Toward a science of learning systems: a research agenda for the high-functioning Learning Health System

Charles Friedman,1 Joshua Rubin,1 Jeffrey Brown,2 Melinda Buntin,3 Milton Corn,4 Lynn Etheredge,5 Carl Gunter,6 Mark Musen,7 Richard Platt,2 William Stead,3 Kevin Sullivan,8 Douglas Van Houweling1

ABSTRACT
Objective The capability to share data, and harness its potential to generate knowledge rapidly and inform decisions, can have transformative effects that improve health. The infrastructure to achieve this goal at scale—marrying technology, process, and policy—is commonly referred to as the Learning Health System (LHS). Achieving an LHS raises numerous scientific challenges.

Materials and methods The National Science Foundation convened an invitational workshop to identify the fundamental scientific and engineering research challenges to achieving a national-scale LHS. The workshop was planned by a 12-member committee and ultimately engaged 45 prominent researchers spanning multiple disciplines over 2 days in Washington, DC on 11–12 April 2013.

Results The workshop participants collectively identified 106 research questions organized around four system-level requirements that a high-functioning LHS must satisfy. The workshop participants also identified a new cross-disciplinary integrative science of cyber-social ecosystems that will be required to address these challenges.

Conclusions The intellectual merit and potential broad impacts of the innovations that will be driven by investments in an LHS are of great potential significance. The specific research questions that emerged from the workshop, alongside the potential for diverse communities to assemble to address them through a ‘new science of learning systems’, create an important agenda for informatics and related disciplines.

BACKGROUND AND SIGNIFICANCE
Vision of a Learning Health System
The capability to share data—and harness its potential to generate knowledge rapidly and inform decisions—can have transformative effects on complex systems that produce goods and provide services. Within specific sectors of the economy, individual organizations and collaborating groups have leveraged their data to increase productivity, gain competitive advantage, and revolutionize business models.1–4 Domains ranging from transportation to agriculture, to environmental sciences, to supply chains, to military services and intelligence have demonstrated the capacity and need to derive significant benefits from leveraging data, systems, and human interconnectedness on ever-increasing scales.4

However, attempts to employ these types of approaches to realize transformative impacts for the most challenging societal problems—including health and education—have not enjoyed similar success.5 While the health sector can point to examples of large clinical research networks, as well as increasing adoption of electronic health record (EHR) systems and other information technologies, this sector has not undergone the type of IT-enabled transformation visible across other industries.6

One widely-held vision for realizing the transformations previously described, at significant scale and scope in the health domain, is generally known as the Learning Health System (LHS). A series of publications from the Institute of Medicine (IOM)7–8 has described the LHS and documented the need for it. The IOM defines the LHS as a vision for an integrated health system “…in which progress in science, informatics, and care culture align to generate new knowledge as an ongoing, natural by-product of the care experience, and seamlessly refine and deliver best practices for continuous improvement in health and healthcare.”7

Though articulated in various forms, the underlying concept is straightforward: harness the power of data and analytics to learn from every patient, and feed the knowledge of “what works best” back to clinicians, public health professionals, patients, and other stakeholders to create cycles of continuous improvement.

Consistent with the characteristics of a continuously learning system as articulated by the IOM in 2012,9 the LHS can also be thought of as common infrastructure, governance, and incentive structures coupled with shared values that enable such a health system and culture. The LHS also promises to empower a new era of personalized medicine.10

Realizing the promise of personalized medicine requires a key component of what the LHS enables—“knowing what works, understanding why it works, learning for whom it works, and applying that knowledge to address patient needs…”11

The Office of the National Coordinator for Health Information Technology (ONC) set achievement of rapid learning as the pinnacle goal of its 5-year strategic plan (2011–2015).12 In 2014, ONC identified the national-scale LHS as its 10-year strategic goal and stated that, “This ‘LHS’ should also enable lower health care costs, improved population health, truly empower consumers, and drive innovation.”13

The LHS imperative
The imperative to achieve an LHS is anchored in a number of realities of the current health system
that render its status quo unsustainable. The USA spends roughly 18% of its gross domestic product (GDP) on healthcare, and, by some projections, could almost double that share by 2040. Healthcare spending per capita by the USA greatly exceeds that of other ‘first world’ countries, while underperforming these nations on many indices of population health status. As one study notes, “While the U.S. performs well on breast and colorectal cancer survival rates, it has among the highest rates of potentially preventable deaths from asthma and amputations due to diabetes, and rates that are no better than average for in-hospital deaths from heart attack and stroke.” Other studies suggest that as much as $750 billion annually, approximately 25% of total expenditures, is literally wasted on administrative inefficiencies, fraud, abuse, and treatments that do not work. Even with such spending, it has long been recognized that patients do not receive over 45% of recommended care and quality is improving slowly. Such urgency itself underpins a series of LHS-related IOM reports dating back to the original ‘The Learning Healthcare System’ vision report of 2007.

