VETERANS ENGINEERING RESOURCE CENTER
THE DREAM PROJECT

by

Rabab A. Bennis

MBBCh, University of Tripoli, Libya, 2008

Submitted to the Graduate Faculty of

Health Policy and Management

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Public Health

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This essay is submitted

by

Rabab A. Bennis

on

April 21, 2014

and approved by

Essay Advisor:
Wesley Rohrer, PhD ______________________________________
Assistant Chair, Health Management Education
Assistant Professor of Health Policy & Management
Graduate School of Public Health
Department of Health Information Management
School of Health and Rehabilitation Sciences
University of Pittsburgh

Essay Reader:
Valerie Watzlaf, PhD, RHIA, FAHIMA ______________________________________
Associate Professor
Department of Health Information Management
School of Health and Rehabilitation Sciences
University of Pittsburgh

Essay Reader:
Robert Monte, RPh, MBA ______________________________________
Director Veterans Engineering Resource Center (VERC)
VA Pittsburgh Healthcare System

Wesley Rohrer, Ph.D.

VETERANS ENGINEERING RESOURCE CENTER: THE DREAM PROJECT

Rabab A. Bennis, MPH

University of Pittsburgh, 2014

ABSTRACT

Due to technological advances, data collected from direct healthcare delivery is growing by the day. The constantly growing data that was collected from various resources including patient visits, images, laboratory results and physician notes, though important, has no significance beyond its satisfying reporting and/or documentation requirements and potential application to specific clinical situations, mainly due to the voluminous and heterogeneous nature of the data. With this tremendous amount of data, manual extraction of information is expensive, time consuming, and subject to human error. Fortunately, information technologies have enabled the generation and collection of this data and also the efficient extraction of useful information. Currently, there is a broad spectrum of secondary uses of this clinical data including clinical and translational research, public health and policy analysis, and quality measurement and improvement.

The following case study examines a pilot project undertaken by the Veterans Engineering Resource Center (VERC) to design a data mining software utility called Data Resource Engine & Analytical Model (DREAM).This software should be operable within the VA IT infrastructure and will allow providers to view aggregate patient data rapidly and accurately using electronic health records.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

All of the work presented henceforth was conducted at the Veterans Engineering Resource Center (VERC). I was a team member on the DREAM project, where I was responsible for physician's interviews, designing modules for the second version of DREAM and interviewing team members for the purpose of this paper. I would like to thank my committee, Dr. Rohrer and Dr. Watzlaf for their amazing help and endless support. I would also like to Thank Mr. Monte, the VERC's Director for giving me an internship opportunity with the center.

1.0 INTRODUCTION

Before summer 2014 interns begin their internship, Robert Monte, Director of the Veterans Engineering Resource Center must decide whether or not DREAM will be one of the projects he will assign for the new interns. The project that started as an idea for a potentially powerful tool for healthcare providers and researchers is facing a critical decision point to determine if it is feasible to continue.

The DREAM project is one of the more interesting technical applications in the Veterans Engineering Resource Center's history. The center had intended to create a search engine that would allow clinical practitioners to develop both field and text based searches as well as providing the option to conduct statistical analysis on the data obtained. This search engine will be searching the VistA electronic health records at the Pittsburgh data warehouse. The main goal was to allow clinical practitioners to search for information on two levels, a population search (all patients who share certain criteria that the practitioner specifies), and an individual patient search. The resulting system would allow clinical practitioners to better utilize their time in monitoring clinical quality indicators efficiently.

Although two different teams have worked on DREAM so far, both had the same mix of professional backgrounds. The DREAM team included IT programmers, developers, clinical practitioners and statisticians. While some members embraced the project idea, others were more skeptical about the feasibility of the project, especially concerning the amount of data the

Veterans Healthcare Administration possesses as well as the VA’s outdated IT infrastructure which was built decades ago and does not allow the rapid sharing of information or accommodation of new users’ applications (Walters, 2009).

1.1 DATA MINING TECHNOLOGY

Data mining is a process of pattern and relationship discovery within large sets of data. It is also one step in a multiple step process called knowledge discovery and data mining (KDD). KDD is a non-trivial extraction of implicit, previously unknown and potentially useful information from data that is located in databases (Milovic, 2012).The process of data mining consists of three stages (StatSoft,2013): (1) Initial exploration : This stage usually starts with data cleaning and transformation. Traditional-analytical tools (e.g. statistics) are used to explore data. (2) Model building and validation: This stage involves considering various models and choosing the best one based on its predictive performance. (3) Deployment: The use of mining results by exporting data into database tables or into other applications, for example, spreadsheets

Analytical techniques

A number of underling analytical techniques can be applied to various functions of data mining. Neural networks, decision tree, traditional statistics and visualization are the most popular techniques (Rogers, SAS Institute). The one aspect that linked these techniques and large data bases is a cheaper storage space and processing power (Milovic, 2012). Other analytical techniques are described below:

1. Neural networks

   An analytical technique composed of groups of connected input/output units where each connection has its own weight. Capable of predicting changes and events after process of learning to adjust network weights to produce optimum predictions (Rogers, SAS Institute).

2. Decision trees

   A graphical representation of relations that exist between data in the database, a technique mainly used for classification and prediction (Milovic, 2012)

3. Visualization

   A visual interpretation of complex relationships in multidimensional data. This technology offers immediate graphical identification of patterns (StatSoft, 2013).

1.2 DATA MINING IN HEALTHCARE

The valid and easily understood knowledge that resulted from the application of information technologies has a broad and constantly growing usage in healthcare settings. Healthcare organizations are mainly interested in using this knowledge to enhance physician practices, disease management, and resource utilization (Hardin, 2006). The driving force for this change has been the pressure to decrease the constantly increasing cost of care (Kraft, 2012).

In spite of the numerous uses of the medical records data, accessing it has always been surrounded by multiple issues. Time spent, number of resources needed to identify the target cohort of patients, EHR software used and ability to access data in narrative text were among the most important issues. Facilitating access to this data along with an organized way of presenting information will have a tremendous impact on willingness of usage by physicians, researchers, and public health professionals.

Information technology and its impact on the quality of healthcare was mainly achieved through providing assistance in three services: monitoring, decision support, and tracking adverse events (Bates, 2003).

