Patient Portal Messaging Safety Analysis for Reliant Medical Group

A Major Qualifying Project Report

submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in part of fulfilment of the requirements for the

Degree of Bachelor of Science

by

____

Chuqi Cai

____

Matthew Harrington

____

Qiuyi Hong

____

Kyle Orfan

____

Bing Yang

Mar.3, 2015

Sponsored by Reliant Medical Group

Dr. Lawrence Garber

____

Professor Diane Strong, Major Advisor

Table of Contents

Table of Figures

Abstract

The goal of this project is to explore any potential risk related to the messaging component of MyChart, the PHR System that Reliant Medical Group currently utilizes. Dr. Lawrence Garber and the staff of Reliant Medical Group are concerned that some patients may miss important messages sent through MyChart and therefore lead to possible increase in health-related risk and medical cost. Our analysis revealed a group of patients that were inactive or problematic MyChart users that Reliant should target for improvement. We provide a detailed plan for identifying problematic users.

Acknowledgements

This project would not be possible without the help of many individuals. As a group, we are incredibly grateful for those who provided guidance and support throughout the entirety of the study. Notably, we would like to acknowledge the following individuals for working with us tirelessly and offering what our group needed:

- Dr. Larry Garber, and Ms. Devi Sundaresan, our project sponsors at Reliant Medical Group. They worked tirelessly in assisting us with whatever we requested and we are so appreciative for everything they did for us.
- Professor Diane Strong, our project advisor. Professor Strong set the bar high for our group from the start and continued to push us to strive to take our research as far as we could.
- Robert Brown, Senior HPC System Integrator at WPI Academic & Research Computing (ARC) department. Mr. Brown setup a Linux Machine for the team to work on and provided excellent support for the project. The team will not be able to perform the cleaning and organization of the data as smoothly without his help.
- Siamak Najafi, Director at WPI ARC department, and Raffaele Potami, Research Data Scientist at WPI ARC department. For legal reasons, the data Reliant provided has to be securely stored on a server with restricted access. Mr. Najafi and Mr. Potami helped set up a secured research server to store the data.

Authorship Statement

This MQP was conducted by Chuqi Cai, Matthew Harrington, Qiuyi Hong, Kyle Orfan, and Bing Yang. We combined our efforts to analyze the data provided by Reliant Medical Group and then put our methods and findings onto paper.

- Chuqi Cai performed background research, analyzing and visualizing data. She also wrote methodology chapter and result chapter as well as did minor editing on both chapters.
- Matthew Harrington performed background research. He also wrote methodology chapter as well as did minor editing of the paper overall.
- Qiuyi Hong performed background research, organizing and categorizing data. She also wrote methodology chapter and result chapter as well as did major editing to the entire paper.
- Kyle Orfan performed background research and feasibility analysis. He also wrote introduction, methodology, and result chapters as well did minor editing to all three chapters.
- Bing Yang performed background research, organizing and categorizing data and created prediction algorithm. She also wrote methodology chapter and result chapter as well as did major editing to the entire paper.

Executive Summary

Background

Information Technology has been playing an increasingly important role in the U.S. healthcare industry. There is no doubt that healthcare information has been more convenient and accessible with the help of Electronic Health Record systems (EHR) and Personal Health Record systems (PHR). The U.S. government also established incentive programs to stimulate the adoption of EHR systems. However, the potential risk that could occur caught the attention of Reliant Medical Group.

Reliant Medical Group provides comprehensive health and medical solutions across Central Massachusetts and the Metro West region. Currently, Reliant Medical Group utilizes MyChart as its PHR system. In accordance with federal regulations, Reliant must make documented annual efforts to improve the usage of MyChart by both patients and doctors in order to benefit from the government incentive program.

Dr. Lawrence Garber, a lead physician at Reliant Medical Group, and Devi Sundaresen of Reliant's research department seek to identify any existing problems with the messaging component of MyChart and improve upon the current use of MyChart in conjunction with our team.

The team considered a number of methods to achieve this objective. Our team worked closely with Dr. Garber as well as our project advisor, Professor Diane Strong, to provide Reliant Medical Group with the most effective approach to a comprehensive solution. Per the recommendation of our advisers and colleagues, the most effective approach involved identifying problem patients within the MyChart data, and proposing methods to reduce the number of problem patients within Reliant's patient database.

Methodology

Our team chose to approach Reliant Medical Group's problem with a detailed and deliberate methodology. In particular, our team decided, in agreement with Reliant, to proceed in the following manner:

- Clean and organize the 2013 MyChart data provided by Reliant

- Perform a detailed analysis, including statistical analysis and visual analysis, of the cleaned and organized data
- Extrapolate conclusions from the results and create an algorithm for Reliant

In addition, our approach requires that we provide Reliant with a detailed outline of our entire approach so that the research department may reproduce our approach if they need to. Our team took great care to ensure that Reliant would have little trouble running the same procedure with their own resources within the research department.

Results

The team used MySQL in Linux environment to clean and organize the data. After consulting with Reliant, the group decided to only focused on the message type 1 (User Messages, typically test results) and type 11 (Patient Medical Advice Request) which were of the most concern by our sponsor. The results of our thorough analysis on the cleaned and organized MyChart data with SPSS showed that patients' MyChart usage did not vary based on age, gender or health condition. The group then discussed with Prof. Strong and determined that patients' historical MyChart usage could be an indicator for their behavior. We looked at each patient's total message per login and the percentage of problem messages one has. With the help of Receiver Operating Characteristic (ROC) Analysis, the team was able to identify that patients who log in on an average of 1.9 messages received and have more than 46% messages unread after 240 hours or more are most likely to be at risk of missing important messages. Another substantial group the team identified was inactive patients. Although the MyChart system does not consider these users inactive, they seemed to have stopped using MyChart since they never logged in after 2012. The group created an algorithm to help Reliant identify these patients with potential safety risks and tested the algorithm with the 2014 data. Our results showed that approximately 5% of the MyChart users in this study qualify as problem patients. These patients only have an approximate 50% likelihood of opening any MyChart messages that they receive.

Recommendation

The group recommended running the algorithm we created on the backend of Reliant’s existing system monthly and push the results to the production server so that the problem patients are flagged. A separate follow up study is recommended for analyzing the results of this change.

1.0 Introduction

Healthcare is a vital industry in today's global economy, and just like other leading industries, healthcare continually benefits from the advances made possible by modern technology. In particular, our group is studying the ways in which technology can benefit both doctors and patients by analyzing the utility of electronic health records (EHR) within a healthcare network. We are analyzing the potential safety risks associated with electronic health records and personal health records, and as a result develop a number of recommendations to help combat those risks.

Electronic health records are now an integral component in many advanced healthcare networks across the nation. In the U.S., every patient has the right to access his or her medical data as part of the Health Insurance Portability and Accountability Act, or HIPAA. Until recent years, this often meant that patients within most healthcare networks would need to request paper copies of their records from their provider's office. Each patient's health records contain all of the information that their healthcare provider documents, from personal information, medical test results, prescription information, and even data from routine office visits.

In recent years, however, healthcare providers across the United States began the transition to electronic health records. This is largely in response to the incentives and support given by the Centers for Medicare and Medicaid Services, or CMS. This U.S. government organization offers technical and fiscal support for healthcare providers that embrace meaningful and effective use of electronic health records within their networks.

Reliant Medical Group is a prominent healthcare provider in Central Massachusetts, with an experienced staff of physicians, nurses, and other medical professionals providing state-of-the-art care to all patients within their network. In order to remain a leader in the competitive healthcare market, Reliant not only strives to provide the best care, but also the best resources for their patients to manage their health. Several years prior, Reliant Medical Group began the implementation of MyChart, a patient access portal that allows doctors and patients to connect and share medical record information electronically. The implementation began as part of the EHR Incentive Program provided by the CMS.

MyChart helps doctors and patients share information regarding patient health in a manner that is comprehensive and easily accessible for all parties. The patient access portal displays all information that a physician has for a given patient. Per HIPAA requirements, only the patient and those medical professionals directly associated with the patient may view, edit, or maintain any of the information within the account. In MyChart, doctors may post medical test results, lab results, prescription information, and future appointment dates. Doctors and patients also have access to a messaging system that allows them to communicate discreetly through the MyChart portal. This system makes it far easier for doctors and patients to access all medical data in a much more convenient and updated manner. In theory, electronic medical records allow patients and doctors to have instant access to the most updated and comprehensive information available.

Reliant Medical Group is particularly interested in the messaging component of the MyChart portal and how Reliant's medical staff and patients utilize the features that it provides. The messaging system notifies the recipient, either a patient or a doctor/nurse, via their personal email address when they have a message waiting in their MyChart inbox. The user must then log in to MyChart and open the message in order for MyChart's delivery system to mark the message as read. Reliant Medical Group is concerned that patients and doctors may not open messages in a timely manner, or they may not open them at all. This could prove troublesome for Reliant, as their messaging system is a primary method of notifying patients of crucial medical updates and test data. This is a safety risk for doctors and patients alike, as there is a significant chance that patients may not receive critical health information. In addition, the usage of MyChart's messaging system is a reflection of the efficacy of electronic health records within the entire healthcare network.

Currently, Reliant Medical Group has a number of detailed datasets collected from the MyChart portal. The data contains information regarding patient and doctor usage of the messaging component of MyChart. In compliance with HIPAA regulations, our sample data does not identify specific individuals and only characterizes the dataset in very basic ways. These data entries do not identify the specific contents of the message or any other personally identifying characteristics, but simply give a number of characteristics that our team can use to establish connections to the data.

Reliant Medical Group wanted to know whether or not there were specifically measurable problems present within the network's usage of MyChart's messaging client. Reliant Medical Group's staff first aims to discover if the data collected provides evidence of repeating problems, and what those problems are. Their concerns are the safety concerns that accompany unanswered MyChart messages. Patients may miss important messages from health providers and therefore leads for potential health risks. Analysis of the data can help promote solutions in the case that a problem does become apparent.

To help provide Reliant Medical Group with the proper analysis and suggestions for improved usage of MyChart, the patient portal system employed by Reliant, our team performed a series of analytical procedures on each dataset. Each set requires proper organization, categorization ("binning"), and visualization followed by a full analysis. Our team then employs a number of methods of statistical analysis and prediction in order to assess the safety concerns regarding MyChart's patient messaging system. In the end, we provided Reliant Medical Group with prediction logic to help them identify the patients who are most unlikely to read their messages within a reasonable timeframe.

1.1 Feasibility Analysis

1.1.1 Cost-Benefit Analysis

Table 1: Cost-Benefit Analysis

	Year 0	Year 1	Year 2	Year 3
Benefits				
Savings from reduced number of unnecessary visit		36,122.82	54,184.23	72,245.64
Total Benefits		36,122.82	54,184.23	72,245.64
Costs				
Development Cost				
Project Cost	(15,000.00)		(15,000.00)	
Development Labor	(3,600.00)		(3,600.00)	
Implementation Cost				
Implementation Labor	(18,000.00)		(18,000.00)	
Total Costs	(36,600.00)		(36,600.00)	
Total Benefits - Total Cost	(73,200.00)	36,122.82	(19,015.77)	72,245.64
Cumulative Net Cash Flow	(73,200.00)	(37,077.18)	(56,092.96)	16,152.68

As we depicted in the accompanying Excel sheet, a cost-benefit analysis is a standard component of each Major Qualifying Project at WPI. Our team included this cost-benefit analysis in order to highlight the changes that may result from the implementation of our suggested approach. We do not have any control over the method in which Reliant Medical Group implements the results of our work, or if they implement them at all. Therefore, the costs and benefits that we outline are strictly our best estimates and are subject to variation dependent on a number of factors.

First, it is critical to note that our team's partnership with Reliant Medical Group was uniquely symbiotic, as WPI typically charges participating companies for MQPs, and Reliant charges in a similar manner for research projects. The standard charge associated with a Major Qualifying Project from WPI is approximately $15,000. This charge reflects the combination of our efforts as students, the efforts of our adviser, and the use of WPI's resources, such as research servers and support from Academic Research personnel. By the same logic, Reliant Medical Group typically charges a standard rate for their participation in research endeavors. As such, Reliant bills out the time involvement of Devi Sundaressen, data manager/analyst at its Research Department, on a standard hourly basis. As is apparent in the Excel sheet, these costs surmount very quickly and it is evident that this undertaking involves a significant number of costs. In our analysis, we are assuming that the development labor is eighty hours billed at forty-five dollars per hour. Our project team and Reliant Research Department reached a mutual agreement that the development labor will be donated and the project fee will be waived.

The implementation of our proposed methods would require an entirely new set of costs from Reliant Medical Group. It will take approximately 6-12 month before full implementation. That incurs an estimate of $18,000 in labor cost. Estimated labor cost information all come from Glassdoor (Reliant Data Developer Salary, 2015).

Although the costs of pursuing this endeavor do not seem overwhelming, the resulting benefits are exceptionally strong for all parties involved. Reliant Medical Group will be the prime benefactor of our team's efforts in the undertakings outlined ahead. At the inception of the project, Reliant expressed interest in determining whether or not there was a safety risk associated with their MyChart PHR system. Our analysis yields the benefit of reducing this risk, as we help identify who may or may not be a problem patient in MyChart. Additionally, our research will help to increase overall patient satisfaction with the MyChart experience and the efficacy of MyChart. This, in turn, will help to better establish Reliant Medical Group's reputation as a leader in the healthcare industry and give the company an image of unfaltering customer satisfaction. Reliant will also reap the benefit of increased physician efficiency due to the reduction in patient conflicts.

Furthermore, Reliant Medical Group will receive a number of tangible benefits from our team's efforts as well. As previously alluded, Reliant does not need to pay the standard fee for MQP involvement, as our collaboration with Reliant Medical Group is a mutualistic collaboration. In addition, by identifying problem patients and preventing their recurrence, Reliant Medical Group will be able to reduce the costs that typically follow unnecessary office visits. The projected cost savings are calculated based on the national annual average medical cost of $8,000 per patient and the national annual average of 332.2 physician visits per 100 patients (CDC, 2015). Our team is assuming that with the successful implementation of our recommendation, the number of unnecessary office visits due to negligence of important messages or medical advices will decrease by a certain percentage every year. In our calculations we assumed 1% decline in unnecessary office visits per year.

1.1.2 Organization Feasibility

Normally when a new system or major changes to existing system appear, several barriers may exist before the implementation becomes successful. However, the nature of our team's suggestion is unique. Our proposal will yield powerful changes for Reliant, but will not require any difficult or extensive changes to daily operations or to physician behavior. Additionally, the area of study in which our partnership with Reliant exists is fairly new territory for research at WPI and Reliant Medical Group alike. This field of study is ripe with opportunities for exploration, and as such Reliant Medical Group will find great value in the findings we present.

Furthermore, the continuation of our work does not interfere with the day to day job of the health providers at Reliant. It is a procedure that will run on the back-end of Reliant's database system. This means that doctors and other staff at Reliant will not need to change any of their daily habits in order for the research department to continue utilizing the procedures outlined in our study. Reliant Medical Group's executive staff may reserve the right to implement changes in the company or suggest that doctors influence patient usage of MyChart as a result of the information gained from the research, but our team does not currently suggest any such action.

Our team would be remiss in neglecting to acknowledge the effects of our study upon the users of MyChart, as the intention of the project is to provide benefits to those users. There is no way to effectively guarantee that patients will see direct benefits from this study when using MyChart. Patients that already use MyChart in a healthy and effective manner are likely to continue doing so, regardless of any knowledge of our efforts. Doctors, however, will experience a slight change in their routines when using the MyChart messaging system. Our project provides for the implementation of a new status label for the problem patients. Doctors will need formal education by Reliant on what a problem patient is and how they should approach problem patients. There is the possibility that physicians may be reluctant when first implementing these changes, as they will need to seek alternative methods of informing these particular patients about checking their MyChart messages. Their efforts will certainly not go unrewarded, as they will avoid the unnecessary phone calls and office visits that may associate with poor MyChart usage.

These changes are certainly not a simple task for any organization, but our team has little doubt regarding Reliant's ability to successfully implement the entirety of our suggestions. Dr. Lawrence Garber of Reliant Medical Group championed much of the research process, along with Devi Sundaresen of Reliant's research department. The research department at Reliant Medical Group handles projects of similar scope and size on a frequent basis, and understands the measures necessary to make those projects successful in an organization such as this. With a physician such as Dr. Garber championing our efforts, we feel that implementation will run smoothly and effectively.

1.1.3 Technology Feasibility

Generally speaking, the technical nature of the methods that our team examined required that we also examine the feasibility of Reliant Medical Group implementing similar hardware and software, as well as choosing personnel to undertake the continuation of our work.

As a team, our individual skill sets and project experience combine to give us a strong foothold for the data analysis as well as research and development of a proper proposed solution. We are a group of five knowledgeable students that have a strong background in statistical analysis. Additionally, we have the expertise of Dr. Diane Strong, who has a wealth of experience in the fields of data analysis and especially data studies related to healthcare industry. We believe that our skills with software such as IBM's SPSS suite, SAS, and R will help us to visualize and analyze the data provided. Furthermore, our team boasts a strong background in SQL programming and the usage of Linux operating systems. Our technical knowledge will couple well with the resources available to us on campus at WPI. We will utilize the server capabilities and state-of-the-art computing resources at WPI to complete this project and provide meaningful and easily understandable results that Reliant can implement for daily use.

Our team also investigated the current state of the computer and network hardware in use at Reliant Medical Group's research facility in Worcester. As of February 2015, Reliant Medical Group's research facility boasts a number of powerful desktop machines, as well as notebook PCs for physicians, all using Microsoft operating systems. In addition, the facility has a powerful network, including reliable wired and wireless connections for all staff members. In accordance with HIPAA regulations, Reliant's network is very secure in order to protect patient data. In respect to hardware, Reliant Medical Group is operating at the current state of the art and updates their hardware frequently.

The aforementioned hardware analysis is critical to ensure the continuation of our team's work, as we needed to determine that Reliant Medical Group's hardware can effectively support the software used in our analysis. In particular, we determined that IBM's SPSS software suite will run smoothly on the hardware currently in use, allowing for continued analysis.

Finally, our team considered the easiness of the Reliant staff replicating our methods if they wish to know how patient behaviors have improved or need to update the cutoff points in the algorithm in the future. Given the detailed description of our process in the Methodology section of this paper, we believe that the staff of Reliant Medical Group will not have any substantial trouble repeating the same analysis that we did if they need to. The SPSS software has a very easy learning curve, and per the recommendation of Professor Diane Strong, it is the best choice for completing detailed statistical analyses without significant coding or programming knowledge.

2.0 Background

In the background section, we cover general information regarding the healthcare industry, our project sponsor, and some more specific information about the technology we are working with during this project.

2.1 General Overview of Healthcare Industry

The healthcare industry is among the fastest growing and most costly fields in the United States. To see how expansive industry growth has been, look at the number of jobs healthcare has created. Data from the Bureau of Labor Statistics shows that employment within the healthcare and social assistance industry grew annually at a rate of 2.6% from 2000 to 2010. This was the second highest growth rate of any industry in America during that period. It has been estimated that annual growth will reach 3.0% between 2010 and 2020 (Henderson, 2012). Interestingly, while the gross amount spent by consumers annually on healthcare was not the highest overall; the rate of change in healthcare expenses between 2011 and 2013 was the largest of any industry.

Although the U.S. health system spent a higher portion of its gross domestic product than any other country, it still ranked 37th out of 191 countries based on its overall healthcare performances in 2000 (World Health Organization, 2013). A report “Mirror, Mirror on the Wall” produced by the Commonwealth Fund, compares the U.S. healthcare system with ten other countries. As shown below in Figure 1, the U.S. had the worst overall ranking of the 11 countries despite spending the most on health expenditures per capita in 2011.

