Improved Cardiac Care via Automated Mining of Medical Patient Records

R. Bharat Rao
Computer-Aided Diagnosis & Therapy Group
Siemens Medical Solutions, USA, Inc
bharat.rao@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 over 5,000,000 patients across the US.

1. Introduction

Cardiovascular Disease (CVD) is a global epidemic that is the leading cause of death worldwide (17 million deaths) [78]. The World Health Organization estimates that CVD is responsible for 10% of "Disability Adjusted Life Years" (DALYs) lost in low- and middle-income countries and 18% in high-income countries. (The DALYs lost can be thought of as "healthy years of life lost" and indicate the total burden of a disease as opposed to counting resulting deaths.)

Section 2 motivates our research by describing how current technologies are unable to combat the CVD epidemic. We begin by describing the cardiology burden faced today, with an emphasis on the United States, and discuss some of the factors contributing to the further deterioration of the CVD epidemic. A number of CVD treatment guidelines have been developed by health organizations to assist the physician on how to best treat patients with CVD. Yet adherence to these guidelines remains poor, despite studies overwhelmingly showing that adherence to these guidelines reduces morbidity and mortality, improves quality of life, and dramatically reduces healthcare costs.

One of the most promising ways to improve the quality of healthcare is to implement these guidelines within healthcare IT systems. Unfortunately, as we discuss in Section 2, due to the poor quality of healthcare data in medical patient records (the "Data Gap"), most healthcare IT systems are unable to provide significant support for CVD care: this is particularly true in chronic CVD situations which contribute to the majority of costs. Furthermore, this "Data Gap" is not likely to improve with the introduction of the Electronic Health Record (EHR),

and is further hampered by the lack of standards for clinical data, and the fragmented nature of the healthcare IT industry. Medical patient data is typically scattered in multiple sources and most of the information about the clinical context is stored as unstructured free text – these are dictated by physicians at different time points over the continuum of care delivered to the patient. It is important to note that the data is only "poor" from the point of view of automated analysis by computers; it is of high-enough quality for physicians to document and summarize the delivery of healthcare over multiple patient visits with different physicians. Many of the patients we have analyzed already have electronic data documenting their medical histories for more than 5 years (some going back even 20 years). Over time, exponentially increasing electronic data will be available for analysis for more and more patients. Analyzing this data will allow us to improve the healthcare of individual patients and also to mine new population-based knowledge that can be used to develop improved healthcare methodologies.

In Section 3, we introduce our solution for bridging the "Data Gap," the REMIND algorithm for Reliable Extraction and Meaningful Inference from Nonstructured Data. REMIND is a probabilistic framework for automatically extracting and inferring high-quality clinical data from existing patient records – namely, from patient data collected by healthcare institutions in the day-to-day care of patients, without requiring any additional manual data entry or data cleaning. We discuss the business decisions that influenced the design and development of the REMIND platform – namely, the need to rapidly deploy REMIND in diverse healthcare IT situations, for different clinical applications, and for different diseases, and to easily plug in different analysis algorithms for natural language processing and probabilistic inference. In Section 4 we briefly review the details of the REMIND algorithm [59]. Our goal is not to build a solution for a single application (e.g., implement a particular Heart Failure guideline) but to build a general solution that support multiple different applications for different diseases. Although REMIND was initially developed for automated guideline compliance, many other clinical applications are also supported by our solution, both at the individual patient level and the population level. These include automated methods for: therapy selection for individual patients [26]; patient identification for clinical trials; data extraction for research [67]; quality assurance; and relating financial and clinical factors [57].

In Section 5 we describe a number of successful deployments of our solution for the various applications listed above. This section illustrates that the REMIND platform can be deployed on the entire range of healthcare IT systems in use today, from relatively simple physician office systems, to some of the most complex hospital databases in existence. Further, our solution provides value in both chronic and acute care settings; can support all aspects of physician workflow (screening, diagnosis, therapy and monitoring) and healthcare administration; and provide research support, both in academic institutions and for ongoing pharmaceutical and medical device clinical trials [58][62]. The results provided have been rigorously verified by clinicians and scientists. In this paper we have focused solely on cardiac applications from clinical data. REMIND is currently deployed across multiple disease areas on a total of over 5,000,000 patients.

We review related research in the field of medicine and probabilistic inference in Section 6, We discuss some future applications of REMIND in Section 7, and conclude in Section 8 with our thoughts on further research.

2. Motivation

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.

2.1. 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 [9]. 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.

2.2. 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][27][28].

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] 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 include 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.

2.3. 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[20]. 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 [76]) and ICD-10 [77]. 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 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 [44] – 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.

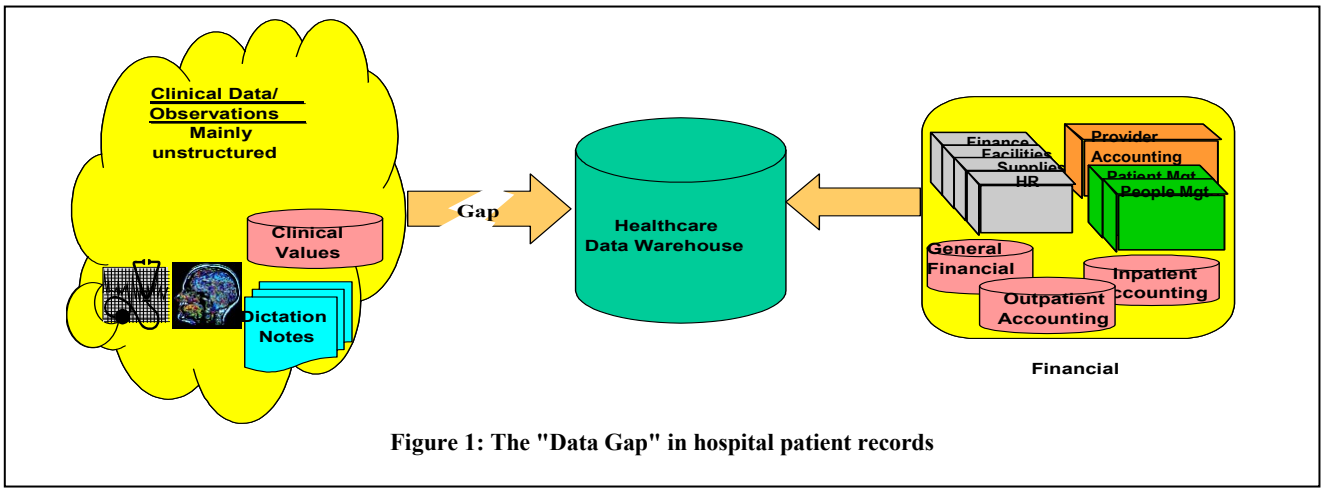

Figure 1: The "Data Gap" in hospital patient records

2.4. 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 [36], 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. Figure 1 illustrates the "*Data Gap*" in EHRs that prevents decision-support tools from assisting the physician in providing guideline-directed high-quality care to the patient.

2.5. 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 [57]. 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,[11][31][48] 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.

3. Automated Inference from Medical Patient Records

The "Data Gap" illustrated in Figure 1 explains the inability of automated decision-support systems to assist the physician in providing high-quality care to the patient. As noted earlier, a computer system could not reliably answer from most electronic patient records the question, "Should this patient be given beta-blockers?" However, if the same question is posed to a physician, it is very likely that they (given sufficient time) can answer the question correctly. This means that the information needed to answer the question does exist (or can be inferred) from the electronic patient data, and the "data gap" exists only for computer systems that requires data to be entered in a structured form for analysis.

