Unlocking New Opportunities for AI-enabled Diagnosis

February 2022

The Gordon and Betty Moore Foundation

The Center for Clinical Informatics and Improvement Research

UCSF

Table of Contents

Executive Summary

Applications of data and technology in the diagnostic domain have significantly advanced over the past decade, but progress has been uneven. A notable gap is the lack of AI tools in use at scale to support clinicians as they undertake the diagnostic process. Given the foundational role of the diagnostic process, it is critical to understand what may be impeding routine use of such tools. With support from The Gordon and Betty Moore Foundation, four workgroups - with experts in data science, diagnosis, and artificial intelligence from multiple sectors of the healthcare industry - were convened to discuss this topic. Emerging from workgroup discussions was the insight that current AI tools are largely built to answer the binary question of whether a given patient has diagnosis XYZ (relying on the diagnostic label). This approach ignores the diagnostic process itself - the information gathering, processing, and decision-making that the clinician engages in upstream of arriving at the diagnosis.

To develop tools that support the clinician's diagnostic process instead of the final label, workgroup discussions were adapted into a framework that identifies the key activities that comprise the clinician's dynamic, longitudinal diagnostic process. The three core activities are 1) information gathering, organization, and prioritization; 2) information integration and interpretation; and 3) formulation of next steps. Using this framework, we illustrate opportunities to develop AI "wayfinding" tools that better support the diagnostic process. Next, we identify the types of data assets that need to be available to develop such tools. These include two broad categories of data - traditional patient-centric (clinical) data and new clinician-centric data that reflects their actions during the diagnostic process and the contextual factors surrounding the clinician and patient during this diagnostic process. Lastly, we address broader healthcare system drivers that could speed development of AI tools that support the diagnostic process. These include strengthening incentives specifically for improved diagnostic performance and better characterizing the dynamic diagnostic refinement processes for different common presenting symptoms. Taken together, these ideas can be taken forward to help spur varied stakeholders to more effectively work towards advancing AI tools that support the diagnostic process, thereby improving health system performance.

Background & Introduction

The availability of digitized health data has expanded extensively in the past decade alongside significant growth in associated technologies - spanning broad enterprise EHR platforms to targeted point-of-care prediction tools. These data and technologies are being applied in myriad ways to assess patient health status and inform care. However, applications of data and technology to *support clinicians during the diagnostic process* are underdeveloped. There are few examples of diagnostic clinical decision support in routine use. This represents a missed opportunity to improve diagnostic performance. It is estimated that nearly all Americans will experience a diagnostic error in their lifetime and nearly 12 million experience a diagnostic error each year.[1,2]

Applications of data and technology in the diagnostic domain have advanced unevenly. There are many examples of solutions for well-characterized, stand-alone diagnostic questions - such as whether an x-ray shows the presence of a pneumothorax. However, these solutions do not address the more common scenario of a patient presenting with symptom(s) where the clinician iteratively works through possible diagnoses to direct treatment. It is unclear why applications of data and technology to *support clinicians during the diagnostic process* have not been adopted at scale.

At first glance, it does not appear to be due to lack of effort: academic researchers and private-sector companies are actively pursuing the analysis of data and development of new technologies – particularly those leveraging machine learning, or more generally artificial intelligence (AI) – to improve diagnosis. For example, the Viz.ai Large Vessel Occlusion Stroke Platform received FDA approval in 2018 for its use of AI to detect and diagnose strokes using brain imaging.[4,5] Similarly, IDx-DR offers an AI-based software program that analyzes images of patients' eyes to diagnose diabetic retinopathy at the point-of-care.[6] Other algorithms rely on continuous collection of glucose levels to aid the diagnostic process, such as the Guardian Connect system for diabetes and an array of academic- and private sector-developed sepsis tools.[7,8] Large delivery systems, such as the Mayo Clinic and HCA Healthcare have partnered with technology companies such as Google to modernize how they store and control access to health care data in order to enable a range of technology solutions -- including AI-enabled digital diagnostics.[9]

To better understand current barriers in the application of data and technology to improve the diagnostic process and to develop strategies to speed progress in this area, we convened workgroups in March through July of 2021. Sponsored by the Gordon and Betty Moore Foundation, workgroups included experts in data science, diagnosis, and artificial intelligence from multiple sectors of the healthcare industry. We assigned each workgroup a topic area framed around the question of how to speed the application of data-driven technologies such as AI to improve the diagnostic process. (We used AI as shorthand to refer to the broad scope of data-driven technologies, ranging from rule-based decision-making aids to more complex machine learning algorithms.) The workgroup topics were as follows:

1. **The Diagnostic Process:** How is the diagnostic process currently understood and where are the shortcomings in current formulations that may be impeding development of AI tools to support it?
2. **Data Assets to Support Development of AI:** How can we create data assets that attract academics, industry, and others to work on diagnostic challenges?
3. **Incentives and Alignment to Speed AI Solutions that Meet Frontline Clinician Needs:** How can we establish strong alignment between frontline diagnostic needs/problems and technology-based (and in particular, artificial intelligence-based) solutions?