EXHIBIT ES-1. OVERALL RANKING

COUNTRY RANKINGS
- Top 2*
- Middle
- Bottom 2*

	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US
OVERALL RANKING (2013)	4	10	9	5	5	7	7	3	2	1	11
Quality Care	2	9	8	7	5	4	11	10	3	1	5
Effective Care	4	7	9	6	5	2	11	10	8	1	3
Safe Care	3	10	2	6	7	9	11	5	4	1	7
Coordinated Care	4	8	9	10	5	2	7	11	3	1	6
Patient-Centered Care	5	8	10	7	3	6	11	9	2	1	4
Access	8	9	11	2	4	7	6	4	2	1	9
Cost-Related Problem	9	5	10	4	8	6	3	1	7	1	11
Timeliness of Care	6	11	10	4	2	7	8	9	1	3	5
Efficiency	4	10	8	9	7	3	4	2	6	1	11
Equity	5	9	7	4	8	10	6	1	2	2	11
Healthy Lives	4	8	1	7	5	9	6	2	3	10	11
Health Expenditures/Capita, 2011**	$3,800	$4,522	$4,118	$4,495	$5,099	$3,182	$5,669	$3,925	$5,643	$3,405	$8,508

Notes: * Includes ties. ** Expenditures shown in $US PPP (purchasing power parity); Australian $ data are from 2010.
Source: Calculated by The Commonwealth Fund based on 2011 International Health Policy Survey of Sicker Adults; 2012 International Health Policy Survey of Primary Care Physicians; 2013 International Health Policy Survey; Commonwealth Fund *National Scorecard 2011;* World Health Organization; and Organization for Economic Cooperation and Development, *OECD Health Data, 2013* (Paris: OECD, Nov. 2013).

Figure 1 – Health System Rankings (Mirror, Mirror On the Wall, Commerwealth Fund)

The same report also indicates the U.S. had an outstanding performance in delivering preventative and chronic care to patients. Interaction with patients, including communication, engagement, and prioritizing patient preferences was a bright spot. Granted, enhancement of coordinated care was necessary in order to improve the overall quality of healthcare.

In summary, the U.S. healthcare system is struggling with access, efficiency, equity, and healthy living. All of these factors have contributed to the low score of the U.S. received for healthcare. To provide better healthcare to patients, U.S. still has a long way to go.

2.2 IT and Healthcare

2.2.1 Electronic Medical Record (EMR)

An electronic medical record is a collection of electronic standard medical and clinical data about individual patients (Gunter, 2005; Haupt, 2011). This data has typically been paper-based in the past; EMRs are becoming more widely used as IT is playing a more important role in the healthcare industry over the past few years (HealthIT EHR, 2014). Some benefits and advantages of electronic medical records over paper-based medical records are:

- It is easier to track patient data over time. A provider/staff would not need to go through piles of paper to try to find information about a surgery a patient had years ago.
- It is easier for the provider/staff to identify those patients that are due for routine preventive visits and screenings such as annual physicals and immunizations. Paper cannot remind a staff member to contact a patient regarding these routine check-ins but EMRs can send notifications to providers to remind them of anything due in a specified timeframe.
- It is easier for providers to tell how patients measure up to certain parameters, such as BMI, blood pressure readings, heart rate, oxygen level, etc. When these values are only taken down by paper, a health provider might not be able to tell if they are off right away, especially those values that need calculation based on a few different measures, whereas electronic medical record could create an alert right after the values are entered.
- Based on what has been mentioned above, basically electronic medical records can help improve the overall quality of care in a practice.

2.2.2 Electronic Health Record (EHR)

While many people confuse Electronic Health Record (EHR) with EMR and use the two terms interchangeably, officially there are some essential differences between the two. As mentioned earlier, medical records used to be recorded on paper and now more providers are recording them with electronic software. EMR is only referring to this format of recording medical records electronically whereas EHR goes beyond the data collected in any provider’s office and includes a much more comprehensive health history of a patient (Haupt, 2011). EHR is designed for providers across certain areas or even the nation to share information involved in a patient’s care with the patient’s authorization.

For example, if a patient has a primary care doctor, an allergy specialist, an eye doctor, and a dentist, EHR allows these four healthcare providers to share necessary health records about this patient. The primary care doctor will know what allergy the patient has and avoid giving her medications that interfere with that allergy even if the patient forgot to mention it to the doctor. If the patient moves to a different state and needs to change doctors, her new healthcare providers will also be able to get access to all her past health history through EHR. The patient would not need to retell their medical history to the new office and possibly forget something important. There would be less chance of error in the process of the data transfer with EHR than if using paper records.

Another example of the benefit of EHR would be that if a patient lives in Massachusetts and is travelling in Florida and happened to have an accident in Florida where she completely lost consciousness, the doctor in Florida would be able to pull her health information through EHR and provide care to her. The patient might not remember any details about her own accident but her doctor back in Massachusetts would have full knowledge of what happened and how she was treated through the convenience of EHR. The major benefits and advantages of EHR shown through these examples are:

- Accurate and complete information about a patient's health condition is available to all related healthcare providers
- Health providers are able to provide quick care to patients when emergencies happen.
- Better coordination of care can be given to the patient by different providers.
- Practice efficiency can be increased and cost savings can be realized through improved diagnostics and outcomes, including less duplication of diagnostic tests.

2.2.3 Personal Health Record (PHR)

Personal health record (PHR) functionality is usually part of an EHR package. It allows patients or an authorized third person to access their own or their family members' health information through a secured portal. Originally PHR was introduced to healthcare providers and patients as part of a better quality healthcare experience. Recently, it has become one of the requirements under the meaningful use of the government incentive program (See Section 2.3 for more information about the Government Incentive Program). As healthcare software companies are dedicated to make healthcare quality and the experience better for patients, more user-friendly features are appearing in the patient portal. For example, there are features such as e-visits where a patient could guide herself through simple symptom diagnostics to find out what she could do if she is not feeling well. There are also patient-provider interactions through these secured portals. Patients could schedule appointments or send provider messages regarding their health conditions online.

2.2.4 IT in Healthcare

Much like other industries around the world, healthcare has incorporated the use of information technology to improve the quality of healthcare and increase efficiency of day-to-day operations. This technology (healthcare technology) is more advanced than the basic registration of patients and scheduling their appointments. The new technology includes tools that benefit both patients and medical organizations. Electronic accessibility of medical records and information can be invaluable if patients and physicians utilize it consistently.

2.2.4.1 Organizational Level

There are many different types of information from a medical record that must be stored and kept for future reference. For instance, information on a medical record includes previous diagnoses, medications, immunizations, allergies, medical imaging, and past visits to doctors and clinics. It also includes old lab results from hospital or office labs (HealthIT Benefits, 2014). Having all of this information in an electronic health record (EHR) improves the accessibility of the information to the doctors treating the patient.

A vital aspect of an EHR is it contains all information entered at all healthcare locations. This ensures a doctor treating a patient at one organization will see the information entered by another doctor at a different location from a previous visit. No information gets lost and all treatments and notes are there for the treating doctor to see. (See more information about EHR in Section 2.2.2)

2.2.4.2 Patient Level

Similar to the EHR, a personal health record (PHR) offers the same medical information to an individual patient that a doctor sees in a specific EHR. In a PHR, an individual can see their old medical history, allergies they have, and medicines they are prescribed. Ideally, it also allows patients to reach out to their doctor and schedule an appointment or ask a particular question regarding their health.

According to a doctor at the Cleveland Clinic, a clinic that implemented IT working in conjunction with healthcare, PHRs increase patient engagement while increasing their awareness of their medical situation (Shehadi, 2012). He expects organizations to look at new ways to reach out to patients, including, potentially, the use of social media. (See more detailed information about PHR in Section 2.2.3)

2.2.4.3 Negative Reviews

Not everyone is buying into this idea of EHRs and PHRs however. According to a study done in 2012, 30% of doctors polled said EHRs had a negative impact on doctor-patient relations (Pulley, 2012). One anonymous employee working in healthcare was quoted in the report, saying "staring at a screen leads to a more impersonal encounter." Other complaints include doctors getting frustrated with the inability to adapt to the technology and the thought of "…entering data rather than interacting with the patient" (Reese, 2012). The same report went on to say that 26% of doctors polled said they were less productive when using the EHRs (compared to 23% saying their production improved).

Ron Sterling, author of Keys to EMR Success and nationally acclaimed EHR expert, believes the statistics in this report show the positive effect of EHRs in the grand scheme of things. "The decrease in productivity is really about the doctors… [Productivity] is contingent on how well the doctor worked that EHR into their patient model" (Reese, S., 2012). In simpler terms, he is saying the decreased productivity among individual doctors reflects their own negative experience with it, while the 23% of those polled who said EHRs increased office efficiency is a more appropriate assessment of the positive experience of the overall practice.

Admittedly speculating, Sterling felt some doctors' negative experience might be caused by their unwillingness to learn a new technology or system when they are accustomed to the old-fashioned paper filing system. The thought is, while it may take a physician longer to electronically enter medical information in the short-term, it ensures others in the practice will manage and utilize the records more efficiently. Whatever one's opinion is about EHRs and PHRs in healthcare, there is no doubt that IT has had a profound effect on healthcare practices.

2.3 Government Incentive Program

Health IT has been the mainstream trend in the healthcare industry for the past 5 years. In 2011, the U.S. government started two programs called the Medicare EHR incentive program and the Medicaid EHR incentive program to stimulate adoption of EHR systems in the healthcare industry. Centers for Medicare & Medicaid Services (CMS) is responsible for both programs. These programs provide incentive payments to eligible professionals, eligible hospitals, and critical access hospitals when they adopt, implement, or upgrade their EHR systems or demonstrate meaningful use of all the certified EHR technologies. Eligible professionals can only choose one of the programs in which to participate. Although these two programs are similar in many ways, they have some significant differences from each other (see table 1).

Table 2: Difference between Medicaid and Medicare EHR incentive Program

Medicaid EHR Incentive Program	**Medicare EHR Incentive Program**
Every state runs its own program	Run by CMS
Program runs from 2011 through 2021	Program runs from 2011 through 2016
Maximum incentive amount is $ 63,750 (across 6 years of program participation)	Maximum incentive amount is $44,000 per physician (across 5 years of program participation)
No Medicaid payment reductions if you choose not to participate	Payment reductions begin in 2015 for providers who are eligible but choose not to participate
In the first year, providers can receive an incentive payment for adopting, implementing, or upgrading a certified EHR	In the first year and all remaining years, providers must demonstrate meaningful use of certified EHR technology to get incentive payments

After President Obama issued an Order Implementing Sequester on March 1st, 2013, Medicare incentive program started to reduce payments to eligible professionals and hospitals. This reduction does not apply to Medicaid EHR incentive program since it is exempt from the mandatory reductions.

To receive a consecutive incentive payment, providers in both programs have to attest that they use their EHR system meaningfully every year. Both programs consist of the same 3 stages for meaningful use. Different stages focuses on different aspects of EHR use. Stage 1 focuses on data capturing and sharing; Stage 2 focuses on advanced clinical processes; Stage 3 focuses on improving health outcomes. Every year, eligible professionals have to follow different rules to attest their meaningful use status. During the first year of each stage, providers only need to meet all the requirements and report a consecutive 90-day data during the calendar year to demonstrate their meaningful use status. In the following years, they have to provide data for the full calendar year to demonstrate they are using EHR system meaningfully.

To ensure all providers use EHR systems meaningfully, CMS has a standard including all the requirements for providers to follow. Providers must meet all of the requirements successfully. Otherwise, they cannot receive a payment and there are no partial payments. For Stage 1, the providers have to meet all requirements including 15 core objectives, five of ten menu objectives with at least one public health-related objective, and 6 Clinical Quality Measures (CQMs) to show the meaningful use status. For Stage 2, the providers have to meet all the requirements including 15 core objectives and three of six menu objectives (See Appendix I for all the objectives' details).

Eligible professionals in the Medicaid EHR Incentive Program can receive an incentive payment during the first year by just demonstrating they are adopting, implementing, or upgrading to a certified EHR system. For adopting a certified EHR system, providers only need to provide evidence of EHR system installation. For implementing a certified EHR system, providers need to provide evidence that they have started using this system. For example, providers can provide training for their staff or enter data for patient demographic information into the systems. For upgrading a certified EHR system, providers can provide evidence that they have added some new features to current systems (CMS Introduction to Medicaid, 2012; CMS Introduction to Medicare, 2014).

Both programs encourage patient use of a PHR system. Starting in 2014, CMS have added additional requirements for eligible hospitals and professionals. Providers who are in Stage 1 must meet Measure #1 and Stage 2 must meet Measure #1 and #2 (CMS Access Tipsheet, 2014).

- Measure #1:
  - Eligible Professionals: At least 50% of the unique patients can access their report and health information in 4 business days after the information is available.
  - Eligible Hospitals: At least 50% of the unique patients who are discharged from the inpatient or emergency department can access their information online within 36 hours of discharge.

- Measure #2:
  - Eligible Professionals: At least 5 percent of all unique patients can view, download, or transmit their health information to a third party.
  - Eligible Hospitals: At least 5 percent of all unique patients who are discharged from the inpatient or emergency department can view, download or transmit their health information to a third party.

2.4 Reliant Medical Group

In 1929, Dr. John Fallon founded the Fallon Clinic, the first group medical practice established in Central Massachusetts, which was renamed Reliant Medical Group in 2011. Reliant has 13 Primary Care offices around the region and they specialize in more than 30 critical medical fields, including cardiology, physical therapy, and neurology. The Reliant mission statement reads:

> "The mission of Reliant Medical Group is to maximize the health of our patients and the community through expert medical care, compassion, innovative delivery models, medical research and education, and the appropriate use of healthcare resources" (Reliant Medical Group About Us, 2014).

Through patient-centered, coordinated healthcare, Reliant strives to become the preeminent healthcare provider in the market.

Dr. Lawrence Garber, the medical director of informatics at Reliant Medical Group, is a big part of what Reliant does to separate themselves from their competitors. Dr. Garber and Reliant utilize secure electronic medical records in order to improve efficiency of day-to-day operations. In addition to their EHR, Reliant has a PHR, called MyChart, which provides patients with a secure portal to view their own medical records and reach out to their doctors if the need arises. Dr. Garber has spearheaded this movement and continues to search for ways to improve healthcare for both physicians and patients.

2.5 Key Competitors to Reliant Medical Group

2.5.1 UMass Memorial Health Care System

The oldest member in the UMass Memorial Health Care system (UMMHC) is Memorial Hospital, which was founded in 1871. Mr. Ichabod Washburn, a Worcester industrialist, endowed it in memory of his wife and daughter through bequest (Lamar Soutter Library Psychiatry Resources, 2014). Now, UMMHC is the largest health care system in Central and Western Massachusetts with more than 12,000 employees. As the clinical partner of University of Massachusetts Medical School since 1998, UMMHC consists of UMass Memorial Medical Center in Worcester and four member hospitals (Clinton Hospital in Clinton, HealthAlliance Hospitals in Leominster and Fitchburg, and Marlborough Hospital in Marlborough) across MA (About UMMS, 2014). The health care services that UMMHC supports in their hospital centers involve heart and vascular care, orthopedics, cancer, diabetes, surgery, newborn intensive care, children's services, women's services, emergency medicine and trauma. UMMHC also provides home health and hospice programs, behavioral health programs and community-based physician practices (About UMass Memorial Health Care, 2014).

Aspiring to deliver improved patient care and respond to growing physician needs, UMMHC began to set up a new electronic medical record (EMR) and patient management solution from Allscripts in 2009. As a result, 2,000 affiliated physicians from UMMHC would have access to digital records (Hardy, 2009). This implementation was also a response to the federal HITECH Act (Health Information Technology for Economic and Clinical Health), which is part of the American Recovery and Reinvestment Act of 2009. HITECH Act provided around $44,000 to EMR-using physicians beginning in 2011 (UMass Memorial Selects Allscripts, 2009). Allscripts would offer training including onsite, post-implementation and customization to help employees gain a better understanding of how EMR could enhance their services. The company would also integrate with the previous information system to minimize the potential for errors. Another reason thar UMMHC chose Allscripts was their identification of a strategic partner with dbMotion, who created the interoperability platform, allowing physicians to have access to all available information about a patient. Having such a platform minimizes the gaps of health care between hospital-based and office-based providers (Prestigiacomo, 2010).

According to former associate CIO Karen Marhefka, not only is UMMHC working on improving their patient experience from an ambulatory perspective, but also on the organization level. The program Karen's team has been working on is named Cornerstone, which is "a multi-year, multi-entity corporate initiative to implement common patient clinical and financial systems and to reduce variability in processes and workflows across the system." (Brogna, 2010). The project aimed to integrate all the data from the seven various systems that UMMHC has used as sources – ambulatory EMR, inpatient EMR, perioperative system, enterprise master person index, clinical portal, health information record management system, and emergency department system (Prestigiacomo, 2010). The success of this project would change the ways of interaction and communication between health care and patients (Brogna, 2010).

2.5.2 Saint Vincent Medical Group

Saint Vincent Medical Group (SVMG) is a multi-specialty physician group with 13 locations in the Central Massachusetts area. SVMG is affiliated with Saint Vincent Hospital located at Worcester Medical Center (Saint Vincent Medical Group, 2014).

On their website, SVMG hosts Athena, another powerful EHR client, under the learning center for educating patients (Saint Vincent What are EHR, 2014). All of their providers have access to patients' data in an EHR to get complete medical histories as to provide more practical plans and decrease medical risks.

2.5.3 Central Massachusetts Independent Physician Association

Central Massachusetts Independent Physician Association (CMIPA) was established in 1998 and is based in Worcester County. It aims to help independent physicians in MA to provide quality services to patients and support each other using this network. CMIPA now has more than 200 community-based members and represents most sub-specialties. As a vision they've talked about providing tools to their members with best-available technology. One of the technology enhancements that CMIPA provides for their members is an EMR system (CMIPA Facts, 2014).

In 2013, CMPIA was called by the state Health Policy Commission (HPC) as witness for the annual hearing regarding the healthcare cost trend in the Commonwealth. In the notification letter, the HPC attached a list of questions regarding CMPIA's current state regarding healthcare cost, their related costs and future plans (Seltz, 2013). Required by HPC, CMPIA handed a written testimony with all the responses. In the reply letter, CMPIA identified EMR as a limitation factor to improve the quality and efficiency of care within the organization. According to CMPIA, different medical groups and hospitals in MA utilized disparate clinical systems, which also led to the difficulty for the systems to exchange information with one another. There were not any effective solutions to advance this situation while possible candidates were cost prohibitive and not completely functional. They highly recommend that EHR vendors develop a universal platform which grants easy communication with other systems. On the other hand, the cost of extracting data from various EHR systems was expensive. Inside CMIPA's provider group, there were 17 different systems while no existing policy introduced a price range of charges from vendors for such standardization. Due to the strong need of organization to reinforce the data into common platforms, it is necessary to universalize an EHR system template (Glazier, 2013).

2.5.4 Massachusetts Health Information Exchange

In 2012, Massachusetts launched an electronic medical record exchange program to afford healthcare providers the opportunity to share the resource of patient medical records (Mass eHealth Institute Research Study, 2014). The motivation behind this project is to strengthen patient medical care experience. According to a research study generated by Massachusetts Health Information Highway (Mass HIway) on the topic of provider and consumer health IT, the participation percentage of transmitting patient information electronically is only 26 for the general healthcare providers. The programs of those who engaged differed from the government project and private exchanges.

2.6 Epic Systems

Reliant Medical Group, the sponsor of this project, acquired its EHR and PHR from Epic Systems (See Appendix II for a list of major customers of Epic). Epic, founded in 1979, is a private company that makes electronic health record software for mid-size and large medical groups, hospitals, and integrated healthcare organizations. All of their applications are developed, installed, and supported in-house. With Epic's small client base, 315 customers, they are focused to provide their customers with personal service and support—from initial implementation and training to ongoing support and optimization (Epic Client Services, 2014). The key to Epic's success is that they have clinicians, developers, and process experts in their leadership team. It is not focused on technology only; instead, they have people who are deeply experienced in both patient care and healthcare technology (Epic About Us, 2014). It has different software dedicated to physician groups and hospitals. Epic also provides Mobile & Portal access, Patient and Revenue Management, Enterprise Intelligence, and other healthcare related software.

Epic's hard work has earned several awards and recognitions:

- #1 Overall Software Suite for 2013 by KLAS (based on 25 separate performance measures across multiple enterprise categories) (Epic Recognition Klas, 2013)
    - Best Ambulatory EMR
    - Best Surgery Management

  - Best Radiology
  - Best Patient Accounting and Patient Management
  - Best Practice Management
  - Best Health Information Exchange
- Epic's customers have won more Davies Awards than any other vendor (Epic Recognition Davies, 2014)
  - HIMSS Nicholas E. Davies Award of Excellence is an award that recognizes excellence in the implementation and value from health information technology (EHR).
  - See Appendix II and Appendix III for a list of award winning Epic customers and the definition of the stages.

