Information technology for healthcare transformation

J. P. Bigus
M. Campbell
H. Chang
C.-H. Chen-Ritzo
J. M. Jasinski
R. K. Sorrentino
S. Ebadollahi
D. Gotz
T. W. A. Grandison
P.-Y. S. Hsueh
J. Hu
M. S. Kohn
C. Neti
S. Ramakrishnan
J. Sun
R. Yaesoubi
X. Zhu
B. Carmeli
A. Farkash
H. Nelken
H. Neuvirth
Y. Peres
M. Rosen-Zvi
A. Shabo
M. Cefkin
W. F. Cody
A. Evfimievski
S. Glissmann
D. Gruhl
P. J. Haas
J. H. Kaufman
C. A. Kieliszewski
S. E. Knoop
P. P. Maglio
R. L. Mak
S. Renly
P. Selinger
T. Syeda-Mahmood
W.-C. Tan
M. J. H. Hsiao
Y. Pan
Y. Y. Y. Tao

Rising costs, decreasing quality of care, diminishing productivity, and increasing complexity have all contributed to the present state of the healthcare industry. The interactions between payers (e.g., insurance companies and health plans) and providers (e.g., hospitals and laboratories) are growing and are becoming more complicated. The constant upsurge in and enhanced complexity of diagnostic and treatment information has made the clinical decision-making process more difficult. The legislative environment has created a system rooted in liability-focused practice management. Medical transaction charges are greater than ever. Population-specific financial requirements are increasing the economic burden on the entire system. Medical insurance and identity theft frauds are on the rise. The current lack of comparative cost analytics hampers systematic efficiency. Redundant and unnecessary interventions add to exorbitant medical expenditures. Contemporary payment models are antithetic to outcome-driven medicine. The rate of medical errors and mistakes is high. Slow inefficient processes and the lack of best practice support for care delivery do not create productive settings. Many of these concerns can be solved through information technology. This paper describes IBM Research's approach to solving these issues, i.e., the evidence-based healthcare platform.

IEEE Proof

Introduction

Cost, quality, and productivity are the dominant driving forces behind the transformation of the healthcare domain.

Digital Object Identifier: 10.1147/JRD.2011.2160684

©**Copyright** 2011 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the Journal reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied by any means or distributed royalty free without further permission by computer-based and other information-service systems. Permission to republish any other portion of this paper must be obtained from the Editor.

0018-8646/11/$5.00 © 2011 IBM

Figure 1

Evidence ecosystem.

High charges, disorganized and inefficient delivery systems, high medical and medication error rates, poor communication and care coordination, the lack of information support for effective decision making, and counterproductive payment incentives characterize the present state of healthcare systems [1]. In order to improve this situation, information technology (IT) has to be used to spearhead healthcare transformation, which involves replacing poorly coordinated acute-focused episodic care with coordinated management for preventive, acute, chronic, long-term, and end-of-life care.

Over the last few decades, IBM Research (in China, Israel, Taiwan, and the United States) has been utilizing technology to collect, collate, and provide (hidden or buried) information or evidence to improve system workflows and to provide insight from the application of advanced analytics on clinical, claims, and other health data. These transformative technologies stem from IBM Research's evidence ecosystem vision illustrated in **Figure 1**.

At the highest conceptual level, Figure 1 presents the team's framework for business process modeling and transformation that enables the delivery of consumable evidence at the point of care and the introduction and integration of novel incentive and outcome-based payment models. Providers consume evidence from the evidence aggregators and pay for evidence in order to get better clinical decision intelligence and improved outcomes. Providers get paid by payers for improved outcomes and usage of best practices. Payers consume evidence on best practices and appropriate treatments in order to pay providers based on appropriate use.

Evidence generation requires aggregation of patient records, claims data, wellness data, and randomized clinical ~~trials.~~ The comparative effectiveness engine in Figure 1 compares outcomes among patients with similar disease progressions (i.e., cohorts). The evidence registry stores curated evidence, and the service provision mechanism delivers evidence at the point of care.

The process of delivering evidence for end-user consumption requires clinical decision intelligence enhancements that allow the analysis of current interaction and process, the attachment of supporting provenance information (i.e., evidence source data), and the definition and utilization of mechanisms for presenting the evidence in a highly consumable way that is appropriate to the cognitive capacity of the clinical professional. The monitoring of evidence use helps in the evaluation of the meaningful use of evidence for improving outcomes and lowering costs.

The evidence ecosystem (see Figure 1) builds upon work in medical informatics [2–19], collaborative care [20–27], data mining [28–33], modeling [34–40], software engineering [41–47], usability and user interface design [48–57], and decision theory [58–66]. Of these foundational topics, the areas that seek to address a similar space as the evidence ecosystem are patient-centered medical home (PCMH) [2] and evidence-based medicine [3–13].

PCMH is defined as an approach to providing comprehensive primary care that facilitates partnerships between individual patients and their personal providers and, when appropriate, the patient's family. IBM Research's evidence ecosystem platform is a technical means for instantiating the PCMH vision. An evidence ecosystem is a technology-based solution that provides powerful analytic tools to process data and to create actionable information.

Evidence-based medicine refers to a methodology for applying the best available evidence gained from the scientific method to clinical decision making. Evidence-based medicine attempts to assess the strength of evidence of the risks and benefits of treatments (including lack of treatment) and diagnostic tests. Evidence quality ranges from meta-analyses to systematic reviews of double-blind placebo-controlled ~~randomized clinical trials~~ to conventional wisdom. The scientific concerns related to ~~randomized controlled trials (RCTs)~~ stem from their external validity and their generalizability. Although RCTs are valuable and a critical source of knowledge and internal validity, both of the aforementioned concerns often limit the scope of use of the generated evidence output from RCTs at the point of care. By leveraging the advanced information acquisition and integration of IBM Research's evidence ecosystem, the team increases the relevant evidence available to the practitioner at the point of care.

In the remainder of this paper, we present the details of the evidence ecosystem platform, describe sample results of the platform's deployment, highlight future work, and conclude.

Platform

As aforementioned (see Figure 1), the evidence ecosystem platform has three key pillars, namely, evidence generation (EG), evidence use (EU), and evidence-based incentive payment models (IPMs). EG is responsible for extracting useful evidence from existing data repositories, sanitizing it, transforming it into a form that is amenable to integration (with other pieces of evidence), and processing it for consumption by higher-level services. EU is responsible for the manipulation of EG artifacts into a state and visualization that is appropriate for the end user. IPM is responsible for the evidence-based analysis and update of the financial models that engender the best possible care outcomes. In the following sections, we present each component of the platform in greater detail.

Evidence generation

The EG platform (see **Figure 2**) consolidates heterogeneous clinical data and provides state-of-the-art analytic services that support evidence-based practice, population analytics, best practice guideline support, decision support, and translational medicine. A key difficulty in implementing an EG system stems from working with data that was collected for patient care, but not for research or analysis. Such data is scattered throughout a healthcare organization and spans multiple modalities (or data types), ranging from structured database records to raw text on paper to magnetic resonance images on modern picture archiving and communication systems machines.

Additionally, EG systems must preserve the provenance and track the uncertainty of the data they curate. They need to facilitate proper role-based access to protected health information (PHI) and to support the continuing evolution of data types as new medical testing technologies emerge. The main challenge faced by the team when building the EG system was in creating technology and tooling that meet these requirements and that can reduce the time it takes to transform data into usable evidence in the clinical setting. This task is further complicated by the fact that clinical systems often lacked well-defined application programming interfaces (APIs), that archives had differing uptimes, and that data may be corrupted after years on legacy content management systems. Furthermore, access to data is often gated by the systems' highly distributed nature, the lack of a coherent "single sign-on" within these environments, and the requirement that access does not interfere with the primary use (e.g., clinical patient care) of these systems. These hurdles are further augmented with the facts that data sets are often quite large, e.g., hours of video and audio transcripts and that, in many cases, the time and processing limits (particularly in emergent care situations) require "best effort" optimization of data acquisition rather than a complete draw or ingest. Fortunately, the team was able to leverage and customize its prior work on data acquisition technologies [41–46].

