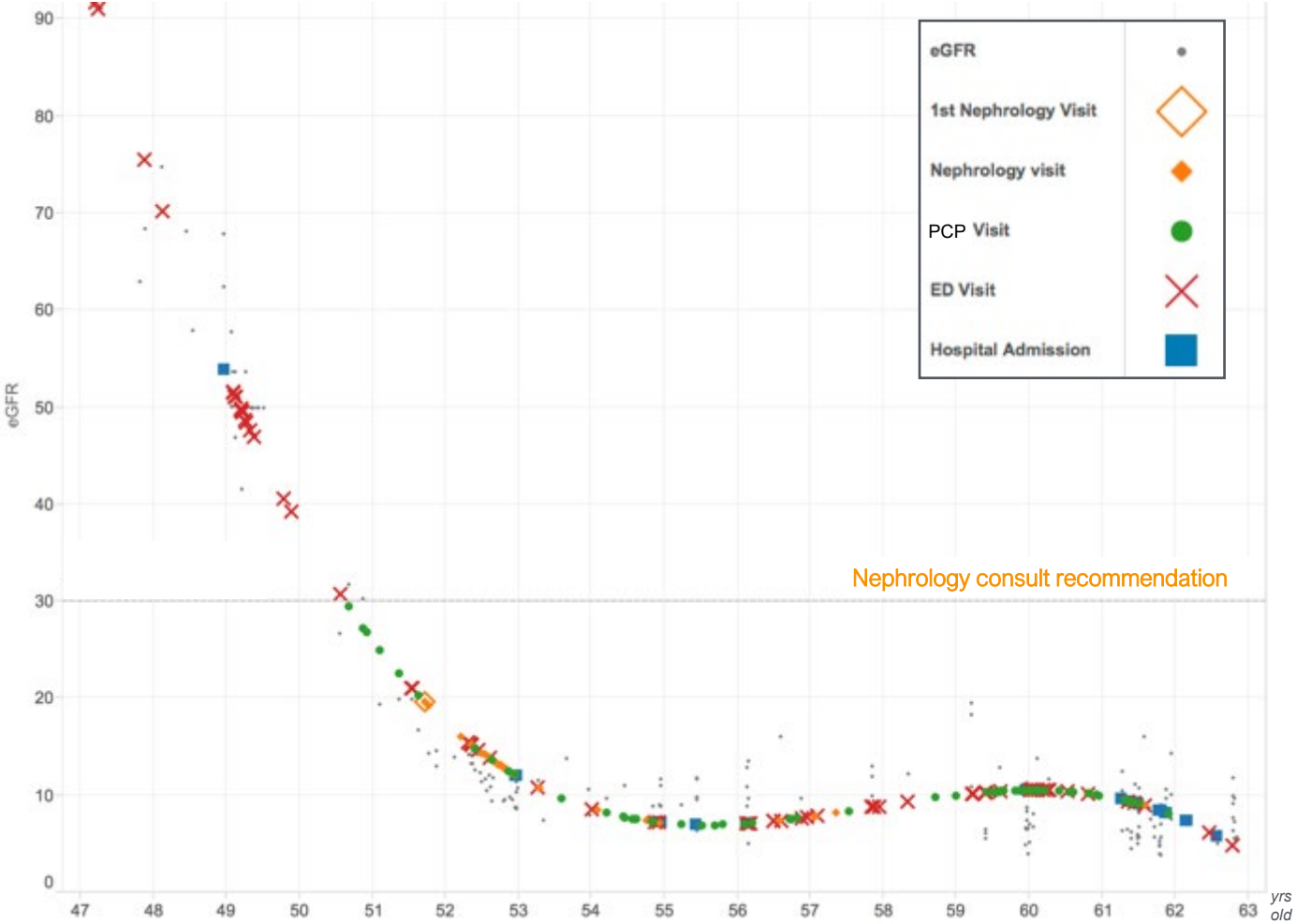


Credit: RJ Andrews

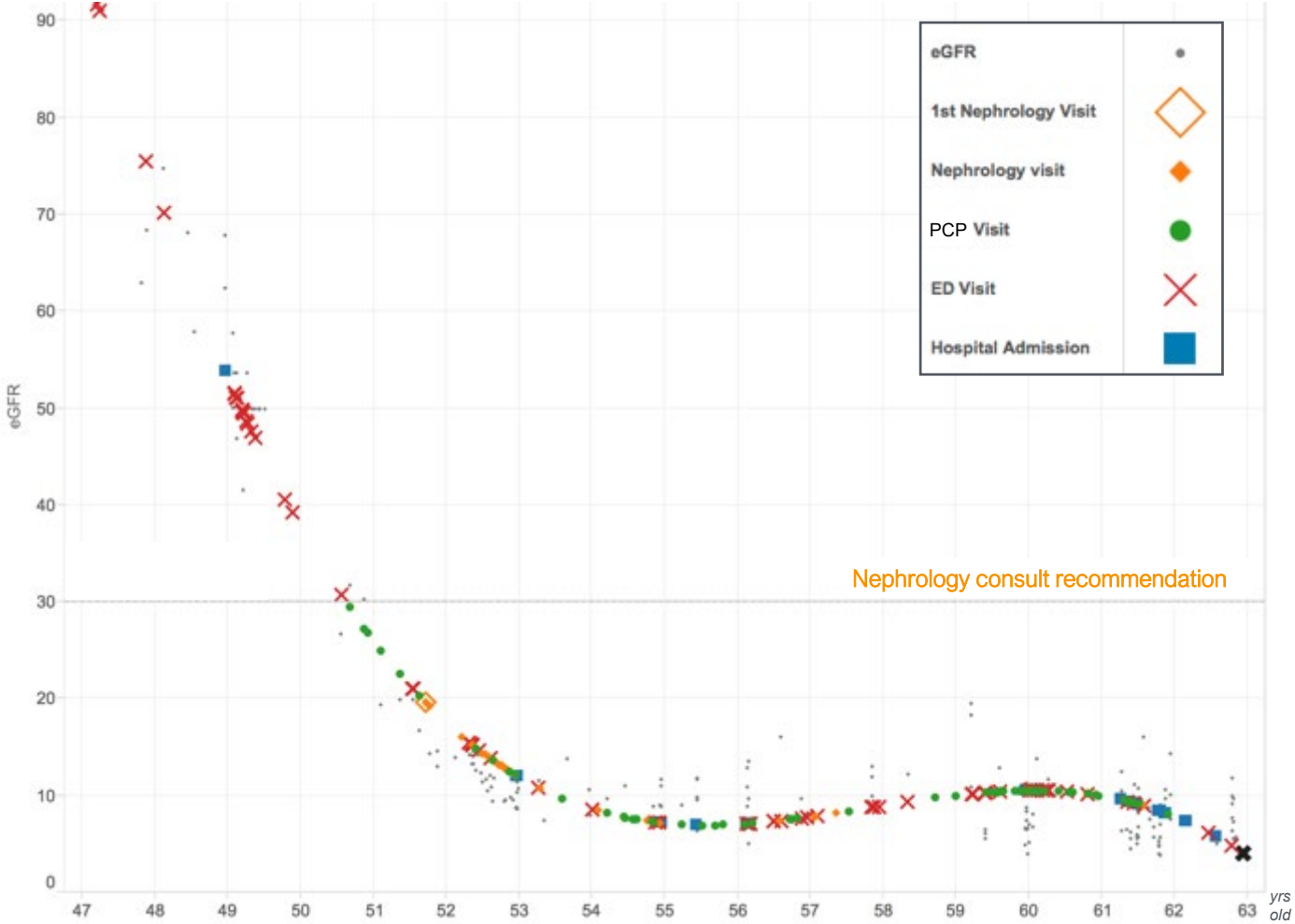


Credit: RJ Andrews

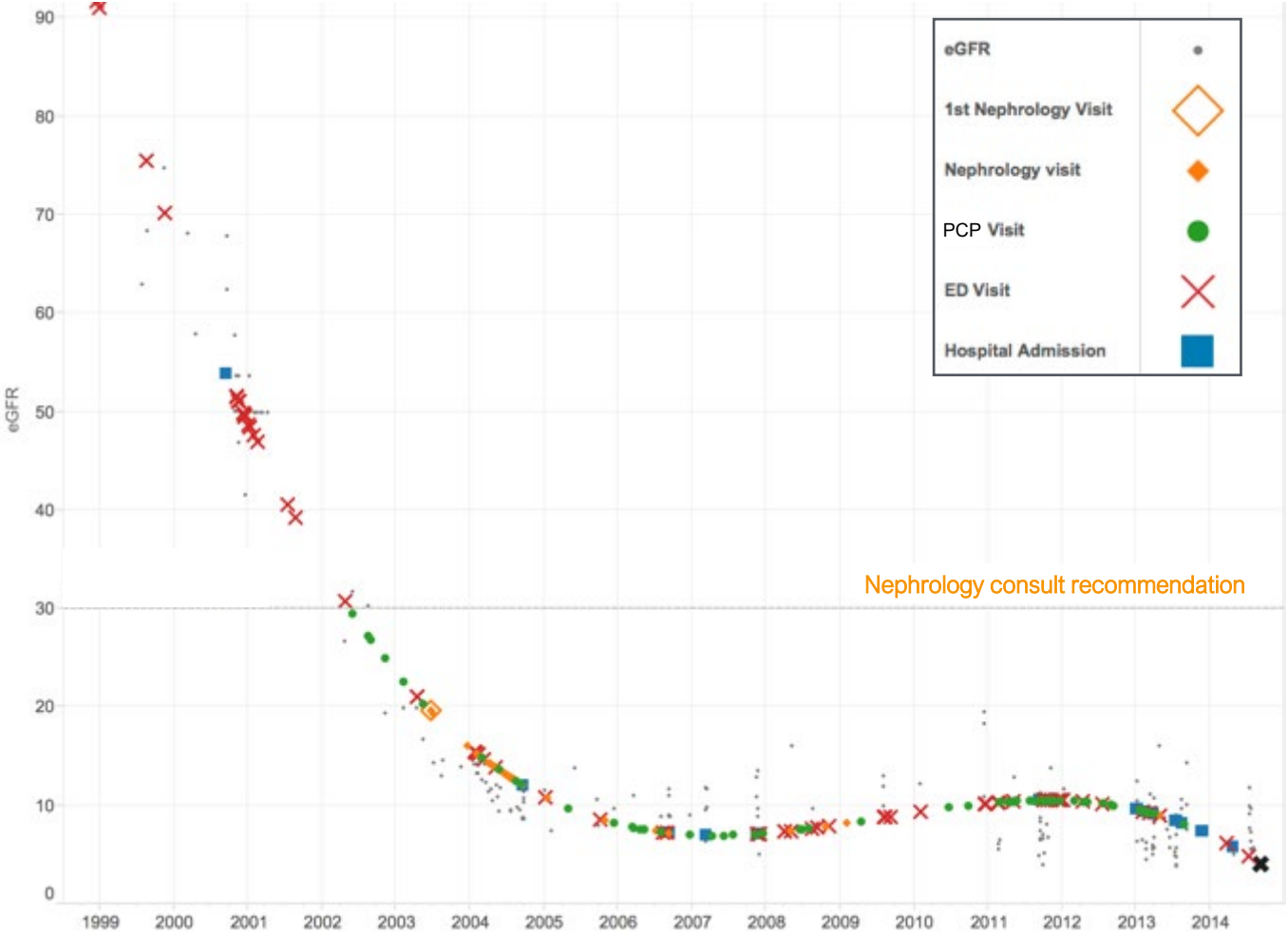
Credit: RJ Andrews



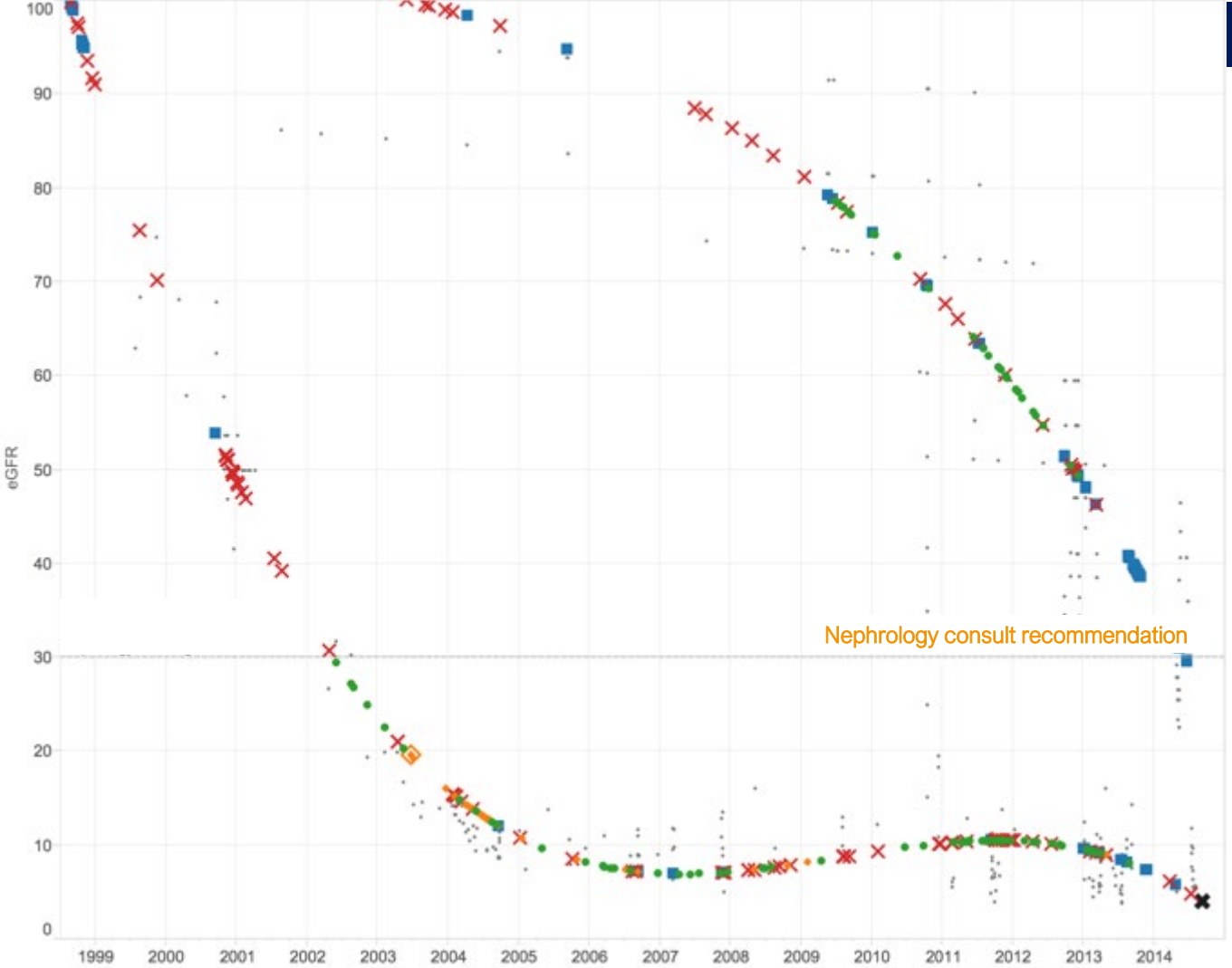
Credit: RJ Andrews



Credit: RJ Andrews



Credit: RJ Andrews



Credit: RJ Andrews

“Doc, why didn’t anyone tell me sooner?”

Validated Measures

A Predictive Model for Progression of Chronic Kidney Disease to Kidney Failure

Navdeep Tangri, MD, FRCPC
Lesley A. Stevens, MD, MS, FRCPC
John Griffith, PhD
Hocine Tighiouart, MS
Ognjenka Djurdjev, MSc
David Naimark, MD, FRCPC
Adeera Levin, MD, FRCPC
Andrew S. Levey, MD

Context Chronic kidney disease (CKD) is common. Kidney disease severity can be classified by estimated glomerular filtration rate (GFR) and albuminuria, but more accurate