2.7 Key Competitors to Epic Systems

According to the Top 100 EHR Company list from Medical Economics, the major competitions for Epic Systems in the industry include Cerner Corporation, Allscripts, NextGen Healthcare Information Systems Inc., GE Healthcare, and U.S. Department of Veterans Affairs (Medical Economics Top EHR Companies, 2013).

2.7.1 Cerner Corporation

Cerner Corporation is a global company supplying healthcare information technology solutions, services, devices, and hardware. It mainly provides one EHR system and one EMR system.

- Ambulatory EHR which is designed for physicians with comprehensive yet flexible documentation, and automates reporting processes for business. It can support more than 40 specialties (Cerner Ambulatory EHR, 2014).
- Acute Care EMR which is a database with a comprehensive set of capabilities. It can provide real-time access to patient and clinical information securely across disciplines and departments (Cerner Acute Care EMR, 2014).

2.7.2 Allscripts Healthcare Solution

Allscripts Healthcare Solution is an American company, based in Chicago, Illinois, that provides physician practices, hospitals, and other healthcare organizations with electronic healthcare records and practice management technology. There are three different EHR system products focusing on different target audiences from Allscripts.

- Allscripts Professional EHR is designed for small to mid-sized physician practices. It can provide all basic needs for physicians with an affordable price. In addition, it has hundreds of pre-loaded templates and more than 20 specialties. It also connects to over 50,000 pharmacies (Allscripts Professional EHR, 2014).
- Allscripts TouchWorks EHR is designed for large, multi-specialty physician practices. It provides very comprehensive solutions with automated clinical decision support to satisfy different users' specific needs. It also offers nearly 800 multispecialty Care Guides to support effective and efficient patient care in the ambulatory setting (Allscripts TouchWorks EHR, 2014).
- Sunrise Acute Care EHR is designed for hospitals and health systems. It includes Electronic Health Record and Computerized Physician Order Entry (CPOE), which is the solution for minimizing the medication ordering errors. This product is designed to manage large hospitals and health systems by coordinating care across location and departments, supporting critical decision-making, and automating processes for accuracy and safety (Allscripts Sunrise Clinical, 2014).

2.7.3 NextGen Healthcare Information Systems

NextGen Healthcare Information Systems is an American company supplying EHR software, financial, and Health Information Exchange (HIE) solutions for hospitals, health systems, physician practices, and other healthcare organizations (NextGen, 2014). Its EHR system, NextGen Ambulatory EHR, provides more than 25 specialists. It includes ICD-10, which is a system that physicians and other professionals currently use to code all diagnoses, symptoms, and procedures recorded in hospitals, physician practices, and other healthcare systems (ICD-

10, 2014). Furthermore, it has been certified for Meaningful Use of Stage 2 (NextGen Ambulatory EHR, 2014).

2.7.4 GE Healthcare

GE Healthcare is a subsidiary of General Electric (GE). Its business mission is to "provide transformational medical technologies and services that are shaping a new age of patient care" (GE Healthcare About Us, 2014). GE Healthcare provides three electronic healthcare system solutions:

- Centricity Practice Solution is designed for providing secure clinical records storage to capture, store and transmit essential patient data. It provides both EMR Module and Practice Management (PM) Module, so that the users can not only have customized features while implementing EMR system but also have the PM service to support the seamless interoperable work flow. In addition, it includes ICD-10.
- Centricity EMR offers users "an ambulatory EMR that integrates well with revenue cycle management and practice management solution". The advantages of this EMR system is being easy to use, being easy to maintain, and being connected with other data systems. It also support patients to schedule appointments, receive billing information and lab results, and communicate with providers.
- Centricity Enterprise is an advanced EHR system which is designed for large healthcare groups such as community hospitals, academic medical centers, and integrated delivery networks. This product can help users to reduce the potential errors and increase the billing efficiency across the delivery network.

2.7.5 U.S. Department of Veterans Affairs

U.S. Department of Veterans Affairs (VA) who operates the largest integrated health care system in the United States, developed and implemented Veterans Health Information Systems and Technology Architecture (VistA). This system helps VA to deliver the best quality medical care to Veterans by providing an integrated inpatient and outpatient electronic health record. VA is developing the next generation of Vista, which is called HealtheVet. It will enhance "capabilities and flexibility to adapt to health care and technology innovation, and continually improve health care for the nation's Veterans" (VistA, 2014).

2.8 MyChart

Reliant Medical Group faces a unique challenge, they have the technological resources in place to help patients and doctors share vital medical information, but they need to find a way to ensure that the patients and doctors use the technology to its full capabilities. The group currently employs the MyChart system, a product of Epic Software Group. MyChart digitizes the medical records that doctors make available to patients, often referred to as "charts". Reliant Medical Group hopes that MyChart will make these charts more accessible to patients, and therefore prompt a stronger connection between doctors and patients, as well as a stronger patient interest in their health.

MyChart boasts a number of features, including the ability for doctors to send various medical test results directly to patients. Dr. Lawrence Garber, a physician of nearly 30 years at Reliant, is a strong proponent of the benefits that MyChart offers both doctors and physicians. Dr. Garber stresses the importance of the messaging component of MyChart. He notes that the MyChart system allows for physicians to see directly if a patient opens a message in the MyChart portal. For each message in the MyChart portal, a patient will receive an email message noting that they have a message in their MyChart inbox. This message does not contain any sensitive data, it simply notifies the patient of the status of the message. Dr. Garber and Reliant Medical Group are concerned with the unintended risks that may arise from the use of a messaging system such as MyChart.

There is currently a significant compilation of clickstream data regarding the patient portal and the usage of the messaging system. Reliant Medical Group seeks to analyze trends in this data in order to discover if there are significant safety concerns within the MyChart messaging system and if there are others methods might benefit specific patients.

2.9 Key Competitors to MyChart

Currently, MyChart serves as Epic Software's personal health record, or PHR. Personal health records differ greatly in nature and content, but offer similar features to those outlined in the

MyChart description. As this would suggest, MyChart has a number of strong competitors in the market today. Some of these competitors are proprietary to the companies that Epic competes with, and others bear no specific connection to healthcare providers or EHR companies. Regardless of how patients use PHR software, MyChart must remain competitive with the other PHR offerings in the market in order for Epic to remain a viable option for health providers across the globe.

Allscripts, one of Epic Software Group's largest competitors, offers a proprietary PHR tool to its customers. This application, named FollowMyHealth, offers features similar to those that MyChart provides, such as patient history tracking, prescription information, and contact information for physicians (Allscripts FollowMyHealth, 2014). FollowMyHealth requires significant input from patients in order for the system to work properly, meaning that it can prove difficult for healthcare providers to persuade patients to invest the required time into the system.

Almost every electronic health record software provider offers a unique incarnation of a personal health record system. These records contain much of the same features that MyChart offers. Some of these in-house solutions offer easy ways for doctors and patients to access the data that health providers already have, while others require that patients add their own data. These solutions are all viable competitors to Epic's MyChart.

2.10 IT and Safety Concerns

As mentioned in Section 2.3, the U.S. government started EHR incentive program to support professionals, hospitals, and other healthcare organizations with the adoption of EHR systems to provide better healthcare services and reduce medical errors. There is significant evidence showing that Health IT has improved the quality of healthcare and reduced medical errors successfully, yet it may also be causing some harm. With widespread adoption of EHR, EMR, and PHR systems, some professionals started to be concerned about potential harm from the use of health IT. Some reasons why there are potential risks in utilizing IT in the healthcare industry are:

- There could be absence of data measures and a central data warehouse to collect and analyze.
- There are contractual barriers in gathering data, such as nondisclosure and confidentiality clauses. Because of these barriers, patients and providers may hesitate to share IT related data and adverse events.
- Some vendors include certain clauses and terms in their contract, such as "hold harmless clauses", and try to escape responsibility for errors or defects in their systems. (Chou, 2012)

Besides the reasons mentioned above, a key risk that brought our sponsor, Reliant Medical Group, to attention is that software-related safety issues, such as software coding errors within the systems or human errors while using the software, may exist. Even though these might not happen very often, additional things may go wrong on the health providers' side, including:

- Relying too heavily on electronic processing. A health provider might click through default settings without paying attention which could result in wrong patient information.
- Sending test results or care summaries to the wrong person. This might be caused by human error or system error.
- Not sending urgent messages to patients in time. A health provider could be very busy and might not have time to review and send patient the test results for days. Urgent messages might get delayed in the process.

On the other hand, patients could also behave differently than they are expected to and therefore result in risks:

- If a patient is an active user of the patient portal and rely only on the information through the portal, she might miss important information if an error occurred on the provider's side.
- If a patient uses the portal from time to time, there is a huge chance that she might miss urgent messages from the health provider.
- If a patient is not familiar enough with the personal health record system, she might misuse the system.
- For features like E-Visit in MyChart (see more details about MyChart in Section 2.8), a patient might mislead herself if she misunderstands the terminology.

3.0 Methodology

In this section, we describe the steps we took as a group to develop our research. From cleaning the data to analyzing it, the following subsections go into detail regarding what we did and how we did it.

3.1 Data Characteristics

The data set we received from Reliant Medical Group consists of four tables:

- **IB_Message** includes a higher level overview of all messages going around the system. The IT specialist from Reliant Medical Group who works with us on this project has already trimmed the table to only messages relevant to MyChart. This table includes some key measures, such as Message Type, Message Priority, Message Due Date, and so on. This table also has a MyChart_Message_ID field which can be used when linking to the more detailed MyChart Message table.
- **MyChart_Message** is a table containing more detailed information about MyChart messages specifically, such as what type of MyChart message it is or if the message is from a patient to the provider or from the provider to the patient.
- **Msg_Stat_Audit** contains information about the status changes of a message. There are 6 statuses that a message can possibly have: create, sent, pend, done, retract, and edit. We are most concerned about two of them, sent and pend. When the receiver of the message opens it, the status changes to "pend". What we would like to analyze, is the time between message status changes from "sent" to "pend", which captures the length of time between when a message was sent and when it was viewed by the receiver.
- **MyChart_Patient_Demographics** includes some basic demographics of the patients we are studying. The data provided to us has been de-identified to make sure that there is no way to identify any single patient through the data set. Basic demographic information includes: gender, year of birth, alive or deceased as of Sep 2014. These records will help us with identifying patient behavior patterns.

- **LOGINDATA_2013** 2013 is a table processing all users' login data within MyChart messages. It includes the time that each user had logged into their MyChart account, as well as the type of message they received. With each user's unique StudyID, we are able to link the patient's login time with the type of message they received in order to find whether they responded to the message or not.
- **Proxy** includes patients' StudyID and whether they have a proxy or not. A patient may have a designated proxy who has full access to his or her account. Children under the age of 12 are required to have proxies, normally their parents or legal guardian, and only their proxies can access their accounts. Adolescents between the age of 12 and 17 are still required to have a proxy but are also allowed to access their own accounts. All adults over the age of 18 will be the primary account user but have the right to choose a proxy if they want to. However, in the database we used from Reliant, information regarding proxy log-ins into patient accounts as it relates to "Read Messages" is unclear. This made it impossible for us to run an analysis using proxy as a factor.
- **RAFScore** is a table containing patients' StudyID and their RAF Score. RAF Score is a measure the system calculated for patients based upon their health status. A RAF Score less than one indicates a patient in a healthy condition. A RAF Score higher than one indicates the patient in an unhealthy condition.

3.2 Organize and Categorize Data

Below we describe how we organized and categorized the data. The results of our organization and categorization are presented in Section 4.1.

3.2.1 Clean Data

The first step our group needed to take was to clean the data. The data we received included a lot of information that would not be utilized in our research, such as messages between nurses and doctors. It was important that we use only the necessary data so the server would not be slowed down. As a result, the group ran a selection query on the entire database we received from Reliant and kept only the MyChart-related records. Cleaning the data removed 80% of the data we received, all of which was deemed irrelevant to our study.

3.2.2 Organize Data

After successfully cleaning the data, we had to organize it in a way that would make analysis of the data easier. Our group is concerned with two types of Reliant system communication: messages from providers to patients, and messages from patients to providers. Splitting up the data into these two communication pathways seemed like the most sensible way to organize the information.

3.2.3 Categorize Data

With the data organized, we could categorize it even further, based upon whatever information we may need at any point during our analysis. Our group categorized the patients into two groups, "good patients" – patients without any messages unread after 10 days, and "potential problem patients" – patients with any messages unread after 10 days. Other characteristics, such as age and gender, are also considered for later stages of the analysis process.

3.3 Analyze and Visualize Data

Below, we explain the process of analyzing the data, including our use of visuals throughout said process.

3.3.1 Analyze Data

As a group, our goal is to perform a detailed and useful analysis of the data supplied to us by Reliant Medical Group. The quality and content of that analysis are a direct result of the methods chosen to complete the analysis procedure. For statistical analysis, our team used SPSS to perform cluster analysis.

3.3.1.1 Cluster Analysis

As mentioned earlier, statistical analysis is critical in our process of targeting and characterizing problematic MyChart users. In particular, this is very important for the characterizing of potential bad users in the future. Cluster analysis is one of the most academically endorsed methods of characterizing these patients given the size of the data set.

Cluster analysis is particularly useful for large sets of data, and our "good patient" and "bad patient" data sets were excellent candidates for the use of such analysis. Furthermore, cluster analysis is most helpful for establishing groupings that would otherwise remain unknown to the analyst. The groupings are based off of pre-established variables within the dataset, all of which are categorical, but don't provide a direct cluster for the dataset.

There are a variety of accepted methods for cluster analysis in the statistical community, and each method offers a number of advantages and disadvantages based on the types of data used and the nature of the dataset itself. After researching the usage of K-Means cluster analysis, Hierarchical cluster analysis, and Two-Step cluster analysis, our team determined that a Two-Step cluster analysis would prove to be the most efficient and effective option for creating a detailed cluster analysis.

The Two-Step cluster analysis is especially effective for datasets that contain multiple categorical variables that an analyst wishes to create clusters with. In addition, the Two-Step cluster analysis allows for the creation of any number of clusters, all based upon a moving Euclidean average algorithm.

3.3.2 Visualize Data

After categorizing the data and computing basic statistical analysis of all variables individually, the next step is visualization. Data visualization can help to show differences in patient behavior. This process can be based on one variable or multiple variables.

3.3.2.1 Histograms and Scatterplots

- Histograms

  Histograms show how data varies within groups of similarly categorized characters, like gender. Different patterns in histograms carry different meanings (Stocks, 2012).

  - A symmetrical graph indicates that the dataset has an intensive performance over a specific variable. Figure 2 shows two typical symmetrical models. Both suggest a stable performance of one or two results.

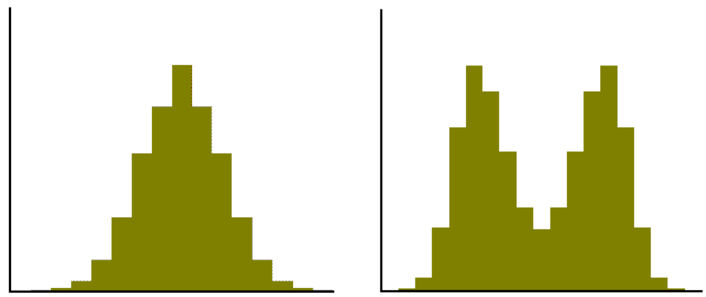

Figure 2 – Symmetrical Graphs (https://www.msu.edu/user/sw/statrev/images/norbih01.gif)

- A skewed graph indicates a certain performance trend over a given variable. The following are two typical skewed models. Both suggest either a positive or negative trend in the data.

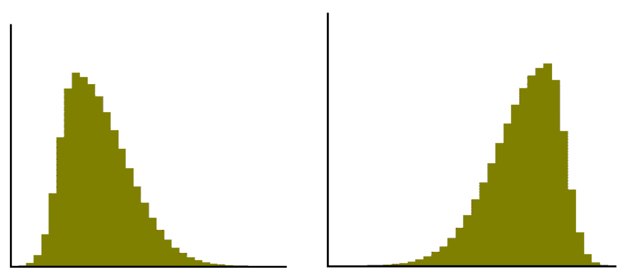

Figure 3 – Skewed Graphs (https://www.msu.edu/user/sw/statrev/images/posskwh1.gif)

- A multimodal graph is a combination of symmetrical, skewed and other typical statistics models. This model suggests that there is no noticeable connection between the variable and the dataset.

- Scatterplots

Scatterplots can be used to show how data changes in a group with two quantitative variables, such as age and years attending school. Different appearances of scatterplots carry different messages (Stangor, 2014).

- A linear graph indicates a linear relationship over two variables. The following are two typical linear models. The graph on the left represents a positive linear relationship, and the graph on the right represents a negative linear relationship.

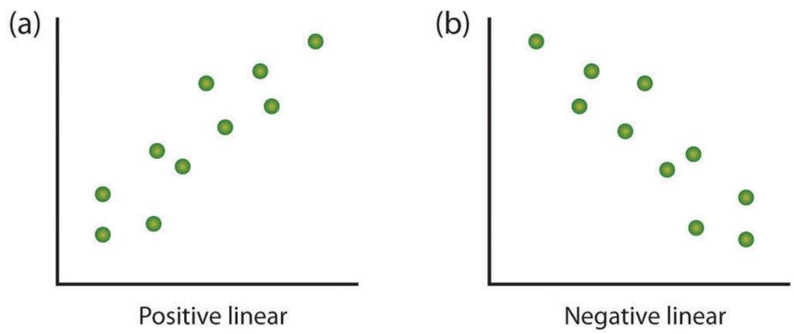

Figure 4 – Linear Graphs(https://new.edu/resources/psychologists-use-descriptive-correlational-and-experimental-research-designs-to-understand-behavior)

- A curvilinear graph indicates a performance trend with regard to two variables. The graph on the left represents a positive curvilinear relationship, and the graph on the right represents a negative curvilinear relationship.

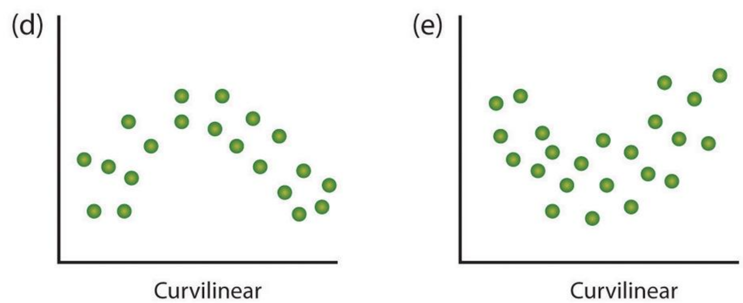

Figure 5 – Curvilinear Graphs (new.edu)

- An independent graph indicates no clear relationship between two variables. The following is a typical independent model.

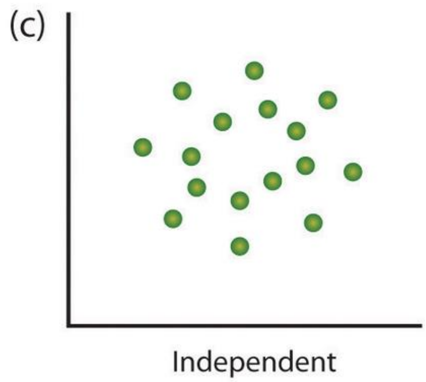

Figure 6 – Independent Graph (new.edu)

3.3.2.2 Receiver Operating Characteristic (ROC) Curve

In order for a statistical analysis of nearly any type of dataset to be truly effective, certain procedures are necessary in order to establish cutoff points among the data. These cutoff points help determine the levels, based on any number of variables, at which the data is no longer statistically significant. This is similar to determining outliers in a dataset, but with a mathematical calculation rather than a visual determination.

The curve generated from the Receiver Operating Characteristic (ROC) is the most effective method for establishing a cutoff point in a specialized set of data, such as the Reliant patient data. Technically speaking, the curve plots sensitivity, or the rate of “true positives”, against 1-specificity, or the rate of “false positives”. For the purpose of this study, it is critical to establish the understanding that the ROC analysis creates a curve that predicts the number of false positives and true positives for each given point. In most standard statistical cases, plotting sensitivity against 1-specificity allows for a test to exclude a certain condition, or essentially prove it irrelevant. Conversely, plotting specificity against 1-sensitivity allows for a test to confirm the relevance of a factor in a given dataset. These tests operate on the theory that the “true” condition, whether positive or negative, will always outweigh the “false” condition.

