Information Transmission in the MIMIC II Clinical Database

By
Rajesh Mishra
B.Sc., Computer Science
Rensselaer Polytechnic Institute, 1992

Submitted to the System Design and Management Program
In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Engineering and Management

At the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2010

Signature of Author ______________________________
Rajesh Mishra

Certified by ______________________________
Peter Szolovits
Professor of Computer Science and Engineering
Head Clinical Decision-Making Group
Computer Science and Artificial Intelligence Laboratory (CSAIL)
Massachusetts Institute of Technology

____
Roy Welsch
Eastman Kodak Leaders for Global Operations (LGO) Professor of Management
Professor of Statistics and Engineering Systems
Sloan School of Management and Engineering Systems
Massachusetts Institute of Technology

Accepted by ______________________________
Patrick Hale
Director
System Design and Management Program

Information Transmission in the MIMIC II Clinical Database

by Rajesh Mishra

Submitted to the System Design and Management Program
In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Engineering and Management

Abstract

The promise of the Electronic Medical Record (EMR) to store, retrieve, and communicate medical information effectively for a caregiver team has remained largely unfulfilled since its advent in the late 1960's. Previous studies have cited that the communication function of the EMR is critical to its successful adoption. Based on Mediated Agent Interaction theory, this study proposes a message-based model of transmission of clinical information in the EMR. This model is implemented on an existing ICU clinical database, MIMIC II, to produce a database of transmission events. Three metrics for information transmission are derived from exploratory and object-attribute analyses: transmission volume, duration, and load (or rate). Also derived is a set of features that includes patient's clinical conditions (with acuity scores and mortality), caregiver type and distribution, care-unit locations, duration of care, and types of clinical records. This list of features is reduced to a set of explanatory variables using correlation and univariate logistic regression. Bayesian Network (BN) models are constructed to predict levels of the transmission metrics. BN models show high prediction accuracy for measuring various levels of messaging volume and load, but marginal accuracy for messaging duration. Results from these methods suggest that the volume of information transmitted in the ICU for adult patients is primarily through charts entered by nurses and respiratory technicians (RTs). The amount of data recorded by RTs increases for patients with higher acuity scores, but transmission from nurses decreases for these patients. The rate at which information is transmitted in the ICU for adult patients is directly related to the rate at which notes and charts are entered, as well as the care-unit location where the data is recorded. Further study is required to investigate factors influencing the length of time information is transmitted in the ICU. This study's findings are based on data recorded by caregivers as clinical observations. Further study is necessary to corroborate these results with clinical communications data, including evidence of reception of clinical information by caregivers. The model proposed by this study may be used as a basis for future research and to discover other patterns of clinical communications.

Thesis Supervisor:	Peter Szolovits
Title:	Professor of Computer Science and Engineering
Thesis Supervisor:	Roy Welsch
Title:	Professor of Statistics and Engineering Systems

Acknowledgements

There are several people I would like to acknowledge that have supported me greatly during the process of completing this work. Firstly, I would like to thank Professor Peter Szolovits for introducing me to the field of biomedical informatics and guiding me throughout my thesis-work. I would also like to thank Professor Roy Welsch for taking interest in this research area and also including me as a student in his course.

I would also like to express gratitude to Dr. Marco Ramoni for his instruction and support in my research efforts. This work would not have been possible without the permission of Dr. Roger Mark and Dr. Gari Clifford to use the MIMIC II database, for which I am grateful. In addition, I would like to thank Mauricio Villarroel for his assistance with the database and answering all of my questions at the oddest possible hour.

My family could not have been more supportive during my student life at MIT. My wife, Jyoti, has made balancing work, student, and family life seem effortless. Her inspiration has encouraged me to complete this thesis and course-work. I would also like to express my appreciation to my daughters, Anisha and Arya, for their understanding of my time away from them during these past two years.

Table of Contents

Tables

Figures

Introduction

Although computerization of medical records has been a "work-in-progress" for over four decades (P.C. & McDonald, 2006), its adoption by clinicians has recently received far greater attention in the United States than in the past. Healthcare administrators tout the benefits of the Electronic Health Record (EHR) by proposing that it will make clinical information accessible to the providers whenever and wherever needed in a comprehensible format, thereby reducing the likelihood of medical errors and lowering the overall costs of healthcare services (Halamka, 2008). Clinicians, however, argue that current electronic medical record systems are complicated to use, may divert their attention away from patients at the point of care, and are expensive to incorporate into their practices (Bhattacherjee & Hikmet, 2007). This longstanding debate, however, has not prevented the US congress from taking legislative action to incentivize all clinicians to employ EHRs in their practices. The Health Information Technology for Economic and Clinical Health Act (HITECH) provision of the 2009 American Recovery and Reinvestment Act (ARRA) rewards healthcare providers who adopt "meaningful use" of EHR systems through Medicare and Medicaid reimbursements. Clinicians who choose not to adopt EHR systems by 2015 will be subject to reduction of such reimbursements progressively over time (HHS, Breach Notification for Unsecured Protected Health Information, 2009). As a consequence, the clinician's dilemma regarding EHRs is transforming from whether or not to adopt this technology to how to make effective use of this technology for clinical work. In January 2010, the US Department of Health and Human Services (HHS) released a document which defines "meaningful use" of Electronic Health Records (HHS, Proposed Rules for Electronic Health Record Incentive Program, 2010). A key element in this definition, cited in Table 2 of this document (and restated in Appendix A: Federal Guideline for Meaningful Use of EMRs related to Clinical CommunicationA, is the technology's effect upon coordination of patient care with the goal to "exchange meaningful clinical information among professional healthcare team" (HHS, Proposed Rules for Electronic Health Record Incentive Program, 2010).

This declaration highlights one of the key uses of the medical record - communication between providers. As clinical care has evolved from a patient being seen by a single doctor to that of a caregiver team serving the needs of the patient, the need for coordination of care has been steadily on the rise (Miller, Scheinkestel, & Michele, 2009). As medical professionals become increasingly specialized in their areas of expertise (for example, nurses as respiratory therapists), sources and volume of clinical information will also grow and require proper integration to ensure effectiveness. In such a clinical environment, effective communication between providers is critical to the overall care for the patient. The repercussions of inadequate caregiver communication are several-fold:

(a) Clinical outcome - An Australian survey found that communication issues are the most common cause of preventable disability or death, ranked higher than technical incompetence and neglect (Coiera, Jayasuriya, Hardy, Bannan, & Thorpe, 2002). A 2005 study by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) reported that approximately 70% of all causes of sentinel events were due to lack of proper communication (JCAHO, 2005).
(b) Cost of service - a recent study estimated that the American Health System could save $10B a year by improving communication (Sands, 2008).

(c) Quality of care - the efficiency of delivering health services is impacted by communication errors, which can lead to both patient and provider dissatisfaction (Edwards, et al., 2009).
(d) Technology adoption – previous studies claim that interpersonal communication can serve as a barrier to adoption of technologies such as Electronic Health Records, as these technologies currently do not address the requirements of this use-case adequately (Coiera, When Conversation Is Better Than Computation, 2000).

The last item noted above underscores the importance of the communication function of the electronic health record. Clinical communication, as described in detail in the next section, requires both transmission and reception of clinical information between caregivers. This paper proposes a model for analysis of clinical communication in the EHR. It primarily focuses on the transmission aspect of communication in the EHR, by exploring patterns of transmission of clinical information.

Thesis Organization

This paper consists of six main sections, starting with this introduction to the topic and concluding with a summary of the results of this research and possible avenues for future research.

The second section surveys the background knowledge on which this thesis research is based. It describes an existing theory of multi-agent social interaction, the computing technologies underlying electronic health records, and an introduction to predictive, probabilistic models known as Bayesian belief networks.

The third section serves as the foundation for this thesis work. It proposes a model for transmission of clinical information based on the principles outlined in the second section. It lists the assumptions that this model makes. It depicts how transmission can be viewed, analyzed, and measured in the electronic health record through this model.

The fourth section describes the tools and methods used to validate the proposed model. The methods described include the implementation of the model on an existing clinical database and how the model is used to determine key factors influencing transmission of clinical information in an Intensive Care Unit (ICU).

The fifth section evaluates the model against its implementation on an existing clinical database. It describes the features discovered by the model that influence transmission and the performance of predictive models based on these features.

The sixth section states conclusions drawn from the observations made in the previous sections. It also cites the limitations of the analysis in this research and suggests possible areas for future research based on this work.

The appendices support all sections mentioned above with evidential information. The bibliography cites previous work referenced in this study.

Background

Many studies have been conducted on doctor-patient communications, whereas, until recently, there has been relatively scant research on interactions between collaborators in a caregiver team. This may be partially attributed to only recent development of the evolving role of nurses and technicians from that of subordinates to medical doctors and administrators (playing the "Nurse-Doctor game" , as Leonard Stein famously noted in 1968 (Stein, Watts, & Howell, 1967)) to partners in clinical decision making as nurse practionners, respiratory therapists, and other specialized professional roles (Germov & Freij, 2009). The earliest work in clinical communications between providers can be found in publications by Enrico Coiera in the mid to late 1990's. As the model proposed by this thesis is based on much of his work, a brief introduction to his theory of mediated agent interaction is provided. A glossary of clinical communication terms discussed in this section is also included in Appendix B: Glossary of Clinical Communication Terms.

Mediated Agent Interaction Theory

Studies in clinical communications have their foundations in computer and social sciences, from information science, telecommunications, and artificial intelligence to sociology, psychology, and linguistics. In his seminal work on mediated agent interaction, Coiera proposes a computational model for interaction between agents (Coiera, Interaction design theory, 2002).

In this model, an agent is operationally defined as an object, either human or non-human, that is capable of interacting with another. An interaction occurs when multiple agents exchange messages with one another. A message is an object that encapsulates information and originates from a transmitting agent and is destined for either one or multiple receiving agent*s*. communication is the process of delivering information between agents and requires a channel, a medium upon which information is passed (e.g., telephone, e-mail, and face-to-face are various types of media). Figure 1 depicts these basic components of agent-based communication.

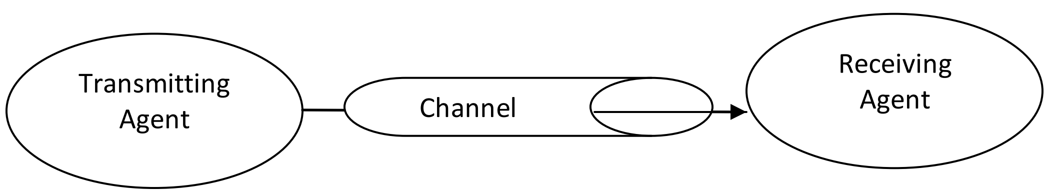


Figure 1: Simplest model of agent-based communication: messages are said to be unicast from the transmitting agent to the receiving agent.

When messages are sent from one transmitting agent to one and only one receiving agent, the communication mode is unicast. When messages are sent by one transmitting agent and received by multiple receiving agents, the communication mode is multicast (as shown in Figure 2). When messages are transmitted to all agents in the communication environment, the communication mode is broadcast.

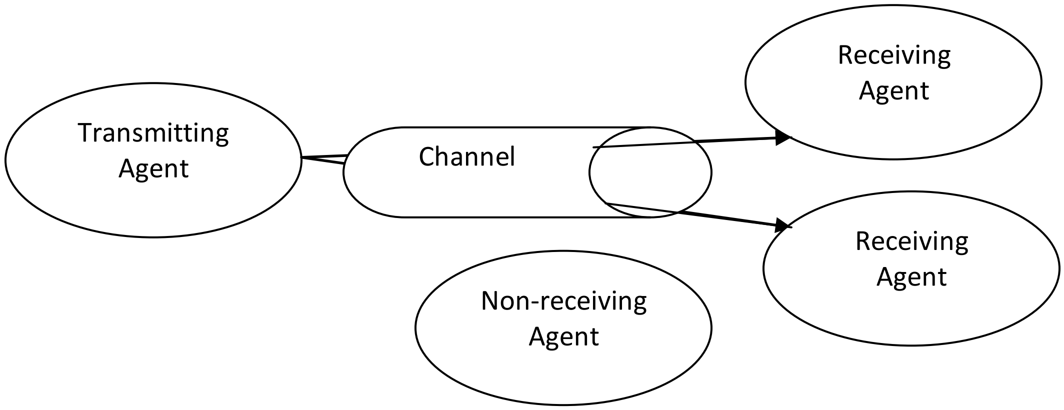


Figure 2: Multicast agent-based communication.

Coeira's Mediated Agent Interaction Theory differs from basic social interaction theory in that it makes a distinction between the agent and the cognitive apparatus through which the agent interprets information as knowledge. According to this view, the channel is not the only source of error in communication. Differences in perception of information between agents can also contribute to communication error, as well as the environmental context in which the interaction occurs and the relationship between agents. For example, the physician (or transmitting agent) sends a message "Check Pressure" to a nurse (or receiving agent) on a note (the channel). Even if the message is not altered in the channel (e.g., the text remains legible and is read correctly), the physician may have intended to communicate "Enter the patient's systolic and diastolic blood pressure readings into the chart", whereas the nurse may have interpreted the message as "Make sure the patient's ocular pressure is within the nominal range". These two perceptions of the message content are markedly different between the agents in this example, as the informational context is not apparent within the message itself (i.e., the type of pressure is not being explicitly referred to) and the functions of the roles of the agents is not mutually understood (i.e., the physician assumes the nurse will make a clinical decision whether the pressure reading fits within the nominal range).

For the purpose of retaining a broader definition of mediated agent interaction, information is strictly defined as un-interpreted data passed between agents. Information can reside on any medium in either structured or unstructured forms. A datum is a single observational point (for example, a systolic blood pressure reading) (Shortliffe & Barnett, 2006). Each agent possesses a model which serves as its own container of knowledge, or interpreted information. In this way, knowledge is regarded as information structured within an agent's model such that it can be interpreted according to the agent's own perception.

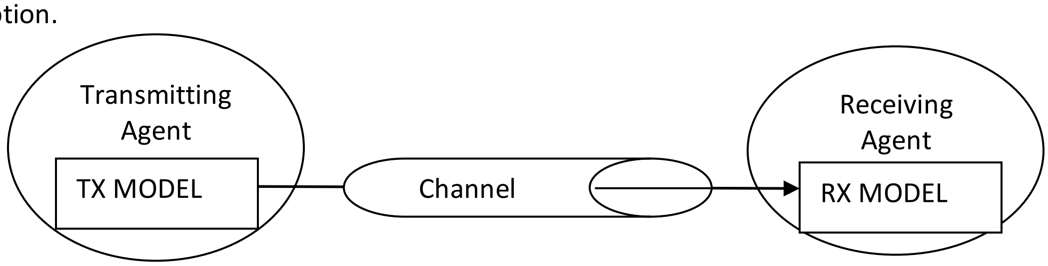


Figure 3: Simple Model of Mediated Agent Based Communication.

Knowledge is said to be shared between agents, also known as common ground, when at least one of the agents' models has gained knowledge as a result of their interaction. The process of checking whether agents' models agree with one another (i.e., validating shared knowledge) is known as grounding. The grounding process has some associated cost, which comes in two forms: (a) background information passed between agents, and (b) maintaining common ground between agents. One purpose of grounding is to progressively facilitate interactions between agents, such that the cost of interaction is reduced over time. grounding efficiency is the degree to which common ground conserves the amount of information sent during communication. Interaction is ground-positive when grounding has shortened the message; ground-negative when it has lengthened the message; ground-neutral when message size is unchanged. Coeira claims that interactions between agents in a given environment evolve to a state that has the lowest grounding cost, known as the Law of Mediated Centre (Coiera, Mediated Agent Interaction, 2001). The underlying assumption is that at this equilibrium point, the cost-to-benefit ratio of grounding is globally optimal for all interacting agents.

Clinical Databases

Central to the model proposed by this thesis is the recording and interpretation of clinical information. Although, in practice, clinical databases consist of a blend of medical observations, such as body temperature readings, and opinions of medical professionals, such as diagnoses and prognoses, this paper adheres to the strict definition of information described in Mediated Agent Interaction Theory. In this context, clinical information is defined as data that consists of a set of observations. Each datum (singular of data) is regarded as a single clinical observation (for example, blood pressure reading, and patient's age) (Shortliffe & Barnett, 2006). Each observation is recorded only once, but may be viewed multiple times. A clinical record is a collection of observations that has a common attribute. For example, a patient's demographic record consists of observations related to the patient's biographic information, such as age, gender, and expiration status ("dead" or "alive"). Records may be organized in a class-object hierarchical structure with "is-a" and "has-a" relationships with other records. For instance, a patient's record may consist of demographic, medication list, hospital admission records, which in turn consist of birth/death records, drug records, and census records, respectively. Data within a record may be recorded at arbitrarily different times, not necessarily in chronological order. Data may also be viewed as part or whole of the record. A clinical database is a collection of clinical records, which can be appended, modified, viewed, and searched at arbitrary times by its users. Individual observations may be grouped across various records to form a new record. However, these summarized records are not considered part of the clinical database, rather a part of the clinical knowledgebase (Shortliffe & Barnett, 2006). The database only consists of records that collectively comprise a unique set of observations. Although there may be inherent redundancy in the knowledge derived from clinical information, each clinical observation is considered distinct on its own, with its own attribute, value, and recording time.

Figure 4 describes an object-process model view of a medical record system, which illustrates the relationships between the key concepts described above. Clinical observations regarding the patient are made outside the medical record system and only enter the system when they are stored as clinical records. The internal value functions within a database benefit the caregiver, not only by providing means of organizing clinical data (inserting, updating processes) and making it accessible (reading, searching), but also by providing a means of communication with other caregivers (inserting and updating by transmitters and reading and searching by receivers). This latter aspect of the clinical database is the subject of this thesis.

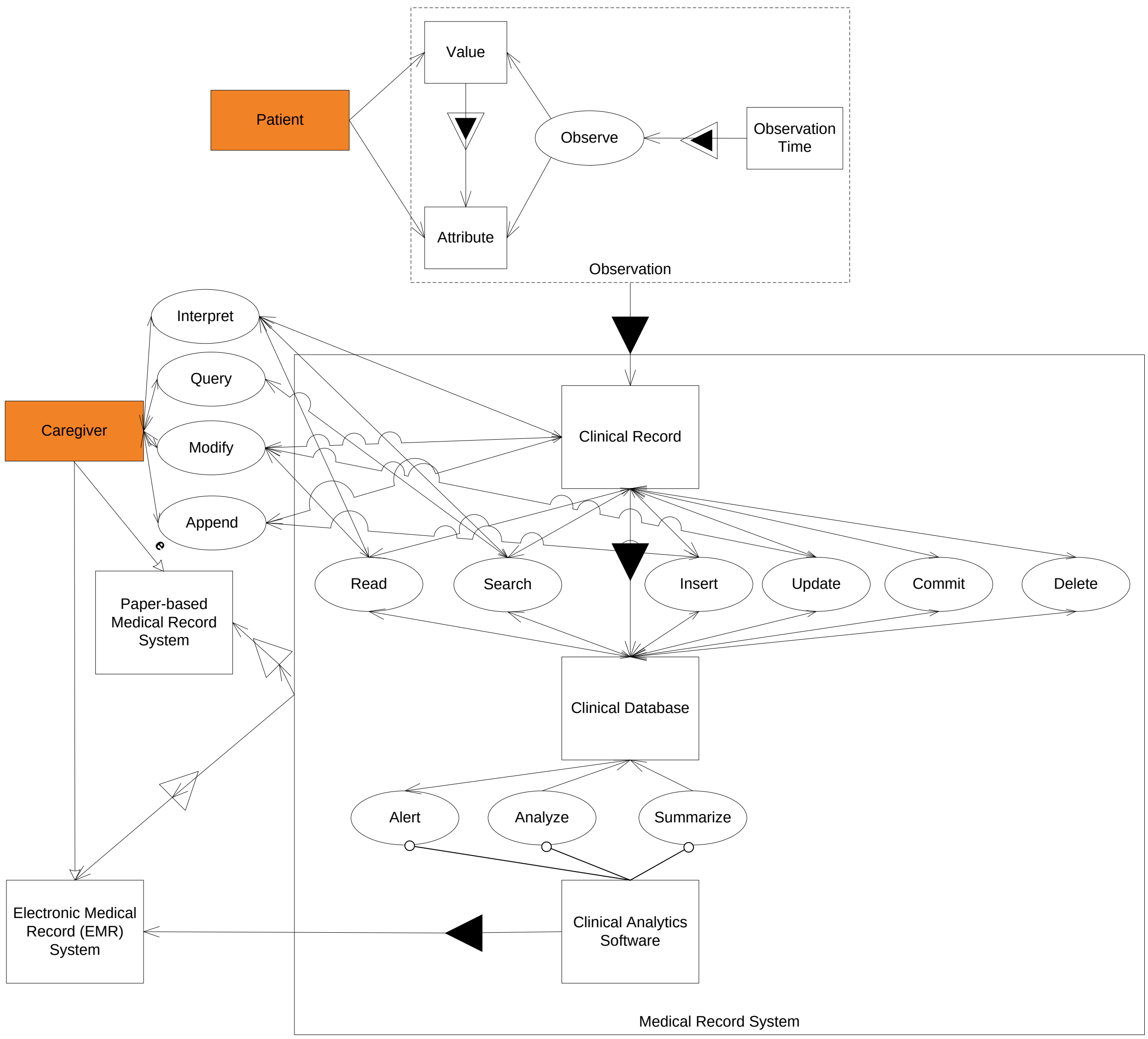


Figure 4: Object-Process Model Abstraction of a Clinical Database within a Medical Record System

There are two types of medical record systems in common use today: (a) paper-based medical record system and (b) electronic. The electronic Medical Record has the additional value-related functions (summarization, analysis, and alert processes) to provide clinical decision support for the caregiver team. The EMR clinical database is the focus of this study.

Bayesian Belief Networks

Bayesian Belief Networks (BBNs), also known simply as belief or influence networks, describe a set of probabilistic inter-dependencies between an arbitrary set of random variables. BBNs have been commonly used in bio-medical informatics as predictive models. They are especially effective in prediction when the prior distributions of the explanatory variables are known (Ramoni & Sebastini, 2003).

A BBN consists of a set of nodes that represent variables and a set of directed edges between these nodes that connote conditional dependence relationships between nodes. An example of such a structure, also known as a directed-acyclic graph (DAG) is provided in Figure 5.

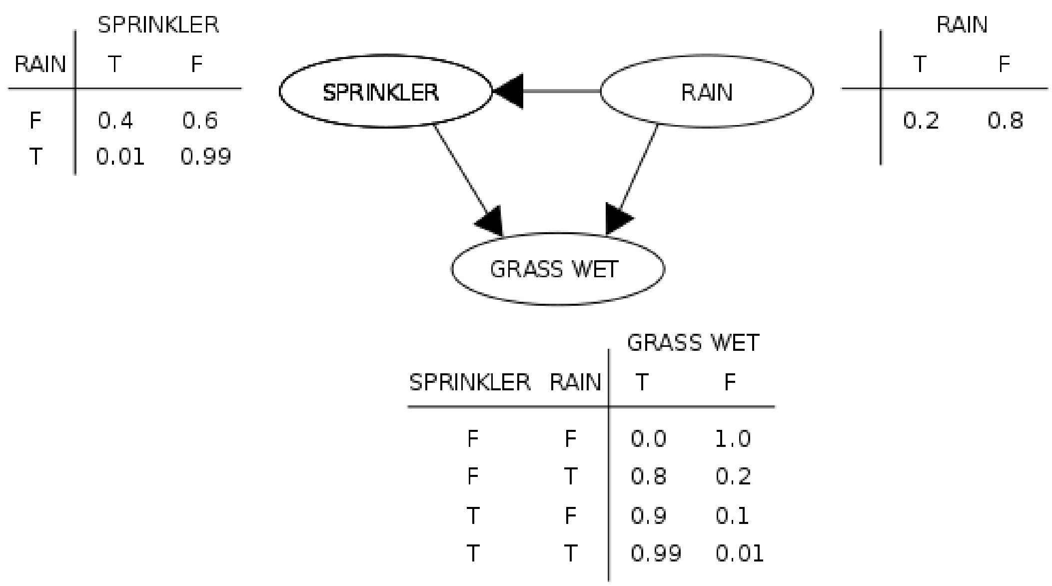


Figure 5: Example of a Bayesian Belief Network (AnAj, 2006)

In the example above, the object-state "Sprinkler ON" is conditionally dependent only upon "Raining Weather" and "Wet Grass" is conditionally dependent upon both "Raining Weather" and "Sprinkler". The conditional probability tables (CPTs) along with the network model can be used together to predict, for example, the likelihood of it raining, the grass being wet and the sprinkler being on, by calculating the joint probability as follows: multiplying the conditional likelihood of each node given its parent, starting at the leaf node and progressing toward the root node. The joint probability for this state of this example network is:

P(Grass Wet=T,Sprinkler=T,Rain=T) = P(Grass Wet=T|Sprinker=T,Rain=T)*P(Sprinkler=T|Rain=T)*P(Rain=T) = 0.002

In addition to analysis, the structure of a Bayesian network can be synthesized from underlying training data and its CPTs can be derived from this training dataset. Once such a network model is learned, it can be used as a predictive model and/or used to analyze the conditional dependence relationships between attributes. This thesis will develop BBNs for both purposes – to determine the key factors influencing transmission of clinical information and validating the model by evaluating their predictive performance.

Clinical Communication Models

The models proposed by this thesis comprise of: (1) an overlying architecture and framework for the study of clinical communications through the clinical database (herein referred to as the OSI Model of Interaction) and (2) an underlying model that describes the essential components of clinical communications (herein referred to as the Interaction Event Model). Both models are based on the theories described in the previous section.

OSI Interaction Model

Agents in any communications system are said to interact with one another when they are exchanging messages with each other. Messages contain content in the form of structured and unstructured information. As a result, knowledge is shared (also known as common ground) when information that is interpreted by the transmitting agent is equivalent to that interpreted by the receiving agent.

This thesis extends this notion of shared information and knowledge in Interaction Theory and applies it to clinical communications. It proposes a layered model for clinical communication, similar to the OSI model used in telecommunications. Figure 6 depicts this layered model, with each layer described from bottom up, as follows.