1. **Monitoring**

Due to the increased amount of clinical data to be collected, it is hard to sift through them to ensure adherence or detect unconformities. Very often, this comparison between quality of care and guidelines happens retrospectively. Ensuring adherence to guidelines in real time during the course of treatment will be very useful in detecting problems as they happen. This will contribute to shaping a more proactive healthcare system instead of a reactive one. A study that examined adherence to heart failure quality of care indicators in 223 US hospitals showed that across all hospitals, the median rate of conformity was 72% for the use of angiotensin converting enzyme inhibitors in patients with left ventricular systolic dysfunction and 43% for smoking cessation counseling (Fonarow, 2005). Another study showed that only 38% of patients with venous thromboembolism risk factors received the prophylaxis treatment mandated in the clinical guidelines (Khan, 2004). These are just a few examples evidencing the gap between best practice guidelines and actual performance.

One of the first usages of data mining in healthcare quality management was done by United Healthcare Corporation. This managed care company developed a Quality Screening and Management (QSM) program to analyze the healthcare provided by its own health plans. QSM compares the care received by patients to practice guidelines using claims, administrative data and Medical records reviews (Leatherman S, 1991). QSM examined 15 measures for patients with chronic diseases and results were used to direct appropriate quality management actions as well as to identify new strategies for improvement.

According to Center of Medicare & Medicaid (CMS), ensuring adherence to guideline results in effective, safe, efficient, patient-centered, equitable and timely care. In July 2010, CMS started an incentive program to provide incentive payments to eligible professionals, eligible hospitals, and critical access hospitals (CAHs) as they adopt, implement, upgrade or demonstrate meaningful use of certified EHR technology. Simply, "meaningful use" means that providers are using a certified EHR technology in ways that can be measured significantly in quality and quantity. To receive the incentive, providers have to show they are "meaningfully using'' their EHRs by meeting both a core and a menu set of objectives (CMS.gov).

2. Decision support

For decades, information systems have been used for clinical support through a wide variety of activities. Such as making key information available (e.g. lab results), calculating weight-based dosage of medication or red-flagging patients for whom a certain drug may be inappropriate (Bates, 2003). Now, computerized algorithms and neural networks can assist in the prediction of clinical outcomes while considering many factors simultaneously. A computerized decision support system can synthesize and integrate patient specific information, and perform complex evaluations that can be presented to clinicians in a timely fashion (Sen, 2012).

3. Tracking adverse events

Computerized tools with electronic health records can identify, alarm, and track the frequency of adverse events. These tools are also capable of detecting drug allergies, and drug-to-drug interaction.

Due to complexity of the healthcare system, mining data in this environment faced multiple challenges in the process of extracting knowledge from large and diverse data. Two of the main challenges are lack of standardization and mining free text clinical data.

Lack of standardization

Lack of coding standards of medical information is one of the huge obstacles facing data mining in healthcare. Although the ICD-9-CM (International Classification of Diseases, 9th Revision Clinical Modification) was mainly developed to track diseases, relying on it alone might be accurate in identifying a specific diagnosis but not in excluding it. When feasible, review of charts should be used to confirm a diagnosis (Birman-Deych, 2005). At Columbia University Medical Center, 48.8% of patients with ICD-9-CM code for pancreatic cancer did not have corresponding disease documentation in pathology reports (Botsis, 2010). Errors in ICD diagnostic coding process arises from the quality and amount of information at admission as well as the coder training and familiarity with the illness.

On October 1, 2014, the ICD-9 code sets will be replaced by ICD-10 code sets. The new code sets consists of two parts.(1) The ICD-10-CM, for diagnosis coding, which uses 3 to 7 digits instead of 3 to 5 digits used with ICD-9-CM. (2) ICD-10-PCS for inpatient procedure coding, which uses 7 alphanumeric digits instead of the 3 or 4 numeric digits used under ICD-9-CM procedure coding(CMS,2013). This new structure provides greater specificity for diagnosis and inpatient procedures which will provide better data for data mining in addition to improving the predictive accuracy.

Free text clinical data

The unstructured clinical notes are a valuable source of information that is hardly found in any other section in the electronic health record. It includes physical findings, symptoms, and medication’s side effects. These notes are very difficult to standardize and thus a challenge for data mining (Cios, 2002). To save time, healthcare providers usually use abbreviations that are often difficult to interpret. The use of synonyms to describe a disease or test and misspelling of words also contributes to this difficulty.

Although difficult to mine, there is an increased need for extracting information from clinical notes. A recent study showed the possibility of detecting drug safety signals after transforming clinical notes to a feature matrix encoded using medical terminology (LePendu, 2013).

Another study that used natural language querying strategy, concluded that the free text of EMR is a viable source of quality reporting on evidence of foot examination for patients with diabetes (Pakhomov, 2008).

1.3 RELATED WORK

Medical Language Extraction and Encoding System (MedLEE)

Created by Carol Friedman at Columbia University. This system uses natural language processing methods to automatically encode data that is in textual form. This program was used in daily operations at Columbia Presbyterian Medical Center (CPMC) to structure and encode clinical reports (Friedman, 1996). A study in 2008 showed positive results using MedLEE for assessing quality of care for cardiovascular diseases (Chiang, 2010).

SAS Software

The SAS data mining tool was developed by SAS Institute and had been used by Oxford Health Plans. The typical user of this tool is a physician or executive with no IT background. The goal was to provide a dynamic interactive tool that provides key indicators involving quality outcomes and encounters of care (Rogers).

Soarian Quality Measures

In 2007, Siemens Medical Solutions introduced a healthcare data mining tool powered by Siemens award winning REMIND technology (Reliable Extraction and Meangiful Inference from Non-structured Data) that can help healthcare providers realize quality improvements by replacing time-consuming chart reviews with highly accurate automated chart abstraction of quality measures(Siemens, 2007). In 2011/2012, Siemens received the Office of the National Coordinator's Authorized Testing and Certification Body Certification for this quality reporting tool.

A guideline adherence study for patients with non-ST Elevation myocardial infarction (MI) at the Veterans Health Administration Hospital, Pittsburgh, presented analysis of 327 patients' records by both manual abstraction and REMIND technology. These patients were studied to see if they were treated properly for each of these medication classes: Aspirin, Beta blockers, ACHE Inhibitors, and Glycoprotein Receptor antagonists, per the American College of Cardiology's guidelines. Results showed that REMIND works as well as manual abstraction. REMIND took 4.5 hours while the abstractor took 176 hours to complete the analysis manually (Rao, 2005).

1.4 CLINICAL DECISION SUPPORT SYSTEMS (CDS)

Decision support systems refers to class of computer –based systems that aid the process of decision making. The CDS provides healthcare professionals with the knowledge and person-specific information, filtered and presented at appropriate time to enhance health and healthcare (Berner, 2009).

There are two major CDS, one of which employs data mining tools and the other, rule-based expert systems in the knowledge engine (Hardin, 2006). While the rule-based expert system must be supplied with facts which requires a broad knowledge from the decision maker in order to provide right answers to well informed questions, the data mining tool doesn't require any previous knowledge (Hardin, 2006).

