Research Challenges in Measuring Data for Population Health to enable Predictive Modeling for Improving Healthcare

Bruce Schatz
Department of Medical Information Science
University of Illinois at Urbana-Champaign
Urbana IL 61801
schatz@illinois.edu

Clay Marsh
Center for Personalized Health Care
Ohio State University Medical Center
Columbus, Ohio

Kevin Patrick
Center for Wireless and Population Health Systems
University of California at San Diego

Santosh Kumar
Department of Computer Science
University of Memphis
Memphis, Tennessee

David Gustafson
Center for Health Enhancement Systems Studies
University of Wisconsin at Madison

Jerry Krishnan
Office of Population Health Sciences
University of Illinois Hospital and Health System, Chicago

Noshir Contractor
School of Communication
Northwestern University
Evanston, Illinois

ABSTRACT

At the core of the healthcare crisis is a fundamental lack of actionable data, needed to stratify individuals within populations, to predict which persons have which outcomes. A new health system with better health management will require better health measurement, to improve cost and quality. It is now possible to use new technologies to provide the rich datasets necessary for adequate health measurement, which enables new information systems for new health systems.

This report is a summary of a workshop on *Measuring Data for Population Health*, sponsored by the NSF SmartHealth program with assistance from the NIH mHealth initiative, held on January 12-13, 2012 in Washington DC. There were 42 attendees, including invited researchers from academia, government and industry, plus program officers from NSF and NIH.

The workshop had background talks by leaders in health systems and information systems, followed by breakout discussions on future challenges and opportunities in measuring and managing population health. This report describes the observations on what problems of health systems should be addressed and what solutions of information systems should be developed. The recommendations cover how new information technologies can enable new health systems, with support from future initiatives of federal programs.

The workshop and its report identify research challenges that utilize new computing and information technologies to enable better measurement and management for practical healthcare. The measurement technologies focus on deeper monitoring of broader populations. The management technologies focus on utilizing new personal health records to provide personalized treatment guidelines, specialized for each population cohort. This would enable predictive modeling for health systems to support viable healthcare at acceptable cost and quality.

A workshop website contains background and discussion notes: https://wiki.engr.illinois.edu/display/hiworkshop/NSF+Workshop+Population+Health

Keywords

health systems, information systems, health determinants, measuring health, population health, mobile devices, smart phones

1. INTRODUCTION

Healthcare is the economic crisis of our time, there is no viable infrastructure for health systems. Simultaneously, there is greater demand for healthcare due to the aging population demographics and greater discontinuity for infrastructure due to a shift from acute care to chronic care. At the core is a fundamental lack of actionable data, it is not possible today to stratify individuals to predict which persons have which outcomes within populations.

A new health system with better health management will require better health measurement, enabling higher quality at lower cost. New technologies can now provide the rich datasets necessary for adequate health measurement, which can enable predictive modeling for improving practical healthcare.

Currently, there are Internet services that daily service hundreds of millions of persons, such as Google for information search or FaceBook for social networks. Billions of persons worldwide have mobile devices, such as cell phones and music players, which contain measurement sensors connected to networks. These devices could enable revolutionary new levels of status data for population health – more features for more persons. New information systems could then provide adequate measurement to support new health systems with viable management.

Ironically, current technologies for health measurement are decades old. For example, the BRFSS (Behavioral Risk Factor Surveillance Survey) is a telephone survey developed by the CDC, asking 100 health questions to 350K persons per year, demographically representative of the population. Paper Quality of Life questionnaires, such as the SF-36 developed by the Medical Outcomes Study in the 1980s, still dominate the measurement of population health.

The near future of health measurement is in commercial devices, such as smart phones, or emerging technologies, such as smart clothes and smart homes. It will soon be feasible to measure every person every day across the entire spectrum of health activities, recording hundreds of thousands of features instead of hundreds across populations covering hundreds of millions of individuals instead of hundreds of thousands.