information regarding risk for progression to kidney failure is required for clinical decisions about testing, treatment, and referral.

Objective To develop and validate predictive models for progression of CKD.

Design, Setting, and Participants Development and validation of prediction models using demographic, clinical, and laboratory data from 2 independent Canadian cohorts of patients with CKD stages 3 to 5 (estimated GFR, 10-59 mL/min/1.73 m^2) who were referred to nephrologists between April 1, 2001, and December 31, 2008. Models were developed using Cox proportional hazards regression methods and evalu-

5 Year Risk of ESRD Progression - JAMA, 2011

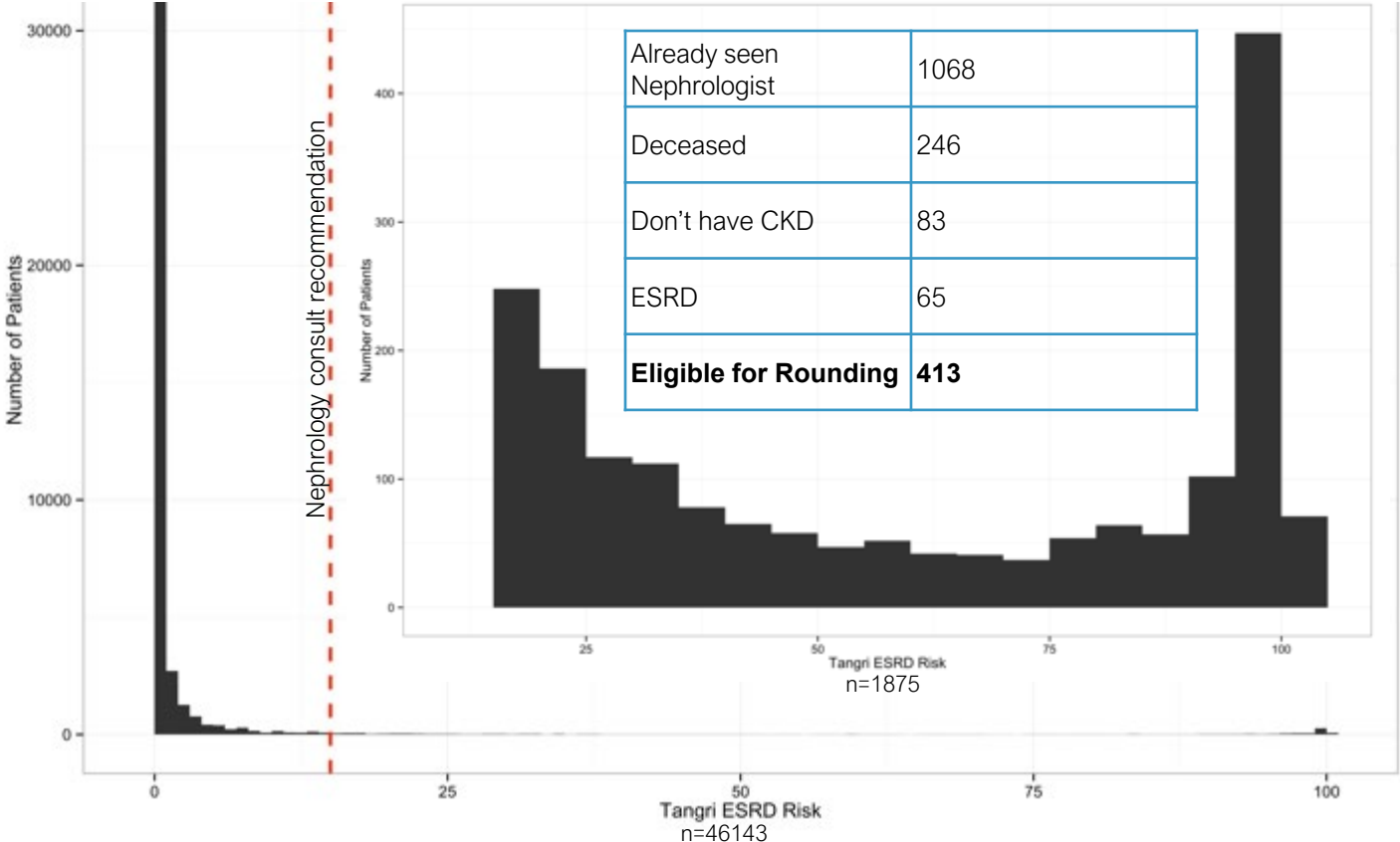
Decline in Estimated Glomerular Filtration Rate and Subsequent Risk of End-Stage Renal Disease and Mortality

