From Data to Predictions and Decisions:
Enabling Evidence-Based Healthcare

Eric Horvitz
Microsoft Research

Computing Community Consortium
Version 6: September 16, 2010[1]

Nearly 2500 years ago, Hippocrates kicked off a revolution in healthcare by calling for the careful collection and recording of evidence about patients and their illnesses. This call—which first introduced the goal of sharing data among physicians to provide the best care possible for patients—established a foundation for the evolution of modern healthcare. Although 25 centuries have passed since Hippocrates' call, we have not yet attained the dream of true evidence-based healthcare. Large quantities of data about wellness and illness continue to be dropped on the floor, rather than collected and harnessed to optimize the provision of care. We are simply not yet doing the best that we can.

We now stand at the brink of a potential revolution in data-centric healthcare, enabled by advances in computer science. Such a revolution promises to enhance the quality of healthcare while cutting costs, and, more generally, enabling physicians to do the very best that is possible with realistically bounded healthcare resources. Doing the best that can be done with available resources aligns with the core promise that all physicians make when they solemnly raise their hand and recite the Hippocratic Oath upon receipt of their medical degree.

Enabling this vision of true evidence-based healthcare will require critical investments for translating key methods and insights into working systems, as well as for advances in core computer science research and engineering to address key conceptual bottlenecks and opportunities.

Collecting and analyzing data collected on health and illness promises to enhance the quality and efficacy of healthcare, and to enhance the quality and longevity of life. The collection and analysis of data can provide new insights about wellness and illness that can be operationalized. Data-centric methods allow us to transform *data* into *predictive models*. Predictive models can be used to generate forecasts with well-characterized accuracies about the future—or diagnoses about states of a patient that we cannot inspect directly. Such forecasts or diagnoses can be harnessed within procedures that generate recommendations for *actions in the world*, and decisions about *when it is best to collect more information about a situation before acting*, considering the costs and time delays associated with collecting more information to enhance a decision.

The pipeline of *data to prediction to action* can be used to automate or provide decision support for accurate triage and diagnosis, to generate well-calibrated predictions about health outcomes,

---

to produce effective plans for chronic disease management, and to formulate and evaluate larger-scale healthcare policies.

Collecting and analyzing large quantities of data also play a critical role in clinical and biomedical discovery. Beyond the daily practice of healthcare and wellbeing, methods for learning from data can provide the foundations for new directions in the clinical sciences via tools and analyses that identify subtle but important signals in the fusing of clinical, behavioral, environmental, genetic, and epigenetic data. Computational procedures can help with building insights via analyses and visualizations, as well in playing an active role in framing and designing clinical studies, and in the proposal and confirmation of biomedical hypotheses—with such methods as those that identify statistical associations among events or observations and help to confirm causal relationships. There is much to gain from the cost-effective optimization of healthcare delivery, and even more value can come by enabling fundamental scientific breakthroughs in biomedicine.

Key Ingredients—and a Bottleneck

Key computational ingredients for enabling evidence-based medicine have become available over the last 25 years. These ingredients include affordable large-scale computation and storage resources, connectivity among computing systems and devices, and the development of computational procedures for machine intelligence, including methods for learning and predicting, planning, and decision making. These core ingredients can be used to build platforms for information systems and applications that can provide assistance with such challenging problems as triage and diagnosis, acute therapies, long-term disease management, prediction of outcomes, and the formulation of ideal healthcare policies that weigh the costs and benefits of different actions.

While data has become available in limited settings, a *critical standing bottleneck* is the lack of data—based both in inadequate capture and in difficult challenges with sharing clinical data for research and development.

A torrent of valuable clinical data capturing information about patients' medical histories, symptomatology, diagnoses, outcomes, and responses to treatments and therapies is being lost rather than captured and curated. The wave of efforts to promote the installation of electronic health record (EHR) systems and the storage of healthcare data promises to change the availability of data that can drive analyses, decision support, and policy. Beyond installation of such systems, successful collection and use of data will depend on deeply embedding the use of such systems into the information and workflows of clinical medicine, and enhancing the efficiency with the capture of medical information. In addition to in-hospital data collection, new sources of data with relevance to wellness and disease are enabled by the growing ubiquity of capable smartphones, inexpensive sensors, and portable or embedded medical instruments, communicating with web services.

