

IBM's Health Analytics and Clinical Decision Support

M. S. Kohn¹, J. Sun², S. Knoop³, A. Shabo⁴, B. Carmeli⁵, D. Sow⁶, T. Syed-Mahmood³, W. Rapp⁷

¹ Jointly Health (formerly IBM Research)

² College of Computing, Georgia Institute of Technology, Atlanta, Georgia (formerly IBM Research)

³ IBM Almaden Research Center, San Jose, CA, USA

⁴ Records of Health (formerly IBM Research), Haifa, Israel

⁵ IBM Haifa Research Lab, Haifa, Israel

⁶ IBM Watson Research Center, Yorktown Heights NY, USA

⁷ IBM Watson Solutions Development, Rochester MN, USA

Summary

Objectives: This survey explores the role of big data and health analytics developed by IBM in supporting the transformation of healthcare by augmenting evidence-based decision-making.

Methods: Some problems in healthcare and strategies for change are described. It is argued that change requires better decisions, which, in turn, require better use of the many kinds of healthcare information. Analytic resources that address each of the information challenges are described. Examples of the role of each of the resources are given.

Results: There are powerful analytic tools that utilize the various kinds of big data in healthcare to help clinicians make more personalized, evidenced-based decisions. Such resources can extract relevant information and provide insights that clinicians can use to make evidence-supported decisions. There are early suggestions that these resources have clinical value. As with all analytic tools, they are limited by the amount and quality of data.

Conclusion: Big data is an inevitable part of the future of healthcare. There is a compelling need to manage and use big data to make better decisions to support the transformation of healthcare to the personalized, evidence-supported model of the future. Cognitive computing resources are necessary to manage the challenges in employing big data in healthcare. Such tools have been and are being developed. The analytic resources, themselves, do not drive, but support healthcare transformation.

Keywords

Big Data, healthcare transformation, evidence-supported decision making, healthcare analytics

Yarb Med Inform 2014;154-62

<http://dx.doi.org/10.15265/Y-2014-0002>

Published online August 15, 2014

Introduction

There are many challenges to healthcare. Improving healthcare, both in terms of quality and cost including the reduction of waste, is a global imperative. There are estimates that 21%–47% of what is spent on healthcare in the United States is for interventions of no value [1]. The percentages may be different in other countries, but the problem is a global problem. One concept that can help address current limitations is called personalized or precision healthcare. We know that standard treatments for many diseases are not effective in all patients. Some patients receive no benefit from, and are possibly harmed by, routine interventions. The inability to identify patients who need an alternate treatment accounts for some of the poor clinical and economic results in the current healthcare model. With personalized healthcare, we learn enough about a patient, and relevant healthcare information, to help make choices that are more likely to benefit that patient. For example, if we can predict which diabetic or cancer patient needs a different therapy then we may improve outcomes and save money by not employing an ineffective or potentially dangerous treatment. A concomitant of personalized healthcare is the need to make evidence-supported decisions. We will only transform healthcare if we can effectively use all the information available to us to make better decisions. Although personalized healthcare is often discussed in the context of genomics, the idea is more than 40 years old and much broader than just genomics [2]. Using existing

information, whether it is recorded in a medical record, a research journal, or a gene sequence is part of personalized healthcare.

The volume of healthcare data available is huge, varied, challenging to use, and as a consequence, described as “Big Data”. The data may be unstructured, sometimes called free-text or natural language, as in journal articles, textbooks, guidelines or the narrative parts in electronic health records (EHRs). Additional unstructured data includes stored images, such as x-rays or echocardiograms. There is also structured data, such as numerical entries in EHRs, genomic sequences and streaming data, such as physiologic monitoring in an intensive care unit (ICU). There is so much data, in so many forms, that individuals are only able to use small amounts of the data. One of the most challenging characteristics of big data is variability. Big data is marked by ambiguity, conflict, and inconsistency. Conventional programmatic computing, where a computer is programmed to process a known data set, is not adequate for managing the volume and inconsistency in big data. As much as 80% of the world's data may be uncertain by 2015 [3]. Big data requires cognitive computing, using data-centric, probabilistic approaches to data, where, after a fashion, the computer “thinks.” Based on human reasoning, cognitive computing identifies complex associations, draws inferences, and learns from experience [4]. It is designed to navigate complex, dynamic, uncertain environments [5]. IBM has developed an array of cognitive analytic tools to gain insight from all types of healthcare

information. We divide these resources into two broad categories. The first is knowledge-driven decision support, designed to gather insight from existing vetted knowledge, such as journal articles, textbooks, guidelines, or protocols. The second is data-driven decision support, which looks for patterns in real world, existing data, predominantly structured or image data, but could include text such as the narrative part of EHRs. The two forms of decision support can overlap and augment each other, and some resources will involve a fusion of the two forms, in the effort to provide meaningful insight for decisions at the point of care.

This paper describes the importance of big data in healthcare. The authors seek to present examples of cognitive computing resources developed by IBM that can be used to analyze and draw inferences from the different kinds of data to help achieve evidence-supported decisions. The tools described here address specific areas of the use of data for decision support. They are not all encompassing, but are important in that they address currently existing sources of information applicable to healthcare. As more and different kinds of information become available, new tools will be necessary to incorporate them into decision support. Decision support is only one component of clinical informatics, which is, in turn, one component of the transformation of healthcare. Issues such as healthcare financial coverage, access, workflow, political obstacles, and methods to encourage clinicians and patients to focus on better outcomes also need to be addressed. For example, brain-mapping techniques have the potential to improve diagnosis and management of behavioral and neurologic diseases [6]. Organized medicine in the US has recognized the importance of clinical informatics by creating the medical subspecialty of Clinical Informatics in 2013, under the guidance of the American Board of Preventive Medicine.