3.1. Exploiting redundancy in patient records

Patient records typically contain multiple redundant pieces of information about a disease, or indeed any medical situation associated with the patient. For instance, a doctor (or a computer program) could infer that a patient had a particular diagnosis (for example, *is diabetic*) in many different ways and from different data sources:

- Billing codes (ICD-9 of 250.xx for diabetes)
- A transcribed free text dictation that identifies a diagnosis (History and Physical, Discharge summary)
- Symptom (Blood sugar values > 300 in labs)
- Treatment (Insulin or oral anti-diabetic administration in Pharmacy)
- A complication associated with disease (e.g., diabetic nephropathy)
- Other relevant information about the diagnosis (e.g., some steroids elevate blood sugar)

3.2. High-level Requirements

In the above example, the diabetes diagnosis can be inferred from many different data sources and by different methods. For instance, natural language processing could help extract information from free text extraction. A critical component in any successful system must be the use of medical domain knowledge to draw inferences from data.

Fundamentally, any solution must (a) be patient-centric, (b) combine information extracted from all available patient data, and (c) be guided by medical domain knowledge. It follows, therefore, that the system must be able to handle and reason with information in different formats, for instance, doctor's notes in free text and financial, pharmacy, and lab databases from the diabetes example (and in future applications: images, proteomic and genomic data). Further, this information may be contradictory (or indicate the presence or absence of diabetes to varying degrees). Therefore, rather than relying on individual data elements to be extracted correctly, probabilistic reasoning is needed to deal with missing, incorrect, and imprecise information in the clinical repository. Finally, a patient history is not static – symptoms, diagnoses, and treatments may all vary, and temporal inferences will be needed to deal with this added complexity.

There are some additional business drivers to consider. First, as medical guidelines change periodically, the medical knowledge associated with our solution will need to be easily modified. Second, once we have a guideline implemented at one institution, we may wish to implement other guidelines at the same institution with minimal effort. Third, as many institutions implement the same guideline, we may wish to transfer our guideline-solution from one institution to another with minimum effort. Fourth, our solution must scale easily from hundreds of patients to millions of patients. Fifth, our solution must handle the data privacy issues inherent to medical data [72].

Flexibility is a key design requirement – our system must be able to easily incorporate new algorithms to meet the needs of future applications and leverage new technologies, for example, new NLP or probabilistic inference methods. Finally, we should be able to support other decision support applications, both at a patient (e.g., patient identification for clinical trials) and population (e.g., quality assurance) levels.

3.3. System Overview

The REMIND algorithm consists of 3 steps. In the *extraction step*, information is extracted from every part of the patient records in isolation, e.g., from every row in a database table, from every phrase in every sentence. Obviously, several thousand such pieces of information can be extracted from a single patient record, many of which may be incorrect and/or

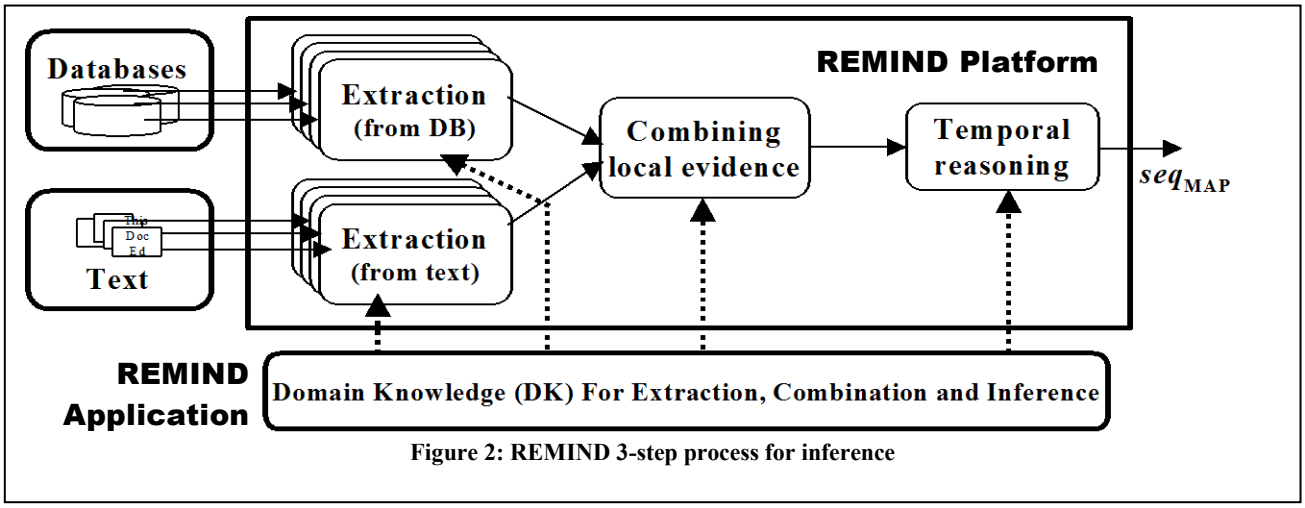

Figure 2: REMIND 3-step process for inference

inconsistent, due to errors in the original data or due to the extraction algorithms (for instance, natural language processing is by no means perfect).

In the *combination step*, all observations about a single variable at a single point in time are combined to produce a single observation or a distribution over many different values.

In the *inference step*, different variables are combined across time to infer the values of the variables needed for determining guideline eligibility or compliance. All 3 steps are configured by the medical domain knowledge needed by REMIND to extract the necessary information and arrive at the desired conclusion.

This algorithm is illustrated in Figure 2. The "REMIND Platform" contains all the code needed to implement the 3 steps described above. A key design decision was to implement the medical domain knowledge to drive REMIND as external modules (the "REMIND application") that plugs into the Platform. These REMIND applications are implemented as XML files which configure how the REMIND platform processes data. Therefore, to switch from implementing, say, a heart failure guideline, to recruitment for a coronary artery disease clinical trial, to monitoring radiation treatments for breast cancer, will not require writing any code – each REMIND application will simply consist of a configuration file.

The REMIND domain knowledge to configure the platform is of two types. First, institution-specific domain knowledge describes how the institution's data is organized, where each kind of data is found and under what format, and how to retrieve all data associated with a patient. The second type of knowledge is application-specific. Note that application-specific domain knowledge can be transferred easily from one institution to another – minor retuning would be needed to deal with the differences in the types of data and data quality at each institution, but in general the process of moving an application from one institution to another is low-effort. Similarly, once the REMIND platform is configured for a particular institution (i.e., the institution specific domain knowledge has been created) it also relatively straightforward to implement new application-specific domain knowledge upon the existing configuration.

4. The REMIND Algorithm

In this section we describe one specific application of REMIND (Reliable Extraction and Meaningful Inference from Nonstructured Data). Our goal is to infer disease progression; whether a patient has a particular disease at different points in time, and if so what stage (degree of severity) of the disease. Our medical knowledge about the disease includes knowledge about legal disease sequences – for instance, it may be legal to go from stage 0 to stage 1, but not from 1 to 0, and also information about expected transition times (gathered from the medical literature and survival curves from clinical trials) from one disease stage to another.

