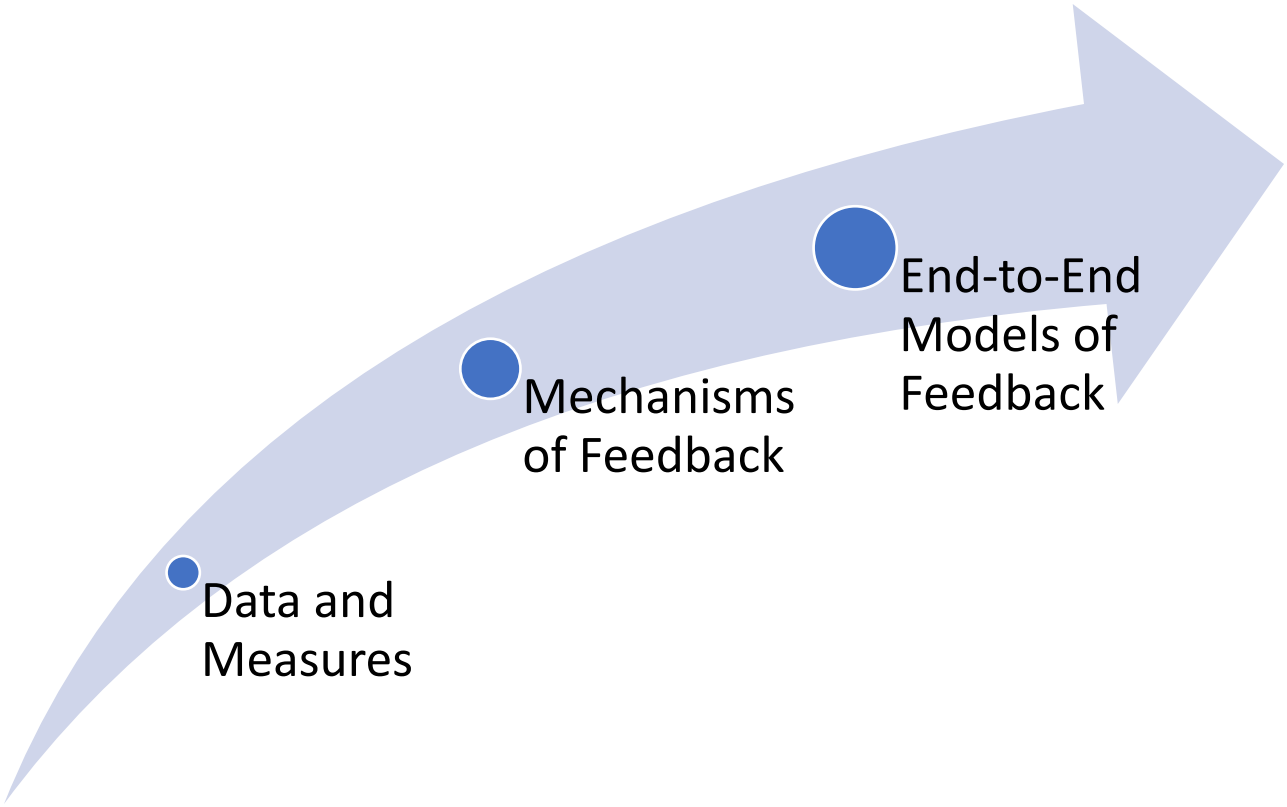
Leveraging Informatics Solutions to Support Clinician Feedback

Julia Adler-Milstein, PhD
UCSF
Chief, Division of Clinical Informatics & Digital Transformation
Director, Center for Clinical Informatics & Improvement Research

MPOG Retreat
October 13, 2023

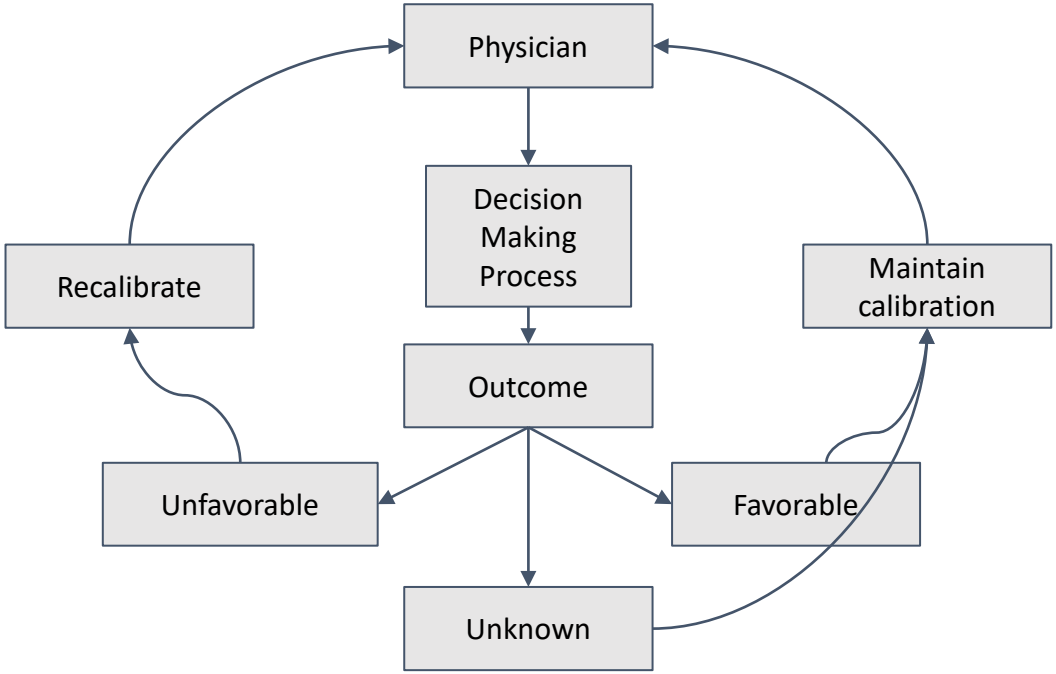
Overview of Talk



Feedback is the de facto means to **improve performance** in many fields of human performance.