This white paper begins by summarizing the framework that emerged from the Diagnostic Process workgroup, which was recently published in JAMA.[10] The framework advances current understanding of the process by identifying the key activities that comprise *the clinician's dynamic, longitudinal diagnostic process,* with a focus on information-related activities. This framework illustrates how many current AI tools fall short by focusing on the final diagnostic label and not on the clinician's upstream process. We use the framework to suggest opportunities to develop tools that better support these upstream activities.

The Dynamic Diagnostic Refinement Process

The diagnostic process is a foundational component of clinical medicine and one that has remained remarkably consistent over time. Perhaps the best-known representation of this

process, published by the National Academy of Medicine in 2015, depicts a trajectory beginning with a patient experiencing a health problem and ending with the outcomes from treatment (**Figure 1**). The core of the process is an iterative cycle in which a clinician gathers information and integrates it with their knowledge and experience; that triggers potential diagnoses to explain the patient's presenting symptoms, and determines *possible* courses of action and the *best* course of action (e.g., gathering more information by ordering further tests to refine the working diagnosis; determining potential treatments).

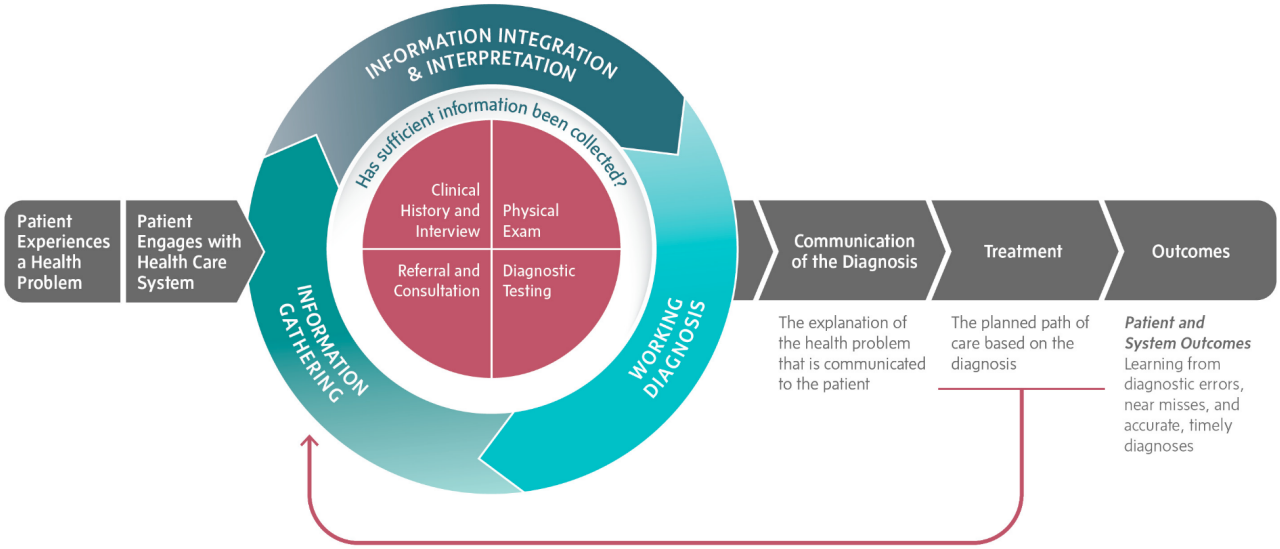

Figure 1. National Academies of Sciences, Engineering, and Medicine conceptualization of the diagnostic process. Taken from *Improving Diagnosis in Health Care*, 2015.

While this representation effectively summarizes the process, it does not capture the experience of the individuals involved in it - with the patient and the clinician as the two central players. From the clinician's perspective, the diagnostic process is more akin to a ***wayfinding experience***. As described in the JAMA piece, wayfinding is the process or activity of ascertaining one's position and planning and following a route. Extending this metaphor, the experience is one in which the patient presents with a set of symptoms, thereby prompting the clinician to engage in Information Gathering, Organization, and Prioritization based on what is known at that moment in time ("where are we now?"). Via Information integration and Interpretation, the clinician is able to Formulate Next Steps, often with a general direction or multiple potential

destinations in mind. As those next steps are followed, they generate new information, kicking off another round of the cycle. With each iteration, the clinician reduces uncertainty, and the destination becomes clearer. Arrival at the destination signals the moment that the clinician determines the diagnosis (Clinician-confirmed Leading Diagnosis) and documents it (usually via a diagnosis code that represents the Final Diagnostic Label). The final set of activities is to begin treatment planning and execution (Development of Care Plan). **Figure 2** below depicts this process.[10]

During the wayfinding process, the path can be short, clear, and straightforward(i.e., one cycle), or stretched out over a long time horizon with many cycles. There are also times when the clinician thinks they have arrived at the destination, only to learn that it was not the correct one, prompting a return to wayfinding. This process is dynamic, changing as new information becomes available or as the patient's condition and symptoms evolve. With new information, the plan for the path forward becomes increasingly refined, targeted, and direct. Understanding diagnosis as a dynamic refinement process shifts the emphasis away from "the diagnosis" - the ultimate endpoint of the wayfinding - and towards the substantive, information-intensive activities involved in the process used to get there.

