Toward a Learning Ecosystem for Diagnostic Excellence

Katherine Satterfield
Joshua C. Rubin
Charles P. Friedman
University of Michigan

A White Paper Prepared for the Gordon and Betty Moore Foundation
March 2018

I. Introduction

Diagnostic excellence is a Grand Challenge Problem requiring a comprehensive and coordinated approach.[1–5] This White Paper will make will make an initial case that application of methods associated with Learning Health Systems (LHSs)[6,7] can improve diagnosis through creation and support of a learning ecosystem. The vision advanced in this White Paper will be refined through deliberations at an invitational conference to take place in the spring of 2018.

The envisioned ecosystem will aim to ensure that:

1. Efforts to improve diagnosis address this grand challenge from multiple complementary perspectives.
2. These efforts foster improvement through capabilities that both create new knowledge and directly apply that knowledge to care.
3. Improvement is a continuous, ongoing process.
4. Mechanisms are in place to ensure that these efforts, and the communities participating in them, learn from one another.
5. Economy of scale and scope is achieved through shared use of common social and technical infrastructure.

Achievement of these aims in unison can bring about transformational improvements in diagnostic processes and health outcomes.

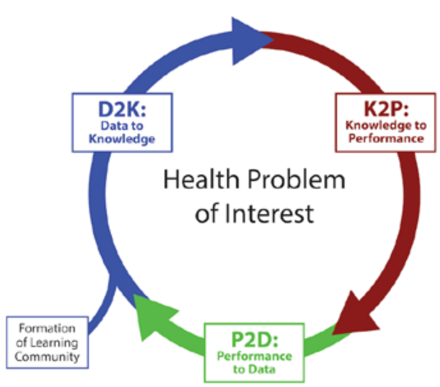
To make the case for a learning ecosystem, this White Paper will begin by introducing the basic concepts of LHSs. Following this section, we will report the results of semi-structured interviews conducted with 32 select members of three important groups: researchers currently focused on diagnostic excellence, researchers in machine learning and artificial intelligence applied to health, and researchers focused on Learning Health Systems themselves. Key concepts and insights extracted from the interviews will inform this White Paper's concluding section, which will describe the general features of a learning ecosystem for diagnostic excellence that will frame deliberations at the invitational conference.

II. Rudiments of Learning Health Systems (LHSs)

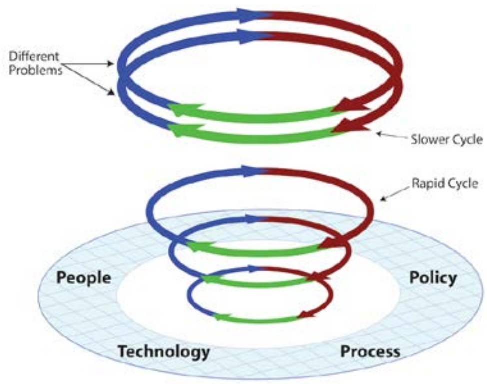
An LHS approach to health improvement rests on three key precepts:

1. *Learn from every patient:* Each patient's health experience can be captured as data and analyzed for purposes of improving individual and population health.
2. *A system problem needs a system solution:* Interventions aimed at individual people will not succeed nearly as well as interventions addressing organizational and cultural contexts that shape how people work together.
3. *Shorten by orders of magnitude the latency between creation of new knowledge and its application to care:* Analyses of data that generate new insights must, in addition to publication in journals, be directly applied to efforts to improve health behavior and care practice.

LHSs rest on the concept of complete "virtuous" cycles of discovery and change, leading to improvement around each cycle's specific target problem. Over time, improvement against the target problem develops through successive iterations of each cycle. As illustrated below, complete learning cycles consist of three phases: data to knowledge (D2K), knowledge to performance (K2P), and performance back to data (P2D). An initiative that applies LHS methods will ensure that new knowledge to improve diagnoses finds its way into practice, instead of accumulating in the literature where it customarily sits for many years before implementation.[6–9] It also formally identifies knowledge learned through practice that can fill gaps in clinical literature.



In an LHS, each learning cycle is an exercise in co-production.[6,10,11] Each cycle is executed and governed by a multi-stakeholder community sharing interest in the specific problem on which the cycle is focused. Representation in each community of all relevant stakeholders—Including patients, a range of health professionals, and care managers—is critical to the learning and improvement process. Broad representation ensures that relevant data are collected in the D2K component of the cycle; that analytic results are interpreted from a complete range of viewpoints as D2K transitions to K2P; and that interventions through the K2P component are broadly supported in the settings where they are implemented. Above all, community governance and execution of learning cycles builds essential trust among all stakeholders and helps ensure that improvement processes will be sustained.



In a functioning LHS, multiple learning cycles will operate simultaneously, each addressing its own problem of interest and each executed by its own multi-stakeholder community. Beyond these learning cycles, what makes an LHS a "system" is the existence of shared infrastructure to achieve economies of scope of and scale necessary for sustainability. All learning cycles follow the same D2K-K2P-P2D flow. For this reason, as illustrated above, common infrastructural services, forming a platform, support all learning cycles. As shown in the figure, the platform is sociotechnical. It comprises people, policy, process, and technology. In addition, LHSs require governance and coordination to ensure that learning cycles complement one another and that the participating communities can learn from one another.

III. Viewpoints from Three Key Groups

To appreciate a vision for how the LHS approach can promote diagnostic excellence in medicine, we interviewed 32 experts focused on diagnostic excellence (n=18), machine learning (ML) and artificial intelligence (AI) in healthcare (n=6) and Learning Health Systems (n=8). Each conversation was approximately 30 minutes in duration and gathered insights into progress made and pressing questions, barriers to diagnostic excellence, examples of translating new knowledge into practice, and conceptions of an appropriate use of learning health systems and machine learning applied toward more timely and accurate diagnoses. Interviews were semi-structured and therefore adaptable to the directions that experts took the conversation. Interviews for each group were guided by a set of questions attached as Appendix A. The names of the interviewees are included in Appendix B.

The work of each group is guided by differing sets of underpinning sciences and associated methods. In general, the diagnostic excellence group utilizes methods associated with clinical research largely as informed by behavioral and cognitive sciences. The ML/AI group is firmly grounded in computer science. The LHS group rests on a mélange of biomedical informatics and system, policy, and implementation sciences.

Diagnostic Excellence Group

General Perspectives of this Group: The researchers we interviewed are proud of the community they have established over the last decade and of subsequent achievements in putting diagnostic error on the radar as a significant patient safety concern. They are riding the energy of the National Academy of Medicine's 2015 report *Improving Diagnosis in Healthcare*[12] which is widely considered a breakthrough for highlighting concerns about diagnostic error in the broader medical community. With a few exceptions, there is general agreement that diagnostic error must be addressed separately from quality and safety;

otherwise, diagnostic concerns would continue to be overshadowed by patient safety problems that are more easily identified and corrected. This view is accompanied by a general sense that their community has not fully defined the problem of diagnostic error. Within this forward momentum, experts described different priorities for future efforts: whether the community focus should approach diagnostic needs for specific diseases, address changing procedures for the care team and the importance of changing care team culture, or emphasize enhancement of prompts and information available at the point of care.