There is a growing recognition of the US healthcare system’s inability to routinely study its own behaviour; an LHS would provide such capability, and would significantly address many of the current challenges faced by the system. For example, researchers required 4 years to recognize that a commonly prescribed drug was leading to tens of thousands of deaths. In the presence of an LHS infrastructure to support routine surveillance leveraging timely data from a population the size of the USA, this safety signal could have been detected in approximately 2 months. Researchers who developed and tested analytic methodologies for learning from EHR data, utilizing the UK’s General Practice Research Database (GPRD), highlighted the importance of learning at large scale. These researchers concluded that, “Ideally future databases should be much larger than GPRD, which includes about eight million patients. On the basis of our work to date, we estimate that 40–50 million patients are needed for the breadth of future studies we can envisage.”

The growing recognition of the importance of an LHS is seen in an increasing number of publications, amplifying the calls from the IOM. For example, a 2013 editorial in the New England Journal of Medicine cites the LHS as one of three essential steps needed to reduce healthcare’s GDP Footprint, while another has challenged all academic health centers to become LHSs, so they can be drivers of needed change in health research and health care delivery. Further illustrating the emerging national imperative, the entire July 2014 issue of Health Affairs was devoted to the potential of ‘big data’ to transform healthcare and health, including an extensive discussion of a new rapid-learning agenda.

Overall, realizing the LHS on a national scale by harmonizing and synergizing multiple efforts has taken on many features of a “Big Hairy Audacious Goal”, which “changes the time frame and simultaneously creates a sense of urgency...”.

Pathway and progress toward an LHS

A foundation for the LHS is emerging as the nation’s healthcare system becomes increasingly digital. This is due in part to an estimated over $30 billion federal investment to promote meaningful use of health information technology through the HITECH Act provisions of the American Recovery and Reinvestment Act of 2009. Averaging several indices, the nation’s health system is approximately 35% digital now but is expected to be 80% digital by 2019. Interoperability requirements built into HITECH will make significant but imperfect progress toward standardizing data stored in EHRs.

Other foundational elements of a national LHS include organizations that, within their own boundaries, have become LHSs. While several organizations have made significant progress toward learning status, Kaiser Permanente, Geisinger Health System, and the US Veterans Administration stand out as examples of institutions that are already harnessing the power of data to improve the health of the populations they serve. In addition, federated learning networks among independent organizations have formed to share data and harvest the benefits of learning from it, with the HMO Research Network, the Vaccine Safety Datalink, the Food and Drug Administration (FDA) Mini-Sentinel, and CancerLinQ as key examples. These developments both document the benefits to be derived from a nationally-scaled LHS and suggest that a national-scale LHS is technically and organizationally achievable. A recent paper examining four emerging data networks concludes: “these four programs [are] examples of the first stage in the development of a shared national big-data resource that leverages the investments of many agencies and organizations for the benefit of multiple networks and users.”

From these developments, it is evident that achievement of a national-scale LHS will not be the work of a single organization, stakeholder group, or governmental entity. Rather, it is anticipated that the LHS will require active participation of and cooperation among multiple and diverse stakeholders, nationwide and ultimately globally. While there is increasing national focus in the USA, the LHS is increasingly a global phenomenon. There is increasing collaboration among international and multinational initiatives including TRANSFoRm (which “aims to develop the technology that facilitates a learning healthcare system” for the European Union), with explicit recognition that key LHS enablers such as standards should be developed in ways conducive to global harmonization.

The LHS research agenda

The LHS is a consummate challenge for researchers across a range of scientific disciplines. Progress at each stage of development will require the best possible answers to an enormously wide range of questions.