Josef Coresh, MD, PhD; Tanvir Chowdhury Turin, MD, PhD; Kunihiro Matsushita, MD, PhD; Yingying Sang, MSc; Shoshana H. Ballew, PhD; Lawrence J. Appel, MD; Hisatomi Arima, MD; Steven J. Chadban, PhD; Massimo Cirillo, MD; Ognjenka Djurdjev, MSc; Jamie A. Green, MD; Gunnar H. Heine, MD; Lesley A. Inker, MD; Fujiko Irie, MD, PhD; Areef Ishani, MD, MS; Joachim H. Ix, MD, MAS; Csaba P. Kovesdy, MD; Angharad Marks, MBBCh; Takayoshi Ohkubo, MD, PhD; Varda Shalev, MD; Anoop Shankar, MD; Chi Pang Wen, MD, DrPH; Paul E. de Jong, MD, PhD; Kunitoshi Iseki, MD, PhD; Benedicte Stengel, MD, PhD; Ron T. Gansevoort, MD, PhD; Andrew S. Levey, MD; for the CKD Prognosis Consortium

2 Year eGFR Change - JAMA, 2014

Adapt Workflows, Roles, and Organization

Don't Rely on Existing Workflows to Solve Problems



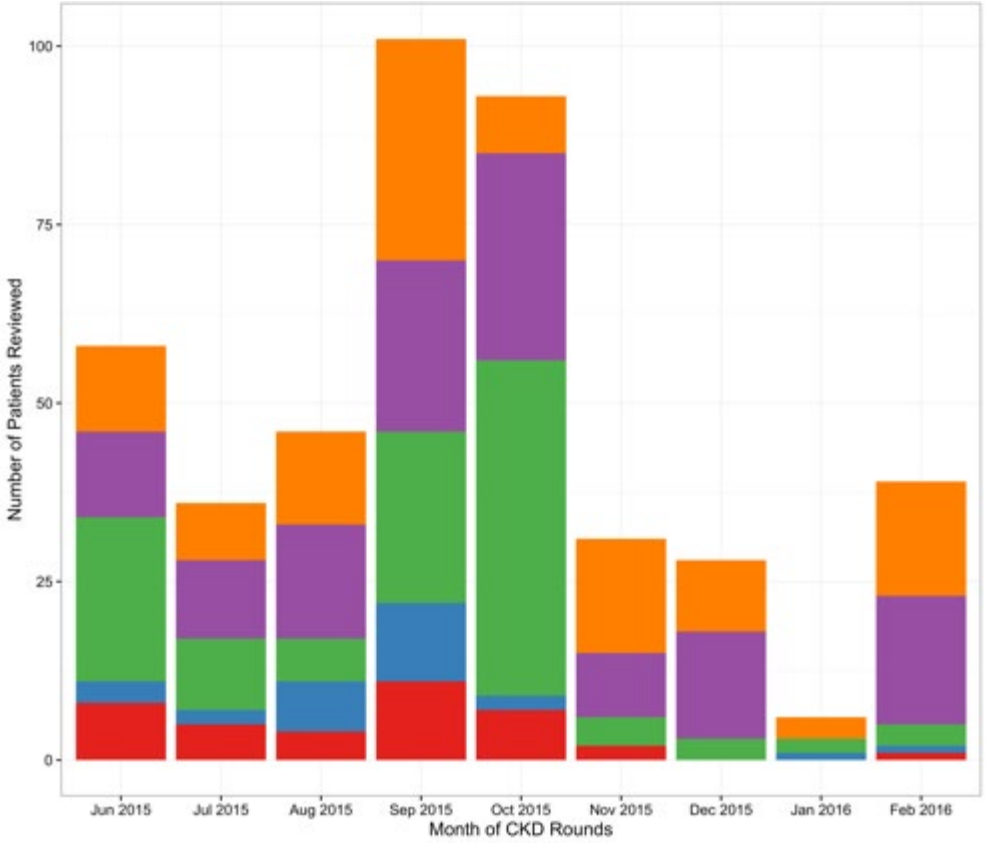
Already seen Nephrologist	1068
Deceased	246
Don't have CKD	83
ESRD	65
Eligible for Rounding	**413**

Adapt Workflows, Roles, and Organization

Don't Rely on Existing Workflows to Solve Problems

Adapt Workflows, Roles, and Organization

Don't Rely on Existing Workflows to Solve Problems



Intervention	N	% of Total
Nephrology Appointment	84	0.72
PCP Appointment	21	0.18
Lab Order	15	0.13
Medication Change	10	0.09
Care Management	7	0.05
Total	137	

OUTCOME
- DECEASED
- DIALYSIS (NO ACTION)
- SEEING NEPHROLOGY (NO ACTION)
- OTHER (NO ACTION)
- DUKEWELL INTERVENTION

Adapt Workflows, Roles, and Organization

Don't Rely on Existing Workflows to Solve Problems

Now extended and applied to:

- **Non-alcoholic fatty liver disease (NAFLD)**
- **Peripheral artery disease**
- **Community-based palliative care**

Adapt Workflows, Roles, and Organization

Don't Rely on Existing Workflows to Solve Problems

Now extended and applied to:

- **Non-alcoholic fatty liver disease (NAFLD)**
- **Peripheral artery disease**
- **Community-based palliative care**

"The difference in [*algorithm*] performance is negligible compared to the difference that a good physician champion makes, or a good intervention plan makes. Those are by far and away the most important things to the success of a project. The actual model itself is, as much as I might delude myself or whatever, it's actually not that important."

- *Technical Stakeholder*

Adapt Workflows, Roles, and Organization

Restructure Organization to Create Alignment

Dr. Kevin Schulman, https://dcricollab.dcri.duke.edu/sites/NIHKR/KR/GR-Slides-1-19-18.pdf

- Duke - Moved Rapid Response Team out of Cardiac ICU to create Patient Response Program with new reporting structure
- Duke - Moved care management function and ACO under newly created Population Health Management Office
- NYC – Moved Test + Trace out of Public Health Department directly into City Hall

8 Key Decision Points in AI Adoption Process

Procurement

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution

Development & Adaptation

3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. **Evaluate pre-integration safety and effectiveness of the AI product**

Clinical Integration

6. Execute change management, workflow integration, and scaling strategy

Lifecycle Management

7. Monitor and maintain the AI product
8. Update or decommission the AI product

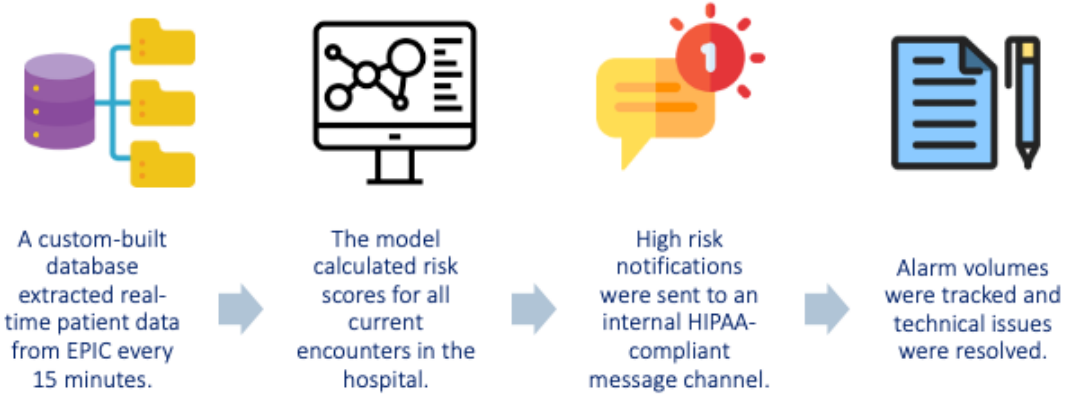
Identifying Label Leakage During a Silent Trial