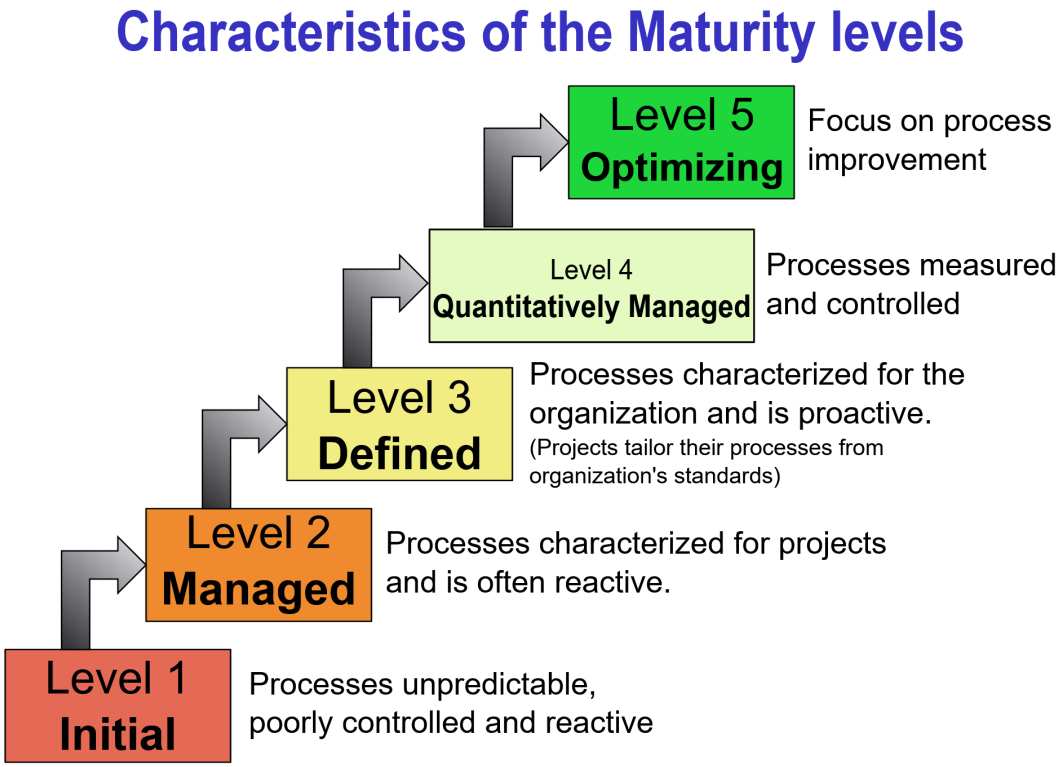
Fernandez Branson C, Williams M, Chan TM, Graber ML, Lane KP, Grieser S, et al. Improving diagnostic performance through feedback: the Diagnosis Learning Cycle. BMJ Qual Saf. 2021 Dec;30(12):1002–9.