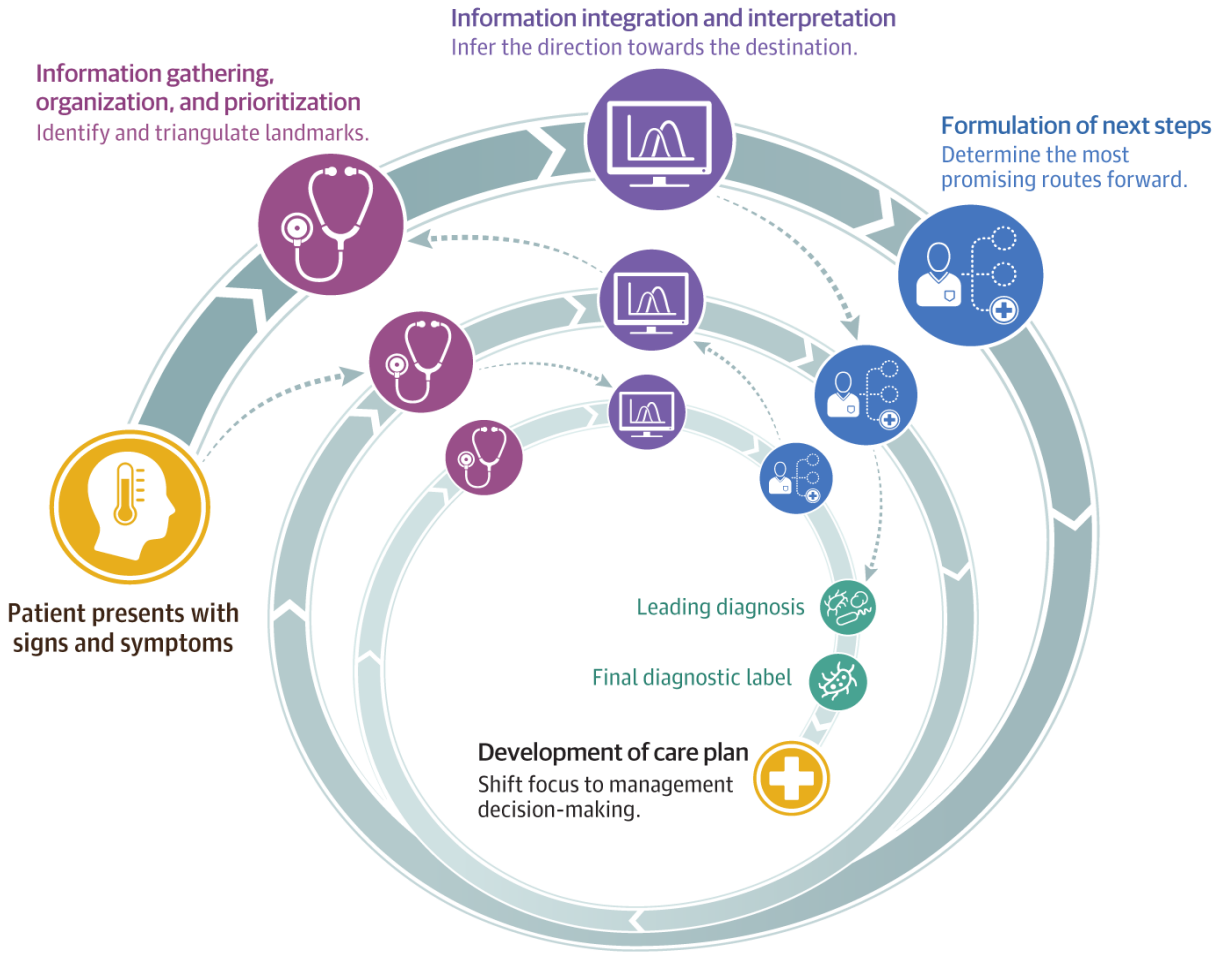

Figure 2. The Dynamic Diagnostic Refinement Process. The solid arrows illustrate the diagnostic process, and the dashed arrows illustrate how new information can alter the process. New information may increase uncertainty, causing a return to an earlier point and consideration of a broader set of possible next steps, or it may also enable jumping ahead in the process, avoiding an additional cycle. The diagnostic label is final in that it meets the administrative requirement for entering a code for billing purposes, but ongoing revision of the diagnosis, which drives treatment, is possible.