Collecting and storing healthcare data is an essential step; such data can often be exploited within organizations in a variety of ways to optimize the delivery of healthcare. However, further steps are needed to promote innovation: Computer scientists with skills in harnessing healthcare

data for building real-world decision support systems, and for analyzing the relative value of different interventions and policies, face hard challenges with gaining access to data resources. Whether data is collected in the course of daily activities or by EHR systems as part of care and workflow at hospitals and private practices, key technical and legal challenges remain with sharing data sets for research and development. Developing and fielding mechanisms and incentive programs that enable (and motivate) the sharing of data in a manner that respects the privacy preferences of patients and physicians will be essential. We will return to this challenge in recommendations below.

Blossoming of Computational Methods in Healthcare

Healthcare challenges have served as a focus of computer science research for decades within the realm of *biomedical informatics*, an interdisciplinary field at the crossroads of computer science and medicine. Indeed, core ideas and efforts in machine learning, reasoning, and discovery with probabilistic models—methods now front and central research topics in computer science—were developed initially within biomedical informatics in a surge of effort in the 1980s. These explorations included difficult challenges on the automation of decision support systems for medical diagnosis, testing, and therapy planning.

The 1980s were marked by the blossoming of computing research on probabilistic models and methods. For all of the innovation, data sets were rarely available, so researchers relied on domain experts for data. That is, in lieu of having access to collections of data, expert clinicians provided assessments of important distinctions and probabilities. Even through great data paucity, computational methods were developed for performing automated diagnosis from symptoms, signs, and other information about patients. Tractable computational procedures were also developed for identifying the next best test or sequence of tests to perform—in pursuit of a diagnosis. The latter procedures consider the current uncertainty in a diagnosis and balance the expected informational value and costs of tests. Computer-based representations and methods were also developed performing automated decision analyses for recommending ideal actions in accordance with the preferences of patients. These methods consider the inferred uncertainties in situations and outcomes.

A variety of investigations were undertaken with the use of probabilistic and decision-analytic models, and prototypes were developed, addressing a variety of challenges, such as the design of interfaces for physicians, explaining inferences and recommendations to physicians and patients, and such focused efforts as optimizing under uncertainty, the transportation of trauma casualties in a county.

Machine learning methods, building upon and extending older powerful statistical procedures for making inferences from data, blossomed in the late-1980s and 1990s. The methods were stimulated by—and in turn influenced—the amassing of large amounts of data in a variety of fields. Machine learning and data mining evolved into an active and vigorous subdiscipline of computer science with a passionate community of researchers.

We now have available an array of computational procedures for automatically constructing diagnostic and predictive models from case libraries of data. These models sit at the heart of

diagnostic and forecasting systems that make predictions about health states and outcomes for individuals and populations. The predictive models created via machine learning can be used with the computational methods for inference and decision support developed over the last two decades—methods that were developed and refined with knowledge bases constructed with the assistance of experts. However, problems with the capture and availability of medical data limit the impact and usage of the computational methods.

Opportunities on the Horizon

The *data to prediction to action* pipeline promises to catalyze transformative changes in the cost-effective delivery of quality and personalized healthcare—spanning applications with the handling of acute illnesses, longer-term disease management, and the promotion of health and preventative care. Algorithms for constructing predictive models and for identifying ideal actions under uncertainty can also have broad application with the critiquing and generation of healthcare policies. Predictive models can be applied to grapple with some of the most expensive healthcare problems we face, and to develop understandings and policies that enhance care while minimizing costs. We will now highlight several promising opportunities as examples of directions in evidence-based healthcare. In particular, the examples span opportunities for formulating healthcare policies that enable *patient-specific interventions and investments*, enhance the efficacies of managing chronic disease, improve hospital operations and safety, and leverage real-time sensing and prediction in critical care settings.

