

Meta-Analysis of Clinical Cardiovascular Data towards Evidential Reasoning for Cardiovascular Life Cycle Management

Lakshmi Narasimhan and Di Wu
Western Kentucky University,
Bowling Green, KY, USA

Narinder S Gill
The Medical Center,
Bowling Green, KY, USA

Lakshmi.narasimhan@wku.edu
di.wu@wku.edu

gilldoc7@yahoo.com

Abstract

The cardiovascular disease is one of the serious and life-threatening diseases in the developed world. One aspect of medical treatment is using drugs with blood pressure reducing or cholesterol lowering functions. Importantly, such treatment needs to be individually tailored and is significantly correlated to the particular conditions of individual patients. However, such pathologies and mechanisms are still only under investigation. Several novel and unique computational methods, called meta-analyses techniques, for formatting and analyzing a wide variety of cardiac datasets are discussed in this paper with the aim of building cardiovascular database and related patient life-cycle management services. In this paper we also present an overview of a second order inference engine underlying the meta-analyses, which yields evidenced-based reasoning that is more likely to better assist decision-making on the effectiveness of cardiovascular treatment than what is available currently. Furthermore, the software architecture and other details of such a medical informatics system tailored to cardiovascular disease are also described. Research and development work on this project yields itself to application to many other areas, such as disease control and prevention in Epidemiology, and dietics. The system can therefore make a profound impact to medical informatics.

Key Words: Meta-analysis, Cardiovascular life cycle management (cardiac clinic), Data mining

Introduction

Computers and information technology (IT) have been widely used in medical research, particularly for providing several types of service. A remarkable advantage of IT is that long-term investigation on patient-care, diagnosis and treatment of diseases can be done efficiently through large-scale data analysis and data mining of personalized historical data sets. Thus for instance, a clinical

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Publisher@InformingScience.org to request redistribution permission.

decision support system provides diagnosis, treatment, follow-up on a number of issues, including the option to create information anchor desks for a wide variety of practitioners, such as, general practitioners, pharmacists, nutritionists, and the like (Kaplan, 2001, Trivedi, 2004, Payne, 2000, Sim, 2001, Purcell, 2005). Such a system will not only benefit patient care, but also provide physicians a platform to study the

medical issues at a broader level. However, few systems have been successfully developed and tailored to handle a broad spectrum of diseases (Economou, 2001, Cornelia, 2003, Bobb, 2004). This is mainly due to the fact that designing and providing a reliable and knowledge-based medical informatics environment, like medical disease decision support system, is still very challenging in terms of collecting large amount of medical data, large scale data mining and management, information extraction including text and graphics, Human Computer Interface (HCI), development of statistical tools, and decision making tools (Kuperman, 2003, Clercq, 2004). In particular, when such tools are put together over a hardware-software architecture, it requires a competitive multi-disciplinary team with experts drawn from each field to manage, massage and exploit data sets effectively and efficiently.

From the medical perspective, cardiovascular disease is one of the serious and life-threatening diseases in the developed world. Causes of cardiovascular disease are many, including genetics, food habits, hardening of the arteries, high blood pressure and others. In recent years, a number of drugs have been discovered or synthesized, with capabilities to soften the arteries and control cholesterol and blood pressure, and some of them are widely used in medical treatment. However, it has been observed that the effectiveness and outcomes of the treatment are surprisingly inconsistent, even among “*similar patients*”. More and more literature show that such diverse reasons result in many problems in understanding the causes of the disease, patient treatment, drug design and disease prevention in individual level (Reddy, 1998, Fuster, 2005, Lewis, 2002). For example, Lipitor (Black, 1998) – a cholesterol-reducing drug/satin has profound impact on certain patients, but limited or no impact on other patients. Further, certain post-operative patients, who perform routine exercise show significant improvement in their life-style, but others who maintain such or similar kind of exercise regimen, seem to have life style problems. Further critical research on data available openly on landmark studies in cardiology reveal that medical treatment needs to be greatly tailored to specific individual patients. However, these pathologies and mechanisms are still under discovery, which seriously affects the accuracy of decision-making on treatment; occasionally such trial and error processes may also lead to medical accidents. Theoretically, such a study can be prosecuted through a careful and complex investigation of patients using computational and statistical tools. However, in practice, it has many technical difficulties, two of them being: i) the need to focus any study to a few measurable parameters and ii) *ipso facto* the need to manage inconsistent medical information and data from several clinical studies. The former problem requires meta-analyses procedures to be developed as envisaged in this proposal, and the latter needs metadata converters (called *form fillers* here) so that medical information data from different landmark studies may be combined in a manner that meets the needs of cardiovascular disease management purposes. These datasets include textual, structured, signal and image data. Unfortunately, to the best of our knowledge, there is no complete cardiovascular life-cycle disease management system available, which can provide systematic study and service of the nature envisioned here.