At the lowest (or physical) layer, the apparatus used to exchange messages is described. For instance, caregiver A transmits a message by writing a nursing note, which is read by caregiver B. The internal processes for reading and writing the note are described within the physical layers of the transmitting and receiving agents. The passage of the note through a message bin considers the note and the bin as components of the physical channel.

At the second (or data-link layer), each message is framed within a start and a stop signal. These signals are asserted by the transmitter and sensed by the receiver to indicate the boundaries of a message. In this way, the integrity of the message can be assessed. In the nursing note example, the starting and ending letters of the note represent the start and stop signals, respectively.

The third (or network layer) defines the path for the messages and includes all intermediary agents involved in the interaction. In this layer, messages may traverse multiple agents and may be transformed during transport. For instance, the nursing note once written by caregiver A may be placed on a message bin, which is later sorted by caregiver C into a message box corresponding to caregiver B. Note that there are two separate channels in this example, the note in the message bin and the note in the message box. Involvement of caregiver C is only to serve as a router of the message, but not as the receiving agent for the message.

The bottom three layers require explicit indication of the channel underlying the communications protocol. As such, information passed across these layers may be subject to channel noise, which can modify the original informational content created by the transmitting agent. The top four layers reside within the individual cognitive apparatus of the agents. As such, they may include corrective mechanisms to overcome channel issues in communication.

The fourth (or transport layer) describes mechanisms to ensure reliability of communication between the transmitting and the receiving agent. These mechanisms may be incorporated through a pre-defined protocol between the agents. These protocols may be as complex as requiring several messages to be passed between agents, or may be more simply defined by inclusion of additional meta-data in the messages. For instance, caregiver A may sign the nursing note before submitting it. Caregiver B may check who signed the note in order to verify its authenticity and who to send a response to. The implicit protocol here is that caregiver A must always sign the note before sending it.

The fifth (or session) layer describes the starting and ending points for the entire interaction. As social interactions may occur either formally or informally, these terminal points for a session may be as explicit as particular events that lead to the series of exchanges and terminated them, or they may be implicitly defined by absolute points in time. When interactions occur in a formal session acknowledged by all agents involved, the session is said to be a communication script. An example of such a script is patient discharge. This session may include a series of messages shared among several members of the caregiver team, including recording observations from laboratory tests, implementing prescription orders, drawing up a discharge summary and instructions, and recording discharge time during release of admission. Note that the starting and ending points of the session are defined by the institutional procedure for discharge that must be recognized by all agents involved.

The sixth (or presentation layer) includes cognitive functions to interpret clinical information presented in the session layer as clinical knowledge. If the interpretations are equivalent between transmitting and receiving agents, the clinical information is said to be understood as shared clinical knowledge; otherwise the information is misunderstood by the transmitting, receiving, or both parties. Note that effects of channel noise in the lower three layers can also contribute to misunderstanding. However, the misinterpretation of clinical information is the realization of this misunderstanding. An example of such a misunderstanding is illegibility of a nurse's progress note. Illegibility may have had its source in either the process of writing the note, reading the note (physical layer), truncation of the note at either end (data-link layer), or deterioration of the note as it was scanned by the digital scanner (network layer). However, the receiving agent may have either interpreted the information correctly or incorrectly depending upon cognitive function. If the agent realizes that the information in the message is flawed, the agent may choose to either discard the information or seek an alternate session for acquiring the information.

The topmost (or application) layer fits in the context of the role fulfilled by the agent to perform his/her service goal. The application layer may introduce a priori expectations of the types of clinical information and knowledge available and comprehensible by the agent. For instance, a respiratory therapist may be engaged in several communication sessions for a variety of patients during the course of a day to communicate knowledge specific to his/her area of expertise. However, he/she may not comprehend or become involved in sessions that are related to physical therapy.

Although this model describes a framework to study all aspects of clinical communications, this thesis is selectively focused on information transmitted across the data-link and physical layers in a clinical database. The next section describes how the model for transmission of clinical information maps to the Interaction Model.

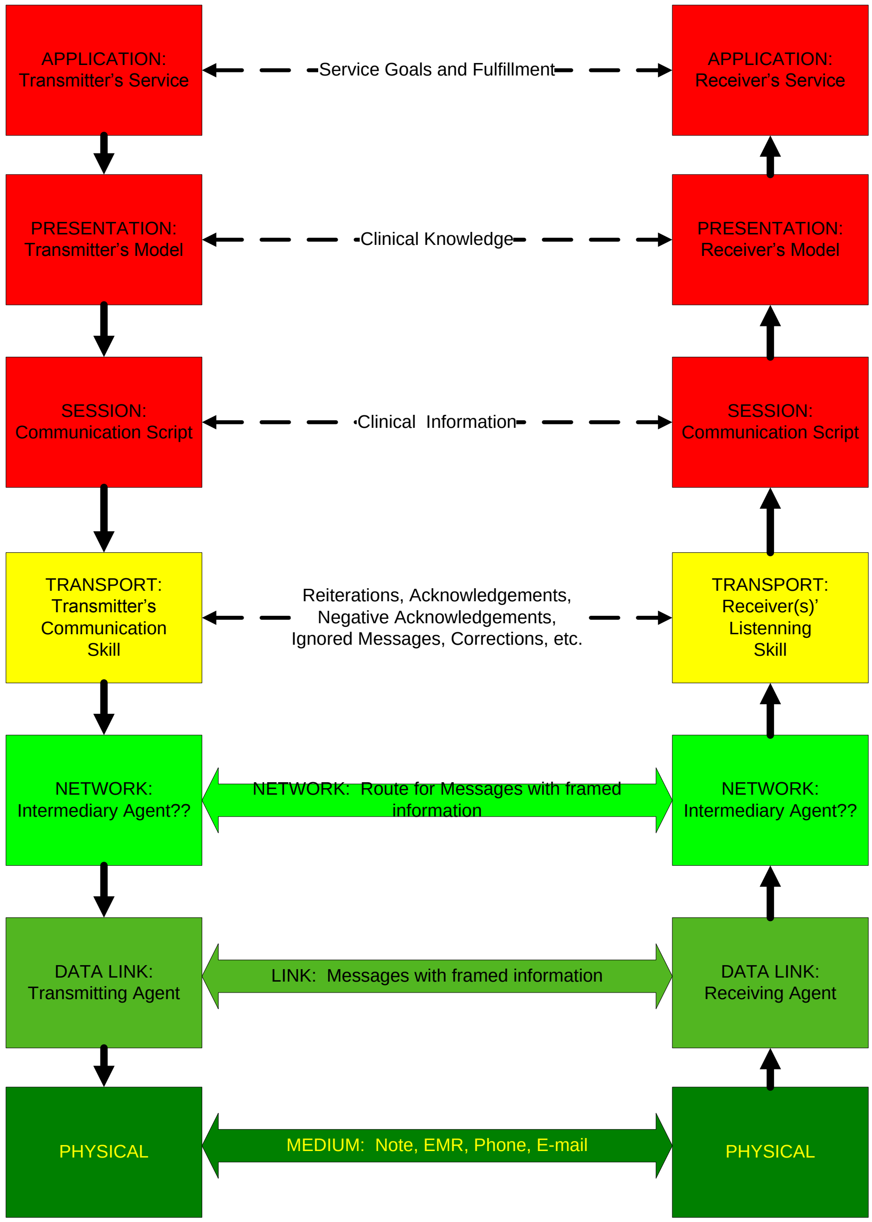


Figure 6: OSI Interaction Model for Clinical Communications

Interaction Event Model

This thesis proposes that the components essential to understanding communications in the clinical database at Layers 1 and 2 of the Interaction Model are the same as those described by Interaction Theory. Since clinical information is not being interpreted at these lower layers and it is assumed that the EMR does not transform the record content, there is no mediation of interaction in this model. Table 1 lists each component of Interaction Theory and its corresponding analog in the clinical communications domain.

Interaction Model Component	**Clinical Database Communications Analog**	**Component Sub-system**
Transmitting Agent	Caregiver	Agent
Receiving Agent	Caregiver	Agent
Channel	Clinical Database	Channel
Communication Modes	Unicast, Multicast, Broadcast	Agent, Channel
Message Content	Record of clinical observations	Message
Message Type	Record Type: one of Note, Chart, Report, Admission, etc.	Message
Message Source	Location where observation recorded	Channel

Table 1: Mapping of Clinical Database Communications to Interaction Model

Following the mapping described above, the clinical database is transformed into an Interaction Event database, where each record corresponds to an interaction message. Whenever a caregiver records a set of observations, he/she is transmitting a message. Whenever a caregiver views a record in the database, he/she is receiving the corresponding message. Whenever a caregiver modifies (or updates) a record, he/she is receiving the message and transmitting another message in its place (known as "read-modify-write").

Message Format

In order to realize this model described above, the clinical database must include meta-information to support identification and characterization of interaction events. This meta-information is captured within fields included in the message format described in Figure 7.

Msg ID	CGTX	CGRX	Start Time	End Time	Location	Subject ID	Record Type	Record ID

Figure 7: Interaction Event Message Format

Each field within a message is described as follows:

- Msg ID – unique identifier for the message.
- CGTX – unique code identifies the transmitting agent.
- CGRX – unique code identifies the receiving agent(s). Lists of agents can be defined as unique codes for multicast messages. A universal code can be assigned for broadcast messages.
- Location – unique code identifies where the message was transmitted from.

- Subject ID – unique code identifying the patient for which the message pertains. This field is not functionally required, as the record ID may prove sufficient. However, it is provided to facilitate implementation, in case record ID's are duplicated across patients.
- Record Type – identifies the type of record or set of observations. This field is not functionally required, as the record ID may prove sufficient. However, it is provided to facilitate implementation, in case record ID's are duplicated across record types.
- Record ID – pointer to locate the actual content in the message payload. If coded uniquely across all patients and record types, it can eliminate use of the subject ID and record type fields. If not, it can be used in conjunction with these fields to location the set of observations in the message.

If CGTX and CGRX are equivalent, then the message is said to be *self-referential*. If the "Start Time' and "End Time" fields are equivalent, then the message denotes a specific instant in time. The "Location" field pertains to the source of the message, as the model assumes that the location of the destination is irrelevant information. The "Record Type" may also be coded as categories of records (for instance, nursing notes, chart entries, medication lists, etc.).

Messaging Objects and Attributes

The Interaction Event Database provides the basis for analyzing clinical communication. There are two types of analyses that can be performed given this database:

(a) temporal/causal analysis – because messages have associated time-stamps, the temporal relation between messages can be analyzed. For example, the dynamic interaction between a set of caregivers can be analyzed regarding a particular patient case. Abstractions can be hypothesized and tested for these patterns of dynamic interactions, such as communications scripts.

(b) object-instance-based analysis – a specific communication component can be chosen as a reference object and its relationships to other communication objects and their attributes can be examined. For example, a specific patient can be chosen as the reference object and messaging attributes pertinent to that patient can be analyzed.

This thesis will focus on the later type of messaging analysis, as it will explore the key objects and attributes relevant to transmission of clinical information. Study of causal relationships discovered in messaging patterns is deferred for future work.

Object-instance-based analysis requires transforming the messaging information in the Interaction Event Database. This transformation, also known as summarization, derives a set of attributes from the relevant objects. Table 2 lists examples of a few such attributes.

Attribute	Object	Messaging Component
Location ID	Visit	Message Source
Location Type	Visit	Message Source
Caregiver ID	Caregiver	CGTX, CGRX
Caregiver Role	Caregiver	CGTX, CGRX
Caregiver Experience	Caregiver	CGTX, CGRX
Hospital Admission ID	Visit	Record Locator
Patient ID	Patient	Record Locator
Patient Age	Patient	Message Content
Patient Sex	Patient	Message Content
Patient Conditions	Patient	Message Content
Expiration Status	Patient	Message Content
Acuity Score	Patient	Message Content
Record Type	Record	Record Locator
Record ID	Record	Record Locator
Day of Week	Visit	Message Time
Season of Year	Visit	Message Time

Table 2: Example List of Objects and Attributes for Clinical Messaging

The list of objects and attributes specified above are not exhaustive by any means. The attributes derived from these objects may generate additional objects and attributes. The effect of attributes on other attributes and objects on other objects can then be analyzed using data modeling tools, some of which are described in the next section.

Methods and Tools

As stated in previous sections, this thesis proposes to examine patterns of information transmission in a clinical database and derive the key objects and attributes that significantly influence various metrics of message transmission. This section details the clinical database used to implement an Interaction Event Database. It also describes the techniques used to detect patterns of transmission in the database. In addition, it specifies data analysis methods to determine the objects and attributes that have an influence upon these patterns. Figure 8 illustrates the entire process followed by this thesis from data collection to analysis and conclusions.

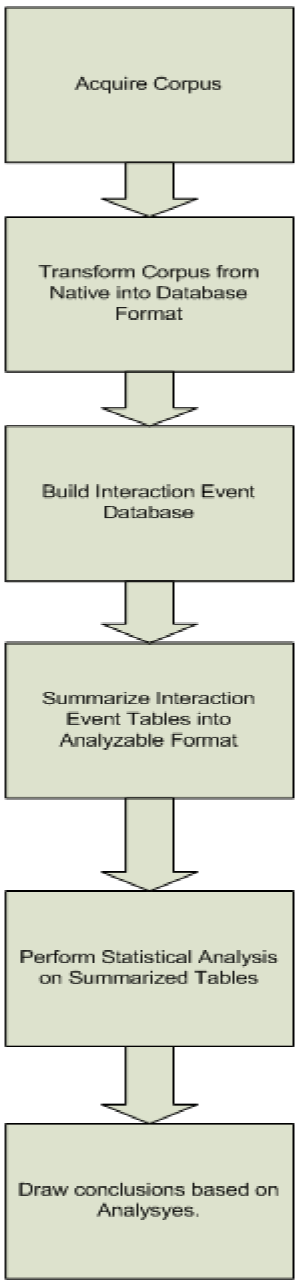


Figure 8: High-level Methodology

Database Description

This study makes use of the MIMIC II (Multi-parameter Intelligent Monitoring in Intensive Care) database, a collection of clinical records collected at the ICU in the Beth Israel Deaconess Medical Center in Boston, Massachusetts (Goldberger, et al., 2000). These records have been made available for bio-informatics research as part of the initiative to advance patient monitoring systems. MIMIC II consists of two separate databases and a collection of dictionaries: (a) a waveform database that consists of acquired discrete signals captured from monitoring instruments, (b) a clinical database that is composed of ICU clinical data, and (c) hospital archival information that includes definitions for codified data in the clinical database (Clifford, Scott, & Villarroel, 2009). This study makes use of the clinical database and hospital archived information only.

Corpus

The MIMIC II clinical database is organized by patient records, where a patient directory consists of files pertaining to a particular patient's records of a given record type. Records in all files had been already de-identified for public use. As such, COUHES authorization was not required to conduct this study. At the time of use, the entire database comprised of records for approximately 26,655 total admitted patients, segregated into 5 batches of approximately 5000 patients each. This study uses only two of these batches for a corpus with data for approximately 10,000 patients.

Observation Types

The corpus consists of the following types of observations for each patient. All observations documented for a particular type are stored in the same file. All files pertaining to a patient are stored in the same directory, with name assigned the patient identifier.

- Additives – record of additives combined with medication to administer to the patient (for example, IV drip) during the patient visit.
- Admissions – record of when patient was admitted to hospital and discharged and expiration status at time of discharge.
- Chart Events – record of charted observations.
- IO Events – record of items given taken in (such as fluids) and/or output from the patient (such as urine, stool, and blood). This record type includes additives and deliveries.
- Medication Events – record of all medications given during the patient visit.
- Census Events – record of entry and exit events in all care-unit locations visited by the patient.
- Deliveries – record of items administered to the patient during visit (e.g., IV infusion).
- Patient Demographic Info – age, sex, and expiration status.
- ICD-9 code – patient condition codes as per standard.
- Note Events – nurse's progress notes, discharge summaries.
- Report Events – X-ray, EKG, EEG reports.
- Total Fluid Balance Events – record of all fluids taken or given from/to the patient.

Dictionaries

The following dictionaries are available for all coded data.

- Caregiver – unique identifier for every caregiver referenced in the records, along with title (e.g., "RN", "MD").
- Care-units – unique identifier for every care-unit location referenced in the records, along with location type; e.g., CCU, SICU, MICU, T-SICU.
- Chart items – unique identifier for every type of chart event observed with text label and category; e.g., SAPS-I score recorded by Laboratory of Computational Physiology (LCP), T3 by Chemistry.
- I/O items – unique identifier for every type of item taken out or given to patient, along with text description and category; e.g., CO Fluid as IV Infusions, Urine out.
- Medication items – unique identifier for every type of medication administered, along with text description and category; e.g., Vivonex.

The following dictionary was constructed manually:

- Caregiver Roles – association of caregiver title to experience and role, along with text description. Each unique title from the caregiver dictionary was entered into this dictionary. For each entry, several caregivers with the same title were searched through the notes and reports records to determine the type of role and experience indicated by the title. For example, title of "MedStu" (a.k.a. medical student) denotes the role of "doctor" with experience as "student" and "LPN" (a.k.a. licensed practicing nurse) denotes the role of "nurse" and experience as "experienced". Appendix C: Caregiver Role Dictionary displays all entries used in this dictionary.

Data Extraction

Figure 9 shows the process used to extract messaging content from the MIMIC II clinical database. The process can be divided into three stages, all of which are described in sub-sections below.

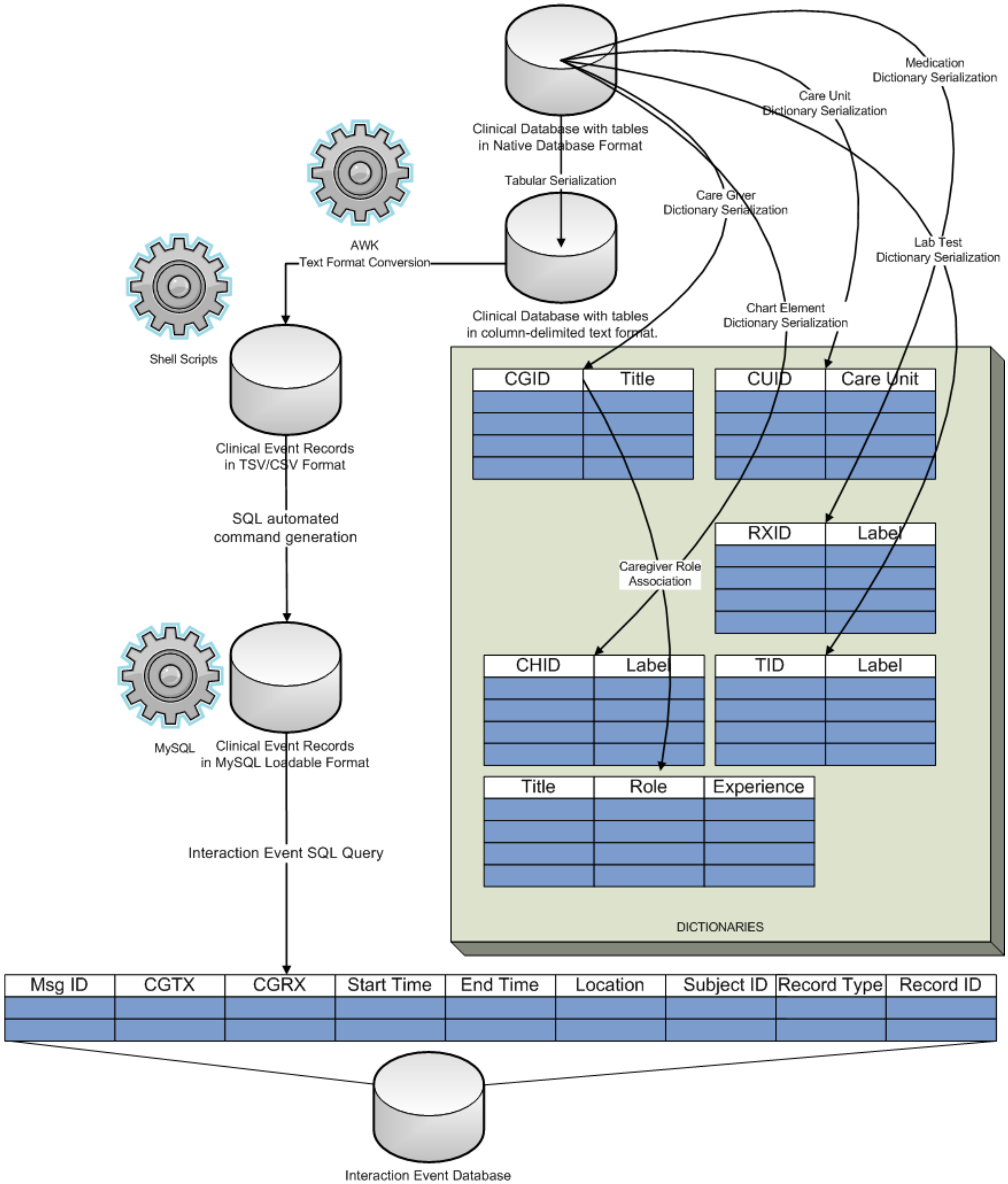


Figure 9: Data Extraction Process

Database Format Conversion

All records and dictionaries of the MIMIC II clinical database are downloaded in tar-gzip compressed format from the physionet.org web-site (Mark, 2009)and installed on a local personal computer. Although the original repository is in Oracle database format, all record files are provided as flat text files in tabular format with columns delimited by four bars ('||||'). The structured records (chart entries, additive list, medication list, I/O events, delivery events, fluid balance events, admission, and patient demographic records are converted easily to tab-separated format (TSV). On the other hand, notes and reports records contained plain-text within the four-bar delimiters and could span several lines. An AWK script embedded in a UNIX bash shell script parses these notes and reports and extracts the plain text as a quoted string in a tab-separated column to allow it to be recognized as a table entry.

Once all the records were converted to TSV format, they were loaded into mySQL database (MySQL, 2009) using an SQL command file, which was generated by a shell script for an entire batch of patient records. The SQL command was then executed within mySQL application to load the database.

Message Extraction

The Interaction Event database was constructed as a "big" Interaction Event Table within mySQL. This table was created through automated generation of mySQL commands. An excerpt of these commands is provided in Appendix D: MySQL Commands to Construct Interaction Event Table. The following assumptions were made in construction of this table:

- MSGID is the primary key in the Interaction Event Database and was automatically incremented for each entry added.
- CGTX was derived from the caregiver entering the observation into the record.
- CGRX (Receiving Caregiver Agent) was assumed to be everyone, denoted as a signed 16-bit value of -1, as the messages are assumed to be broadcast. This assumption had little effect on this study, as it is focused on transmission only.
- StartTime was extracted as the start time of the observation. In cases, where there is only a single timestamp provided for the observation, the StartTime is equivalent to the EndTime.
- EndTime was extracted as the completion time of the observation. In cases, where there is only a single timestamp provided for the observation, the EndTime is equivalent to the StartTime.
- Location was derived from the care-unit provided where the observation was made.
- Subject ID was extracted directly from the patient identifier included in the observation.
- Record Type was assigned to the type of record being extracted.
- Record ID was extracted from the item identifier, which was provided for every record type.

The SQL command file for patients within each batch of records was automatically generated using a shell script to extract the messaging content from the mySQL database. This command file was executed within mySQL to perform the construction of the Interaction Event Table.

Summarization of Tables

In order to perform object-attribute relation analysis within practical limits of computational time and memory capacity, the “big” Interaction Event Table was reduced (or “summarized”) to a set of attributes and transmission metrics, as described in Figure 10.

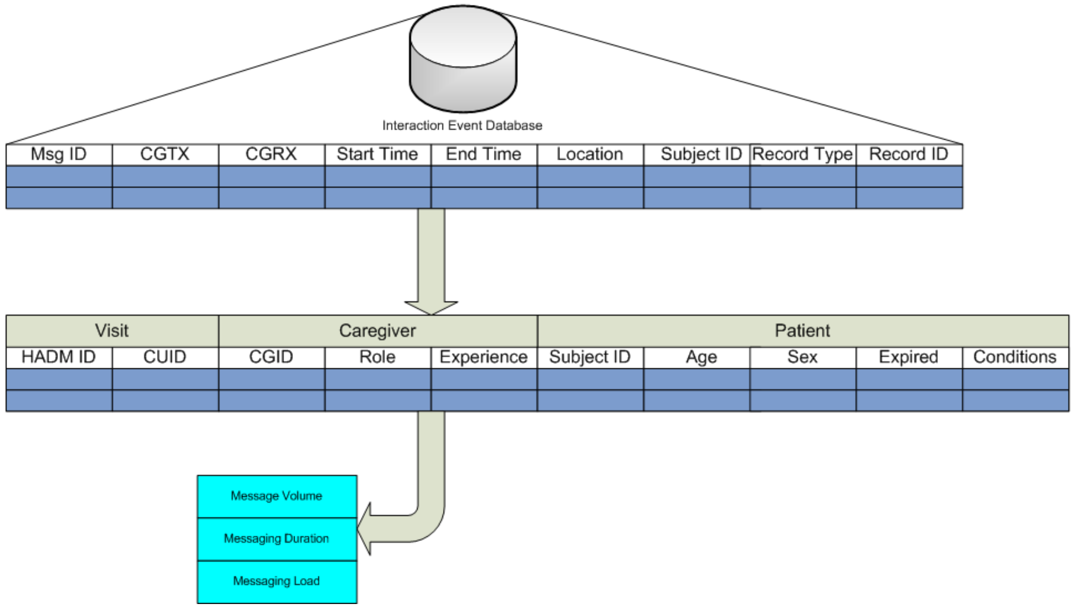


Figure 10: Summarization Process

The objects that were hypothesized to most influence message transmission patterns are:

- Patient – certain characteristics of the patient may affect transmission of clinical information, such as clinical conditions, acuity scores, age, gender, etc.
- Hospital Visit – circumstances of the hospital stay, including care-unit location visited by the patient, and total length of stay may provide additional context for transmission of information.
- Caregiver – who is involved in the care of the patient (and their role and experience) may also affect the transmission of clinical information.