Barriers to Diagnostic Excellence: Interviewees identified several barriers to diagnostic excellence. One recurring theme was the need to standardize and tighten the feedback loop of patient outcomes to clinicians who provided their care. As one interviewee phrased it, "If an individual doctor makes a diagnosis and doesn't ever hear that the patient turned out to have a different diagnosis and got re-admitted, or it turns out their diagnosis was wrong, then [the doctor] will never get better." Experts commented on the inherent difficulty in delivering this feedback for diagnoses that occur through the care of another clinician, or even more so, in another health system. Because of this major barrier to creating feedback, several respondents indicated they would like to seek ways to better engage the patient who has the power to "tell us if we got something right." Overall, these sentiments underscore the need for clinicians to know the distal results of their care decisions.

This need for feedback is predicated on other oft-cited barriers to diagnostic excellence. Interviewees repeatedly called for measures to both define and track error. At the foundation of this request is a shared view that the community has not been able to agree on comprehensive definitions of diagnostic quality or error. One interviewee recalled research that found existing reporting systems often identify cases that might be called diagnostic error but are inconsistently coded in other ways. This lack of consensus both undermines the ability to measure diagnostic error and compromises efforts to convince organizational leadership to take action that would support the creation and implementation of improved measures. This was echoed with calls for mechanisms that seek and create public accountability in the diagnostic process through transparent measures, analogous to efforts in the patient safety movement.

Furthermore, three interviewees directly connected the need for measures to a missing "business case" among leadership in healthcare organizations (HCOs). Without leadership support, the data needed to compute measures or create feedback to clinicians will remain inaccessible, and efforts to make improvements will fail. Other interviewees called for HCOs to make a broader commitment to safety-promoting values and for investment in both known and innovative processes that further diagnostic safety. One clinician-researcher, describing what should happen after a recommendation is made to HCOs, emphasized that following those recommendations "requires work" to bring different departments together and assign responsibilities for maintaining guidelines. He added, "I think overall safety requires more resources and we just don't have the resources to do safety work." These interviews did not reveal deeper insights into how to gain commitment from HCOs, though barriers were largely presumed to be fiscal. However, one participant hypothesized that difficulty in funding research in diagnostic error may also be impeded by clinicians involved in funding review, as they may feel criticized by the notion that they and their peers are not diagnosing patients adequately.

In that regard, many interviewees referenced various ways that clinicians may inadvertently impose barriers to making necessary improvements. At the core of these barriers is that physicians don't widely recognize that their diagnostic processes may be sub-optimal. A clinician spoke about approaching doctors to think about diagnostic improvement: "A lot of them have their hackles go up when you start talking about error. They all know about it because they all know about malpractice and the risk of malpractice, but they all think it's somebody else that's not as good as them, or not as careful, or not as well trained." The interviewees who spoke about this issue cited several drivers, going back to medical school training, where diagnostic processes—and identifying why a physician may decide on a particular diagnosis—is not explicitly taught. Instead, diagnostic processes exhibit more like "private mental events" without an explanation. The interviewees argued that—together with the conception that doctors can and should "memorize it all"—physicians in this medical culture miss opportunities when the patient might be better served if the physician would consult other doctors or diagnostic support systems.

Another clinician-driven barrier is sub-optimal physical exams and history-taking to support thorough differential diagnoses. Some of this concern could be attributed to design of electronic health records and subsequent data completion. The physicians we interviewed seek more comprehensive, longitudinal health data and better ways to interact with relevant aspects of health history that have already been recorded. However, they also stressed the need to prompt clinicians to be more comprehensive in their diagnostic workups. One physician claimed that 80% of the problem is about bed-side diagnostic operations, including failures in history-taking, examining patients, choosing tests, interpreting test and exam results, and integrating them into a differential diagnosis. Of course, clinicians can't make these improvements on their own. One clinician-researcher reiterated a message from Dr. Berwick's keynote at the 2017 Diagnostic Error in Medicine conference: "Doctors may cause errors, but we have to look to systems to solve them."

Collaboration Potential: When prompted to speak about learning health systems, interviewees in the diagnostic excellence group responded neutrally or positively. Some had stronger conceptual grasp of an LHS and comfortably integrated their actionable agendas for diagnostic excellence into the scope of an LHS. For example, one physician offered, "A learning health system should be able to do its own proactive risk assessment: identify gaps—where the deficiencies are—and what they need to do." More generally, there was endorsement of the concept of learning at the organizational level: that healthcare delivery organizations can and should study and learn from their prior efforts to improve the accuracy and timeliness of diagnoses.

Researchers focused on diagnostic excellence held stronger, and varied, opinions about the potential to apply "big data", ML, and AI to their field. Some expressed general skepticism about the value of these methods beyond existing achievements in imaging-related fields such as radiology, dermatology, and pathology. On the whole, while interviewees in the diagnostic excellence group acknowledged the potential for machine learning applications, they expressed apprehension in several areas. First and foremost was concern about the quality of available data from administrative and electronic medical record systems that would feed these algorithms, including concerns about extracting meaningful data from free text, and the effects this would have on the validity and utility of ML algorithms. Many interviewees asserted that any ML algorithms that might be applied to healthcare must be designed to be understandable by the clinicians making care decisions, and not opaque to them. This approach would not only lead physicians to informed opinions about how to use the results of ML algorithms, but could also bolster insight into which recorded health data is the most meaningful for defining a diagnosis or predicting outcomes. In a final caveat, one physician argued for multi-stakeholder development of criteria that artificial intelligence applications must meet in order to qualify for deployment in the healthcare setting.

Machine Learning and Artificial Intelligence Group

General Perspectives of This Group: The six computer scientists we interviewed have recently applied a variety of methodologies toward improving healthcare delivery and administration. With research interests grounded in a desire to answer complex questions through data, several of these scientists referenced how the nature of healthcare poses unique challenges that influence their approach. They criticized how some of their colleagues approach healthcare by seeking new datasets to test algorithms they have already developed, which results in methodologies that may not address problems that clinicians see as important. They advocated for new algorithms that focus on generating insight into what caused an outcome rather than simply increasing accuracy of prediction. One researcher pointedly observed, "I care more about actionability and lead time than positive predictive values." This is echoed in a set of guiding characteristics that another researcher identified through direct work with clinicians: that models must be actionable, or have utility in practice; that they are robust as clinical conditions and context change; and that models should be made more credible by leveraging what is already known from the literature.

Specific Approaches Employing ML and AI: Through the interviews, we identified four categories of approaches, methodologies, and objectives that can be integrated with diagnostic excellence initiatives:

1. Natural Language Processing algorithms have the potential to synthesize several types of free text into information that is either more digestible for a human (i.e. help a physician make sense of a complete electronic health record) or more computable so that data can be applied into another model. Another example of this is to apply NLP to published journal articles to both collect relevant information and assess the reliability of a study's methods—steps toward bringing clinical knowledge closer to the exam room.

2. Patient-Centered Data Mining takes advantage of information about health that resides in EHRs, patient portals, and other personal health applications. Interviewees

expressed the potential to mine this data to learn more about patients' health over time and how this additional resource can be used to inform a diagnosis.

3. Research on Predictive Modeling and Causal Inference is primarily dedicated to accounting for confounding variables to enable causal inferences in observational data and data from non-randomized experiments. Because of the amount of observational data now available, it seems imperative to use it to improve care. This viewpoint engages with a perception that models made from sufficiently large observational datasets can be more generalizable than RCTs in the literature, which use limited and often unrepresentative samples. With the right motivation, use of these methods could shift to better serve diagnosis, such as predicting progression toward a more serious diagnosis or using precision medicine concepts to better characterize a diagnosis for tailored patient care.

