Data Mining for Improved Cardiac Care

R. Bharat Rao, Sriram Krishnan and Radu Stefan Niculescu
Siemens Medical Solutions, Computer Aided Diagnosis and Therapy Group
51 Valley Stream Parkway, E51
Malvern, PA, 19355
{bharat.rao, sriram.krishnan, stefan.niculescu}@siemens.com

ABSTRACT

Cardiovascular Disease (CVD) is the single largest killer in the world. Although, several CVD treatment guidelines have been developed to improve quality of care and reduce healthcare costs, for a number of reasons, adherence to these guidelines remains poor. Further, due to the extremely poor quality of data in medical patient records, most of today's healthcare IT systems cannot provide significant support to improve the quality of CVD care (particularly in chronic CVD situations which contribute to the majority of costs).

We present REMIND, a Probabilistic framework for Reliable Extraction and Meaningful Inference from Nonstructured Data. REMIND integrates the structured and unstructured clinical data in patient records to automatically create high-quality structured clinical data. There are two principal factors that enable REMIND to overcome the barriers associated with inference from medical records. First, patient data is highly redundant – exploiting this redundancy allows us to deal with the inherent errors in the data. Second, REMIND performs inference based on external medical domain knowledge to combine data from multiple sources and to enforce consistency between different medical conclusions drawn from the data – via a probabilistic reasoning framework that overcomes the incomplete, inconsistent, and incorrect nature of data in medical patient records.

This high-quality structuring allows existing patient records to be mined to support guideline compliance and to improve patient care. However, once REMIND is configured for an institution's data repository, many other important clinical applications are also enabled, including: quality assurance; therapy selection for individual patients; automated patient identification for clinical trials; data extraction for research studies; and to relate financial and clinical factors. REMIND provides value across the continuum of healthcare, ranging from small physician practice databases to the most complex hospital IT systems, from acute cardiac care to chronic CVD management, and to experimental research studies. REMIND is currently deployed across multiple disease areas over a total of 5,000,000 patients across the US.

1. INTRODUCTION

Since 1990, more people have died worldwide from CVD than from any other cause. Clearly CVD is an international crisis; however, since all applications described in this paper are from US healthcare institutions, we focus on the United States.

1.1 CVD In The United States

In the United States, an estimated 70 million people have some form of CVD. CVD accounts for roughly one million deaths per year (38% of all deaths), and is a primary or contributing cause in 60% of all deaths [4][1]. CVD claims as many lives per year as the next 5 leading causes of death combined. Unfortunately, a number of trends suggest that the problems of cardiovascular disease will only be exacerbated in the future. First, the aging of the U.S. population will undoubtedly result in an increased incidence of CVD [8]. Second, there is an explosive increase in the number of Americans that are obese or have type 2 diabetes; these conditions result in increased cardiovascular complications.

In addition to being a personal health problem, CVD is also a huge public health problem. In the United States, it is estimated that $394 billion will be spent in 2005 on treatment and management of cardiovascular disease. By comparison, the estimated cost of all cancers is $190 billion. By any measure, the burden of CVD is staggering.

Most patients with CVD will never be cured; rather, their disease must be managed. Often, people with CVD will live for 10 or 20 years after initial diagnosis. A significant portion of the costs associated with CVD comes about when the chronic disease is not managed well, and the patient comes to the emergency room of a hospital with an acute disease, such as a heart attack or stroke. This is further exacerbated by the shortage in the number of cardiologists in the United States. Of the approximately 18,000 practicing cardiologists in the US, over 5,000 are above the age of 55, and 400-500 will retire every year, while less than 300 will enter the workforce. This highlights the need to better manage CVD patients after diagnosis – particularly to provide tools to help the overburdened cardiologist improve the quality of care delivered to CVD patients.

1.2 CVD Guidelines

As the problem of CVD has exploded, so has medical knowledge about how to best diagnose and treat it. New diagnostic tests and therapies are constantly being developed. These tests have shown great promise for both improving the quality of life for the CVD patient, and reducing the burden of health care by reducing the incidence of acute episodes. In an attempt to improve the quality of care for patients, national health organizations, such as the American Heart Association (AHA) and the American College of Cardiology (ACC) have created expert panels to review the results of various clinical trials and studies, extract out best practices, and then codify them into a series of *guidelines*. These guidelines attempt to assist the physician on how to best treat patients with CVD. (This process is not unique to cardiovascular disease, but happens in every branch of medicine.)

Recent studies have shown that strict adherence to these guidelines result in improvements at a personal level, including reduced morbidity and mortality and improved quality of life, as well as reduced costs to the overburdened health care system. Based on these studies CMS (the Center for Medicare & Medicaid Services) has begun a series of programs to reward physicians and hospitals who comply with guidelines in an attempt to improve guideline adherence. These "pay-for-performance" schemes are intended to provide a direct financial incentive to healthcare

providers – in this case, CMS is working with hospitals to promote the adoption of the heart attack component of the AHA and ACC cardiac treatment guidelines, which recommend that physicians prescribe a medicine called a beta blocker early after an acute heart attack and continue the treatment indefinitely in most patients. Beta blockers are prescription medicines that help protect the heart muscle and make it easier for the heart to beat normally. Despite being well-known, compliance to this guideline in the U.S. is estimated to be below 50%.