The following metrics for measuring various aspects of message transmission are proposed to be:

- Message Volume – a count of the number of messages in the attribute relation, measures the level of transmission across these set of attributes.
- Messaging Duration – time duration between the first and last message in the attribute relation, measures the transmission period across the set of attributes.
- Messaging Load – message volume per unit time, the average rate of message transmission (or transmission flow rate)calculated as the number of messages counted during the time interval between the first and last message transmitted for the attribute relation during the patient visit (which approximates to the length of stay). . This method of load calculation tries to achieve close to an overall expected value for the message load during a patient visit.

The object features in the summarization table described here serve as a starting point for feature generation and selection, described in the next section.

Data Preparation

The summarized tables produced by the data extraction process are loaded as object relational structures into the 'R' statistical application (Hornik, 1998) in preparation for data analysis.

Cleaning Process

The statistical and graphical tools of the built-in 'R' packages are used to visualize the "raw" output of the data extraction process. Table 3 summarizes the characteristics of the data-set and the figures that follow show the results of this visual analysis for the key object-attributes in the summarization table.

Object/ Attribute	"Uncleaned" Data	"Cleaned" Data (without SAPS I)	"Cleaned" Data (with SAPS I)	Content Description
Patients	4814	4608	3159	Age < 90
Newborns	1223	1215	144	Unknown age removed
Non-Newborns	3591	3393	3015	
Admissions	4814	4608	3159	One-to-one with patients
Caregivers	2541	2496	1835	Unknown removed
Caregiver Roles	10	9	9	Doctor, Nurse, Respiratory Technician, Administrator, Social Worker, Pharmacist, Patient Care, Rehab, Associate
Caregiver Experiences	5	4	4	Experienced, Senior, Junior, Student
Locations	18	16	11	From Care-unit Dictionary
Location Types	11	11	7	SICU, NICU,MICU,CCU,...
Conditions	2968	2913	2527	ICD-9 codes
Genders	2	2	2	Male/Female
Record Types	8	7	7	Report Events removed, Nurses Notes, Del., Meds, I/O, chart entries, fluid-bal, additives
Expired	2	2	2	Yes/No
Days of Week	7	7	7	Calculated from Admission Time
Seasons of Year	4	4	4	Categorized By Date

Table 3: Characteristics of dataset before and after "cleaning process".

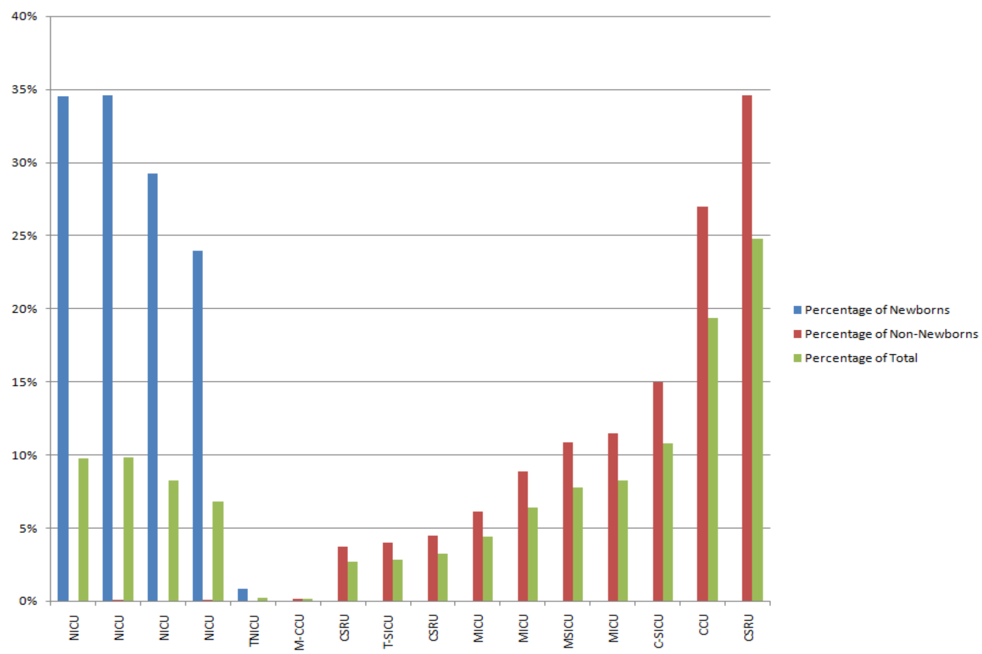


Figure 11: Patient distribution by age segmentation (newborns vs. non-newborns) and Location[1]

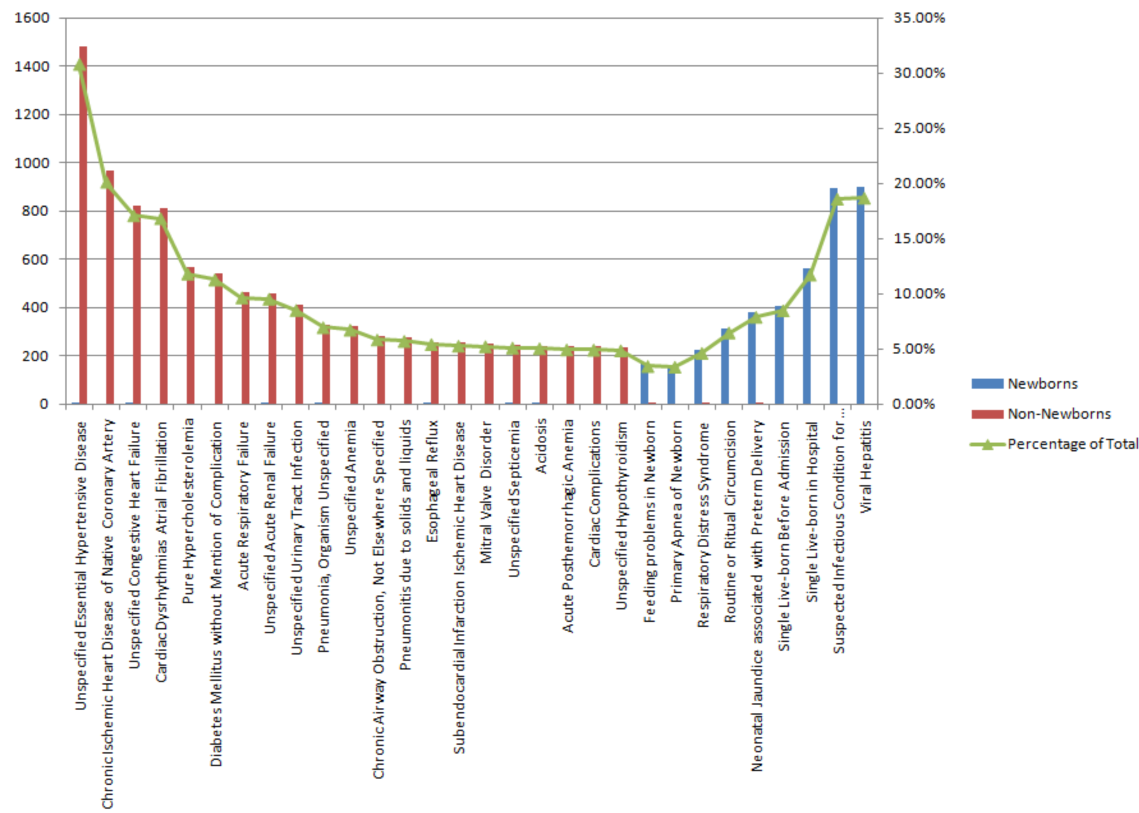


Figure 12: Patient distribution by age (newborns vs. non-newborns) and clinical condition (ICD-9 code).[1]

[1] Ordered by number of newborns from left to right and by adults from right to left

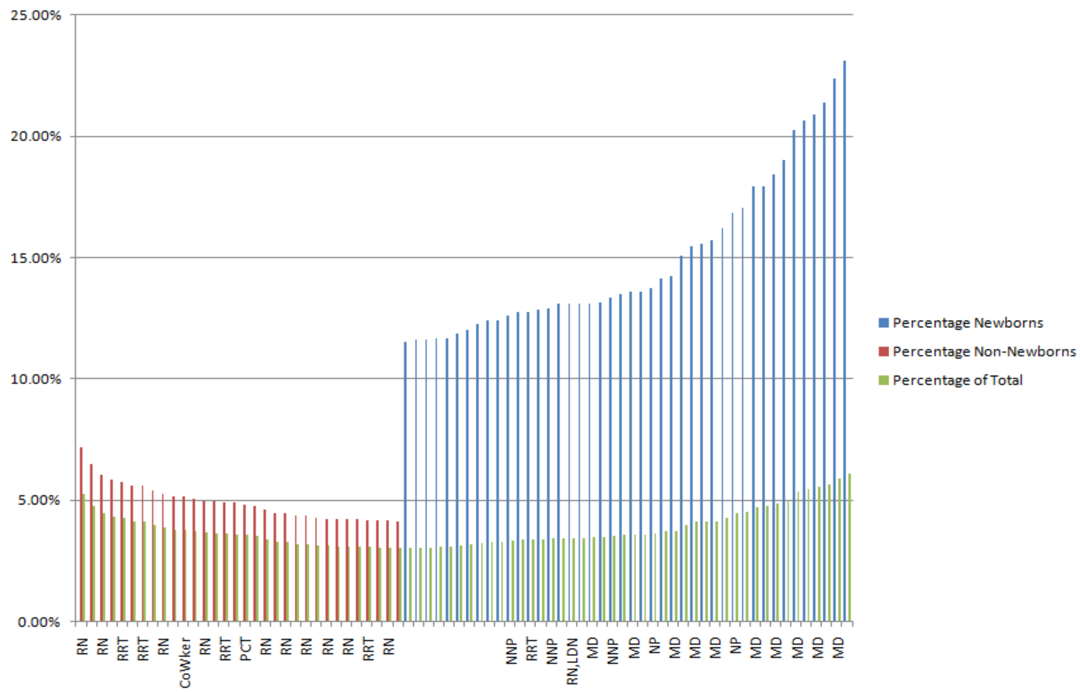


Figure 13: Patient distribution by caregiver involvement.[2]

A sharp distinction between two sub-populations of patients in the ICU database – newborns and adults – can be drawn from Figure 11 to Figure 13. Newborns are seen predominantly by specialized nurses and doctors in the NICUs (neo-natal intensive care units) for a vastly different set of clinical conditions than adults. As a result, the newborn population was cleaned from the dataset, leaving only the adult population as basis for analysis.

Examination of the dataset and dictionary tables revealed that not all caregivers were assigned titles and not all care-units were assigned to records. As such, the role and experience of these caregivers and locations where these messages were transmitted from could not be determined. As a result, these entries were scrubbed from the summary tables. In addition, a special caregiver title and care-unit location designated "LCP" (abbreviation for Laboratory for Computational Physiology) was also ascribed for specific sets of records (these records were also removed from the summary tables). This designation was found to be assigned to chart entries corresponding to SAPS I scores, a measure of the acuity level of the patient's conditions. As this metric could have an effect on transmission characteristics, it was included as a feature in the data analysis. However, since not all patients were assigned these scores, only those that were assigned were chosen in the final dataset for analysis.

In summary, the following "cleaning" operations are performed on the summary tables:

- Newborns of age zero were removed.

[2] Ordered by number of newborns from left to right and by adults from right to left.

- Patients with age designated 200 yrs were removed (these patients were found to have been deliberately assigned for de-identification purposes for those above 90 years in age).
- Care-unit and caregiver designated as “LCP” (code 20001) were removed.
- The following caregivers for non-newborns were found to have unknown titles and were removed from the analysis: 4562, 4745, 4522, 4785.
- Records with unassigned care-unit and caregiver identifiers were also removed from analysis.
- Only patient s with recorded SAPS I scores were retained for analysis.

Exploratory Analyses

Following the “cleaning” process, the summarized tables were examined in two ways: (a) histogram (or frequency distribution) analysis and (b) visual analysis, both of which are described below.

Histogram Analysis

The set of features described in Table 3 are first analyzed by examining frequency distributions of adult patients, as illustrated in Figure 14 and Figure 15.

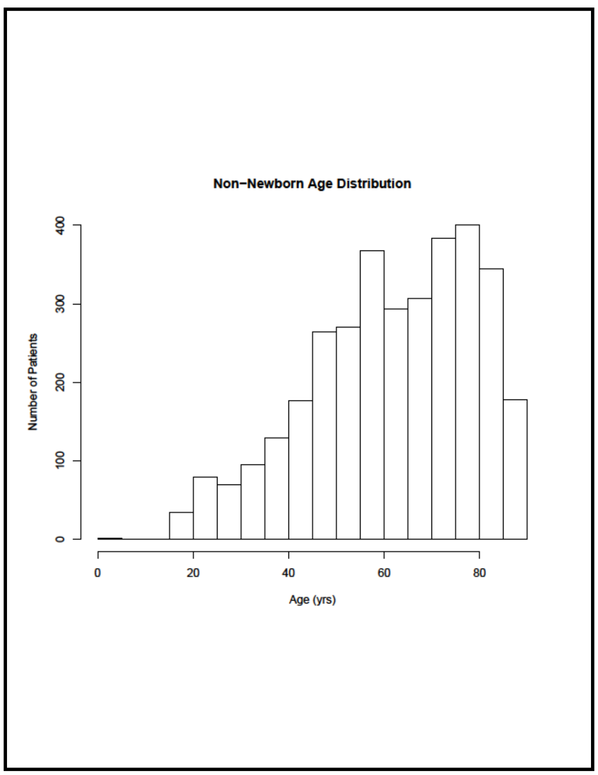


Figure 15: Adult Patient Distribution by Age

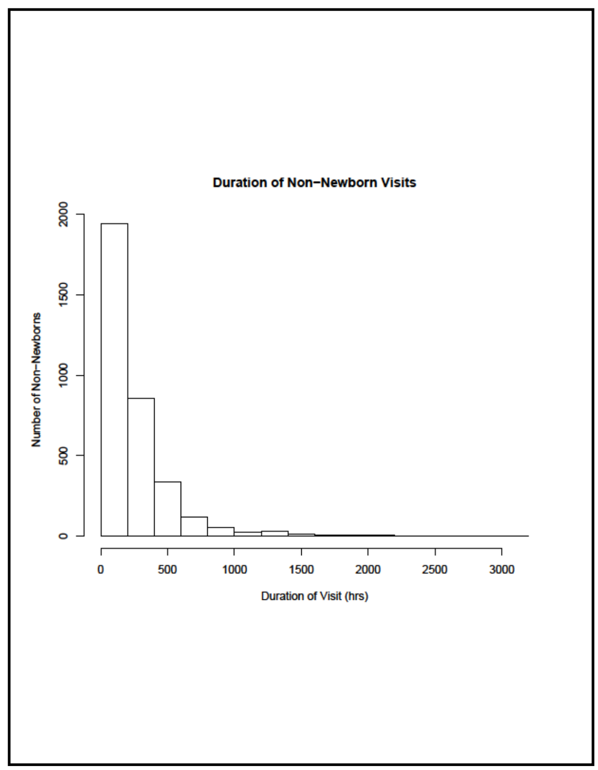
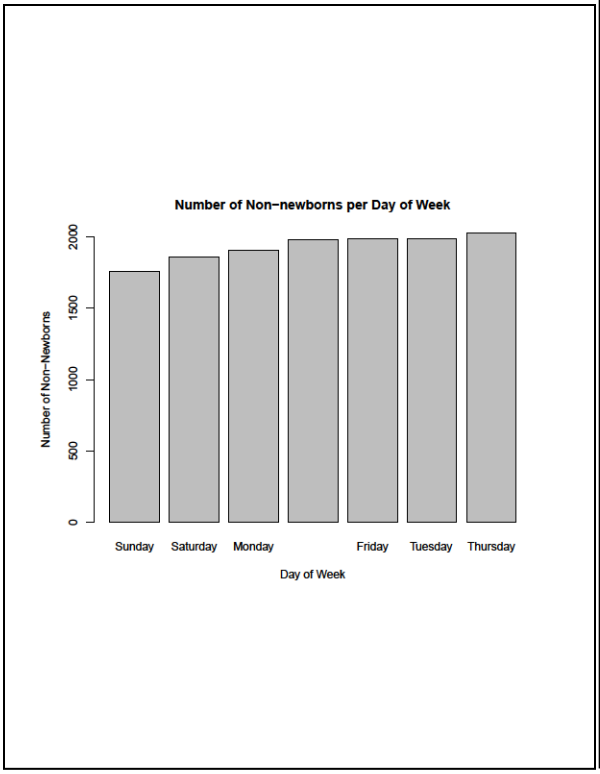


Figure 14: Adult Patient Distribution by Length of Stay

The age distribution among the adult patients in this dataset is negatively skewed toward the senior group and is representative of the age distribution for the entire corpus of 25,852 patients (as specified in the MIMIC II User’s Guide). This suggests that adults visiting the ICU are predominantly the elderly. Whether the age of the patient affects information transmission is unclear from this exploratory analysis and, so, is included as a feature for examination.

The patient's duration of hospital visit (or length-of-stay – LOS) in this dataset is positively skewed toward the lower visit times and representative of the overall corpus. This may be expected as admission for most ICU patients is for acute set of conditions, rather than longer periods of care required for chronic illnesses. A logarithmic transformation is needed to create a more normal distribution for further analysis.



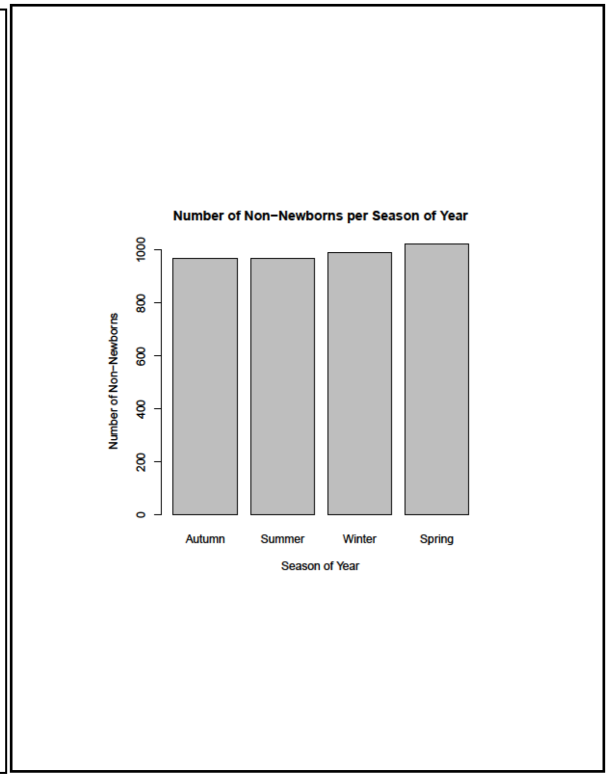


Figure 16: Adult Patient Distribution by Day-of-Week

Figure 17: Adult Patient Distribution by Season-of-Year

The days of the week and seasons of the year were preserved across the date-shift algorithm used to de-identify the database. The distribution of patients in this data-set is slightly higher during mid-week and a bit lower during the week-end, however not substantially. It is worth exploring further whether information transmission is affected by the day-of-the-week, as the patient load appears to be slightly different.

In this dataset, there also appears to be slightly more patients during the winter and spring than in the summer and autumn. Like day of week, it is also worth exploring whether the season has an effect on message transmission due to this difference in patient load.

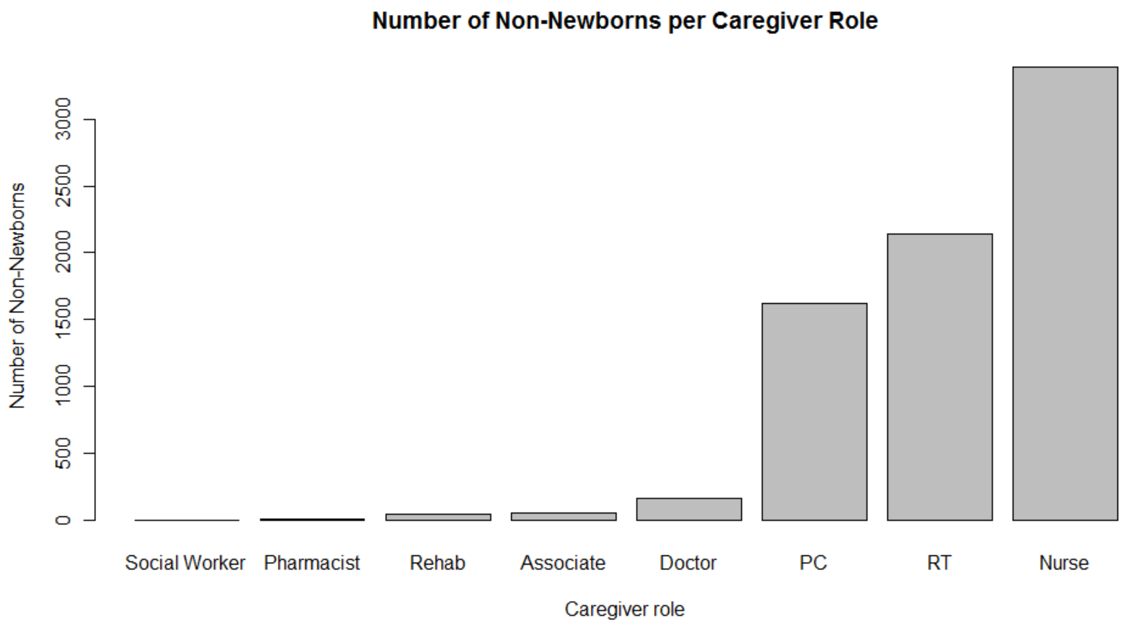


Figure 18: Adult Patient Distribution by Caregiver Role

The number of nurses involved in adult patient care in the ICU is far greater than any of the other caregiver types. Somewhat surprising is the lower ranking of doctors, even lower than respiratory technicians (RT) and patient care associates (PC). This result raises the question: does the nature of the caregivers job function affect transmission? Further analysis for this feature is required to answer this question.

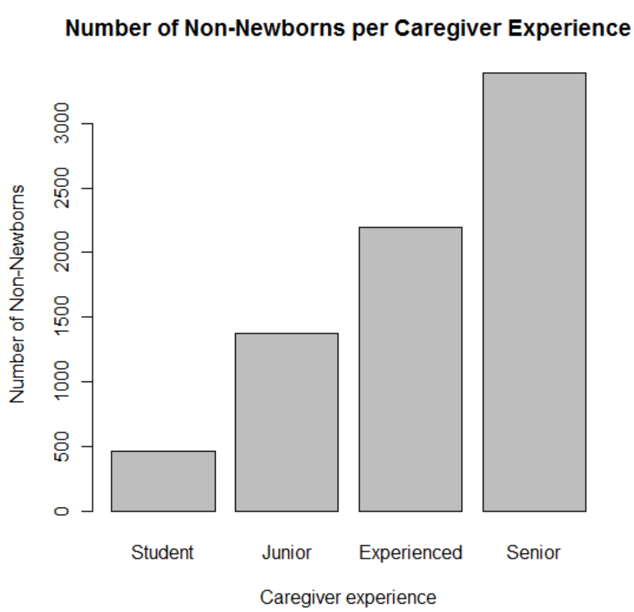


Figure 19: Adult Patient Distribution by Caregiver experience

Adult patients seen by senior and experienced professionals far outnumber those examined by their junior and student counterparts. If the number of patients seen by the more experienced professionals is far greater, then do the amount, time, and rate of information differ for cases in which the students and junior professionals are engaged? To investigate this further, the caregiver experience is retained as a potential explanatory variable.

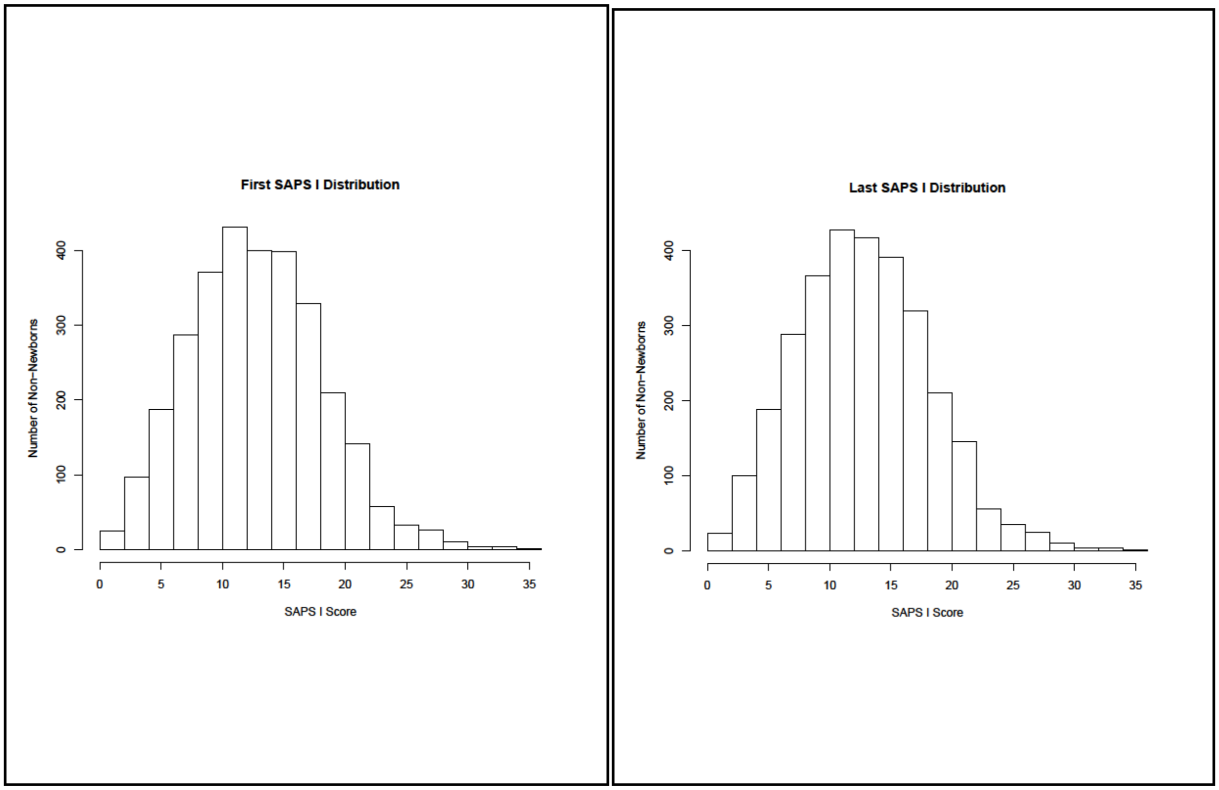


Figure 20 - Figure 21: Patient Distribution according to first and last SAPS I scores taken.