These tools are developed for use by clinicians, but ideally in concert with patients. A goal for the future of healthcare is sometimes described as the empowered, knowledgeable patient. It is thought that such a patient, as an active participant in the decision making and care planning process, will also benefit from decision support tools. The patient will, hopefully, become a better manager of his or her health.

Natural Language Processing Analytics for Unstructured Data

Keeping up with the vast amount of literature published each year is a major challenge for clinicians. In the year 2010, the National Library of Medicine in the United States catalogued 699,000 new articles [7]. How much important published information was never used for clinical decisions? Most physicians have fewer than three to five hours a week to read, and usually only read from two or three journals [8]. Before the internet, collecting articles about a subject was a time consuming challenge. A physician had to go to the library and spend hours using an array of indices and bibliographies to find a handful of articles to read to study a subject in any depth. Today the problem is reversed. On-line search engines can overwhelm a reader with thousands of links or articles based on keywords entered. The challenge now is not amassing references, but processing, filtering and analyzing potentially thousands of sources to find the really helpful insights. Watson is a resource that processes those thousands of sources.

Watson was originally developed to prove that a computer could understand natural language, the language of communication, generate and evaluate hypotheses, and adapt and learn with interaction, outcomes and new information. It demonstrated that ability by successfully playing the television quiz game "Jeopardy!" in early 2011. It uses what its inventors call massively parallel probabilistic algorithms, designed to analyze and understand the English language [9]. "Jeopardy!" was intentionally chosen as the arena in which to demonstrate Watson's skill. If Watson could understand the arcane language used for the "Jeopardy!" clues and create an appropriate response then its ability would be clearly demonstrated. Now, Watson for Healthcare is being developed to make it easier for clinicians to use material such as journal articles combined with historical clinical knowledge to achieve evidence-supported decisions. Clinicians are challenged by the overwhelming amount of published healthcare information. Healthcare professionals might like to read and remember

more of the available literature in order to make better decisions. Watson can read and analyze concepts in millions of pages of medical information in seconds, identify information that could be relevant to a decision facing a clinician, and offer options for the decision maker to consider. Thus, Watson will be the physician's assistant to give the advantage of the recall of the information in literature that a provider cannot get herself as well as help in analyzing the literature.

Watson reads and understands concepts in English. Currently, Watson is learning to help oncologists consider therapies for cancer patients. It can improve its performance through machine learning, a process by which Watson teaches itself. Watson is provided information from the patient's EHR. Through training by expert clinicians, it identifies the critical attributes, and can then review relevant literature, including care guidelines. Watson processes all this information and then provides a ranked list of possible therapy options for the oncologist to consider with her patient. Watson is also being taught to consider patient preferences in evaluating options. Watson is not prescriptive. It provides a list of evidence-based hypotheses for the decision maker to consider, not a dictated prescription. It serves as decision support, not as a decision maker.

Watson is still learning and the technology behind Watson continues to evolve at a rapid pace. Architecturally, Watson is a pluggable solution that can be easily expanded with new or updated algorithms, as they are needed. Just like humans use multiple learned techniques to observe the world and solve problems, Watson is a collection of overlapping reasoning algorithms that address specific portions of the pipeline used for problem understanding and problem-solving.

For example, a specific instance of Watson might have specialized algorithms for understanding the question (Natural Language Processing, query expansion with synonyms, dictionaries, ontologies, language translation, speech translation, spelling correction), making hypotheses (indexing a corpus of data, searching for relevant passages, concept annotations, passage expansion, passage filtering, passage scoring), answer selection and scoring (deep parsing,

semantic matching, answer similarity, lexical matching, temporal reasoning, geospatial reasoning, negation, knowledge graphs), machine learning (logistical regression, Bayesian networks, similarity learning), and dialoging with the user (resolving missing and conflicting information, disambiguation, providing suggestions, providing supporting evidence). Some of these algorithms may be generic enough for all applications, and others may be optimized for a specific domain or sub-specialty area.

As Watson is used in more and more engagements, these algorithms improve and their scope broadens, which will allow Watson to be used in a large variety of situations and a broad range of industries. The future is indeed bright for cognitive computing.

Watson accepts “questions” in natural language, so you don’t have to rely on expressing your question in structured data. In some cases, such as the oncology treatment solution, under development, the question is implied: “What are the recommended treatment options for this patient?” The keywords are the facts of the case that are extracted from the patient’s medical records. The clinician using the tool has the option of reviewing these keywords and making last minute changes that may not be reflected in the case – for example, the patient may no longer experience nausea. Armed with this information, Watson can use its trained models to weigh all these facts against numerous treatment options specified by national guidelines, insights from medical experts, and other medical information, and rank those options appropriately. Watson also serves as a discovery tool, by “showing its work”. Watson can show a user the documents (including guidelines, articles, text books, and other knowledge sources) it used to arrive at its hypotheses as well as the key supporting evidence or refuting evidence that was used to rank these hypotheses. Being able to see these details goes a long way toward understanding the rationale used and thus the options presented. Watson also learns and improves through training and repetition from clinician selections and responses. Just as it improved its skill at “Jeopardy!” by getting feedback about the usefulness of the hypotheses, adjusting its

algorithms and rating its sources of information; Watson is similarly improving its ability to identify relevant treatment options for certain types of cancer.