Overemphasis on the Diagnostic Label Results in Missed Opportunities

With the dramatic increase in the availability of digitized data - from electronic health records, patient devices, and other sources - there is unprecedented opportunity to understand the

dynamic diagnostic refinement process and to develop tools to support it. However, existing efforts to advance AI to support diagnosis have centered on predicting the *diagnostic label* which is the *endpoint* of the diagnostic process that is often administratively captured by the clinician selecting from a list of defined ICD codes. AI tools are largely built to answer the binary question of whether a given patient has a certain diagnosis (relying on the diagnostic label). While the appeal of focusing on this simple representation is evident, especially when developing AI tools that need a clear outcome to predict, tools that are designed to support the diagnostic process but ignore the upstream wayfinding process are of limited clinical utility. Said another way, any tool that predicts your destination at the start of your journey isn't very helpful if it tells you nothing about how to get there. Yet there are few tools that are designed to answer the wayfinding questions - where are we now and what should be done next?

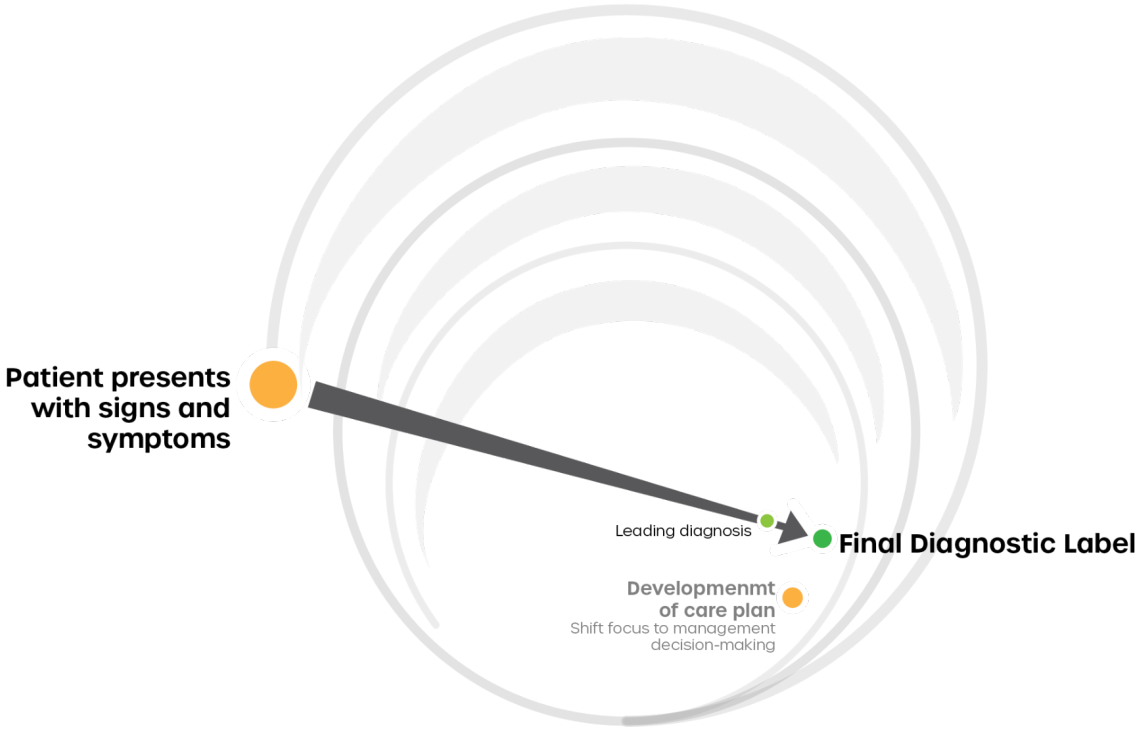

Figure 3. Current model of AI for Diagnosis. AI-based tools typically predict the final diagnostic label using data about the patient and provided care prior to the stay and during the stay, such as history from prior visits, vital signs, lab test results, and more.

What might AI tools that support clinicians in the dynamic diagnostic refinement process actually do? Effective wayfinding helps the clinician understand where they are in the process at a given point in time and then identify and prioritize the potential paths that lie ahead. To support the former, AI tools can search the EHR to identify and organize relevant information (e.g., signs and symptoms from the patient's history that may inform an understanding of the patient's current state). To support the latter, AI tools can identify the diagnostic tests or other assessments that are most likely to reduce uncertainty and move towards the final determination of the diagnosis. One specific field of AI and machine learning that may be well suited to this framework is *reinforcement learning*, where decisions can be suggested based on the current state of the patient and the range of possible next steps. While reinforcement learning applications are largely nascent because clinical decision spaces (i.e., what to do next given what's known) are large and complex, this also represents opportunity for future development, particularly within specific diagnostic areas.

More broadly, this shift could help produce AI tools that clinicians more readily embrace. There has been understandable hesitancy to relinquish decision-making to AI. However, AI tools that help clinicians with wayfinding still leave clinicians to analyze data and make decisions. Specifically, wayfinding AI is explicitly designed to augment human decision-making and focuses on supporting discrete tasks involved in the decision-making process. As a result, clinicians remain actively involved in the process, can more easily verify the accuracy and value of a prediction (which could lower perceived and actual risk of use), and cumulatively should experience a reduction in the associated cognitive burden of decision-making. In addition, current AI tools are frequently designed to find "zebras" - rare cases that may be easier to miss as they are less expected. Wayfinding AI tools that support the diagnostic process will aid clinicians in not only finding "zebras" but also finding "horses" - the more common cases - more efficiently and with less cognitive burden. Associated benefits may feel more tangible because they are experienced more often, with such tools ideally offering useful guidance for every diagnostic encounter.