Realizing this goal requires enabling technologies for supporting the rich data feeds, but in an actionable way. The data collected must correspond with feasible treatments for human health and the data analyzed must deal with haphazard collection of uncertain quality. There are great challenges in computer and information science to be met in filtering noisy data from sensors and in mining diverse information from monitors.

2. HEALTH DETERMINANTS

The determinants of health span from the bodies of individuals to the societies of populations. Such determinants can be summarized by a series of concentric rings. For example, consider Figure 1, which is the framing diagram from the recent book by Bruce Schatz, PhD, and Richard Berlin, MD, entitled *Healthcare Infrastructure: Health Systems for Individuals and Populations* (Springer series in Health Informatics, 2011). This ring diagram is evolved from that in the classic government study sponsored by the Institute of Medicine, *The Future of the Public's Health in the 21st Century* (National Academies Press, 2003).

In measuring health, it is important to include factors ranging from the bodies of individuals to the societies of populations. These include internal functioning {Inner Rings} considered by personal medicine, such as blood and breath (Biology), and metabolism and motion (Body). But it is equally important to include external functioning {Outer Rings} considered by public health, such as Social Networks (families and communities) and Societal Conditions (living and working). These have strong effects, slower in progress but greater in impact, shown as Figure 2.

In-between internal body and external society is stress and behavior of daily life {Middle Ring}, such as diet and exercise whose effects integrate genes and environment parts into a whole.

Enabling technologies to support this full spectrum exist in research prototypes today. But these systems fail to be scalable for whole populations, remaining at the bench rather than at the bedside. On inner rings, data sensors can simulate pulse oximeters for blood and breath, as well as measure gait analyzers for metabolism and mobility. On outer rings, data sensors can locate social interactions while text parsers can extract personal narratives for societal conditions. In the middle, combinations of data and text in mobile devices can effectively record and analyze the diet and exercise from daily lifestyles of heterogeneous persons in heterogeneous environments.

Figure 1. Health Determinants for Rings of Human Status.

5 *Societal* conditions for Living and Working
4 *Social* networks for Family and Community
3 *Stress* and *Behavior* (smoking & alcohol, diet & exercise)
2 *Body* for Physical (motion) and Mental (memory)
1 *Biology* for Blood and Breath (acute rather than chronic)

Figure 2. Ring numbers reflect influence power and speed.

Practical systems measuring real populations have yet to be tested at this breadth and depth for data capture and analysis. For acute care, continuous monitors are the norm in the hospital, but for chronic care, continuous monitors are uncommon in homes. There are substantial datasets beginning to be recorded within Inner Rings, such as personal genomes, and within Outer Rings, such as risk factors. But such datasets are static, measured perhaps once a year, rather than dynamic, measured perhaps once a minute. Chronic conditions change dynamically over time. Such measurement technologies need to be extended from small research prototypes to large population testbeds, to enable predictive modeling for improving health systems.

3. WORKSHOP ORGANIZATION

The agenda schedule and attendee list are available on the workshop website, along with slides and discussion summaries: https://wiki.engr.illinois.edu/display/hiworkshop/NSF+Workshop+Population+Health .

As summarized next, there were a series of background talks on health systems and information systems. These identified the problem and the solution under discussion. Then there were breakout sessions, broken into three groups along the health rings: Inner – Middle – Outer. Finally, all groups met as a whole to discuss overall goals and workshop recommendations.

Common themes from the breakout sessions are discussed below, followed by research challenges and workshop recommendations. The themes and the challenges represent suggestions for new problems and approaches to the research community. The recommendations represent suggestions to funding agencies, many of whose program officers participated in the workshop, along with key members of the scientific community.

4. THE PROBLEM IS NEW HEALTH SYSTEMS

The background talks reinforced the messages that all rings are important determinants of health quality and cost. Three medical research leaders spoke, who are involved in measuring outcomes from health systems spread across the country.