Once data from different systems has been acquired, ~~they need~~ further transformation before it can be used for analytics. The different vocabularies and coding systems

IEEE Proof

Figure 2

Evidence generation platform.

used by different sources must be harmonized so that data is available for analysis in a standard coherent form. Data from multiple sources is combined and reorganized into logical groupings that reflect how the data is to be used, rather than which particular system or device it came from. Additional analytic routines may be run to extract additional information from the source data. For example, entity extractors can tag clinical terms in textual data, and feature extractors can identify structures in images. Since such analyses are often very specific to a particular form of data and new analytics are constantly being devised, the system must allow new analytics to be easily added via a pluggable standardized interface.

Although most analysis routines will consume specific data artifacts and produce new ones, a central point is needed to organize and summarize all of the information about a particular patient into a consistent and coherent longitudinal record. Our approach is to use documents based on the HL7 Clinical Document Architecture (CDA) standard, maintaining links from the summary document to the underlying artifacts used to create it and also preserving provenance chains for the artifacts themselves. This gives investigators the ability to determine the ultimate source of the information they are utilizing. There are numerous challenges in constructing such a summary, including the need to resolve conflicting information from different sources or to represent uncertainty when conflicts cannot be resolved. The summary provides not just the patient's current state but a snapshot of the patient's state at any point in time.

It should be noted that an increase in the use of standards-compliant electronic medical records will make data acquisition for secondary use (and thus evidence generation) easier over time. However, despite evolving standards and coding systems, the need to integrate historical data from legacy systems persists, and the importance of not interfering with the primary use of systems suggests that evidence generation will remain a difficult task for the foreseeable future.

After solving the myriad issues involved with evidence generation, a solid corpus of data exists in the clinical content management database (see Figure 2). At this point, specialized advanced analytics, based on the intended usage, are executed over the data set to create actionable information that will be presented.

Evidence use

EU has two ~~phases, i.e.,~~ advanced analytics and presentation. EU advanced analytics are highly specialized purpose-driven

IEEE Proof

tasks that transform the data received from the EG platform into the consumable units required by the EU presentation component. As such, the artifacts in the leftmost side of the framework in Figure 2 represent end products of this advanced analytics process. To further elucidate the nature of EU advanced analytics, point examples are highlighted in upcoming sections.

The EU presentation phase entails the processes, procedures, and technologies required to ensure that the right information gets to the right person, at the right moment in time, at the right point in the clinician's workflow, and in the right form. This phase is complicated by the fact that EU is governed by constraints on time, collaboration, transparency, and the nature of intended use.

Time—The current modus operandi is that clinicians quickly work in outpatient facilities to see as many patients as possible. Generally, hospital staff members are even more challenged by urgent treatment decisions that must be quickly and accurately made. Thus, user interfaces must be designed to surface summarized patient information and evidence that is most relevant to the current clinical task. The optimization of the EG platform to produce the most current and "best effort" ingests from available systems helps us to meet this constraint. The EU advanced analytics are precisely developed to ensure that dashboarding techniques adapt to the changing condition of a patient, which may cause different data to become relevant over time.

Collaboration—The health of a particular patient is managed by a large team of professionals (e.g., nurses, physicians, and social workers) who must collaboratively work to obtain the best possible outcome for a patient [67]. These professionals typically work asynchronously, seeing patients at different times and often in different facilities. Effective sharing of information across teams is essential, and the presentation layer must have asynchronous collaboration as a fundamental requirement. The design and operation of the platform (in Figure 2) helps ensure that clinicians can collaborate not only over a patient's personal health record but also around the relevant medical evidence used to justify various treatment decisions.

Evidence transparency—It is critical that any displayed evidence provides clear information about the reliability and provenance of the information. Reliability data is garnered from the observed downtimes and ingest rates of the data sources observed by the EG platform, whereas provenance data is acquired from the clinical content management database. Statistical confidence or other reliability data ensures that users of the data understand the weight with which the evidence should influence a decision.

Nature of intended use—Ideally, the evidence displayed to a clinician is exactly what is needed and is highly accurate. In practice, however, expert users such as physicians and nurses will often need to look at evidence from different perspectives before making a decision. This implies that the presentation layer must be able to quickly and smoothly provide different views of the data. Additionally, visualization recommendation techniques [68] may be leveraged to further lighten the cognitive load on healthcare professionals.

As with advanced analytics, the presentation phase is highly dependent on the application being developed, the deployment environment, and the way the practitioners work. Fortunately, active monitoring of how evidence is used is less specialized, and the examination of the patterns of use enables us to determine the interventions that yield desired outcomes, affording us the opportunity to modify the IPMs to better align financial reward with ever-increasing quality of care.

Evidence-based IPMs

Health policymakers (in the United States) are encouraging provider organizations to deliver healthcare services that, among other things, include payments that reward specified improvements in quality and costs [69]. The premise behind this effort is that health recommendations are more likely to be implemented when the clinical evidence supports them and sufficient financial incentives exist for providers to follow them. In line with this, the EE platform (see Figure 1) supports practice environments where providers combine their knowledge of both the clinical evidence and the financial incentives to set treatment strategies for their patients (see Figure 1). Currently, with few exceptions [70], practical guidelines for implementing the necessary provider payment models are unavailable, as are tools for forecasting their impact over time. Although our work on evidence-based IPMs is still in its early stages, upcoming sections outline our initial thoughts on model design and software tooling.

Model design

To model the impact of a given payment incentive for healthcare delivery and health outcomes, we consider four key aspects for analysis, ~~i.e.,~~ contract design, decision making, resource allocation, and patient–provider attribution. In this context, contract design involves determining the essential components of a financial agreement between a payer and a provider that will encourage a provider behavior that is consistent with evidence-based practices. Important dimensions of analysis in contract design include the setting of performance measures, payment structures, and payment rates. More generally, contracts may also be written between payer and patient and between provider organizations and individual providers. Decision making involves the determination of the actions of a decision maker (e.g., payer,

IEEE Proof

provider, and patient) that will result in the optimization of the decision maker's preferences. It provides the logical basis for the payer, provider, and patient responses to a given contract. Resource allocation is an important aspect for analysis when decisions may be influenced by a limitation in critical resources (e.g., time, staff, beds, rooms, equipment, and other medical supplies). *Provider-patient attribution* refers to the problem of assigning accountability for the cost of care and health outcomes of particular patients to particular providers. This assignment is reflected in the contract design since it will affect payments. All four aspects are closely related and contribute to the quality of the analysis of payment incentives.

The use of agency theory for the contract design is well known in economics and business [71, 72]. The general relevance of agency theory in the healthcare industry is well documented [59–61], particularly its specific applicability to physician reimbursement contracts [73–76]. The majority of these efforts make limited, if any, use of empirical data to validate modeling assumptions. A notable exception is the work done by Lee and Zenios [77] who used a data-driven evidence-based approach for designing reimbursement contracts in the context of care delivery for Medicare patients with end-stage kidney disease. This paper attempts to be another validation point for evidence use in incentive and payment model reform.

With regard to decision making, decision theory, which includes expected utility theory [78] and its many generalizations [79] and alternatives [80], allows us to develop mathematical models of the beliefs and preferences of physicians, patients, and payers. Examples of the use of decision theory in medical decision making are provided in [62–64]. We adapt these models by parameterizing them based on an analysis of evidence data collected using survey methods and observational data. Additionally, these preference models may be combined with normative models for determining behavior (e.g., choice of treatment). Methods for formulating and solving normative decision models are drawn from machine learning, control theory, and operations research [81–84]. Dynamic modeling approaches such as system dynamics [85] and stochastic dynamic programming [86] may be used to study the time-dependent effects of alternative payment models.

The attribution of patients to providers, who are accountable for their care, can be done prior to the delivery of care (i.e., prospectively) or after the delivery of care (i.e., retrospectively). If prospectively done, the problem can be formulated using mathematical optimization methods (e.g., a two-stage stochastic decision model [87], where in the first stage, the decision maker selects the attribution pattern, and in the second stage, the decision maker selects the treatment choices). If retrospectively done, statistical models can be used to construct accountability networks of physicians by analyzing observed referral patterns and claims data.

In ongoing client-driven initiatives, these thoughts are currently being tested in the field and have yielded promising initial results. However, acquiring tool support to enable hypothesis testing remains a significant concern.