Case Study

Current State Dynamic Diagnostic Refinement Process for a Patient with a Chief Concern of Progressive Shortness of Breath

Over the course of a few days, a patient notices that they need to breathe more heavily than normal when doing routine tasks like going up stairs. The patient presents to the clinician in a primary care clinic with *progressive shortness of breath* as the chief concern for the visit. This initiates the clinician's dynamic diagnostic refinement process. The clinician's cognitive process starts by summoning their medical knowledge and past experience about various symptoms and conditions related to the chief concern and coordinating that with the information about the patient. The clinician triggers their diagnostic schema for dyspnea, which includes cardiac, pulmonary, and other (e.g., anemia) causes with special attention paid to high morbidity conditions in each category. The clinician gathers information to see which category of disease (heart, lung, anemia) is most likely. This usually includes a series of questions directed to the patient – beginning with open ended questions about their symptoms, the time course, and behaviors (e.g., smoking status). Specific phrases such as "I'm having chest pains…I can't breathe…" trigger the clinician to narrow down to more targeted, closed-ended questions that allow them to move down the schema algorithm and prioritize certain illness scripts (mental models of specific diseases) while simultaneously determining that other branches of the schema(s) are unlikely and should be excluded from further consideration. Additional information gathering further helps make progress in navigating the diagnostic schema via searching for information in the EHR, trying to identify and review any information that may be potentially relevant (e.g., cardiology clinic notes, prior echocardiogram results, or pulmonary function tests). Finally, if the patient has recorded relevant data via a paper log or device (e.g., blood pressure, heart rate, weight, oxygen saturation, temperature), the clinician reviews the patient data to identify relevant trends.

As these sources of information accumulate, the clinician engages in integrating and interpreting the information. Suspecting that the most relevant domain of the diagnostic schema is cardiac, the clinician then proceeds with a physical exam that includes taking the patient's blood pressure, heart rate, weight, oxygen saturation, temperature, and listening to the patient's lungs. The clinician further integrates and interprets this new information. For example, the

clinician synthesizes observations that the patient doesn't have a fever, has gained significant weight, has lower extremity edema, that there are crackles when the clinician listens to the patient's lungs, and that the patient's neck veins are dilated, and reaches the conclusion that the patient is suffering from volume overload. The clinician also considers less likely but still possible causes that have not been definitively ruled out, including pneumonia. The clinician then identifies the best next steps that help narrow uncertainty. For volume overload, the next steps might include a point of care ultrasound and a BNP blood test. To assess for pulmonary edema as well as rule out pneumonia, the clinician might consider a chest x-ray and white blood cell or procalcitonin blood test. The clinician therefore checks a bedside echocardiogram. Once the clinician confirms that the patient's ejection fraction is low and determines the findings are consistent with congestive heart failure, the clinician then moves into management reasoning – where the decisions revolve around appropriate treatment.

Process Step	**Opportunity for AI Support**	**Example**
Information Gathering	Suggest clinical interview questions related to chief concern based on *EHR chart review*	Automatic chart search for risk factors of other common conditions, perhaps with atypical presentations, that do not fit the diagnostic schema – such as pulmonary embolus and immunodeficiency. No further action prompted if no risk factors found.
Information Gathering	Suggest clinical interview questions related to chief concern based on *patient responses*	A patient responds "yes" to a question about weight loss. The AI system invites the clinician to ask about related topics relevant to the diagnosis that are not raised by the end of the interview, such as access to food and drink, infection (fevers/chills), hyperthyroidism (e.g., tremor, palpitations), and cancer (e.g., severe, unremitting, new pain).

Information Gathering	Organize EHR display around diagnostic schemas; within each schema, order information from most-to-least relevant	Color code all information related to heart schema "red"; within heart schema, list cardiology notes first.
Information Gathering Integration	Identify trends and patterns in patient provided data; suggest questions, assessments to explore potential causes	Identify patterns of weight gain over the past 5 days measured and transmitted by a smart scale. Correlate daily weights with lower oxygen saturation levels over the past two nights measured and transmitted by the patient's smart watch.
Information Gathering Integration	AI-enabled stethoscope (with 2 lead ECG) used during physical exam	Analyze rhythms to detect arrhythmia, heart failure and murmurs.
Information Gathering Integration	AI-enabled x-ray; point of care ultrasound	Analyze chest x-ray images to detect pleural effusions, cardiomegaly, and pulmonary edema.
Integration Best Next Steps	Generate list of best next steps within each diagnostic schema with rationale, based on all available information	Suggest bedside echo and BNP blood test as best next steps related to volume overload; suggest chest x-ray and CBC to rule out pneumonia.