Computerized clinical support systems showed considerable effectiveness in enhancing clinical performance. A study for CDS application in checking drug allergies and drug interactions, showed 83% reduction in overall rate of medication errors (Bates, 2003).

One of the most well-known CDS is MYCIN system. MYCIN was developed in 1972, at Stanford University. The primarily function was to help clinicians to choose appropriate antibiotics for bacteremia and meningitis. MYCIN is also capable of providing the reasoning behind its recommendations. However, it was never widely used due to difficulties incorporating the system to clinicians' workflow (Pusic, 2004).

1.5 DATA MINING & PUBLIC HEALTH

The data mining applications have a very broad use in public health. These applications affected clinical and non-clinical aspects of public health. While it allowed the surveillance of disease and infections, it also affected the health policy making in public health. Data mining applications allowed evidence based medicine which resulted in an improvement of health outcomes through ensuring adherence to clinical guidelines. Data mining also affected public health research by accelerating the time needed to identify a research population and by enabling the study of large group of patients retrospectively as they were treated for multiple conditions. Data mining not only affected public health by detecting or preventing diseases, but also by predicting new trends and uncovering situations associated with these events.

2.0 THE VETERANS HEALTH ADMINSTRATION (VHA)

The Veterans Healthcare Administration has one of the most developed electronic medical record systems in the world. About 8.8 million veterans' clinical information is recorded. Table data are generated for each veteran including health summaries, appointment lists, progress notes, discharge summaries, consult request, active medications, lab reports and imaging. As one of the US's largest integrated healthcare systems, the VHA has always been a source of information for researchers and healthcare providers. To be able to obtain customized data, researchers as well as VHA leaders and managers rely on computer programmers and other IT specialists. Retrieving information from data warehouses is expensive, time consuming and subject to human error (Pakhomov, 2008). In addition, unified understanding between researchers and computer programmers about which data is needed and which data is available has always been a major obstacle. After adding the time needed to analyze the data obtained to analyze relationships and examine trends and patterns, this whole process can take anywhere from a few days to several months. These studies very often provide the information retrospectively, and cannot be used for clinical monitoring during the course of treatment and cannot be relied on for timely access to needed information.

Though very helpful information is stored in tabled data, considerable and more important information waits to be analyzed in the free-text portion of electronic health records.

The unstructured clinical notes that are typed in free text (i.e., natural language without well-defined structure), contain patient's symptoms, physical findings, treatments, medications and laboratory results.

2.1 VETERANS ENGINEERING RESOURCE CENTER (VERC)

The VERC has maintained a strong quality improvement culture that had been implemented by its senior management. The Center's mission is to "lead the continuous improvement of healthcare in the VHA through the application of knowledge and expertise in systems engineering and operations management". Many of the employees have a quality or industrial engineering background. The center employed 10 full time VA staff, 5 fellows, and 25 part time contractors.

To expand the level of knowledge and expertise, the Center partnered with several academic institutions in the Pittsburgh area including the University of Pittsburgh, Graduate School of Public Health. They offer summer internships and fellowships which are primarily of interest to industrial engineers, IT programmers and healthcare administrative students. The Center also offers its expertise for training at the facility and VISIN levels.

The VERC culture encourages individual initiative, especially regarding innovative ideas for healthcare solutions. Summer interns are advised that it is their responsibility to dive in and take the lead in assigned projects. Accordingly, VERC is the home of many prototypes and startups in healthcare technology. Based upon the concept of facilitating learning through games and real life simulation, a simulation game was designed to help new administrators implement access strategies and reduce the waiting time for appointments. Another example is a Tele-medicine program for aphasia treatment called PIRATE (Program for Intensive Residential Aphasia Treatment and Education). The prototype of the PIRATE program was built using a smart phone and IPad.

2.2 THE VETERANS INTEGRATED SYSTEM TECHNICAL ARCHITECTURE (VISTA)

VistA is an integrated Electronic Health Record (EHR) information technology system with application packages that share a common data store and common internal services. The data store and VistA kernel are implemented in the Massachusetts General Hospital Utility Multi-programming System (MUMPS or M) computer language, and the Computerized Patient Record System (CPRS) graphical user interface (GUI) is implemented in Delphi (US Department of Veterans Affairs, 2013).

VistA/CPRS is used to record all clinical and administrative information related to the care of the veterans. VistA is deployed universally across VHA at more than 1,500 sites of care, including each Veterans Affairs Medical Center (VAMC), Community Based Outpatient Clinic (CBOC) and Community Living Center (CLC), as well as at nearly 300 VA Vet Centers.

The VistA databases can be searched by a programmer who can write queries in the M language. VistA can also be searched using VA FileMan which is the VistA database management system (DBMS) that runs in any American National Standards Institute (ANSI) environment. The majority of VHA clinical data is stored in VA FileMan files and is retrieved and accessed through VA FileMan Application Program Interfaces (API) and user interfaces. Although FileMan allows retrieving considerable amount of data, it is not capable of completing complex data searches (U.S. Department of Veterans Affairs, 2013).

3.0 THE DREAM PROJECT

The first prototype of DREAM was created by a multidisciplinary team of VERC's summer 2012 interns with the help of VHA physicians from different specialties. The team held meetings with physicians to better understand their needs and the best way to address them. With the broad scale of diseases and medications in the clinical field, the team decided to address the design by specialty. Due to increased interest and enthusiasm from endocrinologists as well as the number of clinical indicators available for diabetes mellitus management, the team decided to focus on the endocrinology department.

Based on the interviews, physicians were mainly interested in a timely clinical quality indicators screening for active patient care. For example, one relevant indicator is the number of current diabetic patients actively seen in the clinic who received the routine retinal check. Physicians were also interested in using free-text clinical notes for Pharmacovigilance. Pharmacovigilance is defined as the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem (WHO). Analyzing the Free-text notes enables detecting drug- adverse event association which can be used for hypothesis generation as well as prediction of adverse event risk (LePendu, 2013).

Although the team had experience using and access to a variety of software utilities, including Microsoft Visual Studio, they chose to use Microsoft Access software, and Microsoft Visual Basic Access (VBA) for the ubiquity of the VA IT infrastructure. The engine was designed to allow searching for a single patient (Patient Search) as well as group of patients (Population Search). The patient search is a query for a single patient's medical record separated into lab, notes, and medication, while Population search will search for group of patients based on demographics, diagnoses, lab results, notes, and active medication classes. (See Figure 1)

By the end of summer 2012, the DREAM prototype was capable of identifying lists of patients based on the criteria selected. The search options would only return positive matches for active Veterans. The results would then be presented in an Excel spreadsheet with specific information about patients who matched the selected criteria. The population search was also capable of matching HbA1c with diabetic drug classes in a specified time range. The engine allowed medication- notes searching which would match specific drugs with a note keyword in a specific time range. The end result would be a list of patients' Information Exchange Network (IEN) that matched the criteria.