Our approach to inference with this multi-source data is to model the data as arising from a generative process, and combine prior knowledge about this process with observations for a specific patient using Bayesian techniques. Patient data is collected in a medical institution at arbitrary points in time (i.e., not at regular intervals but at patient visits only), and these sampling instants vary from patient to patient. Hence, we model the processes of progression of patients' diseases and the collection of this data as continuous time processes that may be sampled at arbitrary instants. We consider a model wherein a patient has a

state (for the disease of interest), and observations about the state and related variables are stored in and may be collected from various data repositories.

4.1. Problem Definition

Let **S** be a continuous time random process taking values in Σ that represents the state of the system; note that **S** may be a combination of multiple variables. Let $\mathbf{T}=\{t_1, t_2, \ldots, t_n\}$, where $t_i<t_{i+1}$, be the n "times of interest when **S** has to be inferred. Let $\mathbf{S}_i$ refer to the sample of **S** at time $t_i \in \mathbf{T}$. Note that **T** and n can vary for different realizations of the process.

Let **V** be the set of variables that depend upon **S**. Let **O** be set of all (probabilistic) observations for all variables, $\boldsymbol{v} \in \mathbf{V}$. Let $\mathbf{O}_i$ be the set of all observations "assigned" to $t_i \in \mathbf{T}$; i.e., all observations about variables $v \in \mathbf{V}$ that are relevant for this time-step t_i. Similarly, let $\mathbf{O}_i(v)$ be the set of observations for variable v "assigned" to t_i.

Let *seq* be a random variable in Σ^n; i.e., each realization of *seq* is a specific (legal) sequence $< \mathbf{S}_1, \mathbf{S}_2, .. \mathbf{S}_n >$.

In the case when we are interested only in the value of a variable at a point in time (e.g., in the AMI example, we simply wish to know if the patient really had an AMI), our goal is to estimate:

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

When we wish to track the patient's progress over time, our goal is to estimate the most likely state sequence, seq_{MAP}, the maximum a-posteriori estimate of *seq* given **O**:

$$seq_{MAP} = \arg\max_{seq} P[seq \mid O]$$

4.2. Overview of Approach

We view **S** as a continuous time Markov process from which we observe non-uniform samples. Our implementation of REMIND assumes that **S** is a stationary Markov process, whereas variables, $v \in \mathbf{V}$ that depend on **S** have conditional distributions (on the parent variable) that are non-stationary. However, our framework can be extended to handle even non-stationary Markov processes.

REMIND's 3-step process that estimates the distribution of the variable of interest V_{MAP} (or seq_{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 $v \in \mathbf{V}$. 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 $o \in \mathbf{O}$ is assumed to be potentially incorrect.

(2) **Combination** step: each observation is assigned to one time of interest, $t_i \in \mathbf{T}$. Then each state, $\mathbf{S}_i$ is estimated from all of the observations $\mathbf{O}_i$.

(3) **Inference** step: the inferences are propagated across time and the posterior probabilities for each variable computed.

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

4.3. 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 $v \in \mathbf{V}$, 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.

4.4. Combination & Inference

The primary focus is estimating what happened to the system (*e.g.,* disease evolution) across the duration of interest. Hence, a natural abstraction of the problem is to look for the best estimate of the sequence of system states across time, and the maximum *a posteriori* (MAP) estimate is the one whose probability is maximal. Hence, given the observations that we have extracted, we would like to estimate the *a posteriori* probability of each legal state sequence and pick the most probable one. This can be done in two steps, the first of which is combination of observations at a fixed

point in time and the propagation of these inferences across time.

We use a Markov Model to estimate the evolution of the patient's state. As the observations about patients are spaced non-uniformly across time, the standard discrete-time Markov approximations are not necessarily justifiable. In order to overcome this shortcoming, we model the process of evolution of the patient state as a continuous-time Markov process from which we get to observe non-uniform samples. More specifically, the parameters we need to model are the dwell time in each state and the transition rates from each state to every other. In our current implementation, we consider the state to be a stationary Markov process whereas the other variables that depend on it can have conditional distributions that are non-stationary. Our framework, however, can be modified to handle even the case of non-stationary state processes.

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 at a given point in time 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:

$$\Pr[seq \mid Obs] \propto \Pr[S_0] \cdot \prod_{i=2}^{n} \Pr[S_i \mid S_{i-1}] \cdot \prod_{i=1}^{n} \Pr[Obs_i \mid S_i]$$

$$\propto \prod_{i=2}^{n} \frac{\Pr[S_i \mid S_{i-1}]}{\Pr[S_i]} \cdot \prod_{i=1}^{n} \Pr[S_i \mid Obs_i]$$

We model the relationships between the set of all variables of interest using a Bayesian Network, which is used to infer the posterior distributions of all the variables at a given point in time given all the information at that time. For inference across time, we may now use a standard dynamic programming based approach (e.g. the Viterbi algorithm [56]).

Because we model the state process as being Markov, we have the following equation that connects the *a posteriori* probability of a sequence of samples of the state process given all the observations to the temporally local *a posteriori* probability of the state given all observations at each time instant.

$$\Pr[v \mid O_i^1(v),...,O_i^k(v)] \propto \Pr[v] \cdot \prod_{j=1}^{k} \Pr[O_i^j(v) \mid v] \propto \frac{\prod_{i=1}^{k} \Pr[v \mid O_i^j(v)]}{\Pr[v]^{k-1}}$$

4.5. Domain Knowledge in REMIND

This includes y the state **S** (the variables we wish to infer), **V** (and the data sources for each variable), institution-specific domain knowledge which describes the institution's data structure and access mechanisms, extraction knowledge (e.g, NLP and database queries), dependencies between **S** and **V**, and the dwell times and transitional probabilities.

Despite the seeming complexity, most of the domain knowledge (DK) in REMIND is fairly simple. The clinical application defines S, the variables in V can be elicited fairly easily, and institution-specific knowledge is a one-time implementation effort across many applications at that institution. DK for extraction can be fairly complex, but we have investigated ways to learn this from data. In other medical Bayesian applications [5][38][47], the actual probability values for the dependencies within S and 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 [62]. Similarly, we roughly estimate dwell times and transition probabilities from survival curves in medical literature. Experiments in [62] also show that REMIND is also insensitive to variations in the temporal parameters.

That said, obviously a big part of the success of any application is the careful tuning that must be done to ensure success. Because of the nature of the application and its potential impact, even if REMIND is inferring information at very high accuracies, there is often value in further improving the end result. All REMIND applications are validated at multiple institutions before release. (The actual DK constitutes the entire REMIND application and is proprietary.)

5. Real-world deployments of REMIND

Here we describe actual deployments of REMIND in various clinical scenarios. Each deployment is characterized by the following variables:

- Name of Institution
- Acute or clinical setting
- IT System, # of physicians and patients supported.
- Population analyzed (may be subset of total)
- Goal (Additional secondary goals are described in parenthesis.)
- Electronic Data Available – at a minimum this will include Billing, demographics, and transcribed free text. Additional specialized databases and free text specialist reports may be available.