Another issue that arises as a result of the above scenario is that several sources of data have now become available from several stakeholders, who include publicly funded projects (whose results are available over the WWW typically), second party funded projects (e.g., pharmaceutical company funded works, whose results have limited applicability and availability), and local sources (typically in a hospital environment). The ways and means to bring and marry these datasets is a non-trivial problem, even though the medical community realizes that there is significantly useful information available in them as a composite collection. Consequently, developing medical informatics tools based on meta-analysis of clinical cardiovascular data towards evidential reasoning for disease management can be one powerful weapon to attack these problems. Medical service providers such as, Cardiologist, Pharmacologist, Specialist internist, Nutritionist need their own specialized anchor desk that are customized to meet their needs.

In this paper, we propose solutions aimed at meeting some of these challenges on technical and medical aspects, and envision several preliminary studies towards developing an evidential reasoning system using statistical inference techniques, so that the medication management can be tailored to a specific individual. The analysis environment needs a competitive and multi-disciplinary team, consisting of cardiologists, computer scientists and information scientists. Additional useful specialists include professionals from the field of bioinformatics to address these challenging issues.

The rest of the paper is organized as follows: the second section provides an overview of the technical architecture of the system, while the third section deals with impacts on cardiovascular disease treatment. The conclusion summarizes the paper and provides pointers for future work that has been planned on the system.

The Meta-Analyser Environment

Cardiovascular like cycle management system presented in this paper is capable of extracting clinical information including text and graphics, constructing data warehouse and providing support for decision making. It has come to our knowledge that such a life cycle management approach is new and to the best of our knowledge only Mount Sinai Hospital, New York City, (<http://www.mountsinai.org/msh/msh-home.jsp>) is attempting a research similar to ours. The system has the capability to handle high-order references from different resources, with differing formats and meta-information. The system presented below employs the outcomes of multi-disciplinary research in many different areas. It has the following three major goals:

1. Address cardiovascular patient information database requirements: Currently available datasets are from varied sources with different formats, which are inconsistent with each other. A systematic information/fact extraction and formatting toolset - so-called Fact eExtractors (FXs) – is being developed.
2. Investigate several computational methods and tools and deploy aid multi-source data from landmark¹ studies: These metadata analyzers have yielded advanced technical information on the treatment and evaluation of various drugs and their effectiveness, wherein datasets have been collected from varied sources/studies.
3. Provide an intelligent system able of providing data information and making accurate decisions on treatment. An internationally competitive team of experts from several domains is being assembled.

Several challenges arise in constructing such system exists in many perspectives, such as how to collect large amount of information, how to do the intelligent information extractions, how to build the data warehouse, and how to make accurate decision-making on treatment. In this paper, an overview of the meta-analysis environment and its various components are described (see Figure 2 at end of paper also):

Data Collection

The cardiovascular patient information is being collected mainly through collaborated hospitals. Data such as, gender, height, weight, medication, ECG record, MRI record and related medical text information are also being obtained through appropriate automated information extractor software systems. In addition, additional data sets can also be obtained from several useful public

¹ A landmark study is typically a longitudinal covering people from multiple ethnicity, race and national origin. An example of such a study is presented in (Mazzei, 1999, Miller, 2002).

domain websites. Besides these, pharmaceutical companies also undertake studies on various drugs and collect several types of information (which unfortunately can be limited in their accessibility). Note that data from each of these sources may not be in the same format, and further, some data may also be missing or has its inherent biases. Table 1 shows the potential data collection through these three main approaches.