The patient search is IEN specific that would search the patient's file for specific keyword or lab value. (See Figure 2) The engine would show the patient's clinical notes with the keyword in the time range selected. The result would not highlight the keyword in the clinical notes.

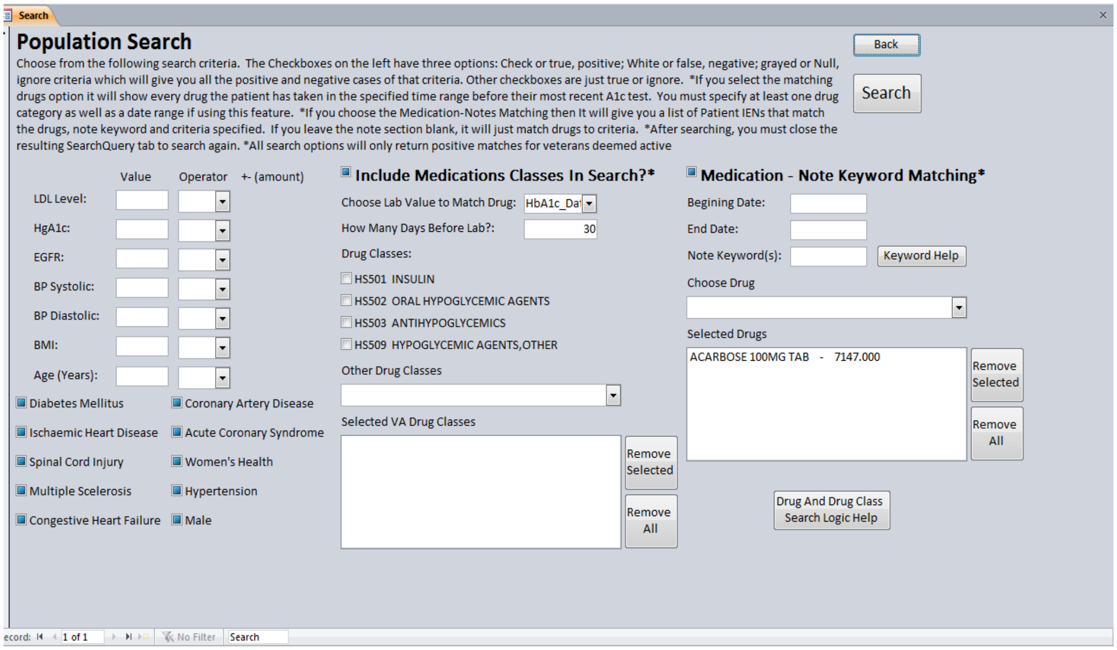

Figure 1: DREAM screen shot of population search

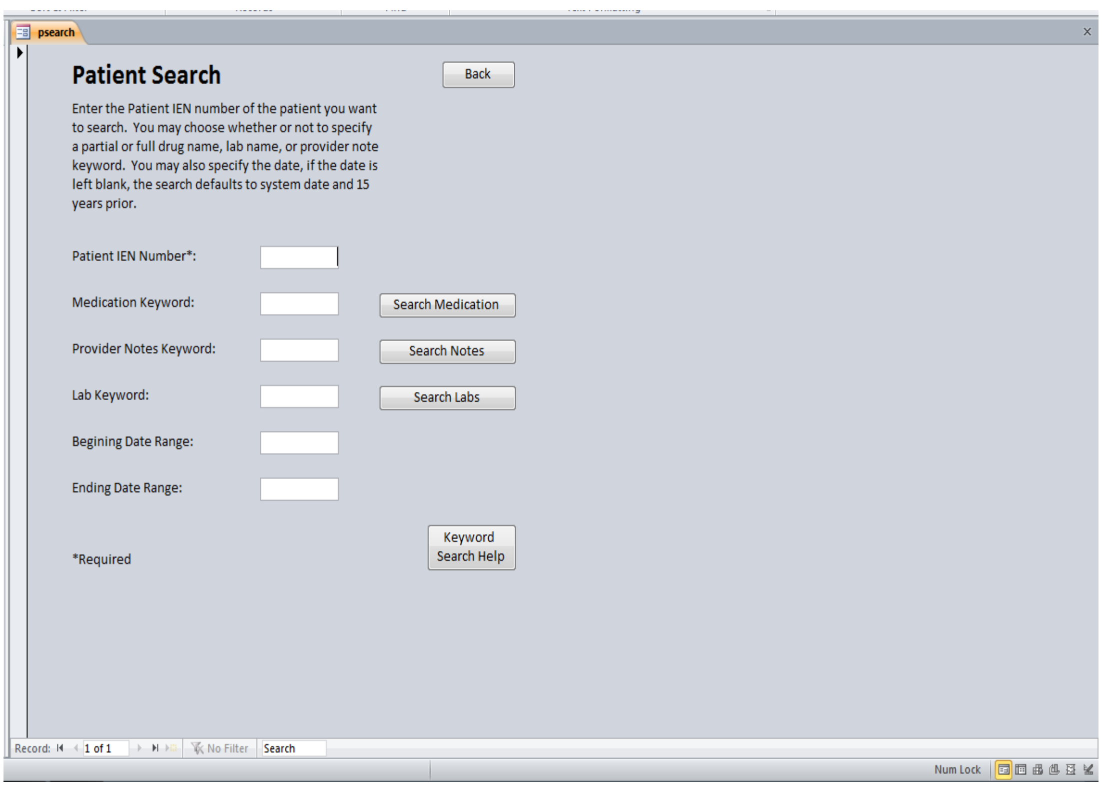

Figure 2: DREAM screen shot of patient search

3.1 THE PROJECT PERFORMANCE

At the VERC, the development of the DREAM project was affected by two major constraints. First, the security clearance required for handling sensitive information such as patient health records, limited the option of outsourcing the project to an external vendor. It also mandated the IT developers' usage of VA computers to build the engine. The limitation of computers to be used affected the software selection due to the uniqueness of the VA IT infrastructure.

The option of providing de-identified patient information to external programmers, though attempted later, was primarily avoided due to the extensive amount of data associated even with a very small sample of patients. In addition, the free-text clinical notes could not be released due to the high possibility that they would still contain patient identifiers.