As a first example, let us consider the prospect of generating evidence-based recommendations for guiding the allocation of limited resources in a patient-specific manner—focusing in particular on the opportunity to *reduce rehospitalization.* Tremendous cost, morbidity, and mortality can come with rehospitalization based on inadequate care and follow-up by patients who have been hospitalized. Rehospitalization rates have come to the attention of policy makers and to the public. The 30-day readmission rates for several diseases are now used as key indicators of hospital quality—and these metrics are now the basis for gating the Annual Payment Update (APU) received by hospitals from Medicare. A recent study of Medicare patients hospitalized in 2004[2] showed that approximately 20% of patients were rehospitalized within 30 days of their discharge from hospitals and that 35% of patients were rehospitalized within 90 days. Rehospitalizations alone are estimated to have cost the nation $17.4 billion during 2004.

Predictive models, learned from large-scale hospital data sets, can be used to identify patients who are at the highest risk of being rehospitalized within a short time after they are discharged. These automated forecasts can serve as the heart of automated decision analyses that are aimed at making decisions about allocating post-discharge support in the most efficacious manner. A recent analysis applied machine learning to a large multi-year data set of patient hospitalizations in the Greater Washington, DC, Metropolitan area. The resulting predictive models infer the likelihood of rehospitalization within 30 days (or other horizons) and these inferences can be coupled with automated decision analyses to weigh the costs and benefits of intervening with special post-discharge support programs. Moreover, by using the tools to ideally allocate

---

[2] S.F. Jencks, M.V. Williams, and E.A. Coleman. *Rehospitalizations among patients in the Medicare fee-for-service program.* The New England Journal of Medicine, 360(14):1418–1428, 2009.

resources for post-discharge support programs to a cohort of patients, we can minimize re-admissions, saving on costs and enhancing the quality of care.[3] Such analyses can be operationalized and integrated as a critical part of the ongoing learning, insight generation, and decision making within health centers—and methods, results, and insights can be shared among hospitals.

As another opportunity area, data-centric analyses and decision models can also play a major role in reducing costs and enhancing the quality of care for the difficult and ongoing challenge of *managing chronic disorders* like congestive heart failure (CHF). CHF is debilitating, prevalent, and expensive. The disease affects nearly 10% of people over 65 years and has nearly a 10% annual mortality rate. Medical costs and hospitalizations for CHF are approximately $35 billion per year in the U.S. alone. CHF patients often hover at the edge of physiological stability. Slight mismanagement of salt in their diet, a sudden increase in ingested liquids, or minor forgetfulness with their compliance with their medication schedule can start an expensive and dangerous downward physiologic spiral ending in a "crash" in the emergency department. Such a crash is typically followed by a week stay as an inpatient in a hospital where the patient is "tuned up" and discharged. Data collection from these patients when they are on their own and when they are hospitalized can be used in the development of predictive models that can loft these patients into stable flight and give them the tools to stay in the air. Predictive models can be coupled with decision analyses for allocating resources to education, sensing, and ongoing support. Data-centric models can guide the formulation of outpatient programs and systems, including the application of real-time systems that provide ongoing monitoring and guidance on the management of a chronic disease. Such programs for CHF can include the ongoing use of devices to sense a patient's hydration, e.g., with an apparatus as simple as a scale that is networked via home wifi or GPRS.

As a third example of a high-payoff direction, machine learning and reasoning offer great promise for addressing the difficult challenge of *keeping hospitals safe and efficient*. Consider specifically the case of hospital-acquired infections. These infections affect 10% of people who are hospitalized and constitute the sixth leading cause of death in the U.S. Nearly everyone knows a friend or family member who ended up wrestling with a hospital-acquired infection such as *C. difficile*, MRSA, and VRE. Frequently, the hospital-acquired infection becomes as challenging a health problem as the one that had originally brought the patient to the hospital. On average, hospital-acquired infections lead to an average increase in hospitalization time by 10 days, and an additional cost of $40,000 per patient, or nearly $7 billion annually in the U.S. The CDC has been estimated that 90% of deaths due to hospital-acquired infections can be prevented. A key direction is the application of predictive models and decision analyses to address this opportunity. Recent work has highlighted the potential for using predictive models to better understand patients' risk factors and to guide surveillance and other preventative actions.[4,5] The

---

[3] M. Bayati, M. Braverman, E. Horvitz, *Policies for Minimizing Rehospitalizations for Congestive Heart Failure,* forthcoming.