Table 1. Opportunities for AI to Support the Dynamic Diagnostic Refinement Process

Data Assets to Support Development of AI-Assisted Dynamic Diagnosis Tools

Existing health and healthcare data assets are rich but also fragmented and varied along many dimensions, including 1) where the data originate, 2) whether and how they are shared and

linked, 3) data storage and access models, and 4) the content and structure of data. For example, data assets might consist of national registry data, EHR data, wearables data collected by patients and technology device companies, environmental/exposure data, claims data, and genomics data. EHR data is most often used to support assessments of diagnostic performance and development of diagnostic AI tools. However, EHR data is usually limited to a single health system and is rarely integrated with the other types of data that may be relevant to diagnosis - spanning the environment to the gene.

A major barrier to developing AI tools that support the dynamic diagnostic refinement process is comprehensive data about that process. There are two broad categories of data that are needed to capture the diagnostic process - patient-centric data and clinician-centric data. The former reflects what information is known about the patient while the latter reflects both clinicians' actions as they wayfind and the contextual factors surrounding the clinician and patient during the diagnostic process, including how the clinician interacted with, and made decisions based on, that information. (See examples in **Table 2**).

Common types of patient-centric data	**Potential types clinician-centric data**
• Longitudinal health record • Patient-generated health data • Exposure data • Genomics data	• Chart review & documentation actions • Prior knowledge of the patient • Patient load • Care team composition, experience

Table 2. Data to capture the diagnostic process.

Both categories of data need to be captured over varied time scales to reflect the fact that some diagnostic processes occur in minutes and others in months (or even years). For example, in an intensive care setting, characterizing sepsis requires data captured over a time scale of hours to days, sampled on the order of minutes to hours, and is typically captured comprehensively for a given patient with a single EHR. In contrast, detection of atrial fibrillation requires monitoring for weeks or even months, with only a few variables being captured (though capture of higher resolution data at a few hundred samples a second may be required to confirm the diagnosis).

However, these data would also usually be available for a given patient within a single EHR, though wearables data for atrial fibrillation detection (e.g., Apple watch) may not be in the EHR, nor the proprietary products that are used (e.g., Zio patch). The journey to a cancer diagnosis often entails a longer time scale that involves assessment from multiple providers in different health systems resulting in relevant information spread across disparate EHRs. These examples illustrate how building comprehensive data assets that include data on the appropriate time scale is complex.

While efforts are needed to advance the availability of both patient-centric and clinician-centric data - with a particular focus on expanding the size and representativeness of the data, and on linking across complementary sources - current efforts are more focused around patient-centric data. Where clinician-centric data exist, they are usually a byproduct of patient-level data (i.e., with indicators of which clinician(s) were involved in the care of a given patient, such as who entered a given order). Addressing this gap requires envisioning and then creating datasets that integrate the patient's trajectory with the clinician's actions (and broader clinician context). Such datasets would effectively reflect a timeline on which the patient's clinical status is interwoven with the clinical actions in response to it. On this foundation, it becomes possible to develop AI tools that support the dynamic diagnostic refinement process, as the points of opportunity for diagnostic guidance stem from changing the clinician's actions.

As a first step towards envisioning the types of *data* and associated *labels* that characterize the clinician actions, we return to the framework and the three core activities within it. A data asset that captures Information Gathering, Organization, and Prioritization will require EHR data describing the types of information available and labels capturing what subset of information was retrieved by the clinician (as a proxy for what the clinician considered important). More nuanced data could further capture the order in which information was retrieved, how long it was viewed, and what the next sought EHR data element was that was viewed. A data asset that captures Information Integration and Interpretation might require free-text clinical notes that capture the diagnostic reasoning or structured fields with preliminary diagnoses. Finally, a data asset that captures Formulation of Next Steps would include the diagnostic tests, referrals, or other assessments that the clinician initiates. (Table 3) At a high-level, building data assets that include these types of data is feasible as EHRs capture detailed metadata about clinician actions. Thus, there is immediate promise to expand existing data assets to include these types of data.

Information Gathering, Organization, and Prioritization	Information Integration and Interpretation	Formulation of Next Steps
• EHR data describing the types of information available for the clinician to view/access • Labels capturing the subset of information that was viewed/accessed by the clinician, in what order, etc.	• EHR data including free-text clinical notes, preliminary diagnoses, etc. • Labels capturing cognitive constructs, such as uncertainty, hedging, etc.	• EHR data on actions taken by clinician, such as diagnostic tests, referrals, assessments • Labels capturing subset of actions relevant to the diagnostic process