Table 1: Typical data sources

Source	Data
Collection I (Private)	Patient info data (by hospitals): Gender, height, weight, medication, ECG, MRI,...
Collection II (2 nd party)	Published limited patient info data (by pharmaceutical companies)
Collection II (Public)	Other available public data (by non-profit institutions, etc)

Fact Extractors

The patient information data usually includes general information such as physical information, text information including doctor’s prescription and patient status, clinical information data such as ECG records and medication information including blood pressure and others. Analysis and storage of these large-scale data requires efficient and specific information extraction. Basically, there are two fundamental problems which need to be solved in this context, fact extraction of text and image:

1. **Semi-formatted text and free text analysis.** Even though several systems are available for creating semi-formatted texts, parsing the text in order to pick out the relevant bits of information is the key behind our system. Several commercial products also exist for such applications, but most of them lack generality of applications. Compared to semi-formatted text, free text is more difficult to analyze and among the approaches, three schools of thought are more popular, namely, the pure linguistic community, the computational linguists and the shallow analyzers. It is well recognized that the first approach is suited to analyzing intent and is much more profound, but at the same time, it is expensive both in terms of time and human and computational efforts. The second approach is good, more as a coarse-grain filter than as an accurate collector. The last approach is generally good for collecting rough information, which is not good for noting of intent, but is useful for trend analysis and corporate knowledge upkeep. The last approach is useful, where it is recognized that that the text contains less than 10% useful information. This approach is error-prone, but over a large window, the errors can be significantly reduced through appropriate data mining techniques. In addition, through a process, which we call language engineering, errors can be further reduced. Further reduction of errors is also possible, when the language engineering process can be tailored to particular domains. Such research is being investigated further as they apply to the current domain in the system.
2. **Image processing and pattern recognition.** Some of the patient information such as, ECG and MRI records, are in signal and image formats respectively. An important feature of those records is that in the entire record, only a small part provides the critical information regarding patient medical status. The crucial questions are how to efficiently and accurately store, retrieve, identify and exploit such useful information to the benefit of treating a given patient. Several technologies and some products for studying this issue are available, but

nearly all of them are still in their infancy and need to be tailored to specific applications. We employ a series of signal fact extractors (*a la* composite filters) to extract various records.

Form Fillers and Data warehouse Development

‘Cleaned data’ after using fact extractors get deposited into specific databases corresponding to each collection. Form filling is a process by which the collector assembles extracted data for storage over a data warehouse (Figure 1). The storage process will use a suitable ontology (classification mechanism) and metadata (data about the datasets). We also note that the data retrieval process will also use a similar kind of form fillers, this time the information requirement being addressed by the user community. The retrieval process also calls for suitable metadata to be stored so that retrieval efficiency and recall rates are high. This project will provide tools and mechanisms to handle ontology and the differing types of metadata² from any two different communities of specialists.

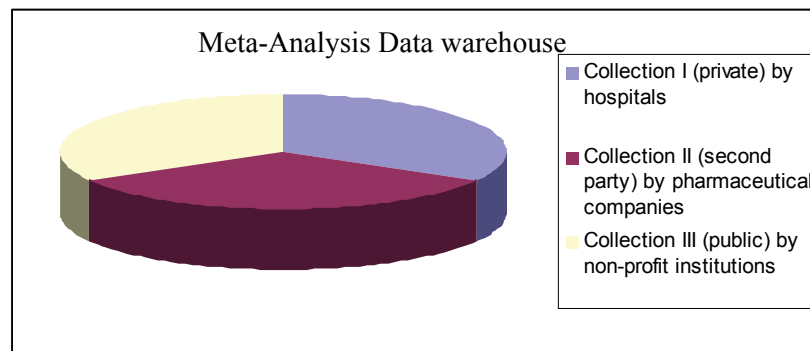


Figure 1: Data warehouse

High-order Inference Engine

Unlike traditional decision-support systems, our system employs a high-order inference engine, whereby a decision is made or predicted using a variety of inference mechanisms over all three types of collections. The main study under this system is aimed at on collection I, while others will be supplementary. Several service tools that are being added include, statistical and data mining using principal component analysis, hidden Markov model, clustering, and maximum likelihood analyzers. Those tools have been shown to be of significant aid in other medical decision support systems (Fritsche, 2002, Rose, 2005).