4. Control Theory, Q-Learning, and Back Propagation techniques can assist in refining predictive models and identifying pivot points or errors in a diagnostic, prognostic, or therapeutic pathway. In retrospective use, this approach might help a care team or health system learn from and improve diagnostic processes.

Collaboration Potential: A critical component of integrating ML/AI methods with healthcare is establishing meaningful and mutually beneficial collaborations among researchers, clinicians, and others involved in the care process. These experts have sought collaborations in multiple ways: bringing physicians in their labs as colleagues, training physicians as doctoral students in computer science, consulting with clinicians, and offering programming assistance that fit a health system's needs. Such approaches have helped computer scientists identify clinically relevant questions, conduct clinically accurate analyses, and obtain access to data. Access to healthcare data has been historically challenging for computer scientists. "Shareable data," one researcher said, "have been key to allowing researchers to really benchmark their data and risk-stratification algorithms, which has led to significant advancements." Even with access to data, there are still difficulties in circling their work back into clinical settings. A few researchers described licensing their algorithms to third-party companies as an avenue to put these products into meaningful practice, though one recounted being rebuffed by a major producer of digital medical products out of disinterest in gaining FDA approval.

Evidence emerged from these interviews that the ML/AI community and the diagnosis excellence community have not achieved significant levels of interaction. One researcher described the importance of creating a dialogue to counter the "myth that AI algorithms are black boxes". Of particular importance, the researchers we interviewed did not readily offer many examples of ML/AI work specific to diagnostic excellence, and two researchers very vocally emphasized treatment over diagnosis as the problem to be addressed. One explained: "Typically, you are providing support to a highly educated individual who does not want your help. … The other sort of larger reason is that's not where the majority of impact is going to be. We train these humans 10 years to diagnose. Yes, they make mistakes, and yes, medical misdiagnosis is common enough, but the heterogeneity…in treating is an even bigger problem." Another believed that "Interesting questions lie in the iterative improvement of the treatment plan, rather than in the diagnostic process." Recognizing that this sample is limited to six computer scientists, it seems clear that application of ML and AI methods toward diagnostic excellence requires increased dialogue among interested parties.

Learning Health System Group

General Perspectives of This Group: In interviews with eight experts in LHSs, we discussed how the LHS model has progressed and how it might support diagnostic excellence in collaboration with the other two communities. Results reflected the interdisciplinary background of LHS researchers and advocates, including former physicians, computer scientists, and individuals with expertise in health information technology, health policy, and population health. As such, their insights affirmed many themes expressed by the diagnosis and ML/AI communities, incorporating calls for feedback to clinicians, actionable algorithms, and shared data. Underlying these themes rested a necessity to form a sociotechnical system solution that links the efforts and interests of all stakeholders before and after diagnoses are made. As described by one expert, success is predicated on "ensuring that the technology we have is fully interoperable and…functionally accessible to clinicians and patients and families alike. And that the culture of the care process is one in which decisions are made in a team fashion rather than in a linear, traditional, clinician-to-patient unidirectional fashion."

Specific Approaches: During interviews, experts cited several examples of nascent learning health systems, but qualified that observation with reminders that they have yet to see mature development of an LHS with all requisite components. Notably, these examples were almost exclusively scaled to single health systems, and interviewees generally recognized that the U.S. has not yet developed the facility to network a larger-scale LHS. At the same time, interviewees expressed belief that implementation of small-scaled LHSs remain an important stepping-stone toward interoperability on a much larger scale. These interviewees hope to see a few innovative health systems invest in advancing a complete LHS concept. This advancement would concretely demonstrate feasibility, generate data on the persistence and effects of the organizational learning that occurs, and attract more adopters of an LHS model.

Interviewees agreed that technical issues were not the major challenge to LHS development. As one former computer scientist stated: "[emergence of a large scale LHS] was unlikely to be achievable using traditional software and systems engineering thinking and [his ideas were] more focused on establishing conditions under which the right sorts of infrastructures and dynamics would emerge." In other words, the development of larger scale systems is contingent on incentivizing cooperation between and within stakeholders in the health industry. This requires deeper understanding of who benefits and who pays for a new model so that HCOs begin to loosen the reins that are currently preventing collaboration and co-production. For example, American health systems currently retain clinical data in proprietary EHRs, and there is reluctance to create systems that promote secure access to a wider system of outsiders. Another LHS expert highlighted concerns that stakeholders must distinguish between traditional clinical research and iterative health system learning. These distinctions must be made for ethical reasons to define what external oversight is required. Negotiations of these and other details are requisite for achieving scalable, effective learning health systems.

Collaboration Potential: Interviewees pointed to the need for a dialogue that connects the LHS to both diagnostic excellence and ML/AI. One researcher described the potential collaboration of these three communities as the 'poster child' for how a learning health infrastructure must complement existing human-completed processes. In describing developing an LHS for diagnosis, one expert echoed the calls of diagnostic researchers to create interactive, interpretable systems. Based on decades of work on diagnostic support tools, he emphasized that providing explanations for "black box" algorithms is critical for forming trust in these methods. Another interviewee expressed the importance of work to render knowledge in a form where clinicians can review, edit, and disseminate it. Finally, several of the LHS researchers emphasized the necessity of bringing multiple stakeholders and scientists representing many disciplines together to address system-level changes and concerns. One expert specifically called for building teams with expertise in both basic research and organizational processes.

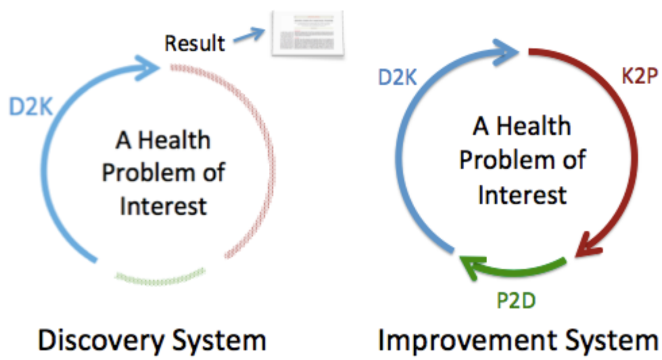
IV. Envisioning a Learning Ecosystem for Diagnostic Excellence

Key Insights from the Interviews

Several themes derived from the interviews both motivate an LHS approach for diagnostic excellence and suggest the shape such an initiative might take. Many of these themes reflect views that were shared across the three interview groups.

The Importance of Connecting the Loop: As described above, interviews of diagnostic excellence researchers revealed several key barriers to improving diagnostic processes and outcomes in practice, such as absent feedback loops of patient outcomes and sub-optimal diagnostic processes. These barriers correspond directly to components of a learning cycle that are currently absent in the research and implementation of health services. For example, research that informs diagnostic best practices has had difficulty achieving successful uptake into clinical practice (a weakness in knowledge-to-practice, or K2P). Additionally, performance-relevant information (patient outcomes post-diagnosis) is not converted back to data for clinical reflection (performance-to-data, or P2D).