Software tooling

Studying the dynamics of modifications to one's model is traditionally done using computer-based simulation software, which has typically used discrete-event simulation [65] and agent-based simulation [66, 88, 89] for the healthcare domain. Because a standard healthcare setting is complex, tooling that enables the construction and analysis of multiple factors in an interconnected network of simulation models is required. A significant technical challenge is that of the interoperability of models, which may be implemented using different programming languages and paradigms. This is an important challenge since the management of many chronic diseases such as obesity is multifaceted, involving cultural, social, political, industrial, and infrastructural issues. Models representing different aspects of the community may independently exist and would need to be integrated in order to perform a more complete analysis of the disease. Toward this goal, the Smarter Planet* Platform for Analysis and Simulation of Health framework [90] is being built.

As the EG and EU components of IBM Research's evidence ecosystem are more mature than the IPMs component, we outline the EG and EU deployment results first.

Platform deployment results

From the prior discussions, it should be apparent that IBM Research's evidence ecosystem (see Figure 1) directly and explicitly addresses the problems that have precipitated the need for healthcare to be transformed. It should also be clear that the platform may be (and needs to be) instantiated into point solutions to demonstrate its value via the EU component. It is in its deployment that one recognizes the true potential for the evidence ecosystem to create and deliver new and intriguing health systems.

Here, we present mature instantiations of the evidence ecosystem, where the EG and EU components, along with specialized task-specific advanced analytics, are used to solve difficult problems that were previously thought to be too intractable to be addressed, namely, 1) providing timely and detailed snapshots of patients for cardiologists, 2) enabling the delivery of care in a collaborative environment, 3) reducing the rates of essential hypertension, 4) detecting adverse drug reactions, 5) preventing the spread of epidemics, and 6) helping to determine the best treatment strategies for human immunodeficiency virus (HIV).

IEEE Proof

Cardiac decision support

The team at the IBM Almaden Research Center [91–94] leveraged the evidence ecosystem to acquire, analyze, and correlate information from patients' electronic medical records and from their echocardiogram (i.e., echo) data, along with associated reports, to provide cardiologists with up-to-date and accurate information on the cases in their current roster. The specialized analytics in the EU component examined the echo data to calculate measurements (i.e., ejection fraction and ventricular volume) that were used to estimate heart performance and to find similar patients from a prediagnosed patient population based on raw clinical data of new undiagnosed patients. Collaborative filtering was used to derive the distribution of dominant diagnosis, treatment, and outcomes from the similar patient population. This solution enabled cardiologists to speed up the time taken to get near-complete information on a patient from weeks to hours, which had the positive side effect of reducing time wasted and unnecessary costs expended due to missing (or still being processed) data.

Collaborative care

The IBM Thomas J. Watson Research Center team [95–97] used the ecosystem to leverage the collective wisdom of all the patients in an institution to drive more informed decision making among a collaboration of healthcare practitioners. They developed patient similarity analytics that address the issue of how to identify patients that are clinically similar to an index patient, given longitudinal medical records of both the index patient and other patients in the population, i.e., patients belonging to the same PCMH. The insights drawn from the cohort of similar patients are used to provide support for diagnostic, treatment, and care management, particularly for complex patients with multiple comorbidities who do not fall neatly into categories defined in published clinical practice guidelines. One of the key challenges was the identification of the correct similarity metric to use in order for the results to have a clinical meaning. The team developed machine learning methodologies that are used to automatically adjust the similarity metric and the weighting among different features by leveraging domain knowledge. Client evaluations of the system highlight the fact that the display of cohorts helps to guide a better treatment process and improves quality of care.

The team at the IBM China Research Laboratory [98–102] extended the collaborative care model by first developing technologies that are directed toward patient engagement (called evidence-based patient-centered collaborative care), and then by developing technologies that instantiate both the EE and the "Internet of things" vision to provide personalized health service (called *Connected Health*).

The IBM Taiwan Collaboratory team [103, 104] instantiated the evidence ecosystem in the form of their *wellness cloud*, which enables patients who are normally monitored by sensors to get personalized feedback on their progress in achieving their health goals. The approach has been proven to improve health outcomes and lower costs [105].

Reducing essential hypertension

The team at the IBM Haifa Research Laboratory [106] utilized the evidence platform to help first in the construction of a genetic-epidemiological model for essential hypertension, and its related organ damage, which is known to affect 50% of the population older than 60 years of age, then in the improvement of diagnostic accuracy, and finally in the introduction of new strategies for early detection, prevention, and therapy selection. This paper forms the foundation for more personalized treatment for such diseases.

Adverse drug event detection

The IBM Almaden Research team leveraged the EE to develop technology that mines adverse drug event (ADE) data to discover evidence of new ADEs and drug-to-drug interactions in large clinical databases. The effort addressed the challenges of utilizing temporal relationships between events, measuring and ranking the strength and the confidence of associations between drugs or drug combinations and diagnostic events, and limiting the bias caused by confounding factors such as preexisting medical conditions. Grounded in survival analysis, the team's method defines a statistical model for patient history records, which specifies one or more parameters for every association that is measured. The statistical data model extends the Cox method [32, 33] to simultaneously describe all significant influences of current and prior drug treatments, as well as of prior clinical events, on the hazards of subsequent clinical events at any point of time, as long as they make a specified frequency threshold in the population sample. This paper has significantly mitigated the effects of overfitting and confounding factors, thus producing better predictions.

Preventing the spread of epidemics

The team at the IBM Almaden Research Center [107–109] utilized the evidence ecosystem to allow scientists to reuse data and disease (propagation) models, to rapidly build upon the work of others, and to accelerate the development and testing of new simulation software. This technology was used in April of 2009 when Mexico City recorded a 3-week surge of influenza accounting for 90,000 visits in 220 health units and 20 hospitals. The U.S. Centers for Disease Control and Prevention declared the virus a new H1N1 virus, which meant that no one knew the virus's potential impact. The team collaborated with Mexico City to avert a possible fourth wave of H1N1. Public health officials wanted a quantitative assessment of the policies put into

IEEE Proof

Figure 3

Social distancing policies put in place in Mexico City in the spring of 2009 lowered transmission by approximately 22%. Time units are in days since January 1, 2009.

effect in Mexico City in the spring of 2009. Deidentified H1N1-positive cases from March to October 2009 were imported into the system and analyzed.

In **Figure 3**, the H1N1 incidence predicted by the base model (red) is shown along with the actual incidence data for Mexico City (blue). The team then introduced and tested an extended model against the base model. The extended model was designed to evaluate the possible effect of the social distancing policies put in place. In a blind experiment, an automated experiment was allowed to add a finite time window placed at any start date. It was shown that, within the window, the minimum error corresponded to a reduction in transmission of approximately 22%. The results demonstrated that school closure and other social distancing measures used in Mexico City can be effective in lowering influenza transmission in the midst of a pandemic.

Optimizing HIV therapy strategies

The IBM Haifa Research Laboratory team [110–113] has been working in a consortium with a number of European partners called EuResist on accumulating *in vitro* patients' data and creating a clinical decision support system that is based on the analysis of not only genome data but also clinical and demographic data. A number of mechanisms were used to create an accurate classifier, i.e., a specialized EU advanced analytic, which achieves quality recommendation results.

In order to validate the system's recommendations, a test was conducted where the team set aside treatment outcome and asked physicians and the automated server machine to estimate whether a given treatment is efficient or not [113]. **Table 1** shows the results and highlights the fact that the system almost always outperforms experts and, thus, can be used to provide better care.

All of these instances of the evidence ecosystem provide good starting points in the transformational use of IT to improve the quality of care, reduce medical costs and errors, and decrease the complexity inherent in the medical field. However, there is a lot more work to be done in the future.

Future work

For EG, the next step is the facilitation of the ingest of more complex modalities such as audio and the use of

IEEE Proof

Table 1 Comparative analysis of EuResist to industry experts.

Rater	*Ability to call failure (specificity)*	*Ability to call success (sensitivity)*	*Overall accuracy (1-error_rate)*
EuResist	70.0%	80.0%	0.76
Average expert	62.0%	64.66%	0.636
Expert 1	80.0%	53.3%	0.64
Expert 3	70.0%	53.3%	0.60
Expert 4	50.0%	66.7%	0.60
Expert 5	70.0%	66.7%	0.68
Expert 6	50.0%	73.3%	0.64
Expert 7	30.0%	93.3%	0.68
Expert 8	60.0%	60.0%	0.60
Expert 9	70.0%	80.0%	0.76
Expert 10	80.0%	26.7%	0.48
Expert 11	60.0%	73.3%	0.68

advanced natural language processing techniques to better handle a wider range of free text documents and unstructured data from the Internet, for example, healthcare social networks. For EU, more advanced analytics need to be developed and deployed in order to create critical mass for the system such that there will be a pool of resources available that would allow easy customization to new environments.