Anchor Desks

An important feature of this system is the provision of anchor desks. Even though there are many reasons for the cause of cardiovascular diseases, through different perspectives drawn from a confluence of decision support mechanisms, effective treatment can be achieved. For example, the nutritionists will pay more attention to the diet of patients, while pharmacologists would like to look at the family history, previous medication of patients or side-effects. This system will provide doctors in different disciplines including cardiologist, pharmacologist, special internist, and nutritionist, a platform or opportunity to approach cardiovascular diseases from multiple angles and perspectives.

² The metadata sets for a collector is different to those of the retriever (or user) and this is known as metadata semantic gap. Hence there is a need to have two sets of metadata for a given collection so that the person who stores and the person who retrieves the data can work with the same system efficiently.

Impacts on Cardiovascular Life Cycle Management

As mentioned before, the study of cardiovascular life cycle problem is a very challenging subject due to the complexity and multiplicity of issues involved, including the presence of cardiovascular pharmacology and mechanisms, open-heart surgery related issues, clinical medication issues, patient care and recovery, and nutrition and disease prevention issues. Some of these issues could be highly related, or counteractive, and several mechanisms are still under investigation. It has been understood that personalized historical data sets have rich information for approaching this problem. Thus, solutions to this problem definitely require not only crossing disciplinary collaboration, but also large-scale and even multi-dimensional data analysis and mining. The system presented in this paper is tailored to cardiovascular life-cycle management and has great impacts in the following manner:

Information Extraction and Data Warehouse

The fundamental problem in medical informatics is how to recognize important patterns or information from collected medical or patient records. Considerable challenges exist in handling semi-text, free text as well as imagery data sets and several novel techniques are being investigated in this system. Further, a data warehouse is employed to handle large-scale as well as multi-dimensional data sets; the architectural implications of the data warehouse are also being studied in order to make the system applicable to a wide variety of medical scenarios.

Data Mining

In order to utilize the rich information existing in data warehouse of our system, several types data mining techniques are being employed. For instance, our system has large-scale and high dimensional data sets, and further, high-inference engine in this system to provide more accurate decisions or results; and, these issues will challenge current available data mining techniques. Through the study of this system, a toolbox of software packages for multi-dimensional data mining, will be developed in order to provide an efficient and accurate analysis weapon.

Anchor Desks

The important feature of our system is to provide an opportunity and platform to study cardiovascular life-cycle treatment through many viewpoints, particularly as they correspond to studies involving issues in medication, pharmacy, surgery, nutrition management and the like. The provision of an anchor desk will provide an account of knowledge or information in related disciplines and the ability to transduce such knowledge across sub-specialties, thereby hopefully providing a comprehensive mechanism for decision support. One of the anchor desks will be dedicated to the patient themselves, so that their data sets are available to them (and also become portable).

Conclusions

In this paper, we have presented a system that performs meta-analysis of clinical cardiovascular data towards evidential reasoning for cardiovascular life-cycle management. The system addresses several issues such as, techniques for handling semi-text, free text and imagery data sets, constructing large-scale data warehouse, investigating novel data mining techniques, and designing anchor desks. The system incorporates a high-order inference engine, which allows the understanding of cardiovascular life-cycle management issues through multiple angles. We are hoping that the system will provide medical doctors and patients with valuable information for improving cardiovascular disease treatment, post-surgery recovery, patient care and further, disease prevention. More details on this work in progress, including issues in testing the system, will become available in due course.