There is overwhelming evidence showing the huge benefits of following these guidelines, from the perspective of the patient, physician, hospital, and public health. Yet overall guideline adherence remains woefully low. There are 3 principal factors which contribute to this lack of compliance.

First, in recent years, there has been an explosion in guidelines. In the United States, the National Guideline Clearinghouse (www.guideline.gov) has almost 1000 guidelines for physicians to follow. These guidelines are often modified on a periodic basis, such as every year, in response to new medical knowledge. A quick search on Google or Med-Line for heart failure guidelines returns several hundred references – some heart failure guidelines, with subsequent modifications are defined in [1][2][3][13][15].

Second, with the growing trend of HMOs, and the economic realities of medicine today, physicians are forced to see more and more patients in a limited amount of time. Often, physicians will only average 10-18 minutes per patient, and carry a patient load of 20-30 patients per day.[1]

Third, there are often multiple physicians and nurses who interact with the patient, and there is often poor communication between these health care workers with regards to the patient. In such a hectic and chaotic environment, it is impossible to (manually) consistently and accurately identify and follow the specific guidelines for that patient among the hundreds of ever-changing requirements in use. Unless the proper clinical guideline is identified and followed at the point of care (that is, when the patient is with his physician), it is not useful.

1.3 Electronic Health Records (EHR)

The electronic health record (EHR) is increasingly being deployed within health care organizations to improve the safety and quality of care [12]. Because a guideline is simply a set of eligibility conditions (followed by a set of recommended treatment actions) it appears fairly straightforward to determine guideline eligibility by evaluating a guideline's inclusion and exclusion criteria against an EHR. Unfortunately, as discussed below and later in Section 5.3, even the best EHRs in the world do not fully capture the information needed to support automated guideline evaluation.

Medical patient data in electronic form is of two types: financial data and clinical data. Financial data consists of all the information required to document the physician's diagnoses and the procedures performed, and is collected primarily for the purpose of being reimbursed by the insurance company or the government. Financial data is collected in a highly structured, well-organized, and normalized fashion, because if it were not in this form, the payers would not reimburse the institution or physician. This data can, therefore, be analyzed, dissected, and summarized in a variety of ways using well-established database and data warehousing methods from computer science.

In addition to structured information about patient demographics, this "financial data" also includes standardized patient diagnoses which are classified according to the internationally accepted standards, ICD-9 (International Classification of Diseases, 9th Revision [40]) and ICD-10 [41]. Many of the criteria used to determine if a patient is eligible for (and therefore should be treated according to) a particular guideline, are based upon diagnostic information. Therefore, it appears as if these structured diagnosis codes would be a rich source for data mining, and particularly for determining whether a patient was eligible for a particular treatment guideline.

Unfortunately, these ICD-9 (and ICD-10) codes are unreliable from the clinical point of view. Various studies have shown that the clinical accuracy of ICD codes is only 60%-80% [7]; in other words, when an ICD code is assigned, the patient will have that corresponding clinical diagnosis only 60-80% of the time. The principal reason for this is that billing data reflects financial rather than clinical priorities.

In the United States, reimbursement is based primarily on the severity of diagnosis; for example, although the patient treatments for AMI (heart attack) and Unstable Angina (a less severe cardiac illness) are virtually indistinguishable, the former diagnosis code generates twice the reimbursement for the institution. There have been several well-publicized cases, where institutions have received hefty fines for "over-coding" (i.e., assigning higher diagnosis codes than is justified). Alternately, billing codes may be missing, or "under-coded", so that institutions are not accused by insurance companies of fraudulent claims. Furthermore, at least in the US, this coding is done by medical abstractors, who although trained to do this coding, typically lack the medical training to assess the clinical data and arrive at the correct diagnosis.

Clearly, financial data alone is insufficient for any kind of patient-level clinical decision support (including determining guideline eligibility), because the errors will multiply when multiple such diagnoses are jointly needed to make a decision (for instance to determine eligibility for a guideline).

Operational clinical systems have very poor data quality from the standpoint of access and analysis. The structured clinical data in clinical repositories (labs, pharmacy, etc.) is sparse with gaps in data and in time, inconsistent due to variations in terminology, and can be clinically misleading. Key clinical information is stored in unstructured form in the clinical repository, typically as unstructured free text in patient history and physicals, discharge summaries, progress notes, radiology reports, etc. Further, the

[1] 10-20 minutes per patient appears reasonable, but it includes *all* activities associated with the patient visit, including: reviewing previous patient history; talking with the patient about their symptoms and history; examining the patient; arriving at a diagnosis; ordering additional tests and procedures; determining what drugs the patient is currently taking; prescribing treatment and medication; explaining the diagnosis and treatment to the patient; counseling the patient on the risks and rewards of the therapy; and ordering referrals if needed; this time also includes time needed for the physician to record all the details of the patient visit including positive and negative findings, impressions, orders, final instructions, and finally signing off on the patient bill.

nature of the relationships within data are not well defined, and causal relationships and temporal dependencies cannot be unearthed without medical knowledge; for example, it may not be immediately clear to which diagnosis a procedure "belongs". Efforts to extract key clinical information based on natural language processing alone have met with limited success [25] – and for even slightly complex decisions like guideline eligibility, reliability is very poor. Simply put, the data in clinical repositories is often messy, and thus only a small fraction of the clinical data is available for analysis.