For incentive payment modeling, the models must be validated across a wide cross section of settings and feedback provided on the effect of localized evidence. On usability, the utilization of many novel *omic* (as in genomic) tests as outcome indicators has become part of the common practice for diagnosis, prognosis, and treatment [114]. These tests may include 1) the traditional procedure of sequencing a virus and identifying mutations in case of infection, 2) transcriptomic tests for breast cancer, and 3) characterizations of chromosomal aberrations in myeloma cells using comparative genomic hybridization. The interpretation of such *omic* tests requires expertise in epigenetic, evolution, or, more generally, bioinformatics; all are research areas that are not part of the knowledge of common physicians, and thus, decision support systems that integrate these tests will be critical in the future. Our ongoing work in this area involves creating data-driven tools and associated presentation technologies that learn from examples of similar patient cohorts and produce personalized recommendations that are based on the complex relation found between the genomic and clinical data.

The above set of future initiatives is not exhaustive and represents only the preliminary set of work items that the team will focus on.

Conclusion

IT holds the promise of transforming the healthcare industry into an evidence-enhanced practice that enables better analytics and decision support for improved outcomes through the automated derivation of new and nonobvious insights from previously siloed repositories, improved treatment processes through the creation of personalized recommendations that incorporate targeted treatments based on individual characteristics, more complete wellness and outcome management through patient-centered and collaborative care, and improved payment systems that reward practitioners based on performance rather than service volume.

In this paper, we have presented IBM Research's vision of evidence-based healthcare and highlighted how the team of global researchers in China, Israel, Taiwan, and the United States is working to realize this vision, creating new classes of systems, techniques, and tools for care delivery. In the process, IBM Research has demonstrated how IBM's evidence-based ecosystem can be used to reduce transaction costs, improve quality of care, lower the number of medical errors and mistakes, and increase practitioner productivity.

*Trademark, service mark, or registered trademark of International Business Machines Corporation in the United States, other countries, or both.

References

1. R. J. Blendon, C. Schoen, C. DesRoches, R. Osborn, and K. Zapert, "Common concerns amid diverse systems: Health care experiences in five countries," *Health Affairs*, vol. 22, no. 3, pp. 106–121, May/Jun. 2003.

IEEE Proof

2. J. Adams, P. Grundy, M. S. Kohn, and E. L. Mounib, *Patient-Centered Medical Home—What, Why and How?* [Online]. Available: ftp://public.dhe.ibm.com/common/ssi/ecm/en/gbe03207usen/GBE03207USEN.PDF
3. G. N. Fox, "Evidence-based medicine: A new paradigm for the patient," *JAMA*, vol. 269, no. 10, p. 1253, 1993.
4. F. Davidoff, B. Haynes, D. Sackett, and R. Smith, "Evidence based medicine," *BMJ*, vol. 310, no. 6987, pp. 1085–1086, Apr. 1995.
5. C. D. Naylor, "Grey zones of clinical practice: Some limits to evidence-based medicine," *Lancet*, vol. 345, no. 8953, pp. 840–842, Apr. 1995.
6. M. R. Handley and M. E. Stuart, "An evidence-based approach to evaluating and improving clinical practice: Guideline development," *HMO Pract.*, vol. 8, no. 1, pp. 10–19, Mar. 1994.
7. D. L. Sackett and W. M. Rosenberg, "On the need for evidence-based medicine," *Health Econ.*, vol. 4, no. 4, pp. 249–254, 1995.
8. F. Davidoff, K. Case, and P. W. Fried, "Evidence-based medicine: Why all the fuss?" *Ann. Intern. Med.*, vol. 122, no. 9, p. 727, May 1995.
9. A. S. Navathe and P. H. Conway, "Optimizing health information technology's role in enabling comparative effectiveness research," *Amer. J. Manag. Care*, vol. 16, no. 12 Suppl HIT, pp. SP44–SP47, Dec. 2010.
10. A. Lyles, "Comparative effectiveness research: NICE for the NHS, but false starts for the US," *Clin. Ther.*, vol. 30, no. 9, pp. 1702–1703, Sep. 2008.
11. S. D. Horn and J. Gassaway, "Practice-based evidence study design for comparative effectiveness research," *Med. Care*, vol. 45, no. 10 Suppl 2, pp. S50–S57, Oct. 2007.
12. A. D. Naik and L. A. Petersen, "The neglected purpose of comparative-effectiveness research," *N. Engl. J. Med.*, vol. 360, no. 19, pp. 1929–1931, May 2009.
13. H. P. Selker, "Comparative effectiveness research: Medical practice, payments, and politics: The need to retain standards of medical research," *J. Gen. Intern. Med.*, vol. 24, no. 6, pp. 776–778, Jun. 2009.
14. H. De Vries, S. Kremers, T. Smeets, J. Brug, and K. Eijmael, "The effectiveness of tailored feedback and action plans in an intervention addressing multiple health behaviors," *Amer. J. Health Promot.*, vol. 22, no. 6, pp. 417–425, Jul./Aug. 2008.
15. P. Elder, G. Ayala, and S. Harris, "Theories and intervention approaches to health-behavior change in primary care," *Amer. J. Prev. Med.*, vol. 17, no. 4, pp. 275–284, Nov. 1999.
16. N. Janz and M. Becker, "The health belief model: A decade later," *Health Educ. Q.*, vol. 11, no. 1, pp. 1–47, 1984.
17. A. Bandura, *Social Foundations of Thought and Action*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
18. V. Gonzalez, J. Goeppinger, and K. Lorig, "Four psychosocial theories and their application to patient education and clinical practice," *Arthritis Care Res.*, vol. 3, no. 3, pp. 132–143, Sep. 1990.
19. F. Kanfer, "Self-management methods," in *Helping People Change*. New York: Pergamon, 1975, pp. 309–316.
20. C. Silagy and T. Lancaster, "The cochrane collaboration in primary care: An international resource for evidence-based practice of family medicine," *Fam. Med.*, vol. 27, no. 5, pp. 302–305, May 1995.
21. K. Zander, *Managing Outcomes Through Collaborative Care: The Application of Care Mapping and Case Management*. Chicago, IL: Amer. Hospital Pub., 1995.
22. M. L. S. Etheredge, *Collaborative Care: Nursing Case Management*. Chicago, IL: Amer. Hospital Pub., 1989.
23. L. Kimball, "Collaborative care: A quality improvement and cost reduction tool," *J. Healthc. Qual.*, vol. 15, no. 4, pp. 6–9, Jul./Aug. 1993.
24. J. Abbott, A. Young, R. Haxton, and P. Van Dyke, "Collaborative care: A professional model that influences job satisfaction," *Nurs. Econ.*, vol. 12, no. 3, pp. 167–169, May/Jun. 1994.
25. M. Aiken, "Collaborative care in the acute care setting," *Nurse*, vol. 68, no. 3, pp. 10–11, 1993.
26. "Case management thrives in collaborative care model," *Hosp. Case Manag.*, vol. 1, no. 12, pp. 209–220, Dec. 1993. AQ1
27. R. A. Jones and C. W. Mullikin, "Collaborative care: Pathways to quality outcomes," *J. Healthc. Qual.*, vol. 16, no. 4, pp. 10–13, Jul./Aug. 1994.
28. R. Nisbet, J. Elder, J. F. Elder, and G. Miner, *Handbook of Statistical Analysis and Data Mining Applications*. New York: Elsevier, 2009.
29. W. DuMouchel, "Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system," *Amer. Statistician*, vol. 53, no. 3, pp. 177–190, Aug. 1999.
30. J. Almenoff, J. M. Tonning, A. L. Gould, A. Szarfman, M. Hauben, R. Ouellet-Hellstrom, O. Ball, K. Hornbuckle, L. Walsh, C. Yee, S. T. Sacks, N. Yuen, V. Patadia, M. Blum, M. Johnston, C. Gerrits, H. Seifert, and K. LaCroix, "Perspectives on the use of data mining in pharmacovigilance," *Drug Saf.*, vol. 28, no. 11, pp. 981–1007, 2005.
31. M. Lindquist, "VigiBase, the WHO Global ICSR database system: Basic facts," *Drug Inf. J.*, vol. 42, no. 5, pp. 409–419, 2008.
32. D. R. Cox, "Regression models and life-tables," *J. R. Stat. Soc. Ser. B*, vol. 34, no. 2, pp. 187–220, 1972.
33. D. R. Cox and D. Oakes, *Analysis of Survival Data*. London, U.K.: Chapman & Hall, Jun. 1984, 208 pages.
34. G. S. Ginsburg and W. Huntington, *Essentials of Genomic and Personalized Medicine*. New York: Academic, 2009, 800 p.
35. J. Arino and P. van den Driessche, "A multi-city epidemic model," *Math. Population Stud.: Int. J. Math. Demography*, vol. 10, no. 3, pp. 175–193, Jul.–Sep. 2003.
36. V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, and A. Vespignani, "Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions," *PLoS Med.*, vol. 4, no. 1, p. e13, Jan. 2007, DOI:10.1371/journal.pmed.0040013.
37. M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ: Princeton Univ. Press, 2008.
38. F. Carrat, J. Luong, H. Lao, A. Sallé, C. Lajaunie, and H. Wackernagel, "A "small-world-like" model for comparing interventions aimed at preventing and controlling influenza pandemics," *BMC Med.*, vol. 4, p. 26, 2006, DOI:10.1186/1741-7015-4-26.
39. A. Flahault, S. Letrait, P. Blin, J. Hazout, J. Ménarés, and A. J. Valleron, "Modelling the 1985 influenza epidemic in France," *Statist. Med.*, vol. 7, no. 11, pp. 1147–1155, Nov. 1988.
40. I. M. Longini, Jr., "A mathematical model for predicting the geographic spread of new infectious agents," *Math. Biosci.*, vol. 90, pp. 367–383, 1988.
41. V. Bhagwan, T. Grandison, A. Alba, D. Gruhl, and J. Pieper, "MONGOOSE: Monitoring global online opinions via semantic extraction," in *Proc. SQAM Workshop*, Bangalore, India, Sep. 21–25, 2009. [Online]. Available: http://www.almaden.ibm.com/cs/projects/iis/hdb/Publications/papers/SQAM2009_MONGOOSE.pdf
42. V. Bhagwan, T. Grandison, and D. Gruhl, "Ingest and integration of medical data in a world with very little DICOM," in *Proc. 13th World Congr. MEDINFO*, Cape Town, South Africa, Sep. 12–15, 2010.
43. T. W. A. Grandison, V. Bhgwan, and D. Gruhl, "Simulating collaboration from multiple, potentially non-collaborative healthcare systems to create a single view of a patient," in *Proc. 6th Int. Conf. Collaborative Comput.: Netw., Appl. Worksharing*, Chicago, IL, Oct. 9–12, 2010, pp. 1–6.
44. A. Alba, V. Bhagwan, J. Grace, D. Gruhl, K. Haas, M. Nagarajan, J. Pieper, C. Robson, and N. Sahoo, "Applications of voting theory to information mashups," in *Proc. ICSC*, 2008, pp. 10–17.
45. D. M. Russell, A. Dieberger, V. Bhagwan, and D. Gruhl, "Large visualizations for system monitoring of complex, heterogeneous systems," in *Proc. INTERACT*, 2005, pp. 938–941.