Individual Performance Improvement - Feedback



Adapted from: Croskerry, Acad Emerg Med, Nov 2000, Vol 7, No 11, 1232-8

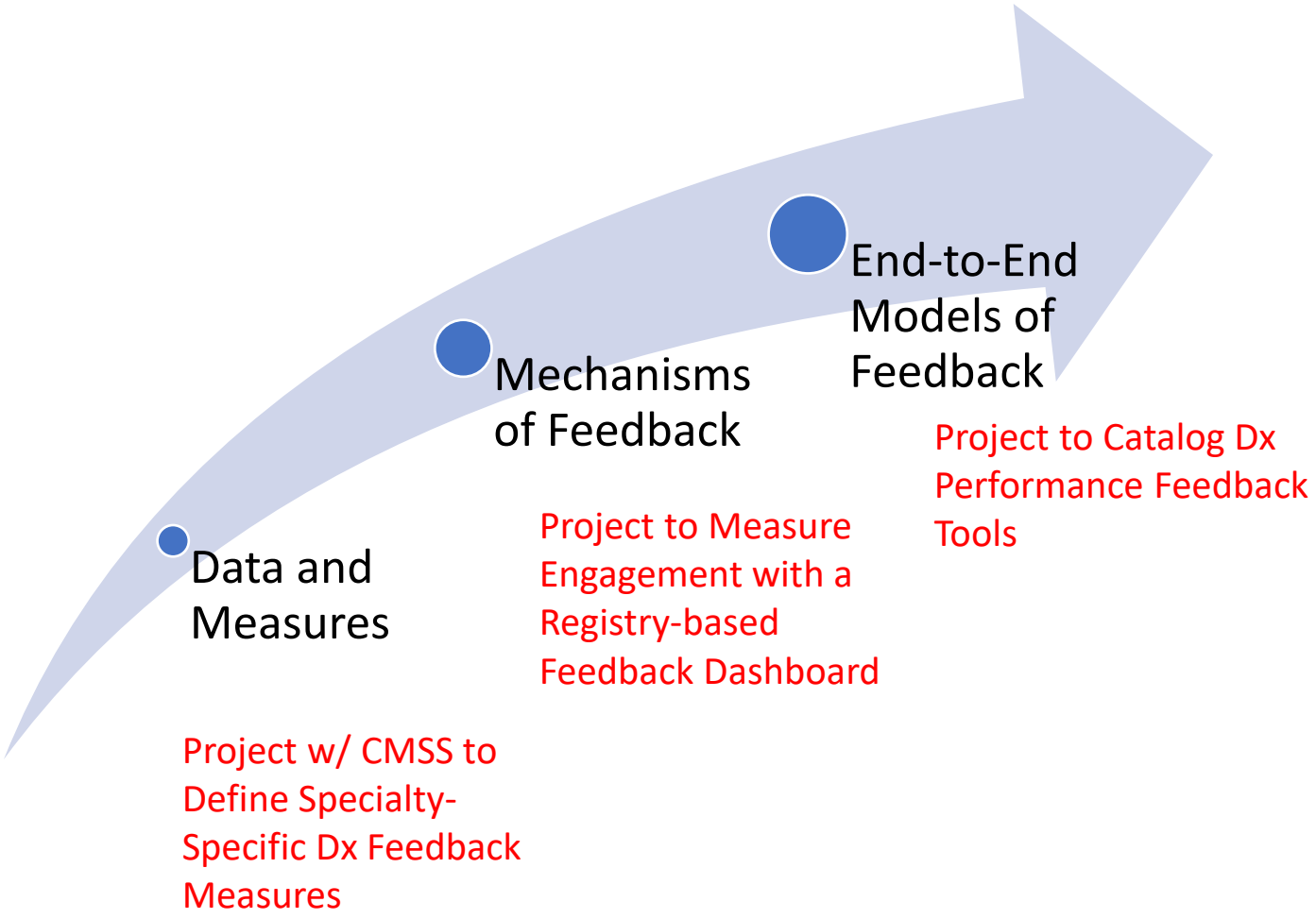
Organizational Performance Improvement - Feedback



The many data & measurement components of feedback (clinician-level)

ID the diagnostic encounter	ID the diagnostician	ID unclosed loops on diagnostic testing	ID diagnoses and potential missed opportunities for diagnosis (MODs)	Built in mechanism to adjudicate (or support adjudication) of missed opportunities for diagnosis (MODs)	Calculate diagnostic performance	Render performance (metrics or visualization)	Built in mechanism to share performance or results with reviewer or diagnostician
The resource is capable of identifying an encounter as one involving diagnostic decision-making (as opposed, for example, to one in which only chronic condition management is carried out.)	The resource is able to correctly identify the diagnostician who has diagnostic responsibility for the diagnostic encounter, and to whom diagnostic performance feedback should be sent.	The resource is able to identify, for example, test results that are completed, but which have not been reviewed or acted on by the diagnostician.	The resource is able to identify potentially correct and incorrect diagnoses.	The resource has mechanisms to check and/or review algorithmically-detected potential missed opportunities for diagnosis so that performance results that are shared back to the diagnostician are as accurate as possible and minimize false positive and false negative signals.	The resource is able to provide a calculated performance metric related to diagnostic accuracy across a population of patients.	The resource is able to represent diagnostic performance in a visualizable graphic or table.	The resource has a built in mechanism to return the rendered performance back to the diagnostician or health system leader.

Three Projects on Feedback



Building a Library

Develop an online, publicly available, free-to-use diagnostic performance feedback resource library for the public good

- Single, unbiased, continuously evolving location to find implementable resources for diagnostic performance feedback
- User experiences posted from others with similar job titles de-risks implementation decisions

Engaging a Broad Set of Collaborators

Johns Hopkins Kelly Gleason, Allen Kachalia	AHRQ Margie Shofer
University of Michigan Zach Landis-Lewis	Premier Blair Childs and Mike Grow
Department of Veterans Affairs, Baylor College of Medicine Ashley Meyer	Vizient, Inc. Ellen Flynn
University of Minnesota Andrew Olson	Mayo Clinic Platform John Halamka
University of California, San Francisco Gurpreet Dhaliwal, Julia Adler-Milstein, Glenn Rosenbluth, Andrew Auerbach, Ben Rosner	HCA Healthcare Chris Ott
University of California, San Diego Rob El-Kareh	CRICO Dana Siegel
Kaiser Permanente Michael Kanter	The Doctors Company Leslie Castaneda
Gordon and Betty Moore Foundation Daniel Yang	Intermountain Healthcare Mike Woodruff

Example

Hip x-ray was obtained to evaluate for hip pain. Was read as negative. Radiologist missed a lytic bone lesion. Led to nearly 2-month delay in new diagnosis of likely metastatic cancer.

Relevant tool to support feedback:

RADPEER

RADPEER is a tool that allows peer review to be performed during routine image interpretation. Discordant interpretations will be marked. After submission of practice data, the group chair or medical director can access the reports online at any time which include summary statistics and comparisons by modality for every participating physician, group summary data by modality, and data summed across all RADPEER participants.

Source: Gooddx.org

*Facts contained in illustration 2 have been altered from the original case

Curating the Collection

Any tool, framework, program, or technology, that has been or can readily be automated or semi-automated, and that leads to an endpoint of measuring and providing individual and/or aggregate diagnostic performance feedback to the clinician, healthcare team, and/or system.

This excludes interventions that reduce diagnostic errors without providing feedback for the purpose of clinician recalibration.

We are live!