Second, most of the IT developers who worked on DREAM had no previous experience using the VA informational health system. While some developers had limited familiarity, the expert VA IT programmers on the team were only available for guidance due to commitments to other projects. The extent of the IT programmers' learning curve definitely affected the time allocated for building the prototype.

Since the project had to be completed in phases, transitioning the project from one team to another was affected by their experience and their translational capabilities. The variability in experience among IT programmers proved an obstacle to the continuity of the project and delayed its development.

3.2 REFLECTIONS ON THE PROJECT

Though the design team agreed on the importance and potential of the DREAM project, there was some disagreement regarding whether to continue the project and how many resources the VERC is willing to allocate for it. While some team members felt that the prototype showed great potential and that it just requires more time to develop it fully, others were much less convinced of the value of the tool considering the time and resources it consumed.

Robert Monte believes that deficiencies in the initial design prior to creating the prototype adversely affected the project. Monte commented “everyone perceives the engine differently and acts based on his own perception”. However, Monte is a strong believer of IT developers’ freedom and believes that a solid initial design would have definitely affected interns’ creativity.

Some team members were strong believers in using the engine in clinical settings while others believed it is more appropriate for research applications because data obtained would still eventually need to be analyzed.

Defining the project’s success was a point of discrepancy as well. Some managers believed that it was good enough to provide a learning experience for the team and that it provided useful information about the VA IT structure while providing feedback on design parameters and the ability to design an effective engine. However, others defined success as reaching the initial end goal of full implementation in which the engine is used by healthcare providers to manage active patients in a clinical health settings.

4.0 A NEW STRATEGY AND A NEW DESIGN

After summer 2012 had ended and students were back to school, work on the DREAM project was put on hold. The available prototype, although still missing many features, was working and required testing and evaluation. When summer 2013 new interns arrived, the fever of DREAM caught them as well. The new team started compiling future steps needed in order to reach the original goal. These steps included adding as well as editing current features. First, the current system lacked the option of the name of the provider or clinic as a selection criterion which resulted in a negative impact of availability of its use by physicians who wanted to use the tool to monitor their own patients seen in their clinics. Second, the engine did not create an index of searches and one search had to be concluded before starting a new one. Third, the time frame for drugs selected and diagnoses was limited and did not allow recognition of a hierarchy of diagnoses to detect primary vs. secondary diseases. Finally, the text search will only look for the exact word entered and cannot recognize different phrases used for same meaning.

The new team faced an obstacle trying to translate queries and previously built table data in order to add or edit features. With the absence of a detailed installation and configuration guide, editing the engine was a major challenge that could not be fixed within the limits of a 3-month internship.

During fall 2013, a decision was reached to outsource DREAM project to Carnegie Mellon University students as part of their information system project. The main drive for this decision was to get a prototype that other teams can easily build on. The Carnegie Mellon team did not have access to VA computers, and due to the long process of obtaining security clearance it was decided that a sample of 10 patients' information be de-identified and outsourced to be used externally. Due to the possibility of patient identifiers being in the free-text clinical notes, fake clinical notes were created and attached to patients' files. Since the prototype was to be built using computers outside the VA system, the CMU team had more freedom in choosing the software to use. Time constraint was a huge challenge since students were working on this project while having full-time class schedules.

The second prototype was designed as multiple search modules. The first module; Demographics, was designed for patients' vitals, demographics, and clinic. The results could be viewed as a pivot table or line chart. (See figure 3)

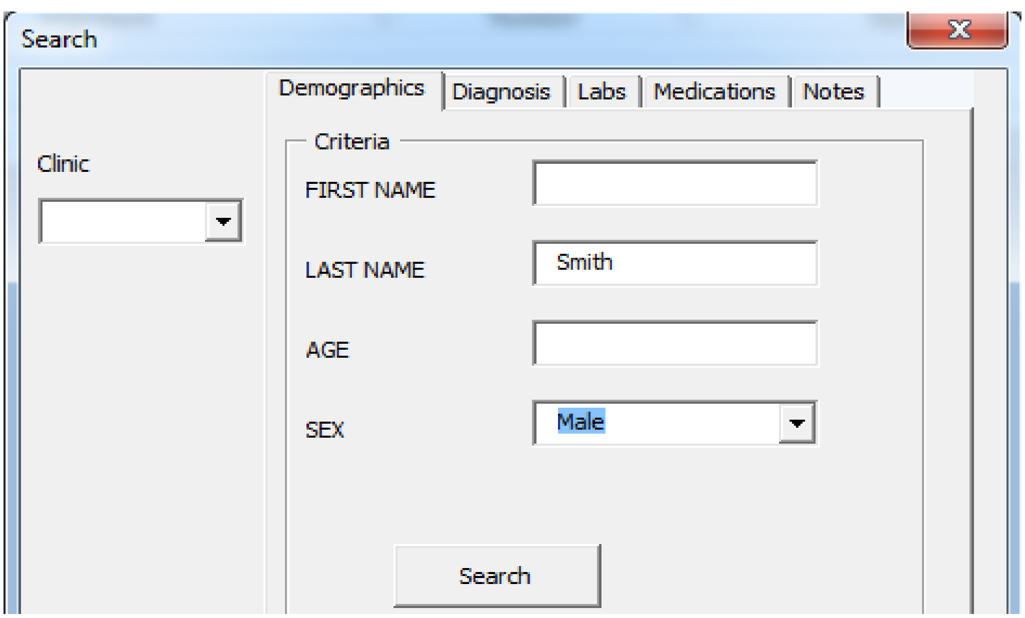

Figure 3: DREAM, 2nd version demographics modules

The second module; Diagnosis, allowed for identifying one single diagnosis or two diagnoses in relation to each other in time, i.e. primary diagnosis might have an earlier date than a secondary diagnosis. (See figure 4)