1.4 The "Data Gap" in medical records

Consider the extremely simple guideline: "If a patient is admitted with a heart attack, they should be prescribed beta blockers upon discharge."

In order to assess compliance, it would appear to be sufficient to determine if the patient was admitted with an AMI (acute myocardial infarction or heart attack) and if they were prescribed beta-blockers. Unfortunately, as discussed earlier, even if the patient has an ICD-9 code for an AMI it may not be clinically accurate. The patient may choose to fill a prescription for a beta blocker at a retail pharmacy, so the institution's pharmacy system (if it has one) will have no record of a beta blocker. Most importantly, even if it were possible to determine if the patient did have an AMI this visit and was (or was not) prescribed beta blockers, there are no data fields to determine if beta blockers are contra-indicated, that is, should not be prescribed due to some other reason, such as other medications, complications, or if the patient is known to be allergic to that drug. To receive certification from JCAHO [20], hospitals hire trained nurses to manually extract information from a random sample of 75 emergency room patients about appropriate beta blocker prescription (and a few other very simple guidelines). In short, this cannot be automatically determined using naïve approaches.

1.5 Automated Patient Data Analysis

Currently there are 3 main ways to perform automated data analysis, discussed below:

1) The most common method, "Limited automated extraction of structured elements only", brings over only the coded financial information (e.g., ICD-9 codes), and loses much of the required clinical information. Further, the coding process has a surprisingly high fraction of errors [29]. Doctors are very pressed for time in the 10-20 minutes they have per patient. If a system alerted a physician about guidelines based on a patient's ICD-9 codes, it would have so many false alerts that the physician would turn it off. (This is not to indicate that billing data is useless. It is used for aggregate level analysis for epidemiological, quality of care, and cost studies [10][17][27] by hospitals, insurers, the US Dept. of Health Care and CMS. And furthermore, REMIND also leverages this data. The key point is that billing data *alone* is useless for decision support.)

2) "Manual conversion of data by medical experts" leads to high-quality clinical data. But, this is expensive, time consuming, and is only possible for a small subset of patients or at institutions with a strong research focus. It is infeasible for routine clinical use.

3) "Forcing doctors to provide structured input." Currently physicians document their observations as dictated free text, and are extremely efficient at doing so. Taking several minutes (out of the 10-20 m/patient) to additionally fill in specific values in a database can lead to physician resentment, wastes valuable physician time and still leads to missing information (fields may not be provided for all needed information in advance). More clinical data will become available in structured form as EHRs get more accepted. But it will take several years before EHRs will be in routine use for a large fraction of the patient population.

The bottom line is that clinical data is complex, non-uniform and non-homogenous. Automated clinical data analysis of the kind associated with financial data, is almost impossible today. There is a desperate need to create highly-structured clinical data from existing patient records collected by the institution in its day to day practice without requiring any manual data entry or change in physician workflow. Our solution works in the current scenario with poor data quality. However, it is designed to be scalable with respect to the volume and quality of data. REMIND will further benefit as better quality data becomes available, via EHRs or by manual methods.

2. THE REMIND ALGORITHM

In this section we briefly describe the problem and the algorithm employed by REMIND (Reliable Extraction and Meaningful Inference from Nonstructured Data) in order to solve it. Our goal is to infer the values of several medical outcomes, described by a set of variables of interest. Examples of such variables include: whether a patient has a particular disease, whether a patient has received a certain type of medication, lab recordings for blood glucose, whether a patient has specific contraindications for a class of medication.

Our approach to inference with this multi-source data is to model the data as arising from a generative process, and combine prior medical knowledge about this process with observations for a specific patient using Bayesian techniques. The medical prior knowledge is encoded in both a Bayesian Network that relates variables of interest as well as in the form of probabilistic rules, as we will see next.

2.1 Problem Definition

Let **V** be the set of variables of interest for a patient. Let **O** be set of all (probabilistic) observations for all variables, $\boldsymbol{v} \in \mathbf{V}$. Similarly, let $\mathbf{O}(\boldsymbol{v})$ be the set of observations for variable $\boldsymbol{v}$. We assume the relationships within **V** are described by a Bayesian Network.

Since we are interested in the most likely value for our variables given the observations extracted from the patient data, our goal is to estimate:

$$V_{MAP} = \arg\max_V P[V \mid O]$$

2.2 Overview of Approach

REMIND's 3-step process that estimates the value of the variables of interest V_{MAP} is summarized below. Our goal is to extract and combine information from all data sources.

(1) **Extraction** step: observations are gathered from the data sources. These observations provide the basic information about the variables $\boldsymbol{v} \in \mathbf{V}$. Operationally; they are converted into a uniform representation, called *probabilistic observations*. These play the same role as likelihood findings in standard Bayesian

reasoning. Note that every observation o ∈ **O** is assumed to be potentially incorrect.

(2) **Combination** step: each observation is assigned to its corresponding variable and a posterior of the observation vector associated with the variable is computed locally.

(3) **Inference** step: the local inferences are propagated across the Bayesian Network that describes the relationships among **V** and the posterior probabilities for the variable vector are computed.

These steps are in direct correspondence to the different propagation steps of the belief propagation algorithm, well known in the probabilistic inference literature.