Because these needs directly invoke the components of the learning cycle, an LHS approach would fundamentally link processes involved in knowledge generation and clinical translation, and these coordinated components structurally reduce barriers that halt the pursuit of diagnostic excellence. In a particularly salient example, an LHS enables discovery systems—that only publish new results in journals—to become improvement systems that give equal emphasis to directed efforts that apply what is learned (shown below).



To the extent that the diagnostic excellence group consists largely of clinicians, engaging them with the ML/AI group will address the deeply held and legitimate need for these computer scientists' work to be more immersed in clinical settings and influenced by the insights of those who directly care for patients.[13] More generally, those advocating for a greater focus and practiced methods to reduce diagnostic error will best be able to interpret that knowledge into medical practice as part of the "K2P" component of learning cycles.

Engaging Multiple Stakeholders: All three groups addressed the importance of engaging multiple health stakeholders in a comprehensive approach toward diagnostic excellence. They recognized that efforts to modify care providers' behavior and reasoning must be shaped and supported by the organizational environments in which these individuals work.[14] The interviewees also recognized that patients could play critical roles in the pursuit of diagnostic excellence, primarily through an invitation to take more active roles in monitoring and managing their health.[14–16] These sentiments directly invoke the multi-stakeholder feature of communities that govern and execute learning cycles in an LHS.

The multi-stakeholder aspect of learning systems and the potential it holds for building trust within communities could also address the shared concern regarding the abstract nature of predictive models and other algorithmic results from machine learning approaches.[17] Greater involvement of patients, clinicians, and managers in the creation and validation of these models would engender shared appreciation of their strengths and limitations, and the best ways to implement them.[18] Components of LHS infrastructure that curate, manage, and disseminate these models would help ensure that they can find their way into practice.

The Importance of Measures: Many interviewees stressed the need for standardized measures to both document the prevalence of preventable, sub-optimal diagnosis and to track improvements. In particular, publications from interviewees and a recent report by the National Quality Forum have highlighted the gap in measures that are sensitive to patient engagement, diagnostic processes, and organizational and policy opportunities.[18,19] This resonates with operation of LHSs and their dependence on data. As part of a learning cycle, the community governing that cycle must identify target measures to document progress in addressing the problem of common interest.[18,20] It follows that an LHS approach for diagnostic excellence will, through its natural mode of operation, both identify measures of success and build consensus around them. An LHS "ecosystem" connecting these cycles, as described below, will enable sharing of these measures and their dissemination to wider communities.

Cultural Change: Many interviewees invoked the concept of culture and the need for cultural change to support successful interventions to improve diagnostic timeliness and accuracy.[21] The LHS approach is one that challenges prevailing culture, as seen through the three LHS precepts cited earlier. In particular, the "learn from every patient" precept grounds the LHS in a safety culture that values all care experiences as opportunities to learn and improve. This precept also grounds the LHS in an empirical culture that values accurate data as the raw material that enables improvement, new knowledge as the pathway to change, and measured results as a demonstration that improvement has occurred.[22] The precept of "system problems need system solutions" places value on pre-competitive environments that develop and share resources to produce results that are greater than the sum of their parts.

Experience to Build On, Spanning the Learning Cycle: From the perspective of the LHS community, diagnostic excellence represents a key, "poster child" health challenge amenable to LHS methods. This interview group pointed out that, while a complete LHS tool kit does not yet exist, an LHS approach for diagnostic excellence can rest on the considerable and growing experience of this community.

Additionally, the interviewees as a group represent a broad range of clinical and scientific perspectives that map naturally to the sciences invoked by the three components of the learning cycle. The expertise of the ML/AI interviewees invokes the D2K component of the cycle, and their engagement with an LHS approach for diagnostic excellence brings novel methods to this process. The diagnostic excellence group spans D2K and K2P. The LHS group contributes perspectives on infrastructure, sustainability, and scale. While it is important that additional stakeholders join a comprehensive effort to improve diagnostic outcomes, the three groups participating in the interviews form a firm foundation, and the interviews suggest a complementarity of interests that can bond them in shared effort. The planned meeting in the Spring of 2018 will take important steps in this direction.

V. Characteristics of the Envisioned LHS Ecosystem

This section offers a preliminary portrayal of important features of a learning ecosystem for diagnostic excellence. The concepts and plans put forward here will be discussed and refined as part of the invitational meeting planned for March 2018.

As noted at the beginning of this White Paper, the envisioned ecosystem will aim to ensure that:

1. Efforts to improve diagnosis address this grand challenge from multiple complementary perspectives.
2. These efforts foster improvement through capabilities that both create new knowledge and directly apply that knowledge to care.
3. Improvement is a continuous, ongoing process.
4. Mechanisms are in place to ensure that these efforts, and the communities participating in them, learn from one another.

5. Economy of scale and scope is achieved through shared use of common social and technical infrastructure.

And to these ends, the ecosystem will have the following general characteristics:

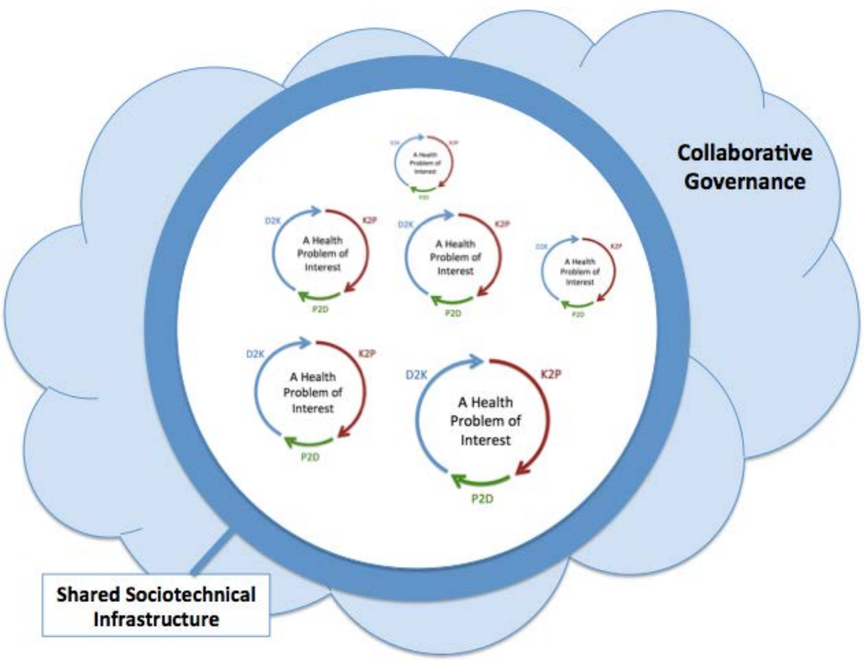
A complete LHS-inspired approach for diagnostic excellence will encompass a set of complete learning cycles, each focused on a specific diagnostic problem area and executed by a multi-stakeholder community. Complete cycles will include D2K, K2P, and P2D components—ensuring that discovery is translated directly into practice. Networking of these learning cycles and the communities that execute them—along with a governance and coordination function and shared infrastructure—creates what might best be called a learning ecosystem focused on the Grand Challenge of diagnostic excellence. The specific diagnostic problems that can be explored in the ecosystem are not pre-determined; they evolve from the interests, passions, and experiences of individuals who form communities to address them. These communities can self-organize; but their formation can also be stimulated through the governance of the ecosystem.