Imagine a world of diagnostic excellence fueled by continuous diagnostic feedback.

GoodDx.org

Source: Gooddx.org image used with permission

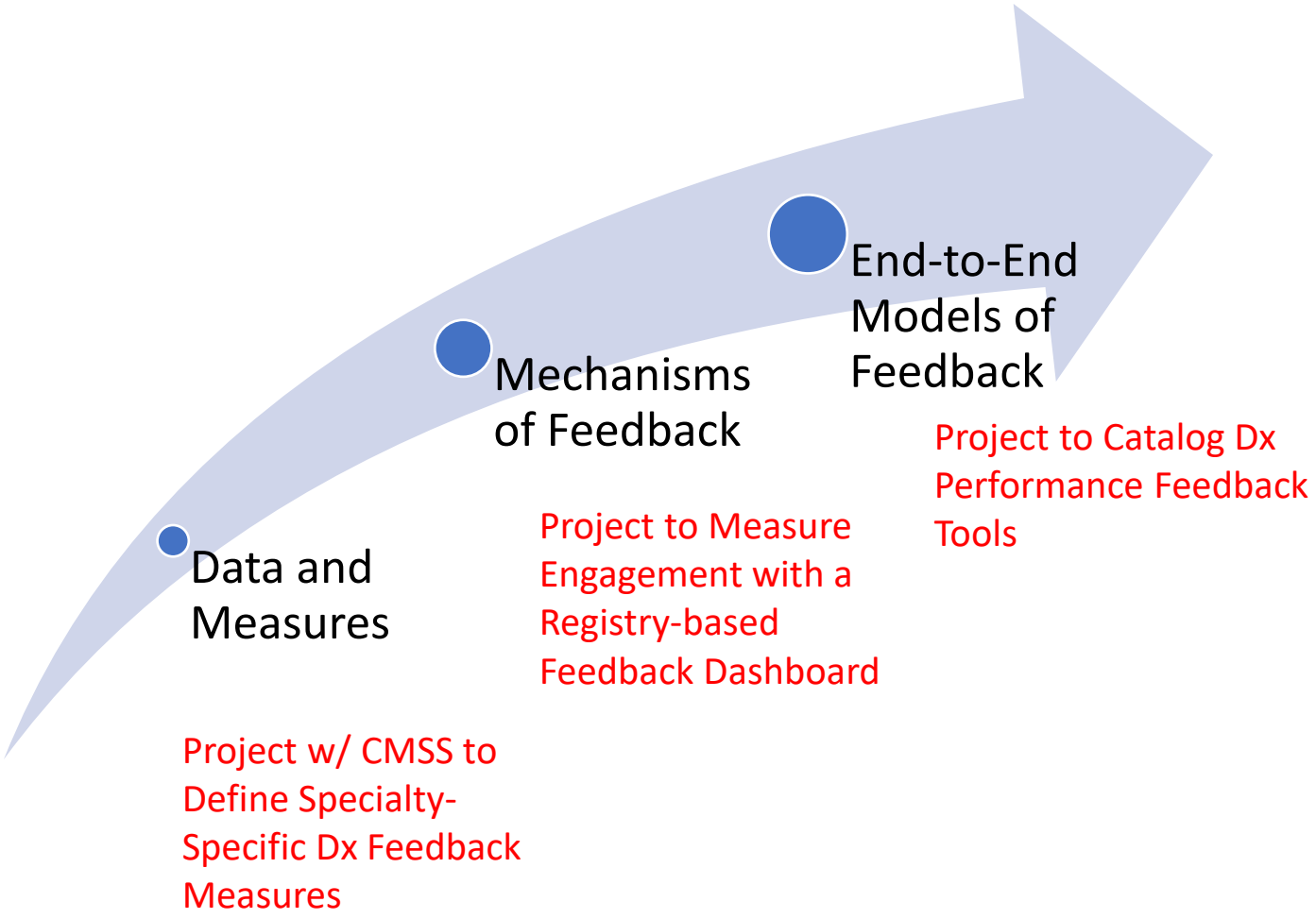
52 Resources Identified and Included

	ID the diagnostic encounter ⓘ	ID the diagnostician ⓘ	ID unclosed loops on diagnostic testing ⓘ	ID diagnoses and potential missed opportunities for diagnosis (MODs) ⓘ	Built in mechanism to adjudicate (or support adjudication) of missed opportunities for diagnosis (MODs) ⓘ	Calculate diagnostic performance ⓘ	Render performance (metrics or visualization) ⓘ	Built in mechanism to share performance or results with reviewer or diagnostician ⓘ
RapRad, Radiology Dx for Pneumothorax ⓘ	✓	✓	−	✓	✓	✓	✓	✓
Coverys (Diagnosis Related Claims Analysis) ⓘ	✓	✓	✓	✓	✓	◐	−	✓
Symptom-Disease Pair Analysis of Diagnostic Error (SPADE) ⓘ	✓	✓	−	✓	✓	✓	✓	−
Candello Explore by CRICO ⓘ	✓	✓	−	✓	✓	−	−	✓
Case Share App - Bay State Medical Center ⓘ	✓	✓	−	✓	✓	−	−	✓

Source: Gooddx.org image used with permission

https://gooddx.org/overview

Three Projects on Feedback



Project Overview

Advancing National-Scale Diagnostic Feedback to Frontline Clinicians

Problem Statement	There are few systematic solutions for clinicians to receive consistent and timely diagnostic performance feedback.
Approach	Illustrate the ability to solve this problem through supporting the development of scalable and durable solutions for the use of diagnostic performance feedback in clinical practice.
Target Participants	Specialty societies have unique infrastructure and expertise to support the development of full-scale solutions (breadth and uptake).