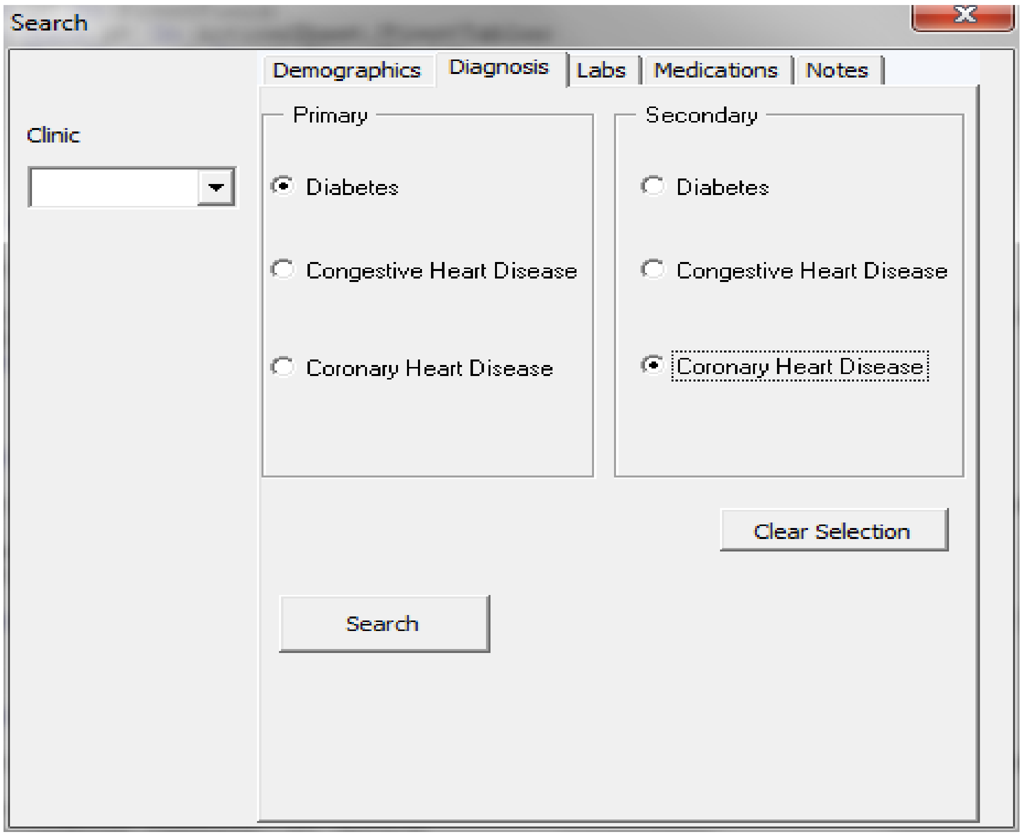

Figure 4: DREAM 2nd version diagnosis module

The third module; Lab, was designed for laboratory tests in a specified time frame. In addition, the results would be color coded, red for abnormal values and blue for normal ones. (See figure 6 &7)

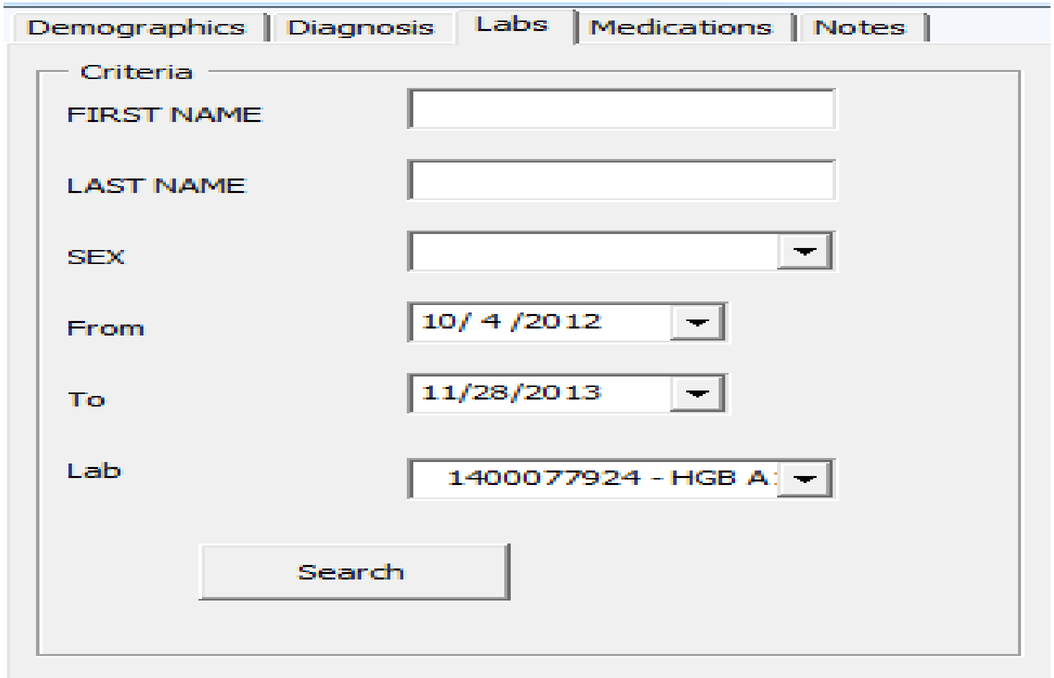

Figure 5: DREAM 2nd version, labs module

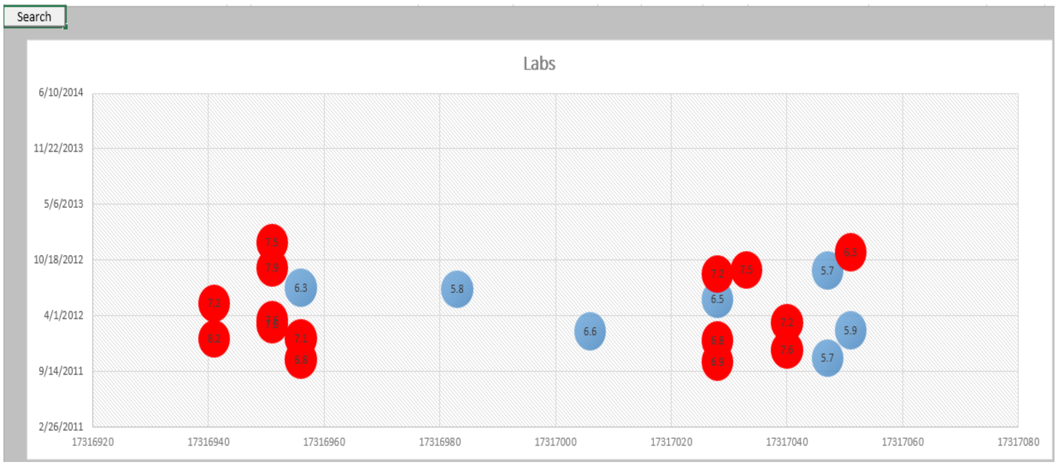

Figure 6: DREAM 2nd version, lab results example

The fourth module; Medications, was designed for active patient's medications. Results will also show an issue date for each medication. (See figure 8)