Brent James from Intermountain Healthcare based in Utah stated that causes of healthcare are Behavior 40% (middle), Genetics 30% (inner), Environment (20%) (outer), Healthcare (10%). He gave detailed examples of how increasing quality and decreasing cost in his health system was due to carefully measuring the outcomes of different procedures in order to manage different cohorts with different treatments.

Clay Marsh from Ohio State University Medical Center further confirmed this by stating that personalized medicine was a combination of Genes x Environment x Behavior. He then explained that the complex problems required complex solutions, involving networks, ecosystems, and modules. That stratification of populations into different cohorts with different treatments was necessary to provide effective and efficient healthcare.

Jerry Krishnan from University of Illinois Hospital and Health System further confirmed this by stating the comparative effectiveness of medical treatments was specialized to demographics. That is, the best treatments or best care for a particular condition depend on the population cohort, and large numbers are necessary to properly account for individual variation within populations.

Discussion following the health system background talks centered on how to improve healthcare with better predictive modeling. This might be termed the "3M" strategy for population health: Monitor – Measure – Manage. In order to *manage* populations to decrease cost and increase quality, it is necessary to *measure* populations across all the rings to determine health status. To *measure* population health deeply, it is necessary to *monitor* individuals on a continuous basis.

An expansion of the 3M Strategy, which came up subsequently in the discussions, was along the lines of Who, What, When, Where, How. These enable evaluation of potential scale of technology solutions. For example, Who: Persons, What: Features, When: Continuous, Where: DailyLife, How: MobileDevices.

5. THE SOLUTION IS NEW INFORMATION SYSTEMS

The background talks reinforced the messages that new technologies are able to support continuous measurement across all the rings. Three informatics research leaders spoke, who are involved in developing research prototypes of continuous monitors across whole populations.

Bruce Schatz from University of Illinois at Urbana-Champaign described the state of population monitoring from its present physical questionnaires to its future digital sensors. He emphasized that there were already wearable fitness devices that could measure features across all the rings, and that research experiments were already proceeding to support these measures with mass mobile devices for continuous vital signs.

Kevin Patrick from University of California at San Diego then discussed measured relationships of internal rings, such as heart rate, to external rings, such as environmental location. This research was deployed to tens of persons, as supported by the NIH Genes and Environment Initiative (GEI). He found strong correlation between genes and environment, for example, time outdoors in green places correlated with better metabolic rates.

Santosh Kumar from University of Memphis, also supported by the NIH GEI, discussed similar field trials with tens of persons. He used custom wearable devices to measure stress in daily life. He found that signatures of measures such as heart rate and respiration pattern were accurate in predicting amount of stress. He also discussed the work of Sandy Pentland from MIT, originally scheduled to present, which progressed in measuring social interaction from custom devices to smart phones, utilizing whatever sensors were available on these mass devices.

Discussion following the information system background talks centered on how to support better predictive modeling. In particular, strategies were discussed to increase the numbers for population trials with continuous monitoring of health status. Current research uses custom devices to measure populations on tens of persons. Such scientific studies are limited in scope by the manufacturing and training to support the custom devices.

The sweet spot in the near future appears to be smart phones, whose sensors can approximate many of the measures needed, including inner rings using the camera and microphone for heart rate and respiration pattern, outer rings using the GPS for location stress and social interaction, middle ring using the camera and accelerometer for diet and exercise. It is feasible now to develop and deploy research testbeds to enable larger scale field trials, enabling much greater measurement into feature variation of health status. Utilizing mass mobile devices could enable development of research testbeds, supporting deployment for thousands of persons. Building the testbeds is more engineering than science, but will enable new science at new scale.

The scale of Information Systems to Measure and Manage Health Systems can be calibrated along a number of different dimensions. Relevant parameters for populations include more than simply the number of persons. That is, the number of persons is important, but equally important for research testbeds is the number of health features, which can be effectively measured by inexpensive scalable devices, such as smart phones and music players. The requirements of running continuously for long durations in the home field of daily life also place stringent constraints on device reliability and energy consumption.