IEEE Proof

46. D. Gruhl, L. Chavet, D. Gibson, J. Meyer, P. Pattanayak, A. Tomkins, and J. Y. Zien, "How to build a WebFountain: An architecture for very large-scale text analytics," *IBM Syst. J.*, vol. 43, no. 1, pp. 64–77, Jan. 2004.
47. J. P. Bigus, D. A. Schlosnagle, J. R. Pilgrim, W. N. Mills, III, and Y. Diao, "ABLE: A toolkit for building multiagent autonomic systems," *IBM Syst. J.*, vol. 41, no. 3, pp. 350–371, Jul. 2002.
48. J. Silva, N. Seybold, and M. Ball, "Usable health IT for physicians. Smart point-of-care and comparative effectiveness research form the true basis for meaningful use," *Healthc. Inform.*, vol. 27, no. 7, pp. 40–43, Jul. 2010.
49. A. F. Rose, J. L. Schnipper, E. R. Park, E. G. Poon, Q. Li, and B. Middleton, "Using qualitative studies to improve the usability of an EMR," *J. Biomed. Inform.*, vol. 38, no. 1, pp. 51–60, Feb. 2005.
50. S. Rinkus, M. Walji, K. A. Johnson-Throop, J. T. Malin, J. P. Turley, J. W. Smith, and J. Zhang, "Human-centered design of a distributed knowledge management system," *J. Biomed. Inform.*, vol. 38, no. 1, pp. 4–17, Feb. 2005.
51. B. T. Karsh, "Beyond usability: Designing effective technology implementation systems to promote patient safety," *Qual. Saf. Health Care*, vol. 13, no. 5, pp. 388–394, Oct. 2004.
52. K. Zheng, R. Padman, and M. P. Johnson, "User interface optimization for an electronic medical record system," *Stud. Health Technol. Inform.*, vol. 129, pt. 2, pp. 1058–1062, 2007.
53. J. B. Starren, P. R. Payne, and D. R. Kaufman, "Human computer interaction issues in clinical trials management systems," in *Proc. AMIA Annu. Symp.*, 2006, p. 1109.
54. A. Kushniruk, E. Borycki, S. Kuwata, and J. Kannry, "Predicting changes in workflow resulting from healthcare information systems: Ensuring the safety of healthcare," *Healthc. Q.*, vol. 9, pp. 114–118, Oct. 2006.
55. R. Blaser, M. Schnabel, C. Biber, M. Bäumlein, O. Heger, M. Beyer, E. Opitz, M. Lenz, and K. A. Kuhn, "Improving pathway compliance and clinician performance by using information technology," *Int. J. Med. Inform.*, vol. 76, no. 2/3, pp. 151–156, Feb./Mar. 2007.
56. Y. Senathirajah and S. Bakken, "Architectural and usability considerations in the development of a Web 2.0-based EHR," *Stud. Health Technol. Inform.*, vol. 143, pp. 315–321, 2009.
57. K. C. Murphy, "Assessing software impact on clinical workflow and resource utilization," *Stud. Health Technol. Inform.*, vol. 143, pp. 309–314, 2009.
58. C. Chaix-Couturier, I. Durand-Zaleski, D. Jolly, and P. Durieux, "Effects of financial incentives on medical practice: Results from a systematic review of the literature and methodological issues," *Int. J. Qual. Health Care*, vol. 12, no. 2, pp. 133–142, Apr. 2000.
59. D. Dranove and W. D. White, "Agency theory: New insights into the health care industry," *Inquiry*, vol. 24, no. 2, pp. 405–415, 1987.
60. D. Dranove and W. D. White, "Agency theory: New insights into the health care industry," *J. Med. Pract. Manage.*, vol. 4, no. 3, pp. 165–169, 1989.
61. R. A. McLean, "Agency costs and complex contracts in health care organizations," *Health Care Manage. Rev.*, vol. 14, no. 1, pp. 65–71, 1989.
62. G. Parmigiani, *Modeling in Medical Decision Making: A Bayesian Approach*. Chichester, U.K.: Wiley, 2002.
63. G. B. Chapman and F. A. Sonnenberg, *Decision Making in Health Care: Theory, Psychology, and Applications*. New York: Cambridge Univ. Press, 2003.
64. E. Karni, "A theory of medical decision making under uncertainty," *J. Risk Uncertainty*, vol. 39, no. 1, pp. 1–16, Aug. 2009.
65. S. H. Jacobson, S. N. Hall, and J. R. Swisher, "Discrete-event simulation of health care systems," in *Patient Flow: Reducing Delay in Healthcare Delivery*, F. S. Hillier and R. W. Hall, Eds. New York: Springer-Verlag, 2006, pp. 211–252.
66. J. H. Christiansen and A. P. Campbell, "HealthSim: An agent-based model of simulating health care delivery," in *Proc. Int. Microsimul. Conf. Population, Ageing Health: Modelling Our Future*, Canberra, Australia, 2003.
67. J. G. Baggs, M. H. Schmitt, A. I. Mushlin, P. H. Mitchell, D. H. Eldredge, D. Oakes, and A. D. Hutson, "Association between nurse–physician collaboration and patient outcomes in three intensive care units," *Crit. Care Med.*, vol. 27, no. 9, pp. 1991–1998, Sep. 1999.
68. D. Gotz and Z. Wen, "Behavior-driven visualization recommendation," in *Proc. ACM Int. Conf. Intell. User Interfaces*, Sanibel, FL, 2009, pp. 315–324.
69. J. Feder and D. Cutler, "Achieving accountable and affordable care: Key health policy choices to move the health care system forward," Center Amer. Prog., Washington, DC. [Online]. Available: http://www.americanprogress.org/issues/2010/12/pdf/affordablecare.pdf
70. F. de Brantes, M. B. Rosenthal, and M. Painter, "Building a bridge from fragmentation to accountability—The prometheus payment model," *N. Engl. J. Med.*, vol. 361, no. 11, pp. 1033–1036, Sep. 2009.
71. M. Harris and A. Raviv, "Some results on incentive contracts with application to education and employment, health insurance, and law enforcement," *Amer. Econ. Rev.*, vol. 68, no. 1, pp. 20–30, Mar. 1978.
72. J. W. Pratt and R. J. Zeckhauser, "Principals and agents: An overview," in *Principals and Agents: The Structure of Business*, J. W. Pratt and R. J. Zeckhauser, Eds. Boston, MA: Harvard Bus. School Press, 1985.
73. T. G. McGuire, "Physician agency," in *Handbook of Health Economics*, A. J. Culyer and J. P. Newhouse, Eds. Amsterdam, The Netherlands: Elsevier, 2000, pp. 461–536.
74. P. C. Fuloria and S. A. Zenios, "Outcomes-adjusted reimbursement in a health-care delivery system," *Manage. Sci.*, vol. 47, no. 6, pp. 735–751, Jun. 2001.
75. K. Eggleston, "Multitasking and mixed systems for provider payment," *J. Health Econ.*, vol. 24, no. 1, pp. 211–223, Jan. 2005.
76. R. Yaesoubi and S. D. Roberts, *Coordinating Contracts in a Preventive Health Care System With Capacity Constraints*, Edward P. Fitts Dept. Ind. Syst. Eng., North Carolina State Univ., Raleigh, NC, 2011, working paper.
77. D. Lee and S. Zenios, *Evidence-Based Incentive Systems With an Application in Health Care Delivery*, School Manage., Yale Univ., New Haven, CT, 2011, working paper.
78. J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavioral Research*. Cambridge, U.K.: Cambridge Univ. Press, 1947.
79. M. Machina, "Choice under uncertainty: Problems solved and unsolved," *J. Econ. Perspectives*, vol. 1, no. 1, pp. 121–154, 1987.
80. G. Loomes and R. Sugden, "Regret theory: An alternative theory of rational choice under uncertainty," *Econ. J.*, vol. 92, no. 368, pp. 805–824, Dec. 1982.
81. M. L. Littman, "Algorithms for sequential decision making," Ph.D. dissertation, Dept. Comput. Sci., Brown Univ., Providence, RI, Mar. 1996, Tech. Rep. CS-96-09.
82. F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research,* 8th ed. New York: McGraw-Hill, 1990.
83. S. P. Sethi and G. L. Thompson, *Optimal Control Theory: Applications to Management Science and Economics*. Dordrecht, The Netherlands: Kluwer, 2000.
84. M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*. New York: Wiley, 1993.
85. J. Sterman, *Business Dynamics: Systems Thinking and Modeling for a Complex World*. New York: Irwin/McGraw-Hill, 2000.
86. M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: Wiley, 1994.
87. J. Birge and F. Louveaux, *Introduction to Stochastic Programming*. New York: Springer-Verlag, 1997.