5.1. Process Control for Diabetics with AMI

Institution: University of Pittsburgh Medical Center
Setting: Acute Care
IT System: Large hospital IT system supporting several hundred physicians, multiple specialities, multiple locations, 2 Million patients.
Population Analyzed: ER Patients admitted with a diagnosis of Acute Myocardial Infarction (heart attack) in a 3 year period (3000 patients).
Goal: Guideline compliance: proper monitoring of diabetics who had AMI. (Also, clinical and financial outcomes analysis)
Electronic Data: Billing, demographics, pharmacy DBs and transcribed free text (history & physical, progress notes, discharge summaries, ECG and ultrasound reports) from the Medical Archival System (MARS)

Despite the best quality of care provided to cardiology patients, it is inevitable that some people will still face acute episodes and will be rushed to hospital. One of the most common such problems is an acute myocardial infarction (AMI), or heart attack. In fact, for many people, a heart attack is the first symptom that a patient even has cardiac problems. For proper care of patients, it is important that patients who are brought to an emergency room are first properly diagnosed that they have an AMI. Just as critical is to ensure that patients who are diagnosed with an AMI are also assessed for diabetes, and to ensure that their blood sugar is monitored and treated properly, that is, given proper glycemic control (AMI patients with diabetes have much better outcomes if the diabetes is also treated). However, as discussed earlier, these billing codes are inaccurate from the clinical point of view, and are used primarily for reimbursement. Finally, the issue of whether diabetic patients are treated properly for glycemic control cannot be evaluated from structured data alone, but must be inferred from the clinical record.

To address these issues, a study was conducted with the University of Pittsburgh Medical Center[57]. The main purpose of this study was to answer the following three questions:

1. Did patients who were being admitted to the UMPC Intensive Care Unit (ICU) with AMI really have an AMI?
2. Did these patients, who had an AMI, also have diabetes?
3. If the patient had an AMI and diabetes, were they given proper glycemic control?

To address these questions, data was collected from patients who were admitted to the UPMC ICU with a principal diagnosis of AMI (that is, with a principal ICD-9 billing code of 410.xx) in the year 2001. From over 1000 records, 52 were selected randomly.

Next, clinical definitions of AMI and diabetes were provided from internationally accepted criteria [71] and coded into REMIND. The diagnosis of AMI depends on the unequivocal presence or absence of a combination of three factors upon which the diagnosis rests: symptoms of cardiac pain, abnormalities in the electrocardiogram (ECG), and enzymes released by injured heart muscle. The degree to which those factors meet criteria, individually and in combination, determine the certainty of the AMI diagnosis ("definite", "probable", or "possible"). Next, each factor was further defined. For example, for various enzymes released by injured heart muscle, such as troponin, CPK, and CK-MP, various ranges corresponding to abnormal, equivocal, and normal ranges were defined. Similarly, ECG changes and cardiac pain were further defined. For these two cases, the clinical definitions had to be inferred from free text. Diabetes could be inferred either from mention by the physician in their reports, or from either administration of insulin or other oral agents specific to diabetes, or from the presence of lab records showing 2 random blood sugars above 300 mg/dl. Glycemic control was assessed by monitoring blood sugar levels for these patients in the hospital. REMIND was run on 52 patients to answer the 3 questions.

A physician at UPMC, blinded to the results from REMIND, then reviewed the patient record manually for these 52 patients, and then answered the same 3 questions listed above. In making a determination of AMI and diabetes, the physician looked at the entire patient chart (including portions not available to REMIND), and made a clinical diagnosis. The reason was that some parts of the medical record were not in electronic form, and therefore were inaccessible to REMIND. Therefore, the conclusions reached were independent of the domain knowledge and rules provided to REMIND.

Using the physician reads as ground-truth, Table 1 compares the hospital billing codes with Ground Truth, and also REMIND with ground truth for diagnosis of AMI and Diabetes Mellitus (DM). Whereas the diagnostic accuracy of the coded information is only 83% for AMI and 90% for DM, results based on

Table 1. Accuracy for AMI & Diabetes for 52 patients

Diagnosis	ICD-9 CODES			REMIND		
	FP	FN	Acc	FP	FN	Acc
AMI	0	9	83%	1	2	94%
DM	1	4	90%	0	1	98%

REMIND are much closer to Ground Truth (90% and 95%, respectively). Of the 52 patients coded as AMI, only 43 actually fit the MONICA criteria for AMI (Definite, Probable or Possible). In comparison, REMIND correctly identifies 8 of the 9 patients with No AMI. Of the 52 patients, 19 had diabetes, based on the Ground Truth. REMIND makes only one diagnostic error, compared to 5 in the ICD-9 codes.

Further, REMIND was able to assess whether a patient was given proper glycemic control. It was found that of the 53 patients analyzed, 13 patients had diabetes. In the critical period 24 hours after admission, of these 13, 6 had excellent control of their blood sugars, 1 had moderate control, 1 had poor control, and surprisingly 5 patients were not assessed for blood sugar at all. For their entire stay, 4 had excellent control of their blood sugars, 5 had moderate control, 3 had poor control, and 1 patient was not measured at all. Note that such clinical assessments of process such as glycemic control would be impossible by just looking at billing codes, and would be extremely time consuming (and expensive) for a physician to perform. For instance, the manual chart review averaged 30 minutes per patient, while REMIND ran in mere seconds over all 1000 patients.

Additional Results from the UPMC Analysis:

As mentioned earlier, one of the advantages of our solution is that once it is implemented on the institution's data, it is very easy to extract additional information to support other clinical applications.

One of the most valuable tools available to the hospital administrator is outcomes analysis – namely, analyzing the available data, slicing and dicing it different ways using different database and OLAP tools, to determine the impact of different variables on outcomes. The problem, however, is that the only available data for analysis is the financial data (with diagnosis and procedure codes), and as we show that analysis can lead to incorrect *clinical* conclusions [57].

Table 2 compares the impact of incorrect coding on two key financial outcomes: Length of Stay (LOS) and Charges. (These are good surrogates for quality of care, because in general patients with better care will have shorter hospital stays, and fewer complications, leading to lower charges.) LOS derived from coded information in all 52 patients coded, as having an AMI is about 0.5 days less than the Ground Truth. (This is because 9 patients who actually don't have an AMI, but have been incorrectly coded as having an AMI, are included in computing the Average LOS) Table 2 shows that using the diagnosis extracted by REMIND achieves much greater accuracy, being only 0.1 days off the truth. Similarly, coded information leads to an underestimation of charges incurred in AMI patients by about $5000, whereas REMIND is only off by $1500.

Table 2. Outcomes on AMI patients

Outcomes	ICD-9 Codes	Ground Truth	REMIND
LOS (days)	7.54	7.93	8.05
Charges ($)	$89673	$94688	$96379

There is an additional subtle problem beyond the under-estimation of poor outcomes. Suppose the administrator is considering hiring a diabetic nurse for the ER for the purpose of providing better treatment to diabetics – the next step would be to analyze the available data to determine exactly how much poorer the outcomes were for AMI diabetics versus non-diabetics, and then determine if the potential impact would justify the resources needed to increase staffing.

Table 3. Impact of diabetes on Financial Outcomes

Outcomes	Patient-type	CODERS	Truth	REMIND
LOS (days)	Diabetics	7.13	11.00	11.67
	Non-diabetics	7.70	6.60	6.60
Charges ($)	Diabetics	70,854	105,100	114,887
	Non-diabetics	97,302	90,175	88,976

Table 3 Shows that the errors in coded information regarding AMI and DM compound the underestimation of both LOS and Charges in diabetics with AMI. Thus, coded information would lead to the conclusion that LOS for diabetics was 0.6 days *less* than for non-diabetics, and charges incurred were *lower* by ~$26,000. In actual fact (Ground Truth), diabetics stayed an average of ~4.5 days *longer*, and incurred an additional ~$15,000 in *extra* charges. REMIND was much closer to Ground Truth, correctly identifying that diabetics both stayed longer (by ~5 days), and incurred higher charges (by ~$21,000). Table 3 demonstrates the value of REMIND in correctly identifying specific diagnostic categories of patients for outcomes research. They also show the hazards of plotting cost-saving strategies and resource allocations based purely on electronically coded information. For instance, ground truth (and REMIND) reveals exactly the opposite Conclusion about LOS and Charges for diabetics with AMI. This establishes the utility of REMIND, which paralleled Ground Truth, in correctly identifying and analyzing outcomes in a large cohort.