CMSS | Council of Medical Specialty Societies

Project Goals

1. Demonstrate that specialty-tailored, systematic solutions for diagnostic feedback (for the purpose of ongoing diagnostic recalibration) are **feasible at scale** in a select number of specialties.

2. Determine which **data source**(s) and **data infrastructure**(s), including clinical registries, exist to support real-world implementation of diagnostic feedback at scale

3. **Develop infrastructure** and **incentives** to promote the **routine review** of diagnostic performance by clinicians.

CMSS | Council of Medical Specialty Societies

High-level Overview of Capabilities from RFI

Number of registries per submission	9 out of 10 submitted information on 1 registry 1 does not have a registry
Provides individual patient-specific diagnostic feedback	4 reported yes 6 reported no
Provides aggregate diagnostic feedback	4 reported yes 6 reported no
Registry focus	Specialty wide – 6 Multi-specialty – 1 Disease/symptom-specific – 2
Data sources	Predominantly EHRs and practice management but some registries use other data sources
Longitudinal data	7 of 9 registries reported yes
Non-registry data sources available	4 reported yes 6 reported no
Technical capability	8 registries were hybrid solutions (build/buy) 1 was in house (build only)

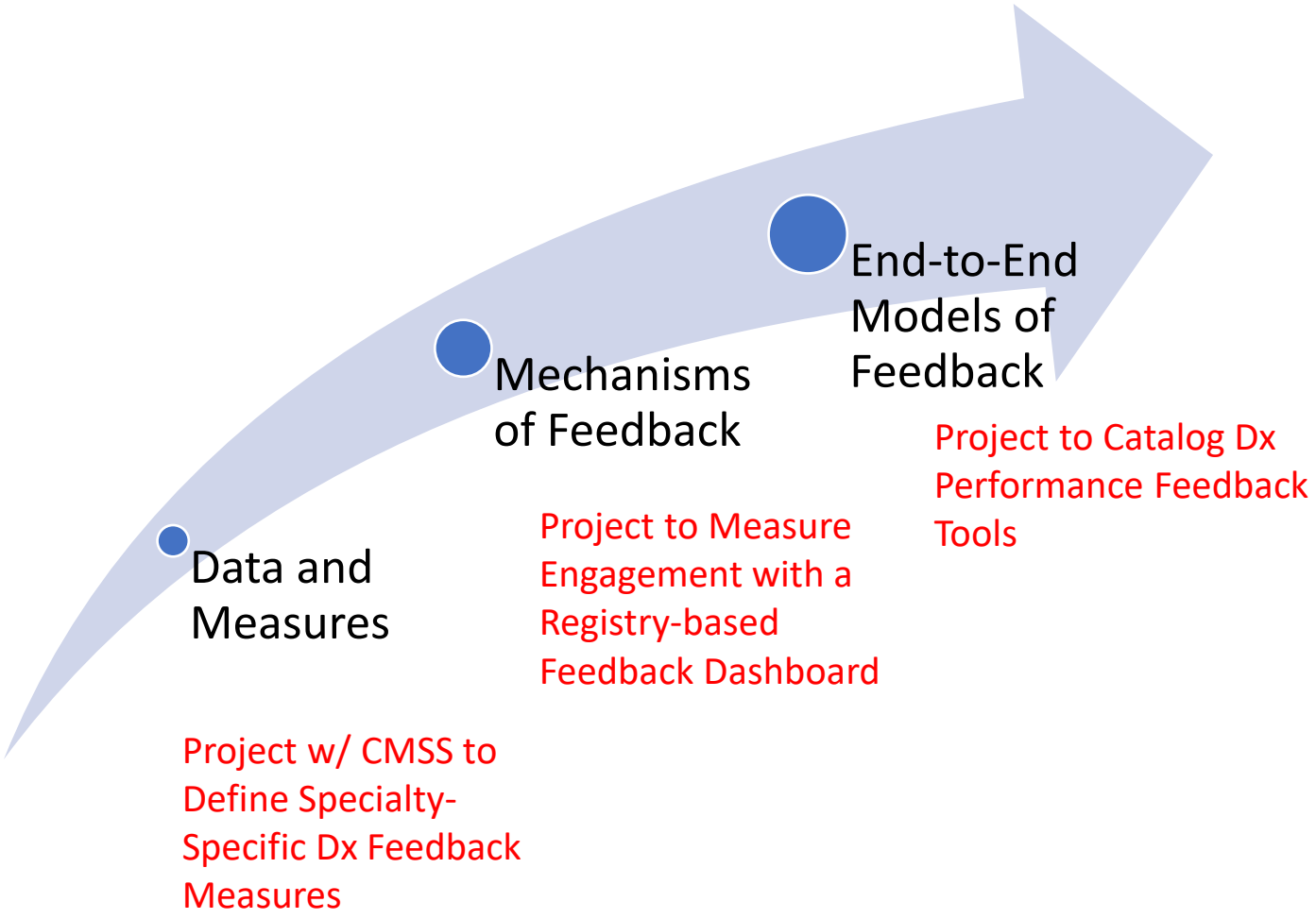
RFI Responses: Opportunities for individual clinician performance/individual patient tracking

- Use registry dashboard and patient level data to identify performance gaps
- Develop quality measures to evaluate and compare clinicians and track performance on individual patients
- Leverage longitudinal data to track practice patterns