[4] Ercole FF, et al. *Applicability of the National Nosocomial Infections Surveillance System Risk Index for the Prediction of Surgical site Infections: A Review*. Brax J Infect Dis. 2007 Feb;11(1):134-41.

[5] Nseir, S. *Accuracy of American Thoracic Society/Infectious Diseases Society of America criteria in predicting Infection or Colonization with Multidrug-Resistant Bacteria at Intensive-Care Unit Admission*. Clinical Microbiology and Infection. July 2010.

data to prediction to action pipeline for hospital-acquired infections will likely have to be implemented locally given the variance and dynamics of colonization and risks over locations and time. Beyond the use of predictive models constructed from data for designing surveillance policies and for engaging in such preventative measures as changing a patient's medications, discontinuing a device, moving the patient to another location, or modifying a therapeutic plan, there is potential to employ algorithms to help with discovery, so as to generate insights about the source of infection. Machine learning algorithms can identify sets of evidence about patients, locations, procedures, and healthcare staff that are linked to the risks of acquiring an infection, and then help with moving from suspicious correlation to causation—by working to identify causal factors that raise the risk of acquiring an infection.

As fourth area of opportunity that is ripe for evidence-based methods is the harnessing of *predictive models* constructed from multiple sources of evidence, including patients' physiological data, for forecasting future states of a patient's physiology. Such models will likely have value in systems that have the ability to detect potential future challenges and needs and, coupled with decision models, they can guide alerting and intervention. For example, predictive models may be able to forecast if and when patients will enter unstable or dangerous physiologic regimes. As an example application, data collected from patients in intensive care units (ICU) may one day be used to predict when the patients may be best weaned from respirators, or more generally, when they will be ready to be moved out of the ICU and into more appropriate and less costly settings at the hospital. Likewise, multiple pieces of evidence can be used to predict when a patient on the floor may have to be moved to the ICU. Such reasoning can help with the planning of resources and also assist with recommendations for real-time and proactive interventions that can enhance patient outcomes while minimizing unnecessary expenditures.

These examples are only a few of an array of promising applications and directions in evidence-based healthcare enabled by computational methods, systems, and infrastructure. We will now turn to some specific recommendations for moving forward.

Sensing, Acting, and Interacting in the Course of Daily Life

Moving from hospital- and practice-centric systems to the pockets and homes of the well and ill, there are promising opportunities ahead for collecting health-related data from individuals and populations in the course of daily activities via mobile devices and inexpensive fixed sensors. Mobile devices could be used to collect data from populations of healthy people interested in assisting with studies of preventative healthcare, as well as by cohorts suffering with intermittent and chronic ailments. Data sets collected from mobile devices and inexpensive sensors at home can provide new insights about wellness and disease. Such data sets can also be used to create predictive models and decisions systems that provide new kinds of services—some of which can be targeted for fielding on mobile devices.

Stationary and mobile devices can sense in an ambient manner from onboard sensors (e.g., GPS devices and accelerometers) and specialized off-board sensors (e.g., wearable PaO2 sensors and ECG devices) with wireless connections to the devices. Ambient sensing can provide large case libraries with data acquired through automated observations and via manual entry by their owners about situations and health statuses—in both a push and an active pull manner via

probing device owners in different contexts based on sensed data.