As only the patients with recorded SAPS I scores were selected in the dataset, several patients were recorded more than once, but the vast majority had only one SAPS I score recorded. For the former cases, only the first and last score were taken. For the latter cases, the first and last scores were assigned to the same value. The patient distribution was, for the most part, preserved across first and last, probably due to the bias toward recording of a single score only. However, the overall distribution is also similar. Based on the distribution described in the MIMIC II User's Guide, it is also representative of the overall corpus.

Data Visualization

In addition to histogram analysis, the features in the dataset are also visually examined using the Visualization toolset in the Weka suite of software tools (Hall, et al., 2009). Variables are graphed in three dimensions (2 as vertical and horizontal axes and 1 using color) to search for possible patterns associated with transmission metrics, as illustrated from Figure 22 to Figure 34.

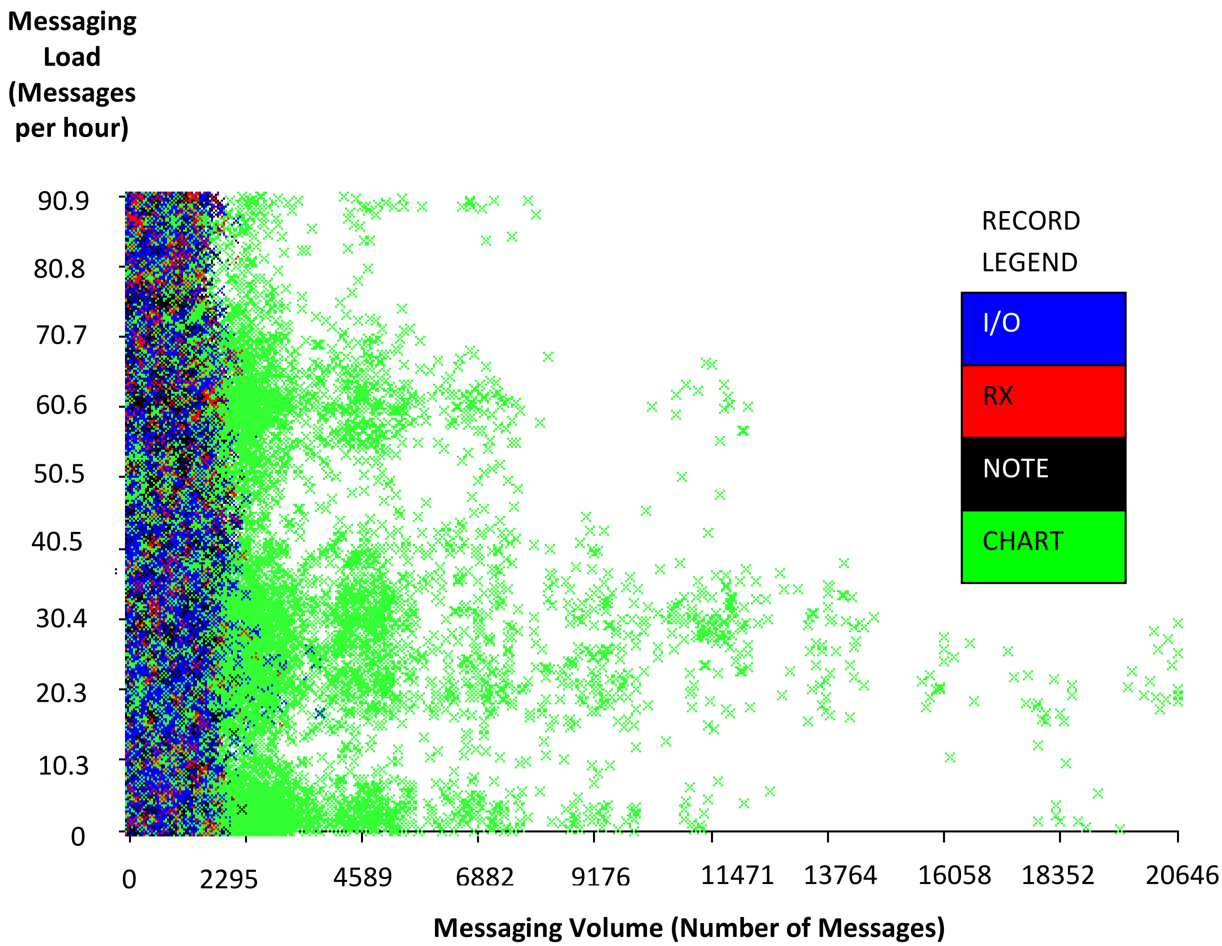


Figure 22: Message Load vs. Message Count by record type.

Figure 22 is a plot of the number of messages recorded per hour against the number of messages, with those samples color-coded pertaining to the following record event categories described as follows: :

- IO [in BLUE] – additive, delivery, I/O, and total fluid balance events.
- Chart [in GREEN] – chart events.
- Rx [in RED] – medication events.
- Note [in BLACK] – note and report events.

As shown in Figure 22, the message load appears to be spread uniformly across all record types. However, at volume above 2200 messages, only the number of chart entries during patient visits is visible. This result indicates that charting load may has a significant influence (and perhaps sole influence) on the overall message volume and load at higher volumes. As such, it should be included as an explanatory variable for further analysis.

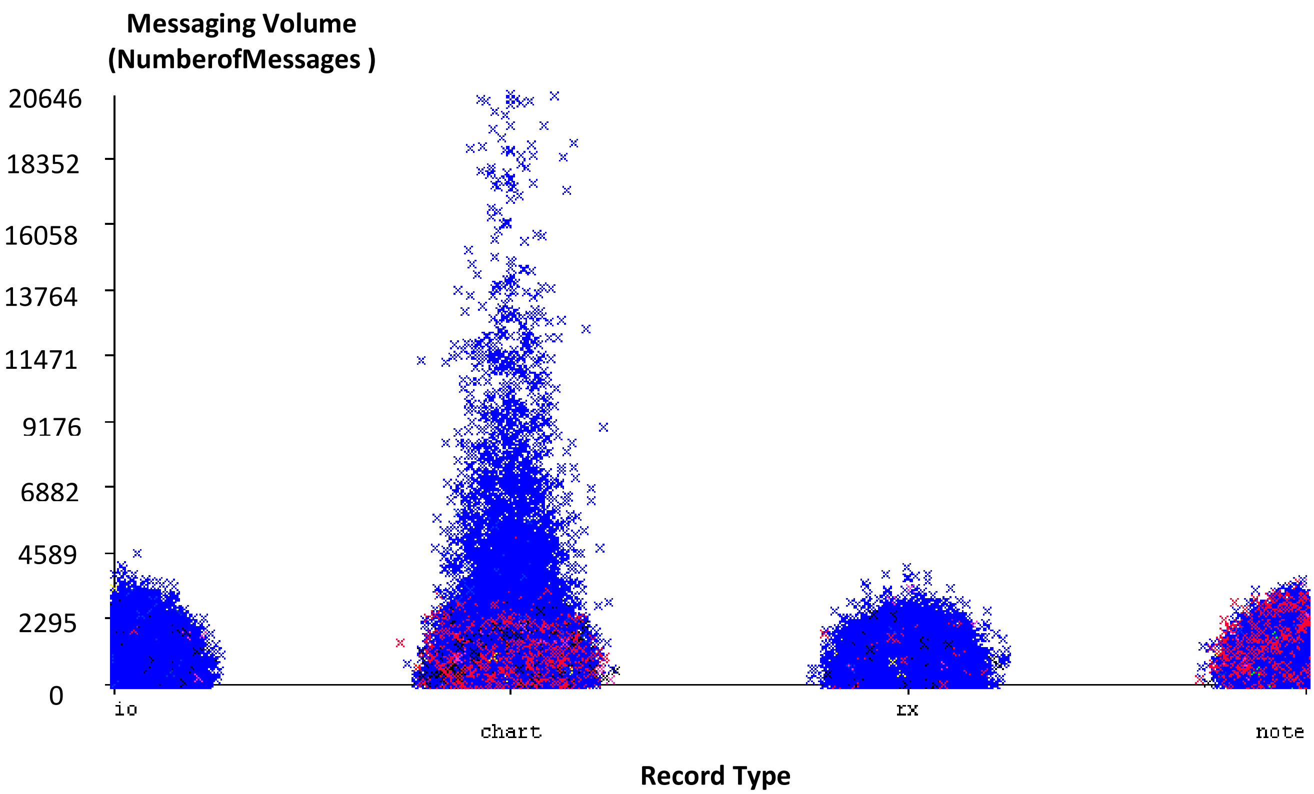


Figure 23: Message Count vs Record Type by Caregiver Title

Figure 23 is a plot of the count of the messages transmitted against the type of record with those samples pertaining to a particular caregiver color-coded according to the legend.. Clearly, the chart entries reach higher messaging volumes, while the other record types do not extend beyond approximately 4000 messages. Plausible explanation for this pattern could be that the amount of charting may be at the discretion of the caregiver, whereas institutional procedural constraints may be limiting the other types of recording. In addition, nurses [in BLUE] are pervasive in transmitting information for all record types, whereas respiratory technicians (RT) [in RED] only record chart entries and notes. Examining the nature of these technicians' notes, it appears that notes titled as "Respiratory Care" notes have been entered, which implies that these entries may be procedurally entered by these professionals, just as physicians enter discharge summaries.. In addition, presence of entries recorded by clinical associates is visible for all types, albeit sporadic. Influence on message transmission at higher levels may be due to the record type as a chart and/or job function as a nurse. Both these features, record type and job function, are retained for further analysis.

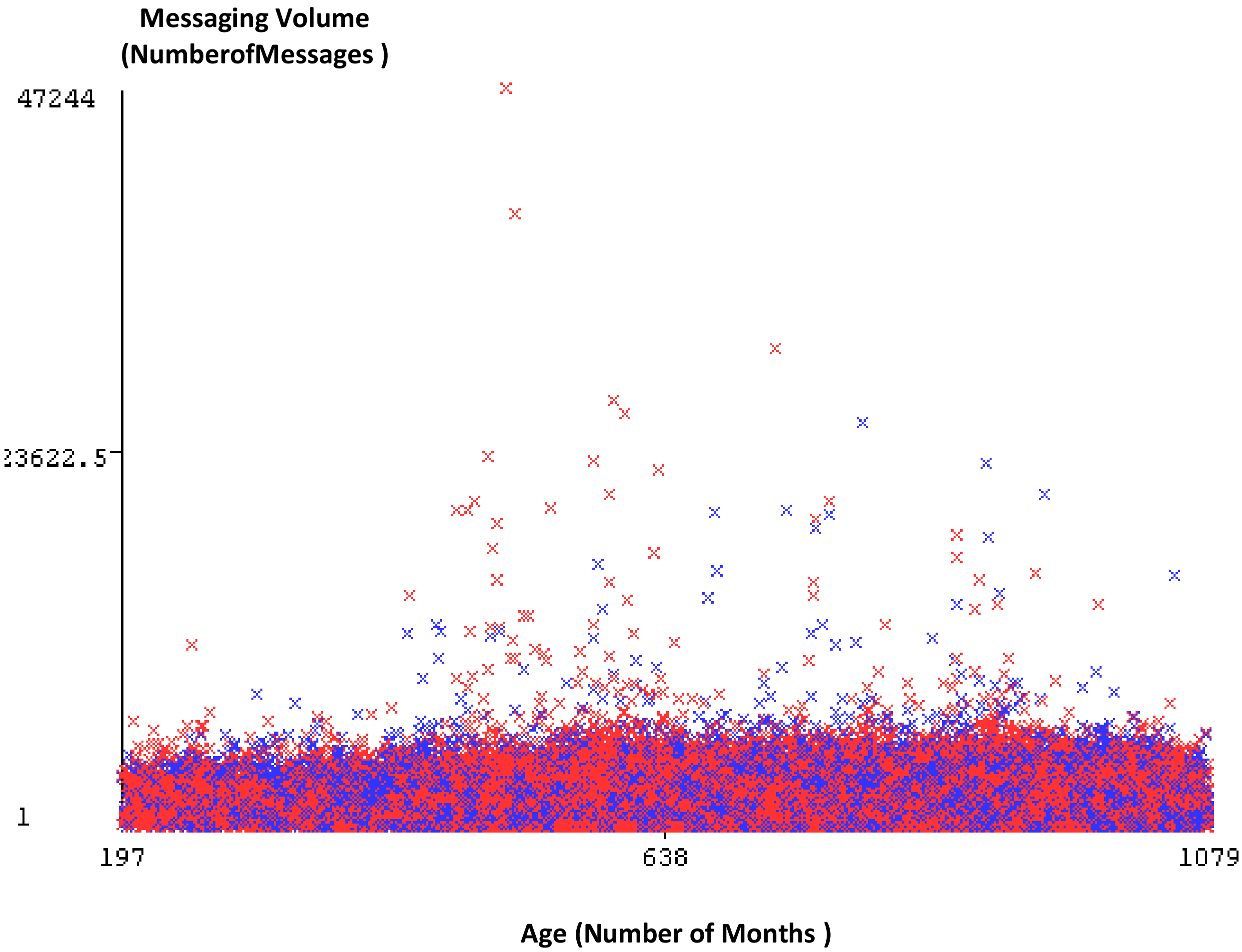


Figure 24: Message Count vs. age by patient gender (Male: RED, Female: BLUE).

Figure 24 is a plot of the number of messages transmitted against the age of the patient. The message volume appears to be uniform across all ages and by gender and restricted to below, suggesting that these features may not have a significant effect on messaging volume. Many of the outliers are in the middle-aged adults between 45 and 60 years in age, which could indicate that certain event(s) may have triggered the onset of extra recording for these patients. Further investigation regarding this pattern is outside the scope of this study.

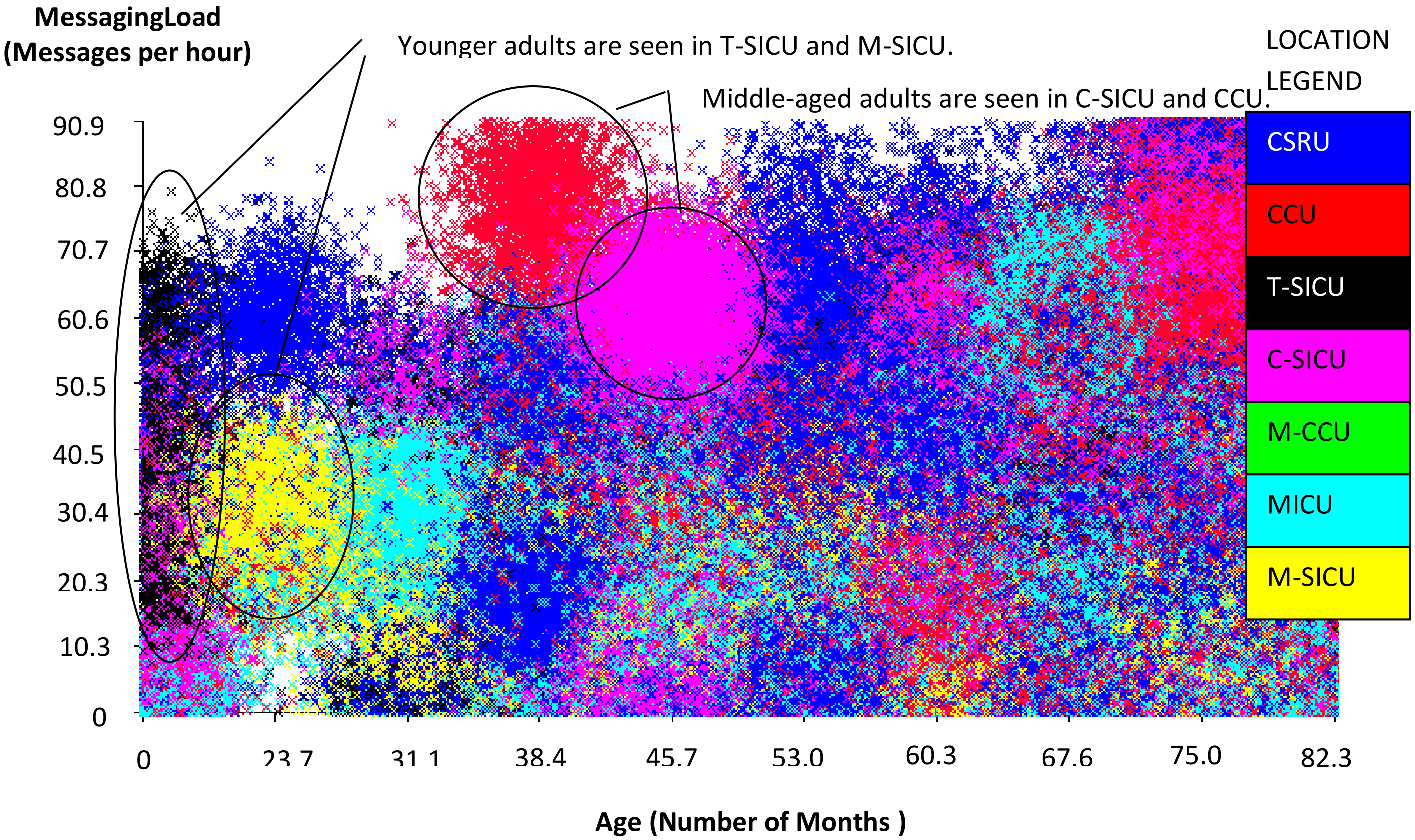


Figure 25: Message Load vs. Age by care-unit location

Figure 25 is a plot of the average number of messages per hour during a patient visit against the age of the patient, color-coded by the care-unit in which the patient was cared for. From the clustering of patients evident in this plot, age and message load appear to vary by type of care-unit location. Trauma patients [in BLACK] appear to be younger in age. On the other hand, the oldest patients tend to visit the cardiac units. There is also a large population of middle-aged adults in their late thirties and forties who are seen in the coronary and cardiac surgical ICUs. These adults appear to receive a lot of attention as the rate at which information is gathered is high compared to patients of other age groups. For further analysis of whether age and care-unit location influence message load, a breakdown of care-unit types is necessary to draw out statistical associations.

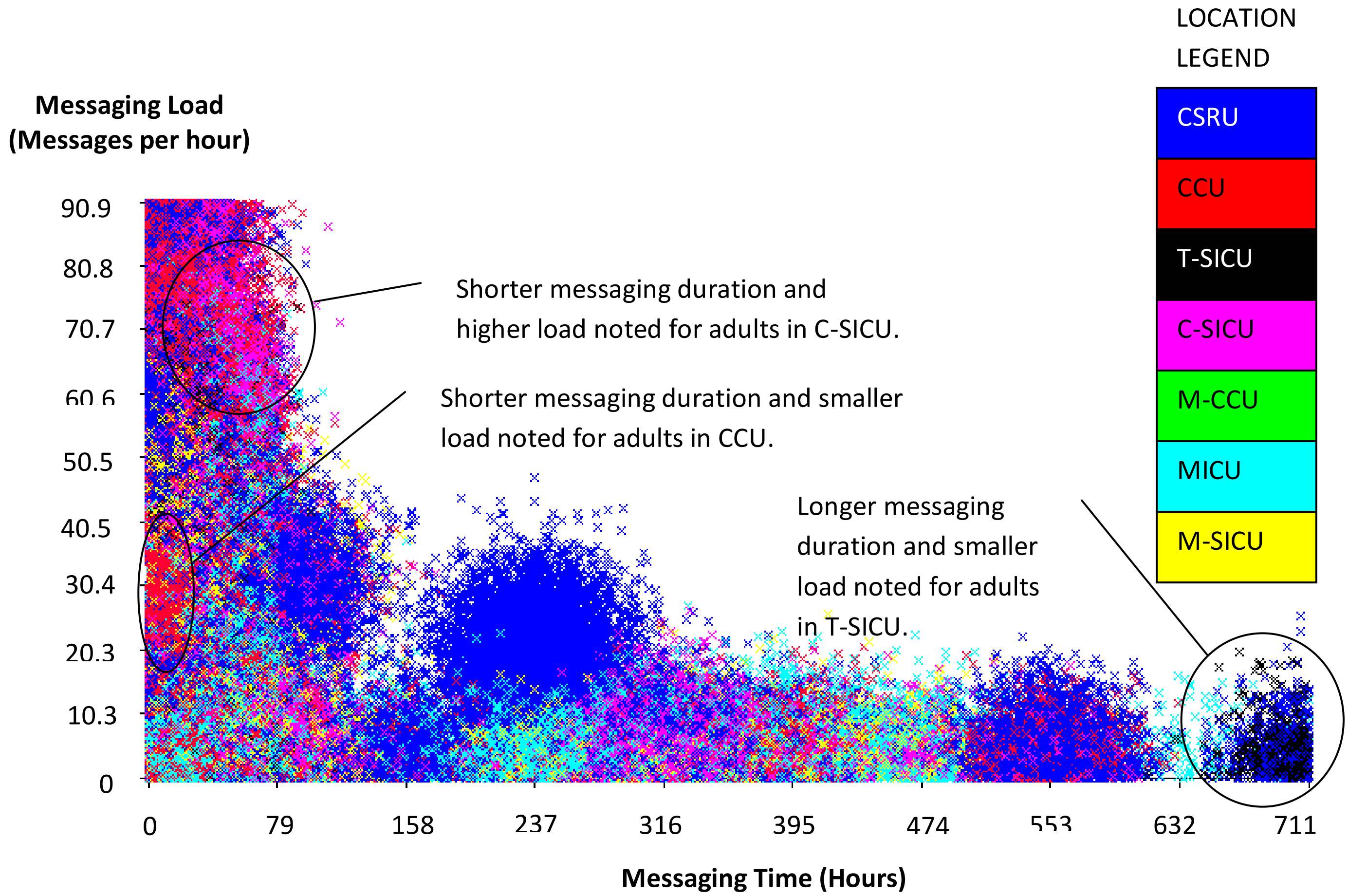


Figure 26: Message Load vs. Messaging Time by care-unit locations.

Figure 26 is a plot of messaging rate against messaging time for adults in various care-unit locations (color-coded). The general trend observed is that messaging time and load appear to be inversely related, which supports the calculation of the load as (= volume / time).

In addition, certain care-unit locations appear to have bias toward messaging duration and others for messaging volume. For instance, messaging load appears lowest and messaging time highest for those adults visiting the T-SICU [in BLACK], which suggests that time may be more influential than volume for these patients. In contrast, messaging duration is much lower for those visiting the CCU [in RED], with messaging load almost the same, which suggests that volume may be the over-riding factor. These results suggest that the care-unit location has some degree of influence on these metrics, and should be retained for further analysis.

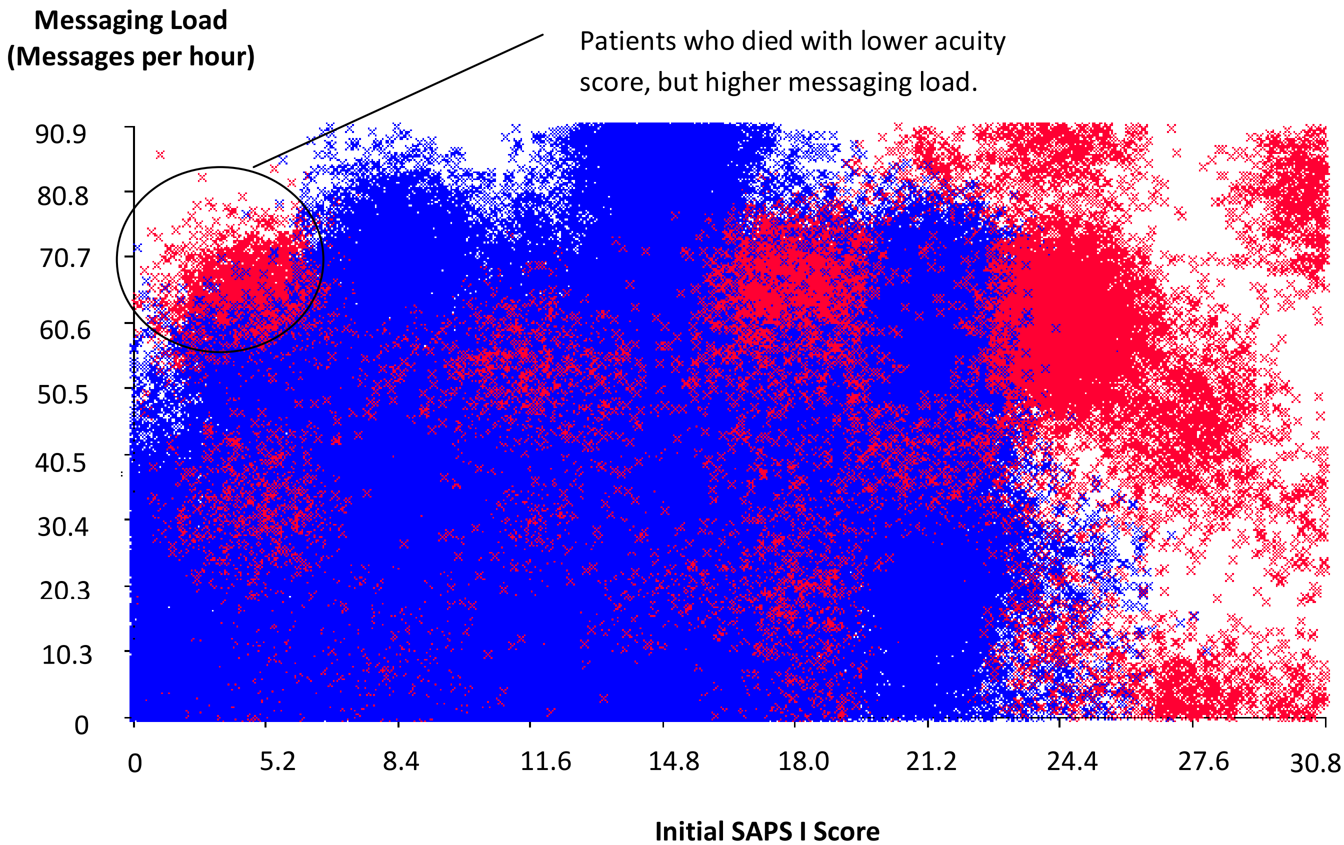


Figure 27: Message Load vs. Initial SAPS I score by Expiration Status (Died: RED, Survived: BLUE)

Figure 27 is a plot of the messaging load against the initial SAPS I score assigned to the patient, color-coded with the mortality status of the patient at the time of discharge. As depicted, the expiration status is clearly associated with higher SAPS I scores. This is expected, as SAPS I measures the acuity level and risk of mortality. With respect to message load, those patients that died and had a higher SAPS I score did not necessarily have high transmission loads, as there are quite a few cases with low and medium message loads on the right side of the plot. Relative to patients with lower SAPS I scores, there does appear to be a concentration of patients that had passed away with higher message loads (top-right of the plot), which suggests that the acuity score may not have been fully indicative of the mortality of these patients, but rate of transmission of information was high. Although this sub-population of patients is not focused upon in this thesis work, it poses an interesting area for further research to characterize this set of patients.