- Pediatric sepsis prediction
  - Outcome definition: Blood Culture ∩ Antibiotics for 4 days ∩ Acute organ dysfunction
  - LSTM with 6-hour prediction window and 3-hour snooze
  - Retrospective training set: 17,491 unique encounters for children between 30 days old and 18 years old between November 1, 2016 – December 31, 2020
  - Temporal validation set: 6,545 unique encounters for children between 30 days old and 18 years old between January 1, 2021 – June 30, 2022

Identifying Label Leakage During a Silent Trial

	AUROC	AUPRC	PPV at 20% sensitivity (with 3hr snooze)	PPV at 50% sensitivity (with 3hr snooze)
Retrospective test set	0.816	0.483	0.769	0.612
Temporal validation	0.862	0.386	0.851	0.611

Silent Trial Design



Identifying Label Leakage During a Silent Trial

- Silent trial results
  - Model ran on 1,475 unique encounters over 2 months
  - Model generated 30 alarms per day >> 2 alarms per day expected
  - Model fired alarm on almost all patients in ED within first hour of arrival

Identifying Label Leakage During a Silent Trial

- Label leakage due to layer normalization in LSTM
  - In retrospective training data:
    - set maximum encounter length to 168 hours
    - truncated sepsis encounters at time of sepsis
  - Shorter encounter → more padding of encounter hours with 0s → smaller mean after layer normalization
  - Longer encounter → less padding of encounter hours with 0s → larger mean after layer normalization
  - In retrospective data, model learned to associate early hours of encounter with sepsis

Identifying Label Leakage During a Silent Trial

- Retrained LSTM without layer normalization using the same hyperparameters

	AUROC	AUPRC
Retrospective test set (with layer normalization)	0.816	0.483
Temporal validation (with layer normalization)	0.862	0.386
Retrospective test set (without layer normalization)	**0.782**	**0.01**

8 Key Decision Points in AI Adoption Process

Procurement

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution

Development & Adaptation

3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product

Clinical Integration

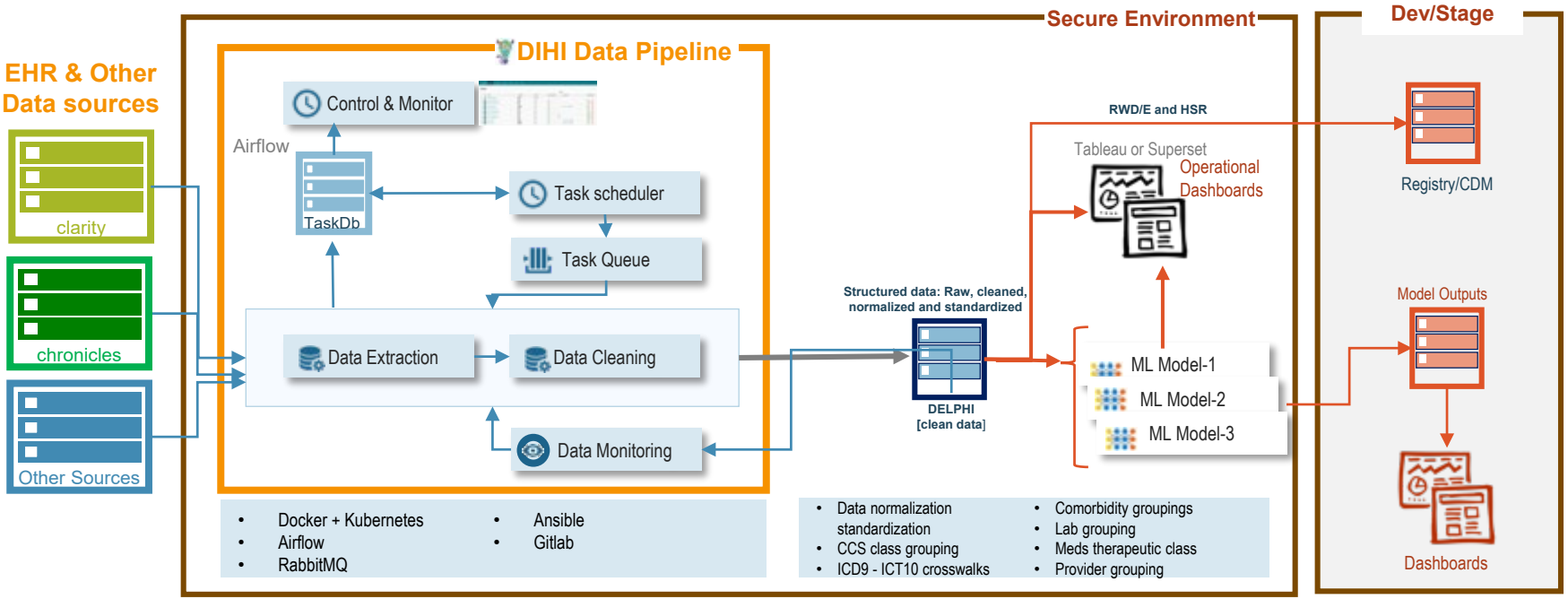
6. **Execute change management, workflow integration, and scaling strategy**

Lifecycle Management

7. Monitor and maintain the AI product
8. Update or decommission the AI product

Build Modular Infrastructure to Support Many Projects

Flexible Data Pipeline Technology Infrastructure



Model Labels

Model Facts **Model name:** Deep Sepsis **Locale:** Duke University Hospital

Approval Date: 09/22/2019 **Last Update:** 09/24/2019. **Version:** 1.0

Summary

This model uses EHR input data collected from a patient's current inpatient encounter to estimate the probability that the patient will meet sepsis criteria within the next 4 hours. It was developed in 2016-2019 by the Duke Institute for Health Innovation. The model was licensed to Cohere Med in July 2019.

Mechanism

- **Outcome** ....sepsis within the next 4 hours, see (1) for sepsis criteria
- **Output** ....0% - 100% probability of sepsis occurring in the next 4 hours
- **Patient population** ....all adult patients >18 y.o. presenting to DUH ED and admitted
- **Time of prediction** ....every hour of a patient's encounter
- **Input data source** ....electronic health record (EHR)
- **Input data type** ....demographics, analytes, vitals, medication administrations
- **Training data location and time-period** ....DUH, 10/2014 – 12/2015
- **Model type** .... Recurrent Neural Network

Validation and performance

	Prevalence	AUC	PPV @ Sensitivity of 60%	Sensitivity @ PPV of 20%
Local Retrospective	18.9%	0.88	0.14	0.50
Local Temporal	6.4%	0.94	0.20	0.66
Local Prospective	TBD	TBD	TBD	TBD
External	TBD	TBD	TBD	TBD

Uses and directions

- **Operational use case(s):** Every hour, data is pulled from the EHR to calculate risk of sepsis for every patient at the DUH ED. A rapid response team nurse reviews every high-risk patient with a physician in the ED to confirm whether or not to initiate treatment for sepsis.
- **General use:** This model is intended to be used to by clinicians to identify patients for further assessment for sepsis. The model is not a diagnostic for sepsis and is not meant to guide or drive clinical care. This model is intended to complement other pieces of patient information related to sepsis as well as a physical evaluation to determine the need for sepsis treatment.
- **Examples of appropriate decisions to support:** Patient X has a high risk of sepsis according to the model. A rapid response team nurse discusses the patient with the ED physician caring for the patient and they agree the patient does not require treatment for sepsis.
- **Before using this model:** Test the model retrospectively and prospectively on local data to confirm generalizability of the model to the local setting.
- **Safety and efficacy evaluation:** Analysis of data from clinical trial (NCT03655626) underway. Preliminary data shows rapid response team, nurse-driven workflow was effective at improving sepsis treatment bundle compliance.

Comment | Open Access | Published: 23 March 2020

Presenting machine learning model information to clinical end users with model facts labels

Mark P. Sendak, Michael Gao, Nathan Brajer & Suresh Balu

npj Digital Medicine **3**, Article number: 41 (2020) | Cite this article

Warnings

- **General warnings:** This model was not trained or evaluated on patients receiving care in the ICU. Do not use this model in the ICU setting without further evaluation. This model was trained to identify the first episode of sepsis during an inpatient encounter. During long inpatient stays with multiple sepsis episodes, model accuracy needs to be further evaluated. The model is not interpretable and does not provide rationale for high risk scores. Clinical end users are expected to place model output in context with other clinical information to make final determination of diagnosis.
- **Examples of inappropriate decisions to support:** This model may not be accurate outside of the target population, primarily adults in the non-ICU setting. This model is not a diagnostic and is not designed to guide clinical diagnosis and treatment for sepsis.
- **Discontinue use if:** Clinical staff raise concerns about utility of the model for the indicated use case or large, systematic changes occur at the data level that necessitates re-training of the model.

Other information:

- **Outcome Definition:** https://doi.org/10.1101/648907
- **Related model:** http://doi.org/10.1001/jama.2016.0288
- **Model development & validation:** arxiv.org/abs/1708.05894
- **Model implementation**: jmir.org/preprint/15182
- **Clinical trial:** clinicaltrials.gov/ct2/show/NCT03655626
- **Clinical impact evaluation:** TBD
- **For inquiries and additional information:** please email mark.sendak@duke.edu

8 Key Decision Points in AI Adoption Process

Procurement

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution

Development & Adaptation

3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product

Clinical Integration

6. Execute change management, workflow integration, and scaling strategy

Lifecycle Management

7. **Monitor and maintain the AI product**
8. Update or decommission the AI product

AI System Monitoring at DIHI

Effective monitoring of AI/ML solutions also requires multidisciplinary combination of technical and human capabilities, including expertise in engineering, data analysis, AI/ML, and clinical domain knowledge employed during the solution development phase.