IEEE Proof

88. R. Andrews, *The Effect of Pay-for-Performance Design Features on Hospital Quality and the Distribution of Bonuses: An Agent-Based Modeling Approach*, Weill Cornell Medical College, New York, 2010, working paper.
89. M. Charfeddine and B. Montreuil, "Integrated agent-oriented modeling and simulation of population and healthcare delivery network: Application to COPD chronic disease in a Canadian region," in *Proc. Winter Simul. Conf.*, Baltimore, MD, 2010, pp. 2327–2339.
90. M. Cefkin, S. Glissmann, P. Haas, L. Jalali, P. Maglio, P. Selinger, and W. C. Tan, "Splash: A progress report on building a platform for a 360 degree view of health," in *Proc. 5th INFORMS Workshop Data Mining Health Inform.*, Austin, TX, 2010.
91. D. Gruhl, A. Amir, D. Beymer, G. Julia, H. Greenspan, A. Hobbs, K. Pohl, T. Syeda-Mahmood, J. Terdiman, and F. Wang, "AALIM: A cardiac clinical decision support system powered by advanced multi-modal analytics," in *Proc. Int. Med. Inform. Conf. (MEDINFO)*, Cape Town, South Africa, 2010.
92. D. Gruhl, A. Amir, D. Beymer, G. Julia, H. Greenspan, A. Hobbs, K. Pohl, T. Syeda-Mahmood, J. Terdiman, and F. Wang, "AALIM: A cardiac clinical decision support system powered by advanced multi-modal analytics," *Stud. Health Technol. Inform.*, vol. 160, pt. 2, pp. 846–850, 2010.
93. T. Syeda-Mahmood, F. Wang, D. Beymer, A. Amir, M. Richmond, and S. N. Hashmi, "AALIM: Multimodal mining for cardiac decision support," in *Proc. CinC*, Durham, NC, 2007, pp. 209–212.
94. D. Beymer, T. Syeda-Mahmood, A. Amir, F. Wang, and S. Adelman, "Automatic estimation of left ventricular dysfunction from echocardiogram videos," in *Proc. IEEE Comput. Soc. Workshop MMBIA*, 2009, pp. 164–171.
95. S. Ebadollahi, J. Sun, D. Gotz, J. Hu, D. Sow, and C. Neti, "Predicting patient's trajectory of physiological data using temporal trends in similar patients: A system for near-term prognostics," in *Proc. AMIA Annu. Symp.*, Washington, DC, 2010, pp. 192–196.
96. J. Sun, D. Sow, J. Hu, and S. Ebadollahi, "Localized supervised metric learning on temporal physiological data," in *Proc. 12th ICPR*, 2010, pp. 4149–4152.
97. J. Sun, J. Hu, S. Ebadollahi, and D. Sow, "A system for mining temporal physiological data streams for advanced prognostic decision support," in *Proc. ICDM*, 2010, pp. 1061–1066.
98. S. Liu, H. Li, and G. Xie, "A pilot system for regional healthcare collaboration based on open standard," *Chin. J. Health Inform. Manage.*, 2010. ~~AQ2~~
99. S. Liu, H. Li, and G. Xie, "Design and implementation of standard-based electronic health record system," *Chin. J. Health Inform. Manage.*, 2010. ~~AQ3~~
100. H. Chen and G. Xie, "The use of web ontology language and other semantic web tools in drug discovery," *Expert Opin. Drug Discov.*, vol. 5, no. 5, pp. 413–423, May 2010.
101. Y. Ni, G. Xie, L. Zhang, J. Mei, H. Li, S. Liu, H. Liu, and Y. Pan, "Semantic search for clinical evidence using PICO framework," in *Proc. MEDINFO*, Cape Town, South Africa, 2010.
102. H. Liu, J. Mei, G. Xie, Y. Pan, J. Wen, Z. Gao, X. Sun, and X. Chen, "Development of a guideline-based decision support system prototype for collaborative primary healthcare," in *Proc. MEDINFO*, Cape Town, South Africa, 2010.
103. P. Hsueh, R. Lin, J. Hsiao, L. Zeng, S. Ramakrishnan, and H. Chang, "Cloud-based platform for personalization in a wellness management ecosystem: Why, what, and how," in *Proc. IEEE Collaborative Comput.: Netw., Appl. Worksharing*, 2010, pp. 1–8.
104. L. Zeng, P. Hsueh, H. Chang, C. Chung, and R. Huang, "GreenOlive: An open platform for wellness management ecosystem," in *Proc. IEEE/INFORMS Conf. Service Oper. Logist., Inform.*, 2010, pp. 88–93.
105. H. van Keulen, I. Mesters, J. Brug, M. Ausems, M. Campbell, K. Resnicow, P. Zwietering, G. van Breukelen, W. van Mechelen, J. L. Severens, and H. de Vries. (2008, Jun.). Vitalum study design: RCT evaluating the efficacy of tailored print communication and telephone motivational interviewing on multiple health behaviors. *BMC Public Health* [Online]. *8*, p. 216. Available: http://www.ncbi.nlm.nih.gov/pubmed/18565222
106. S. Bianchi, A. Burla, C. Conti, A. Farkash, C. Kent, Y. Maman, and A. Shabo, "Biomedical data integration—Capturing similarities while preserving disparities," in *Proc. Conf. IEEE Eng. Med. Biol. Soc.*, 2009, pp. 4654–4657.
107. J. Kaufman, S. Edlund, and J. Douglas. (2009). Infectious disease modeling: Creating a community to respond to biological threats. *Stat. Commun. Infectious Diseases*~~. article 1.~~ [Online]. *1(1)*. Available: http://www.bepress.com/scid/vol1/iss1/art1/ ~~AQ4~~
108. S. Edlund, M. Bromberg, G. Chodick, J. Douglas, D. Ford, Z. Kaufman, J. Lessler, R. Marom, Y. Mesika, R. Ram, V. Shalev, and J. Kaufman, "A spatiotemporal model for influenza," in *Proc. HIC, Frontiers Health Inform.*, 2009. AQ5
109. J. Lessler, J. H. Kaufman, D. A. Ford, and J. V. Douglas, "The cost of simplifying air travel when modeling disease spread," *PLoS ONE*, vol. 4, no. 2, p. e4 403, 2009, DOI:10.1371/journal.pone.0004403.
110. M. Rosen-Zvi, A. Altmann, M. C. F. Prosperi, E. Aharoni, H. Neuvirth, A. Sönnerborg, E. Schülter, D. Struck, Y. Peres, F. Incardona, R. Kaiser, M. Zazzi, and T. Lengauer, "Selecting anti-HIV therapies based on a variety of genomic and clinical factors," in *Proc. ISMB*, 2008, pp. 399–406. [Online]. Available: http://engine.euresist.org/
111. A. Niculescu-Mizil, C. Perlich, G. Swirszcz, V. Sindhwani, Y. Liu, P. Melville, D. Wang, J. Xiao, J. Hu, M. Singh, W. X. Shang, and Y. F. Zhu, "Winning the KDD cup orange challenge with ensemble selection," *J. Mach. Learn. Res. W&CP*, vol. 7, pp. 23–34, 2009.
112. A. Altmann, M. Rosen-Zvi, M. Prosperi, E. Aharoni, H. Neuvirth, E. Schulter, J. Buch, D. Struck, Y. Peres, and F. Incardona, "Comparison of classifier fusion methods for predicting response to anti HIV-1 therapy," *PLoS One*, vol. 3, no. 10, p. e3 470, 2008, DOI:10.1371/journal.pone.0003470.
113. M. Zazzi, R. Kaiser, A. Sonnerborg, D. Struck, A. Altmann, M. Prosperi, M. Rosen-Zvi, A. Petroczi, Y. Peres, E. Schulter, C. Boucher, F. Brun-Vezinet, R. Harigan, L. Morris, M. Obermeier, C. F. Perno, R. Shafer, A. Vandamme, K. van Laethem, A. Wensing, T. Lengauer, and F. Incardona, "Prediction of response to antiretroviral therapy by human experts and by the EuResist data-driven expert system (the EVE study)," *HIV Med.*, vol. 12, no. 4, pp. 211–218, Apr. 2011.
114. A. C. Culhane and J. Howlin, "Molecular profiling of breast cancer: Transcriptomic studies and beyond," *Cell. Mol. Life Sci.*, vol. 64, no. 24, pp. 3185–3200, Dec. 2007.