In conclusion, this study showed that REMIND was able to successfully aid in both diagnosis of AMI and diabetes, and in assessing the quality of care for these patients in at least one aspect (glycemic control) in the

acute (ICU) environment. The results showed that diagnosis was significantly superior to the use of structured data (i.e. billing codes), and allowed for fast assessment of process quality that could not be assessed using structured data alone.

In many of the deployments that follow, REMIND is used for clinical applications beyond the primary one described. In the interests of brevity, we restrict the description to the primary application.

5.2. Therapy recommendation for patients at risk for Sudden Cardiac Death (SCD)

Institution: South Carolina Heart Center
Setting: Chronic Cardiac Care
IT System: Physician practice IT system supporting 24 cardiologists, single location, 61,027 patients.
Population: All patients.
Goal: Guideline compliance & Therapy Recommendation: identify patients who are at risk for Sudden Cardiac Death, and assess them for defribillator implantation
Electronic Data: Billing and demographics DBs and transcribed free text (history and physical reports, physician progress notes, and lab reports).

Cardiac patients who have had a prior MI are at risk for sudden cardiac death (SCD). Each year, SCD claims the lives of 300,000 Americans. In 1997, a trial was conducted to study the efficacy of using implantable cardioverter defribillators (ICDs) to help prevent sudden cardiac death [49]. The Multicenter Automatic Defibrillator Implantation Trial II (MADIT II), showed that patients who had a prior MI and had low ventricular function, had their 20 month mortality rate drop from 19.8% to 14.2%, a significant 31% reduction in mortality, when an ICD was implanted. The trial was stopped in 2001, with a recommendation to implant ICDs in these patients [50].

Afterwards, there was a need to rapidly identify patients who met these criteria, and evaluate them for ICD implantation. Ordinarily, this could be done in one of two ways. One approach would be to review several thousand patient records manually to assess whether a patient was eligible for an ICD. This approach would be extremely time-intensive and laborious. Another approach could be to evaluate patients as they come in for regular check-ups with their cardiologist. Unfortunately, this would result in needless deaths as patients would only be evaluated if they had a check-up, not to mention the possibility that this new guideline may not be evaluated among the several hundred that the physician must consider. Therefore, it is critical to rapidly assess whether patients were eligible for ICDs, as every month of delay would result in an increased chance of SCD.

Working with the South Carolina Heart Center (SCHC), we implemented REMIND to identify patients who were eligible for an ICD as per the MADIT II study, and who had not yet received an ICD. A total of 61,027 patients were analyzed from the practice for eligibility of an ICD per MADIT II guidelines. REMIND identified 383 patients of the 61,027 as being eligible for an ICD. The total processing time for REMIND for all 61,027 patients was 5 hours on a Pentium M 1.4 GHz laptop.

These 383 patients were mixed with 383 patients randomly drawn from the rest of the population (i.e., MADIT-II ineligible as per REMIND), and 150 of these patients were randomly re-selected. An electrophysiologist manually reviewed the charts for each of these 150 patients to assess MADIT-II eligibility. The reviewer was blinded to the results of REMIND at the time of this determination.

The concurrence between the REMIND system and the manual chart review for eligibility for MADIT-II trial was 94% (141/150). The sensitivity and specificity of REMIND to identify patients were 99% (69/70) and 90% (72/80) respectively. "*Conclusion: REMIND can automatically identify patients who meet definable clinical guideline inclusion/exclusion criteria with a high degree of accuracy. REMIND could be used to improve quality of care and outcomes for patients at risk for cardiovascular disease*."[26]

5.3. Guideline Adherence Study for Patients with Non-ST Elevation MI

Institution: Veterans Health Administration (VHA) Hospital, Pittsburgh
Setting: Acute Care
IT System: Large hospital IT system supporting several hundred physicians, multiple specialities, multiple locations, 7 Million patients across the US.
Population Analyzed: ER Patients admitted with a diagnosis of unstable angina or non-ST elevated MI over the last 3 years (1400 patients).
Goal: Guideline compliance with ACC guideline.
Electronic Data: Tremendous amounts of structured and unstructured information (see below).

The Veterans 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 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.

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 [10] (http://www.acc.org/clinical/guidelines/unstable/unstable.pdf.). (Another type of myocardial infarction, MI with ST elevation, is treated differently.)

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 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 [10]. 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 v0.5 took 4.5 hours to extract the values of the 4 variables (see Table 11) 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 [67].

Table 4. Accuracy of REMIND vs. trained medical nurse for guideline compliance

	ACCURACY (%) N=327	
TREATMENT	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 4 compares the accuracy of REMIND and manual abstraction for each of the 327 patients. That is, for each patient, this analysis showed what percent of patients were accurately assessed using REMIND and manual abstraction (using the adjudication as a gold standard). Table 4 shows that REMIND works at least as well as manual abstraction in identifying patients who were treated per guidelines for non-ST elevation MI. Note, that the task is different from that shown in Table 1. There, the task was to extract ICD-9 codes, and the comparison was with abstractors who had no medical knowledge. Here, the task is to extract clinical information, and we compare REMIND to a trained nurse with expert medical knowledge; this task is much harder because, as discussed earlier, these medical inferences require subtle inferences to be drawn, particularly for determining contra-indications.

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.

The Veterans Health Administration (VHA) operates 172 medical centers, more than 800 ambulatory care clinics, and provides healthcare for 7 million veterans making it the largest integrated health system in the United States. Further, the VHA mandates that the databases in all institutions are identical. This means that now that REMIND can work successfully at VHA Pittsburgh, it should apply to all the VHA medical centers, with no changes in domain knowledge. This can dramatically increase the impact of our system at the VHA.

5.4. Patient Identification for Clinical Trials

Institution: Nebraska Heart Institute (NHI)
Setting: Chronic Care
Physician practice IT system supporting 32 cardiologists, 4-5 locations, 208,000 patients.
Population: All patients.
Goal: Automatically identify patients who are eligible for clinical trials
Electronic Data: Billing and demographics DBs and transcribed free text (history and physical reports, physician progress notes, and lab reports).

Introducing medical advances into practice is a risky endeavor. Clinical Trials are a critical component to managing this risk. Clinical Trials are required for drugs and medical devices (as an aside, including machine learning-based software to perform Computer-Aided Detection[14][21][61]). Each clinical trial has its own inclusion and exclusion criteria (www.clinicaltrials.gov), which are used to identify eligible patients, and then enroll them, for the trial.

Clinical Trials are very expensive. Pharmaceutical companies spend $20 Billion/yr in the US [43]. Patient recruitment is roughly 10% of a trial's costs ($3.6B in 2002 alone). Difficulty in recruiting patients has been identified as the top cause of delay in Clinical Trials [42]. A key factor in determining the length and cost of a trial is the time it takes to identify and sign up eligible patients for a trial. (For a blockbuster drug, every day's delay in releasing the drug, costs the company, $2Million/day [33].)