Model Monitoring

- Data quality monitoring
  - Input data accurate, complete, and up-to-date
  - Entity/grouper monitoring
  - Continuous monitoring
- Performance comparison
  - auroc, auprc wrt. training
  - Analysis cadence: M/Q/Y
- Output drift monitoring
  - Data distribution
  - Category distribution

Solution Monitoring

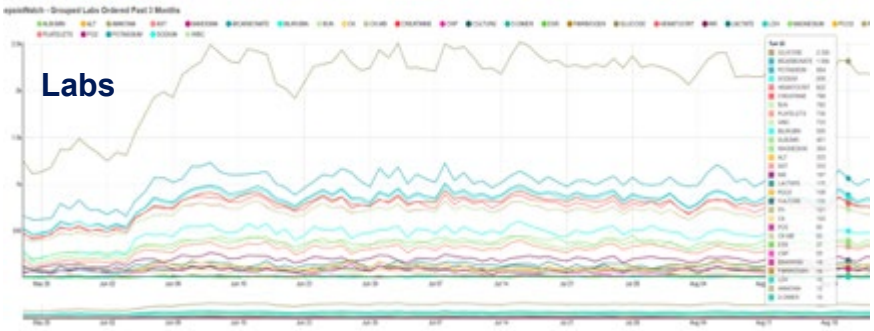
- Outcome monitoring
  - Project specific measures
  - Bi-annual for most solutions
- Workflow changes
  - Observation / documentation
- Usage monitoring
  - UI tools/dashboard usage
  - Secondary data analysis
- User feedback
  - Survey for model & solution usability and refinements

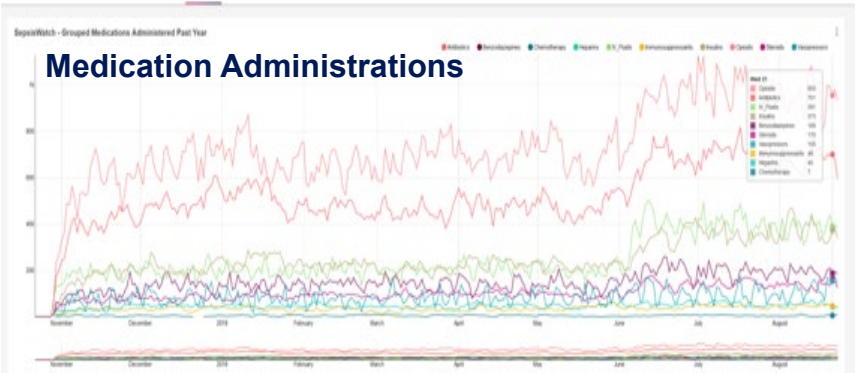
Operations Monitoring

- Alerting & notification
  - Flexible rules-based engine for alerting
  - Used in clinical workflow
  - Email/page/spok/sms etc.
- Technical monitoring
  - Model run times, failures etc.
  - Service level monitoring
- Regulatory & Policy
  - Compliance monitoring for regulation & Duke policies
  - Ethical and legal standards

Solution and Input Data Monitoring

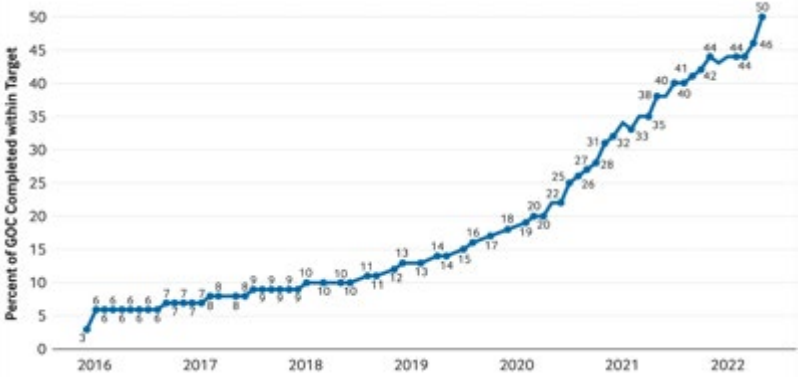
Continuous monitoring to ensure safety and quality of data used in model inputs





SEP-1 bundle compliance | Sepsis Watch model

Goal-concordant care outcome | HealthGuard model



8 Key Decision Points in AI Adoption Process

Procurement

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution

Development & Adaptation

3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product

Clinical Integration

6. Execute change management, workflow integration, and scaling strategy

Lifecycle Management

7. Monitor and maintain the AI product
8. **Update or decommission the AI product**

Sepsis Watch Post-Integration Lifecycle Management

	Monitoring & Evaluation	Update	Operational Management
Event based	• Debug issues that arise (e.g., data endpoint unexpectedly goes down)	• **Customize the UI for different user groups** • Train new versions of the model for new clinical settings	• Update user access • **Update reporting functionalities to support clinician management**
Recurring	• Monitor technical elements of the model and source data in pipeline • Monitor changes that affects work environment and use of model	• **Regularly scheduled maintenance (e.g., update groupers every 6 months)**	• **Conduct bi-annual end user training to ensure baseline knowledge of AI system**
Semi-Recurring	• **Audit the solution for impact on clinical and operational outcomes and impact on work environment**	• Improve the UI (e.g., add comment feature, automatically check boxes) • Scale to different use cases	• Convene governance committee monthly • Secure ongoing funding for AI system use
One-off	• Create channels for end users to report issues and provide user support services	• Create process and criteria to scope responses to user requests	• Determine ownership of model (e.g., clinical lead, technical lead)

8 Key Decision Points in AI Adoption Process

Procurement

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution

Development & Adaptation

3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product

Clinical Integration

6. Execute change management, workflow integration, and scaling strategy

Lifecycle Management

7. Monitor and maintain the AI product
8. Update or decommission the AI product

Duke Institute for Health Innovation	**2 mins**
Health AI Partnership	**2 mins**
Safe, Effective, and Equitable AI Translation	**20 mins**
Health Equity Across the AI Lifecycle (HEAAL)	**8 mins**

Health AI Partnership Inaugural Workshop

We are all invited to collaboratively develop a framework that addresses core technology evaluation domains across both case studies. While grounded in two real cases studies, the framework should be generalizable.

The framework should answer the question: **“our health system is considering adopting a new solution that uses AI; how do we assess the potential future impact on health inequities?”**

Health AI Partnership Inaugural Workshop

case 1: NYP Post-partum depression	
1:10 – 1:20 PM	NYP team presents case 1
1:20 – 1:50 PM	Breakout group activity - Participants expect to report back - Observers can take break or work on activity without need to report back
1:50 – 2:20 PM	Breakout rooms report back, Q&A
2:20 – 2:35 PM	Expert panel remarks and discussion
2:35 – 2:45 PM	NYP team presents case 1 learnings and approach

case 2: PCCI KnowThyPatient	
3:00 – 3:10 PM	PCCI team presents case 2
3:10 – 3:40 PM	Breakout group activity - Participants expect to report back - Observers can take break or work on activity without need to report back
3:40 – 4:10 PM	Breakout rooms report back, Q&A
4:10 – 4:25 PM	Expert panel remarks and Q&A
4:25 – 4:35 PM	PCCI team presents case 2 learnings and approach

Health System Partners

DukeHealth

Hackensack Meridian Health

Jefferson Health. HOME OF SIDNEY KIMMEL MEDICAL COLLEGE

KAISER PERMANENTE®

MAYO CLINIC

NewYork-Presbyterian The University Hospital of Columbia and Cornell

OCHIN

PCCI Pioneering New Ways to Health

UCSF University of California San Francisco

UNIVERSITY OF MICHIGAN

Interested & Affected Parties

AMA

Berkeley UNIVERSITY OF CALIFORNIA

DLA PIPER

IDEO·ORG

Patrick J McGovern FOUNDATION

GORDON AND BETTY MOORE FOUNDATION

PLOS

Workshop Feedback

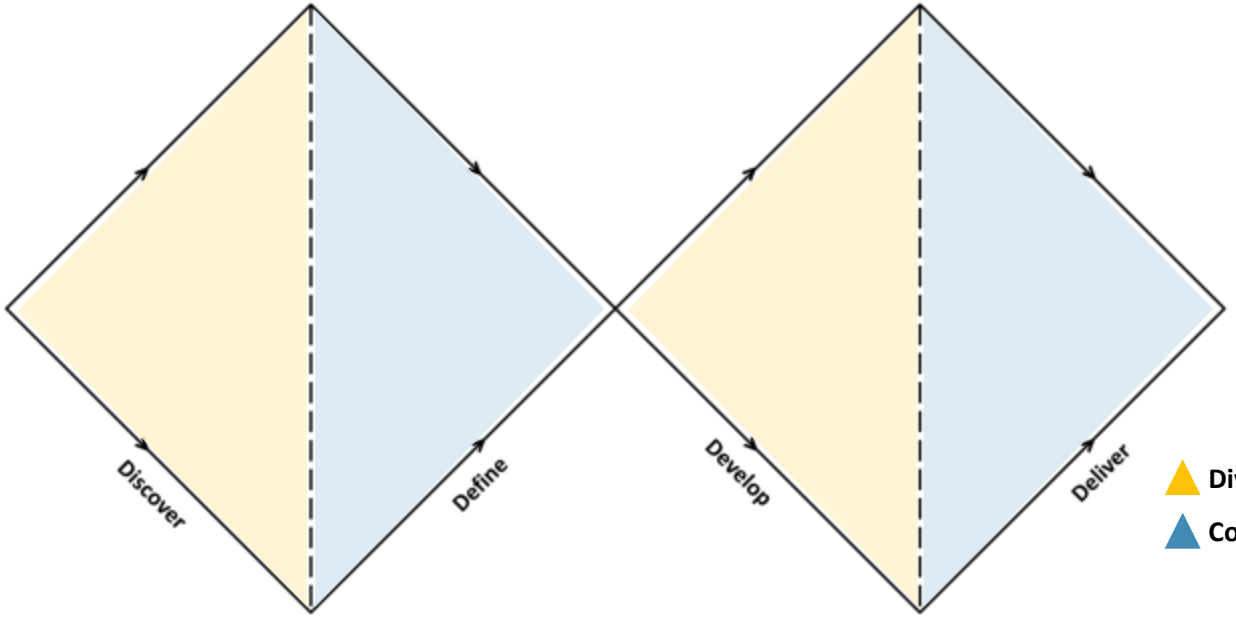
- 77 people attended the workshop (including hosts and the HAIP leadership team), and 30 people provided feedback (~39%)
- Overall Experience: (1 = *Not at all*, 5=*Very much)*