For example, a community might form to address the lack of metrics and methods to measure diagnostic quality. The learning cycle might begin with the "data to knowledge" (D2K) component focused on inventorying existing metrics and methods, and the data about their value and effectiveness. What is learned from this process will shape the "knowledge to practice" (K2P) component of the cycle. If the conclusion of the community is that many measures exist, but they are being underutilized, the K2P component of the cycle would focus of promotion of greater and more appropriate use of the existing tools. However, if the D2K component identifies clear need for entirely new methods and metrics, the K2P effort will stimulate community-driven creation of such tools. The "practice to data" (P2D) component of the cycle would document what new tools were developed, setting the stage for the next for the D2K component of the second cycle which would generate data from the testing of these tools.

At any given time, multiple learning cycles will be ongoing in this ecosystem. Each cycle will exist at a particular level of scale: some operating in single institutions, others in networks or consortia of institutions. The community identified with each cycle will be coterminous with the scale of the cycle: e.g. the community members for a network-based cycle will represent all institutional members of the network. Any individual who is working in the greater ecosystem could be part of one or multiple communities. Over time, learning and improvement will occur with respect to each diagnostic problem area, and in addition, the communities associated with the different cycles in the ecosystem will learn from each other. This cross-community learning will happen in two ways. First, some of the specific strategies to improve diagnostic outcomes, learned in one of the communities, will naturally inform other communities. Second, through a process that might be called "meta-learning", improved methods to execute the D2K, K2P, and P2D components of learning cycles will emerge and will diffuse across communities.

The co-existence of multiple learning cycles focused on diagnosis as a Grand Challenge is necessary, but not sufficient, to form a fully productive ecosystem. Two additional features are required:

1. Governance will ensure that the learning cycles, and the communities that execute them, function as a coherent, complementary and interactive set.
2. Infrastructure will provide sociotechnical services—spanning policy, process, and technology—that support all learning cycles.



The structural features of the envisioned ecosystem are illustrated in the figure below.

<u>Governance</u>: It will be very important that governance strike a balance between openness and control to ensure that innovation is nurtured while coherence is maintained. This will require Internet-like principles of invoking only a minimum set of standards that will enable all learning cycles that are part of the initiative to develop in compatible ways. As emphasized by many of the interviewees, broad stakeholder participation will be a key element of successful governance of the proposed ecosystem.

Governance of the ecosystem will "certify" cycles and their associated communities as members of the ecosystem, will establish and enforce policies to ensure that the cycles function in a coordinated manner, and will ensure that the diagnostic problems addressed by the cycles comprise a complementary "spanning set" relative to the Grand Challenge of diagnostic excellence.

<u>A Spanning Set of Problems</u>: The fundamental governance challenge for the ecosystem rests in the initial selection and evolution of the target diagnostic problems that are the foci of the learning cycles themselves. It is essential that, at any given time, the portfolio of problems addressed span a range of challenges, invoke a range of methods, and engage a broad community. The problem addressed by a learning cycle can be defined along a set of dimensions comprising what might be called the *diagnostic problem space*. Some, but not all of the dimensions that define this problem space include:

Problem Type(s): A cycle may emphasize a specific type of sub-optimal diagnosis; for example, some cycles may focus on diagnostic timeliness while others may focus on specific classes of errors such as cognitive errors or malfunctions in care process.[23,24]
Clinical Scope: A learning cycle may be focused on a specific disease or class of diseases, such as cancer generically or specific type(s) of cancer, or specific syndromes such as sepsis.[14,18,19,25]
Diagnostic Processes: At any given time, a learning cycle may be focused on specific aspects of diagnostic process such as history taking, data integration, or differential diagnosis.[16,19,26,27]
Methodological Approach(es): Some learning cycles may be focused on specific intervention methods or approaches such as algorithmically-based decision support, or problems endemic to the improvement process itself, such as measurement of diagnostic quality.[3,19,28]

As they evolve, learning cycles will inevitably shift position in this problem space. As learning cycles progress through successive iterations, the communities that direct them may intentionally shift their focus along any of the dimensions noted above in response to changing needs. So, as a collection, the cycles comprising the ecosystem will be able to adapt to changing needs and respond to emerging phenomena.

Shared Infrastructural Resources: Even though, as noted in the interviews, there is no fully comprehensive infrastructure supporting LHSs, many services supporting specific components of the learning cycle exist and can be borrowed for the proposed ecosystem from other LHS-focused efforts. Expert assistance to the communities conducting learning cycles will be required to ensure that these communities make best use of the infrastructure.[29,30]

On the D2K side of the learning cycle, platforms for managing clinical data, such as PopMedNet,[6,31–34] comprise important infrastructure.[13,35] These resources enable the sharing of data stored in multiple sites and amalgamating data needed for specific studies. “Big Data” platforms and algorithmic toolkits, routinely used by the ML/AI community, play a similar role.[32,34,36] More conceptually, some diagnostic excellence interviewees advocate for more structured recording of data from the diagnostic process to support the knowledge generation that occurs in the D2K component.[37]

In order to consistently bring that knowledge into practice, infrastructure on the K2P side of the cycle must have sufficiently detailed and unambiguous knowledge representation, be easily updated with new knowledge, and readily integrate into clinical workflows.[38] One important example of infrastructure that supports K2P is the “Knowledge Grid” (KGrid).[39,40] The KGrid provides a library of computable knowledge, which serves both as a repository of what each learning cycle has learned, and as a mechanism for generating tailored advice to drive practice change. Toolkits of change strategies, for example those currently in production by the Society to Improve Diagnosis in Medicine in partnership with the Institute for Healthcare Improvement under a grant from the Gordon and Betty Moore Foundation,[41] can also be seen as important examples of K2P infrastructure.

Inception, Scaling, and Composition: An important feature of a platform-based ecosystem is the capability to start relatively small, develop incrementally, and scale up—evolving from what already exists. At inception, the ecosystem might consist of a relatively small number of learning cycles/communities, assembled from projects already underway and focused on a small but diverse set of diagnostic challenges that can be mapped to the diagnostic problem space. These inaugural projects would be gradually modified to conform to the structure of learning cycles and integrated through the initiation of governance and the adoption of infrastructural services. As the ecosystem grows and develops with the addition of new cycles and communities, and the expansion of existing cycles, governance and infrastructure would evolve correspondingly. Because infrastructure enables economies of scale, the cost of establishing and maintaining each new cycle would reduce as the size of the ecosystem enlarges. In accord with Metcalfe's Law, the value of the ecosystem will increase as the square of the number of learning cycles supported.[42,43]

To fully deliver on the promise of the envisioned learning ecosystem for diagnostic excellence, not only must a number of communities of interest be coalesced, and corresponding LHSs be realized, but foundational elements spanning technology, policy, process, and people need to be in place to ultimately enable, incentivize, and drive some measure of composition into an even larger scale (nationwide or international) LHS for diagnostic excellence.[6] In the United States, such an emergent composition of a nationwide LHS is envisioned in a series of reports of the National Academy of Medicine dating back over a decade, in federal health IT strategic planning, and in the broadly endorsed multi-stakeholder consensus *Core Values Underlying a National-Scale Person-Centered Continuous Learning Health System* of the Learning Health Community incipient global grassroots movement.[44–47] Corresponding and synergistic efforts are underway across Europe and Asia as well.[48–53]