The figure below (figure 7) illustrates the principles behind this process, as well as the conditions that yield each of the following:

- True Positive Rate (TPR): The rate of statistically proven presence of a given trait or factor (e.g., blue eyes)
- False Positive Rate (FPR) : The rate of apparent presence of trait or factor, but no statistical significance
- True Negative Rate (TNR): The rate of statistically proven absence of a given trait or factor
- False Negative Rate (FNR): The rate of apparent absence of a trait or factor, but has statistical significance

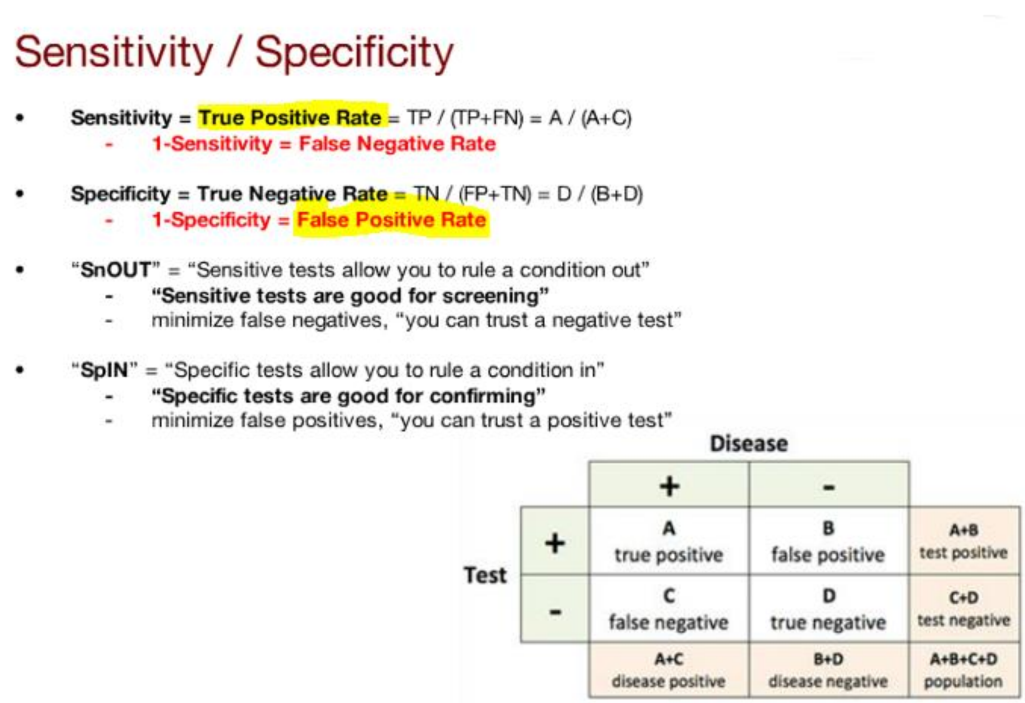

Figure 7 – Principles behind ROC Curve (provided by Dr. Garber, source: AMIA.org)

When analyzing an ROC curve, one must take care to analyze the area underneath the curve. This area is a measure of the credibility of the analysis procedure. Essentially, the greater the curve, the more accurate it is.

An analyst must then assess a number of distinct points in the curve and determine which point will become the cutoff. The cutoff is best determined as the point with the highest sensitivity and lowest specificity, meaning that it has many true positives and few false positives. The point will likely rest in the upper left region of the graph, and the analyst must then find a point to correspond to the visual interpretation (see sample ROC Curve in figure 8 below).

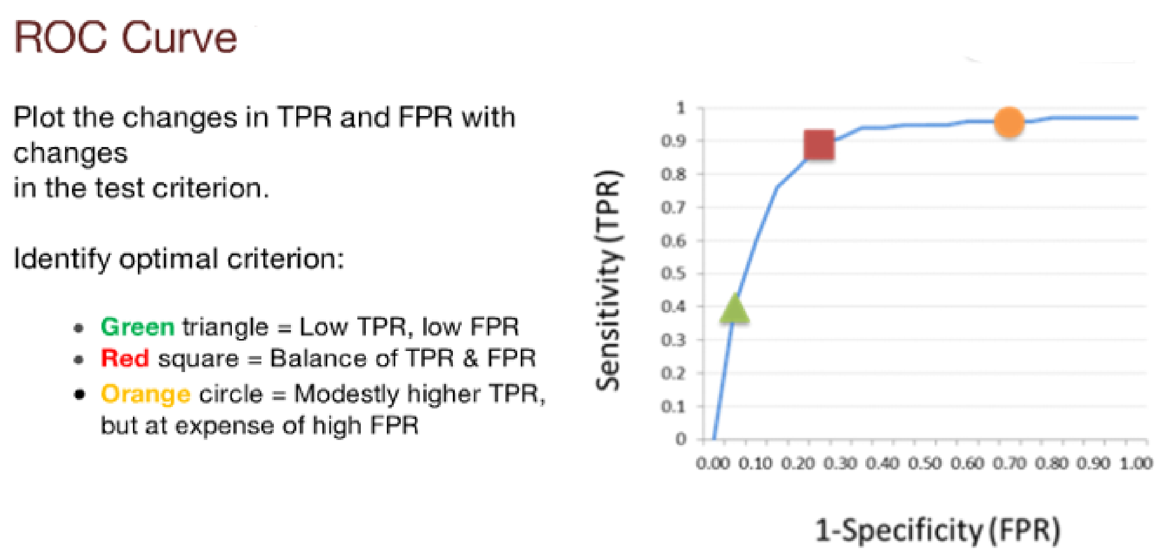

Figure 8 – Sample ROC Curve (provided by Dr. Garber, source: AMIA.org)

3.4 Prediction Algorithm

As discussed earlier, our team's utilization of Reliant Medical Group's data involves the creation of a specially designed algorithm based on a number of important criteria. This algorithm will help our team effectively calculate the likelihood that a given patient from the data set will open a message in the MyChart portal. We are using this algorithm and its results in conjunction with our previously outlined categorization approach to analyze trends in Reliant's current patient data and to predict trends in their patients' future actions.

3.4.1 Variables

There are a number of variables to consider in the process of creating this algorithm. Below are the variables our team suspects bear a connection to the likelihood of a patient reading a message in MyChart before fully analyzing the data. As mentioned in the background chapter, there are a number of patient confidentiality and security concerns that our research must obey in order for our work to remain HIPAA compliant. Each of the following variables is a reflection of those HIPAA standards combined with the data that Reliant Medical Group feels is necessary for the analysis of patient response patterns.

- Message Urgency: Are the contents of the message of high importance? Is it imperative that the patient open the message within 48 hours of sending? Ranked 1 for lowest urgency, 3 for highest urgency.
- Patient Notification of Pending Message: Did a member of the Reliant staff give the patient an advance warning that they would receive a message in MyChart? Is this a routine message that the patient typically expects?
- Time of Delivery: When was the message delivered?
- Time to open: How long did it take for the patient to open the message (if they open it at all)?
- Does patient demographic information affect patients' user behavior? (Age, Gender, Health Condition, etc.)

After carefully analyzing the data, the team found that it is not possible to predict patients' user behavior based on any single factor. As a result, the factors used in our final prediction algorithm are as following:

- Problem Message Percentage: What is the percentage of problem messages (overdue and never-read messages) for a patient? If a patient has only missed one message out of ten, then it is more likely that this patient is an active user and just missed a message due to some random reasons. In contrast, if a patient missed 100% of his messages, this is a good indicator of this patient not being an active user/responder.
- Message per Login Ratio: During the analysis, the team found out that many patients will neglect the messages intentionally when they know what the messages are for. For instance, if a patient is actually an active user of MyChart and knows that he/she is expecting a test result, he/she will log on to MyChart, go straight to the Test Result section and leave the message unread. To avoid including these patients in our algorithm by mistake, the team decided to use message per login ratio to determine the activeness of a patient.

3.5 Testing

Upon creating an algorithm, it is important we test and check it to ensure we have designed it to be consistently accurate and correct. There are a few different ways to test it.

3.5.1 Peer-Checking

The most basic form of testing and checking an algorithm is peer-review. Having another pair of eyes review the algorithm and test it for accuracy is always helpful. If you write the algorithm, you are much less likely to catch any mistakes you have made. Running it by a few people to double check it can only help.

3.5.2 Desk-Checking

Essentially, desk-checking involves a person, acting as the computer, to check the algorithm. The individual plugs in different sets of data into the algorithm and goes through the algorithm in detail each time, carefully seeking out any deficiencies or problems. Any problems that become exposed require us to go back to the drawing board.

Desk-checking generally involves "drawing up a table of the variables and outputs from the algorithm and then working through the algorithm, with sets of test data, recording the changing values of the variables. The final results can then be compared to the expected results to see if the algorithm works as it should" (NSW HSC, 2014).

3.5.3 Walk-Through

A walk-through is a quick method used to determine what the algorithm does and what the logic is behind it. It is done mentally and is just a quick check as to see what is trying to be accomplished.

3.5.4 2014 Data

We were provided with the data from 2014 for the purpose of testing. Running the data through the algorithm we have devised is another way to test the accuracy of our proposal. Any algorithm that is created absolutely must be checked for proof that it works. Every implementation is checked the same way. Take an input, calculate the output by hand, and then compare that output to the one provided by the algorithm.

3.5.5 Comparison & Analysis

We run the same logic and analysis on the 2014 data as we did for analyzing the 2013 data. This is to ensure that all our findings and assumptions still hold and to see how accurate our prediction algorithm is.

3.6 Software to Use

3.6.1 Microsoft Access

Because Reliant uses Microsoft Access to manage some of its own data, and our group has experience working with Access, this was our first choice. Unfortunately, during the early stages of our analysis, Access was not functioning well. For example, when entering the command below, Access slowed down considerably before eventually crashing.

SELECT field x, field y

FROM table a, table b

WHERE a.AID = b.BID

With advice from the Research Department at WPI, we decided to use software other than Microsoft Access.

3.6.2 MySQL

MySQL is an open source database software. It has been a preferred choice for many corporations due to its speed, reliability, and ease of use. MySQL has also been used on websites such as Google, YouTube and Wikipedia (About MySQL, 2014). Although there are many built-in functions for importing and exporting files in Linux and MySQL, we encountered file permission errors when we attempted to use these built-in functions. With assistance from Robert Brown, senior HPC Systems Integrator at WPI Academic & Research Computing, we were able to use Perl, a powerful tool in Linux, to import data from Reliant into MySQL, and export query results from MySQL for further analysis (see Appendix IV and Appendix V for sample Perl scripts).

3.6.3 Linux

Linux is an operating system (OS), much like Windows and OS X are. Unlike Windows and OS X however, Linux is an open OS. It does not have a sponsoring company, meaning Linux is open to everyone to develop or support. Linux can be used on personal computing devices as well as small devices with high mobility (Overview of the Linux Operating System, 2009).

We used a Virtual Linux Server (VLS). Using a physical server would remove any ability to work with the data remotely. Our group did not have any working experience with Linux. When we learned how valuable it could be to us and our research, we reached out to Robert Brown for assistance in using Linux.

3.6.4 Microsoft Excel

Much like Microsoft Access, we were comfortable using Excel based upon past experience using the application. Excel was a useful tool, especially with the amount of data visualization our project required. Excel has many powerful features for analyzing data, such as (Excel Easy Data Analysis, 2014):

- Conditional Formatting
- Charts
- Pivot Table

3.6.5 SPSS

SPSS stands for Statistical Package for the Social Sciences. It is reflected through its name that it is designed to be easy and convenient for social scientist to use. Nowadays, due to its convenience, it has become popular software among other fields as well, including the health sciences (Quintero, 2015).

3.6.3.1 Cluster Analysis

Modern statistical software, such as R, SPSS, and SAS, all offer a Two-Step cluster analysis option. From our research, we understand that SPSS offers the most comprehensive Two-Step cluster analysis suite. Our team believes that the SPSS suite offers the most customizable design and a very detailed output, which will help provide us with strong visual data for our results outline.

Our team used the SPSS software suite to perform the cluster analysis. Our project adviser, Professor Diane Strong, recommended the use of SPSS for most statistical analyses, and the supporting documentation for SPSS allowed us to quickly learn how to perform a detailed and customized cluster analysis for the data set. After creating the automated cluster analysis, our team also performed a number of limited cluster analyses to see if the software suite would match our team's predicted outcome. We found a number of differences in our predictions and the final output, as we explain in the results section.

IBM's SPSS suite offers a powerful two-step cluster analysis tool, where the user must weigh a number of categorical variables against a specified number of continuous variables. The categorical variables, which SPSS uses to create the characteristics of each cluster, are essentially the independent variables of the study. The continuous variables, which SPSS analyzes for each respective cluster, are the equivalent of a dependent variable for the purpose of this study.

3.6.3.2 Receiver Operating Characteristic (ROC) Curve

In the case of the MyChart patient data, we attempted to find a cutoff point for the data in order to determine which patients in our potential problem patient group would statistically qualify for being actually problematic. As mentioned in the background discussion of ROC curves and the analysis of said curves, we searched for a cutoff point in the upper left region of the resulting curves.

Our team has a great deal of experience using IBM's SPSS statistical software suite for a number of applications. SPSS offers a strong utility for ROC analysis, and our team quickly interpreted how to customize the ROC analysis procedure to fit the "actual problem patient" dataset.

Receiver Operating Characteristic Curves are statistical in nature, and therefore the analysis cannot rely on a single curve. In many instances, a study may require a vast sequence of curves to test for a cutoff point incrementally. In order to find the best cutoff point of these incremented curves, one must then make a visual comparison in conjunction with any other study-specific assumptions.

With regard to the aforementioned area under the curve, we can then understand that choosing an accurate point amongst several curves would require finding the point with the greatest area underneath the curve. This is assuming that the curves are similarly shaped.

These curves also operate under a very specific set of pre-defined rules and standards. In particular, the curve must analyze the sensitivity and 1-specificity of either the presence or absence of a specific characteristic. This is similar to a hypothesis test, where one must attempt to determine if a given statement is false or not-false.

4.0 Results

This section goes into detail about our findings throughout the duration of the study. The results from our organizing, categorizing, and analyzing of the data are included, as are images depicting specific coding snippets we used.

4.1 Organize and Categorize Data

4.1.1 Combine IB_Message Table and MyChart_Message Table

As we previously mentioned, the IB_Message table and the MyChart_Message table include unique properties of a message. Thus, we needed to compile all valid data into one table for the purpose of further analysis (see Appendix VI for detailed code). The combined table has the following fields (see figure 9):

- STUDYID is a unique ID for each patient.
- MSG_ID is the IB_Message ID. This field is used in the MSG_STAT_AUDIT table which will help identify the read time of each message.
- MSG_TYPE_C identifies the message type in the Epic system.
- MYC_MSG_ID is the MyChart_Message ID. This ID is related to the next field of this table, TOFROM_PAT_C.
- TOFROM_PAT_C is a unique property in the MyChart_Message Table, which reveals whether a message is sent to the patient by the provider, or sent to the provider by the patient.
- MYC_PARENT_MESSAGE_ID is also a unique property in the MyChart_Message Table. When the recipient of a message replies to it with a message of their own, it will be recorded as a child message of the original one. This will help identify the effect of parent messages on the time it takes the recipient to read the message.
- MSG_TIME is the sent time of a message. We wanted to keep this information accessible in case it is needed at any time during the project.

Field	Type
STUDYID	int(7) unsigned zerofill
MSG_ID	varchar(18)
MSG_TYPE_C	int(11)
MYC_MSG_ID	varchar(18)
MYC_MSG_TYP_C	int(11)
TOFROM_PAT_C	int(11)
MYC_PARENT_MESSAGE_ID	varchar(18)
MSG_TIME	datetime

Figure 9 – Combined Table: IB_Message & MyChart_Message

Within the IB_Message table, there are a number of messages being sent to and from locations within the provider's office. These messages are not going to nor coming from a patient, even if they have related MyChart_Message IDs. Therefore, we discarded the irrelevant data after informing Reliant and receiving clearance to do so. At this point, there were 211,360 messages in this Combined Messages table for the group to analyze.

4.1.2 Determine the Time It Takes Recipients to Read the Message for Each Record

The MSG_STAT_AUDIT table recorded information about message statuses, including the time of each status change. The ID used in this table to identify the messages is the IB_Message ID. Thus, we created a table that links the messages to their sent time and read time (see Appendix VII for detailed code and figure 10 for the fields).

Field	Type
STUDYID	varchar(18)
MSG_ID	varchar(18)
MSG_SENT_TIME	datetime
MSG_DONE_TIME	datetime
TOFROM_PAT_C	int(11)
MYCHART_MESSAGE_ID	varchar(18)
MYC_PARENT_MESSAGE_ID	varchar(18)
MYC_MSG_TYP_C	int(11)
MSG_TYPE_C	int(11)

Figure 10 – MSG_SENT_READ Table

Using this table, our group created a calculated field named "DONETIME" to calculate the time difference between "MSG_SENT_TIME" and "MSG_DONE_TIME" for each message (see figure 11).

Field	Type
STUDYID	int(7) unsigned zerofill
MSG_ID	varchar(18)
MYCHART_MESSAGE_ID	varchar(18)
TOFROM_PAT_C	int(11)
DONETIME	int(11)
MYC_PARENT_MESSAGE_ID	varchar(18)

Figure 11 – Calculated Field: 'DONETIME'

4.1.3 Categorize the Data Based on the DONETIME

Based on the DONETIME calculated, we categorized the data according to the intervals mentioned in the methodology. Our group created two tables, one for messages sent to the patient (see figure 12) and one for messages sent to the provider (see figure 13). Each table has the count of messages read within the time range indicated.

TIME	COUNT
<24 hrs	73021
24-48 hrs	5226
48-72 hrs	2677
72-96 hrs	1784
96-120 hrs	1147
120-144 hrs	864
144-168 hrs	736
168-192 hrs	642
192-216 hrs	440
216-240 hrs	373
>240 hrs	8373
Never Done	7091

Figure 13 – Msg to the Patient

TIME	COUNT
<24 hrs	90251
24-48 hrs	8232
48-72 hrs	5394
72-96 hrs	1728
96-120 hrs	945
120-144 hrs	611
144-168 hrs	393
168-192 hrs	259
192-216 hrs	149
216-240 hrs	128
>240 hrs	814
Never Done	82

Figure 12 – Msg to the Provider

We ran a quick first round analysis and determined that most of the messages are opened within reasonable time. Especially for the messages sent to providers, more than 90% were opened within the guaranteed time. Therefore, the group decided to proceed with the sponsor's suggestion and look into additional details regarding messages sent to patients. In this case, the group concerned most about problem messages that fall under the ">240 hrs" and "Never Done" category.

4.1.4 Categorize the Potential Problem Patients

4.1.4.1 Overdue Messages and Never-Read Messages

To appropriately categorize the potential problem patients, the group created two sets of tables. First, we created two tables for overdue and never-read messages. The overdue message table (N=8,373) includes messages with DONETIME > 240 (see the sample in figure 14). The never-read message table (N=7,091) includes messages that have "NULL" value for the DONETIME field (see the sample in figure 15).