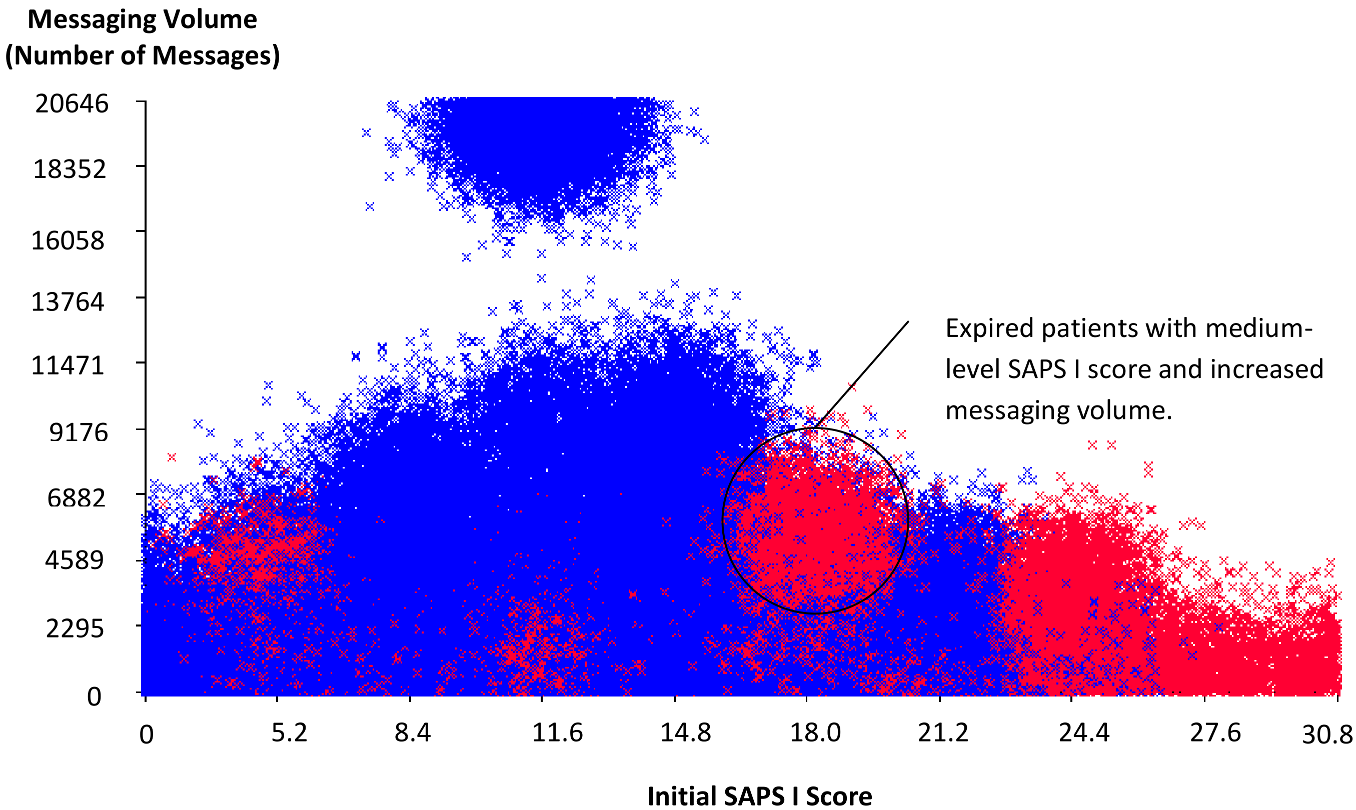


Figure 28: Message Count vs. Initial SAPS I score by Expiration Status (Died: RED, Survived: BLUE)

Figure 28 is a plot of messaging volume against the acuity score of the patient. In contrast to the uniformity of message load at higher SAPS I scores as depicted in Figure 27, messaging volume does appear to be relatively low for these expired patients with high SAPS I scores. This pattern may be attributed to the lower messaging time observed for these patients in Figure 29, as a shorter amount of time would leave little time to gather clinical information for this patients. The messaging volume does appear to be higher for a cluster of patients with mid-ranged SAPS I scores (top-middle of plot). Most, if not all, of these patients appeared to have a positive expiration status. The reason for this pattern is unclear and merits further analysis. Although this sub-population of patients are not investigated in this study, it poses as an opportunity for future research.

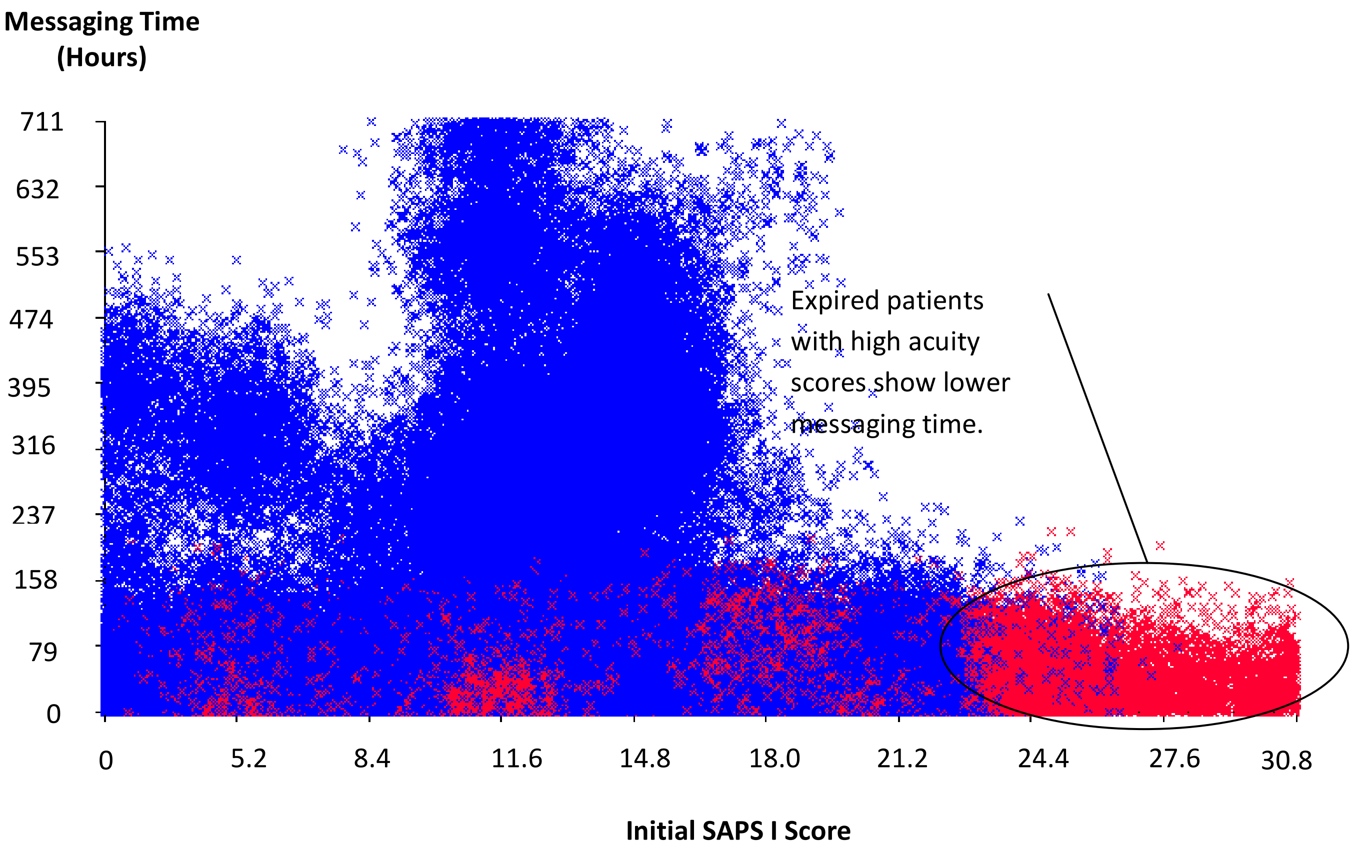


Figure 29: Messaging Time vs. Initial SAPS I score by Expiration Status (Died: RED, Survived: BLUE).

Figure 29 is a plot of the messaging time against patients' acuity levels. Like messaging volume, the messaging time appears to be reduced for patients with higher SAPS I scores (bottom right of plot). This result could be attributed to a variety of reasons, including shorter length of stay for these high-risk patients. Further analysis is necessary to determine if this is the case.

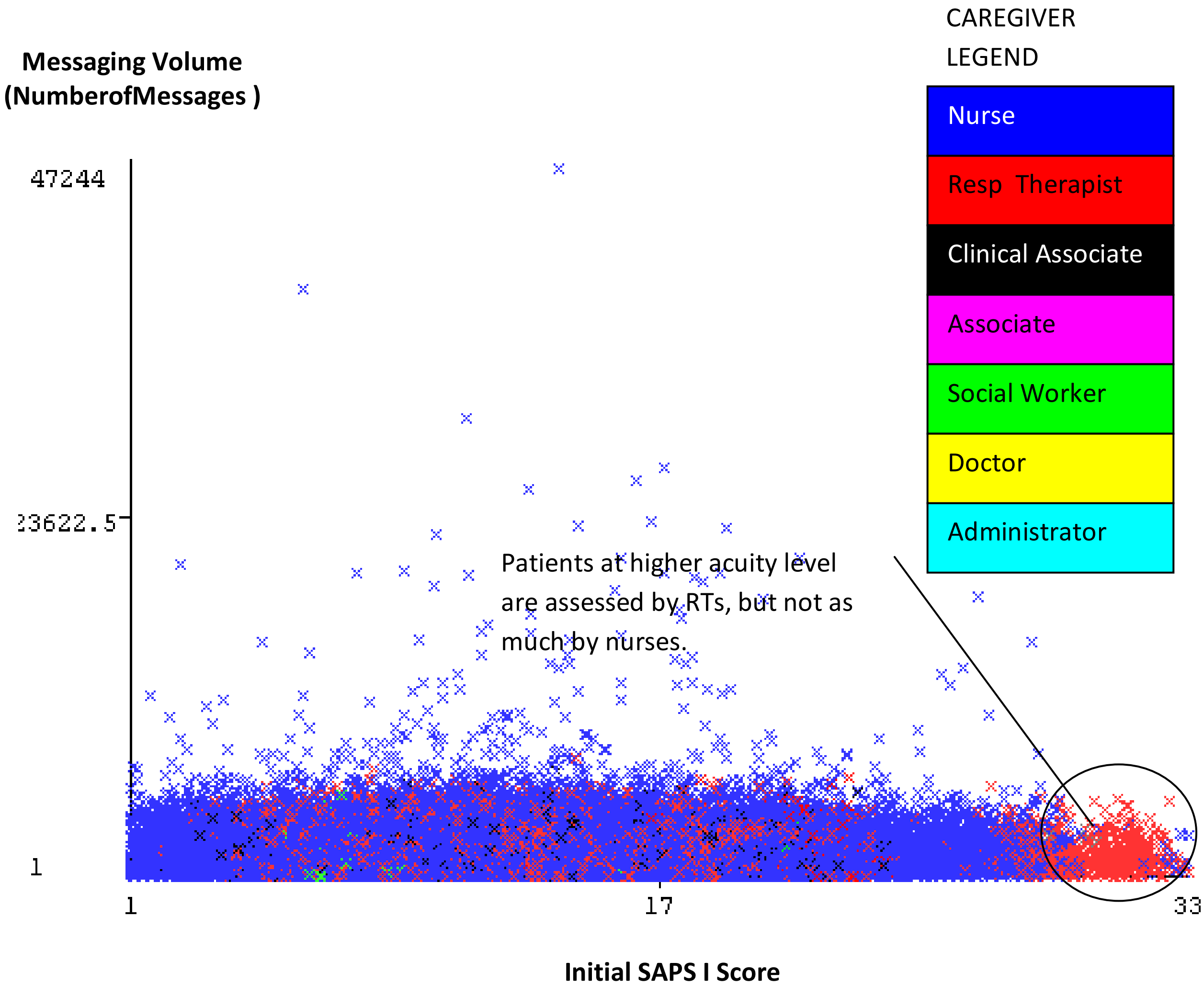


Figure 30: Message Count vs. Initial SAPS I score bycaregiver type.

Figure 30 is a plot of the messaging volume against the initial SAPS I score assigned to the patients, with the designated caregiver illustrated in color. Somewhat surprisingly message count (or volume of transmission) is not affected by the initial SAPS I score, nor the caregiver title, as it appears mostly flat across the entire range of scores (with a slight dip toward the higher end). There does appear to be a predominance of respiratory technicians [in BLUE] for patients with initial SAPS I scores above 30, with reduced transmission from nurses and other caregivers (bottom right of plot). The reason for this is unclear, but merits further analysis. As such, the first and last SAPS I scores are also included as possible explanatory variables.

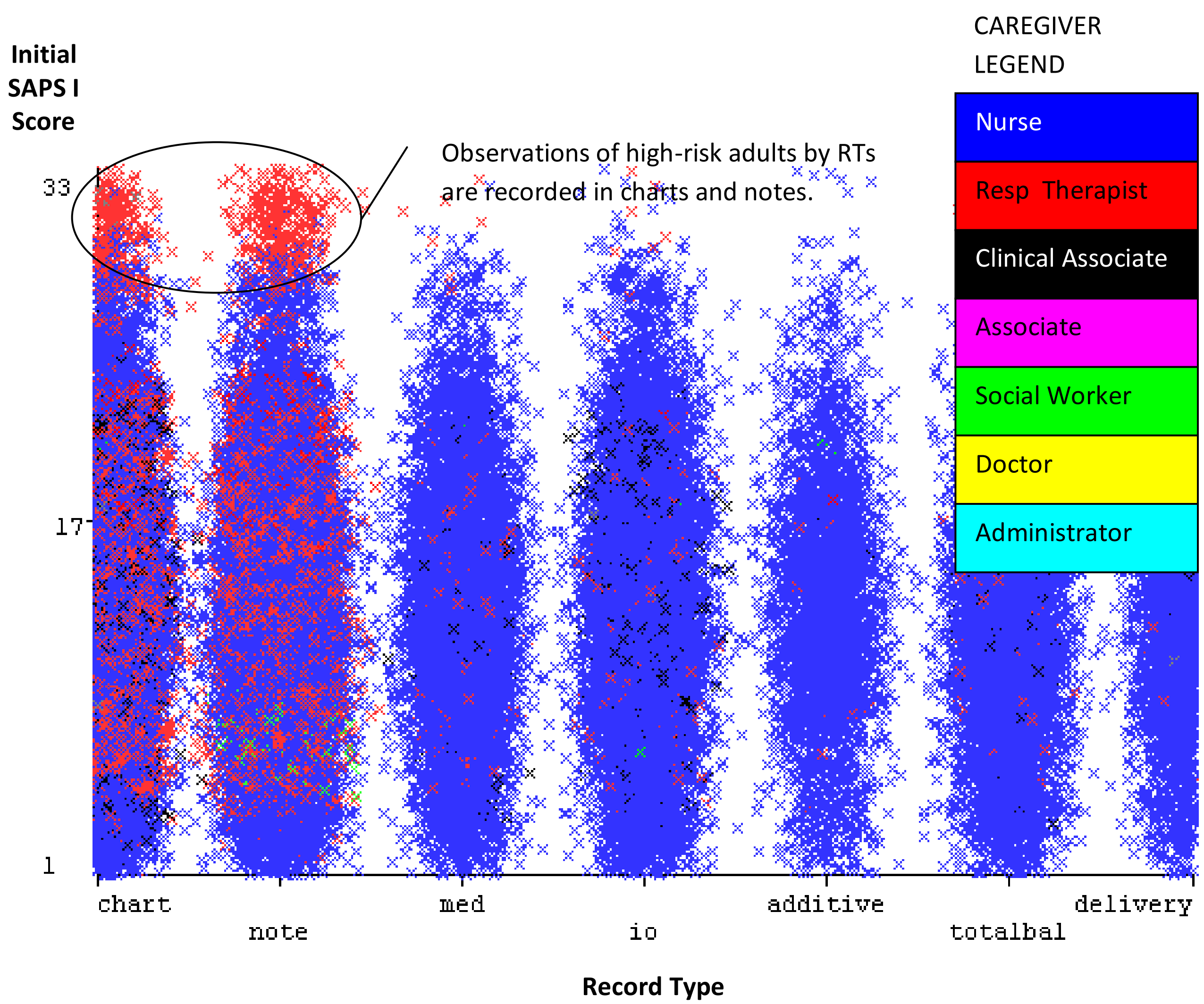


Figure 31: Initial SAPS I score vs. Record type across various caregiver types.

Figure 31 illustrates the combined findings of

Figure 23 and

Figure 30, where respiratory technicians predominantly transmit clinical information through chart entries and respiratory care notes, whereas nurses employ all records without bias toward any specific type, regardless of SAPS I score. In addition, the respiratory technicians are engaged at higher SAPS I scores, more so than other types of caregivers. Clinical care associates transmit information primarily by recording I/O events and chart entries, less so by notes and other forms.

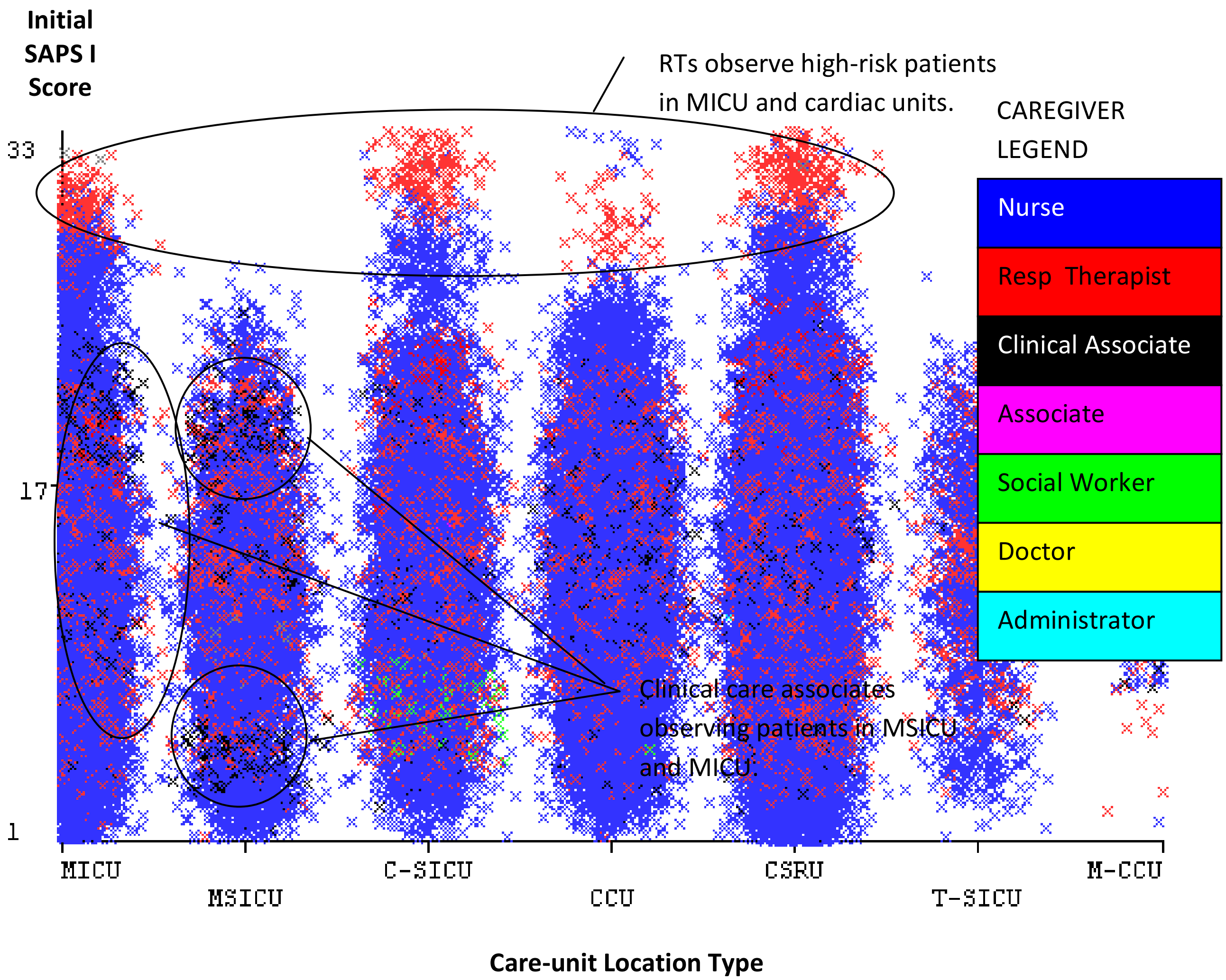


Figure 32: Initial SAPS I score vs. Location Type by caregiver type.

Figure 32 is a plot of the initial SAPS I score assigned to patients against the care-unit in which they are seen, color-coded by the caregiver that observed them. As shown on the right-hand side of the plot, there appears to be a shortage of records for those adult patients visiting the M-CCU. In addition, there appears to be a slightly greater involvement of clinical care associates [in BLACK] in the MSICU and the MICU, as compared to other locations. Also, respiratory technicians are heavily engaged for patients with high SAPS scores in the MICU, C-SICU, CCU, and CSRU, less so in the T-SICU and MSICU. To investigate possible reasons for these patterns, the location type will be included as a feature for further analysis.

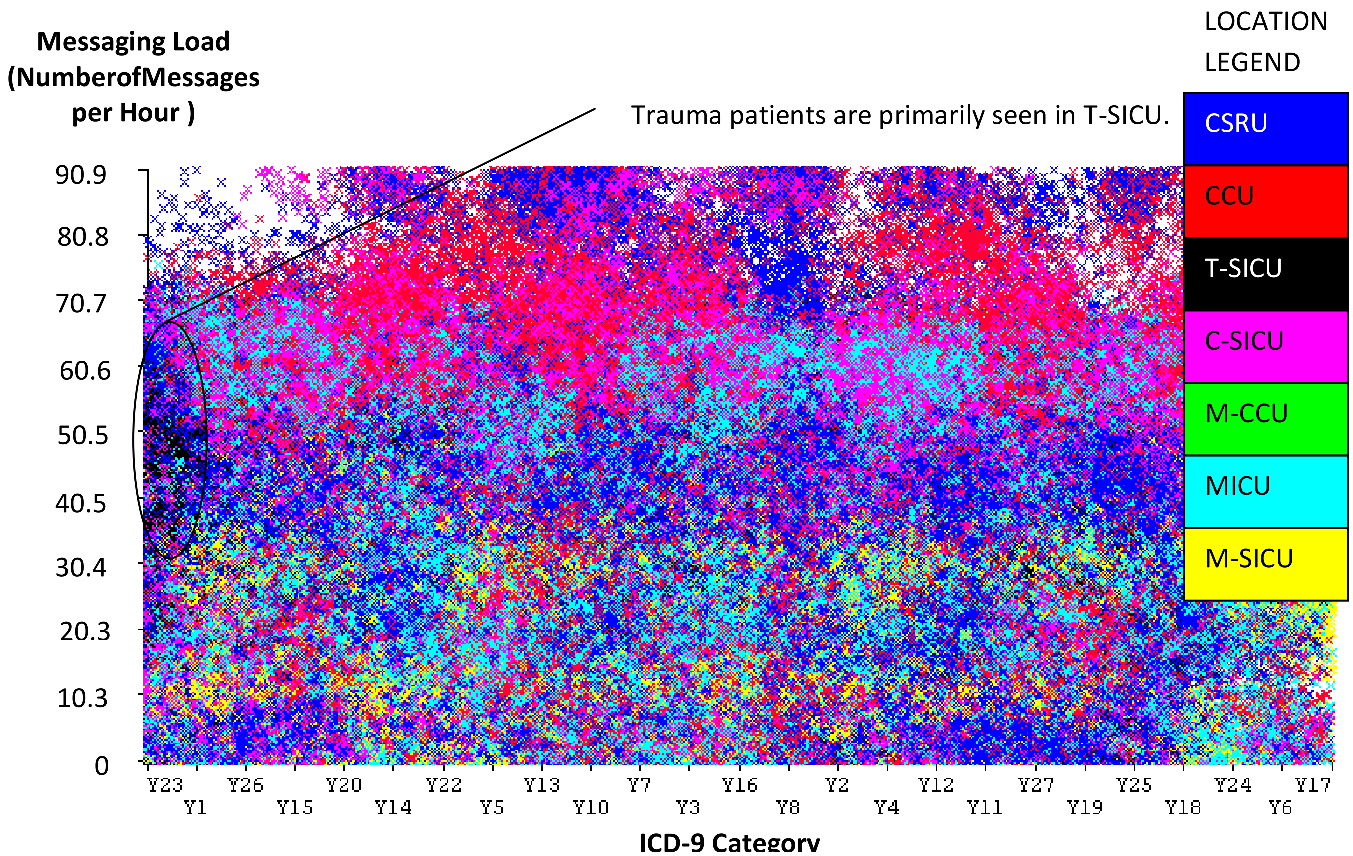


Figure 33: Message Load vs. ICD-9 Category across various care-unit locations

Figure 33 is a plot of the messaging load against the ICD-9 categories to which patients' clinical conditions are assigned. The ICD-9 patient condition codes are categorized using the standard ranges provided in the tabular index (NCHS, 2009). Sub-ranges are included where conditions in these ranges were found to be of more frequent occurrence than the inclusive range. In addition, the most frequent specific clinical conditions (those with a single ICD-9 code describing the condition) are cited in the table within Appendix E: ICD-9 Categories with Codes and Predominant Conditions, which also indicates the label (prefixed with 'Y') assigned to each category.