The scope of these challenges was first suggested at a 2010 IOM workshop that focused on the digital infrastructure for the LHS. Participants in this workshop identified the LHS as an ultra-large scale, ultra-complex, cyber-social-technical system, recognizing that at a national and ultimately global scale, the LHS is a massive system of systems. That same year, in a separate study addressing similar issues, a group of researchers wrote: “The U.S. urgently needs a major initiative to develop software and systems engineering foundations for a national-scale health information network (NHIN). The NHIN will be an ultra-large-scale (ULS) system. An ULS systems perspective therefore must guide these activities.”

It follows that achieving the LHS vision requires more than merely addressing difficult engineering challenges; profound and difficult socio-technical challenges will need to be addressed by bringing together a wide-range of disciplines. Moreover, the LHS is not a ‘build out’ project. Along the way, many challenges—including government policies, competition among healthcare providers, rivalries among technology vendors and developers, concerns related to balancing privacy and security against the value created by secure sharing of data—will need to be addressed. The LHS will need to be sustainable through a sound business and governance model. The complex trans-disciplinary work ahead may build upon emerging specializations of the past.
decade, including “implementation and integration sciences... the theory and methods necessary to tackle complex societal issues and problems”.42

METHODS: THE WORKSHOP
In this spirit, the National Science Foundation convened a 2-day invitational workshop to identify the fundamental scientific and engineering research challenges to achieving a national-scale LHS. The workshop was planned by a 12-member committee and ultimately engaged 45 prominent researchers spanning multiple disciplines over two full days in Washington, DC on 11–12 April 2013. While fully recognizing that the LHS is ultimately a global challenge and that the findings would have potential to generalize across national boundaries, support from a US national agency directed a national focus for the workshop.

The participants were challenged to view the LHS as a ‘system after next’.43 As such, the LHS can succeed only by creating novel combinations of role, process and technology. This must occur by working back from the future, not by figuring out how to fix the various problems with a current system that fails to learn rapidly, routinely, and at scale.

The research questions, that were the focal product of the workshop, developed through two rounds of breakout group deliberations. In each round, the participants divided into four groups with different composition. In the first round, each group examined one of four use cases describing the operation of a high-functioning LHS at scale. The use cases separately addressed LHS operation in support of health care quality, population health, personal health, and biomedical research and illuminated what learning could mean—and how it could ultimately improve health—in each of these contexts. Each group identified the research challenges that must be addressed to realize a very high-functioning LHS capable of executing its assigned use case. The LHS use cases may be found at http://healthinformatics.umich.edu/lhs/nsfworkshop.

In the second round, each group focused on one of four broad system-level requirements for the high-functioning LHS: 1. An LHS trusted and valued by all stakeholders 2. An economically sustainable and governable LHS 3. An adaptable, self-improving, stable, certifiable, and responsive LHS 4. An LHS capable of engendering a virtuous cycle of health improvement.

Each group identified the research challenges that must be addressed to realize a very high-functioning LHS capable of meeting the assigned requirement.

As part of the plenary session closing the workshop, each participant wrote the research challenge question stemming from the 2-day meeting that he/she felt was most important and that he/she could envision himself/herself being personally involved in addressing.

Following the workshop, the planning committee compiled, refined, and organized the research challenges into the form of questions. The questions were circulated back to all participants for review and comment, which resulted in further refinement of the questions. A modification of the four system-level requirements used for the second round of the workshop’s group deliberations formed the basis for organizing the key findings. Ultimately, 106 questions were generated and they were organized into 4 categories and 19 subcategories. In reviewing the report, the 19 subcategories themselves (box 1) may paint the most comprehensive picture of the breadth and nature of the research challenge questions generated through the workshop methodology. The challenges and questions that follow form a broad scientific agenda for realizing a high-functioning LHS. Further, in addition to the granular level of findings of the research challenge questions themselves, there was a second level of transcendent findings, illuminating themes interwoven into a new science of learning systems anchored in the socio-technical challenges identified.

RESULTS: THE RESEARCH CHALLENGES AND QUESTIONS
We provide below an illustrative sample of the research challenge questions generated as a formal result of the workshop. This illustrative sample was selected from among the full set of questions (n=106). The representative sample below includes at least one question from all subcategories, containing a mix of high-level and specific questions, with an intent of conveying the perspectives deriving from the diverse disciplines represented. A full listing is available in the report itself (http://healthinformatics.umich.edu/lhs/nsfworkshop).

Requirement 1: an LHS trusted and valued by all stakeholders

▸ What methods will be needed to measure confidence, trust, and trustworthiness?
What approaches will promote monitoring of confidence and trust at varying levels of scale, and thus enable improvement of the LHS over time?