STUDYID	MSG_ID	MYCHART_MESSAGE_ID	TOFROM_PAT_C	DONETIME	MYC_PARENT_MESSAGE_ID	MYC_MSG_TYP_C	MSG_TIME
[redacted]	[redacted]	[redacted]	1	838	[redacted]	1	2013-01-01 10:29:00
[redacted]	[redacted]	[redacted]	1	307	[redacted]	1	2013-01-01 14:14:00
[redacted]	[redacted]	[redacted]	1	838	[redacted]	1	2013-01-01 17:37:00
[redacted]	[redacted]	[redacted]	1	474	[redacted]	1	2013-01-01 17:53:00
[redacted]	[redacted]	[redacted]	1	314	[redacted]	1	2013-01-02 07:51:00
[redacted]	[redacted]	[redacted]	1	838	[redacted]	1	2013-01-02 08:03:00
[redacted]	[redacted]	[redacted]	1	359	[redacted]	16	2013-01-02 08:42:00
[redacted]	[redacted]	[redacted]	1	838	[redacted]	12	2013-01-02 08:56:00
[redacted]	[redacted]	[redacted]	1	701	[redacted]	16	2013-01-02 09:00:00
[redacted]	[redacted]	[redacted]	1	319	[redacted]	25	2013-01-02 09:27:00

Figure 14 – Overdue Message Table

STUDYID	MSG_ID	MYCHART_MESSAGE_ID	TOFROM_PAT_C	DONETIME	MYC_PARENT_MESSAGE_ID	MYC_MSG_TYP_C	MSG_TIME
[redacted]	[redacted]	[redacted]	1	NULL	[redacted]	1	2013-01-01 17:37:00
[redacted]	[redacted]	[redacted]	1	NULL	[redacted]	1	2013-01-01 17:39:00
[redacted]	[redacted]	[redacted]	1	NULL	[redacted]	1	2013-01-02 06:09:00
[redacted]	[redacted]	[redacted]	1	NULL	[redacted]	16	2013-01-02 08:24:00
[redacted]	[redacted]	[redacted]	1	NULL	[redacted]	16	2013-01-02 08:29:00
[redacted]	[redacted]	[redacted]	1	NULL	[redacted]	1	2013-01-02 09:22:00
[redacted]	[redacted]	[redacted]	1	NULL	[redacted]	16	2013-01-02 11:03:00
[redacted]	[redacted]	[redacted]	1	NULL	[redacted]	16	2013-01-02 11:36:00
[redacted]	[redacted]	[redacted]	1	NULL	[redacted]	1	2013-01-02 11:57:00
[redacted]	[redacted]	[redacted]	1	NULL	[redacted]	16	2013-01-02 11:59:00

Figure 15 – Never-Read Message Table

4.1.4.2 Patients with Overdue Messages and Patients with Never-Read Messages

Based on the data from the two tables mentioned above, we created a table for patients who have overdue messages (N=5,666) and a table for patients who have never-read messages (N=4,389) (see Appendix VIII and Appendix IX for detailed code). Due to the reason that some patients have not logged in at all during 2013, when joining the tables, the value in their login count field is null. In our analysis, there were 822 entries that had null value for this field. Since login count is used in later calculation, the group manually assigned a login count value of '0' for those entries. Another field the team had to manually update was the "RAFSCORE" field which indicates if a patient is considered healthy or not. In the 2013 data we got from our sponsor, 733 entries do not have the RAFSCORE data. For the consistency of our analysis, we manually assigned a RAFSCORE value of '1' for those entries. The reason we chose '1' was because patients with a RAFSCORE greater than 1 are considered unhealthy. We felt unfair to assign a random number for any patient and therefore consulted our project sponsor to determine this value.

4.1.4.3 Added Login Count Metric

As mentioned in Section 3.1, Reliant also provided us with a detailed user activity log. Based on our project sponsor's feedback, we added a metric, count of logins per patient in 2013, to both tables. One of the issues that we encountered while trying to get accurate login count was proxies (refer to section 3.1 for details about proxy). The overall patient base with proxies was small, only 6,697 patients out of 56,243 MyChart users. Of the patients we are targeting in our analysis for this project, only about 400 patients have proxies. The problem with proxy users is that when a proxy logs in and checks messages, the user activity log does not have the log in record. Therefore, for the accuracy of our analysis, we decided to discard patient records with proxies as directed by our sponsor.

4.1.4.4 Re-organized Data with Targeted Message Types

After consulting with our project sponsor, we re-organized the data based on MyChart Message Types. The overdue patient table and never-read patient table have been adjusted and repopulated with data related only to messages with MyChart Message type 1, which are user messages (usually test results), and type 11, which are medical advice requests. The count of total messages received has also been altered to account for the all type 1 and 11 messages received in 2013. The reason why we decided to focus on the two types of messages is that 75% of the problem messages are of either type 1 or type 11. Other message types include medical prescription renewal request, appointment confirmation, e-visit summary, and others. Most of the categories we decided to discard were system-generated confirmations that the health provider is less concerned about if a patient actually read them whereas type 1 and 11 are test results and medical advice that could include important information. The revised overdue patient table has 3,180 records and the revised never-read patient table has 2017 records.

As you can see below in Figure 16 and Figure 17, the overdue patient table includes patient demographic information, rafscore, and login count. It also includes the number of overdue messages, the number of targeted messages (type 1 and type 11), and the number of total message (messages of all type) a patient has during the year of 2013.

Field	Type
STUDYID	varchar(18)
AGE	int(11)
SEX	varchar(2)
PAT_STATUS	varchar(18)
FIRST_MYCHART_LOGINDT	datetime
LAST_MYCHART_LOGINDT	datetime
OverdueCount	int(11)
TARGETMSGCOUNT	int(11)
LOGINCOUNT	int(11)
PROXYYN	char(2)
RAFSCORE	decimal(10,3)
TOTALMSGCOUNT	int(11)

Figure 16 – Patients with Overdue Messages (Fields)

STUDYID	AGE	SEX	PAT_STATUS	FIRST_MYCHART_LOGINDT	LAST_MYCHART_LOGINDT	OverdueCount	TARGETMSGCOUNT	LOGINCOUNT	PROXYYN	RAFSCORE	TOTALMSGCOUNT
10[redacted]	45	F	Alive	2010-12-30 10:02:23	2014-09-29 07:39:48	1	6	17	N	0.593	8
10[redacted]	30	F	Alive	2012-01-04 09:12:00	2014-09-30 15:22:19	1	13	54	N	0.527	13
10[redacted]	66	F	Alive	2012-08-30 16:25:32	2014-09-12 18:52:55	1	4	4	N	1.463	4
10[redacted]	38	M	Alive	2013-01-06 23:19:30	2014-09-17 15:33:22	1	1	8	N	0.124	1
10[redacted]	56	F	Alive	2012-10-04 19:25:36	2014-09-24 19:31:16	1	1	75	N	0.656	1
10[redacted]	80	F	Alive	2010-07-12 11:03:22	2013-05-12 11:58:58	1	1	7	N	1.464	1
10[redacted]	56	F	Alive	2012-01-07 07:38:43	2014-08-04 19:10:07	1	3	3	N	0.656	3
10[redacted]	38	F	Alive	2012-09-25 06:46:30	2014-10-08 19:47:51	1	18	26	N	0.809	18
10[redacted]	37	F	Alive	2009-06-25 12:54:57	2014-09-27 12:23:00	1	21	55	N	0.205	26
10[redacted]	68	F	Alive	2010-08-16 09:43:19	2014-09-28 19:38:08	1	7	41	N	1.925	30

Figure 17 – Patients with Overdue Messages (Sample Data)

Similarly, the never-read patient table includes patient demographic information, RAFscore, and login count. It also includes the number of never-read messages, the number of targeted messages (type 1 and type 11), and the number of total message (messages of all type) a patient has during the year of 2013 (see figure 18 and figure 19).

Field	Type
STUDYID	varchar(18)
AGE	int(11)
SEX	varchar(2)
PAT_STATUS	varchar(18)
FIRST_MYCHART_LOGINDT	datetime
LAST_MYCHART_LOGINDT	datetime
NeverdoneCount	int(11)
TARGETMSGCOUNT	int(11)
LOGINCOUNT	int(11)
PROXYYN	char(2)
RAFSCORE	decimal(10,3)
TOTALMSGCOUNT	int(11)

Figure 18 – Patients with Never-Read Messages (Fields)

STUDYID	AGE	SEX	PAT_STATUS	FIRST_MYCHART_LOGINDT	LAST_MYCHART_LOGINDT	NeverdoneCount	TARGETMSGCOUNT	LOGINCOUNT	PROXYYN	RAFSCORE	TOTALMSGCOUNT
10[redacted]	67	F	Alive	2010-06-02 08:47:26	2014-09-12 17:13:53	1	3	13	N	0.662	7
10[redacted]	51	F	Alive	2013-06-07 21:41:17	2014-09-26 21:37:25	1	1	6	N	0.263	7
10[redacted]	55	F	Alive	2009-02-26 08:42:27	2014-08-17 16:32:43	1	2	7	N	1.076	2
10[redacted]	27	F	Alive	2013-08-14 15:10:14	2014-09-10 12:15:00	1	1	2	N	1.000	1
10[redacted]	35	F	Alive	2013-06-07 18:12:41	2014-09-29 15:27:38	1	1	10	N	1.000	1
10[redacted]	40	F	Alive	2013-01-28 12:18:59	2014-05-31 21:54:17	1	3	19	N	0.205	3
10[redacted]	88	M	Alive	2010-05-04 08:15:30	2014-06-20 08:46:30	1	1	7	N	1.118	1
10[redacted]	76	M	Alive	2010-07-31 13:43:20	2013-04-22 08:07:35	1	2	5	N	2.024	4
10[redacted]	30	F	Alive	2010-03-12 17:27:50	2014-08-24 23:28:42	2	34	31	N	1.114	34
10[redacted]	59	F	Alive	2012-10-30 11:58:22	2014-06-12 17:42:53	1	10	22	N	1.328	10

Figure 19 – Patients with Never-Read Messages (Sample Data)

4.1.4.5 Potential Problem Patients

We initially created the two tables, overdue patients and never-read patients, separately for the purpose of exploring if there are different patterns between the two patient groups. However, after a thorough statistical analysis, the group found that most users who misuse the system follow similar pattern, no matter if they have messages never-read or just overdue. Therefore, we combined the two tables into one in the later stage of the project (see Section 4.2.1 for more details).

4.2 Analyze and Visualize Data

With the dataset provided by Reliant now organized for analysis, our team quickly utilized the resources available to us at WPI to create visuals that displayed the data in a detailed manner. The results of statistical analysis and visual charts are exceptionally useful, as they make our most critical discoveries easy for Reliant and other interested parties to understand. This is in no small part due to the time and effort spent cleaning and organizing the dataset. In addition, the time spent selecting the proper variables to scrutinize gave our team a significant advantage when creating the visualizations for Reliant Medical Group.

As mentioned in the methodology, the team utilized both Microsoft Excel and SPSS for analyzing and visualizing the data.

4.2.1 Re-categorizing the data based on visual results

Reliant Medical Group originally provided the team with a very large and very complex set of patient data from the MyChart system. This data included a wide array of identifying characteristics, some of which appeared to serve as a possible starting point for our process of identifying trends within the dataset. At the inception of our team's command of the MyChart data, we attempted to differentiate various groups of patients based on these characteristics, but quickly adopted new strategies for modifying the dataset.

Initially, the team sought to group patients based on a message by message basis, using the occurrence of never-read and overdue messages as a deciding factor. Per the suggestion of the Reliant staff, our team sought to create tables of patient data that consisted of patients with never-read messages, as well as tables of data that consisted of patients with overdue messages. We quickly discovered that grouping patients on this basis was not only disorganized and difficult for others to understand, but also did not raise any statistical evidence of different behavior patterns within the predefined groups. Therefore, for the sake of replicability and simplicity, our team created a dataset that consisted exclusively of "potential problem patients", which include patients with overdue or never-read messages of type 1 or type 11 and a dataset that consisted exclusively of "good patients", which are patients who always check their messages within 10 days.

4.2.2 Visual Analysis

As we stated in the introduction and background, Reliant Medical Group is concerned with potential risks related to misuse of the MyChart messaging system. In accordance with the binning and categorization procedures previously detailed, our team chose to analyze the pertinent data on a message type and patient-by-patient basis. We started with simple visuals representing the time it takes a patient to open the message, the type of messages that are neglected, and patient characteristics.

4.2.2.1 Message Open Time

The first chart we created was to help us understand how many messages fall into each category, including the messages opened within reasonable time period (within 10 days of receiving the message), messages that are overdue (opened at some point but after 10 days of receiving the message), and messages that have never been read (see figure 20). As shown in the chart, the majority of messages were opened within 240 hours, or 10 days. Meanwhile, there is still a portion of messages that were not opened within this desired timeframe, or even never opened. After consulting our project sponsor, the team decided to look into more details and determine if there were any safety related issues we could find.

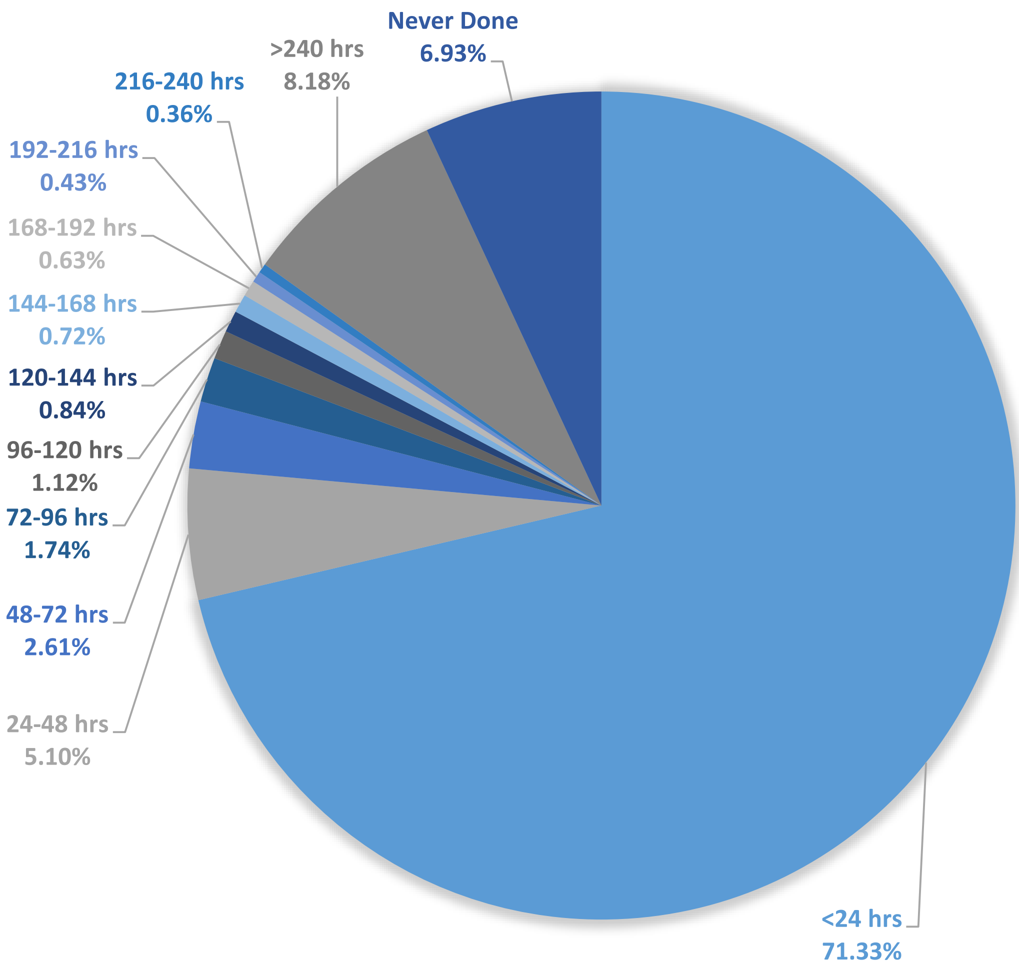

Figure 20 – Message Open Time

4.2.2.2 Message Types

The team created another set of pie charts to try identifying if the messages being ignored actually raise a safety concern (see figure 21 and figure 22 below).

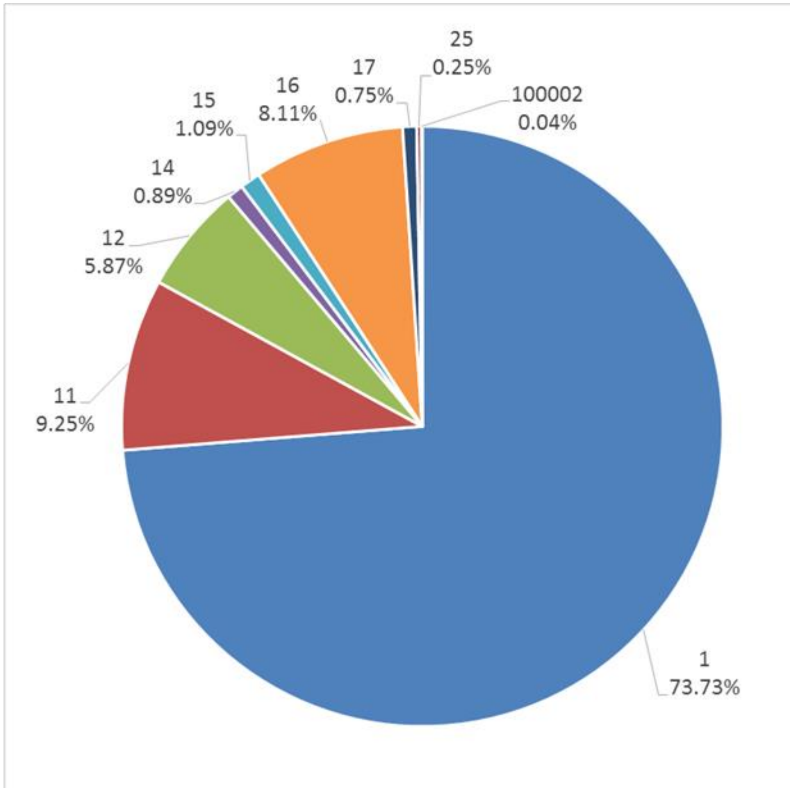

Figure 21 – Never-Read Messages by Message Type

From figure 21 above it is clear that more than 80% of all Never-Read messages are either Type 1 (User Messages, mostly test results) or Type 11 (Patient Medical Advice Request). Some other message types include Type 12 (Patient Appointment Schedule Request), Type 16 (Patient Medication Renewal Request), and other system related requests such as Patient Demographics Update Request and so on.

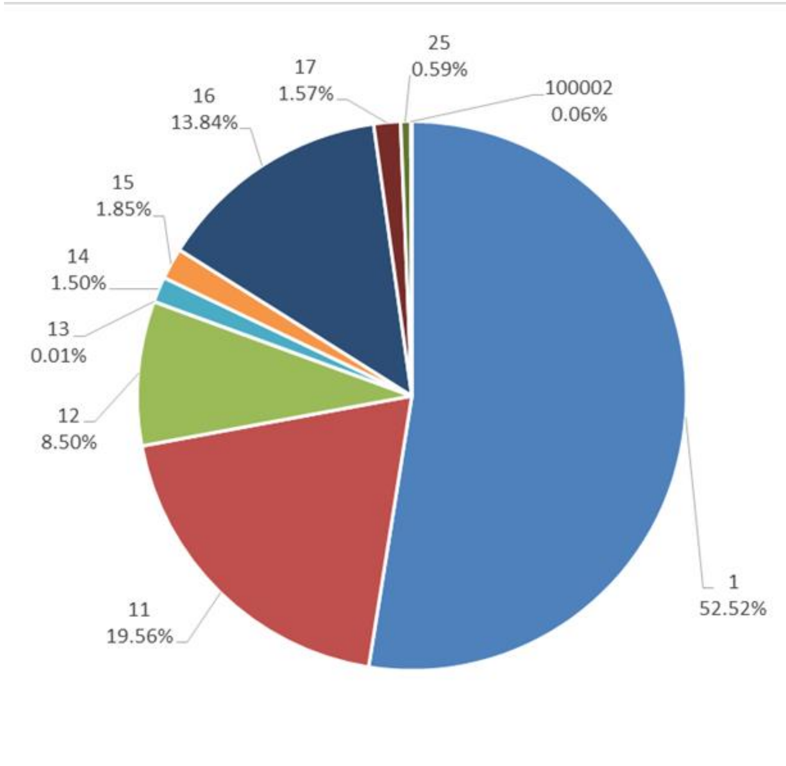

Figure 22 – Overdue Messages by Message Type

Similar to the never-read messages, figure 22 shows that most overdue messages are either of Type 1 or Type 11. Type 12 and Type 16 also makes up a substantial portion.

It is important to note that Reliant Medical Group understands that many messages of such types are neglected intentionally, as the patients are often aware of their account activities. As a result of discussing this chart with our project sponsor, we decided to focus on the Type 1 and Type 11 messages.

4.2.2.3 Patient Types

At this point, as we already know what message types we are pursuing, we shifted our focus to patient characteristics. Our team examined the effects of age, gender, and health condition upon the likelihood of a patient having one or more overdue or never-read messages. The examination of never-read patients and overdue patients is examined independently of the presence of each other at this point, as we were still trying to figure out if patient behavior toward the two types will be different.