2.3 Extraction of Probabilistic Observations From Data

In this step we *produce probabilistic observations, O_i, from data sources*. Each O_i is drawn entirely from a single piece of information in a data source (e.g., from a phrase in a sentence, or a row in a database), and hence is assumed to be inherently undependable (either due to errors in the data or in the extraction process). An observation O_i is of the form <NAME, DATE, DIST> where NAME is an observed variable $\boldsymbol{v} \in \mathbf{V}$, DATE is the date of the observation, and DIST defines a distribution over all possible values that can be taken by NAME given the observation. REMIND currently does extraction from relational databases and free text. Methods from computational linguistics are used to extract information from free text.

These observations generated from the data sources are meant to encode the *a posteriori* distribution of a variable given the section of the data source that they are extracted from, and are subsequently converted into likelihood findings for computation in the Bayesian Network.

2.4 Combination & Inference

The primary focus is estimating the most likely (MAP) state of the variables given the observations extracted in the previous step. This can be done in two steps, the first of which is a local combination/inference of observations for each variable, followed by the propagation of these inferences across the Bayesian Network.

Each piece of information that is extracted in the previous step is in the form of an *a posteriori* probability of a variable given the small context that it is extracted from. We can thus have multiple such assertions from different parts of the same source and from different sources at any given instant in time. All the assertions about a variable are combined into one assertion in a straightforward manner by using Bayes' theorem (under the assumption that the observations are independent given the variable) as follows:

$$\Pr[O_1(v),\ldots,O_k(v)\,|\,v] \propto \prod_{i=1}^{k} \Pr[O_i(v)\,|\,v] \propto \frac{\prod_{i=1}^{k} \Pr[v\,|\,O_i(v)]}{\Pr[v]^k}$$

We model the relationships among the set of all variables of interest using a Bayesian Network, which is used to infer the posterior distribution of all the variables given all the information available:

$$P[V\,|\,O] \alpha \prod P(v_j\,|\,PA(v_j)) \cdot P(O(v_j)\,|\,v_j)$$

In other medical Bayesian applications [5][22][26], the actual probability values for the dependencies within V are typically a huge bottleneck, and require tremendous fine tuning. Because REMIND leverages data redundancy, our systems works well for a wide range of probability values for inference and extraction [33].

3. REAL-WORLD IMPLEMENTATIONS OF REMIND

REMIND has been implemented for a wide variety of different quality metrics with data from a number of different institutions. Some examples include glycemic control for diabetic patients with heart attacks [29] to detection of recurrence of colon cancer [30]. In the following sections, two examples of the use of REMIND for cardiac care are described.

3.1 Quality of Care Analysis for Multiple Institutions

As stated previously, cardiac disease is a major health problem in the United States and throughout the world. Many of the hospital stays associated with cardiac disease occur because of acute incidents that can be avoided if the patient is properly treated and monitored, and several guidelines have been developed to improve the quality of care [1]. To assist these efforts, several leading medical organizations, including the ACC, AHA, and the AMA, have jointly identified key performance metrics to assist with proper monitoring and treatment of heart failure patients. These metrics are designed to assist the cardiologist monitor the health of the patient, and assess whether changes in treatment are needed. In addition, these metrics list key medications that the patient should be taking. The AMA has created PCPI, the Physician Consortium for Practice Improvement, to be responsible to codify and maintain these metrics.

Unfortunately, simply generating a guideline or metric does not guarantee that physicians will follow them. To assist physicians and practices with compliance to these guidelines, REMIND was used on data from two physician practices consisting of a total of 270,000 patients. First, patients with heart failure were identified using both ICD-9 codes as well as by analyzing the physician notes. Then, each of the metrics in the PCPI guidelines were extracted for these heart failure patients.

For example, the PCPI guidelines state that every heart failure patient should have a number of measurements and assessments taken each year, including left ventricular function, blood pressure, signs and symptoms of cardiac volume overload, activity level, etc. Each of these measurements can be done in a number of different ways. For example, left ventricular function can be assessed using various imaging modalities, such as ultrasound, nuclear medicine, etc. Activity level can be assessed through observation of the patient through one of many simple exercises. Sometimes, there will be explicit data on these, but other times the assessment of these things must be inferred from the physician's dicated notes. In addition, the PCPI guidelines state that patients should be on medications such as beta blockers, ACE or ARB, and Warfarin (for patients who also have atrial fibrillation) unless there are contra-indications to these

medications. REMIND was used to assess each of these guidelines at a patient level, and then aggregated to the entire physician practice (for both practices).

A second analysis was done on patients taking a medication called amiodarone. This is an extremely powerful, but toxic, drug used to treat atrial fibrillation, a cardiac condition. In addition to its toxicity, it often can lead to complications in cardiac patients taking other medications. Because of this, it is very important for patients who are taking amiodarone to be monitored periodically (usually every 6 months) for signs of toxicity. The North American Society of Pacing and Electrophysiology (NASPE) has released a set of guidelines for monitoring patients taking Amiodarone [13]. Our system identifies patients who are taking amiodarone, and then within this subset, those patients who are not being treated as per the NASPE guidelines. The goal here is to help reduce the incidence of side-effects due to the toxic nature of Amiodarone.

REMIND was run at both institutions' data with virtually no change in the system. We are in the process of expanding our pool to 1,000,000 cardiology practice patients, and plan to offer a suite of quality of care reports and facilitate benchmarking, both to national standards and across institutions.