All code ranges described in Appendix E: ICD-9 Categories with Codes and Predominant Conditions were plotted against message load and color-labeled with location in Figure 33. Recognizable patterns are:

- Trauma patients (those in range Y23) seen in the T-SICU [in BLACK] is neither associated with a higher nor lower message load, tending rather toward the mid-range.
- Cardio-vascular patients (those in ranges Y7 through Y10) seen in the CCU [in RED] and C-SICU [in MAGENTA] appear to have higher associated message transmission loads.

As particular patient conditions appear to have an influence on messaging load, the ICD9 category was included as a "macro-feature" and specific conditions of higher frequency within these ranges (specified in the table in Appendix E: ICD-9 Categories with Codes and Predominant Conditions) were also included as features. The list of all selected features is described in the following section.

Data Analysis

Detailed Description of Selected Features

Following the exploratory analysis, the list of object-attribute relational features was modified from Table 3 to include additional new attributes and eliminate a few insignificant ones. The table in Appendix F: Selection of Features for Messaging describes the entire set of features used for statistical analysis. These features are also described in the following sub-sections by object category.

Patient Attributes

Age

Patients chosen for data analysis are between 15 and 90 years in age. As this attribute has a continuous value between these limits, this variable was converted into discrete values, as described below.

Sex

Patients chosen are either male (value of 'M') or female (value of 'F').

Expiration Status

Patients chosen had either expired (value of "Yes") or not (value of "No") during the admission period.

SAPS I Scores

Most, but not all, patients were assigned a Simplified Acuity Physiology Score (Type I). Some patients were scored several times during hospital admission, but most patients were scored only once. In this later case, the first and last scores have equivalent values. The dataset consists of patients with at least one score assigned. The SAPS I calculate tries to predict the mortality of the patient based on physiological parameters, such as body temperature, white-blood-cell (WBC) count, and blood pressure. The result is an integral value, with higher scores indicating greater risk of mortality. Prior analyses on the entire cohort of patients in this database have shown that the mortality can be greater than 50% for patients with a SAPS score of above 33 (Clifford, Scott, & Villarroel, 2009).

Clinical Conditions

The clinical database includes ICD-9 codes recorded for each patient, which indicate the clinical conditions experienced by the patient. The dataset utilizes these codes by grouping them by range, as described in Appendix E: ICD-9 Categories with Codes and Predominant Conditions, and then selecting the most frequently occurring codes. The list of specific codes includes: diabetes, sepsis, trauma, hepatitis, HIV, hypercholesterolemia, anemia, alcoholism, tobacco use, Alzheimer's, depression, heart failure, hypertension, acute myocardial infarction, atherosclerosis, cardiac dysrythmia, pneumonia, acute respiratory failure, acute renal failure, urinary tract infection, and cardiac complication. Each of these codes is assigned a binary value: 'TRUE' if patient has the condition, 'FALSE' otherwise.

Caregiver Attributes

The clinical database only provides a single piece of information regarding each caregiver; his/her job title abbreviation is stored in a dictionary. Two attributes were generated from this title information: (a) the role of the caregiver and (b) the professional experience of the caregiver. Both assignments are described in Appendix C: Caregiver Role Dictionary.

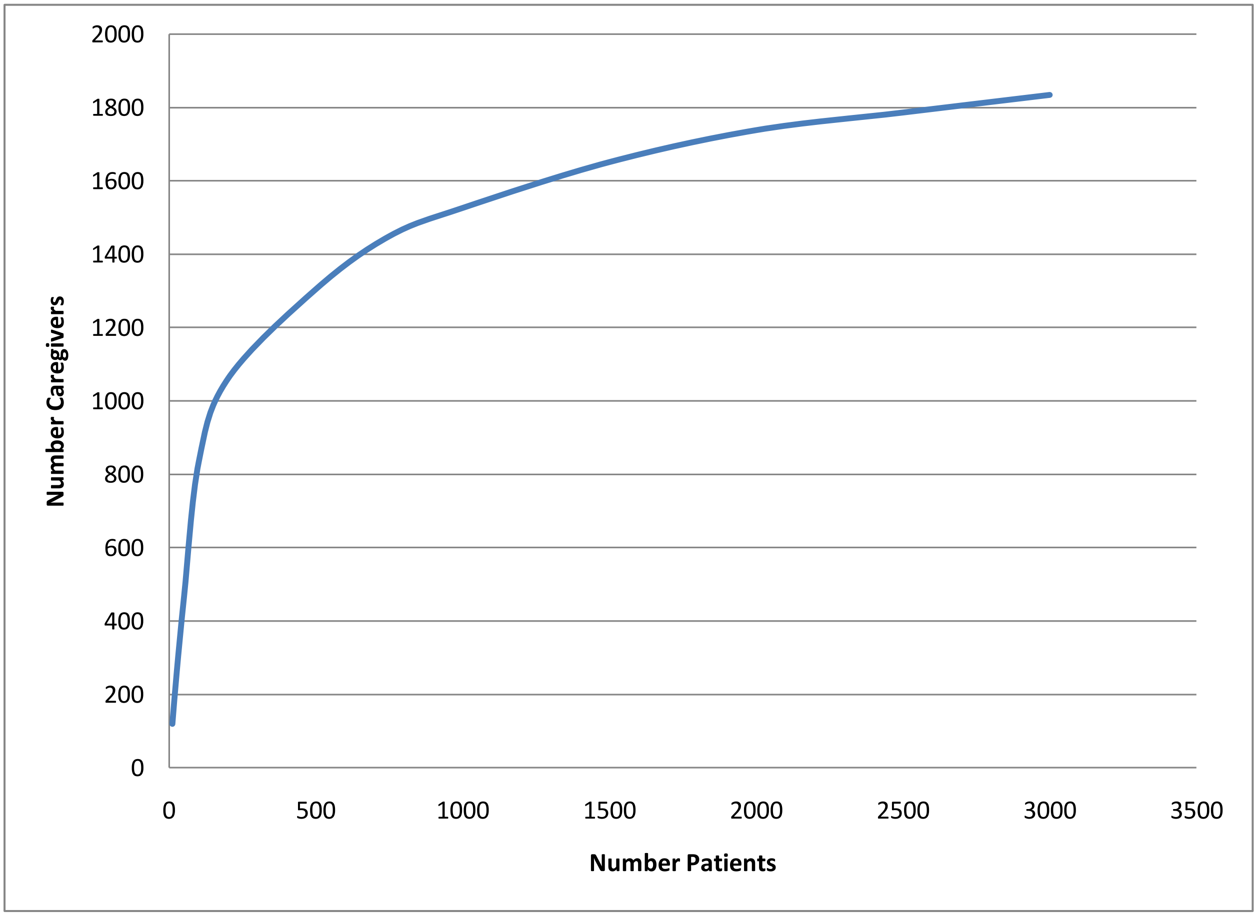


Figure 34: Number of Caregivers Recording Entries vs. Number of Patients

Caregiver Role

All possible roles assigned to caregivers are described in Table 4. The assignment based on title was validated by looking up references to the caregivers assigned the title in the database and checking whether the context of these references fit the role connoted by the title. For instance, "RN" is hypothesized to be an abbreviation for "registered nurse". To validate this assignment, several caregivers holding this title were queried in the database. As many of these caregivers had entered "nurse's progress notes" for several patients, designation of "nurse" as a role for this title was logical. Similarly, "RT" is assumed to be an acronym for "respiratory therapist". Upon verifying against the database, caregivers of this title entered "respiratory care notes" within the record, which provides sufficient evidence for this assignment.

Role	Title(s)	ICU Function
Administrator	Admin, CsMngm, DirHCQ, PrADM	Management
Associate	CRA, NA, PA, PA-C	Physician's, Nurse's Assistant
Dietician	DietIn, MS-RD	Nutritionist
Doctor	DR, DO, MD, MedStu, Resident	Physician
Medical Technologist	BSMT, MTASCP	Technician
Nurse	RN, LPN, NP, CCRN, StuRN	Nurse
Patientl Care Associate	PC, UCO/PC, CoWkr, PCT	Also called Clinical Care Associate
Pharmacist	PhaD, Ph, RPH, RPHS, StPHA	Formulate Drugs
Rehabilitation Specialist	OTR/L, PT, PTA	Physical and occupational therapy
Researcher	LCP	Laboratory for Computational Physiology only
Respiratory Therapist	CRT, RRT, SRT	Respiratory care
Social Worker	LICSW, MSWint	Social Services

Table 4: Caregiver role descriptions

Caregiver Experience

Like the role of a caregiver, the experience of a caregiver is extrapolated from the job title. The designation of "student" is given to a caregiver when the title includes some indication of student (for example, "MedStu"). The designation of an "experienced" professional is reserved for those caregivers that have certifications or licensure to perform their duties. The designation of a "junior" professional is assigned to those that have recently started a salaried position (for example, "interns"). The designation of a "senior" professional is assigned to the remainder.

Visit Attributes

Day of Week

As the day of the week is preserved in the date-shift algorithm used for de-identification of clinical records, it is back-calculated from the date indicated on the record. It is also used to calculate an overall count of the number of days in the week the patient has been admitted in the ICU for the visit. In this way, it is indicative of length of stay, at the granularity of days.

Season of Year

As the season of the year is preserved in the date-shift algorithm used for de-identification of clinical records, it is back-calculated from the date indicated on the record. It is also used to calculate an overall count of the number of seasons of the year the patient has been admitted in the ICU for the visit. In this way, it is indicative of length of stay, at the granularity of two-and-half months.

Length of Stay

The length of stay, also known as the total admission time, is the entire duration of admission from the visit. It lasts from time of admission to discharge. Since it is a continuous variable, variations using logarithmic transformations and nominalization are also used as analogs of this variable.

Locations and Types

The locations within the ICU, or care-units, are coded within the clinical database. In addition, a care-unit dictionary provides a translation from the code to the type of unit. All location types referenced in the datasets are described in Table 5.

Abbreviation	Description	Function
T-SICU	Trauma Surgical Intensive Care Unit	Monitoring and treatment for patients with injuries.
CSRU	Cardiac Surgery Recovery Unit	Monitoring and treatment of patients following surgery for heart diseases, lethal arrhythmias.
CCU	Coronary Care Unit	Care of patients with heart attacks, unstable angina, and other cardiac conditions.
MICU	Medical Intensive Care Unit	Monitoring and care for patients with lung, kidney, liver, GI, blood conditions and cancers.
C-SICU	Cardiac Surgical Intensive Care Unit	Treatment of patients with complex heart procedures, including open-heart surgery, transplantation, and implantation of ventricular assist devices.
M-CCU	Medical-Coronary Care Unit	Care of patients with vascular conditions.
M-SICU	Medical-Surgical Care Unit	Care of patients with severe respiratory failure, septic shock, and surgical procedures, such as intra-abdominal and orthopedic.

Table 5: Types of Care-unit Locations

Each of these location types is included as a feature for each patient with a binomial value ('TRUE' if the patient was seen in this type of location and 'FALSE' if the patient was not seen). In this way, the location type can be categorically analyzed across patients.

Transmission Attributes

Record Types

The types of records in the clinical database are equivalent to the types of observation events recorded by the caregivers. This list of record types is listed in Table 6.

Record Type	Observations contained	Types of Values
Additive	Amount Doseunit	Numeric Nominal
Chart	2 value types 2 values	Coded Numeric
Delivery	Rate Rate Units	Numeric Nominal
I/O	Amount Doseunit	Numeric Nominal
Medication	Volume Dosage Solution Type	Numeric Numeric Nominal
Note	Unstructured, uncoded, unlimited	Plain text
Report	Unstructured, uncoded, unlimited	Plain text

Table 6: Types of Records in the Clinical Database

Each type of record has an associated count, indicating the number of entries recorded for the type. This continuous variable also has other analogs which use logarithmic transformation and discretization of various types.

Volume

Since each observation event or record entry is regarded as transmission of a single message, a count of the number of messages can be tallied during any given time period. The total message count during a visit is equal to the number of recorded entries during the visit. This metric, however, is not equivalent to the amount of information transmitted, as this would require the total number of observations, or observation volume, to be considered. This observation volume is not used in this study, as it was not feasible to code the vast number of note and report records by observation. Rather, the information volume is estimated using the word-count in these plain text records and a value of one (1) unit for all other record types. This gross metric is limited to comparison of relative amounts of information within a certain record type (for example, a note record with word count higher than another is assumed to include more observations). This metric is not used to compare estimates across record types (for example, a single observation recorded in a chart cannot be compared to a report record with a particular word count). Due to the inaccuracy of the estimates of note and report observations, the estimated total observation volume is a gross estimate of volume of information and should not be interpreted in terms of absolute numeric value; rather, it may only be used to compare in relative terms (for example, patient A is in a higher segment of total observational volume than patient B within the same dataset). Contrasting the two metrics, the messaging volume is a more reliable indicator of transmission volume as I/O volume, whereas the estimated observational volume is a less reliable indicator of information volume. This study primarily focuses on message volume.

Duration

Since each observation has an associated timestamp and the admission and discharge times are also known, it is possible to determine the times of the first and last message (or observation event). The absolute difference between these two values indicates the total messaging period during the patient visit. This metric can be regarded as the active transmission period.

Load (or Rate)

The transmission load (or transmission rate) may pertain to either an estimate of the total observational content or the total message content. The former is based on the estimate of the observational volume and the later based on the message volume. In either case, the calculation is based on:

$$\text{Load (or Rate)} = \text{Volume} / \text{Duration}$$

The duration is taken as the time interval between the first and last message. In this way, the load is calculated as the expected average rate of information flow during a patient visit.

Feature Analysis

The features resulting from the exploratory analysis described in the previous section are processed in three stages. First, the continuous response variables (i.e., messaging metrics – volume, duration, and load) are converted to categorical variables covering ranges of values in order to allow them to be eventually input into the Bayesware software package to generate network models. Then, the input features are evaluated individually against these transformed variables using logistic regression, so that they may be individually screened for association with the target variables. The features that are found to be statistically significant are kept for further analysis and those found insignificant are eliminated, as they indicate no effect on the response variable. The remaining list of features is further reduced by removing correlates of the key variables, so that they do not overshadow the effect of the primary explanatory variables. This final set of explanatory variables (listed in Appendix F: Selection of Features for Messaging) is used for construction of the predictive models described in the next section.

Discretization of Transmission Metrics

All three types of messaging metrics are continuous variables that hold positive real values. To facilitate generation of the Bayesian prediction model using the Bayesware software package (Sebastiani & Ramoni, Bayesware, 1999-2000), these variables are discretized (or "binned") into ranges of values. Although the software provides its own utilities for quantization, the discretization step was performed separately at earlier stages of analysis, so that the variables are characterized as early as possible and any biases are discovered earlier rather than later. This binning process is unique to each metric, based on its distribution in the dataset.

Transmission Volume

The transmission volume is measured by the following metrics: (a) the number of messages transmitted during the admission period, and (b) an estimate of the number of observations by record type transmitted during the admission period. The message count metric follows a severely positively skewed distribution, as shown in Figure 35. In order to ensure that the skew does not obfuscate the analysis, a logarithmic transformation was applied, as shown in Figure 36.

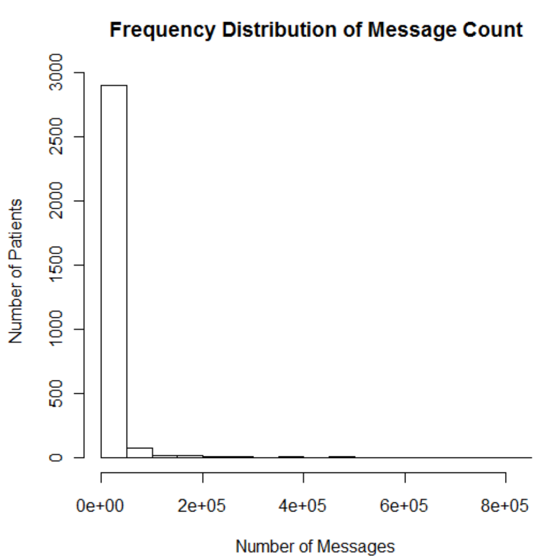


Figure 35: Frequency Distribution of Messaging Volume

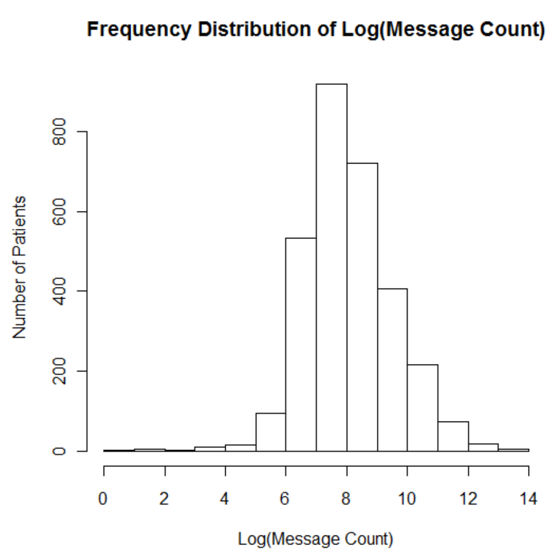


Figure 36: Frequency Distribution of Transformed Message Volume using Natural Logarithm.

The resulting non-skewed natural logarithmic values were then equally split into 5 bins (or quantiles) with equal number of patients (uniform frequency distribution). The lower and upper quantiles were retained and the middle three quantiles were combined to form a mid-range, resulting in a 3-class variable. The majority of the patients in this mid-range are between approximately 650 and 5000 messages.

Similarly, the estimated number of observations is also transformed logarithmically for each record type to account for the positive skew, as shown in Figure 38 to **Error! Reference source not found.**.

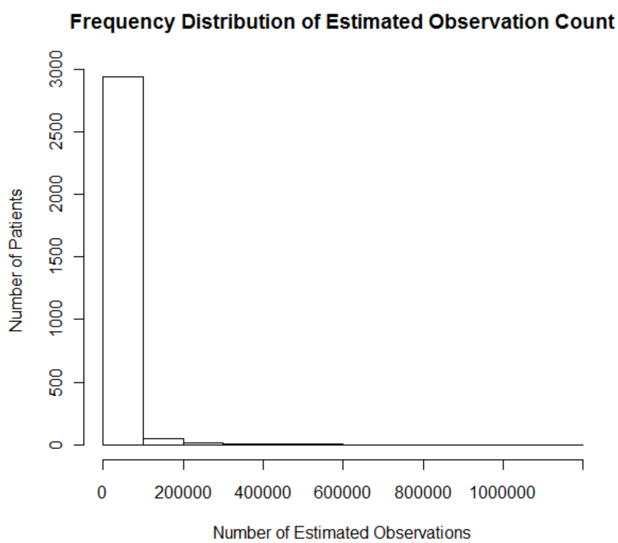


Figure 38: Estimated Total Number of Observations

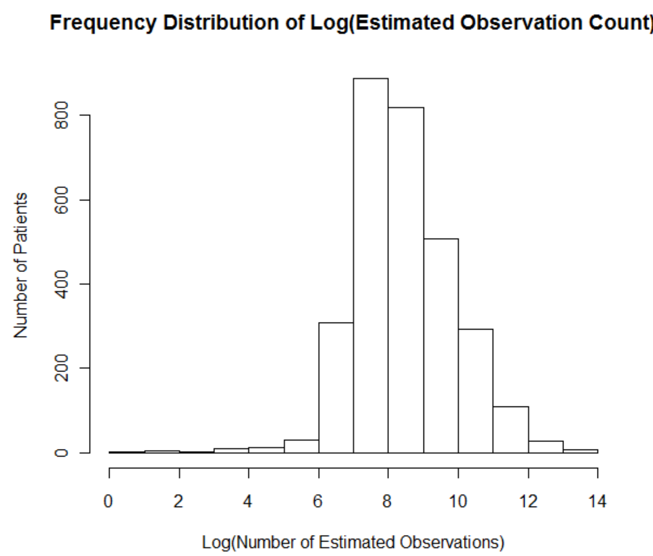


Figure 39: Frequency Distribution of Natural Logarithm of Estimated Total Number of Observations

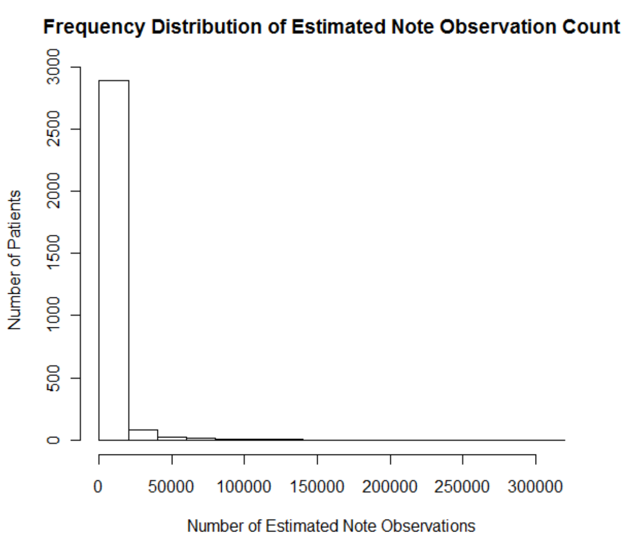


Figure 37: Estimated Number of Note Observations

Transmission Duration

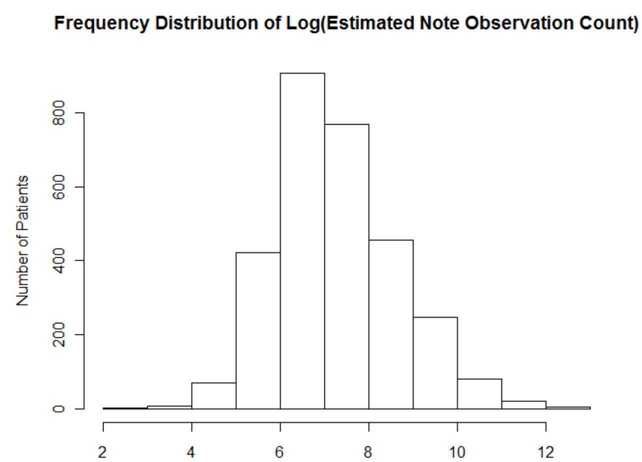


Figure 40: Frequency Distribution of Natural Logarithm of Estimated Note Observations.

The transmission duration is measured as the time period between the first and last message during the patient's visit. Like transmission volume, this metric also follows a highly positively skewed distribution. As such, it requires logarithmic transformation to remove the skew.

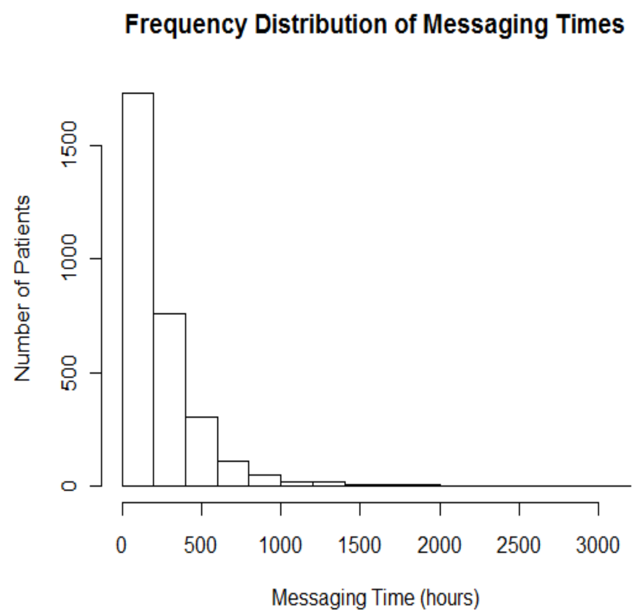


Figure 41: Frequency Distribution of Messaging Time

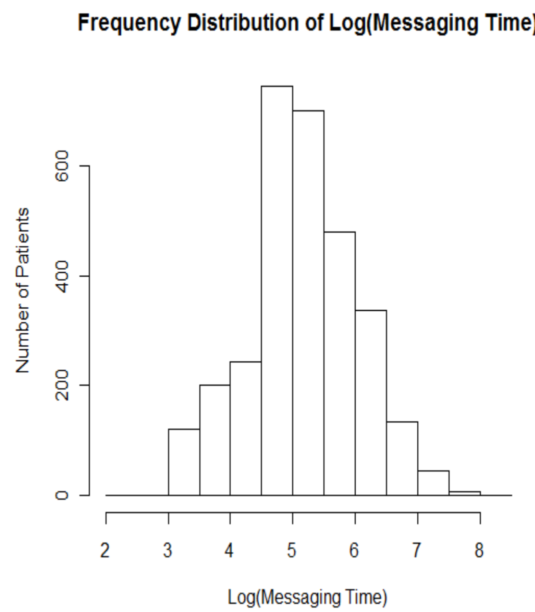


Figure 42: Frequency Distribution of Natural Logarithm of Messaging Time

The transformed messaging duration values were also split into 5 bins (or quantiles) with equal number of patients (uniform frequency distribution). The lower and upper quantiles were retained and the middle three quantiles were combined to form a mid-range, resulting in a 3-class variable. The majority of patient cases in this mid-range had a messaging time between 4 to 10 days, similar to the length of stay.

Transmission Load

Unlike the other two metrics, the transmission load, measured as number of messages over the messaging period, follows a slightly positively skewed distribution. When measured in observations per messaging period, the distribution is normal with high kurtosis (longer tails and less flat on top).