- Age

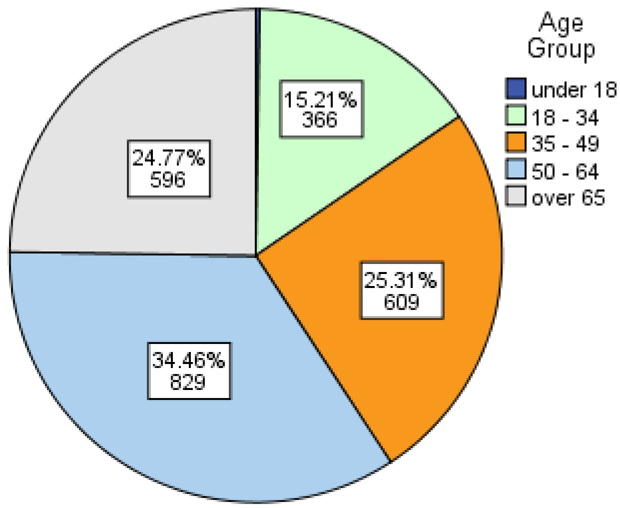

Figure 23 – Never-Read Patients by Age

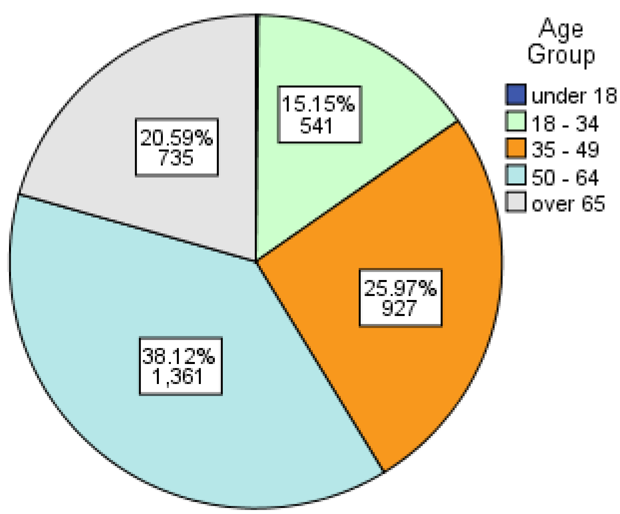

Figure 24 – Overdue Patients by Age

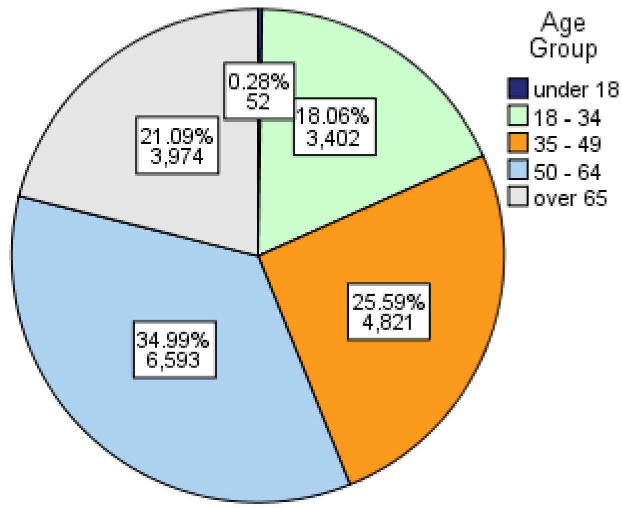

Figure 25 – Total Patients by Age

From the figure 23, figure 24, and figure 25 above, there is limited evidence among the given data to suggest that age is directly related to patients' user behavior. From these charts we noticed that there is little difference. In particular, the largest difference we could see from comparing these charts appeared in the age group of "over 65". However, that difference was only about 4%. The overall patient distribution of age is identical.

- Gender

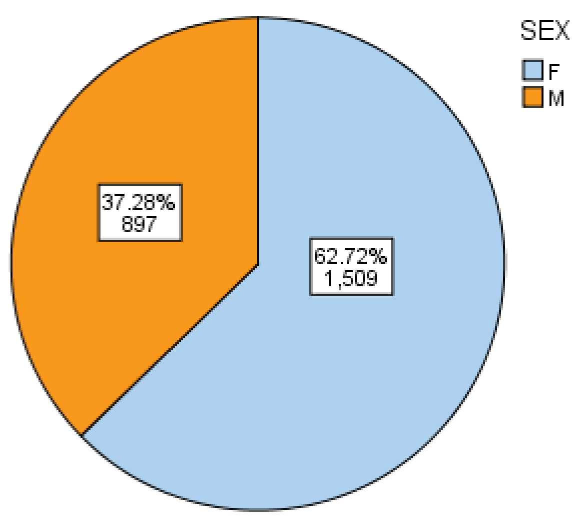

Figure 26 – Never-Read Patients by Gender

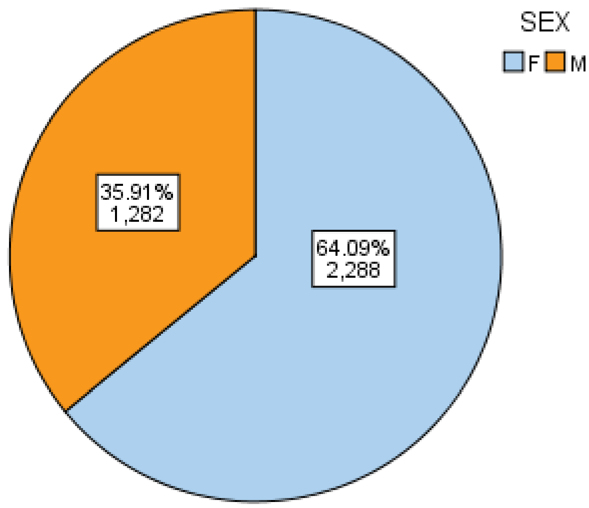

Figure 27 – Overdue Patients by Gender

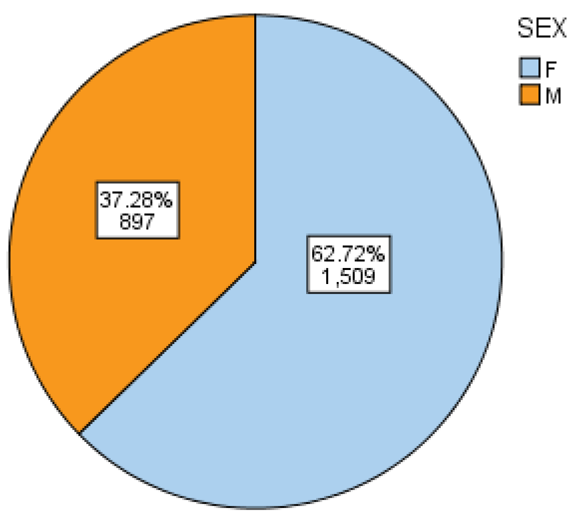

Figure 28 – Total Patients by Gender

Similarly, in regards to gender, from the figure 26, figure 27, and figure 28, there is very little evidence among the given data to suggest that gender is directly related to patients' user behavior. Although there are more female patients in both the overdue and never-read patient charts, there are more female patients overall, so the difference is proportionate. For the total patient sample, there is 62.72% female; for the overdue sample, there is 64.09% female; for the never-read sample, there is 62.72% female.

- Health Condition

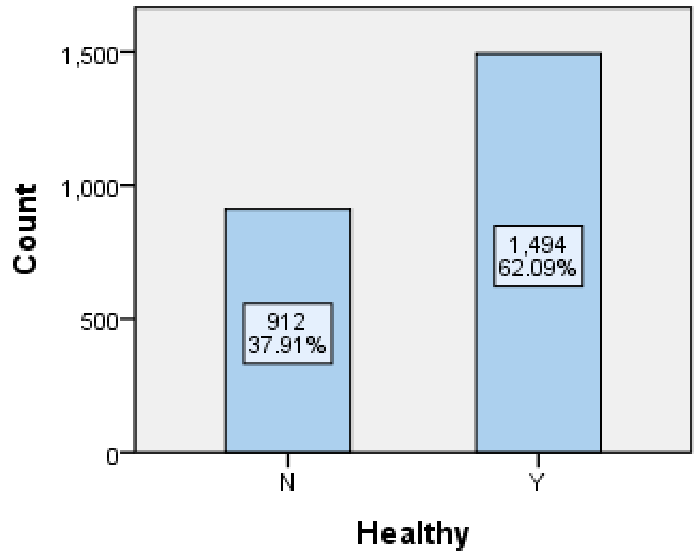

Figure 29 – Never-Read Patients by Health Condition

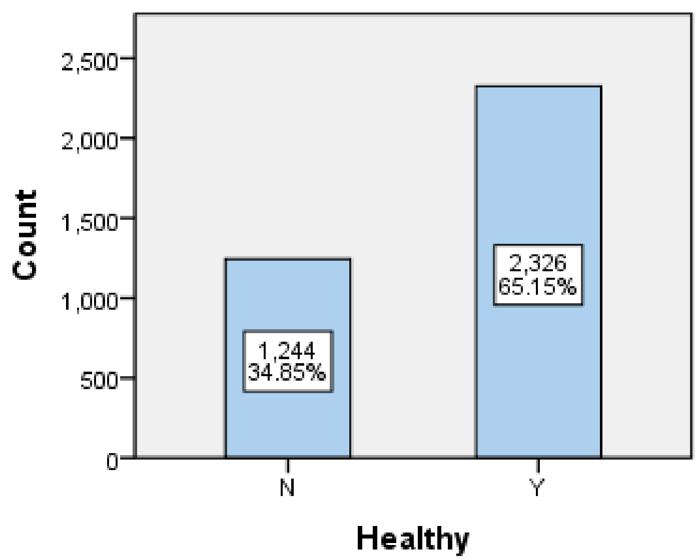

Figure 30 – Overdue Patients by Health Condition

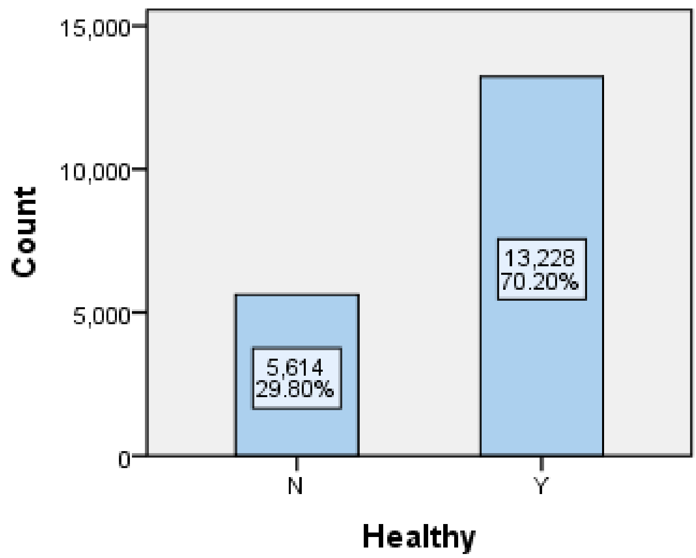

Figure 31 – Total Patients by Health Condition

It is critical to note that RAFscore, a numerical measure of a patient's overall healthiness, was the final categorical variable that our team analyzed before moving on to other statistical methods. From figure 29, 30 and 31 above, we demonstrated that health condition, or RAFscore, is not a reliable indicator of a patient's MyChart usage. On all three charts, the column on the left stands for unhealthy patients (patients with RAFSCORE > 1), and the column on the right stands for healthy patients (patients with RAFSCORE <= 1). The proportions of problem patients with high RAFscore ratings, as well as those with low RAFscore ratings, are not significantly different when compared to the total population of MyChart users based on the same criteria. Although this appears complicated, the visual data holds the true conclusions, as there is no visual difference among the three visual outputs supplied.

According to the charts provided, our group found a basis to suggest that there is no obvious difference between the patient with overdue messages and the patient with never-read messages. Therefore, after addressing this observation with Reliant Medical Group, the team decided to carry on the analysis with only two categories for the patients, "good" patients who have never missed any messages, and "potential problem" patients who have at least one overdue or never-read message.

4.2.3 SPSS Analysis

With the help of the Excel visuals, the team was able to get a clearer focus on the direction. We identified that none of the single factors we considered was likely to help predict patient behavior. We decided to move to use historical MyChart usage to predict future user behavior. As recommended by Prof. Strong, our project advisor, the team did a more thorough analysis with SPSS to back up our points. This involves cluster analysis on all patients, categorized into good or bad patients, and ROC Curve analysis, for determining high risk bad patients.

4.2.3.1 Cluster Analysis

In the methodology section, we noted that our team performed sets of both automated and customized cluster analyses via the SPSS software suite. The SPSS documentation suggests that the automated clustering method is preferable, as it creates the number of clusters suggested by the Euclidean moving average method. Our team also tested the Two-Step cluster analysis method with a pre-defined number of clusters.

We chose to cluster the data from the “problem” MyChart patients based upon the categorical variables of age group, gender, and health status. In each cluster we chose “Problem Message%” as the continuous variable, so as to accurately gauge the percentage of all messages in the cluster that were overdue or never-read (see Figure 32 below). If there are some clusters that have outstanding sample size, then we can identify that the patients who have the characteristics of that cluster will tend to be “potential problem patients”. Otherwise, we can rule out those variables we observed as unlikely to be the key factors that will influence patients’ behavior.

For the “good” MyChart patients, our team took a slightly modified approach, using age group, gender, and health status as categorical variables, but removing problem message percentage as the continuous variable, as “good” patients do not have any problem messages in their records.

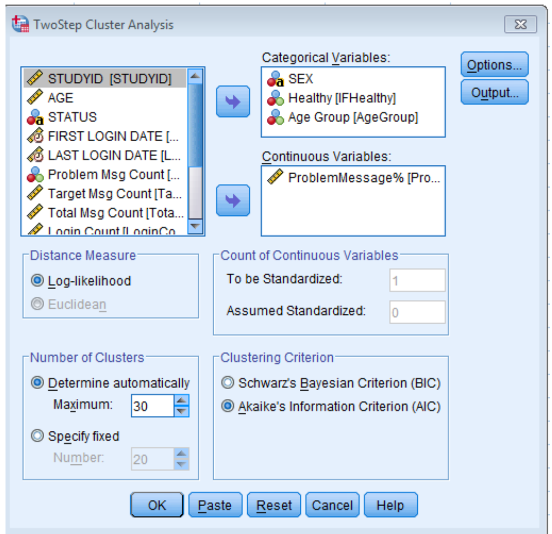

Figure 32 – Cluster Analysis Variables

Pictured below is the output of the SPSS software suite's cluster analysis. SPSS created 12 clusters based on an algorithm using a Euclidean moving average. The patient dataset used for this analysis consisted of only those patients that had problem messages. SPSS then clustered the patients by age grouping, gender, and health status based on RAF score. The clusters do not appear to represent any distinct groupings, since there are no such clusters that have outstanding sample size. This is important as it shows that the influence of those three variables, age group, RAFscore, and gender, did not cause a drastic shift in the distribution of the population. Relatively even distribution of clusters suggests that the chosen variables do not require any further investigation, as they did not make a large impact. Therefore, this leads us to pursue the idea that we should investigate factors other than age, gender, and health status.

Clusters

Input (Predictor) Importance
1.0 0.8 0.6 0.4 0.2 0.0

Cluster	12	11	10	9	7	2	4	6	1	3	8	5
Label												
Description												
Size	13.2% (737)	13.0% (722)	10.8% (603)	9.8% (546)	8.6% (479)	7.7% (427)	7.2% (398)	7.1% (397)	6.7% (374)	6.3% (349)	5.1% (284)	4.5% (250)
Inputs	Age Group 35-49 (100.0%)	Age Group 50-64 (100.0%)	Age Group 18-34 (100.0%)	Age Group 50-64 (100.0%)	Age Group 50-64 (100.0%)	Age Group 35-49 (72.6%)	Age Group 35-49 (100.0%)	Age Group 65+ (100.0%)	Age Group 65+ (100.0%)	Age Group 65+ (58.7%)	Age Group 50-64 (100.0%)	Age Group 65+ (100.0%)
	Healthy Healthy (100.0%)	Healthy Healthy (100.0%)	Healthy Healthy (100.0%)	Healthy Healthy (100.0%)	Healthy Unhealthy (100.0%)	Healthy Unhealthy (100.0%)	Healthy Healthy (100.0%)	Healthy Unhealthy (100.0%)	Healthy Unhealthy (100.0%)	Healthy Healthy (100.0%)	Healthy Unhealthy (100.0%)	Healthy Healthy (100.0%)
	SEX F (100.0%)	SEX F (100.0%)	SEX F (100.0%)	SEX M (100.0%)	SEX F (100.0%)	SEX F (79.2%)	SEX M (100.0%)	SEX F (100.0%)	SEX M (100.0%)	SEX M (100.0%)	SEX M (100.0%)	SEX F (100.0%)
	ProblemMessage% 52.01%	ProblemMessage% 54.59%	ProblemMessage% 56.35%	ProblemMessage% 58.56%	ProblemMessage% 43.73%	ProblemMessage% 46.33%	ProblemMessage% 60.77%	ProblemMessage% 53.65%	ProblemMessage% 54.83%	ProblemMessage% 58.63%	ProblemMessage% 53.95%	ProblemMessage% 55.41%

Figure 33 – Cluster Analysis 2013 Problem Patients

To provide a more accurate analysis, we also clustered the good patients, who do not have any unread or overdue messages. As the figure showed below:

Clusters

Input (Predictor) Importance

1.0 0.8 0.6 0.4 0.2 0.0

Cluster	2	1	4	8	11	7	13	9	3	10	6	12	5
Label													
Description													
Size	14.7% (1945)	13.9% (1840)	13.7% (1815)	9.6% (1280)	7.0% (923)	6.9% (916)	5.8% (775)	5.6% (740)	5.1% (676)	5.0% (668)	4.5% (598)	4.1% (547)	4.1% (542)
Inputs	Age Group 35-49 (100.0%)	Age Group 50-64 (100.0%)	Age Group 18-34 (100.0%)	Age Group 50-64 (100.0%)	Age Group 50-64 (100.0%)	Age Group 35-49 (100.0%)	Age Group 65+ (100.0%)	Age Group 65+ (100.0%)	Age Group 65+ (93.9%)	Age Group 50-64 (77.7%)	Age Group 65+ (100.0%)	Age Group 35-49 (72.6%)	Age Group 18-34 (100.0%)
	Healthy Y (100.0%)	Healthy Y (100.0%)	Healthy Y (100.0%)	Healthy Y (100.0%)	Healthy N (100.0%)	Healthy Y (100.0%)	Healthy N (100.0%)	Healthy N (100.0%)	Healthy Y (100.0%)	Healthy N (100.0%)	Healthy Y (100.0%)	Healthy N (100.0%)	Healthy Y (100.0%)
	SEX F (100.0%)	SEX F (100.0%)	SEX F (100.0%)	SEX M (100.0%)	SEX F (100.0%)	SEX M (100.0%)	SEX F (100.0%)	SEX M (100.0%)	SEX F (97.8%)	SEX M (100.0%)	SEX M (100.0%)	SEX F (100.0%)	SEX M (100.0%)

Figure 34 – Cluster Analysis 2013 Good Patients

The clusters analysis results for all good patients shows that age, health condition and gender are cannot be proved to be factors that will influence patients' behavior in our research, since there is no viable difference among cluster sizes. This is not to suggest that all clusters are even in size and distribution, but that they do not vary by any extreme amount throughout the procedure. Although the largest cluster contains 14.7% of the total population while the smallest contains 4.1% of the population, it is not a tremendous difference for a group with thirteen clusters. The cluster sizes of good patients range from 4.1% to 14.7% while the clusters sizes of problem patients range from 4.5% to 13.2%, with similar distribution of the clusters. To be specific, the largest cluster of good patients consist of healthy females between the age of 35 to 49; the largest cluster of problem patients also consist of healthy females between the age of 35 to 49. The figures above lend credibility to our team's choice to no longer pursue age, gender, or health condition for the prediction of potential problem patients.

4.2.3.2 Scatterplot

As previously mentioned, the team next analyzed prediction factors via patient MyChart usage. Based on our previous visualizations and analyses, we decided to use two factors to predict the true high risk patients.