Data-Driven Decision Support

Healthcare systems generate and store huge amounts of data. There is valuable information hidden in that data, hidden in patterns that cannot be readily recognized by the human eye and brain. Analyzing existing data, sometimes described as secondary use or re-use of data, actually creates new evidence. What important insights could be gleaned from the EHRs of a large population of patients that could be used to make better decisions for individual patients? Detecting novel correlations changes from aserendipitous event to organized discovery. What patterns could a computer detect in the massive stream of physiologic data in an ICU that would allow clinicians to identify serious problems earlier when treatment could be simpler and more likely to be successful?

Efforts to develop computerized applications for clinical decision support (CDS) started decades ago, building on rule-based expert systems [10]. These efforts have not been very successful, mainly due to the difficulty in formulating predefined rules that faithfully and completely describe all possible care processes [11]. This task has become even more complex, because of the enormous amount of new health data (e.g., genomics, sensors, imaging), much of it is of unknown significance or sufficiently ambiguous that it could not be incorporated into authoritative clinical practice guidelines [12]. We describe a spectrum of cognitive computing tools to overcome these difficulties.

Patient Similarity Analytics

One of the limitations of published healthcare studies is that they often address one specific condition. Learning about patients with multiple chronic problems in a real world context can be difficult. Using existing data about other patients that are very similar to the patient provides useful information

that can be leveraged for making better decisions. With the tremendous growth of the adoption of EHRs, various sources of information are becoming available about patients. A key challenge is to identify effective secondary uses of EHR data to help improve patient outcome without generating additional burdens on physicians. Patient similarity derives a relationship measure between a pair of patients based on their EHR data. It enables case-based retrieval of patients similar to an index patient, treatment comparison among the cohorts of patients similar to the index patient and cohort comparison and comparative effectiveness research.

Deriving meaningful patient similarity measures requires integrating physician input. We created a suite of approaches to encode physician input as supervised information to guide the development of the similarity measure to address the following questions:

- How to adjust the similarity measure according to physician feedback?
- How to interactively update the existing similarity measure efficiently based on new feedback?
- How to combine different similarity measures from multiple physicians?

First, physician feedback provides locally supervised metric learning (LSML) [13] to define a generalized Mahalanobis measure to adjust the distance measure among patients consistent with the clinical context. We construct two sets of neighborhoods for each training patient based on an initial distance measure. In particular, the homogeneous neighborhood of the index patient is the set of retrieved patients that are close in distance measure to the index patient and are also considered similar by the physician; the heterogeneous neighborhood of the index patient is the set of retrieved patients that are close in distance measure to the index patient but are considered not similar by the physician. Given these two definitions, both homogeneous (containing true positives) and heterogeneous (containing false positives) neighborhoods are constructed for all patients in the training data. Then we formulate an optimization problem that tries to maximize the homogeneous neigh-

borhoods, and at the same time minimizing the heterogeneous neighborhoods.

Second, the interactive Metric learning (“IML”) method incorporates additional feedback that can incrementally adjust the underlying distance metric based on latest supervision information [14] IML is designed to scale linearly with the data set size based on matrix perturbation theory, which allows the derivation of a sound theoretical foundation. Our empirical results demonstrate that IML outperforms the baseline by three orders of magnitude in speed while obtaining comparable accuracy on several benchmark datasets.

Third, to combine multiple similarity measures (one from each physician), we first construct discriminative neighborhoods from each individual metrics, then combine them into a single optimal distance metric. We formulate this problem as a quadratic optimization problem and propose an efficient strategy to find the optimal solution [15]. Besides creating a globally consistent metric, this approach provides an elegant way to share knowledge across multiple experts (physicians) without sharing the underlying data, thus preserving privacy. Through our experiments on real claims datasets, we have shown improvement of classification accuracy as we incorporate feedback from multiple physicians.

All three techniques address different aspects of operationalizing patient similarity in the clinical application. Locally supervised metric learning can be used to define the distance metric in the batch mode, where large amounts of evidence are first obtained to form the training data. The training data should consist of clinical features of patients such as diagnosis, medication, lab results, demographics and vitals, and physician feedback about whether pair of patients are similar or not. For example, one simple type of feedback is binary indicator about each retrieved patient, where 1 means the retrieved patient is similar to the index patient and 0 means (s)he is not similar. Then the supervised similarity metric can be learned over the training data using LSML algorithm. Finally, the learned similarity can be used in various applications for retrieving a cohort of patients similar to a target patient. The other techniques address related challenges of

using a supervised metric, such as updating the learned similar metric with new evidence efficiently, and how to combine multiple physicians' opinions. Obtaining high quality training data is very important but often challenging, since it typically imposes overhead on users, who are busy physicians. These learning techniques have the potential to minimize the physician burden.

We have conducted preliminary evaluation of all the proposed methods using historical claims data consisting of 200,000 patients over three years from group of primary care practices. Heart failure diagnosis codes assigned by physicians are considered as the supervision information, while all other information (e.g., other diagnosis codes) is used as input features. The goal is to learn the similarity that cluster heart failure patients more closely, while pushing other patients far away from heart failure patients. Classification performance based on the target diagnosis is used as the evaluation metric. Our initial results show significant improvements over many baseline distance metrics in all three settings [13, 14, 15].