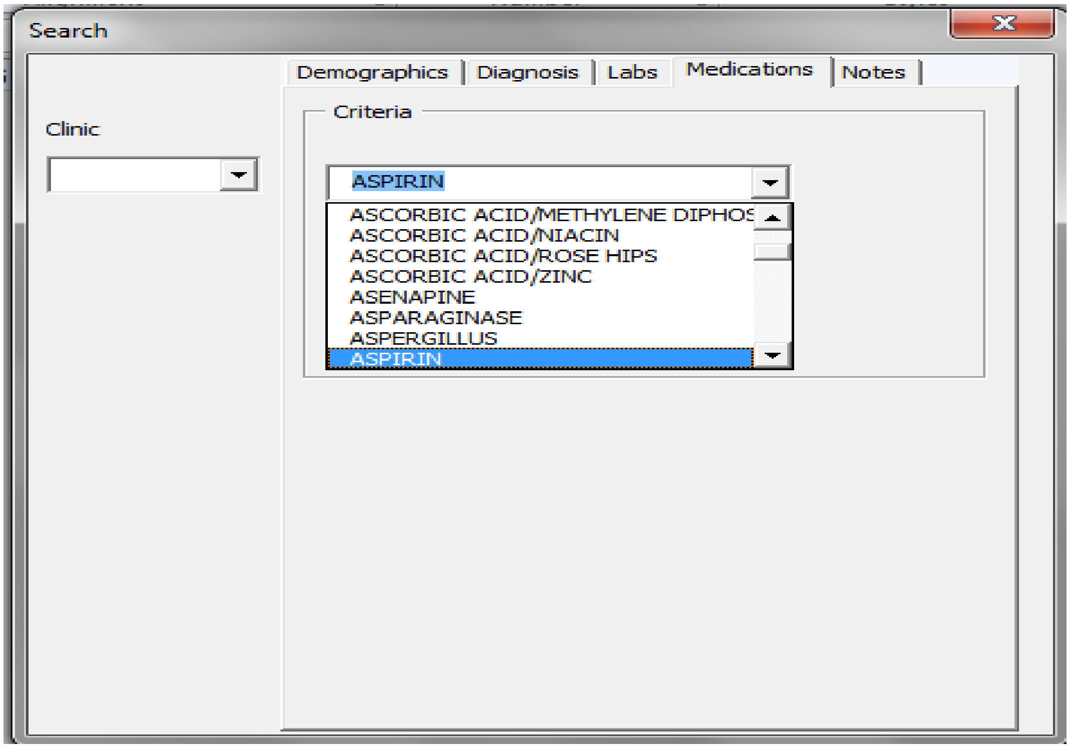

Figure 7: DREAM 2nd version, medications module

The final module; was designed for the free-text search. Due to limited time, only two keywords were used for this prototype, foot exams and eye exam.

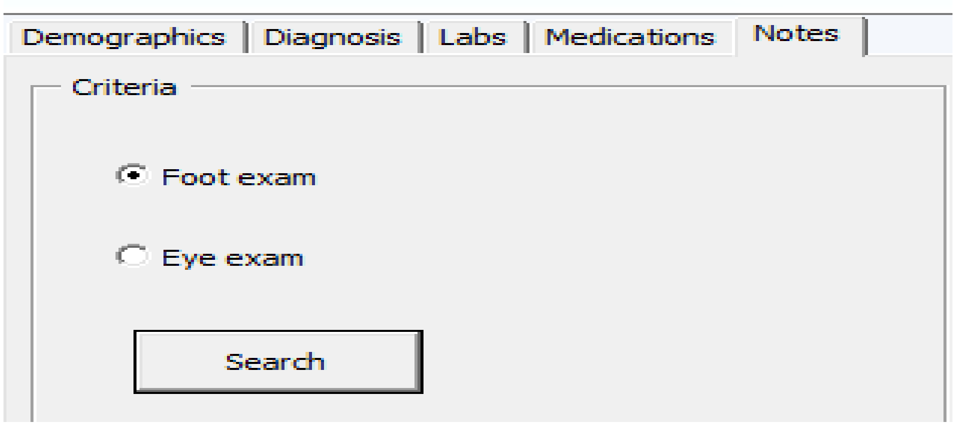

Figure 8: DREAM 2nd version, notes search module

By searching multiple clinical notes, the team compiled familiar phrases that are related to the keyword the researcher is interested in. (See table 1) The results would show patients' IEN who matched criteria. Whenever needed the original clinical note could be viewed for further analysis. (See figure 10)

Table 1: Foot and eye exam common phrases

Foot exam	Eye exam
DM Foot Exam	Refer for DM eye exam
Podiatry Clinic consult ordered	Teleretina exam
Foot ulcers	Diabetic Retinopathy Surveillance program
Foot care	
The 10gm monofilament exam, pedal pulses and foot inspection	
Diabetes foot exam	

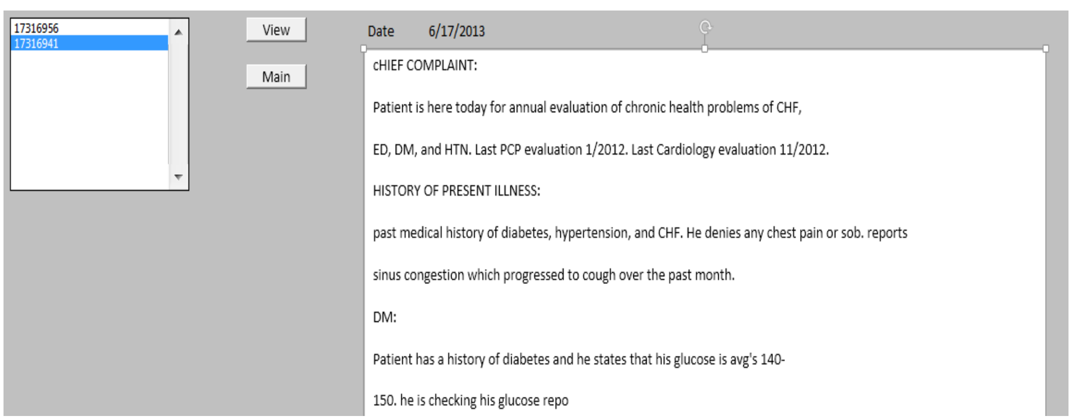

Figure 9: DREAM 2nd version, notes results example

5.0 CONCLUSION & RECOMMENDATIONS

It should be noted that this project was intended for summer interns and is not one of the VERC's main projects. Consequently, there is no master management plan for this project as the basis for comparing results against measurable objectives. Accordingly, this assessment is written based largely on the researcher's own point of view and experience within the VA.

5.1 CONCLUSION

The DREAM project was a great learning experience for the two groups of summer interns who worked on it in addition to the Carnegie Mellon team. The team members with clinical backgrounds learned much about IT technologies and the team members with IT backgrounds learned much about clinical quality indicators as well as medical terminology. Working with no clearly specified design nourished the teams' creativity and gave them an opportunity to explore and learn more about data mining. Team meetings with clinicians and senior managers were also helpful in giving students a snap shot of how external consultants work and provided them with the experience of translating needs into a working tool that will address those needs.