Table 3. Types of data and associated labels that characterize clinician actions

As with any new data type, there will need to be companion work to assess the completeness, reliability, robustness, and caveats to measures of the diagnostic process. It may be necessary to begin by focusing such an expansion on a specific diagnostic domain or clinical setting to curate the data, illustrate the value it brings, and characterize the limitations. It is important to ensure that such a focus does not increase the risk of data bias - a core concern in the development of all data assets, particularly in the context of data-driven algorithm development. Currently, algorithms and the data assets used to develop those algorithms have limited documentation on data provenance, including the populations they were built on. As a result, it is unclear what potential data bias and algorithmic bias may result. In addition, many available data assets and industry-healthcare system partnerships occur at large well-resourced academic medical centers (e.g., MIMIC, NIH's National COVID Cohort Collaborative (N3C)) that are able to invest in technology, research, and staff to build and maintain data pipelines.[11,12] Therefore, many available data assets capture only the segments of the population served by those health centers. For diagnosis, this means a narrow consideration of what types of data may be available

when patients are seen in ambulatory and inpatient settings and what types of next steps are available. Additionally, most of the existing diagnostic AI algorithms reflect the availability of data (i.e., where accessible, curated datasets exist) and not necessarily the clinical utility of that algorithm. Thus, efforts are needed to not only encourage the expansion of data assets but to do so equitably.[13] While we have well-developed notions of how to assess potential bias in terms of patient demographics, it is not well understood how to assess bias in terms of organization- and clinician-centric data assets. The type of clinical setting and the type of EHR used offer reasonable places to start but more nuanced assessment will undoubtedly be necessary.

Strengthening Incentives for AI-Assisted Dynamic Diagnosis Tools

While data assets that enable development of AI tools supporting the dynamic diagnostic refinement process must be made available, availability alone is not sufficient to ensure tools are 1) developed, 2) implemented, and 3) used. There are two key domains of misalignment impeding the needed incentives. First, there is a lack of strong incentives, and arguably even disincentives, for specifically targeting improving diagnostic performance. As a result, developers (especially in industry but also in academic settings) deprioritize diagnostic AI. Even new value-based payment models have not explicitly focused on improved accuracy or timeliness of diagnosis as an outcome. Without clear evidence that poor diagnostic performance is a major driver of spending (at least when compared to known targets such as reducing inpatient utilization), it is difficult to justify focusing improvement efforts under value-based payment on diagnosis. In addition, liability protections and policies around diagnostic AI tools that support clinical reasoning are not well-established, and developers and providers may therefore prefer to invest resources in tools with less uncertainty. The FDA will likely regulate and require pre-market review of AI software that influences clinical diagnosis. This review process may further deter developers. The combination of weak financial incentives, potential for liability risk, and the need for FDA approval are key policy targets to address to spur investment in diagnostic AI tools that support clinician decision-making.

The second area of misalignment is between those developing AI tools and those using them - specifically the need for the developers to understand clinician needs as they progress through the dynamic diagnostic refinement process. In order for AI solutions to be useful, they must be

accurate, timely, and *provide relevant information*. Returning to our framework, each component offers the opportunity to define what it means to meet these criteria. For example, when building a tool to support Information Gathering, Organization, and Prioritization, a key performance criterion is whether the tool identifies the relevant data and excludes the irrelevant information. A second criteria might be when to show a curated list of information to the clinician - is information most timely at the very start of when a chart is opened, or should it come after the clinician has sought preliminary information? These nuanced decisions have implications not only for the AI itself but for its integration into cognitive and clinical workflows within the EHR and illustrate how critical it is to be able to understand the nature of the diagnostic process and how new tools could help improve it.

These challenges are often eased for homegrown AI technologies as developers are within the institution. This proximity can facilitate more direct knowledge of frontline providers' diagnostic needs, easier access to and understanding of the data required to build and test tools, and reduced need for generalizability and scalability. Yet even with these advantages, it is challenging to develop solutions that are widely used given the friction at each step in the tool development process. Commercial AI developers face even more friction as outsiders who lack detailed understanding of the process, often resulting in solutions that do not meet frontline needs.

To help address this gap - and to reveal the most compelling opportunities to develop useful tools - it makes sense to use expanded data assets to characterize current dynamic diagnostic refinement processes for different common presenting symptoms. A data-driven representation would offer an initial "look" into the typical activities undertaken by clinicians, as well as where there is the most variability. This could spur specific discussions with frontline clinicians to identify where variability reflects suboptimal processes and how AI tools could be designed to improve them. A specific focus on developing wayfinding AI tools that tackle discrete tasks to support diagnostic decision-making could further reduce friction as such tools are narrower in scope, more clinically acceptable, and pose less risk (and therefore face less likelihood or need for regulation).

Conclusion

Data-driven technologies such as AI hold promise for helping to improve diagnostic performance. Yet few tools exist to support clinicians as they undertake the dynamic diagnostic refinement process. To address this gap, action must occur at multiple levels - spanning broad incentives to improve diagnostic outcomes to greater availability of data assets characterizing the process itself. A key first step is orienting around a shared understanding of the dynamic diagnostic refinement process and then using that shared understanding to identify 1) the types of activities that need to be supported, 2) the types of data assets that need to be developed, and 3) the approach to ensuring that resulting tools will be adopted and used. Our hope is that the framework we present, and the complementary actions we describe, lay the foundation for this new orientation and can be taken forward by varied stakeholders working to advance AI tools that improve health system performance.

Acknowledgements

This work was supported by a grant to UCSF from the Gordon and Betty Moore Foundation (#10029), with guidance from Daniel Yang and Tommy Wang. The ideas presented stem from the discussions of the three workgroups, with a list of the leads and participants found below.