Medical Sieve

Another big data challenge in medicine is the effort that is required to review, interpret, and extract the maximum relevant information from across the wide variety of healthcare data and use it for decision making. Electronic patient data is distributed in many enterprise systems in hospitals, and obtaining a holistic perspective of patient condition is difficult and time consuming particularly for those specialists that already look at a lot of patient imaging studies such as radiologists. Statistics show that eye fatigue is a common problem with radiologists as they visually examine a large number of images per day. An emergency room radiologist may look at as many as 200 cases a day, and some of these imaging studies, particularly lower body CT angiographies can be as many as 3000 images per study. Due to the volume overload, and limited amount of clinical information available as part of imaging studies, diagnosis errors, particularly related to coincidental diagnoses can occur.

With radiologists being a scarce resource in many countries, it will be even more important to reduce the volume of data necessary for clinicians, especially since it may have to be sent over low bandwidth tele-radiology networks.

IBM Research is developing a new radiologist cognitive assistant called the Medical Sieve, which is an image-guided informatics system that filters the essential clinical information that physicians need to know about a patient for diagnosis and treatment planning. The system gathers clinical data about the patient from a variety of enterprise systems in hospitals including EHR, pharmacy, labs, Admission-Discharge-Transfer system, and radiology/cardiology Picture archiving and communication systems using HL7 and DICOM adapters. It then uses sophisticated medical text and image processing, pattern recognition, and machine learning techniques guided by advanced clinical knowledge, to process clinical data about the patient to extract meaningful summaries for detecting the anomalies. In doing so, it exhibits a deep understanding of diseases and their interpretation in multiple modalities (X-ray, Ultrasound, CT, MRI, PET, Clinical text) covering various radiology and cardiology specialties. Finally, it creates advanced summaries of imaging studies capturing the salient anomalies detected in various viewpoints.

Medical Sieve algorithms were evaluated for anomaly detection accuracy in many diagnostic imaging modalities in specialties ranging from cardiac, to breast, neuro and musculoskeletal imaging. Specifically, we evaluated the accuracy of discrimination between normal and abnormal left ventricular shapes in a recent publication, in which the left ventricle was automatically located in 4-chamber views and was fitted with a prolate spheroidal model [16]. The ellipsoidal model was used to represent a normal left ventricular shape and deviations from the fit were used as features for discrimination. The method was tested on a dataset of 340 patients and 2,158 echocardiographic sequences depicting a variety of cardiac diseases in patients ranging from aneurysms (89), to dilated cardiomyopathy (76), hypertrophies (78) and normal LV size and function (448), etc. Of these, 503

sequences were 4-chamber views including about 138 sequences labeled as normal LV size and function from their corresponding reports. To discriminate between normal and abnormal LV, we used 40% of the normal and abnormal LV cases for training and the remaining for testing. A total of 25,020 feature vectors were generated from these sequences as they were of variable length in heart cycles and averaged about 64 images per sequence. We experimented with different kernels for learning with support vector machines and the best classification performance was obtained with radial basis functions as kernels. The class was decided at the level of the echocardiographic sequence by taking the majority vote from the classification of parametric features of individual images of the sequence within cardiac cycles. The results are summarized in Table 1.

Table 1 Accuracy of discrimination between normal and abnormal left ventricular shapes.

Classifier	Accuracy (%)
Fisher LDA	68.92
Linear SVM	74.26
RBF kernel SVM	82.67
kNN, k=1	64.12
kNN, k=2	69.07
kNN, k=3	68.21

The quality of summaries generated by Medical Sieve was also evaluated in the domain of coronary angiography studies. In this case, the system prepares an automatic summary of an angiography study by retaining those images depicting the coronary arteries with good visibility as key images for interpretation. The vessel visibility is derived using a measure based on the automatic detection of arterial tubular structures in images. We conducted a study to compare the performance of automatic key frame selection with manually identified key frames by clinicians. This study was conducted over a database of 210 video sequences from 70 patients representing a total of 5,250 images. The system was able to reduce the data browsing load by 95% while still selecting the relevant keyframes

within 1-2 frames of the ones chosen by clinicians during a comparison study with the clinicians. Comparison with summaries generated using Frangi filtering of angiography images showed superior performance by a margin of 26% [17, 18].

Metagenomics

The explosion of genomic information has created further complications for decision makers. Whole genomic sequencing means new opportunities to personalize healthcare. However, it also faces us with processing huge amounts of data and the need to better understand the role of DNA and its segments.

New technology has led to dramatic cost reductions in the price of sequencing a genome, even allowing for sequencing of entire organism genomes. Since 2008, the cost of sequencing per megabase has fallen by 5 orders of magnitude and the cost of full genome sequencing by 4 orders of magnitude [19]. Today, the National Center for Biotechnology Information nucleotide database provides a vast and growing collection of sequences from GenBank® (NIH annotated collection of all publicly available DNA sequences), RefSeq (comprehensive, integrated, non-redundant, annotated set of reference sequences including genomic, transcript, and protein), Third Party Annotation (submitter-provided annotation for sequence data derived from GenBank), and Protein Data Bank (repository for three-dimensional structural data of large biological molecules, such as proteins and nucleic acids) and other sources including approximately 4,000 complete bacterial genomes and 40,000 viruses [20]. Microbes compose the largest fraction of species on Earth. They are integral to the health of humans and the environment. Microbial biodiversity affects the health of ecosystems (e.g. marine estuaries), keystone species (e.g. honeybees) and individual human patients. The growing library of reference genetic data affords the opportunity to identify microorganisms from metagenomic samples of amino acid, DNA, and RNA sequences, by comparing primary sequence data to online reference sources. Today this can be accomplished with genome available software such as the Basic

Local Alignment Search Tool (BLAST) [21]. Genome alignment with tools such as BLAST is an efficient approach to characterizing transcript sequences of previously known and sequenced organisms. However, the method is disadvantaged by its inability to account for phenotypic variations such as alternative splicing [22]. To account for the detection of new organisms as well as natural variations within different phenotypes, a range of complementary big data analytic tools will be required including de novo assemblers. A number of assemblers are available today when it is necessary to create a transcriptome without the aid of a reference genome [22, 23, 24].