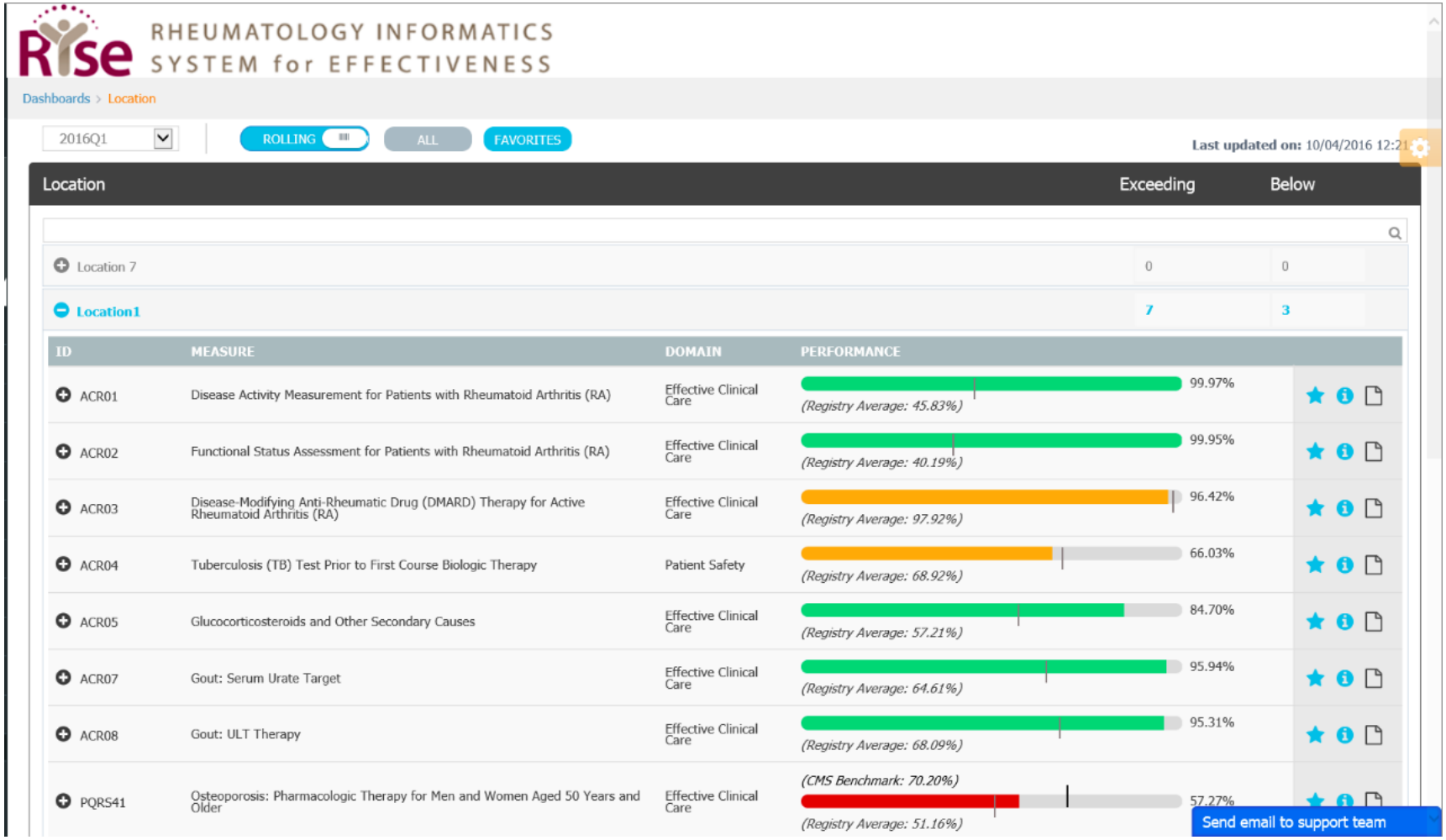
- Use **lab data** to measure not just [XXX] performance, but the performance of ordering providers (appropriate test utilization), as well as patient trajectories through monitoring test results
- Use **discrete data** from the EHR (e.g., blood pressure measurements, abnormal lab values) to identify patients who meet the criteria for a specific diagnosis but do not have one documented
- Use **NLP on non-discrete data** to alert participants of gaps in diagnosing certain conditions and comparisons to guidelines

- Evaluate appropriate testing of patients for common clinical conditions informed by the patient's observable characteristics
- Track patient treatment patterns in the primary care setting following confirmed testing and diagnosis of the clinical condition during the follow-up period

Three Projects on Feedback



Impact of Registry Use on Quality and Outcomes in Rheumatology (QORA)



QORA Aims

- Measure physician and practice-level variation in patterns of RISE dashboard use by analyzing the extensive audit-log data in the RISE data warehouse