3.2 Guideline Adherence Study for Patients with Non-ST Elevation MI

The Veteran Health Administration (VHA) patient database is universally acknowledged as one of the best (if not the best) databases of clinical information in the world. The VHA database is designed to collect a tremendous amount of clinical information in structured form – in addition to the demographics, diagnosis (ICD-9), laboratory, and pharmacy system, many additional clinical variables are recorded in structured form. Additionally, the VHA database has a vast store of unstructured free text, including history and physicals, admission and discharge reports, progress notes, specialist reports, nursing evaluations, and radiology, ECG, and ultrasound reports. In fact, the VHA database is being strongly recommended by CMS as a model for future EHRs.

TREATMENT	ACCURACY (%) N=327	
	REMIND	MANUAL
Aspirin	319 (97%)	314 (96%)
Beta Blockers	319 (97%)	316 (97%)
ACE Inhibitors/ARB	300 (92%)	310 (95%)
Glycoprotein IIb/IIIa Receptor Antagonists	300 (92%)	290 (89%)

Table 1. Accuracy of REMIND vs. trained medical nurse for guideline compliance

It was expectation that with such a tremendous database, the history of quality of care research, and the diligent efforts of the physicians and nurses to keep it current over the last 20 years, there would be little need for automated REMIND analysis. As expected, the support for automated analysis was significantly better than that at any other institution we have encountered. However, somewhat surprisingly we also found that despite the world-class database and research, the available structured data was ineffective for answering questions about the quality of care and compliance.

As discussed previously, one of the big needs in cardiology is to assess whether patients are being treated properly as per established clinical guidelines. The treatment guideline for patients with a certain type of myocardial infarction, in this case patients with non-ST elevation MI was provided by the ACC [9]

The main responses to the guideline are to provide medication to the patient. For each patient, one must select the correct set of medications for the patient. There are four broad classes of medication for these patients: aspirin; angiotensin converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARB); beta blockers; and glycoprotein IIb/IIIa receptor anatognists. For each medication, it is important to figure out if the patient should be taking the drug, and also if a patient has a known contra-indication (allergy) to the drug. For example, ACE or ARBs should only be given to patients with diabetes mellitus, congestive heart failure, left ventricular dysfunction or hypertension. In addition, there are a number of reasons a patient even in these conditions should not be given the medication, such as if the patient is pregnant, has pulmonic or aortic stenosis, renal failure, etc. As one can see, the determination of the appropriateness of each class of medication is quite complex.

The VHA in Pittsburgh, PA has been conducting a retrospective research study on a population of 1400 patients. A trained research nurse manually extracts the information for about 90 variables from these patients. We implemented domain knowledge within REMIND to extract information for about 80 of these variables, and have compared the results of the extraction with the manual extraction on about 1000 patients.

In this paper, we present the results of analysis for a sub-population of 327 patients admitted with non-ST elevation MI. These patients were studied to see if they were treated properly for each of these four classes of medications as per the ACC guidelines [9]. For each patient, the patient record was searched to see if the patient was treated properly for each of these four medications by both REMIND and manually with the manual abstraction. For each patient, any disagreement between REMIND and the abstraction was adjudicated manually by a medical expert. If REMIND and the research nurse's extraction agreed, both were assumed to be correct. Note that the research nurse had access to the entire patient record, which includes information that was not available to REMIND.

REMIND took 4.5 hours to extract the values of the 4 variables (see Table 1) for 327 patients using a Pentium M 1.6 GHz laptop. (The current version of REMIND is expected to be faster by about 2-3 orders of magnitude.) The medical abstractor took 176 hours to complete the analysis manually for the same variables [36].

Table 1 compares the accuracy of REMIND and manual abstraction for each of the 327 patients. That is, for each patient, this analysis shows what percent of patients were accurately assessed using REMIND and manual abstraction (using the adjudication as a gold standard). Table 1 shows that REMIND works at least as well as manual abstraction in identifying patients who were treated per guidelines for non-ST elevation MI

In a controlled study like this, it is possible to spend the time to manually review every patient to assess performance. In reality,

however, it is impractical to expect a medical expert to spend time to manually review every patient chart to study if the patient was treated properly or not. In this study, only non-ST elevation MI was considered. If one includes the full spectrum of cardiac diseases, including ST elevation MI, heart failure, arrhythmias, etc., then one can easily see how daunting a task it would be to review every chart for compliance. By using a tool like REMIND, it would be possible to review patients with many different conditions. This would enable physicians to ensure that patients were treated properly, and hence improve their conditions dramatically.

4. RELATED RESEARCH

Our work draws heavily on earlier work on Bayesian networks and graphical models (see [16][19] for an overview). Probabilistic networks have been used in biomedicine and health-care have become increasingly popular for handling the uncertain knowledge involved in establishing diagnoses of disease, in selecting optimal treatment alternatives, and predicting treatment outcomes in various different areas. For example, DxPlain [5] is a decision support system which uses a set of clinical findings (signs, symptoms, laboratory data) to produce a ranked list of diagnoses which might explain (or be associated with) the clinical manifestations. DXplain provides justification for why each of these diseases might be considered, suggests what further clinical information would be useful to collect for each disease, and lists what clinical manifestations, if any, would be unusual or atypical for each of the specific diseases. Quick Medical Reference (QMR [26]) is a large probabilistic graphical model which combines statistical and expert knowledge for approximately 600 significant diseases and 4000 findings. In the probabilistic formulation of the model [34] the diseases and the findings are arranged in a bi-partite graph, and the diagnosis problem is to infer a probability distribution for the diseases given a subset of findings. Promedas [22] is a patient-specific diagnostic decision support system which produces a differential diagnosis on the basis of a set of patient findings. It also suggests the most informative tests that may be performed to make the differential diagnosis more precise. Promedas is based on medical expert knowledge encoded into a probabilistic graphical model (a Bayesian network), which serves as the inference engine of the system. These systems all require clinical data to be entered in a structured database.