We see the generation of vast amounts of genomic data on microbes and a scientific community struggling to manage its volume, variety, and velocity, and master the knowledge encoded within. New algorithms, computing methods, and data repositories are needed to “connect the dots” between genomic readouts of microbial diversity and personal/environmental health. To spread the benefit of metagenomic data analysis across disciplines, algorithms and data need to be available as computing software and services. Scientists are in the beginning stages of developing laboratory and computational procedures to apply metagenomic DNA sequencing for the benefit of healthcare, science, government, and industries utilizing traditional biological testing methods (healthcare, agriculture, environmental management) [24, 25, 26]. We plan to contribute to unlocking the potential of environmental sequencing through scientific discoveries in large-scale data repository management, large-scale bioinformatics analysis, correlation of genomic signals with traditional sensor networks and other signals (e.g. public healthcare ports) as well as cloud-based software and analytic services.

Evidence-based Case Structuring - Evicase

Once evidence is developed we need to be able to incorporate it into the decision process. Current methods, such as guidelines or protocols are beneficial, but have limitations. Clinical practice guidelines can

be used to formulate the predefined rules for the reasoning engine of a clinical decision support system. Clinical practice guidelines are typically consensus and evidence-based guidance, for treating groups of patients defined by some fixed clinical criteria [27]. These guidelines are developed by medical specialties and sub-specialties (e.g., oncology or breast oncology), and are mostly based on a critical mass of controlled clinical trials and other comparative effectiveness studies that demonstrated value for specific interventions [28]. However, the efficacy and side effects of a new treatment cannot necessarily be generalized to different settings in the real clinical environment, as the studies are based on group outcomes and may not apply directly to an individual patient. The trials often do not account for differences among patients with the same disease, often due to lack of data, e.g., personalized genetic make-up of a patient [29]. Therefore, the next generation of CDS should utilize the latest biomedical discoveries [30] to alleviate the translational barriers from bench to bedside, while feeding back lessons learned by its users. Recent advancements follow this direction, for example, CancerLinQ (developed by the American Society of Clinical Oncology) is a "learning health system" designed to draw insight from the vast, untapped pool of data on "real world" patients [31]. A prototype based on 170,000 de-identified medical records of breast cancer patients provided by oncology practices around the United States, already allows data mining and visualization at the point of care for personalized CDS.

In order to improve the implementation of clinical guidelines for specific individuals (care processes), we need to refine established knowledge through data-driven insights by combining rule-based reasoning with case-based reasoning. To address this need, we suggested an evidence-based case structuring framework to generate multi-tiered statistical structures we call Evicase [32]. An Evicase object integrates established biomedical evidence with insights gained from analysis of patient cases in operational information systems of healthcare providers. Established evidence is refined through machine learning analysis of patient data, resulting in various means for clinicians to retrospectively analyze care processes and

to prospectively answer questions regarding an individual patient.

In our implementation, the Evicase is a three-tiered structured object:

1. Tier-one: Knowledge encapsulation provides the guideline view for the specific patient's presentation. The system's knowledge management module interacts with relevant external resources and encapsulates clinical guidelines and evidences. We developed a set of ontologies, rules, and diffusion processes to effectively anchor the clinical domain knowledge into the Evicase and for generating tier-one.
2. Tier-two: Retrospective analysis incorporates insights that are generated from retrospectively analyzing the organization's patient records and from monitoring and assessing its care processes along the established clinical guidelines and best practices. In particular, such algorithms may suggest rule-based patient-similarity metrics according to the guideline's fixed criteria, as well as statistically-based patient-similarity measures that are refined and adapted to the care organization's data. Such analysis enables applications such as guidelines adherence, outcomes assessment, and cost optimization.
3. Tier-three: Prospective analysis applies statistical analysis and machine learning algorithms to the patient records available at the care organization in order to reveal prospective clinical insights. Analyzing patients' clinical data in the context of similar patients may provide prospective outcome assessment, which in turn can be used for treatment recommendations leading to improved patient outcome.

Designed as a stand-alone multi-tiered structure (combining knowledge and data), Evicase can be used for a range of decision support applications including guideline adherence monitoring and personalized prognostic predictions.

For example, an Evicase for analyzing the treatment of Soft Tissue Sarcoma patients has been developed in collaboration with a cancer center in Italy and consists of three tiers: (1) clinical practice guidelines used in that cancer center classified to standard, individualized, experimental, or deviated;

(2) retrospective analyses of clinical records in the cancer center, resulting in patient groups based on similarity according to local guidelines as well as actual outcomes, using machine learning techniques; (3) probability prediction of each outcome for different possible treatments based on the historical outcomes observed among the group of the most similar patients [33].

The Evicase framework is designed to help physicians make informed decisions when literature-based knowledge is insufficient and textbook guidelines are imprecise. As such, Evicase objects could help clinicians increase effectiveness of treatments through the optimization of care processes for specific patient populations. It can also be exchanged with other providers, allowing comparative effectiveness research as well as bringing new business potential in the form of an Evicase open market.

Finally, Evicase might also be used to generate decision support aids for patients, which could provide clinical benefits as well as cost reduction for the individual patient.