	Satisfaction Overall, how satisfied are you with the current workshop?	**Safeness** How much did you feel that it was a safe space to share your experiences?	**Contribution** How much did you feel like you were able to contribute to the workshop?
Overall	4.40	4.63	3.83
Participant	4.10	4.50	3.70
Case presenter	4.50	5.00	4.25
Expert panelist	4.80	4.80	4.00
Observer	4.45	4.55	3.73

Framework Development Roles

Participant		Role	Responsibilities
C	Case study presenters	3 innovation teams that develop and implement AI solutions in healthcare delivery organizations	Curated a case study, presented it at the workshop and tested out the framework
F	Framework developers	Clinician, community representative, computer scientist, project manager, legal expert, and sociotechnical scholar	Created a scaffolding of the framework and contributed to developing its content
H	HAIP leaders	Clinicians, computer scientists, lawyers, and a community organizer	Evaluated the framework and provided feedback
W	Workshop participants	77 stakeholders from 10 healthcare delivery organizations and 4 ecosystem partners with clinical, technical, operational, regulatory, and AI ethics expertise	Contributed to developing the content of the framework
D	Design researchers	Qualitative research scientist, clinical data scientist, and project manager	Facilitated the co-design process by collecting, iterating, and synthesizing data from all participants

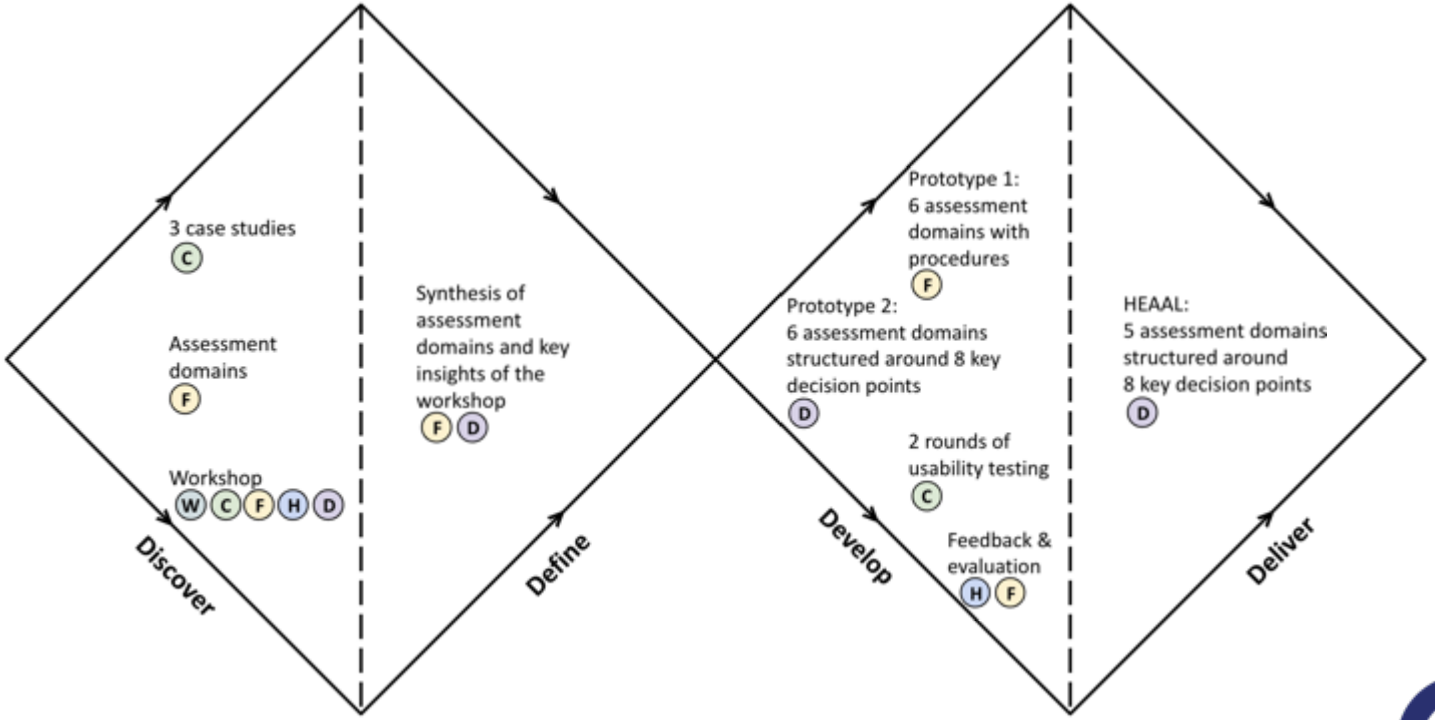
Procedures: Co-design



Divergent thinking process

Convergent thinking process

Procedures: Co-design



Results: Five assessment domains

- 5 assessment domains evaluated across the span of 8 key decision points of AI adoption process

Assessment Domain	Definition
Accountability	Ensures that potential adverse impacts of using the AI solution are overseen by specific stakeholders within healthcare delivery organizations who have clear responsibilities.
Fairness	Ensures that the solution performs equitably across patient subgroups by establishing and evaluating meaningful fairness criteria.
Fitness for purpose	Ensures that the proposed solution solves the identified problem for patient subgroups.
Reliability and validity	Ensures that the solution achieves pre-specified performance targets across technical, clinical, and process measures.
Transparency	Ensures that the processes of model development, implementation, identification of potential risks and harms, and progress towards equity objectives are communicated effectively to end users and patient subgroups.

Results: Procedures

- Detailed step-by-step procedures to conduct in each key decision point
- Procedures tailored to an existing and a new AI solution

Key Decision Point	# of procedures for an existing AI solution	# of procedures for evaluating a newly developed AI solution
1. Identify and prioritize a problem	2	2
2. Define AI product specification	13	5
3. Develop success measures	2	2
4. Design AI solution workflow	5	5
5. Generate evidence of safety, efficacy, and equity	6	11
6. Execute AI solution rollout	3	3
7. Monitor the AI solution	3	3
8. Update or decommission the AI solution	3	3
Total # of procedures	37	34

Results: Key stakeholders

Stakeholder Type	Definition
Strategic (S)	Stakeholders who develop strategic plans and make decisions that align with organizational interests
Operational (O)	Stakeholders who manage workflow and make decisions to integrate
Clinical (C)	Stakeholders who provide clinical care to patients
Technical (T)	Stakeholders who develop the model and its infrastructure
Regulatory (R)	Stakeholders who review the model from regulatory and ethical perspectives
Patient (P)	Stakeholders who receive clinical care and provide insights on their community experiences
Clinical champion	Clinical stakeholders who lead the project and provide clinical expertise in model development
Product manager	Stakeholders who manage the project and communicate with various stakeholders involved in the project

Results: Data sources

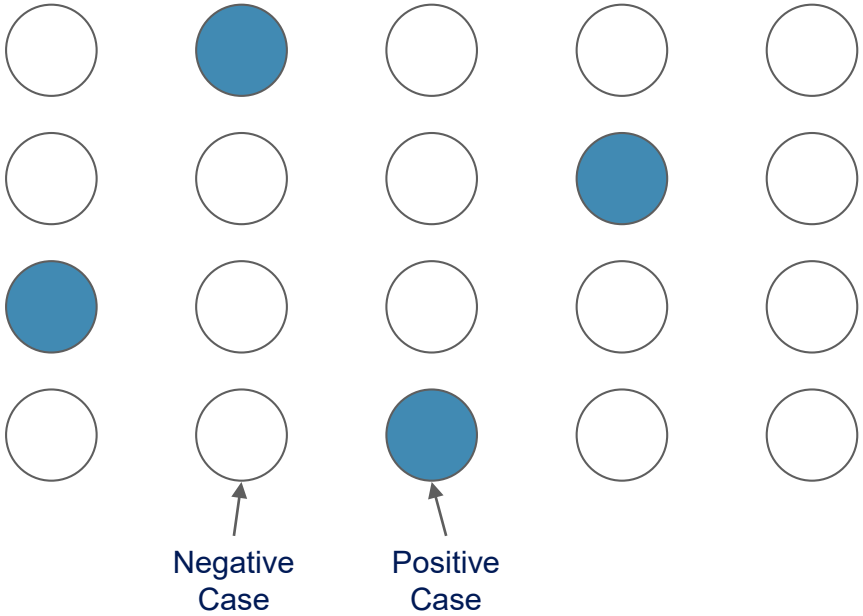
Data Source	Definition
Local healthcare retrospective data	Historical healthcare data that is curated within the primary healthcare delivery organization seeking to adopt an AI product. When a model is internally developed, the local healthcare retrospective data set is used for training the model.
Local healthcare prospective data	Real-time healthcare data that is curated within the primary healthcare delivery organization seeking to adopt an AI product. The local healthcare prospective data set is used for validating a model during a 'silent trial' and for using the model in clinical care.
Local non-healthcare data	Non-healthcare data that is curated within a geographic setting where a healthcare delivery organization is based. The local non-healthcare data can be derived from a variety of external sources, including US Census.
Training data	Data used for training a model. When the model is externally developed, the training data set contains data from an external source.
Literature review	Data collected through reviewing previously published scholarly works on a specific topic.
Qualitative data	Data collected through qualitative research methods, including surveys, focus groups, and interviews.