REMIND has been used at NHI to successfully identify eligible patients for two actively-recruiting clinical trials. Some "recruitability criteria" (beyond the trial criteria) were added into the analysis; there criteria are used by NHI's trial coordinators to further identify which eligible patients are more likely to agree to participate in the trial (depending on the trial, these criteria could include physical fitness, geographic distance from a hospital, age, co-morbidities, etc.) A key metric for success is the *eligibility rate*, namely, the fraction of eligible patients from all patient records examined. Note that eligibility is determined by examining the entire patient record (including non-electronic data). The traditional method for trial recruitment (other than advertising for patients) is manual examination of patient records by the trial coordinators. Automated identification of eligible patients will help NHI to recruit more patients, with less effort, and less time, resulting in faster and cheaper trials.

The first trial is a medical device trial sponsored by a major device manufacturer. This trial has 18 inclusion/exclusion criteria with a target enrollment of 20 patients at NHI. From NHI's population of 208,000, REMIND identified 363 likely eligible patients. Adding the "recruitability criteria," reduced this list to 31 patients. 29 of these patients were then reviewed by the trial coordinators for eligibility (2 patients could not be easily validated as the appropriate records were at an offsite location.) Of the 29, 18 (62%) were confirmed eligible. Of the 11/29 ineligible patients, 5 were determined to be ineligible from (non-electronic) data available to the coordinators, but not to REMIND.

REMIND has performed even better on an actively recruiting drug trial sponsored by a major pharmaceutical company. This trial has 14 inclusion/ exclusion criteria with a target enrollment of 200 patients at NHI. From NHI's population of 208,000, REMIND identified 2,538 likely eligible patients, of which 312 also met the additional recruitability criteria. Of the 286 patients were validated by NHI, 270 (94%) were confirmed eligible by NHI's trial coordinators. Of the 16/286 ineligible patients, 5 were determined to be ineligible from (non-electronic) data not available to REMIND. (In another process, NHI has contacted 215 of these 270 to participate in the trial; 35 of these patients have accepted.)

To put *eligibility rates* of 62% and 94% in perspective, an *eligibility rate* of 10% would be considered extremely good.

5.5. Quality of Care Analysis for Multiple Institutions

Institution: NHI & SCHC (described earlier)
Population: All 270,000 patients from both institutions
Goal: Automatically extract quality of care information for Heart Failure and Amiodarone

Within the realm of CVD, heart failure imposes the heaviest burden on the healthcare system. In the United States, approximately 5 million people have heart failure, with 550,000 more diagnosed each year. Heart failure results in 12-15 million physician office visits and 6.5 million hospital days each year, and accounts for over 50,000 deaths yearly. Heart failure is a chronic disease with no cure. Patients who are diagnosed with heart failure may live for 10-15 years after the initial diagnosis. Many of the hospital stays associated with heart failure 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 [16]. 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, as described earlier, 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, MRI, 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 [25]. 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.

The previous applications have presented REMIND results at a single institution. Recall that one of the key requirements in REMIND was to allow the same applications to be run at multiple institutions with little or no retuning. NHI and SCHC have very different healthcare IT systems (different vendors, different kinds of data stored, different databases, different data formats), but being cardiology physician practices, have very similar needs regarding quality of care.

REMIND was run at both institutions' data with virtually no change in domain knowledge. Although validation is ongoing for amiodarone, initial validation results indicate that – at least for cardiology practices – domain knowledge developed at one institution (NHI) retains an equivalent level of performance when transferred to another institution (SCHC). This is critical for rapid deployment. 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.

6. Related Research

From the earliest days of computing, physicians and scientists have explored the use of artificial intelligence systems in medicine [41]. The original hope was the such systems would become physicians in a box, possibly even surpassing systems in diagnostic tasks [12][61]. Today, the research focus has changed from just diagnosis to support the continuum of healthcare via clinical decision support systems (see, [18][75][63][64]). The fundamental goal of such systems is to reduce costs, improve the quality of care and patient outcomes (see [54] for a summary). Although the impact of such systems on a national scale has been muted, the biggest impact has been made by computerized physician order entry systems that have been shown to reduce medication errors and improve patient outcomes. These systems are based entirely on structured data, and alert the prescribing physician about potentially dangerous drug-drug and drug-disease interactions [24][68].

Another long-standing area of computer research in medicine has been the automated interpretation and analysis of medical images [55]. In the recent past, many such systems have moved out of the realm of research labs into clinical practice, mostly as Computer-Aided Diagnosis systems [21][13] that assist the radiologist in identifying potential cancers in medical images [61][14][8]. We are currently expanding the REMIND platform to include images, and are developing therapy-assistance tools that will help the physician make therapeutic decisions, particularly in the treatment of lung cancer.

From the computer science perspective, our work draws heavily on earlier work on Bayesian networks and graphical models (see [29][34] 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 [47]) 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 [65] 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 [38] 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 [19] 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 [52] 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 [39] use structured information to predict probabilities of survival for ICU patients. Other related research [32][37][40] deals with representing temporal data and enforcing temporal integrity.

As discussed earlier, a fundamental premise of REMIND, is to exploit the redundancy in the medical record. Our initial implementations achieved very high performance despite using very simple methods from computational linguistics. Although Natural Language Processing (NLP) is not the focus of this work, we are leveraging the rich body of research in this area [45]. Consider the falling examples, all drawn from doctors' dictations, that contain the word Aspirin: "Patient is on Aspirin 2 mg daily; Patient was off Aspirin for a while and then resumed; Dr Smith considered Aspirin 2 mg for him; He stopped taking Aspirin post operative; Use of Aspirin 2mg cannot be excluded; Aspirin on Mondays and Wednesdays; He wants to discuss possible contraindications of his Aspirin dose; Dr Jones ruled out Aspirin for him." Clearly simple look-up for the word "Aspirin" will fail to identify all patients currently taking Asprin. Friedman et al [23] discuss the potential of using NLP techniques in the medical domain, and also provides a comparative overview of the state-of-the-art NLP tools applied to biomedical text. [17][23][30] provide a survey of various approaches to information extraction from biomedical text including named entity tagging and extracting

relationship between different entities and between different texts. Clinically relevant observations and features can be extracted with much better accuracy since documents (EMRs) do not have to be treated as bag-of-words ignoring their structure and semantics altogether. For instance, Taira et al [69][35] have done research on automatic structuring of radiology reports. Of direct relevance is the analysis of doctors' dictations by Chapman [15] which identifies the 7 most common uses of negation in doctors' dictations. Augmenting our aliases with a general lexical reference [22] or a medical language dictionary (SNOMED [66]) should improve performance. Furthermore text-mining research to identify relevant documents [46][53] may help eliminate irrelevant documents that are mixed in with doctors' dictations. DISCOTEX [51], like REMIND, extracts information from text, and integrates it via data mining. DISCOTEX focuses on learning rules, whereas REMIND uses domain knowledge for data mining. REMIND is implemented so that text extraction and NLP (and better reasoning) methods can be easily plugged into REMIND.

7. Next Steps

Our immediate next step is to incorporate REMIND into the point of care. Initially, REMIND could be used to alert research coordinators when a potentially eligible patient (for their trial) is being seen by another clinician somewhere else in the clinic. Eligible patients are most likely to enroll if approached at the hospital (since all tests, examinations, and paperwork can be completed on-site, instead of making a separate trip, as would be the case if approached on the phone).

REMIND can also provide point of care support to the physician, for instance, by evaluating the patient against all ongoing open trials and guidelines, and flagging the eligible ones. To this, we are installing REMIND on a multi-million patient database for a large academic medical center.