Streaming Analytics

Big data characteristics include velocity and complexity. In some environments, data arrives quickly and in large amounts that either cannot be adequately stored or should be analyzed in near real-time because of the immediate nature of clinical decisions.

ICUs are data rich environments, where multiple streams of physiological data from sophisticated patient monitoring systems and ancillary devices are collected and interpreted by clinicians. While the outputs of these devices aim at improving patient care by signaling early warnings of complications, they are also creating an information overload problem. A 50 bed ICU may generate a quarter of a terabyte of data on a monthly basis. Despite containing a wealth of information, only a small subset of these data points are currently exploited for the delivery of care in modern ICUs. The rest is simply dropped after a few days. As a result, intensive care is often provided reactively, in response to adverse events buried in these large volumes of data and typically detected after the emergence of

clinical symptoms, or after the interpretation of a clinical test. An opportunity for an earlier, simpler intervention that could avoid a serious problem may be lost.

The management and analysis of these data points is a big data challenge that has the potential to make critical care much more proactive. There are two classes of such problems that we have addressed in our research. For many patients, complications are presaged by signs buried in patient data streams, but with well understood patterns. These complications include hospital acquired infections, as well as respiratory, cardiac, and neurological events. One notable example of an early warning pattern is the reduction of heart rate variability that is known to be associated with early stages of sepsis [34, 35]. For other complications, the specific signature of early signs in physiological data streams is unknown and subject to research. In this case, mining large historical patient-related data sets could lead to the discovery of new early detection patterns. Our research has led us to address both of these classes of problems.

A) Real-Time Analysis for ICU Patient Data Streams

The detection of known patterns in patient monitoring data for the early detection of complications in ICUs is a real-time analytical problem requiring systems able to analyze in a timely fashion structured and unstructured data points produced at large rates. Just like a gold miner setting up filters on a river to extract gold nuggets, big data analysts use a stream computing paradigm to design filtering analytics able to extract nuggets of information from flows of patient data. We have developed at IBM Research the Online Healthcare Analytics (OHA) infrastructure, also known as Artemis [34], which is a programmable framework for real-time analysis of intensive care sensor data leveraging IBM InfoSphere Streams (Streams), a real-time high-performance stream analysis engine. Streams provides a programming and runtime environment, where analytic developers within medical institutions can develop and deploy real-time analytics on large flows of structured and unstructured data.

OHA also leverages different time series, machine learning and data mining technologies in the form of analytical toolkits to facilitate the authoring of complex real-time applications. OHA interfaces Streams with an open set of data collection systems (e.g., Excel Medical Electronics BedMasterEX system, the CapsuleTech data collection system). Although these tools are designed to be intuitive, it still requires some training and commitment to use them effectively.

While the successful use of custom real-time analytic solutions built for the monitoring of specific conditions has been well documented in the literature [35,36], the OHA platform differs significantly from these systems with its programmability and agility. With OHA, analysts within medical institutions are able to compose and deploy an open set of analytics, tailored to address their goals. As they discover new real-time analytics that they would like to deploy, analysts using OHA do not have to rebuild custom solutions bringing these analytics to the bed side. Instead, they can simply deploy these analytics on the OHA platform. The extensibility of the OHA programming model facilitates the inclusion of analytics written in several common languages ranging from high level languages such R and Matlab to lower level languages like Python or even Java, C++ and C.

Different versions of the OHA system are currently in use in live ICU environments under research agreements, in several types of intensive care ranging from neonatal ICUs [34, 37] to neurological ICUs [38]. Different real-time analytics have been deployed on OHA, including heart rate variability analytics aiming at modeling the autonomic nervous system response as a way to detect early signs of inflammatory responses [37], seizure detection analytics on electroencephalograms [39] and analytics monitoring the intra-cranial pressure auto-regulation in neuro-ICU settings [38].

B) Mining Patient Monitoring Data for the Discovery of Early Detection Patterns

The discovery of new interesting patterns in patient monitoring data is intrinsically an “at rest” data analysis problem requiring systems able to analyze large amounts of his-

torical data sets. We have been using an array of offline analytical platforms such as Weka, R, SPSS, and big data platforms like Hadoop for the mining of large volumes of data. We have created applications of Weka analytics to build models able to predict secondary complication in neuro-ICUs [40]. Patient similarity concepts learned on historical data may allow physicians to make clinical decisions leveraging experiences gathered from data from similar patients observed in the past [41]. An in-silico research study using physiological sensor data streams from 1,500 ICU patients obtained from physionet [42] shows that these similarity constructs may be used to forecast the trajectory of blood pressure streams and help predict adverse ICU events such as acute hypotensive episodes.

Conclusion

One of the obstacles to achieving a personalized, evidence-supported future for healthcare is the effective use of the myriad and voluminous data that surrounds us. We have to be able to acquire and analyze huge amounts of often conflicting historical and research data and turn it into actionable information delivered to the decision makers. However, analytic tools are not a panacea for the problems in healthcare. They offer nothing in isolation, only in the context of a commitment to change. Healthcare professionals, patient advocacy groups, policy analysts and economists have all described various paths and challenges for the desired future. Information technology cannot drive change. Healthcare stakeholders must desire and plan for the transformation of healthcare. Once the strategies for transformation are developed, obstacles can be identified. Then, and only then, can technology be an enabler by helping overcome the obstacles.