Combi et al [11] provides an extensive review of temporal reasoning methods in medicine. We briefly list some methods that are similar to REMIND in some aspects. Ngo et al [28] describe a temporal probabilistic reasoning method via context-sensitive model construction. Bellazi et al [6] describe a system that uses a Dynamic Bayesian Network to analyze the blood glucose level of a patient over a time interval. Kayaalp et al [23] use structured information to predict probabilities of survival for ICU patients. Other related research [18][21][24] deals with representing temporal data and enforcing temporal integrity.

5. NEXT STEPS

Our immediate next step is to incorporate REMIND into the point of care. REMIND can provide point of care support to the physician, for instance, by evaluating the patient against all guidelines, and assess treatment against these guidelines.

Other interesting applications include disease surveillance, epidemiological studies, bioterrorism surveillance, and outbreak detection. The RODS [37] (Real-time Outbreak and Disease Surveillance) system mines emergency room data (specifically, 7 fields are provided) and can detect early signs of an outbreak, particularly by detecting spikes in ER admissions. Our approach is complementary, based on a more detailed analysis of individual patient data. We also intend to explore pay-for-performance opportunities with CMS and other payers. Medicine is rich with knowledge bases such as taxonomies (LOINC [32], MeSH [38], and RxNORM), controlled vocabularies (SNOMED CT [35]), and ontologies (UMLS [39]). These systems provide reasoning with crisp logic but unable to handle uncertain knowledge and incomplete/imprecise data. REMIND will incorporate these external sources of knowledge into its inference.

6. CONCLUSIONS

We conclude by re-stating some key points:

Medical data is highly complex and difficult to analyze. Financial data is well organized but has limited clinical value. Clinical data is very poor from the point of view of automated analysis. Systems that collect high-quality data will become part of routine clinical care, but are unlikely to have a large patient impact in 5-10 years.

Methods based on analyzing a single kind of data, for example, billing data alone, or just text data (with NLP) are unlikely to have much success. Each source of data has its unique limitations, which might be overcome by information from another data source.

Our solution, REMIND, overcomes these problems by exploiting the redundancy in patient data, and combining information from multiple sources based on external medical knowledge. A probabilistic reasoning system performs the actions necessary to infer high-quality clinical data despite the contradictions, errors, and omissions in the data (and the data extracts from the patient record).

Here we have only discussed cardiac applications of REMIND. REMIND has been used for other disease areas, including cancer, and efforts are underway to combine images with clinical and financial data to improve analysis. REMIND is currently deployed on a rapidly growing population of over 5,000,000 patients.

7. ACKNOWLEDGEMENTS

This system would not have been successful without the guidance of our clinical collaborators who continue to help shape REMIND: Harsha Rao, MD (Univ of Pittsburgh Medical Center), Colin Germond, MD (Cancer Care Ontario, Canada), Venk Gottipaty, MD, Tim Attebery, and Sherry Schutz (South Carolina Heart Center), Sheryl Dodds, RN, Katie Packard, PharmD, RN, Jacque Taylor, RN, and Kathy Smith, RN (Nebraska Heart Institute), Ali Sonel, MD, Chester Good, MD, Lauren Wall, RN, and Alanna Macioce, RN (VA Hospital, Pittsburgh). Most of all, we are grateful for their support during the early phases of REMIND, and willingness to trust in and help refine a paradigm that is totally different from one traditionally followed in medicine.

We would also like to thank Siemens colleagues: Sathyakama Sandilya, Ph.D., Ingo Schmuecking, MD, William Landi, Ph.D., Alok Gupta, PhD, MBA, Ajit Singh, Ph.D., Arun Goel, Geoff Towell, PhD, Prasad Aloni, Michael Greenberg, Romer Rosales,

PhD, Balaji Krishnapuram, PhD, Harald Steck, PhD, Abhinay Pandya, the entire SISL team and Narasimha Murthy. Without their diligence, research and inventiveness, REMIND would have never transitioned from concept to product.

8. REFERENCES

[1] Advisory Council to Improve Outcomes Nationwide in Heart Failure. Consensus recommendations for the management of chronic heart failure. Am J Cardiol 1999;83 (2A):1A-38A.

[2] American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Guidelines for the evaluation and management of heart failure. Report of the ACC/AHA Committee on Evaluation and Management of Heart Failure. J Am Coll Cardiol 1995;26:1376-98.

[3] American College of Cardiology/American Heart Association Task Force on Practice Guidelines. ACC/AHA Guidelines for the Evaluation and Management of Chronic Heart Failure in the Adult. ACC/AHA Committee to Revise the 1995 Guidelines for the Evaluation and Management of Heart Failure. J Am Coll Cardiol. 2001; 38:2101-13.