Other interesting applications include disease surveillance, epidemiological studies, bioterrorism surveillance, and outbreak detection. The RODS [70] (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 [60], MeSH [73], and RxNORM), controlled vocabularies (SNOMED CT [66]), and ontologies (UMLS [74]). 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.

8. 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 (the "Data Gap" in Figure 1). 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).

Although our system works with poor data and is not an NLP system, better data and better data extraction methods only improve our performance. REMIND is designed to allow multiple analysis algorithms to be plugged into the platform.

Our goal is to build a general framework to perform inference from medical patient data for a variety of applications and diseases. REMIND provides value in different clinical settings for different diseases. Our system has been designed to support quickly adding data from new institutions, and creating new applications (the domain knowledge files).

The key barrier for IT systems to support automated guideline compliance is the lack of high-quality clinical data collected in day-to-day care. Once REMIND automatically extracts this data, then many other applications are enabled, including: trial recruitment, quality assurance, therapy monitoring, etc.

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 current deployed on a rapidly growing population of over 5,000,000 patients.

9. Acknowledgements

This sy stem 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, I am 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.

Any successful commercial product relies on the dedication of a team. I am indebted to my Siemens colleagues: Sathyakama Sandilya, Ph.D., Ingo Schmuecking, MD, William Landi, Ph.D., Sriram Krishnan, Ph.D., Radu Stefan Niculescu, 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:

10. 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] G. O. Barnett, J. J. Cimino, J. A. Hupp, and E. P. Hoffer, "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] C. Benesch, D. M. Witter Jr, A. L. Wilder, P. W. Duncan, G. P. Samsa, D. B. Matchar, "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] L. Bogoni, et al, "CAD for Colonography: A Tool to Address a Growing Need", (to appear in) *British Journal of Radiology*.

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

[10] Braunwald E, Antman EM, Beasley JW, *et al,* "ACC/AHA 2002 guideline update for the management of patients with unstable angina and non-ST-segment elevation myocardial infarctio.: 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.

[11] J. Broderick, T. Brott, R. Kothari, R. Miller, J. Khoury, A. Pancioli, D. Mills, L. Minneci, R. Shukla, "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.

[12] B.G.Buchanan, E.H.Shortliffe (eds.) "Rule-Based Expert Systems: the MYCIN experiments of the Stanford Heuristic Programming Project," *Addison-Wesley*, Reading MA, 1984.

[13] S. Buchbinder, I. Leichter, R. Lederman, B. Novak, P. Bamberger, M. Sklair-Levy, G. Yarmish, and S. Fields, "Computer-aided Classification of BI-RADS Category 3 Breast Lesions1", in *Radiology*, 2004, Vol. 230, pp. 820-823.

[14] P. Cathier, et al, "CAD for Polyp Detection: an Invaluable Tool to Meet the Increasing Need for Colon-Cancer Screening", in the *Proceedings of the 18th International Congress and Exhibition, Computer Assisted Radiology and Surgery (*CARS), Chicago, USA, June 23-26, 2004, pp. 978-982.

[15] Chapman, W., Bridewell W., Hanbury P., Cooper, G., Buchanan, B.G., "Evaluation of Negation Phrases in Narrative Clinical Reports", in the *Proceedings of the American Medical Informatics Association* (AMIA) Symposium , 2001, pp. 105-109.

[16] Chavey et al, "Guideline for the Management of Heart Failure Caused by Systolic Dysfunction: Part I. Guideline Development, Etiology and Diagnosis" *American Family Physician*, Vol. 64/No. 5 (September 1, 2001)

[17] A. M. Cohen, and W. R. Hersh, "A Survey of Current Work in Biomedical Text Mining", in *Briefings in Bioinformatic*s, March 2005, Vol. 6(1), pp. 57-71.

[18] Enrico Coiera, "Guide to Health Informatics – 2nd Edition", Arnold Publishers, December 2003.

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

[20] 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.

[21] M. Dundar, G. Fung, L. Bogoni, *et al* "A Methodology for Training and Validating a CAD System and Potential Pitfalls", in the *Proceedings of the 18th International Congress and Exhibition, Computer Assisted Radiology and Surgery* (CARS), Chicago, USA, June 23-26, 2004, pp. 1010-1014.
[22] Fellbaum, C., WordNet: An Electronic Lexical Database. *MIT Press*, May 1998.
[23] C. Friedman, and G. Hripcsak, "Natural Language Processing and Its Future in Medicine: Can Computers Make Sense out of Natural Language Text", in *Academic Medicine*, August 1999, Vol. 74(8), pp. 890-895.
[24] W. Galanter et al., "A Trial of Automated Decision Support Alerts for Contraindicated Medications Using Computerized Physician Order Entry," *J. Am. Med. Inform. Assoc*. 2005;12:269-274.
[25] N.Goldsclager *et al* for the North American Society of Pacing and Electrophysiology, "Practical Guidelines for Clnicians Who Treat Patients with Amiodarone", *Arch Intern Med.* 2000; 160:1741-1748
[26] V. Gottipaty, et al, "Automated Identification Of Madit-II Eligible Patients Using Remind Artificial Intelligence Software", in the 6th Scientific Forum on Quality of Care and Outcomes Research in Cardiovascular Disease and Stroke, *American Heart Association* (AHA), Washington DC, May 14-16, 2005.
[27] 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
[28] 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.
[29] Heckerman, D., "A tutorial on learning with Bayesian networks", *Microsoft Research Technical Report*, MSR-TR-95-06, 1996.
[30] L. Hirschman, J. C. Park, J. Tsujii, L. Wong, and C. H. Wu, "Accomplishments and Challenges in Literature Data Mining for Biology", in *Bioinformatics*, December 2002, Vol. 18(12), pp. 1553-1561
[31] R. G. Holloway, D. M. Witter Jr, K. B. Lawton, J. Lipscomb, G. Samsa, "Inpatient costs of specific cerebrovascular events at five academic medical centers. ", *Neurology*, 1996 Vol. 46, pp. 854–860.
[32] 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.
[33] Jaeger K.D., "Drug Pricing and Consumer Costs", *Pres to US Senate Commerce Committee,* April 23, 2004.
[34] Jensen, F.V., "An introduction to Bayesian Networks", UCL Press, 1996.
[35] Johnson, D.B., Taira, R.K, Zhou, W., Goldin, J.G., Aberle, D.R., Hyperad, "Augmenting and visualizing free text radiology reports", *RadioGraphics*, 1998, Vol. 18, pp. 507-515.
[36] Joint Commission on Accreditation of Healthcare Organizations (JCAHO), Website: http://www.jcaho.org.
[37] Kahn, M., Fagan, L., Tu, S., "Extensions to the Time-Oriented Database Model to Support Temporal Reasoning in Expert Medical Systems", in *Methods of Information in Medicine*, 1991, Vol. 30, pp. 4-14.
[38] B. Kappen, W. Wiegerinck, E. Akay, J. Neijt, and A. van Beek, "Promedas: A Clinical Diagnostic Decision Support System in Bayesian Modeling Applications", in the *Proceedings of the 15th Belgian-Dutch Conference on Artificial Intelligence* (BNAIC'), Nijmegen, The Netherlands, October 23-24, 2003. pp. 455-456.
[39] Kayaalp, M., Cooper, G. F., Clermont G., "Predicting ICU Mortality: A Comparison of Stationary and Nonstationary Temporal Models", in *Proceedings of American Medical Informatics Association* (AMIA) Symposium, 2000, pp. 418-422.
[40] 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
[41] R. S. Ledley, and L. B. Lusted, "Reasoning Foundations of Medical Diagnosis", in *Science*, 1959, Vol. 130, pp. 9-21.
[42] Lehman Brothers; McKinsey "Parexel Pharmaceutical International Sourcebook, 2000", *CenterWatch,* 2000.
[43] Lehman Bros, "Pharma Outsourcing Digest" 3/23/01.
[44] C. L. Leibson, J. M. Naessens, R. D. Brown, J. P. Whisnant, "Accuracy of hospital discharge abstracts for identifying stroke.", *Stroke*, 1994, Vol. 25, pp. 2348–2355.
[45] Manning, C.D., Schutze, H., "Foundations of Statistical Natural Language Processing", *MIT Press*, Cambridge, Massachusetts.
[46] McCallum A.K., "BOW: A toolkit for statistical language modeling, text retrieval, classification and clustering", 1996,: http://www.cs.cmu.edu/~mccallum/bow.
[47] R. A. Miller, F. E. Fasarie, and J. D. Mayors, "Quick Medical Reference (QMR) for Diagnostic Assistance", *MD Computing*, Sept-Oct, 1986, Vol. 3(5), pp. 34-48.
[48] J. B. Mitchell et al, "What role do neurologists play in determining the costs and outcomes of stroke patients?", *Stroke*, 1996, Vol. 27, pp. 1937–1943.
[49] Moss AJ, Cannom DS, Daubert JP, et al, for the MADIT II Investigators. "Multicenter Automatic Defibrillator Implantation Trial II (MADIT II): design and clinical protocol." *Ann Noninvasive Electrocardiology* 1999;4:83-91.
[50] Moss AJ, Zareba W, Hall J, et al, for the Multicenter Automatic Defibrillator Implantation Trial II Investigators. Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. *New England Journal of Medicine,* 2002;346:877-883
[51] Nahm U.Y., Mooney R.J., "A Mutual Beneficial Integration of Data Mining and Information Extraction", in *Proceedings of American Association of Artificial Intelligence* (AAAI), 2000, pp. 627-632.
[52] 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.
[53] Nigam K., McCallum A., Thrun S., Mitchell T. "Learning to Classify Text from Labeled and Unlabeled