6. COMMON THEMES ACROSS GROUPS

The group discussions compared and contrasted major themes.

As might be expected, each group emphasized that establishing a real-world laboratory is essential to progress into population health. That is, finding a willing cohort that would be instrumented and recorded continuously and longitudinally. The basic problem is that little is known about the variation across individuals in daily life. Since feature baselines are not known, monitors cannot detect health problems. Datasets of normal for demographic groups are needed, to distinguish abnormal deviations from data artifacts. That is, it is essential to establish an observatory testbed to baseline what is signal and what is noise in continuous measures of daily life.

Most discussions focused upon using existing devices as being immediately available to deploy. Thus far, most research has used specialized devices such as arm bands and chest straps monitoring heart rate and breath rate, or GPS locators carried by the person. But there are now enough sensors in smart phones to accurately monitor many of the measures necessary to capture daily life across the health rings. Alternatively, when the persons are within their homes, environmental sensors can record many measures using infrared rays and directional mikes.

Since the measures are thus limited in the short term by what sensors are widely available, once population baselines have been established, there will be much research needed in designing comfortable sensors with a wider range of important measures. That is, once the population testbeds are established, the baselines can be used to compute deviation from norms using better hardware such as wearable vests with sensor grids, or better software for outlier detection against demographic norms.

The existing datasets to leverage vary depending on the health rings. Biology data (Ring 1) is becoming available, but from static genomes with little corresponding environmental context, rather than from dynamic proteomes that profile daily life. Within inner rings, there is extrapolated data from acute care medical records. The Middle Ring is the domain of behavior, which is recorded on national scale by annual processes such as the CDC BRFSS (Behavioral Risk Factor Surveillance Survey). The Outer Rings correspondingly have existing datasets, maintained by CDC and similar EU organizations, relating to aggregate social and societal interactions across national populations.

Generally, the further the ring, the more variation there is across human populations. Acute care has been successful, primarily due to small variation across individuals. So an inner ring scientific experiment to monitor heart rate (Ring 1) or gait pattern (Ring 2) need only measure tens of persons, such as the 60 person demographic sample described in the background talks.

Measuring the middle ring effectively will require a larger scale field trial. The studies measuring stress at the monitor level of 60 persons have used artificial stimuli to create stress. Measuring stress in daily life will likely have completely different continuous curves and significantly larger individual variation. Setting up an observatory testbed, at the scale of present clinical trials, would measure tens of features across thousands of persons. For example, measure the major vital signs and their variants across 2500 persons with different demographics and different lifestyles. This might begin to capture the personal health records with enough variation to predict different responses of different cohorts to different diet and exercise regimes, as an obesity example.

Health Systems can be subsequently improved by using the testbed measures on larger populations coupled with effective management strategies for cohort stratification.

7. RESEARCH CHALLENGES

The workshop and its report identify research challenges and opportunities utilizing new computing and information technologies to enable better measurement and management for practical healthcare. This would enable predictive modeling for health systems, such as:

- What data should be recorded in what standard format to measure daily health?
    - For example, blood pressure or heart rate to diet/exercise/stress to social interaction across the rings.
- How can missing data be best handled, including interpolation and projection?
    - For example, during recording of continuous curve, when is it data signal and when is it noise.
- How should sensor data be classified into actionable data?
    - For example, categorizing temporal patterns in body monitors, or spatial patterns in social monitors, while identifying outliers against normal baselines.
- How should this data be compared and correlated?
    - For example, statistical methods such as data mining patterns or information retrieval clusters, along with customized visualization to improve comprehension.
- How should the diverse sources be judged for quality?
    - For example, evaluating provenance for physician supplied text versus patient supplied text, or for continuous sensor data versus episodic survey data.
- How can population data be transformed into usable knowledge?
    - For example, computer generated classification may differ from human generated classification, requiring revisions to guidelines for diagnosis and for treatment.
- How should the new population data be used for practical health systems?
    - For example, outcomes can be generated from patient-derived text, then utilized for targeted brochures to advise patients, via personalized knowledge modeling.
- How can multiple knowledge sources be integrated for multiple users?
    - For example, different sources have different uncertainties and different users have different preferences. Inference systems must be developed to incorporate data provenance and user beliefs, to adequately capture context.
- How do individuals reconcile new data/information received from different sources?
    - For example, reference websites (e.g., CDC) vs. social networks (e.g., Facebook) vs. other social references (e.g., Wikipedia). Deeper understanding is needed of who/what is best at influencing behavior and inducing change.
- How can existing medical data be leveraged to support customized decision making?
    - For example, proactive detection and remediation of specific medical conditions may require inferences from physician records and patient sensors (food intake and gait speed) using clinical knowledge for optimal interventions.
- What is the impact of new data on health cost and quality?
    - For example, different cohorts have different outcomes, which likely cut across demographic boundaries in different ways than at present, requiring more complex information and intelligent systems for new personal health records.
- What scale of research testbed is needed to produce technology for national deployment?
    - Moving up from a research prototype to an observatory testbed changes scale from 60 to 2500 persons.
    - Moving up another 40 times can measure practical scale for health systems. Human variation across all rings requires 100K persons, at least, with sufficient features on a continuous basis within daily life.

8. WORKSHOP RECOMMENDATIONS

The workshop was intended to make recommendations for future initiatives. These are summarized here, based on the arguments presented during the discussions.

The ultimate goals are to establish the technologies and the processes for a ***Health Moonshot,*** a national scale Framingham 2.0 study with 100K persons. The Moonshot could be targeted towards a major health problem, such as Obesity or Smoking or Aging, or more simply provide the first realistic baseline of human variation across the rings for future studies of population health.

(1) ***Testbeds for Making Baselines.*** [$10M over 5 years]

The primary need in measuring data for population health is for reference datasets, for large testbeds that capture daily life with tens of features for thousands of persons.

The main recommendation is establishing observatory testbeds to record such datasets using different methods with different technologies. Several *National Health Monitoring Observatories* should be funded for creation of datasets unique in scale and duration. Such observatories will likely require public-private partnerships, to capture system dynamics of whole populations, with focus on scalability.

A testbed could simply take demographic samples to measure body functions and environment stress (Ring 2), or target important measures such as diet and exercise (Ring 3) or social and family interactions (Ring 4). Previous initiatives have also emphasized scalable testbeds, such as the NSF/NIH Digital Libraries Initiative.

The resources for a testbed are necessarily larger than prototype projects, since real users over long periods must be supported with new technologies. After the project ends, there must be a *sustainability transfer* into a health system, embedding new information systems into real-world practice.

Based on experiences with previous initiatives, we **recommend projects of at least $10M over 5 years**.

(2) ***Prototypes for Using Baselines.*** [$1.5M over 3 years]

Initial testbeds might focus on existing devices, such as smart phones, as discussed. But a variety of experiments and technologies will become necessary to support long-term research in measuring data for population health.

The Experiments would focus on *Assessing human variability across populations.* For example, verifying commonality of certain measured features across different chronic diseases, or developing computational models for early prediction of disease dynamics.

The Technologies would focus on *Establishing standards and priorities for health monitors.* For example, designing new hardware for inexpensive and comfortable measurement devices, or developing new software for outlier detection of lifestyle risks and disease hotspots. These experiments and technologies must be tested with tens of persons for such prototypes.

We **recommend projects at $1.5M over 3 years**, similar to the current NSF SmartHealth program.

(3) *Education for Human and Data Resources.*

Resources are also needed to insure continuing quality for activities. Support is needed for *Interdisciplinary collaborations* between foundation disciplines such as computer and information technology and also social and behavioral sciences, and between application disciplines, such as population health (public health) and healthcare administration (medicine), including faculty development and student fellowships.