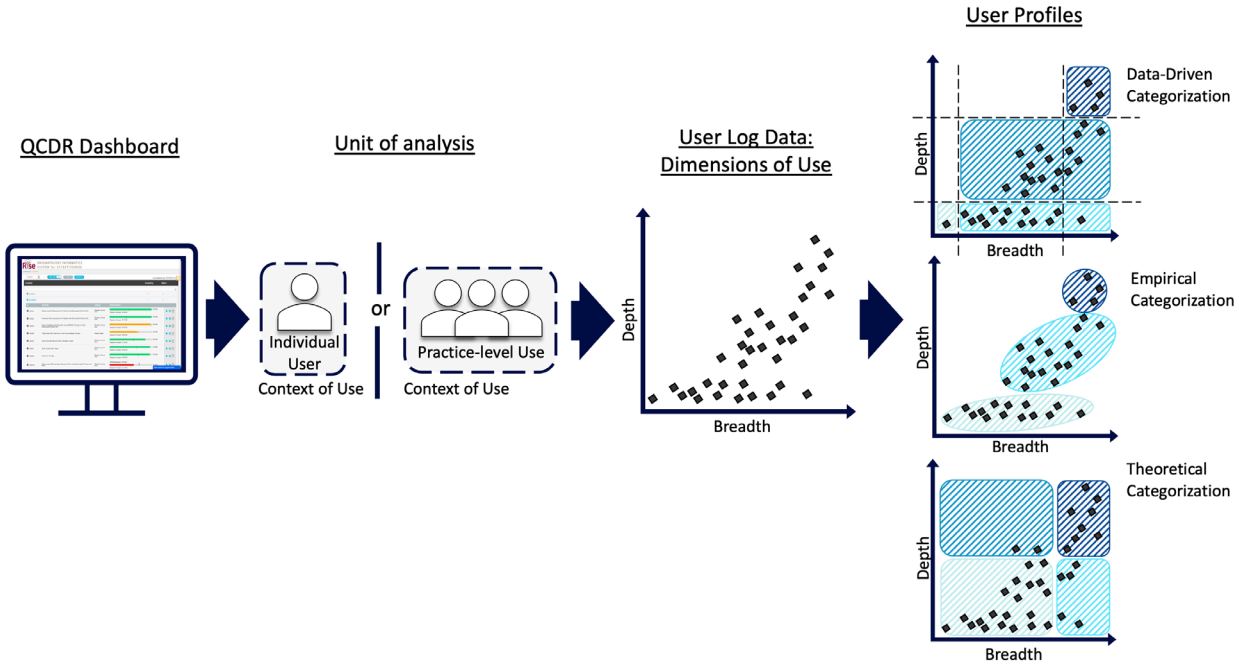
- Investigate the impact of participation and engagement with the RISE registry on quality of care and clinical outcomes

- Identify strategies for redesigning tools available on the RISE dashboard to improve its usability and its effectiveness for local quality improvement.

Breadth-Depth-Context (BDC) Framework: Customizable Domains for defining practice engagement with a QCDR dashboard

Domain	Concept	Considerations for Measurement
Dimensions of use	Breadth	Number of unique dashboard sessions represents one aspect of engagement.
	Depth	Use of additional, advanced functions available through the dashboard represent another aspect of engagement. Users can access advanced functions such as generation of patient-level reports or data exports. These can be performed for selected quality measures or for all quality measures.
	Patterns of Breadth or Depth	Patterns of breadth or depth, such as consistency of sessions or actions (how regularly sessions or actions occur) or temporality (what time of year sessions or actions occur), can also be assessed.
Context of use	Individual Setting	Characteristics specific to individual users, such as user roles, years of dashboard use, and personal beliefs about dashboard accuracy, accessibility, and utility.
	Inner Setting	Characteristics specific to the practice that individual users are affiliated with, such as practice infrastructure, availability of personnel to review practice data, EHR vendor, and patient-case mix characteristics.
	Outer Setting	Characteristics of the sociopolitical environment in which the practice operates, such as policies that incentivize population health management.

Conceptual Approach to Generate Engagement Profiles

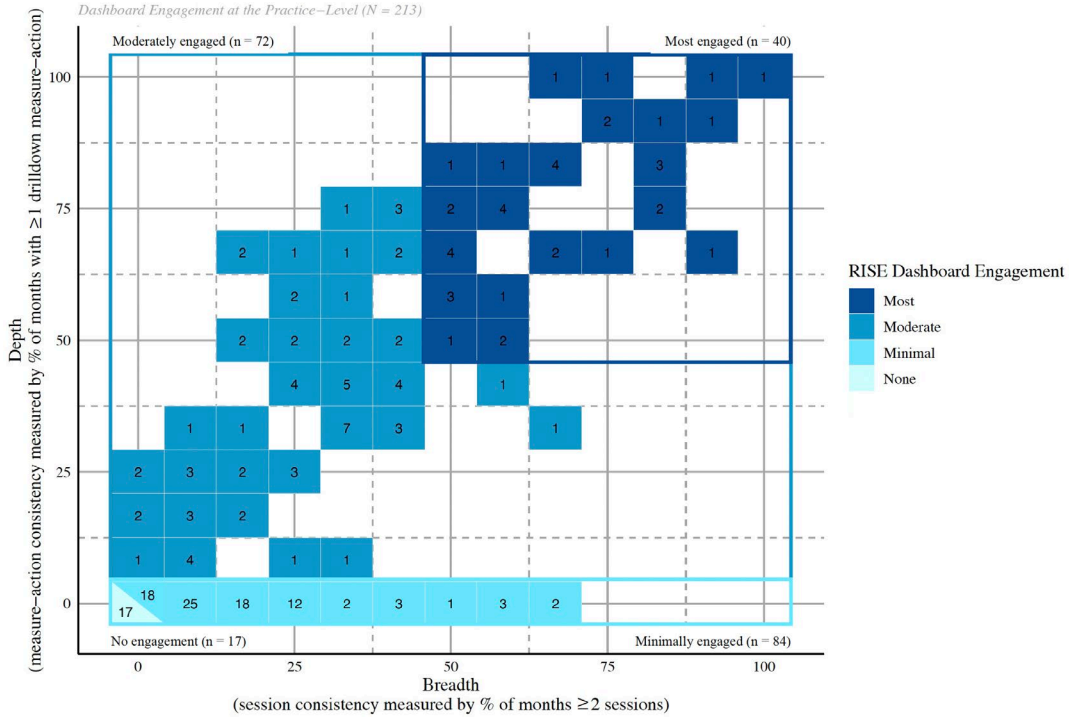


Kersey et al. Under review

BDC (Breadth-Depth-Context) Framework Dimensions of Use Domain: Sample metrics for assessing breadth and depth of engagement