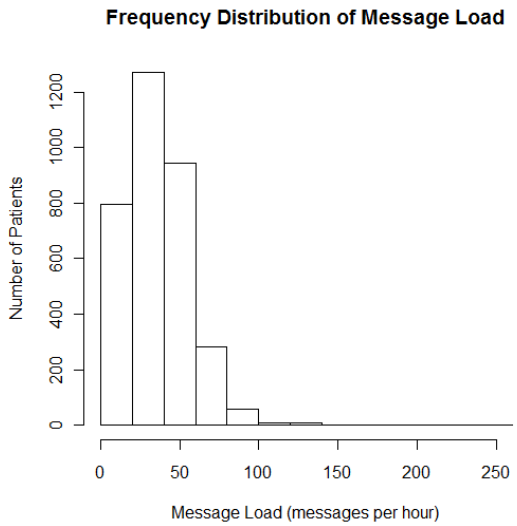


Figure 43: Frequency Distribution of Message Load

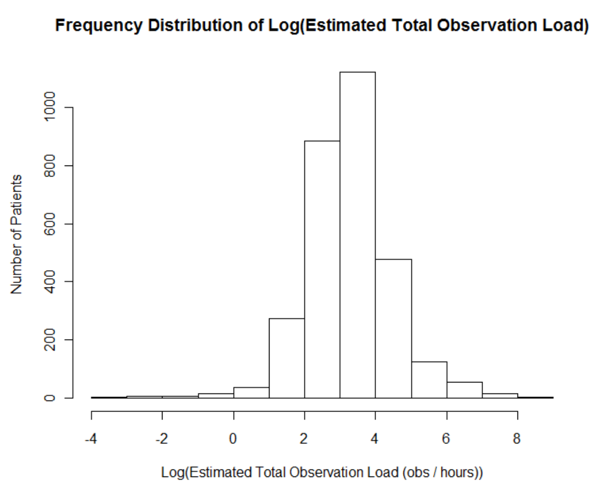


Figure 45: Frequency Distribution of Natural Logarithm of Estimated Observation Load (per Hour)

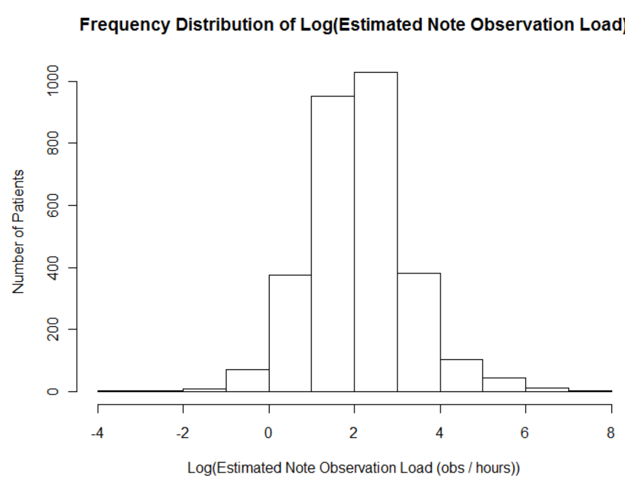


Figure 44: Frequency Distribution of Natural Logarithm of Estimated Note Observation Load (per Hour).

The estimated observational load (based on word count for notes and reports) is in the majority between 15 and 50 per hour, the estimated note observational load mostly between 4 and 20 per hour, and the estimated chart observational load is mostly between 9 and 40 per hour. Due to the longer tails in the observation loads, a log transformation is performed on the "raw" calculated values. These metrics are each directly segmented into 5 equal bins (or quantiles) for subsequent analysis. The lower and upper quantiles were retained and the middle three quantiles were combined to form a mid-range, resulting in a 3-class variable. The message load in this mid-range is between 10 and 50 messages per hour for the majority of patient visits.

Attribute Elimination

Each of the possible explanatory variables is individually evaluated for significance with the transmission response variables. A univariate logistic regression model is used to perform this evaluation, followed by the Wald test to select features based on significance. The Wald test calculates a Z score based on the logit coefficient normalized against the standard error. An associated p value is also computed to test significance of the Wald-Z score. A high absolute Z value with a low p-value indicates that the variable is significantly related to the response metric. A positive Z value indicates that the variable is directly related to the response metric, whereas a negative Z value suggests negative correlation. Those variables that are found to have low significance associated with their Wald-Z scores ($p >= 0.05$) are eliminated and others found to have more significance ($p < 0.05$) are retained; i.e., univariates that are at least 5% likely to be unassociated with the response variable are removed. Although this approach reduces the complexity of the model efficiently, it has the disadvantage of eliminating attributes though they may potentially prove significant in concert with other variables. A forward/backward search algorithm using multi-variate regression is an alternate means of performing feature reduction. Due to time constraints, univariate elimination is selected over the multi-variate method. Following the elimination round, all remaining explanatory variables are ranked by Wald-Z scores.

Attribute Selection

Even after eliminating features based on significance levels, the number of attributes remains too large for a reasonable predictive model. The curse of dimensionality dictates that the performance of a predictive model degrades as the number of features input to the model increases. In addition, several variables may provide the same explanatory effect (for example, number of caregivers and number of caregivers by role). This redundancy is assessed using a ranked-correlation analysis of the features. The Spearman method ranks samples for a given dependant variable by its value and computes a correlation coefficient on the resulting rank order. In this way, the monoticity of the relationships between the features is preserved. A self-referential Spearman correlation matrix is constructed for all attributes. Features are clustered (by interchanging rows and columns) according to the degree of correlation. Features are selected from each cluster that has a unique explanatory effect.

Bayesian Network Prediction Models

The set of features that survived the univariate and correlation analyses are used to construct models to predict the transmission metric values based on these attributes. A predictive model is chosen (rather than a simpler descriptive model) for several reasons: (a) validation of the model would undergo more rigorous treatment, and (b) future simulation work is possible based on the predictive model. In addition, since the response variables are known (the transmission metrics), supervised learning techniques can be utilized to construct the model. Although a wide variety of such modeling techniques are available, including linear and logistic regression, tree-based methods, neural networks, support vector machines, and random forests (Hastie, Tibshirani, & Friedman, 2009), Bayesian Belief networks is the selected method of choice.   A wealth of literature exists in the use of BN models in biomedical informatics (Husmeier, Dybowski, & Roberts, 2004), (Sebastiani, K.D., Szolovits, Kohane, & Ramoni, 2006). The Bayesian method also elicits an intuitive structure of the associations between the model attributes.

Data Set Partitioning

Independent of the learning technique used, the data set is partitioned into two groups: (a) a training set and (b) a validation set. The training set is used to construct the model and the validation set is used to revise the model. In practice, the vast majority of the data is used for training and a smaller amount is used for model validation. Due to the wealth of clinical data in this study, the reverse was employed, where twenty percent (20%) of the data was used for training and eighty percent (80%) was used to validate the BN models. In this way, the models also underwent more rigorous scrutiny with the larger validation sets.   An additional test set is generated from the clinical database for independent testing of the model. The characteristics of all three datasets are described in Table 7. As shown in the table, the distribution of the primary attributes, including those for patient (including expiration status and gender), caregiver roles, and care-unit occupancy are approximately equivelant across all datasets. . The number of condition codes, total number of caregivers, and total estimated observations are slightly disproportionate across the datasets. However, these characteristics are aggregated in large quantities and should have minimal impact, in terms of bias and variance, on evaluation.

Object/ Attribute	Training Set	Validation Set	Test Set
Adults Patients	750	2265	3376
Admissions	750 (1:1)	2265 (1:1)	3376 (1:1)
Male Gender	301 (40%)	957 (42%)	1458 (43%)
Expired	77 (10.26%)	242 (10.68%)	363 (10.75%)
Conditions	1330	2266	2578
Caregivers	1411	1775	1887
Nurses	1020 (72.29%)	1296 (73.01%)	1375 (72.87%)
Respiratory Technicians	270 (19.13%)	322 (18.14%)	328 (17.38%)
Doctors	18 (1.27%)	25 (1.40%)	21 (1.11%)
Patients in Cardiac Units	540 (72.00%)	1629 (71.92%)	2436 (72.15%)
Patients in Trauma Units	33 (4.4%)	91 (4.01%)	145 (4.30%)
Patients in Medical Units	253 (33.73%)	815 (35.98%)	1166 (34.54%)
Total Records (in millions)	8.612	30.977	32.506
Total Estimated Observations (in millions)	9.429 (11:10)	35.517 (7:6)	45.508 (7:5)
Estimated Note Observations (in millions)	2.756 (29.23%)	10.446 (29.41%)	13.432 (29.52%)
Estimated Chart Observations (in millions)	6.051 (64.17%)	22.680 (63.86%)	29.117 (63.98%)

Table 7: Characteristics of Datasets

Bayesian Network Learning

As described in the Background section, Bayesian network models can be used to assess conditional dependence (and independence) relationships between variables. In this way, how the explanatory variables influence the transmission metrics can be explained through the construction of such a conditional dependence network. Network structures are developed using the Bayesware Discoverer software package (Sebastini & Ramoni, 1999-2000). The training data set is first loaded into the software application. The continuous variables are then discretized either by the frequency or the range of values, depending on the attribute. Since the training and validation sets may not be comprehensive in covering all classes for any particular attribute, missing levels are added as possible values for each attribute as part of the discretization process to prepare the model against test samples that may hold these levels not predicted by the model.

The learning strategy used to construct the network is dependent upon: (a) the ordering of the attributes and (b) the algorithm used to select a model.

Ordering of Attributes

The attributes were ordered using the following rules (in order of priority):

- Patient's attributes precede caregiver attributes which precede visit attributes.
- Transmission metrics are specified last.
- Features are sorted according to their Wald-Z scores, from highest to lowest.

Search Algorithm

The Bayesware software package provides three choices for the search algorithm: greedy, arc-inversion, and exhaustive.

The greedy approach, also known as the K2 algorithm, starts with the attribute at the bottom of the order and tries single edges between each of the remaining attributes as a parent to this child node. The parent is chosen with the corresponding model that yields the highest joint probability with a Bayes Factor above 1. The Bayes Factor is a ratio used to assess the predictive capability of two models, by comparing the marginal likelihood of predicting the data given each of the models individually. In this instance, the factor is used to assess whether the model has improved with the addition of the parent node, whether its marginal likelihood has increased (the ratio is above 1). Once a parent node is selected using the Bayes Factor in this way, the algorithm then seeks another parent for the child node following the same process, until it cannot add any more parents. The algorithm may also be constrained by the number of parents that are sought for a particular child node. In this study, however, this parameter is left unrestricted for model generation in this study. Once the child node is processed, the algorithm moves to the next attribute higher up in the ordering. In this scheme, the bottom-most attributes have no capability of becoming parents to those attributes higher up in the order.

The arc-inversion scheme follows the same model selection algorithm as the greedy approach, based on the joint probability and Bayes factor. However, the traversal of the nodes is reversed: from the top-most node to the bottom-most, seeking children for parents rather than parents for children.

The exhaustive approach attempts all possible models for the attributes and chooses the best based on the same criteria as the other two algorithms. This approach is computationally too expensive for models with as many variables as used in this paper.

The algorithm chosen for this application is the greedy (or K2) algorithm, as it has been cited extensively in the literature and used in many other applications.

Cross-validation

Once the network structure has been constructed using the training set, the model is evaluated against the validation set. The transmission metrics are chosen as the target variables for evaluation. The Bayesware application is programmed to use K=5 folds. Predicted outcomes are collected for each sample tested, along with all possible values for the attribute and the joint probability. These values are stored as files for post-processing using an evaluation script. The Bayesware application also calculates an accuracy figure for each target attribute. This figure, along with the marginal likelihood for the model, is used to revise the ordering of the attributes, such that the model with the lowest marginal likelihood and the highest accuracy for the target metric is chosen.

Model Evaluation

Since the transmission metrics evaluated in these models are trinomial in value, the class reference approach is used to evaluate each level for every metric against the other two values that it can hold. Results for both test/validation datasets are processed sample by sample by an evaluation script which counts the number of correct and wrong predictions for each metric-value pair. For each metric-value pair, it tabulates the evaluation counters described in Table 8.

	Predicted = Attribute's Value	**Predicted != Attribute's Value**
Actual Value = Attribute's Value	True Positive (TP)	False Negative (FN)
Actual Value != Attribute's Value	False Positive (FP)	True Negative (TN)

Table 8: Test/Validation Evaluation Counters

For example, the message load metric may hold the possible values: 0-20%_quantile, 20-80%_quantile, 80-100%_quantile. For a test sample, if the predicted value and the actual value are in the 0-20%, then the result is considered a true-positive. If the predicted value is in the 0-20% and the actual value falls in the quantiles between 20-100%, then the result is considered a false-positive with respect to the 0-20%_quantile. If the actual value falls in the 0-20% range and the predicted value falls outside this range, then the result is considered a false-negative with respect to 0-20%_quantile. If the predicted and actual values both fall in any quantile between 20-100%, then it is considered as a true-negative with respect to the 0-20%_quantile.

The evaluation counters are used to calculate the Sensitivity and the False Positive Rates of the predictive models for each attribute-value pair, using the following equations:

Sensitivity = TPR = TP / (TP + FN)

(1 – Specificity) = FPR = FP / (FP + TN)

These values are then plotted for each validation and test fold to generate the Receiver-Operator Characteristic (ROC) curves for each metric-value pair. The ROC curve is a plot of the sensitivity (or TPR) against FPR. An ideal predictive model results in 100% TPR and 0% FPR (top-left corner of the plot). Due to test error, however, even though the rate of correctly predicted positive values may be high, the number of false alarms may be also high. The ROC curve illustrates this cost/benefit trade-off. The area under the ROC curve (auROC) provides a single metric to measure the predictive capability of a model and is calculated using the trapezoidal rule for each curve, as results are generated. This auROC result is used to assess the predictive accuracy of each model, based on its proximity to the value of 1.

Results and Discussion

The transmission of information within the ICU clinical database is analyzed via four methods: (a) data visualization – plots along several dimensions reveal certain messaging patterns in relation to other attributes, (b) histogram/frequency analysis – frequency distributions describe various features , (c) univariate analysis – individual association between the explanatory and response variables are described, and (d) Bayesian belief networks – models predict levels of the transmission metrics, as influenced by explanatory variables and their probabilistic inter-dependencies. The first two analyses provide a descriptive view of transmission data extracted from the clinical database. The latter two analyses suggest a predictive model for how clinical information is transmitted in the ICU. Results from both sets are integrated to form a coherent view of how information is transmitted across the clinical database.

Descriptive Models

Messaging Patterns

Patient Attributes

Although the adult patients in the dataset are close to evenly distributed by gender (females outnumber males 3:2), they tend to be older in age, with the majority between 55 and 85 years. Patients within this age group exhibit a variety of clinical conditions, but predominantly cardio-vascular in nature. These older patients are also seen primarily in the cardiac care units. In contrast, younger patients, between 15 and 25 years, are seen primarily in the trauma and medical units for fractures, burns, wounds, contusions, and other such acute conditions related to trauma. The messaging load (or transmission rate) appears to be much higher for patients seen for cardiovascular conditions in the cardiac units, then the younger patients seen in medical and trauma units. The message count (or transmission volume), however, does not rise for either set of circumstances – it stays roughly constant. In contrast, the messaging time is much higher for patients seen in the trauma units and much lower for patients in the cardiac units. These results suggest that, in the ICU, time duration plays a more significant role than volume of recording activity in driving the rate of transmitted information higher. This applies to the cardiac as well as trauma and medical units, albeit the metrics are inverted in both scenarios.

Most expired patients have higher than average SAPS I scores, as expected, since the SAPS I score tries to predict mortality. The message load for these patients has a wide spectrum of values, whereas message count is slightly less than average and the messaging time is unbiased toward either feature. These patterns suggest that patient's risk of mortality on its own does not have significant influence over transmission of clinical information. Other factors, however, may come into play to affect these response variables.

Caregiver Attributes

The number of caregivers transmitting clinical information per patient grows logarithmically with the number of patients, which implies that the ratio of caregivers to patients is not constant in the record. However, the distribution of caregivers transmitting per role type does appear to be constant across all datasets with nurses being approximately 71% of transmitting caregivers, followed by 17-20% respiratory technicians, and only about 1% doctors.

Messaging patterns appear to be specific to the caregiver type and location. Respiratory technicians record predominantly chart and note entries in the cardiac units, whereas nurses record uniformly across all record types in all locations. Doctors appear to rarely record entries, but when they do, they enter data primarily in the medical units. Patient care associates are most actively transmitting in the medical care-units, but not as much in the cardiac units.

The most striking pattern is of respiratory technicians dominating transmission of clinical information for patients with the highest SAPS I scores, over any other caregiver role, even nurses. Evidence of this pattern is also visible in the MICU, C-SICU, and the CSRU, over other care-unit locations. Reasons for this are explored during discussion of the integrated view.

Visit Attributes

The care-unit locations are, for the most part, functionally distinct, although researchers have found that patients are occasionally admitted into a given care-unit due to capacity considerations, rather than for their clinical condition (Zhang & Szolovits, 2008). The T-SICUs primarily cater to patients suffering from traumatic conditions. The cardiac units, C-SICU, CCU, and CSRU, serve patients with primary cardio-vascular clinical conditions. The medical units, MICU and MSICU, serve patients with a variety of clinical conditions. The MICU and the cardiac units appear to serve patients with the highest SAPS I scores. Although respiratory technicians are transmitting from other units, they are most actively transmitting from these units for these patients. Reasons for this are explored in subsequent analyses.

The time duration for messaging is much lower in the cardiac units then it is in the medical and T-SICU units, despite the length-of-stay varying in these units. This shortened messaging time may be driving the message load to be higher in these units. More evidence is required to determine whether this is a significant trend.

Effects of Explanatory Variables on Transmission

Each explanatory variable is evaluated against each transmission response variable individually and only those found to have significant association are analyzed. The list of significant (and insignificant) features and their “sphere of influence” upon the transmission metrics are illustrated in Figure 46 and described in Appendix F: Selection of Features for Messaging Prediction Models

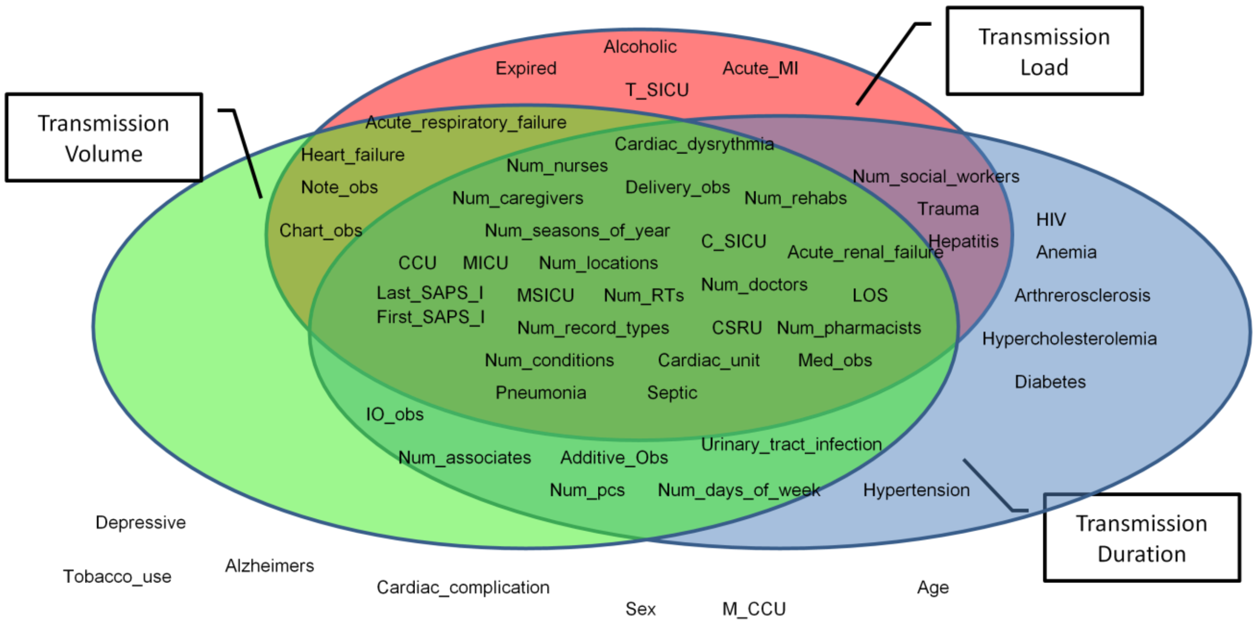


Figure 46: Effect of Explanatory Variables on Transmission Metrics

All care-unit locations, except for the trauma unit (T-SICU) and the medical-coronary care unit (M-CCU), appear to associate with all transmission metrics. The other service type common to all metrics is the length-of-stay. The SAPS I scores also associate with all metrics, even though expiration status only appears to affect messaging load. In addition, the number of particular caregiver roles has an effect on all metrics, including nurses, respiratory technicians, pharmacists, and doctors. The types of records, particularly notes and charts, affect all metrics.

Patient conditions are particular to certain metrics. The messaging time appears to be solely affected by chronic conditions like atherosclerosis, hypercholesterolemia, diabetes, hypertension, and viral infection (including HIV and hepatitis), whereas the messaging volume and load together are affected by more acute conditions like heart failure and acute respiratory failure. A few patient symptoms appear to be affected by all metrics, including pneumonia, sepsis, acute renal failure, and cardiac dysrythmia.

Transmission Volume

As shown in Table 9, the primary factors influencing transmission volume are those observations recorded as chart and note entries. This result also corroborates with the patterns discovered in the previous section. In addition, the number of caregivers involved in a patient case, esp. nurses, respiratory technicians, and patient care associates, has a strong influence on the amount of clinical information transmitted. If the distribution of caregivers are biased toward nurses (over respiratory technicians), however, there is less clinical information transmitted. Although the SAPS I scores are influential, they do not have as strong an association as the caregiver and record type attributes. The care-unit locations and individual patient conditions also have less of an association with transmission volume.

	Estimate	Std. Error	Z value	|Z| value	Pr(>|z|)
chart_est_obs_segment	4.329716	0.236776	18.28613	18.28613	1.07E-74
est_obs_segment	4.144043	0.234118	17.70064	17.70064	4.15E-70
num_days_of_week	0.287036	0.01877	15.29195	15.29195	8.65E-53
note_est_obs_segment	3.341639	0.220179	15.17689	15.17689	5.03E-52
num_nurse_segment	0.800111	0.053407	14.98139	14.98139	9.72E-51
num_caregivers	0.030913	0.002622	11.78886	11.78886	4.46E-32
num_nurses	0.051948	0.004426	11.73581	11.73581	8.35E-32
num_rt_segment	0.345638	0.030728	11.24832	11.24832	2.36E-29
num_rts	0.067475	0.006495	10.38953	10.38953	2.77E-25
num_record_types	0.362782	0.038602	9.398001	9.398001	5.56E-21
num_pcs	0.20526	0.021868	9.386121	9.386121	6.23E-21
total_adm_time	0.001446	0.000167	8.669592	8.669592	4.34E-18
num_locations	0.555924	0.067427	8.244811	8.244811	1.65E-16
est_additive_obs	0.032115	0.00396	8.109262	8.109262	5.09E-16
num_location_types	0.58446	0.073295	7.974098	7.974098	1.53E-15
num_seasons_of_year	0.712502	0.090598	7.86445	7.86445	3.71E-15
est_delivery_obs	0.041347	0.005309	7.788401	7.788401	6.79E-15
est_total_obs	1.48E-05	1.97E-06	7.512632	7.512632	5.80E-14
est_chart_obs	2.31E-05	3.09E-06	7.478425	7.478425	7.52E-14
msg_count	2.05E-05	2.75E-06	7.465033	7.465033	8.33E-14
est_note_obs	4.89E-05	6.60E-06	7.410211	7.410211	1.26E-13
nurse_dist	-1.47471	0.203364	-7.25159	7.251591	4.12E-13
est_io_obs	0.000271	3.87E-05	6.993036	6.993036	2.69E-12
rt_dist	1.42404	0.209744	6.789428	6.789428	1.13E-11
est_med_obs	0.000547	8.22E-05	6.650764	6.650764	2.92E-11
num_conditions	0.057417	0.008883	6.463684	6.463684	1.02E-10
saps_segment	0.323045	0.050823	6.356273	6.356273	2.07E-10
LAST_SAPS_I	0.042744	0.006884	6.20923	6.20923	5.32E-10
FIRST_SAPS_I	0.042776	0.006893	6.205859	6.205859	5.44E-10
num_doctors	0.529351	0.135913	3.894778	3.894778	9.83E-05
acute_respiratory_failure	0.395267	0.106728	3.703484	3.703484	0.000213
acute_renal_failure	0.355567	0.096621	3.680006	3.680006	0.000233
septic	0.417229	0.115284	3.619156	3.619156	0.000296
CSRU	0.270103	0.074679	3.616866	3.616866	0.000298
cardiac_unit	0.231622	0.081285	2.849502	2.849502	0.004379
heart_failure	0.229953	0.087288	2.6344	2.6344	0.008429
pneumonia	0.326792	0.124756	2.619442	2.619442	0.008807
est_chart_obs_load_segment	0.518569	0.206041	2.516825	2.516825	0.011842
msg_load	0.004484	0.001797	2.495296	2.495296	0.012585
MICU	0.200118	0.082735	2.41877	2.41877	0.015573
C_SICU	0.220338	0.099663	2.210825	2.210825	0.027048
CCU	0.167533	0.081419	2.057666	2.057666	0.039622
cardiac_dysrythmia	0.1613	0.079873	2.019444	2.019444	0.043441
urinary_tract_infection	0.228864	0.113799	2.011135	2.011135	0.044311
num_associates	0.571598	0.287972	1.984904	1.984904	0.047155

Table 9: Univariate Effects on Message Count

All features in Table 9 to Table 11 are defined in Appendix F: Selection of Features for Messaging Prediction Models.

Transmission Duration

Similar to transmission volume, the volume of total, notes, and chart observations are the most associated with the time duration. Also, the number and type of caregivers are significantly associated with the messaging time, more so than the length-of-stay (or total admission time). Although not as high ranking, a few chronic patient conditions, such as atherosclerosis, hypercholesterolemia, and hypertension, are shown to be inversely related to the transmission duration, which suggests that the messaging time is shortened for those with long-term conditions.