[4] American Heart Association. Heart Disease and Stroke Statistics – 2005 Update. Dallas, TX, American Heart Association, 2004.

[5] Barnett, G.O., Cimino, J.J., Hupp, A.J., and Hoffer, E.P. DXplain: an Evolving Diagnostic Decision-Support System, JAMA, 1987, Vol. 258(1), pp. 67-74.

[6] Bellazzi, R., Larizza, C., De Nicolao, G., Riva, A., Stefanelli, M. Mining biomedical time series by combining structural analysis and temporal abstractions. JAMIA (symposium supplement), vol. 5 (1998), 160-164.

[7] Benesch, C., Witter Jr, D.M., Wilder, A.L., Duncan, P.W., Samsa, G.P., Matchar, D.B. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. Neurology, 1997, Vol. 49, pp. 660–664.

[8] Bonow, R.O., Smaha, L.A., Smith Jr, S.C., Mensah, G.A., and Lenfant, C. World Heart Day 2002: The International Burden of Cardiovascular Disease: Responding to the Emerging Global Epidemic. Circulation, 2002, Vol. 106, pp. 1602 – 1605.

[9] Braunwald, E., Antman, E.M., Beasley, J.W., et al. ACC/AHA 2002 guideline update for the management of patients with unstable angina and non-ST-segment elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Committee on the Management of Patients With Unstable Angina 2002.

[10] Broderick, J., Brott, T., Kothari, R., Miller, R., Khoury, J., Pancioli, A., Mills, D., Minneci, L., Shukla, R. The Greater Cincinnati/Northern Kentucky Stroke Study: preliminary first-ever and total incidence rates of stroke among blacks. Stroke. 1998, Vol. 29, pp. 415–421.

[11] Combi, C., Shahar, Y. Reasoning and Temporal Data Maintenance in Medicine: Issues and Challenges. Computers in Biology and Medicine, Vol. 27(5), 1997, pp. 353-368.

[12] Committee on Data Standards for Patient Safety, Board on Health Services. Key Capabilities of an Electronic Health Record System: Letter Report. Institute of Medicine of the National Academies, 2004.

[13] Goldsclager, N. et al. Practical Guidelines for Clinicians Who Treat Patients with Amiodarone. Arch Intern Med. 2000; 160:1741-1748

[14] Heart Failure Society of America. HFSA practice guidelines. HFSA guidelines for management of patients with heart failure caused by left ventricular systolic dysfunction--pharmacologic approaches. J Card Fail 1999;5:357-82

[15] Heart Society of America. HFSA Practice Guidelines. HFSA Guidelines for Management of Patients with Heart Failure Caused by Left Ventricular Systolic Dysfunction – Pharmacological Approaches. Pharmacotherapy. 2000; 20(5):495-522.

[16] Heckerman, D. A tutorial on learning with Bayesian networks. Microsoft Research Technical Report, MSR-TR-95-06, 1996.

[17] Holloway, R.G., Witter Jr, D.M., Lawton, K.B., Lipscomb, J., Samsa, G. Inpatient costs of specific cerebrovascular events at five academic medical centers. Neurology, 1996 Vol. 46, pp. 854–860.

[18] Horn, W., Miksch, S., Egghart, G., Popow, C., Paky, F. Effective Data Validation of High Frequency Data: Time-Point, Time-Interval, and Trend-Based Methods. Computers in Biology and Medicine, 1997.

[19] Jensen, F.V. An introduction to Bayesian Networks. UCL Press, 1996.

[20] Joint Commission on Accreditation of Healthcare Organizations (JCAHO), Website: http://www.jcaho.org.

[21] Kahn, M., Fagan, L., Tu, S. Extensions to the Time-Oriented Database Model to Support Temporal Reasoning in Expert Medical Systems. Methods of Information in Medicine, 1991, Vol. 30, pp. 4-14.

[22] Kappen, B., Wiegerinck, W., Akay, E., Neijt, J., and Van Beek, A. Promedas: A Clinical Diagnostic Decision Support System in Bayesian Modeling Applications. Proceedings of the 15th Belgian-Dutch Conference on Artificial Intelligence (BNAIC'), Nijmegen, The Netherlands, October 23-24, 2003. pp. 455-456.

[23] Kayaalp, M., Cooper, G. F., Clermont G. Predicting ICU Mortality: A Comparison of Stationary and Nonstationary Temporal Models. Proceedings of American Medical Informatics Association (AMIA) Symposium, 2000, pp. 418-422.

[24] Larizza, C., Moglia, A., Stefanelli, M. M-HTP: A System for Monitoring Heart Transplant Patients. Artificial Intelligence in Medicine, 1992, Vol. 4, pp. 111-126

[25] Leibson, C.L., Naessens, J.M., Brown, R.D., Whisnant, J.P. Accuracy of hospital discharge abstracts for identifying stroke. Stroke, 1994, Vol. 25, pp. 2348–2355.

[26] Miller, R.A., Fasarie, F.E., and Mayors, J.D. Quick Medical Reference (QMR) for Diagnostic Assistance. MD Computing, Sept-Oct, 1986, Vol. 3(5), pp. 34-48.

[27] Mitchell, J.B., et al. What role do neurologists play in determining the costs and outcomes of stroke patients? Stroke, 1996, Vol. 27, pp. 1937–1943.