Documents", in *Proceedings of the 15th Conference of American Association for Artificial Intelligence* (AAAI), 1998, pp. 792-799 .
[54] J. A. Osheroff, E.A. Pifer, J. M. Teich, MD, et al, "Improving Outcomes with Clinical Decision Support: An Implementer's Guide", by *Health Information & Management Systems Society*, 2005.
[55] K. Preston, Jr., and M. Onoe (Editors), "Digital Processing of Biomedical Images", *Plenum Press*, New York, l976.
[56] Rabiner R. L., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", in the *Proceedings of the IEEE*, Vol. 77(2), pp. 257-286.
[57] R. H. Rao, and R. B. Rao, "Quality Assurance through Comprehensive Extraction from Existing (non-structured) Patient Records", in the *Annual Conference and Exhibition, Healthcare Information and Management Systems Society* (HIMSS), San Diego, California, Feb 9-13, 2003.
[58] R. B. Rao, S. Sandilya, R. S. Niculescu, C. Germond, and A. Goel. "Mining Time-dependent Patient Outcomes from Hospital Patient Records", in the *Proceedings of American Medical Informatics Association* (AMIA) Annual Symposium, San Antonio, Texas, November 9-13, 2002.
[59] R. B. Rao, S. Sandilya, R. S. Niculescu, C. Germond, and H. Rao. "Clinical and Financial Outcomes Analysis with Existing Hospital Patient Records", in the *Proceedings of the Ninth ACM SIGKDD International Conference of Knowledge Discovery and Data Mining* (KDD), Washington DC, August 24-27, 2003, pp. 416-425.
[60] Regenstrief Institute "LOINC: Logical Observation Identifiers Names and Codes", http://www.regenstrief.org/loinc/ LONIC Homepage
[61] J. Roehrig, "The Promise of CAD in Digital Mammography", in the *European Journal of Radiology*, Elsevier, 1999, Vol. 31, pp. 35-39
[62] S. Sandilya, and R. B. Rao, "Continuous-Time Bayesian Modeling of Clinical Data", in the *Proceedings of the fourth SIAM International Conference on Data Mining* (SDM), Lake Buena Vista, Florida, April 22-24, 2004.
[63] E. H. Shortliffe, "Computer Programs to Support Clinical Decision Making", in the *Journal of American Medical Association* (JAMA), 1987, Vol. 258, pp. 61-66.
[64] E. H. Shortliffe (Ed.), L. E. Perreault (Ed.), G. Wiederhold (Asso. Ed.), L. M. Fagan (Asso. Ed.), "Medical Informatics: Computer Applications in Health Care and Biomedicine (Health Informatics) – 2nd Edition", *Springer*; November, 2000.
[65] M. A. Shwe, B. Middleton, D. E. Heckerman, M. Henrion, E. J. Horvitz, and H.P. Lehmann, "Probabilistic Diagnosis Using a Reformulation of the INTERNIST-1/QMR Knowledge Base: I. The Probabilistic Model and Inference Algorithms", in the *Methods of Information in Medicine*, October 1991, Vol. 30(4), pp.241-255.
[66] SNOMED International "SNOMED Clinical Terms", College of American Pathologists, http://www.nlm.nih.gov/research/umls/rxnorm_main.htmlhttp://www.snomed.org/
[67] A. F. Sonel, 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", in the *Annual Scientific Session of American College of Cardiology* (ACC 2005), Orlando, Florida, March 6-9, 2005.
[68] J. Spina et al, "Clinical Relevance of Automated Drug Alerts From the Perspective of Medical Providers", *American Journal of Medical Quality* 2005;20:7-14.
[69] Taira R., Soderland S., Jakobovits R., "Automatic Structuring of Radiology Free Text Reports" *RadioGraphics*, 2001, Vol. 21, pp. 237-245
[70] 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.
[71] H. Tunstall-Pedoe, "The World Health Organization MONICA Project (Monitoring Trends and Determinants in Cardiovascular Disease): A major international collaboration," *Journal of Clinical Epidemiology*, 1988, Vol. 41, pp. 105-14.
[72] United States Department of Health and Human Services, "Summary of the HIPAA Privacy Rule", http://www.hhs.gov/ocr/hipaa.
[73] United States Library of Medicine, "MeSH: Medical Subject Headings", http://www.nlm.nih.gov/mesh/
[74] United States Library of Medicine, "UMLS: Unified Medical Language System", http://www.nlm.nih.gov/research/umls/
[75] G. Wiederhold, Edward H. Shortliffe, L.M. Fagan, Leslie E. Perreault, Lawrence M. Fagan (editors) "Medical Informatics : Computer Applications in Health Care and Biomedicine (Health Informatics)" *Springer*; 2nd edition, November, 2000.
[76] World Health Organization "Manual of the international statistical classification or diseases, injuries, and causes of death", *World Health Organization*, Geneva, 1977.
[77] World Health Organization "Report of the international conferences for the Tenth Revision of International Classification of Diseases", *World Health Organization*, Geneva, 1992.
[78] World Health Organization, "The Atlas of Global Heart Disease and Stroke", *World Health Organization & Center for Disease Control*, 2004.