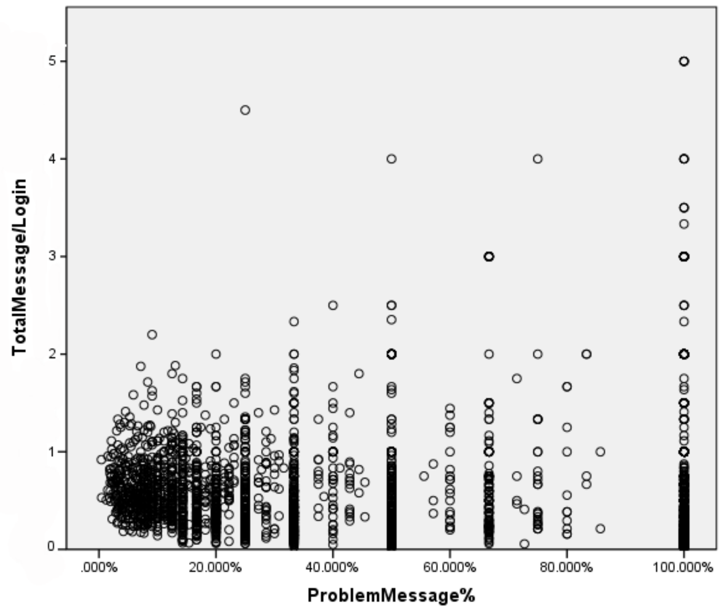

Figure 35 – Scatterplot Message per login / Problem Message %

As shown in figure 35, the team discovered that a combination of message per login and problem message percentage could be a helpful indicator for the true high risk patients. Message per login is calculated through dividing the total message a patient has received by the total login counts of that patient during the same time period, which is the entire year of 2013 in this case. Having a small message per login ratio can mean the patient is a very active MyChart user. When such an active user has overdue or never-read message, it is most likely that this patient neglected the message intentionally. For instance, the patient is aware of a test result coming in and checked out the test result directly without checking the message. Problem message percentage is calculated through dividing the number of total problem messages (overdue or never-read messages of type 1 or type 11) by the total number of target messages (all messages of type 1 and type 11) received by the patient. This can also help indicate if a patient randomly missed a message, or if a patient rarely checks messages.

4.2.3.3 Receiver Operating Characteristic (ROC) Curve

Although figure 35 shows that message per login and problem message percentage give a good indicator, there is no clear cutoff point for either factor. In order to determine the most accurate cutoff points for both parameters, the group conducted ROC Curve analysis based on our sponsor, Dr. Garber's recommendation. The resulting curves represented the incidence of patients in each category (varying by .1 each time), with our focus emanating the messages per login between 1 and 2. As we observed from the ROC curves we made, our team finally chose a cutoff point equal to 1.9 messages/login ratio. From our earlier explanation, this cutoff point suggests that any patient with more than 1.9 messages per login is a potential problem patient. This does not mean that patients that receive more than 1.9 messages in a single login are automatically considered a problem patient, but that those patients who do not log in frequently enough to keep the ratio low have the potential to become a problem patient. We chose this cutoff point, according to three reasons we list below (see figure 36):

1. There is an outstanding point at the left upper corner.
2. Compared to other curves that also have outstanding points, this one has a larger sample size.
3. The area under this point is 0.722, which shows that this ROC curve represents a good test.

Case Processing Summary

Message Login Group	Valid N (listwise)
Positive[a]	201
Negative	5376

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is Greater than 1.9.

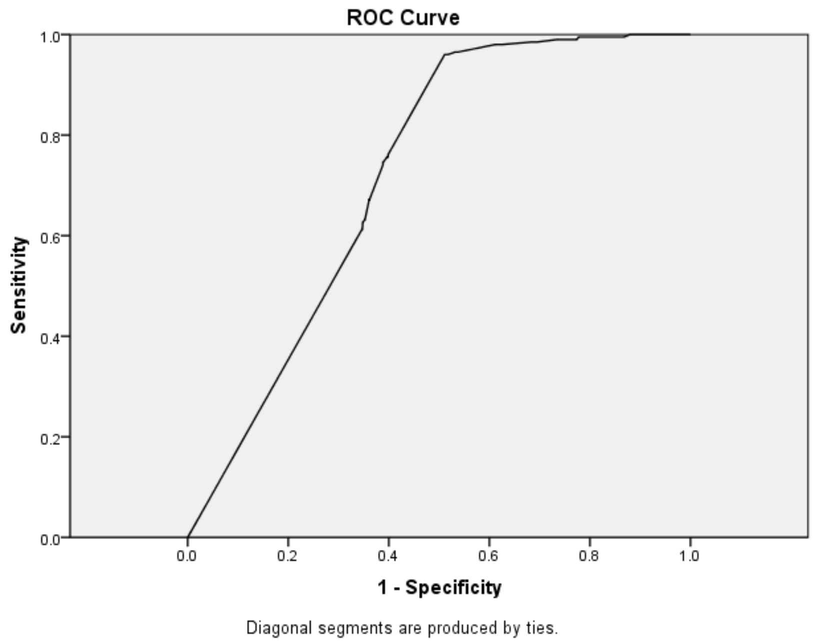

Figure 36 – ROC Curve

Based on the coordinates of the Curve table, we can determine the cutoff point for patient problem messages is 45.8% corresponding to 1.9 message/login ratio. We can see there are two sets of data with identical Sensitivity and 1-Specificity, and we choose the one with a lower problem message percentage—45.8%. In addition, since MySQL only allows two decimal, we used 46% instead of 45.8% as the cutoff point for problem message percentage.

Positive if Greater Than or Equal to a	Sensitivity	1-Specificity
40.58824%	0.960	0.518
41.42157%	0.960	0.518
42.26190%	0.960	0.517
43.65079%	0.960	0.513
44.94949%	0.960	0.512
45.80420%	0.960	0.511
48.01769%	0.960	0.511
52.27273%	0.761	0.399
55.05051%	0.756	0.398

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1.
All the other cutoff value are the averages of two consecutive ordered observed test values

Figure 37 – ROC Curve Sensitivity/Specificity Table

After we determined our cutoff point, we started to only target the patients whose problem message percentage is greater than 46% as well as message-login-ratio is greater than 1.9. This target group contains all the statistically high-risk patients.

These cutoff points are especially useful for Reliant Medical Group to establish an algorithmic method of predicting which of their patients may fall into the high-risk category of MyChart users.

4.3 Prediction Algorithm

Based on the full analysis and consultation with the project sponsor, the group came up with the following logic for the prediction algorithm (See Appendix X for details):

1. Gather data a year back from the date that the procedure is run.
2. Clean up the tables (Discard proxies; Discard type 999 messages).
3. Create a temporary table that combines IB Message table with MyChart Message table to link the messages together.
4. Create a temporary table to determine how long it takes for each message to be read.
5. Create a temporary table to count the total problem messages (overdue or never-read) each patient has within the past year.
6. Create a temporary table to count the total messages each patient has received within the past year.

7. Create a temporary table to count the important messages each patient has received within the past year.
8. Create a temporary table to count the number of logins of each patient.
9. Create a temporary table with all patients who have overdue or never-read messages.
10. Generate a list of patient who are no longer considered active (no login within the year but have unread messages).
11. Generate a list of patients who are active but rarely check their messages (with message per login ratio greater than 1.9 and problem message percentage greater than 46.8%).

Using temporary tables during the procedure is recommended for the purpose of saving space. However, if Reliant Medical Group is interested in keeping the historical data for comparison or other purposes later, they can decide to do that.

4.4 Testing Algorithm

4.4.1 Testing the Algorithm with the 2014 Data

The procedure was run through with the 2014 data. Results are as expected (See Table 2 for a comparison of the 2013 and 2014 data).

Table 3: Comparison of 2013 and 2014 Data

	2013	2014
Number of Total Messages	211,360	270,400
Number of Total Patients who received at least one message	24,050	30,764
Number of Total “Inactive” Patients	822	1,425
Number of Total “High Risk Active” Patients	193	339
Number of Total Patients to be Flagged out	1,015	1,764

4.4.2 Testing with Analysis on 2014 Data

The group also ran the same Cluster Analysis on the 2014 Data to make sure that the findings and assumptions still hold and the prediction algorithm is valid (See figure 38 and figure 39). As shown from the following figures, the clusters show similar distribution of patient characteristics. Although the 2014 bad patient cluster shows that there tends to be slightly more bad patients that are older, that is not enough evidence for age group to be a prediction factor. There are more older patients overall and the number of bad patients comparing with the number of good patients again make it impossible to use age group alone as an indicator of any behavior pattern.

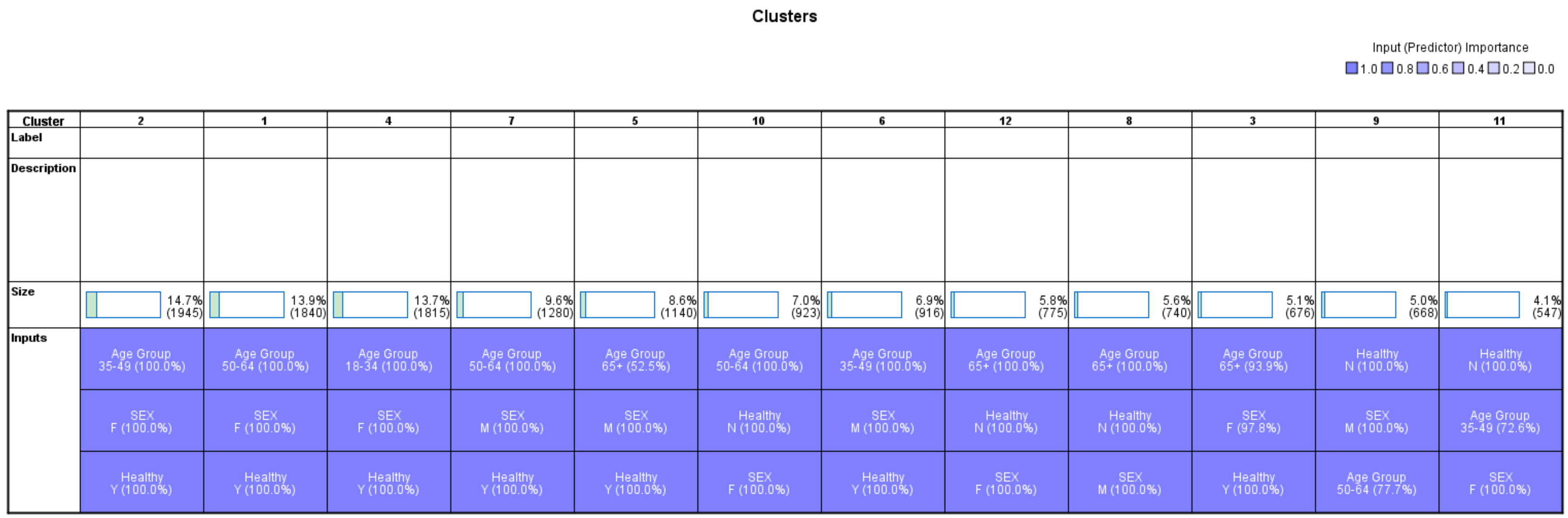

Figure 38 – 2014 Good Patient Clusters

Clusters

Input (Predictor) Importance
1.0 0.8 0.6 0.4 0.2 0.0

Cluster	7	10	9	1	8	4	6	2	5	3
Label										
Description										
Size	14.1% (1386)	12.7% (1247)	12.0% (1177)	11.4% (1117)	10.4% (1026)	9.7% (950)	8.4% (827)	7.5% (739)	7.1% (696)	6.7% (660)
Inputs	Age Group 50-64 (100.0%)	Age Group 35-49 (100.0%)	Age Group 18-34 (99.3%)	Age Group over 65 (55.3%)	Age Group 50-64 (100.0%)	Age Group 35-49 (65.7%)	Age Group 50-64 (100.0%)	Age Group over 65 (100.0%)	Age Group over 65 (99.9%)	Age Group 35-49 (73.6%)
	Healthy Y (100.0%)	Healthy Y (100.0%)	Healthy Y (100.0%)	Healthy N (100.0%)	Healthy Y (100.0%)	Healthy Y (99.9%)	Healthy N (100.0%)	Healthy Y (100.0%)	Healthy N (100.0%)	Healthy N (100.0%)
	SEX F (100.0%)	SEX F (100.0%)	SEX F (100.0%)	SEX M (100.0%)	SEX M (100.0%)	SEX M (100.0%)	SEX F (100.0%)	SEX F (52.9%)	SEX F (100.0%)	SEX F (78.8%)
	Problem Msg % 56.24%	Problem Msg % 54.76%	Problem Msg % 57.30%	Problem Msg % 50.95%	Problem Msg % 59.35%	Problem Msg % 58.65%	Problem Msg % 48.56%	Problem Msg % 53.88%	Problem Msg % 52.89%	Problem Msg % 49.03%

Figure 39 – 2014 Bad Patient Clusters

5.0 Recommendation

Based on our full analysis throughout this project, the team recommends running the prediction procedure once a month with data going back a year from the day the procedure is run. We recommend running the procedure on the backend of the system and then push the results to a production server to flag the patients who are most likely not to respond to the messages within a satisfactory timeframe. This implementation process is most effective if implemented sooner than later, so our team suggests that Reliant do so within 6 to 12 months of the conclusion of the project.

The team also recommends that the provider's office educate the patients more about using their patient portal properly. To be more specific, the office should educate the patients to check their messages regularly and with a timely manner if they receive reminder emails. Doctors should especially remind the patients who are flagged as potential problem patients that they should be using their patient portal more actively. By having the ability to determine which patients are "problem patients", Reliant now has the ability to take any pre-emptive actions that they wish. We suggest that Reliant flag each problem patient in a manner that allows physicians to take any action that they feel necessary. This action could involve targeting the patient for corrective action, or simply trying to increase communication through a more effective medium. These changes are especially important for high-importance messages, and Reliant should take care to educate physicians about the benefits of increased communication.

We also suggest a separate follow analysis to assess the results of our proposed procedure after implementation. In particular, our team strongly suggests that Reliant have a data analyst outside of the current team provide feedback on the efficacy of our team's suggestions after one year and two years of continual usage.

Based on the result of the follow up study, we suggest that Reliant make adjustments accordingly to the procedures that they suggest to physicians. We strongly believe that a pop-up alert system within the provider side of MyChart will help physicians to recall which patients are problematic and provide them with detailed steps on how to help increase patient usage of MyChart and all of the features that it provides.

Bibliography

2013 The Top 100 EHR Companies. (2013). Retrieved from http://medicaleconomics.modernmedicine.com.ezproxy.wpi.edu/medical-economics/content/tags/top100ehrs/top-100-ehr-companies-part-1-4

About MySQL. (n.d.). Retrieved December 18, 2014, from http://www.mysql.com/about/

About UMass Memorial Health Care. (n.d.). Retrieved October 27, 2014, from http://www.umassmemorialhealthcare.org/about-us

About Us. (n.d.). Retrieved October 27, 2014, from http://reliantmedicalgroup.org/about-us/

About Us. (n.d.). Retrieved October 27, 2014, from http://www.epic.com/about-index.php

About Us. (n.d.). Retrieved October 27, 2014, from http://www3.gehealthcare.com/en/about_us

Acute Care Electronic Medical Record. (n.d.). Retrieved October 27, 2014, from http://www.cerner.com/solutions/Hospitals_and_Health_Systems/Acute_Care_EMR/

Allscripts TouchWorks EHR Features. (n.d.) Retrieved October 27, 2014, from http://www.allscripts.com/products-services/products/touchworks-ehr/features

Ambulatory EHR. (n.d.). Retrieved October 27, 2014, from http://www.cerner.com/solutions/Physician_Practices/Ambulatory_EMR_-_EHR/

Centers for Disease Control and Prevention (n.d.). National Ambulatory Medical Care Survey: 2010 Summary Tables.

Center for Medicare&Medicaid Services (2012). An Introduction to the Medicaid EHR Incentive Program for Eligible Professionals.

Center for Medicare&Medicaid Services (2014). An Introduction to the Medicare EHR Incentive Program for Eligible Professionals.

Center for Medicare&Medicaid Services (2014). Patient Electronic Access Tipsheet.

Client Services. (n.d.). Retrieved October 27, 2014, from http://www.epic.com/services-index.php

CMIPA Facts. (n.d.). Retrieved October 27, 2014, from https://www.cmipa.com/facts.php

Consumer Expenditures. (2014): Bureau of Labor Statistics, U.S. Department of Labor.

Data Analysis. (n.d.). Retrieved October 27, 2014, from http://www.excel-easy.com/data-analysis.html

FollowMyHealth. (n.d.). Retrieved October 27, 2014, from http://www.allscripts.com/products-services/products/followmyhealth

Glazier, W. (September 27, 2013). Response to HPC.

Gunter, T., & Terry, N. (2005). The Emergence of National Electronic Health Record Architectures in the United States and Australia: Models, Costs and Questions. J Med Internet Res, 7(1):e3. DOI: 10.2196/jmir.7.1.e3

Hardy, K. (Ed.). (n.d.). (2009). UMass Memorial Health Care Implements EHR and Patient Management Solution. Retrieved from http://www.healthcareitnews.com/news/umass-memorial-health-care-implements-ehr-and-patient-management-solution

Haupt, A. (2011). The Era of Electronic Medical Records.

Health IT and Health Care Quality. (n.d.). Retrieved October 27, 2014, from http://www.healthit.gov/patients-families/health-it-and-health-care-quality

Henderson, R. (2012). Industry Employment and Output Projections to 2020. Retrieved from Monthly Labor Review: Bureau of Labor Statistics, U.S. Department of Labor.

ICD-10. (n.d.). Retrieved October 27, 2014, from http://www.healthcareitnews.com/directory/icd-10-coding

Massachusetts eHealth Institute (2014). MeHI Provider and Consumer Health IT Research Study.

Nextgen. (n.d.). Retrieved October 27, 2014, from https://www.nextgen.com/

NextGen Ambulatory EHR. (n.d.).Retrieved October 27, 2014, from https://www.nextgen.com/Products-and-Services/Ambulatory/Electronic-Health-Records-EHR

Prestigiacomo, J. (2010). UMass Memorial Health Care Invests $100M in HIT Mega Project.

Professional EHR. (n.d.). Retrieved October 27, 2014, from http://www.myendnoteweb.com.ezproxy.wpi.edu/EndNoteWeb.html

Psychiatry Resources. (n.d.). Retrieved October 27, 2014, from http://libraryguides.umassmed.edu/content.php?pid=174874&sid=1813216

Pulley, J. (2012). 30% of Docs Say Electronic Health Records Negatively Affect Care

Quintero, et. al. (n.d.). "Workload Optimized Systems: Tuning POWER7 for Analytics", Abstract.

Recognition Davies. (n.d.). Retrieved October 27, 2014, from http://www.epic.com/recognition-davies.php

Recognition Klas. (n.d.). Retrieved October 27, 2014, from http://www.epic.com/recognition-klas.php

Reliant Medical Group Database Developer Salary in Worcester, MA. (n.d.). Retrieved March 3, 2015 from http://www.glassdoor.com/Salary/Reliant-Medical-Group-Database-Developer-Worcester-Salaries-EJI_IE611458.0,21_KO22,40_IL.41,50_IM941.htm

Reese, S. (2012). EHRs: Which Ones Doctors Like and How Their Lives Changed.

Saint Vincent Medical Group. (n.d.). Retrieved October 27, 2014, from http://www.stvincentmedgroup.com/

Seltz, D. (2013). Letter to CMIPA: The Commonwealth of Massachusetts Health Policy Commission.

Shehadi, R., Tohme, W., & Baker, E. (2012). IT and Healthcare: Evolving Together at the Cleveland Clinic.

Stangor, C. (2014). Introduction to Psychology.

Stocks, J. T. (2003). Basic Statistics Review.

Sunrise Clinical and Revenue Cycle Management. (n.d.). Retrieved October 27, 2014, from http://www.allscripts.com/products-services/products/sunrise-clinical-and-revenue-cycle

Testing an Algorithm. (n.d.). Retrieved October 27, 2014, from http://hsc.csu.edu.au/sdd/core/cycle/3520/test_algorithms.htm

The Clinical Partner of UMMS. (n.d.). Retrieved October 27, 2014, from http://www.umassmed.edu/about/clinical/

UMass Memorial health care develops fully integrated EMR. (n.d.). Retrieved October 27, 2014, from http://www.eurekalert.org/pub_releases/2010-04/uomm-umh040110.php

UMass Memorial Health Care Selects Allscripts Electronic Health Record and Practice Management as Preferred Solution for Community Physicians. (n.d.). Retrieved October 27, 2014, from http://investor.allscripts.com/phoenix.zhtml?c=112727&p=irol-newsArticle&ID=1341295&highlight=

VistA. (n.d.). Retrieved October 27, 2014, from http://www.ehealth.va.gov/VistA.asp

What are Electronic Health Records? (n.d.). Retrieved October 27, 2014, from http://www.stvincentmedgroup.com/learning-center/blog/what-are-electronic-health-records

What is Electronic Medical Record (EMR)? (n.d.). Retrieved October 26, 2014, from http://www.healthit.gov/providers-professionals/electronic-medical-records-emr

What Is Linux: An Overview of the Linux Operating System - The Business of Linux. (2009). Retrieved from https://www.linux.com/learn/new-user-guides/376?start=5

Why Adopt EHRs? (n.d.). Retrieved October 26, 2014, from http://www.healthit.gov/providers-professionals/why-adopt-ehrs

World Health Organization (2010). World Health Organization Assesses the World's Health Systems.