Culture and Values: At its core, the LHS vision at any level of scale is anchored in a cultural commitment to seizing every opportunity to iteratively learn from every interaction and experience, so as to improve the health of individuals and populations, as well as the system itself.[45,54,55] Optimizing opportunities for continuous and rapid improvement of the system as a whole requires not only the technical capability to share data and knowledge, but also a trust fabric, an incentive structure, and a cultural imperative to drive such sharing.[56]

To engender a learning ecosystem for diagnostic excellence, some problems require sharing beyond organizational boundaries. For instance, certain rare diseases may not occur with sufficient incidence in even the largest academic medical centers or regional networks of healthcare delivery systems, to empower researchers with sufficiently large datasets to study and improve diagnosis. In these cases, mobilizing sharing across a network becomes an imperative. Indeed, researchers pioneering statistical methods to learn valid and trustworthy lessons from digitized observational data have long recognized the need for massive data sharing to enable such types of learning and improvement.[57]

Optimizing our capacity for diagnostic excellence requires democratization that can only be underpinned by such sharing. A cultural commitment to principles such as open science, replicability, and scientific integrity, bring about an imperative for creating opportunities for sharing of data, information, and knowledge; many great minds need to be able to examine the data and the knowledge generated from it, so that they can critique and iteratively improve upon it; in turn, processes need to be open to, inclusive of, and participated in by stakeholders spanning the health compass.[45,58] Breaking down traditional silos and barriers to such sharing, collaboration, coopetition, and ultimately democratization, are preconditions to ultimately optimizing our shared potential for continuously, rapidly, and systemically learning and improving diagnosis.

Summary: Looking Ahead to an Invitational Meeting

Grounded in experience to date with Learning Health Systems and the thoughts expressed in 32 interviews, this initial White Paper has outlined an approach that will address five key aims to accelerate progress in diagnostic excellence: 1) incorporating multiple complementary perspectives, 2) fostering improvements by creating and applying new knowledge; 3) making improvement continuous; 4) stimulating co-production and learning across groups; and 5) achieving economy of scale and scope.

These plans require further refinement and much more detail, but many key questions remain open. The invitational meeting planned for March 2018 will strive to provide this refinement and detail. Many of the experts we interviewed will participate in this meeting. This White Paper will be one, of many, sources of information to stimulate discussions at the meeting.

References

1. Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, The National Academies of Sciences, Engineering, and Medicine. 9. The Path to Improve Diagnosis and Reduce Diagnostic Error. In: Balogh EP, Miller BT, Ball JR, eds. *Improving Diagnosis in Health Care*. Washington, D.C.: National Academies Press; 2015:355-401. doi:10.17226/21794

2. Omenn GS. Grand Challenges and Great Opportunities in Science, Technology, and Public Policy. *Science*. 2006;314(5806):1696-1704. doi:10.1126/science.1135003

3. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. *BMJ Qual Saf*. 2017;26(6):484-494. doi:10.1136/bmjqs-2016-005401

4. Sittig DF. Grand challenges in medical informatics? *J Am Med Inform Assoc*. 1994;1(5):412.

5. Wachter RM. Why Diagnostic Errors Don't Get Any Respect—And What Can Be Done About Them. *Health Aff (Millwood)*. 2010;29(9):1605-1610. doi:10.1377/hlthaff.2009.0513

6. Friedman CP, Rubin JC, Sullivan KJ. Toward an Information Infrastructure for Global Health Improvement. *Yearb Med Inform*. 2017;26(01):16-23. doi:10.15265/IY-2017-004

7. Friedman CP, Allee NJ, Delaney BC, et al. The science of Learning Health Systems: Foundations for a new journal. *Learn Health Syst*. 2017;1(1):n/a-n/a. doi:10.1002/lrh2.10020

8. Balas E, Boren S. Managing Clinical Knowledge for Health Care Improvement. *Yearb Med Inform*. 2000;2000:65-70.

9. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. *J R Soc Med*. 2011;104(12):510-520. doi:10.1258/jrsm.2011.110180

10. Batalden M, Batalden P, Margolis P, et al. Coproduction of healthcare service. *BMJ Qual Saf*. 2016;25(7):509-517. doi:10.1136/bmjqs-2015-004315

11. Berwick DM. Era 3 for Medicine and Health Care. *JAMA*. 2016;315(13):1329-1330. doi:10.1001/jama.2016.1509

12. Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, The National Academies of Sciences, Engineering, and Medicine. *Improving Diagnosis in Health Care*. (Balogh EP, Miller BT, Ball JR, eds.). Washington, D.C.: National Academies Press; 2015. doi:10.17226/21794

13. Wiens J, Shenoy ES, Weinstein RA. Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. *Clin Infect Dis*. doi:10.1093/cid/cix731

14. Henriksen K, Dymek C, Harrison MI, Brady PJ, Arnold SB. Challenges and opportunities from the Agency for Healthcare Research and Quality (AHRQ) research summit on improving diagnosis: a proceedings review. *Diagnosis*. 2017;4(2). doi:10.1515/dx-2017-0016

15. McDonald KM, Bryce CL, Graber ML. The patient is in: patient involvement strategies for diagnostic error mitigation. *BMJ Qual Saf*. 2013;22(Suppl 2):ii33-ii39. doi:10.1136/bmjqs-2012-001623

16. Ely JW, Graber ML. Preventing diagnostic errors in primary care. *Am Fam Physician*. 2016;94(6). https://pdfs.semanticscholar.org/434e/70706b77b41b8151aa13ce0db3303ca954aa.pdf. Accessed September 8, 2017.

17. Krumholz HM. Big Data and New Knowledge in Medicine: The Thinking, Training, And Tools Needed for A Learning Health System. *Health Aff (Millwood)*. 2014;33(7):1163-1170. doi:10.1377/hlthaff.2014.0053

18. Singh H, Graber ML, Hofer TP. Measures to Improve Diagnostic Safety in Clinical Practice. *J Patient Saf*. 2017;Publish Ahead of Print. doi:10.1097/PTS.0000000000000338

19. National Quality Forum. Improving Diagnostic Quality and Safety: Final Report. September 2017. http://www.qualityforum.org/Publications/2017/09/Improving_Diagnostic_Quality_and_Safety_Final_Report.aspx. Accessed September 29, 2017.

20. Donabedian A. The quality of care: How can it be assessed? *JAMA*. 1988;260(12):1743-1748. doi:10.1001/jama.1988.03410120089033

21. Singh H, Zwaan L. Inpatient notes: Reducing diagnostic error-a new horizon of opportunities for hospital medicine. *Ann Intern Med*. 2016;165(8):H02-H04. doi:10.7326/M16-2042

22. Lieu TA, Platt R. Applied Research and Development in Health Care—Time for a Frameshift. *N Engl J Med*. 2017;376(8):710-713. doi:10.1056/NEJMp1611611

23. Okafor N, Payne VL, Chathampally Y, Miller S, Doshi P, Singh H. Using voluntary reports from physicians to learn from diagnostic errors in emergency medicine. *Emerg Med J*. 2016;33(4):245-252. doi:10.1136/emermed-2014-204604

24. CRICO Strategies. *Annual Benchmarking Report: Malpractice Risks in the Diagnostic Process.* Cambridge, MA: CRICO Strategies; 2014. www.rmfstrategies.com/benchmarking. Accessed December 8, 2017.

25. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med*. 2015;7(299):299ra122-299ra122. doi:10.1126/scitranslmed.aab3719

26. Singh H, Sittig DF. Advancing the science of measurement of diagnostic errors in healthcare: The Safer Dx framework. *BMJ Qual Saf*. 2015;24(2):103-110. doi:10.1136/bmjqs-2014-003675

27. Meyer AND, Singh H. Calibrating how doctors think and seek information to minimise errors in diagnosis. *BMJ Qual Saf*. 2017;26(6):436-438. doi:10.1136/bmjqs-2016-006071

28. Newman-Toker DE, McDonald KM, Meltzer DO. How much diagnostic safety can we afford, and how should we decide? A health economics perspective. *BMJ Qual Saf*. 2013;22(Suppl 2):ii11-ii20. doi:10.1136/bmjqs-2012-001616

29. Singh H. Improving Diagnostic Safety in Primary Care by Unlocking Digital Data. *Jt Comm J Qual Patient Saf*. 2017;43(1):29-31. doi:10.1016/j.jcjq.2016.10.007

30. Graber ML, Byrne C, Johnston D. The impact of electronic health records on diagnosis. *Diagnosis*. 2017;0(0). doi:10.1515/dx-2017-0012

31. Vogel J, Brown JS, Land T, Platt R, Klompas M. MDPHnet: Secure, Distributed Sharing of Electronic Health Record Data for Public Health Surveillance, Evaluation, and Planning. *Am J Public Health*. 2014;104(12):2265-2270. doi:10.2105/AJPH.2014.302103

32. Ames MJ, Bondy MJ, Johnson MSC, Wade MTD, Davidson PA, Kahn MM. Analysis of Federated Data Sharing Platforms for a Regional Data Sharing Network. *AMIA 2013 Summit Clin Res Inform San Franc CA*.

33. Brown J, Balaconis E, Mazza M, et al. PS1-46: HMORNnet: Shared Infrastructure for Distributed Querying by HMORN Collaboratives. *Clin Med Res*. 2012;10(3):163-164. doi:10.3121/cmr.2012.1100.ps1-46

34. Toh S, Platt R, Steiner JF, Brown JS. Comparative-Effectiveness Research in Distributed Health Data Networks. *Clin Pharmacol Ther*. 2011;90(6):883-887. doi:10.1038/clpt.2011.236

35. Simon GE, Coronado G, DeBar LL, et al. Data Sharing and Embedded Research. *Ann Intern Med*. 2017.

36. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Inform Assoc JAMIA*. 2016;23(5):899-908. doi:10.1093/jamia/ocv189

37. Delaney BC, Kostopoulou O. Decision support for diagnosis should become routine in 21st century primary care. *Br J Gen Pract*. 2017;67(664):494-495.

38. Nurek M, Kostopoulou O, Delaney BC, Esmail A. Reducing diagnostic errors in primary care. A systematic meta-review of computerized diagnostic decision support systems by the LINNEAUS collaboration on patient safety in primary care. *Eur J Gen Pract*. 2015;21(sup1):8-13. doi:10.3109/13814788.2015.1043123

39. Flynn AJ, Shi W, Fischer R, Friedman CP. Digital knowledge objects and digital knowledge object clusters: Unit holdings in a learning health system knowledge repository. In: *Proceedings of the Annual Hawaii International Conference on System Sciences*. Vol 2016-March. ; 2016:3308-3317. doi:10.1109/HICSS.2016.413

40. Flynn AJ, Bahulekar N, Boisvert P, et al. Architecture and Initial Development of a Digital Library Platform for Computable Knowledge Objects for Health. *Stud Health Technol Inform*. 2017;235:496-500. doi:10.3233/978-1-61499-753-5-496

41. Grant Detail: Addressing Diagnostic Error Using the Patient Safety Transformation Model. Gordon and Betty Moore Foundation. https://www.moore.org/grant-detail?grantId=GBMF5757. Accessed November 30, 2017.

42. Metcalfe B. Metcalfe's Law: A network becomes more valuable as it reaches more users. *InfoWorld San Mateo*. 1995;17(40):53.

43. Hendler J, Golbeck J. Metcalfe's law, Web 2.0, and the Semantic Web. *Web Semant Sci Serv Agents World Wide Web*. 2008;6(1):14-20. doi:10.1016/j.websem.2007.11.008

44. Learning Health System Series. *Natl Acad Med*. https://nam.edu/programs/value-science-driven-health-care/learning-health-system-series/. Accessed January 19, 2018.

45. LHS Core Values. Learning Health Community. http://www.learninghealth.org/corevalues/. Accessed January 19, 2018.

46. Endorsers. Learning Health Community. http://www.learninghealth.org/endorsers/. Accessed January 19, 2018.

47. The Office of the National Coordinate for Health Information Technology. Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap. October 2015. https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf.

48. The Learning Healthcare Project. http://www.learninghealthcareproject.org/location_map.php. Accessed January 19, 2018.

49. About the Farr Institute. Farr Institute. http://www.farrinstitute.org/about. Accessed January 19, 2018.

50. Stucki G. Swiss Learning Health Systems (SLHS): A national platform for health systems and services research, policy and practice. January 2016. https://www.slhs.ch/images/Proposal_Swiss_Learning_Health_System.pdf.

51. Delaney BC, Curcin V, Andreasson A, et al. Translational Medicine and Patient Safety in Europe: TRANSFoRm—Architecture for the Learning Health System in Europe. BioMed Research International. doi:10.1155/2015/961526

52. International Symposium on improvement of the medical information infrastructure in Tohoku (Feb. 23-25) | ToMMo. http://www.megabank.tohoku.ac.jp/english/timeline/20150225_01/. Accessed January 19, 2018.

53. 2nd NHRI-ToMMo Conference (July 15-16, 2016) | ToMMo. http://www.megabank.tohoku.ac.jp/english/timeline/20160527_01/. Accessed January 19, 2018.

54. Platt J, Spector-Bagdady K, Platt T, et al. Ethical, legal, and social implications of learning health systems. *Learn Health Syst*. 2018;2(1):n/a-n/a. doi:10.1002/lrh2.10051

55. Rubin JC, Friedman CP. Weaving together a healthcare improvement tapestry: learning health system brings together health it data stakeholders to share knowledge and improve health. *J AHIMA*. 2014;85(5):38–43.

56. Friedman CP, Rubin J, Brown J, et al. Toward a science of learning systems: A research agenda for the high-functioning Learning Health System. *J Am Med Inform Assoc*. 2015;22(1):43-50. doi:10.1136/amiajnl-2014-002977

57. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *BMJ*. 2009;338:b81. doi:10.1136/bmj.b81

58. Rubin JC. Patient empowerment and the Learning Health System. *Learn Health Syst*. 2017;1(3):n/a-n/a. doi:10.1002/lrh2.10030

Appendix A: Interview Guides

Diagnostic Excellence Group

Thank you for taking time today to speak with me. I work for a diagnostic improvement project funded by the Gordon and Betty Moore Foundation. Specifically, this project will explore how three communities of researchers—those currently studying diagnosis, those focused on Learning Health Systems, and those developing machine learning techniques—might come together to address this goal. With this interview, I'm hoping to learn about your own research and explore your views about this potential collaboration.