Metric	Metric calculation	Example
Breadth		
Session count	The total number of sessions logged by a user during the period of interest. Multiple sessions are possible on a given day.	If a user logs into the dashboard twice on Monday and once on Thursday, then their session count is 3 for the week.
Session duration	Total or average time spent accessing the dashboard, across all sessions during the period of interest.	If a user logs into the dashboard for a two-hour session on Monday and a one-hour session on Thursday, then their average session duration is 1.5 hours for the week.
Session consistency	The percentage of time (months, weeks, etc.) during the period of interest for which a user had a given number of sessions.	If a user logs into the dashboard on Monday and Thursday, then their session consistency is calculated as 2/5 of days in a work week (40%).
Depth		
Measure-action count	Since measure-level actions can be applied to any number of user-selected quality measures, a measure-action count quantifies the number of quality measures interacted with during a given action. Thus, a measure-action count equals the number of measure-level actions multiplied by the number of quality measures it was applied to, per session or period of interest. A measure-action count does not apply to a summary-level action (i.e., generation of performance summary report). Summary-level actions automatically compile all aggregated quality performance data without requiring a user to individually select measures or examine patient-level data.	If a user viewed patient-level data for 3 quality measures and exported patient-level data for 1 quality measure, then they completed 4 measure-actions (3 quality measures viewed + 1 quality measure exported) during the session.
Action-type count	A count of the different types of actions that occurred, per session or period of interest for all measure-level or summary-level actions performed.	If a user viewed patient-level data for 3 quality measures, exported patient-level data for 1 quality measure, and generated a performance summary report, then they completed 3 distinct action-types (view + export + summary report) during the session.
Action consistency	The percentage of time (month, weeks, etc.) during the period of interest for which a user completed a given number of action-types or measure-actions.	If a user viewed patient-level data for 3 quality measures on Monday and exported patient-level data for 2 quality measures on Thursday, then this can be calculated as a measure-action consistency of ≥ 3 measure-actions 1/5 (20%) of days in a workweek or a measure-action consistency of ≥ 2 measure-actions 2/5 (40%) of days in a workweek. Alternatively, these actions could represent an action-type consistency of ≥1 action-type 2/5 (40%) of days in a workweek.

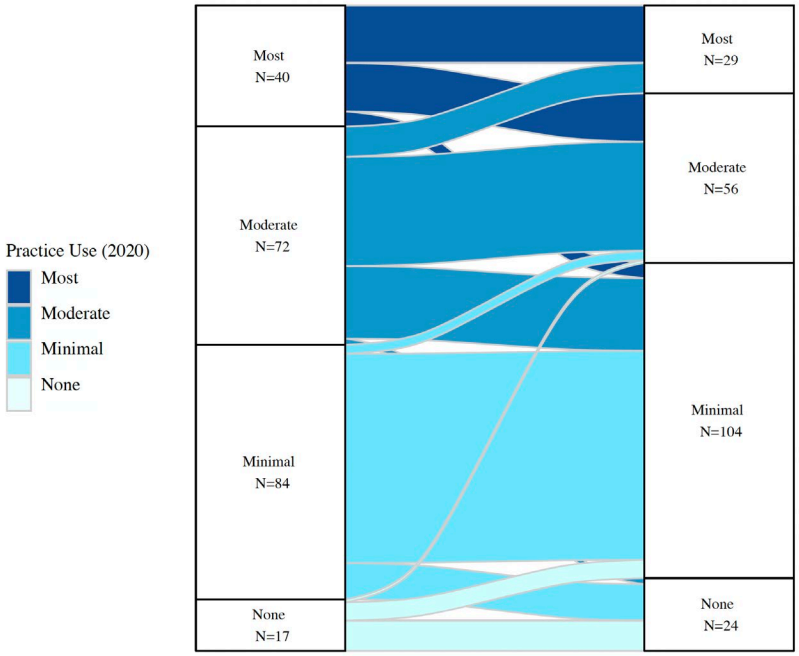
Applying the BD Part of the Framework



- **213 ambulatory practices from the RISE registry in 2020-2021**
- **4 engagement profiles:**
  - **not engaged (8%)**
  - **minimally engaged (39%)**
  - **moderately engaged (34%)**
  - **most engaged (19%)**

Kersey et al. Under review

Change over Time



- **The majority of practices (63%) stayed in the same profile from one year to the next**
- **9% of practices moved into a more engaged profile**
- **28% moved into a less engaged profile**

Kersey et al. Under review

Adding in the "C" Part of the Framework

- In adjusted models, practices had a higher likelihood of being in the most engaged group if they had:
  - more patients (>5000)
  - specific EHR vendors (eClinicalWorks and eMDs)

QORA Aims

- Measure physician and practice-level variation in patterns of RISE dashboard use by analyzing the extensive audit-log data in the RISE data warehouse

- **Investigate the impact of participation and engagement with the RISE registry on quality of care and clinical outcomes – looking promising but stay tuned for final results☺**

- Identify strategies for redesigning tools available on the RISE dashboard to improve its usability and its effectiveness for local quality improvement.

Summary Conclusions