References

- Black, D.M. et al. (1998). An overview of the clinical safety profile of Atorvastatin (Lipitor), a new HMG-CoA reductase inhibitor. *Archives of Internal Medicine*, 158, 577-584.
- Bobb, A. et al. (2004). The epidemiology of prescribing errors The potential impact of computerized prescriber order entry. *Archives of Internal Medicine*, 785-792.
- Clercq, P.A., Blom, J.A., Korsten, H.H., & Hasman, A. (2004). Approaches for creating computer-interpretable guidelines that facilitate decision support. *Artificial Intelligence Medicine*, 1-27.
- Cornelia, et al. (2003). Effects of a computerized system to support shared decision making in symptom management of cancer patients: Preliminary results. *Journal of the American Medical Informatics Association*, 573-579.
- Economou, G.-P., Lymberopoulos, D., Karavatselou, E., & Chassomeris, A. (2001). A new concept toward computer-aided medical diagnosis: A prototype implementation addressing pulmonary diseases, *The IEEE Transactions on Information Technology in Biomedicine*, 55-65.
- Fritsche, L. (2002). Recognition of critical situations from time series of laboratory results by case-based reasoning. *Journal of the American Medical Informatics Association*, 9, 520-528.
- Fuster, V. (2005). Cardiovascular disease in the elderly: A challenge for science and clinical care. *Nature Clinical Practice Cardiovascular Medicine*, 2, 549
- HIPPA (Health Insurance Portability and Accountability Act. (n.d.). Available at <http://www.hipaa.org/> (This is the US Government approved reference source).
- Kaplan, B. (2001). Evaluating informatics applications—clinical decision support systems literature review. *International Journal of Medical Informatics*, 15-37.
- Kuperman, G.J. et al. (2003). Computer physician order entry: Benefits, costs, and issues. *Annals of Internal Medicine*, 31-39.
- Lewis, K. (2002). Epidemiology and prevention of cardiovascular diseases: A global challenge. *Journal of Public Health Policy*.
- Mazzei, W.J. (1999). Maximizing operating room utilization: a landmark study, *Anesth Analg*, 89, 1–2.
- Miller, A.H., Roth, B.A., Mills, T.J., Woody, J.R., Longmoor, C.E., & Foster, B. (2002). Ultrasound guidance versus the landmark technique for the placement of central venous catheters in the emergency department. *Acad Emerg Med*, 9(8), 800-805.
- NIST: National Institute for Standards and Technology -- <http://hcs1.sdct.nist.gov:8080/hcs1/index.htm>
- Payne, T.H. (2000). Computer decision support systems. *The Cardiopulmonary and Critical Care Journal*, 47-52.
- Purcell, G.P. (2005). What makes a good clinical decision support system? *British Medical Journal*, 740-741.
- Reddy, K.S., & Yusuf, S. (1998). Emerging epidemic of cardiovascular disease in developing countries. *Journal of the American Heart Association*, 87, 596-601.
- Rose, C., Smaili, C., & Charpillat, F. (2005). A dynamic Bayesian network for handling uncertainty in a decision support system adapted to the monitoring of patients treated by hemodialysis. *17th IEEE International Conference on Tools with Artificial Intelligence*, 2005.
- Sim, I., et al. (2001). Clinical decision support systems for the practice of evidence-based medicine. *Journal of the American Medical Informatics Association*, 527-534.
- Trivedi, M., et al. (2004). A computerized clinical decision support system as a means of implementing depression guidelines. *American Psychiatric Association*, 879-885.

Biographies



Dr. Lakshmi Narasimhan obtained his Bachelor's degrees in Physics and Electronics Engineering from the University of Madras and Indian Institute of Science respectively. He also obtained his Master's and Phd degrees respectively from the Madras Institute of Technology and University of Queensland, Australia. His research interests include soft engineering, informatics technology and medical informatics.



Dr. Di Wu received his Ph.D. in Bioinformatics and Computational Biology and Applied Mathematics from the Iowa State University. He also received his B.S. in Biology from Fudan University, China. His research interests include bioinformatics, biological system modeling, optimization with applications to biomedical problems.

Narinder S Gill, M.D. obtained his M.D. and M.B.B.S. in the Internal Medicine from V.S.S. Medical College in India. He finished his residency in Mount Sinai School of Medicine, NY.