The important ideas you provide today will contribute directly to a White Paper we will be preparing for the Foundation later this fall.

The interview has two parts. In the first, we'll focus on diagnostic research; in the second we'll talk a bit about Learning Health Systems and machine learning.

Before we begin, do I have permission to record our conversation to aid my notetaking?

Part I: Research on Diagnosis

1. I'd like to start by getting a sense of your personal journey to studying medical diagnosis. How did your interest in this problem develop?

2. Tell me about your research and how your work has contributed to defining the diagnostic error problem and finding or adopting solutions.

3. Looking beyond your own research, what have been the major accomplishments in research on improving diagnosis in the past 10 years?

4. What are the most important questions going forward?

5. Thinking of potential collaborations among researchers in medical diagnosis, what mechanisms, resources, or infrastructure that are not currently in place would—if added—catalyze progress in the field?

6. As the field begins to identify methods to improve diagnosis, the challenge remains to translate that knowledge into practice. What are your ideas for how this might be done? Can you point me to good examples of how this translation is happening now? [*Cue:* Describe how you would prioritize and engage stakeholders—HCOs, providers, patients, payers, HIT?]

Part II: The Other Communities in this Potential Collaboration

7. In 2007 the U.S. Institute of Medicine advanced the concept of the Learning Health System. What does this term mean to you?

    a. *(If interviewee seems reasonably knowledgeable about the LHS)*
    Based your understanding of the term, in what ways might approaches associated with the Learning Health System be applied to improve diagnosis? [*Cue:* How can an LHS support your mission of improving diagnosis? Where do you perceive limitations?]

b. Are you personally involved in any work that relates to the Learning Health System?

8. Tell me about how you perceive the potential for 'big data' and 'machine learning' in improving medical diagnosis.
   [*Cue:* how does this potential make you feel excited, uncertain, or uncomfortable?]

Conclusion: Thank you for your time and your excellent thoughts. I hope we can contact you again as this work proceeds. If you think of anything you wanted to add, feel free to e-mail me.

Machine Learning and Artificial Intelligence Group

Thank you for taking time today to speak with me. I work for a diagnostic improvement project funded by the Gordon and Betty Moore Foundation. Specifically, this project will explore how three communities of researchers—those currently studying diagnosis, those focused on Learning Health Systems, and those developing machine learning and AI techniques applicable to medical diagnosis—might come together to address this goal. With this interview, I'm hoping to learn about your own research and explore your views about this potential collaboration.

The important ideas you provide today will contribute directly to a White Paper we will be preparing for the Foundation later this fall.

The interview has two parts. In the first, we'll focus on machine learning and related methods, particularly as they apply to medical diagnosis; in the second we'll talk a bit about Learning Health Systems.

Before we begin, do I have permission to record our conversation to aid my notetaking?

Part I: Machine Learning and AI Research

1. I'd like to start by getting a sense of your personal journey to applying machine learning and AI to health-related problems. How did your interest in these problems develop?

2. Tell me about your specific research and how your work to date has contributed to solving health-related problems.

3. Looking beyond your own research, what have been the major accomplishments of machine learning and AI applied to medical diagnosis in the past 5 years?

4. Looking forward, what are the most important questions surrounding the application of machine learning techniques to the improvement of medical diagnosis?

5. What progress to do you expect to see in the next 5-10 years? What will emerge as the most important methods contributing to this progress?

6. As your field begins to identify methods to improve diagnosis, the challenge remains to translate that knowledge into practice. What are your ideas for how this might be done? Can you point me to good examples of how this translation is happening now?
   [*Cue:* How would or do you collaborate across disciplines into clinical practice?

What connections, resources, mechanisms, or infrastructure are missing to make that happen?]

Part II: Learning Health Systems

7. In 2007 the U.S. Institute of Medicine advanced the concept of the Learning Health System. What does this term mean to you?

    a. *(If interviewee seems reasonably knowledgeable about the LHS)* Based your understanding of the term, in what ways might approaches associated with the Learning Health System be applied to improve diagnosis?

    b. Are you personally involved in any work that relates to the Learning Health System?

    c. What do you see as the relationship between the Learning Health System and machine learning?

Conclusion: Thank you for your time and your excellent thoughts. I hope we can contact you again as this work proceeds. If you think of anything you wanted to add, feel free to e-mail me.

Learning Health Systems Group

Thank you for taking time today to speak with me. I work for a diagnostic improvement project funded by the Gordon and Betty Moore Foundation. Specifically, this project will explore how three communities of researchers—those currently studying diagnosis, those focused on Learning Health Systems, and those developing machine learning and AI techniques applicable to medical diagnosis—might come together to address this goal. With this interview, I'm hoping to learn about your own research and explore your views about this potential collaboration.

The important ideas you provide today will contribute directly to a White Paper we will be preparing for the Foundation later this fall.

The interview has two parts. In the first, we'll focus on the Learning Health System. In the second, we'll talk more about how the LHS can be applied as a framework toward the goal of improving medical diagnosis and how that might connect to machine learning research.

Before we begin, do I have permission to record our conversation to aid my notetaking?

Part I: Learning Health Systems

1. I'd like to start by getting a sense of your personal journey to working on Learning Health Systems. How did your interest in these approaches develop?

2. Briefly tell me about your specific research and how your work to date has contributed to the development and implementation of LHSs.

3. Looking beyond your own research, what have been the major accomplishments toward implementing an LHS? Cam you point me to good examples of real-world implementations of learning health systems?

4. Looking forward, what are the most important questions in the development of an LHS?

Part II: The Learning Health Systems and Interdisciplinary Collaboration

5. Where do you see the interaction of an LHS framework with mechanisms built specifically to improve medical diagnosis?

    a. As a framework, the LHS necessitates that knowledge is translated into practice. More specifically, how do you see LHS approaches translating knowledge into practice to improve diagnosis?

6. Researchers in diagnostic error focus on cognitive and system processes that can be intervened among three groups of core stakeholders: healthcare organizations, providers, or patients (and to a lesser extent, payers and HIT). How would you prioritize and engage these stakeholders in building an LHS infrastructure?

7. Describe your vision for how LHS researchers and advocates can better tap into the work of our colleagues who work in machine learning techniques and AI? What can we learn from them now and in the future?

    a. In what ways can machine learning and AI work to improve diagnosis?

Conclusion: Thank you for your time and your excellent thoughts. I hope we can contact you again as this work proceeds. If you think of anything you wanted to add, feel free to e-mail me.

Appendix B: List of Interviewees

We appreciate the time and insights provided by these experts from each researcher group:

Diagnostic Excellence
Todd Allen
Paul Epner
William Follansbee
Mark Graber
Kerm Henriksen
Jason Maude
Prashant Mahajan
Kathryn McDonald
Michael Millenson
David Newman-Toker
Shantanu Nundy
Art Papier
Gordon Schiff
Sue Sheridan
Hardeep Singh
Bob Trowbridge
Bob Wachter
Alan Weil

Machine Learning and Artificial Intelligence in Healthcare
Gari Clifford
Noemie Elhadad
Nigam Shah
Peter Szolovits
Byron Wallace
Jenna Wiens

Learning Health Systems
Julia Adler-Milstein
Brendan Delaney
Michael McGinnis
Mark Musen
Richard Platt
Aziz Sheikh
Kevin Sullivan