Can the potential of a dataset to generate valid knowledge (fitness for use) be computed from the dataset itself? What metadata are required?

How can we analyze motives and incentives for security breaches and attacks on the LHS, and best structure governance and security policies and mechanisms to guard against them? What is the relationship between the real and perceived security of the LHS and stakeholders’ levels of confidence and trust in it?

What value metrics will assess the magnitude of improvement in health outcomes and cost-efficiency, as well as social and behavioral impacts, associated with LHS activities?

How can the LHS be designed, engineered, and operated as a self-defending and self-repairing system for purposes of protecting individual and institutional privacy and the integrity of data knowledge against malicious attack and accidental disclosure?

Requirement 2: an economically sustainable and governable LHS

What value is the LHS uniquely capable of producing, for which stakeholders, and how can this value be identified, measured, and incentivized? How can the science of networks, markets, game theory, and mechanism design inform the design and operation of the LHS?

How can predictive models and empirical studies address what ingredients, essential to standing up and sustaining the LHS, have no private rationale for funding or are unlikely to be funded privately, and therefore should be considered for catalysis stemming from public funding? How can lessons from other rapid learning systems and ultra-large scale systems in other sectors be applied to thinking about incentives and markets vis-à-vis the LHS?

How can predictive models and empirical studies address the relationship between incentive structures and data sharing and other key activities essential to the function of the LHS that will advance the public interest?

Requirement 3: an adaptable, self-improving, stable, certifiable, and responsive LHS

What features of the LHS, as a socio-technical ecosystem, will best position it to incorporate innovation in data sources that include real-time monitors, sensors, and devices intended to augment bodily functions and directly improve a person’s health?

How do we combine data generated in learning-from-doing and from observational data with careful experiments to be sure knowledge is scientifically solid and valid for stated purposes, but also as widely useful as possible?

What analysis must be conducted, what evidence must be gathered, and what arguments must be constructed and maintained over time, about the development, design, and operation of the system and about the environment in which it operates to enable system certification for safety, performance, and other critical system properties?

How do we make data sufficiently self-describing so that, for example, the system might be able to identify, without human intervention, data that are most relevant to addressing a question?

How can we understand and develop mechanisms to balance the tradeoff between speed (rapidity) and accuracy?

Requirement 4: an LHS capable of engendering a virtuous cycle of health improvement

What existing analytical and inferential methods serve the needs of the LHS, and what needs of the LHS require new methods? In particular, in the context of an LHS at scale, what methods are needed to address bias, confounding, propagation of erroneous information, spurious correlations, and other potential sources of mis-learning in the LHS?

What rigorous methods for data description and new data quality metrics will capture the semantics that enable rapid learning from (big) heterogeneous data streams? What approaches will describe fitness for use within the context of a specific purpose? For example, what are the semantics for characterizing data and identifying and describing bias?

How do we develop ways to communicate generated results, and surrounding uncertainties, to others who may wish to replicate (or build upon) the work done, as well as to the general public?

How can the LHS become smart enough to detect attempts to answer a question that is not answerable with the resources in the system?

What are the best approaches to measuring and understanding the predictive value of models resulting from big data?

What new methods are needed to determine the value of knowledge generated by the LHS and resulting actions?

To support health care quality improvement, how can the LHS develop a complex multidimensional model of a person’s future health status and communicate it to them or their trusted health-care provider to promote their health and wellbeing?

DISCUSSION: THE LHS AND AN EMERGING SCIENCE OF CYBER-SOCIAL LEARNING SYSTEMS

The challenges and specific research questions—organized into four areas corresponding to requirements a high-functioning LHS must meet—constitute one level of expression of the workshop’s findings. At a higher level, the findings in their totality and their multiple interactions generate a broader perspective, suggesting that rising to the challenge of the LHS may require a novel emergent science of large-scale learning systems best seen as an evolution from the science of information systems, through a science of cyber-physical systems, and ultimately to a science of cyber-physical-social ecosystems. While these concepts were explicitly discussed, the workshop did not describe this new science in detail, but rather was suggestive of it, with the LHS providing the real world, imperative, driving problem that brings this science into sharper focus. A second workshop product was therefore the suggestion of this new perspective, the applicability of which may transcend the health domain that was the original focus of this workshop.