His research interests include Medical Informatics and applications of computer technology to medical research.

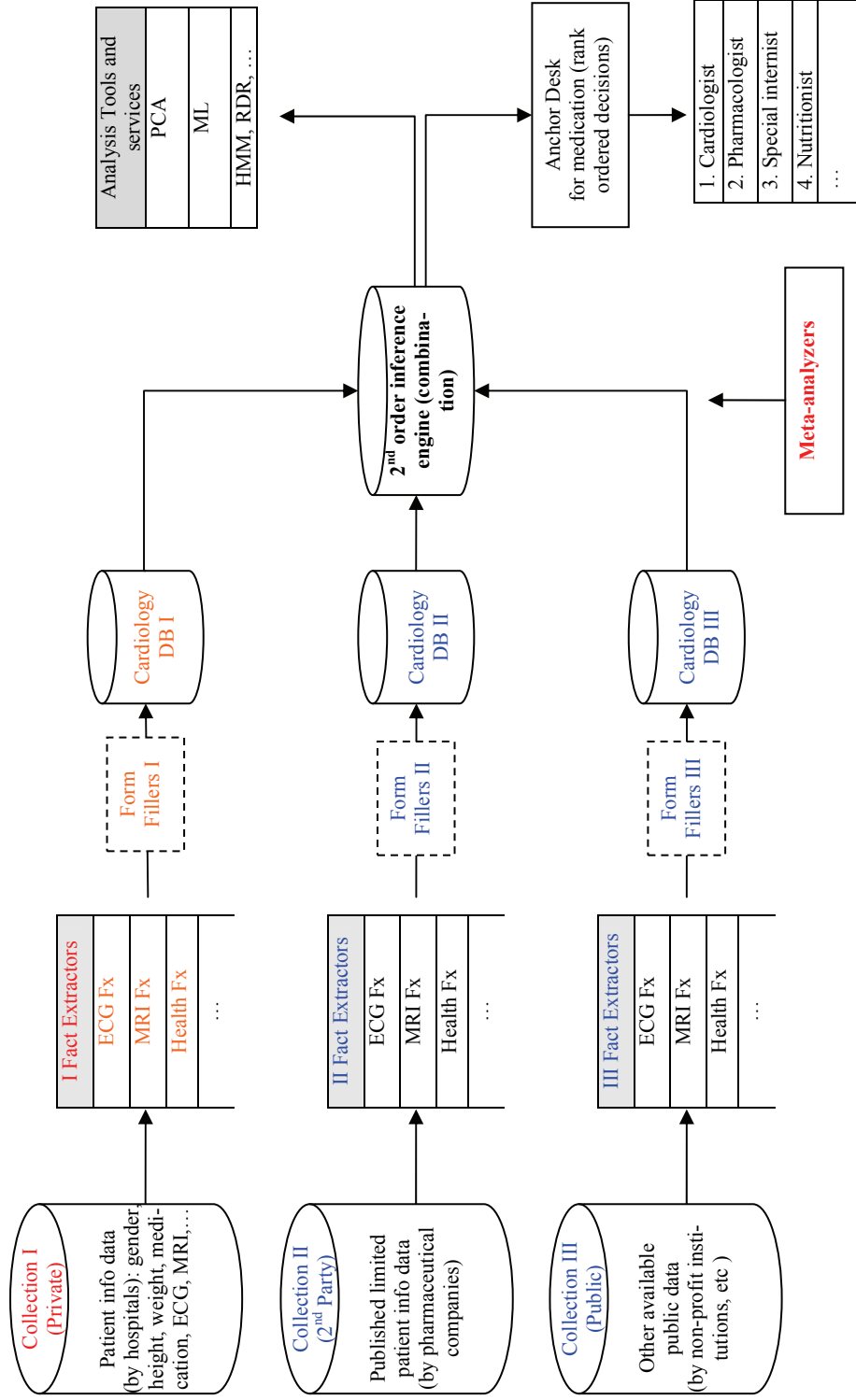


Figure 2: Overall System Architecture for Cardiovascular Meta-analysis Anchor Desk