Received February 5, 2011; accepted for publication March 8, 2011

Joseph P. Bigus *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (bigus@us.ibm.com)*. Dr. Bigus is a Senior Technical Staff Member at the Thomas J. Watson Research Center. His current research interests include agent-based modeling and simulation and their application to understanding healthcare systems payments and incentives.

Murray Campbell *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (mcam@us.ibm.com)*. Dr. Campbell is a Senior Manager in the Business Analytics and Mathematical Sciences Department at the Thomas J. Watson Research Center. His current research interests include healthcare payment models and data analysis.

Henry Chang *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (hychang@us.ibm.com)*. Dr. Chang is the Collaboratory Director of IBM Research Collaboratory in Taiwan and a Senior Technical Staff Member in the Healthcare Transformation Department at the Thomas J. Watson Research Center.

IEEE Proof

Ching-Hua Chen-Ritzo *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (chenritzo@us.ibm.com)*. Dr. Chen-Ritzo is a Research Staff Member in the Business Analytics and Mathematical Sciences Department at the Thomas J. Watson Research Center. Her research interests include applied mathematical optimization and simulation for applications in services.

Joseph M. Jasinski *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (jmjasin@us.ibm.com)*. Dr. Jasinski is an IBM Distinguished Engineer in the Industry Solutions and Emerging Business Department at the Thomas J. Watson Research Center. He has worked on electronic materials, nanotechnology, computational biology, and health informatics.

Robert K. Sorrentino *IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (sorrentino@us.ibm.com)*. Dr. Sorrentino is a Research Staff Member in the Healthcare Transformation Department at the Thomas J. Watson Research Center. His research interests include advanced healthcare analytics and innovative payment models for healthcare.

Shahram Ebadollahi *IBM Research Division, Thomas J. Watson Research Center, Hawthorne, NY 10532 USA (ebad@us.ibm.com)*. Dr. Ebadollahi is the Manager of Healthcare Transformation Research in the Services Research Department at the Thomas J. Watson Research Center. His research interests include multimedia content analysis and indexing, medical imaging, and the general area of medical informatics.

David Gotz *IBM Research Division, Thomas J. Watson Research Center, Hawthorne, NY ~~10589~~ USA (dgotz@us.ibm.com)*. Dr. Gotz is a Research Staff Member at the Thomas J. Watson Research Center who has explored novel visual analytics, information visualization, and intelligent user interface technologies.

Tyrone W. A. Grandison *IBM Research Division, Thomas J. Watson Research Center, Hawthorne, NY ~~10589~~ USA (tyroneg@us.ibm.com)*. Dr. Grandison is the Program Manager for Core Healthcare Services and a Research Staff Member at the Thomas J. Watson Research Center whose research interests include healthcare security, privacy, and information integration.

Pei-Yun Sabrina Hsueh *IBM Research Division, Thomas J. Watson Research Center, Hawthorne, ~~NY10583~~ USA (pshsueh@us.ibm.com)*. Dr. Hsueh is a member of the Healthcare Transformation Group at the Thomas J. Research Center. Her research focuses on active personalization analytics for collaborative care solutions and cloud-based services.

Jianying Hu *IBM Research Division, Thomas J. Watson Research Center, Hawthorne, NY ~~10589~~ USA (jyhu@us.ibm.com)*. Dr. Hu is a Research Staff Member at the Thomas J. Watson Research Center whose research interests include statistical pattern recognition, machine learning, and signal processing, with applications to healthcare informatics, business analytics, document image analysis, and multimedia content analysis and retrieval.

Martin S. Kohn *IBM Research Division, Thomas J. Watson Research Center, Hawthorne, NY 10532 USA (marty.kohn@us.ibm.com)*. Dr. Kohn, MD, MS, FACEP, CPE, is the Associate Director for Healthcare Analytics for IBM Research who practiced emergency medicine for more than 30 years, including several executive level healthcare positions, before joining IBM in 2005.

Chalapathy Neti *IBM Research Division, Thomas J. Watson Research Center, Hawthorne, NY 10532 USA (cneti@us.ibm.com)*. Dr. Neti is currently the Director and Global Leader of Healthcare Transformation at IBM Research.

Sreeram Ramakrishnan *IBM Research Division, Thomas J. Watson Research Center, Hawthorne, NY 10532 USA (sramakr@us.ibm.com)*. Dr. Ramakrishnan is a Research Staff Member in Healthcare Transformation at the Thomas J. Watson Research Center whose research focuses on adapting models and techniques from the manufacturing domain into the services domain, particularly in healthcare.

Jimeng Sun *IBM Research Division, Thomas J. Watson Research Center, Hawthorne, NY 10532 USA (jimeng@us.ibm.com)*. Dr. Sun is a Research Staff Member at Thomas J. Watson Research Center whose research interests include data mining for healthcare analytics, social network analysis, visual analytics, and data streams.