Appendix

Appendix I: Meaningful Use Requirements

- **15 Core Objectives:**
  1. Computerized Provider Order entry (CPOE)
  2. Drug-drug and drug-allergy checks
  3. Maintain an up-to-date problem list of current and active diagnoses
  4. E-Prescribing (eRx)
  5. Maintain active medication list
  6. Maintain active medication allergy list
  7. Record demographics
  8. Record and chart changes in vital signs
  9. Record smoking status for patients 13 years or older
  10. Report ambulatory clinical quality measures to CMS/States
  11. Implement clinical decision support
  12. Provide patients with an electronic copy of their health information, upon request
  13. Provide clinical summaries for patients for each office visit
  14. Capability to exchange key clinical information
  15. Protect electronic health information
- **10 Menu Objectives:**
  1. Submit electronic data to immunization registries
  2. Submit electronic syndrome surveillance data to public health agencies
  3. Drug formulary checks
  4. Incorporate clinical lab-test results
  5. Generate lists of patients by specific conditions
  6. Send reminders to patients for preventive/follow-up care
  7. Patient-specific education resources
  8. Electronic access to health information for patients
  9. Medication reconciliation
  10. Summary of care record for transitions of care

*eligible professionals and hospitals must either objective 1 or objective 2 for their 5 Menu Objectives

Clinical Quality Measures:

Must choose 2 core clinical quality measures and 3 clinical quality measures that you select from an additional list.

- Clinical quality measures:
  1. Hypertension: Blood Pressure Measurement
  2. Preventive Care and Screening Measure Pair ：
     1) Tobacco Use Assessment
     2) Tobacco Cessation Intervention
  3. Adult Weight Screening and Follow-up
- If the data produced by your EHR indicates a zero for the denominator of one or more of the core clinical quality measures, then you must choose one or more alternate core clinical quality measures from this list.
  1. Weight Assessment and Counseling for Children and Adolescents
  2. Preventive care and Screening: Influenza Immunization for Patients 50 years Old or Older
  3. Childhood Immunization Status
- Additional Clinical Quality Measures:

  1. Diabetes: Hemoglobin A1c Poor Control

  2. Diabetes: Low Density Lipoprotein (LDL) Management and Control

  3. Diabetes: Blood Pressure Management

  4. Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB)
  Therapy for Left Ventricular Systolic Dysfunction (LVSD)

  5. Coronary Artery Disease (CAD): Beta-Blocker Therapy for CAD Patients with Prior Myocardial Infarction (MI)

  6. Pneumonia Vaccination Status for Older Adults

7. Breast Cancer Screening
8. Colorectal Cancer Screening
9. Coronary Artery Disease (CAD): Oral Antiplatelet Therapy Prescribed for Patients with CAD
10. Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)
11. Anti-depressant medication management: (a) Effective Acute Phase Treatment, (b)Effective Continuation Phase Treatment
12. Primary Open Angle Glaucoma (POAG): Optic Nerve Evaluation
13. Diabetic Retinopathy: Documentation of Presence or Absence of Macular Edema and Level of Severity of Retinopathy
14. Diabetic Retinopathy: Communication with the Physician Managing Ongoing Diabetes Care
15. Asthma Pharmacologic Therapy
16. Asthma Assessment
17. Appropriate Testing for Children with Pharyngitis
18. Oncology Breast Cancer: Hormonal Therapy for Stage IC-IIIC Estrogen Receptor/Progesterone Receptor (ER/PR) Positive Breast Cancer
19. Oncology Colon Cancer: Chemotherapy for Stage III Colon Cancer Patients
20. Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients
21. Smoking and Tobacco Use Cessation, Medical assistance: a) Advising Smokers and Tobacco Users to Quit, b) Discussing Smoking and Tobacco Use Cessation Medications, c) Discussing Smoking and Tobacco Use Cessation Strategies
22. Diabetes: Eye Exam
23. Diabetes: Urine Screening
24. Diabetes: Foot Exam
25. Coronary Artery Disease (CAD): Drug Therapy for Lowering LDL-Cholesterol
26. Heart Failure (HF): Warfarin Therapy Patients with Atrial Fibrillation

27. Ischemic Vascular Disease (IVD): Blood Pressure Management

28. Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antithrombotic

29. Initiation and Engagement of Alcohol and Other Drug Dependence
Treatment: a) Initiation, b) Engagement

30. Prenatal Care: Screening for Human Immunodeficiency Virus (HIV)

31. Prenatal Care: Anti-D Immune Globulin

32. Controlling High Blood Pressure

33. Cervical Cancer Screening

34. Chlamydia Screening for Women

35. Use of Appropriate Medications for Asthma

36. Low Back Pain: Use of Imaging Studies

37. Ischemic Vascular Disease (IVD): Complete Lipid Panel and LDL Control

38. Diabetes: Hemoglobin A1c Control (<8.0%)

Appendix II: Award Winning Customers of Epic Systems

Organizations at Acute Stage 7
Representing 122 Hospitals

Baylor Scott & White Health
Bon Secours Health System
Cedars-Sinai Health System
Children's Medical Center
Cincinnati Children's
Deaconess Health System
Edgerton Hospital & Health Services
Hawai'i Pacific Health
Hennepin County Medical Center
Hospital Sisters Health System - Eastern Wisconsin
Kaiser Permanente
Lakeland Health
Lancaster General Health
Legacy Health
Nemours
NorthShore University HealthSystem
The Ohio State University Wexner Medical Center
Sentara Healthcare
SSM Health Care
Stanford University Medical Center
Stoughton Hospital
Tampa General Hospital
Texas Health Resources
TriHealth
Tucson Medical Center
UC Davis Medical Center
UCLA Health
UC San Diego Health System
University of Iowa Hospitals & Clinics
UW Health
West Virginia University

Organizations at Ambulatory Stage 7
Representing 985 Clinics

Atrius Health and Reliant Medical Group
Baylor Scott & White Health
Bon Secours Health System
Cincinnati Children's
Dean Clinic
Essentia Health
Hawai'i Pacific Health
Lakeland Health
Legacy Health
MetroHealth (OH)
Nemours
NorthShore University HealthSystem
Novant Health
The Ohio State University Wexner Medical Center
TriHealth
University of Iowa Health Care
West Virginia University

Appendix III: Stages of EMR Adoption

The **EMR Adoption Model**[SM] (EMRAM) identifies and scores hospitals using an 8 step scale that charts the path to a fully paperless environment.

United States EMRAM | Canada EMRAM | NEW! United States Ambulatory EMRAM

United States EMR Adoption Model [SM]

Stage	Cumulative Capabilities	2014 Q1	2014 Q2
Stage 7	Complete EMR; CCD transactions to share data; Data warehousing; Data continuity with ED, ambulatory, OP	3.1%	3.2%
Stage 6	Physician documentation (structured templates), full CDSS (variance & compliance), full R-PACS	13.3%	15.0%
Stage 5	Closed loop medication administration	24.2%	27.5%
Stage 4	CPOE, Clinical Decision Support (clinical protocols)	15.7%	15.3%
Stage 3	Nursing/clinical documentation (flow sheets), CDSS (error checking), PACS available outside Radiology	27.7%	25.4%
Stage 2	CDR, Controlled Medical Vocabulary, CDS, may have Document Imaging; HIE capable	7.2%	5.9%
Stage 1	Ancillaries - Lab, Rad, Pharmacy - All Installed	3.2%	2.8%
Stage 0	All Three Ancillaries Not Installed	5.6%	4.9%
Data from HIMSS Analytics® Database ©2014		N = 5449	N = 5447

PLEASE NOTE: These graphics are an abbreviated version of the HIMSS Analytics EMR Adoption Model. All organizations must secure permission to post our model on any public notices and to obtain their score they must complete the HIMSS Analytics study prior to validation of their score.

Appendix IV: Import into MySQL Perl Script

```
#!/usr/bin/perl
# BB
use strict;

# Define variables
my $f1;
my $f2;
my $f3;
my $f4;
my $f5;
my $f6;

my $one = 0;
#
# Open file
#
open (IP, "/research/Data\ from\ sponsor/MYCHART_DEMO_FINAL.txt");

# Read through 1st file and load associate array "buffer"
while (<IP>) {
   chomp;

   # Skip first row
   $one++;
   if ($one eq 1) {
      next;
   }

   # Read line from file and split into 6 fields seperated by a ","
   ($f1,$f2,$f3,$f4,$f5,$f6) = split /,/, $_;

   # Delete "
   $f1 =~ s/\"//g;
   $f2 =~ s/\"//g;
   $f3 =~ s/\"//g;
   $f4 =~ s/\"//g;

   # Fixup date time
   (my $my_date, my $my_time) =  split / /, $f5;
   (my $my_month, my $my_day, my $my_year) = split /\//, $my_date;
   $my_time =~ s/\n//g;
   $my_time =~ s/\r//g;

   (my $my_datel, my $my_timel) =  split / /, $f6;
   (my $my_monthl, my $my_dayl, my $my_yearl) = split /\//, $my_datel;
   $my_timel =~ s/\n//g;
   $my_timel =~ s/\r//g;

   # Output SQL statement
   printf("INSERT INTO MYCHART_DEMO SET STUDYID='%s', AGE='%s', SEX='%s', PAT_STATUS='%s',
        FIRST_MYCHART_LOGINDT='%s-%s-%s %s', LAST_MYCHART_LOGINDT='%s-%s-%s %s';\n",
        $f1, $f2, $f3, $f4,
        $my_year, $my_month, $my_day, $my_time,
        $my_yearl, $my_monthl, $my_dayl, $my_timel);
}
close IP;
```

Appendix V: Export from MySQL Perl Script

```
#!/usr/bin/perl

use strict;
use DBI;

# MySQL
my $dbh;
my $sql;
my $sth;
my $ref;

$dbh = DBI->connect('DBI:mysql:reliantmqp','root','aj;45l',);

$sql = "SELECT * FROM overdue";
$sth = $dbh->prepare($sql);
$sth -> execute;

while ($ref = $sth->fetchrow_hashref) {
   printf("%s,%s,%s,%s,%s,%s\n",
         $ref->{'STUDYID'},
         $ref->{'MSG_ID'},
         $ref->{'MYCHART_MESSAGE_ID'},
         $ref->{'MYC_PARENT_MESSAGE_ID'},
         $ref->{'MYC_MSG_TYP_C'},
         $ref->{'MSG_TIME'});
}
$sth->finish;
```

Appendix VI: Code for Combining IB_Message Table and MyChart_Message Table

```
CREATE TABLE COMBINEDMESSAGES
(STUDYID VARCHAR(18),
MSG_ID VARCHAR(18) PRIMARY KEY,
MSG_TYPE_C INT,
MYC_MSG_ID VARCHAR(18),
MYC_MSG_TYP_C INT,
TOFROM_PAT_C INT,
MYC_PARENT_MESSAGE_ID VARCHAR(18),
MSG_TIME DATETIME);

INSERT INTO COMBINEDMESSAGES
SELECT a.STUDYID,
a.MSG_ID,
a.MSG_TYPE_C,
b.MESSAGE_ID,
b.MYC_MSG_TYP_C,
b.TOFROM_PAT_C,
NULLIF(b.PARENT_MESSAGE_ID,''),
a.CREATE_TIME
FROM IBMESSAGE a
INNER JOIN MYCHART_MESSAGE b on a.MSG_ID = b.INBASKET_MSG_ID;
```

Appendix VII: Code for Creating MSG_SENT_READ Table

```
CREATE TABLE SENTDONE
(STUDYID VARCHAR(18),
MSG_ID VARCHAR(18),
MSG_SENT_TIME DATETIME,
MSG_DONE_TIME DATETIME,
TOFROM_PAT_C INT,
MYCHART_MESSAGE_ID VARCHAR(18),
MYC_PARENT_MESSAGE_ID VARCHAR(18),
MYC_MSG_TYP_C INT,
MSG_TYPE_C INT;

INSERT INTO SENTDONE
SELECT a.STUDYID,
a.MSG_ID,
b.STATUS_CHG_TIME,
c.STATUS_CHG_TIME,
a.TOFROM_PAT_C,
COALESCE(c.MYCHART_MESSAGE_ID, b.MYCHART_MESSAGE_ID),
a.MYC_PARENT_MESSAGE_ID,
a.MYC_MSG_TYP_C,
a.MSG_TYPE_C
FROM COMBINEDMESSAGES a
LEFT OUTER JOIN MSG_STAT_AUDIT b ON a.MSG_ID = b.MSG_ID AND b.STATUS_AUDIT_C=2
LEFT OUTER JOIN MSG_STAT_AUDIT c ON a.MSG_ID = c.MSG_ID AND c.STATUS_AUDIT_C=4;
```

Appendix VIII: Code for Creating Overdue Patient Table

```
CREATE TABLE OVERDUEPATIENTS
(STUDYID VARCHAR(18) PRIMARY KEY,
AGE INT,
SEX CHAR(1),
PAT_STATUS VARCHAR(18),
FIRST_MYCHART_LOGINDT DATETIME,
LAST_MYCHART_LOGINDT DATETIME,
OVERDUECOUNT INT,
TARGETMSGCOUNT INT,
LOGINCOUNT INT,
PROXYYN CHAR(1),
RAFSCORE decimal(10,3),
TOTALMSGCOUNT INT);

INSERT INTO OVERDUEPATIENTS
SELECT a.*, b.COUNT, c.MsgCount, d.LOGINS,
CASE WHEN e.STUDYID IS NULL THEN 'N' ELSE 'Y' END,
f.RAFSCORE, g.MsgCount
FROM MYCHART_DEMO a
inner join countoverdue b on a.STUDYID=b.STUDYID
INNER JOIN countmessages c on a.STUDYID=c.STUDYID
LEFT OUTER JOIN countlogin d on a.STUDYID=d.STUDYID
LEFT OUTER JOIN PROXY e on a.STUDYID = e.STUDYID
LEFT OUTER JOIN RAFSCORE f ON a.STUDYID = f.STUDYID
LEFT OUTER JOIN count g on a.STUDYID = g.STUDYID;

UPDATE OVERDUEPATIENTS SET LOGINCOUNT=0 WHERE LOGINCOUNT IS NULL;
UPDATE OVERDUEPATIENTS SET RAFSCORE=1 WHERE RAFSCORE IS NULL;
```

Appendix IX: Code for Creating Never-Read Patient Table

```
CREATE TABLE OVERDUEPATIENTS
(STUDYID VARCHAR(18) PRIMARY KEY,
AGE INT,
SEX CHAR(1),
PAT_STATUS VARCHAR(18),
FIRST_MYCHART_LOGINDT DATETIME,
LAST_MYCHART_LOGINDT DATETIME,
NEVERDONECOUNT INT,
TARGETMSGCOUNT INT,
LOGINCOUNT INT,
PROXYYN CHAR(1),
RAFSCORE decimal(10,3),
TOTALMSGCOUNT INT);

INSERT INTO NEVERDONEPATIENTS
SELECT a.*, b.COUNT, c.MsgCount, d.LOGINS,
CASE WHEN e.STUDYID IS NULL THEN 'N' ELSE 'Y' END,
f.RAFSCORE, g.MsgCount
FROM MYCHART_DEMO a
inner join countneverdone b on a.STUDYID=b.STUDYID
INNER JOIN countmessages c on a.STUDYID=c.STUDYID
LEFT OUTER JOIN countlogin d on a.STUDYID=d.STUDYID
LEFT OUTER JOIN PROXY e on a.STUDYID = e.STUDYID
LEFT OUTER JOIN RAFSCORE f ON a.STUDYID = f.STUDYID
LEFT OUTER JOIN count g on a.STUDYID = g.STUDYID;

UPDATE NEVERDONEPATIENTS SET LOGINCOUNT=0 WHERE LOGINCOUNT IS NULL;
UPDATE NEVERDONEPATIENTS SET RAFSCORE=1 WHERE RAFSCORE IS NULL;
```

Appendix X: Logic for Prediction Algorithm

(Please be advised that SQL syntax could be slightly different based on the environment and version of SQL tool. What's provided here is ONLY the logic.)

1. Create a temporary table that combines IB Message table with MyChart Message table to link the messages together (See Appendix VI).

2. Create a temporary table to determine how long it takes for each message to be read (See Appendix VII).

```
INSERT INTO #DONETIME
SELECT STUDYID,
MSG_ID,
MYCHART_MESSAGE_ID,
TOFROM_PAT_C,
HOUR(TIMEDIFF(MSG_DONE_TIME,MSG_SENT_TIME)) AS 'DONETIME',
MYC_PARENT_MESSAGE_ID,
MYC_MSG_TYP_C
FROM SENTDONE;
```

3. Create a temporary table to count the total problem messages (overdue or never-read) each patient has within the past year.

```
INSERT INTO #PROBLEMMSG
SELECT * FROM #DONETIME
WHERE (DONETIME > 240 OR DONETIME IS NULL)
AND MYC_MSG_TYP_C IN (1,11);

INSERT INTO #countproblemmsg
SELECT STUDYID, COUNT(*) AS 'PROBLEMMSGCOUNT'
FROM #PROBLEMMSG
GROUP BY STUDYID;
```

4. Create a temporary table to count the total messages each patient has received within the past year.

```
INSERT INTO #counttotalmsg
SELECT STUDYID, COUNT(*) AS 'TOTALMSGCOUNT'
FROM COMBINEDMESSAGES
GROUP BY STUDYID;
```

5. Create a temporary table to count the important messages each patient has received within the past year.

```
INSERT INTO #counttargetmsg
SELECT STUDYID, COUNT(*) AS 'TARGETMSGCOUNT'
FROM COMBINEDMESSAGES
WHERE MYC_MSG_TYP_C IN (1,11)
GROUP BY STUDYID;
```

6. Create a temporary table to count the number of logins of each patient.

```
INSERT INTO #countlogin
SELECT STUDYID, COUNT(*) AS 'LOGINCOUNT'
FROM LOGIN
WHERE MYC_UA_TYPE_C=1
GROUP BY STUDYID;
```

7. Create a temporary table to count the number of logins of each patient.

```
INSERT INTO #BADPATIENTS
SELECT a.*,
b. PROBLEMMSGCOUNT,
c. TOTALMSGCOUNT,
d. TARGETMSGCOUNT,
ISNULL(e. LOGINCOUNT,0) AS 'LOGINCOUNT',
ISNULL(f.RAFSCORE ,1) AS 'RAFSCORE',
NULL AS MSGPERLOGIN,
NULL AS PROBLEMPERCENT
FROM MYCHART_DEMO a
INNER JOIN countproblemmsg b ON a.STUDYID=b.STUDYID
INNER JOIN counttotalmsg c ON a.STUDYID=c.STUDYID
INNER JOIN counttargetmsg d ON a.STUDYID=d.STUDYID
LEFT OUTER JOIN countlogin e ON a .STUDYID=e.STUDYID
LEFT OUTER JOIN RAFSCORE f ON a.STUDYID = f.STUDYID;

UPDATE #BADPATIENTS
SET MSGPERLOGIN = TOTALMSGCOUNT/LOGINCOUNT;

UPDATE #BADPATIENTS
SET PROBLEMPERCENT = PROBLEMMSGCOUNT/TARGETMSGCOUNT;
```

8. Generate a list of patient who are no longer considered active (no login within the year but have unread messages).

```
INSERT INTO INACTIVEPATIENTS
SELECT * from #BADPATIENTS
WHERE LOGINCOUNT = 0
AND LAST_MYCHART_LOGINDT < DateAdd(yy, -1, GetDate());
```

9. Generate a list of patients who are active but rarely check their messages (with message per login ratio greater than 1.9 and problem message percentage greater than 46.8%).

```
DELETE FROM #BADPATIENTS
WHERE STUDYID IN (SELECT STUDYID FROM INACTIVEPATIENTS)

INSERT INTO HIGHRISKPATIENTS
SELECT * FROM #BADPATIENTS
WHERE (MSGPERLOGIN > 1.9 OR MSGPERLOGIN IS NULL)
AND PROBLEMPERCENT > 0.46;
```