Sciences have been developed and refined to study information systems, systems responsible for the digital representation, processing, storage, transmission, and use of information enabled by advances in computer science and engineering, as well as other fields. A next generation science of cyber-physical systems has also emerged, driven by advances in sensors, the ongoing miniaturization of digital computers, the wireless networking of edge devices to massive cloud computing systems, computer control of robotic machinery, and machine learning and artificial intelligence. After considering the challenges associated with realizing a national-scale LHS, the workshop participants came to recognize cyber-social ecosystems as a next step in this evolution.

A national-scale LHS will have to be understood and designed as such a cyber-social ecosystem: a large-scale, decentralized,
The LHS requires a new and significant crossing of capabilities that today are present only in these largely disconnected communities. Realizing this goal will require the NSF to continue to expand its embrace of health and healthcare as a driving domain for fundamental research, as it will require the NIH to expand its embrace of computing as a transformative power for biomedical research and healthcare innovation and delivery. The potential benefits are manifold. If the LHS is successful, there will be important lessons for how to leverage large amounts of real-world data, mechanisms for learning from such data, feedback components aimed at mobilizing lessons learned to inform decisions and actions, and cyber-social ecosystems—entailing networks of computers, machines, people, and organizations—to improve performance and bring about transformation in many other sectors outside of health. Therefore, additional mechanisms, of which this workshop was one starting point, and resources to bring together the largely disconnected research communities, will be needed to underpin an interdisciplinary, multi-stakeholder journey to realize such promise.

The specific research questions that emerged from the workshop, alongside the potential for diverse communities to assemble to address them through a “new science of learning systems”, were deeply exciting to the workshop participants.

Contributors The authors have participated to varying extents in the preparation of this manuscript, but all have met the four authorship criteria listed below: substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; drafting the work or revising it critically for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding This work was supported by National Science Foundation grant number IIS-1249354.

Competing interests Authors CF, JR, and JB are members of the interim steering committee for the Learning Health Community, a grass-roots, not-for-profit entity seeking to promote a national scale learning health system. Author MM serves on the scientific advisory boards of Apixio, HealthFidelity, and Kyron companies which have granted him stock options.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement A set of documents which reports the findings of the workshop in much greater detail, including all 106 research questions, can be accessed (http://healthinformatics.umich.edu/lhs/nsfworkshop).

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is appropriately cited and resolved.
Research and applications


See e.g.: Collins J, Porras J. Built To Last: Successful Habits of Visionary Companies. 1994.


The TRANSFoRm Project EU. Translational Research and Patient Safety in Europe. Funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no 247787 [TRANSFoRm]. http://www.transfofrmproject.eu/


Gorry GA. Envisioning the system after next. EDUCOM Rev. 1997:30


See e.g.: Cutler D. Where are the health care entrepreneurs? The failure of technological development and demonstration under grant agreement no 247787 [TRANSFoRm]. http://www.transfofrmproject.eu/


Toward a science of learning systems: a research agenda for the high-functioning Learning Health System

Charles Friedman, Joshua Rubin, Jeffrey Brown, Melinda Buntin, Milton Corn, Lynn Etheredge, Carl Gunter, Mark Musen, Richard Platt, William Stead, Kevin Sullivan and Douglas Van Houweling

*J Am Med Inform Assoc* published online October 23, 2014

Updated information and services can be found at: [http://jamia.bmj.com/content/early/2014/10/23/amiajnl-2014-002977](http://jamia.bmj.com/content/early/2014/10/23/amiajnl-2014-002977)

These include:

**References**

This article cites 11 articles, 4 of which you can access for free at: [http://jamia.bmj.com/content/early/2014/10/23/amiajnl-2014-002977#BIBL](http://jamia.bmj.com/content/early/2014/10/23/amiajnl-2014-002977#BIBL)

**Open Access**

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: [http://creativecommons.org/licenses/by-nc/4.0/](http://creativecommons.org/licenses/by-nc/4.0/)

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Topic Collections**

Articles on similar topics can be found in the following collections

Open access (175)

**Notes**

To request permissions go to: [http://group.bmj.com/group/rights-licensing/permissions](http://group.bmj.com/group/rights-licensing/permissions)

To order reprints go to: [http://journals.bmj.com/cgi/reprintform](http://journals.bmj.com/cgi/reprintform)

To subscribe to BMJ go to: [http://group.bmj.com/subscribe/](http://group.bmj.com/subscribe/)