Reza Yaesoubi *IBM Research Division, Thomas J. Watson Research Center, Hawthorne, NY 10532 USA (yaesoubi@us.ibm.com)*. Dr. Yaesoubi is a post-doctoral Research Staff Member in the Healthcare Transformation Department at the Thomas J. Watson Research Center who works on optimal control of infectious disease control under resource constraints and healthcare payment systems.

Xinxin Zhu *IBM Research Division, Thomas J. Watson Research Center, Hawthorne, NY 10532 USA (zhux@us.ibm.com)*. Dr. Zhu is a Research Staff Member in the Healthcare Transformation Department at the Thomas J. Watson Research Center who has substantial work experience in both private and public sectors of the healthcare industry.

Boaz Carmeli *IBM Research Division, Haifa Research Lab, Haifa University Campus, Mount Carmel, Haifa 31905, Israel (boazc@il.ibm.com)*. Mr. Carmeli is a Research Staff Member and the manager of the Clinical Genomics IT Group at IBM Research, Haifa, whose research interest lies in clinical genomics analytics.

Ariel Farkash *IBM Research Division, Haifa Research Lab, Haifa University Campus, Mount Carmel, Haifa 31905, Israel (arielf@il.ibm.com)*. Mr. Farkash is a Research Staff Member and a member of the Healthcare IT Group at IBM Research Haifa whose expertise lies in semantics and interoperability of clinical, genomic, and environmental data.

Haim Nelken *IBM Research Division, Haifa Research Lab, Haifa University Campus, Mount Carmel, Haifa 31905, Israel (nelken@il.ibm.com)*. Dr. Nelken is the manager of the Integration Technologies team at IBM Research, Haifa, which focuses on document processing and management, healthcare and life sciences, and speech technologies.

Hani Neuvirth *IBM Research Division, Haifa Research Lab, Haifa University Campus, Mount Carmel, Haifa 31905, Israel (hani@il.ibm.com)*. Mrs. Neuvirth is a Research Staff Member at the Machine Learning and Data Mining Group whose specializations include machine learning with applications to bioinformatics.

Yardena Peres *IBM Research Division, Haifa Research Lab, Haifa University Campus, Mount Carmel, Haifa 31905, Israel (peres@il.ibm.com)*. Mrs. Peres is a Research Staff Member and Manager of the IT for Healthcare and Life Science Group in IBM Research, Haifa, whose research efforts include data integration and combining viral genomics with clinical data.

IEEE Proof

Michal Rosen-Zvi *IBM Research Division, Haifa Research Lab, Haifa University Campus, Mount Carmel, Haifa 31905, Israel (rosen@il.ibm.com)*. Dr. Rosen-Zvi is a Research Staff Member in the Data Mining and Machine Learning Group at IBM Research, Haifa.

Amnon (Shvo) Shabo *IBM Research Division, Haifa Research Lab, Haifa University Campus, Mount Carmel, Haifa 31905, Israel (shabo@il.ibm.com)*. Dr. Shabo is a Research Staff Member specializing in health informatics and heads the Healthcare and Life Sciences Standards Program of IBM Corporate who specializes in longitudinal and cross-institutional electronic health records and is a pioneer of the independent health record banks vision.

Melissa Cefkin *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (jmcefkin@us.ibm.com)*. Dr. Cefkin is a Research Staff Member in the Almaden Services Research Group. He focuses on studies of work and organization to analyze complex service systems and the social practices of information use.

William F. Cody *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (cody@almaden.ibm.com)*. Dr. Cody is the Senior Manager of the Healthcare Informatics team in the Department of Computer Science at the IBM Almaden Research Center.

Alexandre Evmievski *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (evfimi@us.ibm.com)*. Dr. Evfimievski is a Research Staff Member in the Computer Science Department at IBM Almaden Research Center whose research interests include data mining and data privacy.

Susanne Glissmann *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (smglissm@us.ibm.com)*. Dr. Glissmann is a Research Staff Member at the IBM Almaden Research Center whose research interest is in ontologies, conceptual models, and architectures in different domains, such as healthcare, business, or IT.

Daniel Gruhl *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (dgruhl@almaden.ibm.com)*. Dr. Gruhl is a Research Staff Member in the Computer Science Department at the Almaden Research Center whose research interests include large-scale text analytics systems.

Peter J. Haas *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (peterh@almaden.ibm.com)*. Dr. Haas received a Ph.D. degree in operations research from Stanford University in 1986 and has been a Research Staff Member at the IBM Almaden Research Center since 1987. His research interests include database sampling, automated relationship discovery in data bases, query optimization methods, autonomic computing, and discrete-event simulation.

James H. Kaufman *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (kaufman@almaden.ibm.com)*. Dr. Kaufman is Manager of the Public Health Research project in the Department of Computer Science at the IBM Almaden Research Center whose current research interests include public health, interoperable health information infrastructure, electronic health records, and epidemiological modeling.

Cheryl A. Kieliszewski *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (cher@us.ibm.com)*. Dr. Kieliszewski is a Research Scientist in the Services Research Department, focused on the human element of service system design.

Sarah E. Knoop *IBM Research Division, Almaden Research Center, San Jose, CA ~~95110~~ USA (seknoop@us.ibm.com)*. Ms. Knoop is the Healthcare Systems Research Manager at the IBM Almaden Research Center. She received a B.S. degree and her focus is on research problems related to healthcare data interoperability, integration, analytics and mining.

Paul P. Maglio *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (pmaglio@us.ibm.com)*. Dr. Maglio is a Research Scientist and Manager in the Service Research Department at IBM Research who has worked on programmable web intermediaries, attentive user interfaces, multimodal human-computer interaction, human aspects of autonomic computing, and service science.

Ronald L. Mak *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (rlmak@us.ibm.com)*. Mr. Mak is a Research Staff Member in the Services Research Department of the IBM Almaden Research Center who works on a project that integrates computer-based simulation models to provide insights in healthcare policy decisions.

Sondra Renly *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (srrenly@tus.ibm.com)*. Ms. Renly is a member of the Almaden Healthcare Research team bringing her clinical experience to improving healthcare standards and interoperability, with an emphasis in toward secondary use cases such as public health.

Pat Selinger *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (patseli@us.ibm.com)*. Dr. Selinger is a world-renowned pioneer in relational database management and inventor of the technique of cost-based optimization that has been adopted by nearly all relational database vendors and is now taught in virtually every university database course.

Tanveer Syeda-Mahmood *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (stf@talmaden.ibm.com)*. Dr. Syeda-Mahmood is a Research Manager in the healthcare informatics at IBM Research, Almaden. Her research interests are in advanced multimodal analytics for healthcare applications.

Wang-Chiew Tan *IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (wangchiew@us.ibm.com)*. Dr. Tan is a Visiting Scientist in the Services Department at IBM Research, Almaden, and an Associate Professor in the Computer Science Department at University of California whose research interests are in database systems, with emphasis on data provenance and information integration.

Mark J. H. Hsiao *IBM Taiwan Collaboratory, IBM Taiwan Corporation, Taipei 11073, Taiwan (mhsiao@tw.ibm.com)*. Mark Hsiao is a Research Scientist in the IBM Taiwan Collaboratory who focuses on the cloud-based analytics services, smart nutrition, data mining, and multimedia information retrieval.

Yue Pan *IBM Research Division, China Research Laboratory, Beijing 100193, China (panyue@cn.ibm.com)*. Dr Pan is a Senior Technical Staff Member and Senior Manager of Information and Knowledge at IBM China Research Laboratory whose research interests include ontology technology and semantic web.

Ying Y. Y. Tao *IBM Research Division, China Research Laboratory, Beijing 100193, China (taoytao@cn.ibm.com)*. Dr. Tao is a Research Staff Member at the China Research Laboratory whose research interests include clinical healthcare solutions.

IEEE Proof

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1 = Please provide complete authors of Ref. [26].

AQ2 = Please provide volume, issue number, and page range of Ref. [98].

AQ3 = Please provide volume, issue number, and page range of Ref. [99].

AQ4 = Please provide page range in Ref. [107].

AQ5 = Please provide page range in Ref. [108].

Notes: 1) Note that references [65] and [88] are the same, [66] and [91] are also the same. Therefore, reference [88] and [91] were deleted from the list. Citations were renumbered accordingly. Please check.

2) Note that Ref. [93] was split into two. Consequently, the reference list and bibliographic citations were renumbered. Please check.

END OF AUTHOR QUERIES

IEEE Proof