An additional limitation to the role of analytics is the availability and quality of information. For example, Watson cannot process or use information that isn't published or available. Publication bias, the tendency to publish studies that are positive or are statistically significant, is a recognized phenomenon [43, 44]. Watson is limited by

publication bias just as a clinician would be, but it may be able to mitigate publication bias because of the large volume of articles it can review. All of the tools described depend to some extent on machine learning. Machine learning works well when there is enough training data to cover all of the features used within the machine learning model. The challenge in almost every case is that there is less training data than we would like, so we have to compensate with other clustering and conditioning techniques (based on subject matter expert knowledge) to get to the level of accuracy and precision required in the medical domain.

The variability or uncertainty that is inherent in big data represents another limitation. Published articles can be contradictory or flawed. Data in EHRs can be inconsistent or erroneous. The need to compensate for data limitations is one of the reasons that all these tools thrive on more data. More data gives them more opportunity to identify and compensate for the flaws. The necessity of managing such conflicted and inconsistent data is what mandates cognitive computing.

Not all decision support requires big data, but big data techniques allow us to incorporate more information when it is helpful. Big data has inherent limitations. The process of looking for patterns in big data will yield a large number of statistical associations. However, many of them will be inconsequential with no discernible causal relationship to the outcome being studied. The number of meaningful relationships may be orders of magnitude smaller. Evaluation and feedback from domain experts can help address this problem by helping identifying the meaningful relationships [45]. The hype surrounding big data, creating unachievable expectations, is a problem in itself [46].

Clinical decision support is only valuable if it is used. There are reports that physicians tend to disregard or not use decision support systems, perhaps from a failure of metacognition, the willingness to assess one's thought process and assumptions [47]. We can only expect wide spread use of decision support tool if it provides clear value. The evidence that decision support systems have improved outcomes at this point is limited [48, 49]. Any new techniques need to prove their value in the clinical world.

We have described an array of analytic and clinical decision support tools IBM has designed to help enable evidence-supported decisions. We have shown that computer resources have been or are being developed to use the different kinds of healthcare information, big data, more effectively. Decision support is one component of the broad-based effort necessary to transform healthcare to improve outcomes and control costs.

References

- Berwick D, Hackbarth A. Eliminating Waste in US Health Care. *JAMA* 2012;307(14):1513-16.
- Lathem W. Individual Medicine. *New Eng J Med* 1968;279(1):22-5.
- IBM. Global Technology Outlook; 2012.
- Deloitte. Tech Trends 2014 Inspiring Disruption. Deloitte University Press; 2014.
- McCall R, Franklin S. Cortical Learning Algorithms with Predictive Coding for a Systems-Level Cognitive Architecture. Second Annual Conference on Advances in Cognitive Systems Poster Collection 2013:149-66.
- Thompson PM, Vidal C, Giedd JN, Gochman P, Blumenthal J, Nicolson R, et al. Mapping adolescent brain change reveals dynamic wave of accelerated gray matter loss in very early-onset schizophrenia. *Proc Natl Acad Sci USA* 2001;98(20):11650-5.
- http://www.nlm.nih.gov/pubs/staffpubs/od/ocpl/library_booktrade_almanac_2012.html accessed Oct. 17, 2012.
- Williamson JW, German PS, Weiss R, Skinner EA, Bowes F 3rd. Health science information management and continuing education of physicians. A survey of U.S. primary care practitioners and their opinion leaders. *Ann Intern Med* 1989;110(2):151-60.
- Ferrucci D, Levas A et al. Watson: Beyond Jeopardy!, *Artificial Intell* 2013;199-200:93-105.
- Lundsgaarde HP. Evaluating medical expert systems. *Soc Sci Med* 1987;24(10):805-19.
- Leaper DJ, Horrocks JC, Staniland JR, De Dombal FT. Computer-assisted diagnosis of abdominal pain using "estimates" provided by clinicians. *Br Med J* 1972;4(5836):350.
- Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen, R, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* 2012;148(6):1293-307.
- Sun J, Wang F, Hu J, Ebadollahi S. Supervised patient similarity measure of heterogeneous patient records. *ACM SIGKDD Explorations Newsletter* 2011;14(1):16-24.
- Wang F, Sun J, Hu J, Ebadollahi S. iMet: Interactive Metric Learning in Healthcare Applications. *SDM* 2011;944-55.
- Wang F, Sun J, Ebadollahi S. Integrating Distance Metrics Learned from Multiple Experts and its Application in Inter-Patient Similarity Assessment. *SDM* 2011;59-70.
- Syeda-Mahmood T, Wang Q, Neillie P, Compas C, Beymer D. Discriminating normal and abnormal left ventricular shapes in four-chamber view 2D echocardiography. *Proc. International Conference on Biomedical Imaging (ISBI)*; 2014.
- Syeda-Mahmood T, Beymer D, Wang F, Mahmood A. Automatic selection of keyframes in angiography videos. *Proc. International Conference on Pattern Recognition (ICPR)*; 2010.
- Frangi AF, Niessen WJ, Vincken KL, Viergever MA. Multiscale vessel enhancement filtering. In: Colchester A, Delp SL, editors. *Medical Image Computing and Computer-Assisted Intervention I MICCAI98*. Berlin: Springer-Verlag; 1998. P. 130-7.
- Wetterstrand KA. DNA sequencing costs: data from the NHGRI large-scale genome sequencing program; 2011. Available at www.genome.gov/sequencingcosts. Accessed July 2014
- National Center for Biotechnology Information, U.S. National Library of Medicine. <http://www.ncbi.nlm.nih.gov/nuccore>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215(3):403-10.
- Birrol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R, Stazyk G, et al. De novo transcriptome assembly with ABySS. *Bioinformatics* 2009;25(21):2872-7.
- Miller JR, Sergey K, Granger S. Assembly algorithms for next-generation sequencing data. *Genomics* 2010;95(6):315-27.
- Graham RJ, Bellare P, DeRisi JL. PRICE: Software for the Targeted Assembly of Components of (Meta) Genomic Sequence Data. *G3(Bethesda)* 2013;3(5):865-80.
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature* 2007;449(7164):804-10.
- Tringe GS, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, et al. Comparative metagenomics of microbial communities. *Science* 2005;308(5721):554-7.
- Ansari S, Rashidian A. Guidelines for Guidelines: Are They Up to the Task? A Comparative Assessment of Clinical Practice Guideline Development Handbooks. *PLoS One* 2012;7(11):e49864. doi:10.1371/journal.pone.0049864
- Dreyer NA, Tunis SR, Berger M, Ollendorf D, Mattox P, Glick R. Why Observational Studies Should be Among the Tools Used in Comparative Effectiveness Research. *Health Aff* 2010; 29:1818-25.
- Kawamoto K, Lobach DF, Willard HF, Ginsburg GS. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. *BMC Med Inform Decis Mak* 2009;9:17.
- Doroshov JH. Selecting Systemic Cancer Therapy One Patient at a Time: Is There a Role for Molecular Profiling of Individual Patients With Advanced Solid Tumors, Editorial. *J Clin Oncol* 2010; 28(33):4869-71.
- <http://www.asco.org/quality-guidelines/cancerlinq> Accessed May 8, 2014
- Rinott R, Carmeli B, Kent C, Landau D, Maman Y, Rubin Y, et al. Prognostic Data-Driven Clinical