Training is needed for *Resource datasets and analytic tools*, to leverage existing datasets from CDC and from EMRs, plus provide shared vocabularies and ontologies.

Acknowledgement: *The workshop was funded in part by grant 1146740 from the National Science Foundation, via IIS-CISE.*

About the Authors:

Bruce Schatz is Head, Department of Medical Information Science, and Professor, Department of Computer Science, University of Illinois at Urbana-Champaign.

Clay Marsh is Executive Director, Center for Personalized Health Care, Ohio State University Medical Center, Columbus Ohio, and Vice Dean for Research in the College of Medicine.

Kevin Patrick is Director, Center for Wireless and Population Health Systems, and Professor, Department of Family and Preventive Medicine, University of California at San Diego.

Santosh Kumar is Director, Wireless Sensors and Mobile Networks Laboratory, and Faudree Professor, Department of Computer Science, University of Memphis, Tennessee.

David Gustafson is Director, Center for Health Enhancement Systems Studies, and Professor, Industrial Engineering and Preventive Medicine, University of Wisconsin at Madison.

Jerry Krishnan is Associate Vice President for Population Health Sciences, University of Illinois Hospital and Health System, and Professor of Medicine, University of Illinois at Chicago.

Noshir Contractor is White Professor of Behavioral Sciences, School of Engineering and Applied Science, and School of Communication, Northwestern University, Evanston Illinois.

Appendix: WORKSHOP ATTENDEES

Information Systems Speakers

- Health Determinants and New Technologies (Healthcare Infrastructure) Bruce Schatz, UIllinoisUrbanaChampaign
- Measuring Personal Health with Mobile Devices (Genes and Environment) Kevin Patrick, UCaliforniaSanDiego
- Measuring Social Health with Mobile Devices (Stress and Society) Santosh Kumar, UMemphis

Health Systems Speakers

- Measuring Patient Outcomes in Population Health
  Brent James, Director, Health Delivery Research Institute, IntermountainHealthSystem
- Population Stratification for Personalized Medicine
  Clay Marsh, OhioStateUMedicalCenter
- Comparative Effectiveness Research in Population Health
  Jerry Krishnan, UIllinoisHospitalHealthSystem

INNER Rings (biology, body)
Clay Marsh, MD, OhioStateUMedicalCenter
Kevin Patrick, MD, UCSanDiego
Jonathan Silverstein, MD, NorthShoreHealthSystem
Marjory Skubic, PhD, UMissouri
Ida Sim, MD, UCSanFrancisco
Colleen McBride, PhD, NIH NHGRI
Milos Hauskrecht, PhD, UPittsburgh
David Page, PhD, UWisconsin

MIDDLE Ring (behavior, stress)
Brent James, MD, IntermountainHealthSystem
Santosh Kumar, PhD, UMemphis
David Gustafson, PhD, UWisconsin
Elizabeth Mynatt, PhD, GeorgiaTech
Katie Siek, PhD, UColorado
Genevieve Melton-Meaux, MD, UMinnesota
ChengXiang Zhai, PhD, UIllinoisUrbanaChampaign
Andrew Campbell, PhD, Dartmouth

OUTER Rings (social, society)
Jerry Krishnan, MD/PhD, UIllinoisChicago
Noshir Contractor, PhD, NorthwesternU
Colleen McHorney, PhD, Merck Research
Robert Karasek, PhD, UMassachusetts
Christophe Giraud-Carrier, PhD, BYU
Nathan Cobb, MD, Schroeder Institute
Edward Sondik, PhD, CDC, National Center for Health Statistics

NSF: Vasant Honavar, Misha Pavel, Howard Wactlar, Julia Skapik, Will Barkis, Brooke Coley, Famida Chowdhury

NIH: Wendy Nilsen, William Riley, Robert Kaplan, Holly Jimison, Mike Spittel, Alex Blum, Russ Glasgow, Faheen Akbar, Gordon Willis, Shoshana Kahana, Tisha Wiley