[28] Ngo, L., Haddawy, P., Krieger, R.A., Helwig, J. Efficient Temporal Probabilistic Reasoning via Context-Sensitive Model Construction. Computers in Biology and Medicine, 1997, Vol. 27(5), pp. 453-476.

[29] Rao, R.H., and Rao, R.B. Quality Assurance through Comprehensive Extraction from Existing (non-structured) Patient Records. Annual Conference and Exhibition, Healthcare Information and Management Systems Society (HIMSS), San Diego, California, Feb 9-13, 2003.

[30] Rao, R.B., Sandilya, S., Niculescu, R.S., Germond, C., and Goel, A. Mining Time-dependent Patient Outcomes from Hospital Patient Records. Proceedings of American Medical Informatics

Association (AMIA) Annual Symposium, San Antonio, Texas, November 9-13, 2002.
[31] Rao, R.B., Sandilya, S., Niculescu, R.S., Germond, C., and Rao, H. Clinical and Financial Outcomes Analysis with Existing Hospital Patient Records. Proceedings of the Ninth ACM SIGKDD International Conference of Knowledge Discovery and Data Mining (KDD), Washington DC, August 24-27, 2003, pp. 416-425.
[32] Regenstrief Institute. LOINC: Logical Observation Identifiers Names and Codes. http://www.regenstrief.org/loinc/ LONIC Homepage
[33] Sandilya, S., and Rao, R.B. Continuous-Time Bayesian Modeling of Clinical Data. Proceedings of the fourth SIAM International Conference on Data Mining (SDM), Lake Buena Vista, Florida, April 22-24, 2004.
[34] Shwe, M.A., Middleton, B., Heckerman, D.E., Henrion, M., Horvitz, E.J., and Lehmann, H.P. Probabilistic Diagnosis Using a Reformulation of the INTERNIST-1/QMR Knowledge Base: I. The Probabilistic Model and Inference Algorithms. Methods of Information in Medicine, October 1991, Vol. 30(4), pp.241-255.
[35] SNOMED International. SNOMED Clinical Terms. College of American Pathologists, http://www.snomed.org/
[36] Sonel, A.F., et al. What is the Most Efficient Data Extraction Method for Quality Improvement and Research in Cardiology?: A Comparison of REMIND Artificial Intelligence Software vs. Manual Chart Abstraction for Determining ACC/AHA Guideline Adherence in Non-ST Elevation Acute Coronary Syndromes. Annual Scientific Session of American College of Cardiology (ACC 2005), Orlando, Florida, March 6-9, 2005.
[37] Tsui, F.C., et al. Technical Description of RODS: A Real-time Public Health Surveillance System. Journal Am Med Informatics Assoc 10/5 (Sept/Oct) 399-408, 2003.
[38] United States Library of Medicine. MeSH: Medical Subject Headings. http://www.nlm.nih.gov/mesh/
[39] United States Library of Medicine. UMLS: Unified Medical Language System. http://www.nlm.nih.gov/research/umls/
[40] World Health Organization. Manual of the international statistical classification or diseases, injuries, and causes of death. World Health Organization, Geneva, 1977.
[41] World Health Organization. Report of the international conferences for the Tenth Revision of International Classification of Diseases. World Health Organization, Geneva, 1992.

About the authors:

Dr. R. Bharat Rao is the Senior Director of Engineering R&D, at the Computer-Aided Diagnosis and Therapy (CAD) Solutions Group in Siemens Medical Solutions, Malvern, PA. He received his B.Tech in Electronics Engineering from the Indian Institute of Technology, Madras, India, and his M.S. and Ph.D. (in machine learning) from the Department of Electrical & Computer Engineering, University of Illinois, Urbana-Champaign, in 1993. Dr. Rao's current research interests are focused on the use of machine learning and probabilistic inference to develop decision-support tools that can help physicians improve the quality of patient care and their efficiency. He is particularly interested in the development of novel data mining methods to collectively mine and integrate the various parts of a patient record (lab tests, pharmacy, free text, images, proteomics, etc.) and the integration of medical knowledge into the mining process. In 2005, Siemens honored him with its "Inventor of the Year" award for outstanding contributions related to improving the technical expertise and the economic success of the company. He also received the inaugural IEEE Data Mining Practice Prize for the best deployed industrial and government data mining application in 2005.

Dr. Sriram Krishnan is a Senior Staff Scientist at the Computer-Aided Diagnosis Computer-Aided Diagnosis and Therapy (CAD) Solutions Group in Siemens Medical Solutions, Malvern, PA. He received his Ph.D. in 1997 from the University of Michigan in the area of ultrasound beamforming. From 1997-2002, he worked at Acuson Corporation, performing research in medical ultrasound and beamforming. In 2002, he transferred to the CAD Solutions Group at Siemens Medical Solutions, where he has focused on two main areas: application of machine learning for detection of cardiac images, and application of machine learning techniques for medical data. Dr. Krishnan has numerous publications and 10 issued patents.

Dr. Radu Stefan Niculescu is a Staff Scientist at Siemens Medical Solutions in Malvern, PA. He currently works on data mining large medical records in the form of both structured and unstructured data. In 2005 he earned a Ph.D. degree from the Computer Science Department at Carnegie Mellon University, where his research was focused on incorporating domain knowledge constraints in parameter estimation of Bayesian Networks.