- Decision Support - Formulation and Implications. *Stud Health Technol Inform* 2011;169:140-4.
33. Carmeli B, Casali P, Goldbraich A, Goldstein A, Kent C, Licitra L, et al. Evicase: An Evidence-Based Case Structuring Approach for Personalized Healthcare. *Stud Health Technol Inform* 2012;180:604-8.
 34. Blount M, Ebling M, Eklund M, James A, McGregor C, Percival N, Smith et al. Real-time analysis for intensive care: development and deployment of the artemis analytic system. *IEEE Eng Med Biol Mag* 2010;29(2):110-8.
 35. Lake DE, Richman JS, Griffin P, Moorman R. Sample entropy analysis of neonatal heart rate variability. *Am J Physiol* 2002;283:R789R797.
 36. Moorman JR, Carlo WA, Kattwinkel J, Schelonka RL, Porcelli PJ, Navarrete CT, et al. Mortality Reduction by Heart Rate Characteristic Monitoring in Very Low Birth Weight Neonates: A Randomized Trial. *J Pediatr* 2011;159(6):900-6.
 37. McGregor C, James A, Eklund M, Sow D, Ebling M, Blount M. Real-time Multidimensional Temporal Analysis of Complex High Volume Physiological Data Streams in the Neonatal Intensive Care Unit. *Stud Health Technol Inform* 2013;192:362-6.
 38. Schmidt JM, Doerner J, Sow D, Perrote A, Mayer S. Bedside Application for Assessment of Cerebral Autoregulation and Patient State Changes. *Neuro Critical Care Conference* 2013.
 39. Johnson A, Sow D, Biem A. A discriminative approach to EEG seizure detection. *AMIA Annu Symp Proc* 2011:1309-17.
 40. Schmidt JM, Sow D, Agarwal S, Claassen J, Mayer S. Heart Rate Variability Monitoring for Early Detection of Secondary Complications after Subarachnoid Hemorrhage. *Neuro Critical Care Conference* 2013.
 41. Sun J, Sow D, Hu J, Ebadollahi S. A system for mining temporal physiological data streams for advanced prognostic decision support. 10th IEEE International Conference on Data Mining 2010.
 42. MIMIC (Multiparameter Intelligent Monitoring in Intensive Care) II Database [Online]. Available at <http://physionet.org/physiobank/database/mimic2db/>.
 43. Dwan K, Gamble C, Williamson PR, Kirkham JJ, for the Reporting Bias Group. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias — An Updated Review. *PLoS One* 2013;8(7):e66844. doi:10.1371/journal.pone.0066844.
 44. Song F, Parekh-Bhurke S, Hooper L, Loke YK, Ryder JJ, Sutton AJ, et al. Extent of publication bias in different categories of research cohorts: a meta-analysis of empirical studies. *BMC Med Res Methodol* 2009;9:79. doi: 10.1186/1471-2288-9-79.
 45. Silver N. *The Signal and the Noise*. Penguin Press; 2012.
 46. Marcus and Davis, *NY Times Eight (No, Nine!)* Problems With Big Data. *New York Times* April 7, 2014.
 47. Berner E, Graber M. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008;121(5A):S2-S23.
 48. Sahota N, Lloyd R, Ramakrishna A, Mackay JA, Prorok JC, Weise-Kelly L et al; CCDSSystematic Review Team. Computerized clinical decision support systems for acute care management: A decision-maker researcher partnership systematic review of effects on process of care and patient outcomes. *Implementation Science* 2011;6:91.
 49. Roshanov P, Misra S, Gerstein H, et al. Computerized clinical decision support systems for chronic disease management: A decision-maker researcher partnership systematic review. *Implement Sci* 2011;6:92.

Correspondence to:

Martin S. Kohn, MD, MS, FACEP, FACPE
 Chief Medical Scientist
 Jointly Health
 Big Data Analytics for Remote Patient Monitoring
 120 Vantis, #570
 Aliso Viejo, CA, 92656, USA
 E-mail: marty.kohn@jointlyhealth.com