Bibliography

1. Institute of Medicine, National Academies of Sciences E and Medicine. *Improving Diagnosis in Health Care*. The National Academies Press; 2015. doi:10.17226/21794

2. Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual Amp Saf*. 2014;23(9):727. doi:10.1136/bmjqs-2013-002627

3. NORC at the University of Chicago and IHI/NPSF Lucian Leape Institute. *Americans' Experiences with Medical Errors and Views on Patient Safety*. Institute for Healthcare Improvement and MORC at the University of Chicago; 2017.

4. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *Npj Digit Med*. 2020;3(1):118. doi:10.1038/s41746-020-00324-0

5. Golan D. AI-powered stroke triage system performance in the wild. *J Exp Stroke Transl Med*. 2020;12(3):03. doi:10.37532/1939-067X.12.3.03

6. IDx-DR. Digital Diagnostics. Accessed September 23, 2021. https://www.digitaldiagnostics.com/products/eye-disease/idx-dr/

7. Summary of Safety and Effectiveness Data: Continuous Glucose Monitoring (CGM) System. Published online March 8, 2018. https://www.accessdata.fda.gov/cdrh_docs/pdf16/P160007b.pdf

8. Wu M, Du X, Gu R, Wei J. Artificial Intelligence for Clinical Decision Support in Sepsis. *Front Med*. 2021;8:665464. doi:10.3389/fmed.2021.665464

9. Evans M. Google Strikes Deal With Hospital Chain to Develop Healthcare Algorithms - WSJ. Published May 26, 2021. Accessed July 23, 2021. https://www.wsj.com/articles/google-strikes-deal-with-hospital-chain-to-develop-healthcare-algorithms-11622030401

10. Adler-Milstein J, Chen JH, Dhaliwal G. Next-Generation Artificial Intelligence for Diagnosis: From Predicting Diagnostic Labels to "Wayfinding". *JAMA*. 2021;326(24):2467-2468. doi:10.1001/jama.2021.22396

11. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3(1):160035. doi:10.1038/sdata.2016.35

12. NIH launches analytics platform to harness nationwide COVID-19 patient. National Center for Advancing Translational Sciences. Published June 15, 2020. Accessed July 26, 2021. https://ncats.nih.gov/news/releases/2020/NIH-launches-analytics-platform-to-harness-nationwide-COVID-19-patient-data-to-speed-treatments

13. Larson EB. Learning Health System initiative weaves research into care delivery at KP Washington. Medium. Published July 3, 2017. Accessed July 26, 2021. https://medium.com/@KPWaResearch/learning-health-system-initiative-weaves-research-into-kaiser-permanente-washington-health-care-9afe1d4ed009

Workgroup Leads

Jonathan H. Chen, MD, PhD
Stanford Center for Biomedical Informatics Research + Division of Hospital Medicine, Stanford University

Gari Clifford, PhD
Department of Biomedical Informatics, Emory University
Department of Biomedical Engineering, Georgia Institute of Technology

Christina Silcox, PhD
Duke-Margolis Center for Health Policy, Duke University

Workgroup Participants

Lesley Curtis, PhD
Department of Population Health Sciences, Duke School of Medicine

Josh Denny, MD, MS
All of Us Research Program, National Institutes of Health

Gurpreet Dhaliwal, MD
Department of Medicine, UCSF
San Francisco VA Medical Center

Noémie Elhadad, PhD
Department of Biomedical Informatics, Columbia University

Krzysztof Geras, PhD
School of Medicine, New York University

John Halamka, MD

Mayo Clinic Platform, Mayo Clinic

Alistair Johnson, PhD
Knowledge in Data Lab, The Hospital for Sick Children

Mo Kaushal, MD, MBA
General Atlantic
Clinical Excellence Research Center, Stanford University

Steven Lin, MD
Division of Primary Care and Population Health, Stanford University School of Medicine

Ziad Obermeyer, MD, MPH
School of Public Health, UC Berkeley

Pilar Ossorio, PhD, JD
Department of Medical History and Bioethics, School of Medicine, University of Wisconsin; School of Law, University of Wisconsin

Marc Overhage, MD, PhD,
Anthem, Inc.

Grace Peng, PhD
National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health

Mark Sendak, MD, MPP
Duke Institute for Health Innovation

Paul Tang, MD, MS
Clinical Excellence Research Center, Stanford University
Palo Alto Medical Foundation

UCSF Team

Julia Adler-Milstein, PhD
Center for Clinical Informatics and Improvement Research, UCSF

David Coleman, BA
Center for Clinical Informatics and Improvement Research, UCSF

Jen Gong, PhD
Center for Clinical Informatics and Improvement Research, UCSF

Sarah Rosenthal, BA
Center for Clinical Informatics and Improvement Research, UCSF

Crissy Thao, MPH
Center for Clinical Informatics and Improvement Research, UCSF

Contact information for UCSF team:
Julia Adler-Milstein, PhD
10 Koret Way, #301B
San Francisco, CA 94131
415-476-9562
Julia.adler-milstein@ucsf.edu