Data sets acquired via such community sensing can provide raw material for machine learning and reasoning about wellness and illness, considering such factors as activities and location. The case libraries can serve as novel sources of discovery and insight, with multiple offline and real-time uses spanning a spectrum of applications. These unprecedented data sources and analyses may one day identify previously unrecognized risk factors for illness (malignancy, coronary artery disease, etc.), provide insights about the multifactorial basis of disease exacerbation (e.g., intermittent vertigo, asthma, CHF, etc.), and be harnessed to recognize the early warning signs of pending illness or risks (e.g., the onset of a cardiac arrhythmia, pulmonary problems, etc.)

A wide array of applications that harness predictive models or the results of data-centric studies can provide users with local decision support. Mobile applications might one day work to stabilize physiology (e.g., in CHF patients), nudge users toward establishing healthful habits, and even help researchers with the design and fielding of snap clinical trials.

Key Directions for Attention and Investment

Several key research and engineering priorities in this space can speed our attaining the vision of true evidence-based healthcare. These priorities include focusing attention and dollars on the science and engineering of data capture, pushing on the theory and practice of machine learning and reasoning, innovating on the hard challenges of data privacy, anonymization, and sharing, studying methods for complementing human expertise in healthcare setting, and for working to stimulate and leverage synergies and collaboration among investigators in the core computer science, biomedical informatics, and medical communities.

Science and Engineering on the Challenges of Data Capture

Methods for the capture and sharing of healthcare data are a critical ingredient needed to move forward on the path to impactful evidence-based healthcare. We are already seeing efforts and incentives that are promoting the implementation of electronic health record (EHR) systems. Without electronic encodings of health information, data-centric models and methods are dead in the water. However, developing and implementing EHR systems are only a first step. There is a critical need for developing and fielding methods for capturing healthcare information in an efficient manner as part of the normal course of the workflow of patient care. Methods such as handwriting and speech recognition applied to medical care will reduce the costs of entry and minimize the data-collection blind spots that come with manual entry. Also, programmatic interfaces that allow the efficient and automated flow of data from laboratories, sensors, and devices into healthcare databases need to be developed to automate the storage of health information. On another front, there is a great deal to learn about methods for ideal sensing within embedded sensors at home and from mobile devices that are carried by consumers and patients. Advances include understanding how to manage preferences about privacy and to actively probe based on current goals of a study.[6]

---

[6] A. Kause, E. Horvitz, A. Kansal, F. Zhao. Toward Community Sensing, *Proceedings of IPSN 2008, International Conference on Information Processing in Sensor Networks*, St. Louis, Missouri, April 2008.

Taking a longer-term view, there are opportunities to build new kinds of rich databases by perform extended sensing and encoding of medical data via automated methods for more deeply comprehending clinical context, procedures, situations, and events. It is likely that multimodal combinations of vision, speech, and natural language processing will be valuable in this pursuit. Research is required in such areas as plan recognition, gesture recognition, language understanding, cognitive modeling, human dialog, and rhetorical structure of content, as well as on the core perceptual challenges of visual and acoustical scene analysis. Long-term research on doing such extended data capture will bolster our ability to optimally deliver evidence-based healthcare. As an example that highlights the potential richness of directions in this realm, surgeons have reported a desire to track the nuances of surgical procedures via a means for automatically capturing and segmenting phases of a surgery so as to learn from a large case library about the best surgical methodology (e.g., to minimize bleeding and other risks) for different patients.

Studies in the Science and Practice of Machine Learning

There is much to be done in the realm of automated methods for learning and decision making for healthcare. Areas of investment should include directions with advancing theoretical principles and real-world techniques for doing *transfer learning*. Transfer learning refers to methods that ideally harness and adapt models or data collected at a different location or time to a specific new predictive task at hand. Transfer learning can help a hospital to use a data set or predictive model from elsewhere to be used to "boot up" an analysis or recommendation system when only small amounts of local data have been collected. We also need to *extend active learning* for use in complex, real-world settings, endowing systems with the ability to seek the most important missing data for a prediction task.

Moving to the critical topic of *evidence-based discovery*, while great strides have been made with causal inference, we are only at the beginning of an era of understanding how to distinguish *causality* from probabilistic association and to design studies that identify causal influences. We are also early in our understanding about how to infer and pursue hidden variables from data— variables that might provide new insights and dispel erroneously assumed causal influences.