Health Equity Across the AI Lifecycle (HEAAL) Framework Highlights

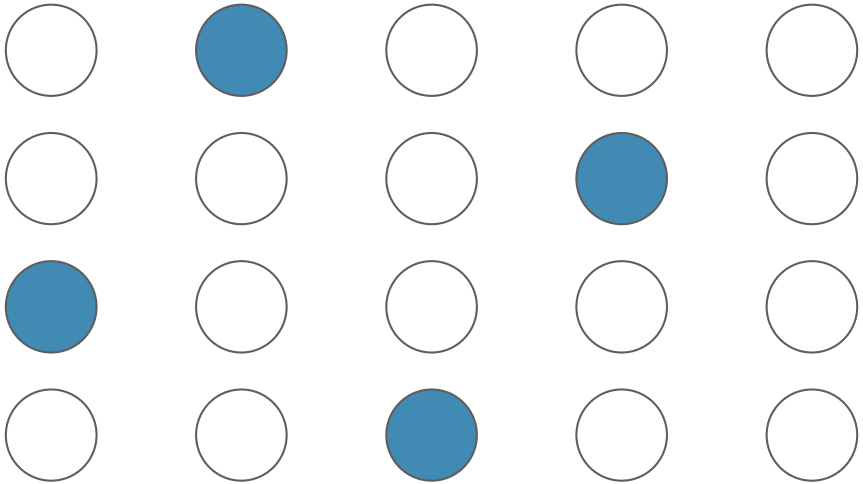
If there's evidence of inequity for the condition of interest in historical data, don't rely on subgroup performance.

Peripheral Artery Disease Case

PAD in white adults

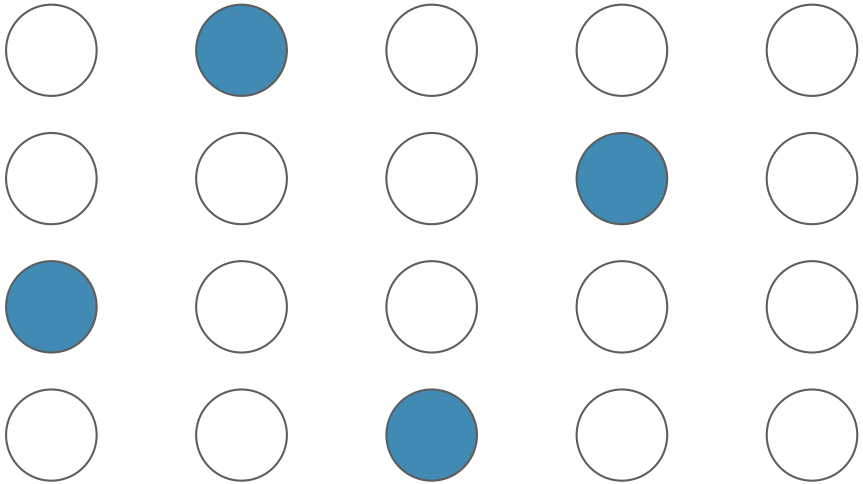


PAD in black adults



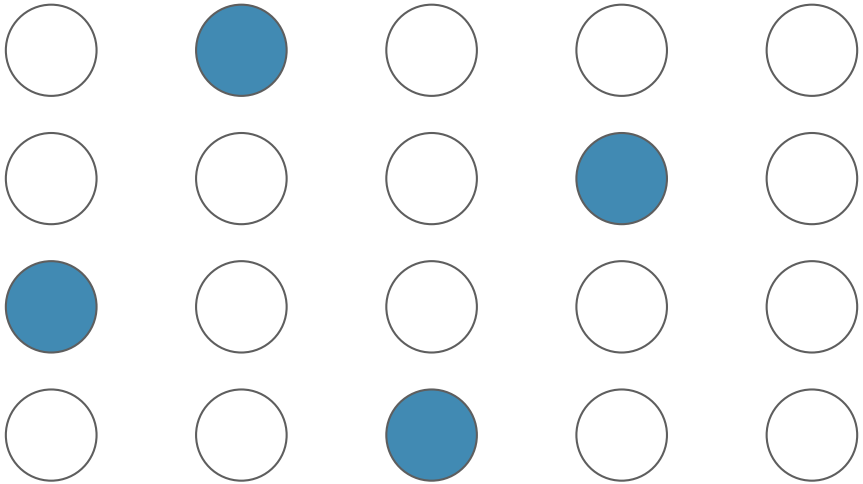
Peripheral Artery Disease Case

PAD in white adults



True prevalence = 4/20 (20%)

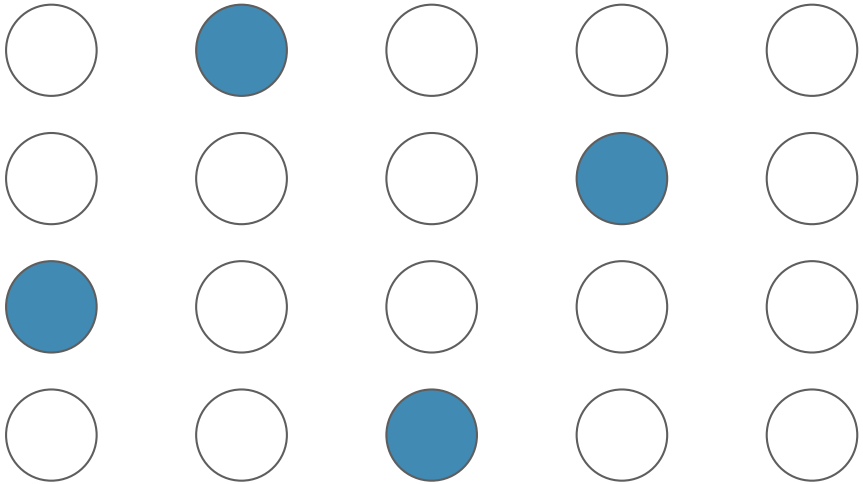
PAD in black adults



True prevalence = 4/20 (20%)

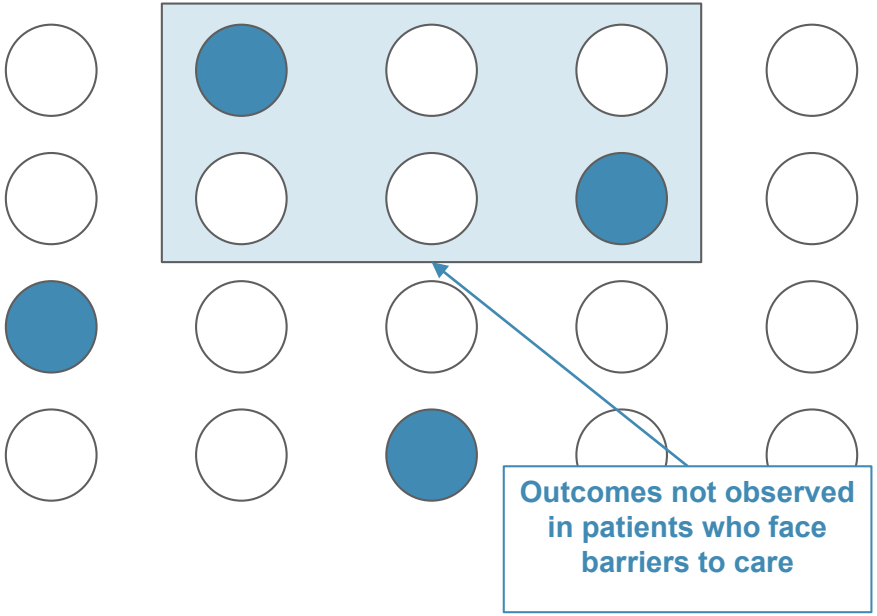
Peripheral Artery Disease Case

PAD in white adults



Observed prevalence = 4/20 (20%)

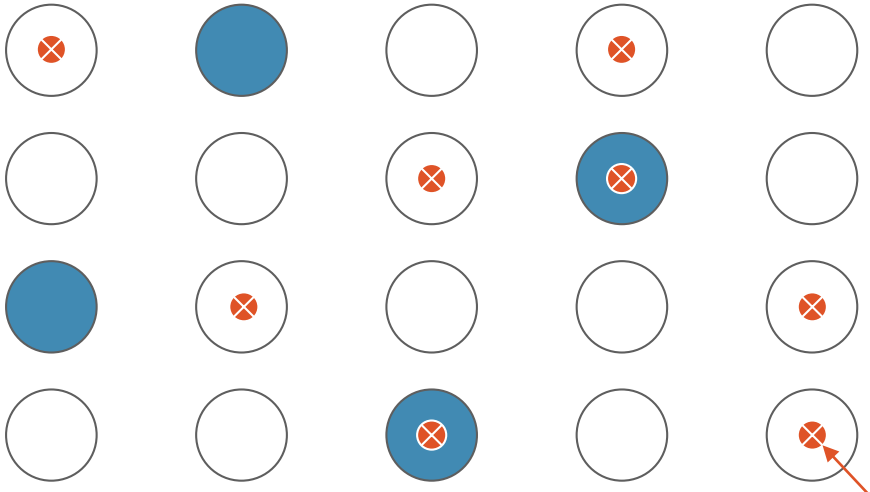
PAD in black adults



Observed prevalence = 2/20 (10%)