The experience associated with working with a unique system such as the Veterans Affairs IT infrastructure forced students to refine their ideas and to discover previously unknown issues and opportunities.

Defining the projects' scope is an early essential step to a successful project. Scope is identified through collaboration between the project's owner, sponsor, and stakeholders. Scope management is defined by PMI (Project Management Institute) as the process of defining what work is required and then making sure all of that work and only that work, is done. Lack of a defined DREAM scope, though this likely encouraged creativity, it also definitely lowered the chances of the project to succeed. Though scope change is a step that is hard to avoid, documenting and getting it approved minimize the risk of questioning and revising the scope when a new team is in charge.

Another area that represented a challenge to this project was risk management. Identifying risks, whether positive or negative, is done early during the initiation of the project. Although planning ahead will not eliminate risks, it will increase the probability for positive risks (opportunities) and lower the probability of negative risks (threats). The DREAM project faced multiple risks that the team was not prepared to handle. For example, the potential of a negative risk that the engine will not reach the speed needed for it to provide clinicians with a timely response was high. The amount of data the engine is searching was the main reason for the low speed and it usually caused the engine to crash. The potential risk of transitioning the project between teams was not fully assessed during the initiation of the project. Although the first prototype had not reached the initial goal, it achieved different results than were expected. Having to start from scratch had a great impact on the project's success and caused major delay in the implementation phase.

Great attention was given to physicians' needs and expectation of this project, however, a comprehensive stakeholder analysis was not performed. Identifying major stakeholders and determining their expectations, influence, and impact is a very important step. From the researcher's point of view, major stakeholders for this project would be the IT Department and the Information Security Office. Although they were consulted on technical issues and constraints, the team did not consult them on feasibility of the goals or the most efficient methods to achieve them. Both stakeholders have great interest in and influence upon outcomes, and their perceptions should be identified and addressed.

5.2 RECOMMENDATIONS

Based on my findings in this study, the following are my recommendations for action:

I. Since the learning curve of teams definitely affected the project's progress, a team with clinical and technological background will be most appropriate for this project. This team will require a project manager from the VERC's staff to facilitate the transition between different teams and to minimize the effect of discrepancy in experience levels among teams.

II. A scope management plan: a document that illustrates how scope will be planned, executed, controlled, and how to obtain acceptance to deliverables. Usually obtained through several meetings with key stakeholders and experts.

III. Stakeholder register and stakeholder analysis: a document that identifies all stakeholders and includes their major requirements, expectations, influences, and phases of the project where they are most interested.

IV. Work break down structure: a detailed description of all tasks required to complete the project and their relationship to one another.

BIBLIOGRAPHY

Bates, D., Gawande, A., (2003) Engl J Med 2003; 348:2526-2534June 19, 2003DOI: 10.1056/NEJMsa020847

Birman-Deych, E., Waterman, A., Yan, Y., Nilasena, D., Radford, M., Gage, B. (2005). Accuracy of ICD-9-CM Codes for Identifying Cardiovascular and Stroke Risk Factors. Medical Care. May 2005, Vol. 243-issue5-pp480-485

Botsis, T., Hartvigsen, G., Chen, F., Weng, C., (2010). Secondary Use of HER: Data Issues and Informatics Opportunities. 1; 2010:1-5.

Chiang, J., Lin, J., Yang, C., (2010) Automated evaluation of electronic discharge notes to assess quality of care for cardiovascular diseases using Medical Language Extraction and Encoding System (MedLEE). J Am Med Inform Assoc 2010; **17:3** 245-252 doi:10.1136/jamia.2009.000182

Cios, K., Moore, G., (2002). Uniqueness of medical data mining. Artificial Intelligence in Medicine. Volume 26 Issue 1-2, pages 1-24

Fonarow GC, Yancy CW, Heywood J, ADHERE Scientific Advisory Committee, Study Group and Investigators. (2005). Adherence to Heart Failure Quality-of-Care Indicators in US Hospitals: Analysis of the ADHERE Registry. *Arch Intern Med.* 2005; 165(13):1469-1477. doi:10.1001/archinte.165.13.1469.

Hardin, M., Chhieng, D. (2006).Data Mining and Clinical Decision Support Systems. 12/2006; DOI: 10.1007/978-0-387-38319-4_3

Khan, SR., (2012). Management and adherence to VTE treatment guidelines in a national Prospective cohort study in the Canadian outpatient setting. The Recovery Study. 108(3):493-8. doi: 10.1160/TH12-03-0169.

Kraft, M., (2002). Data Mining in Healthcare Information Systems: Case Study of a Veterans' Administration Spinal Cord Injury Population. Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)

Leatherman, S., (1991).Quality Screening and Management using claims data in a managed care Setting.17 (11):349-59.

Milovic, B., Milovic, M. (2012). Prediction and Decision making in Health Care using Data Mining. International Journal of Public Health Science (IJPHS) Vol. 1, No. 2, December 2012, pp. 69~78 ISSN: 2252-8806

Pakhomov, S., Hanson, P., Bjornsen, S., Smith, S. (2008). Technical Brief: Automatic Classification of Foot Examination Findings Using Clinical Notes and Machine Learning**.** J Am Med Inform Assoc 2008; **15:2** 198-202 doi:10.1197/jamia.M2585

Project Management Institute (PMI) is a professional site that presents project management tools & templates for it. http://www.pmi.org/

Pusic, M., Ansermino, M., (2004) FFA, MMed, MSc, FRCPC. Clinical decision support systems. BCMJ, Vol. 46, No. 5, June, 2004, page(s) 236-239 — Articles.

Rogers, G., & Joyner, E., Mining you data for health care quality improvement. SAS Institute, Retrieved from: www2.sas.com/proceedings/sugi22/EMERGING/PAPER139.PDF

Sen, A., Banerjee, & A., Atish P. (2012).Clinical decision support: Converging toward an integrated architecture. Journal of Biomedical Informatics - October 2012 (Vol. 45, Issue 5, Pages 1009-1017, DOI: 10.1016/j.jbi.2012.07.001)

(Electronic Version): StatSoft, Inc. (2013). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/.

U.S Department of Veterans Affairs. (2013). VistA Monograph. Office of Information Technology, Office of Product Development and Veterans Health Administration, Office Of Information and Analytics. Retrieved from:

Walters, J., (2009) Transforming Information Technology at the Department of Veterans Affairs. Retrieved from: *https://www.isaca.org/Knowledge.../WaltersVAReport-June09.pdf*