In another area, we have few tools for building predictive models from *multiple non-stationary streams of data over time* that contains important clues at multiple time granularities. The latter will likely be important a broad spectrum of applications extending to predicting which seemingly healthy people are at risk for sudden death to the real-time monitoring of hospitalized patients in support of decisions about moving patients on the ward to the ICU and vice versa.

Multiple current challenges in machine learning are relevant to healthcare and the pursuit of evidence-based healthcare can bring theoretical and practical learning and reasoning challenges into clearer focus.

Privacy and Security for Enabling Data Sharing

Methods and interfaces that allow people and organizations to manage the privacy of data while sharing that data among key stakeholders will be a critical enabler for evidence-based healthcare.

This is a challenge arena that interleaves civil liberties, current legal practices and restrictions (e.g., HIPAA[7]), and directions in technical innovation. Methods need to be developed that legally and technically allow people to manage their preference about privacy and sharing, for protecting privacy, and for enabling people to share data in a comfortable manner with others for their own healthcare and as part of population efforts to promote biomedical science in an altruistic manner. People who would otherwise contribute their data to healthcare studies may often be reasonably concerned with data and computation moves outside of personal computers and devices, outside of familiar organizations and protected networks, and is accessed by unfamiliar organizations in distant locations—and used in ways that they poorly understand.

We do not understand how best to anonymize and encrypt healthcare data for use in data-centric analysis and systems. Continuing to invest in efforts in computer science on security and cryptography—as well as on the human factors of preferences and user interfaces for comprehending and controlling privacy and sharing—will be essential. Several methods are promising and relevant on the broad challenge of accessing and sharing healthcare data for evidence-based healthcare. One promising approach to ensuring the privacy of healthcare data is *differential privacy*. With this method, types of noise are adding to data sets to obscure identities, and the process is designed to limit losses in the accuracy of inferences. In another approach, *privacy enhanced* machine learning techniques allow for the training of models in the absence of a central database. Rather than centralize the data, these methods distribute the computational procedures of machine learning by sharing encrypted intermediate or partial results; the methods may enable systems that learn from data from multiple centers where the data never has to leave the hospital that collected it. Other approaches focus on enhancing the transparency and controllability of sharing and access, bolstering peoples' ability to make decisions about the data they share. Transparent and efficient databases can enable people to specify how and when their data can be used, including the type and quantity of data that may be collected. Investments in research on technical approaches, preferences, and the human factors of privacy and sharing are critical and will likely be necessary for bolstering evidence-based healthcare with large amounts of data while protecting civil liberties.

Complementing Human Expertise

While some methods for evidence-based healthcare generate visualizations, insights, and policies, the methods also lead to systems and services that operate in real time—systems that are designed to work with healthcare practitioners, patients, and other consumers. We have a great deal to learn about the best way to extend and complement healthcare experts with computational support. For example, we have a poor understanding of the best ways to provide awareness in a non-invasive manner to healthcare workers of potentially important evolving, time-critical situations with patients at home or in hospital. We do not know how to ideally remind or how to predict forgetting of important information. We do not know how to detect and counter well-known psychological biases in judgment in decision and decision making that plague experts and non-experts alike. We do not know how to choose the right level of detail to explain a disease process, a care plan, or discharge notes to a patient or to her daughter or father. We also do not know how to best use computing systems to assist healthcare providers with handling the disruption of focus on a patient or situation in rich multitask settings, and how to support the

---

[7] http://www.cms.gov/HIPAAGenInfo/Downloads/HIPAALaw.pdf

ideal recovery and continuity of understanding about a patient's situation—critical for baton passing with changes of teams overseeing patients, as well as for ideal rehydration of understanding about a patient between appointments. New studies and methods for performing *complementary computing* will be critical with the use of predictive models and automated reasoning systems in healthcare. Core areas of research include human-computer interaction, visualization, cognitive science, cognitive psychology, and human factors.