Peripheral Artery Disease Case

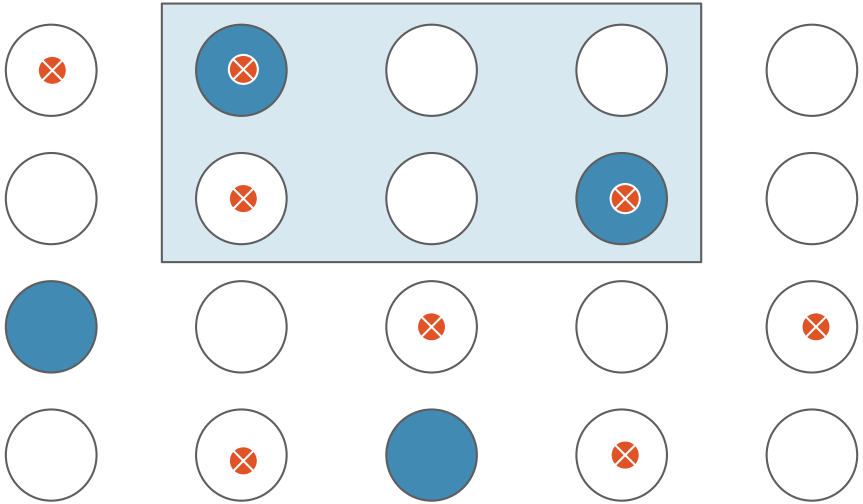
PAD in white adults



Observed PPV = 2/8 (25%)
Observed Sensitivity = 2/4 (50%)

Patient predicted at high risk of PAD

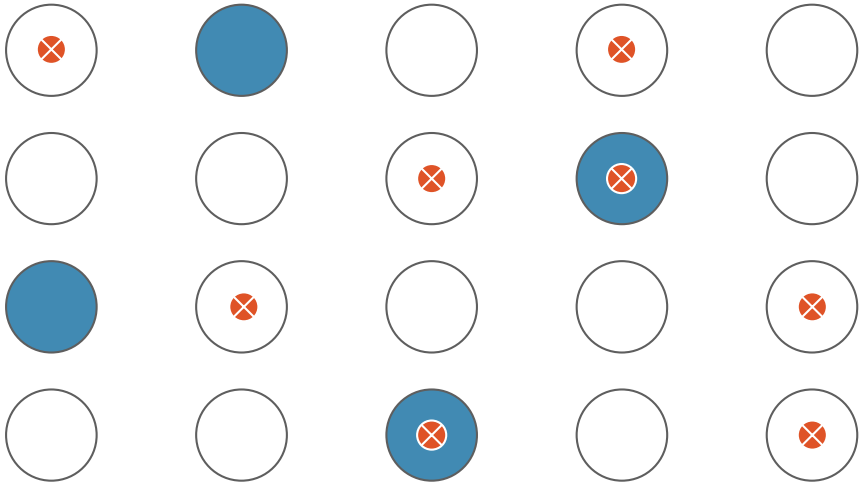
PAD in black adults



Observed PPV = 0/8 (0%)
Observed Sensitivity = 0/4 (0%)

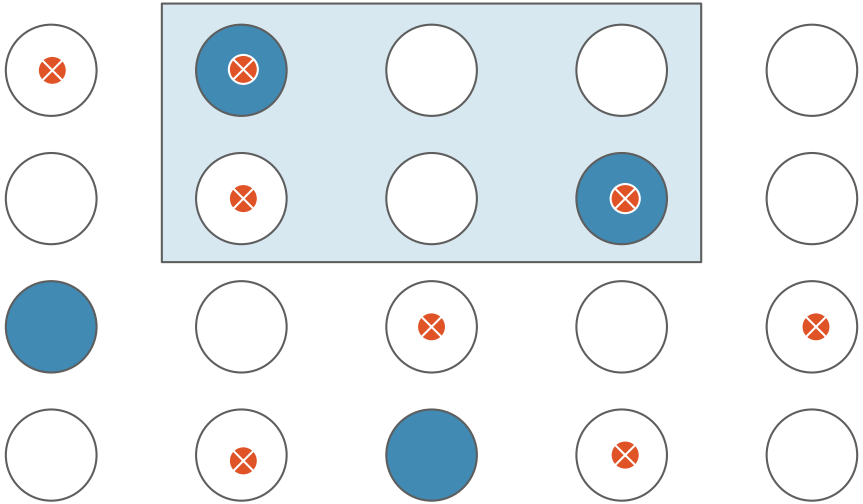
Peripheral Artery Disease Case

PAD in white adults



Observed PPV = 2/8 (25%)
Observed Sensitivity = 2/4 (50%)

PAD in black adults



True PPV = 2/8 (25%)
~~Observed PPV = 0/8 (0%)~~
~~Observed Sensitivity = 0/4 (0%)~~
True Sensitivity = 2/4 (50%)

Peripheral Artery Disease Case

PAD in white adults

PAD in black adults

In this scenario, the model performs worse on Black patients because of a diagnosis inequity. If the diagnosis inequity were addressed, the model performance on Black patients would be the same as on White patients.

In cases like this, you cannot accurately assess model performance within the disadvantaged subgroup. You need to test the model prospectively in a way that addresses inequities to accurately assess performance across advantaged and disadvantaged subgroups.

Observed PPV = 2/8 (25%)
Observed Sensitivity = 2/4 (50%)

True PPV = 2/8 (25%)
~~Observed PPV = 0/8 (0%)~~
~~Observed Sensitivity = 0/4 (0%)~~
True Sensitivity = 2/4 (50%)

Engage with our community of practice!

health ai partnership

AI INITIATIVE

Franciscan Missionaries of Our Lady HEALTH SYSTEM

KAISER PERMANENTE®

OCHIN

Berkeley UNIVERSITY OF CALIFORNIA

AMA

GORDON AND BETTY MOORE FOUNDATION

MAYO CLINIC

ONC Office of the National Coordinator for Health Information Technology *Participating as a federal observer

UCSF University of California San Francisco

DLA PIPER

Hackensack Meridian Health

MICHIGAN MEDICINE UNIVERSITY OF MICHIGAN

PCCI

UWHealth

DukeHealth

Jefferson Thomas Jefferson University

NewYork-Presbyterian

scan HEALTH PLAN

VA

Website healthaipartnership.org

Bibliography

Sendak MP, Balu S, Schulman KA. Barriers to Achieving Economies of Scale in Analysis of EHR Data. A Cautionary Tale. Applied Clinical Informatics. 2017 Aug 9;8(3):826–31. doi: 10.4338/ACI-2017-03-CR-0046

Brajer N, Cozzi B, Gao M, Nichols M, Revoir M, Balu S, et al. Prospective and External Evaluation of a Machine Learning Model to Predict In-Hospital Mortality of Adults at Time of Admission. JAMA Network Open. 2020 Feb 7;3(2):e1920733-14. doi: 10.1001/jamanetworkopen.2019.20733

Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model information to clinical end users with model facts labels. npj Digital Medicine. 2020 Mar 15;3(41):1–4. doi: 10.1038/s41746-020-0253-3

Sendak M, Sirdeshmukh G, Ochoa T, Premo H, Tang L, Niederhoffer K, et al. Development and Validation of ML-DQA -- a Machine Learning Data Quality Assurance Framework for Healthcare. Journal of Machine Learning Research. 2022 Aug 5. doi: 10.48550/arxiv.2208.02670

Sandhu S, Sendak MP, Ratliff W, Knechtle W, Fulkerson WJ, Balu S. Accelerating health system innovation: principles and practices from the Duke Institute for Health Innovation. Patterns. 2023;4(4):100710. doi: 10.1016/j.patter.2023.100710

Kim JY, Boag W, Gulamali F, Hasan A, Hogg HDJ, Lifson M, et al. Organizational Governance of Emerging Technologies: AI Adoption in Healthcare. ACM Conference on Fairness, Accountability, and Transparency (FAccT). 2023 Jun 12. doi: 10.1145/3593013.3594089

Wang SM, Hogg HDJ, Sangvai D, Patel MR, Weissler EH, Kellogg KC, et al. Development and Integration of Machine Learning Algorithm to Identify Peripheral Arterial Disease: Multistakeholder Qualitative Study. JMIR Form Res. 2023;7:e43963. doi: 10.2196/43963

Jee Young Kim, Alifia Hasan, Kate Kellogg, William Ratliff, Sara Murray, Harini Suresh, Alexandra Valladares, Keo Shaw, Danny Tobey, David E Vidal, Mark A Lifson, Manesh Patel, Inioluwa Deborah Raji, William Boag, Linda Tang, Shems Saleh, Suresh Balu, Mark P Sendak. Development and preliminary testing of Health Equity Across the AI Lifecycle (HEAAL): A framework for healthcare delivery organizations to mitigate the risk of AI solutions worsening health inequities. medRxiv 2023.10.16.23297076; doi: https://doi.org/10.1101/2023.10.16.23297076

Thank you

mark.sendak@duke.edu
suresh.balu@duke.edu