Leveraging Skills and Passion across Computer Science, Bioinformatics, and Medicine

While some investigators in the core computer science community have dedicated significant attention to biomedical challenges, there is a great opportunity to seek deeper integration and collaboration between investigators in the mainstream computer science community, researchers in biomedical informatics, and clinicians and other healthcare workers within medical domains. Although a large number of computer scientists have been deeply engaged with challenges in computational biology, there is less engagement with informatics challenges in clinical medicine. Researchers in the bioinformatics community have been focused squarely on opportunities to enhance clinical medicine, often working closely with medical domain experts.

Indeed, a surprising separation has evolved between the core computer science and bioinformatics communities in the realm of clinical medicine. People in the two communities often have different "taste" in problems, largely attend different meetings, and largely publish in different journals. There is tremendous potential upside to bringing these two communities more closely together, along with experts from medical specialties, via active attempts to build integrated teams.

The Need for Federal Investment

Federal programs have fueled remarkable studies of machine intelligence over several decades—including efforts motivated by and focused squarely on healthcare challenges. For example, the NLM has been a key source of innovation in explorations of artificial intelligence in medicine (AIM) over the years, from work in the 1970s on rule-based diagnosis and therapy systems to research in the 1980s that led to an explosion of efforts with the use of probabilistic and decision-theoretic techniques for tackling real-world challenges in healthcare and other domains. However, explicit and implicit borders on topical relevance and appropriateness have limited funding for computing innovations in healthcare.

In the past year, several venues and programs have begun forging new links and opportunities for evidence-based healthcare. For example, an October 2009 workshop on *Discovery and Innovation in Health IT*[8], organized by the CCC and jointly sponsored by the National Science Foundation (NSF), the Office of the National Coordinator for Health Information Technology (ONC), the National Institute of Standards and Technology (NIST), the National Library of Medicine (NLM), the Agency for Healthcare Research and Quality (AHRQ), and the American Medical Informatics Association (AMIA), brought together over 100 computer scientists and bioinformatics researchers to discuss challenges and directions in healthcare broadly. Subsequently, in March 2010, NSF, NLM, and DARPA co-organized a meeting on

---

[8] http://www.cra.org/ccc/healthit.php

Computational Thinking Applied to Big Data in Healthcare, assembling leaders in multiple communities. The discussions at these meetings and several others like them over the past year have focused on basic research questions—including challenges in the area of machine intelligence—that will enable us to harness an increasing wealth of medical data, yielding more reliable, timely, efficient, and effective care.[9] It is critical that Federal agencies continue providing related opportunities for nurturing and sustaining cross-disciplinary healthcare research teams. In addition, following the cultivation of new multi-disciplinary research teams, specific programmatic initiatives are necessary to support the research efforts of these teams.

Earlier this spring, NSF's Directorate for Computer and Information Science and Engineering (CISE) took the lead in issuing a solicitation in the area of *Smart Health and Wellbeing (SHB)*, calling for proposals that seek "improvements in safe, effective, efficient, equitable, and patient-centered health and wellness services through innovations in computer and information science and engineering." This innovative program—nominally funded for FY 2011—is a step forward for health IT research. However, this modest investment should be followed with additional funding that provides multi-year support to large-, medium-, and small-sized cross-disciplinary teams, whether in the form of one program that broadly spans health IT (as presented in [9]), including the evidence-based care work described in this report, or in the form of multiple programs, each with emphases on specific aspects of health IT. But to be most effective, the funding must bridge key stakeholders, including computer scientists, biomedical informaticists, and medical practitioners. Such a multi-disciplinary effort requires collaboration of NSF (including CISE, Engineering (ENG) and Social, Behavioral, and Economic Sciences (SBE) Directorates) and NIH (including NLM and the National Cancer Institute).

---

[9]For more information about the *Discovery and Innovation in Health IT* workshop, including a white paper that describes a broad spectrum of basic research questions in health IT, see: http://cra.org/ccc/docs/init/Information_Technology_Research_Challenges_for_Healthcare.pdf.