	Estimate	Std. Error	Wald-Z	|Z|	Pr(>|Z|)
est_total_obs_segment	3.332584	0.052806	63.11004	63.11004	0
est_note_obs_segment	3.239018	0.051779	62.55409	62.55409	0
num_days_of_week	0.277179	0.004492	61.70824	61.70824	0
est_chart_obs_segment	3.266064	0.053894	60.60129	60.60129	0
num_nurse_segment	0.726233	0.013694	53.03168	53.03168	0
num_caregivers	0.027333	0.000781	34.99447	34.99447	1.93E-225
est_chart_obs_load_segment	2.289921	0.065614	34.89996	34.89996	2.03E-224
num_nurses	0.046971	0.001376	34.1433	34.1433	2.77E-216
num_locations	0.668824	0.021897	30.54464	30.54464	9.57E-179
num_rts	0.058044	0.00192	30.22674	30.22674	1.57E-175
num_location_types	0.708491	0.023947	29.5852	29.5852	4.24E-169
num_seasons_of_year	0.859681	0.029889	28.76219	28.76219	5.86E-161
num_pcs	0.190689	0.006851	27.83179	27.83179	6.48E-152
num_rt_segment	0.272901	0.009822	27.78564	27.78564	1.80E-151
total_adm_time	0.001471	5.39E-05	27.30237	27.30237	7.58E-147
num_record_types	0.311782	0.012415	25.11268	25.11268	1.59E-126
num_conditions	0.05055	0.003031	16.67512	16.67512	8.78E-60
nurse_dist	-1.1118	0.069907	-15.9041	15.90407	9.44E-55
rt_dist	1.037289	0.072553	14.29706	14.29706	6.51E-45
est_additive_obs	0.027207	0.002032	13.38788	13.38788	9.52E-40
saps_segment	0.216197	0.017744	12.18422	12.18422	2.27E-33
LAST_SAPS_I	0.028524	0.002397	11.90191	11.90191	5.95E-32
FIRST_SAPS_I	0.027939	0.002368	11.80057	11.80057	1.89E-31
MICU	0.343914	0.02925	11.75771	11.75771	3.07E-31
est_med_obs	0.00045	4.23E-05	10.647	10.647	5.19E-26
acute_respiratory_failure	0.376614	0.037714	9.986057	9.986057	4.00E-23
acute_renal_failure	0.339502	0.03417	9.93572	9.93572	6.53E-23
septic	0.394149	0.040763	9.669391	9.669391	8.41E-22
num_doctors	0.415363	0.045737	9.081611	9.081611	1.89E-19
CCU	0.227234	0.029019	7.830558	7.830558	6.67E-15
heart_failure	0.239582	0.031086	7.707046	7.707046	1.74E-14
pneumonia	0.302327	0.044428	6.804872	6.804872	1.22E-11
num_associates	0.661571	0.104015	6.360365	6.360365	2.32E-10
CSRU	0.16727	0.026629	6.281599	6.281599	3.83E-10
hypercholesterolemia	-0.22615	0.0367	-6.16218	6.162177	8.13E-10
C_SICU	0.197678	0.035608	5.551505	5.551505	3.08E-08
anemia	0.152379	0.030396	5.013143	5.013143	5.66E-07
cardiac_dysrythmia	0.134715	0.028585	4.712822	4.712822	2.55E-06
urinary_tract_infection	0.190377	0.04068	4.679891	4.679891	3.00E-06
num_rehabs	0.530595	0.114677	4.626858	4.626858	3.87E-06
social_worker_dist	-2.09825	0.462393	-4.5378	4.537801	5.91E-06
cardiac_unit	0.119467	0.02907	4.109571	4.109571	4.07E-05
msg_load	0.002875	0.000701	4.101378	4.101378	4.22E-05
MSICU	0.162189	0.041683	3.890992	3.890992	0.000102
pc_dist	0.518544	0.136055	3.811294	3.811294	0.000141
hypertension	-0.09345	0.026697	-3.50026	3.500257	0.000472
est_delivery_obs	0.038487	0.01108	3.473468	3.473468	0.000521
arthrerosclerosis	-0.10108	0.02992	-3.3782	3.378195	0.000739
hepatitis	0.200047	0.064584	3.097495	3.097495	0.00197
num_social_workers	0.157083	0.056921	2.759656	2.759656	0.005821
num_pharmacists	0.755193	0.290663	2.59817	2.59817	0.009418
trauma	-0.08321	0.036544	-2.27701	2.277012	0.022855
est_io_obs	0.000225	9.91E-05	2.272065	2.272065	0.023153
diabetes	0.069329	0.030894	2.244076	2.244076	0.0249
doctor_dist	1.844632	0.860412	2.143895	2.143895	0.032121
hiv	0.229815	0.113571	2.023537	2.023537	0.043106

Table 10: Univariate Effects on Messaging Time

Transmission Load

Like messaging volume and duration, the chart and note observational load appears to have strong effect on the overall messaging load. However, message load appears to be associated with variables similar to messaging duration. Specifically, the care-unit locations and the SAPS I scores have a strong association with transmission load. Patient cases situated in the medical surgical unit (MSICU) experience lower transmission loads, whereas those in the other cardiac units (CCU, CSICU, and CSRU) and the TSICU have higher transmission loads. The nurse and respiratory technician distribution also appear to have an effect on the load, where the former has a decreased load and the latter an increased load.

	Estimate	Std. Error	Wald-Z	|Z|	Pr(>|Z|)
est_total_obs_segment	2.699304	0.060242	44.80797	44.80797	0
est_chart_obs_segment	2.693079	0.060292	44.66719	44.66719	0
est_note_obs_segment	2.440243	0.060485	40.34466	40.34466	6.24E-285
num_nurse_segment	0.541873	0.015437	35.10282	35.10282	1.30E-226
num_days_of_week	0.18835	0.005493	34.28861	34.28861	7.70E-218
num_caregivers	0.021629	0.000718	30.13155	30.13155	1.43E-174
num_nurses	0.035681	0.00123	29.01083	29.01083	2.10E-163
num_rts	0.048996	0.001858	26.37454	26.37454	4.17E-138
num_rt_segment	0.251873	0.009799	25.70443	25.70443	6.53E-132
num_pcs	0.167941	0.006759	24.84874	24.84874	3.77E-124
num_seasons_of_year	0.677665	0.029225	23.18763	23.18763	1.39E-109
num_record_types	0.279532	0.012253	22.81373	22.81373	2.11E-106
num_locations	0.442093	0.021551	20.51403	20.51403	1.20E-87
num_location_types	0.461147	0.023681	19.47366	19.47366	1.19E-79
nurse_dist	-1.16415	0.068716	-16.9414	16.94144	1.44E-61
rt_dist	1.023331	0.071592	14.29389	14.29389	6.79E-45
est_additive_obs	0.027363	0.002063	13.26291	13.26291	4.63E-39
LAST_SAPS_I	0.029221	0.002362	12.37231	12.37231	2.48E-34
est_med_obs	0.00042	3.51E-05	11.97629	11.97629	2.53E-32
FIRST_SAPS_I	0.027939	0.002368	11.80057	11.80057	1.89E-31
saps_segment	0.191859	0.017559	10.92665	10.92665	2.77E-27
expired	0.37472	0.042016	8.91847	8.91847	8.02E-19
num_doctors	0.394324	0.044824	8.797156	8.797156	2.32E-18
num_conditions	0.025408	0.003006	8.453464	8.453464	4.34E-17
CCU	0.226801	0.028631	7.921655	7.921655	3.27E-15
acute_respiratory_failure	0.290151	0.037278	7.783338	7.783338	9.63E-15
est_chart_obs_load_segm	0.539801	0.072474	7.448153	7.448153	1.23E-13
cardiac_unit	0.20732	0.028577	7.254902	7.254902	5.09E-13
est_delivery_obs	0.032871	0.004682	7.0213	7.0213	2.71E-12
septic	0.281805	0.040243	7.002643	7.002643	3.08E-12
CSRU	0.15683	0.026288	5.96591	5.96591	2.72E-09
acute_renal_failure	0.199833	0.033868	5.900404	5.900404	4.03E-09
num_associates	0.530017	0.101106	5.242206	5.242206	1.70E-07
C_SICU	0.182738	0.035148	5.199041	5.199041	2.14E-07
pc_dist	0.678159	0.134351	5.047653	5.047653	4.74E-07
MICU	0.144962	0.029192	4.965788	4.965788	7.22E-07
est_io_obs	0.000197	4.40E-05	4.480353	4.480353	7.73E-06
total_adm_time	0.000215	4.88E-05	4.40384	4.40384	1.10E-05
num_rehabs	0.449991	0.110864	4.05896	4.05896	5.05E-05
est_note_obs	3.64E-05	9.22E-06	3.951442	3.951442	7.95E-05
heart_failure	0.121219	0.030799	3.935775	3.935775	8.48E-05
pneumonia	0.168163	0.043805	3.838885	3.838885	0.000126
doctor_dist	2.978569	0.862993	3.45144	3.45144	0.000565
social_worker_dist	-1.53579	0.450409	-3.40976	3.409764	0.000659
T_SICU	0.210948	0.065229	3.233962	3.233962	0.001234
est_total_obs	1.10E-05	3.69E-06	2.986734	2.986734	0.002842
msg_count	1.52E-05	5.23E-06	2.917195	2.917195	0.003558
est_chart_obs	1.72E-05	6.24E-06	2.754895	2.754895	0.005906
cardiac_dysrythmia	0.074794	0.028254	2.647185	2.647185	0.008159
num_pharmacists	0.748877	0.286812	2.611034	2.611034	0.009072
MSICU	-0.09433	0.041134	-2.29312	2.293124	0.021909
acute_MI	0.08511	0.037129	2.292315	2.292315	0.021956
hepatitis	0.144923	0.063656	2.276657	2.276657	0.022877
num_social_workers	0.125345	0.056144	2.232546	2.232546	0.025652
alcoholic	0.178191	0.08314	2.143261	2.143261	0.032172

Table 11: Univariate Effects on Message Load

Correlation Analysis

A ranked Spearman correlation analysis is performed on the variables found to be significantly associated with the transmission metrics. The rows and columns of the correlation matrix correspond to the features listed in Appendix F: Selection of Features for Messaging Prediction Models. The features are placed in the same order for both rows and columns. The cells within this matrix, displayed in Figure 47, hold the Spearman correlation coefficient between features in the corresponding row and column and are color-labeled as follows: GREEN indicates the most positive association (close to a value of +1.0), RED the most negative association (close to a value of -1.0), and yellow with low association (close to a value of 0.0).

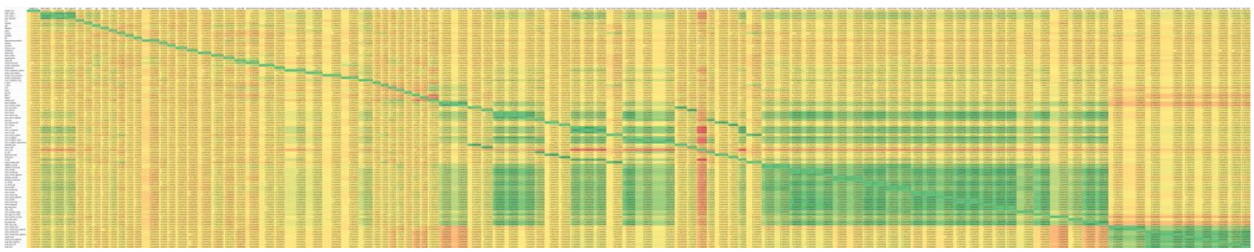


Figure 47: Spearman correlation matrix for significant univariates.

As expected, the diagonal represents perfect correlation between each feature and itself. In addition, there is high correlation between the various metrics for a common attribute (for instance, between estimated chart observations and logarithm of estimated chart observations). From each set, only a single variable is chosen to represent an attribute, the one which has the greatest univariate effect on the transmission metrics (highest Z ranking).

Between individual attributes, there is relatively low association between individual patient conditions and the other explanatory variables. In contrast, the first and last SAPS I scores (second and third rows in the matrix) have a slightly higher correlation with other variables, including estimated observations by type, messaging time, message count, and number and types of transmitting caregivers. In particular, there is a loose inverse relationship between the SAPS I scores and the percentage of nurses transmitting ($r = -0.401$), whereas there is a loose direct relationship with the percentage of respiratory technicians ($r = 0.498$). This result corroborates with the univariate and visualization analyses.

The variables that measure the number of caregivers by type have a significant positive correlation with one another ($r = 0.62$ between nurses and respiratory technicians) and the overall number of caregivers per patient (nurses, $r = 0.62$, respiratory technicians, $r = 0.70$). This result suggests that if the quantity of personnel is higher for a particular case, it is also high per caregiver type; vice versa, if the number of caregivers is low for a particular patient type, it is also low per caregiver type.

The estimated number of observations, message count, and messaging time are highly correlated with the number of caregivers by type (nurses with total estimated observations, $r = 0.91$, respiratory technicians with total estimated observations, $r = 0.70$). These results suggest that patient cases with higher number of caregivers have a higher messaging volume, which may be expected as there are more sources of transmission.

The messaging time is also highly correlated with the message counts (r = 0.89) and number of caregivers (r = 0.85). This result may also be expected as the transmission volume may increase with increased time for transmission. The association between messaging time and number of caregivers is less clear and merits further investigation via predictive models.

Prediction Models

The explanatory features found to have the greatest association with the transmission metrics are used to develop prediction models based on Bayesian networks. The validity of these models is assessed by examining their performance on the test dataset. The network and the associated conditional probability tables (CPTs) are examined to establish relationships with the metric of interest.

Transmission Volume

The target variable chosen for this metric, based on the univariate and correlation analyses, is the trinomial variable corresponding to the logarithmically transformed, message count, also labeled LOG_MSG_COUNT_SEGMENT in Figure 48.

Bayesian Belief Network

The other transmission metrics, messaging duration and load, are included in this model, in order to gain insight into their relative influence on messaging volume. The learned model does show a conditional independence relationship with the messaging duration (LOG_MSG_TIME_SEGMENT) through the parent node for the number of caregivers (NUM_CAREGIVERS). The transmission load (MSG_LOAD_SEGMENT) is conditionally independent through several parent nodes, including estimated chart (EST_CHART_OBS_SEGMENT) and I/O events (EST_IO_OBS), as well as the number of caregivers common to transmission duration. These results suggest that the variables influencing messaging volume may be common to those affecting messaging duration and load. These effects are analyzed further following evaluation of the performance of this model.

Attributes in the resulting model also appear to be clustered around the common object relation . These cliques are demarcated with a dotted boundary around the cluster. Messaging volume appears to be directly influenced by the record clique and other object cliques, such as care-unit and caregiver, are conditionally dependent upon the record clique. This result may be expected, as the amount of transmission is directly related to the number of chart entries, which depends upon who is providing care and where it is being provided. Somewhat surprisingly, the patient condition clique does not directly influence the volume of information generated. It does so primarily through the care-unit in which the patient is admitted.

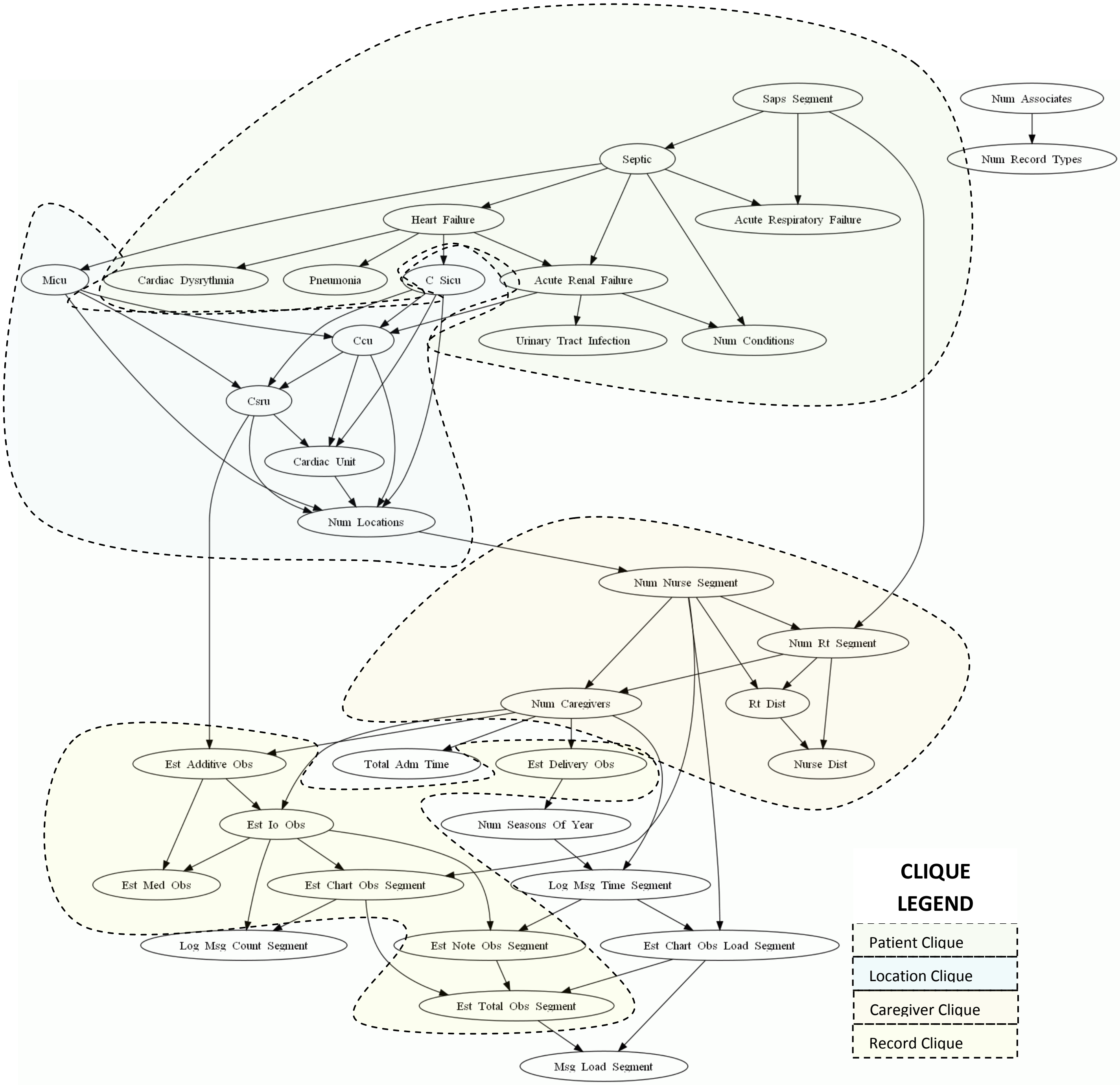


Figure 48: Bayesian Network Model to Predict Message Count

Prediction Performance

The prediction performance for the model in

Figure 48 on the test dataset is evaluated using ROC analysis in Figure 49 through Figure 51 and the performance metrics shown in Table 12.

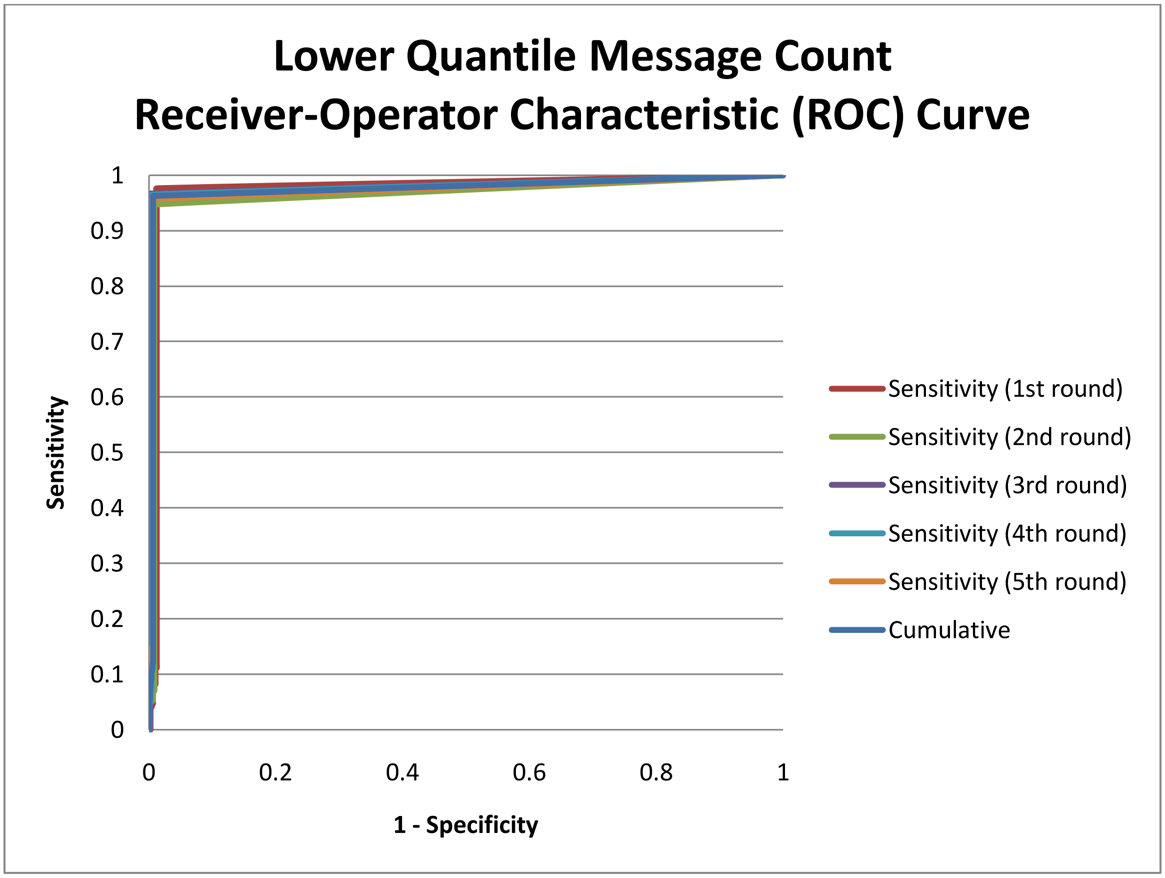


Figure 49: ROC Curve for Prediction of Low Messaging Volume

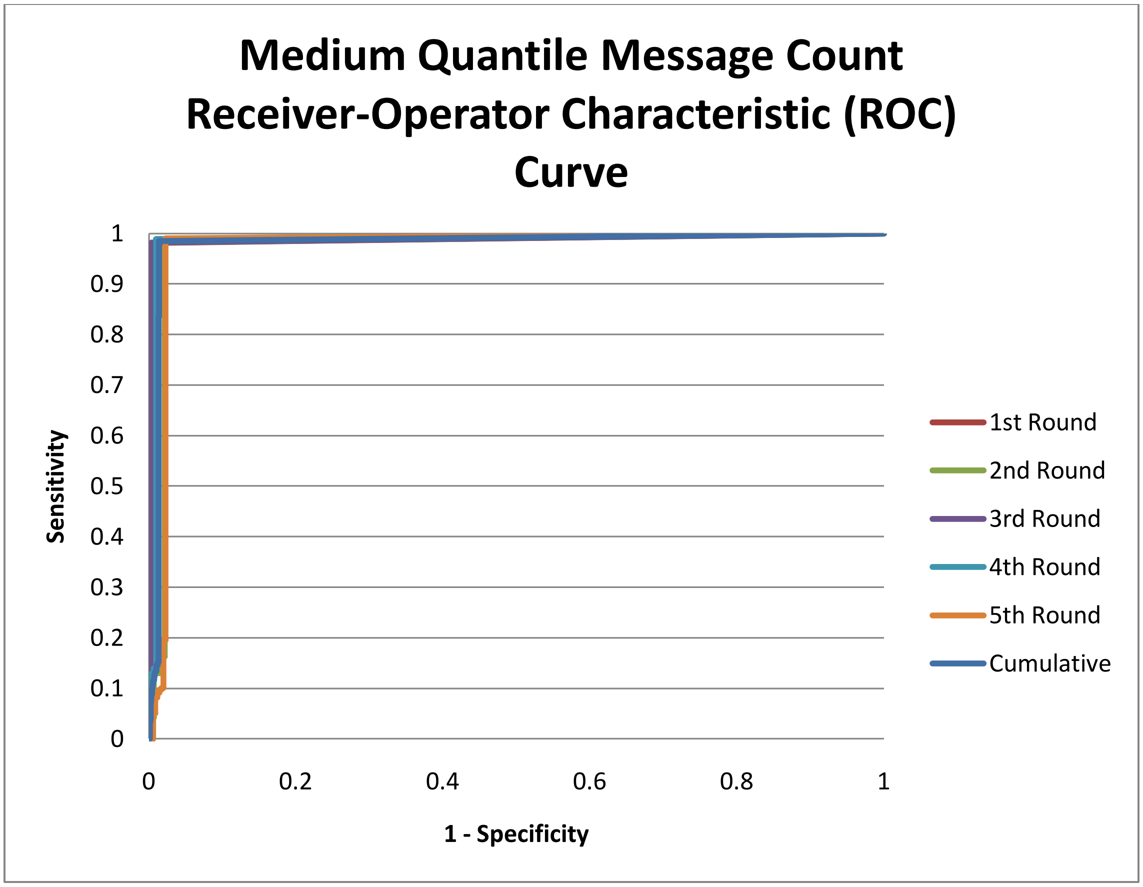


Figure 50: ROC Curve for Prediction of Medium Messaging Volume

ROC curves are generated for five separate partitions of the test data-set, as well as for the entire dataset. All curves tend sharply toward a true-positive rate and low false positive rate (top-left corner of plots), suggesting that the Bayesian classifier described in

Figure 48 is able to discriminate reasonably well on the dataset for each of the three possible values for the bins corresponding to the logarithm of the messaging count. Since the model is not calibrated against the targetted accuracy of the classifier, these models may have some class bias (i.e., more samples holding negative values for each class).

Quantile	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)	True Positive Rate (TPR - Sensitivity)	False Positive Rate (FPR)	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Area Under ROC Curve (AUROC)
low	826	16	2502	32	0.962703963	0.006354249	0.980997625	0.987371744	0.975680925
medium	1667	23	1661	25	0.985224586	0.013657957	0.986390533	0.985172005	0.980894267
high	835	6	2493	42	0.952109464	0.00240096	0.992865636	0.983431953	0.97381849

Table 12: Results of prediction of messaging volume on test dataset.

The medium quantile has a slightly higher performance than the other two classes, which could be attributed to the larger size of the medium bin (20-80%). However, they are approximately equal and high in performance, which allows further analysis of this model.

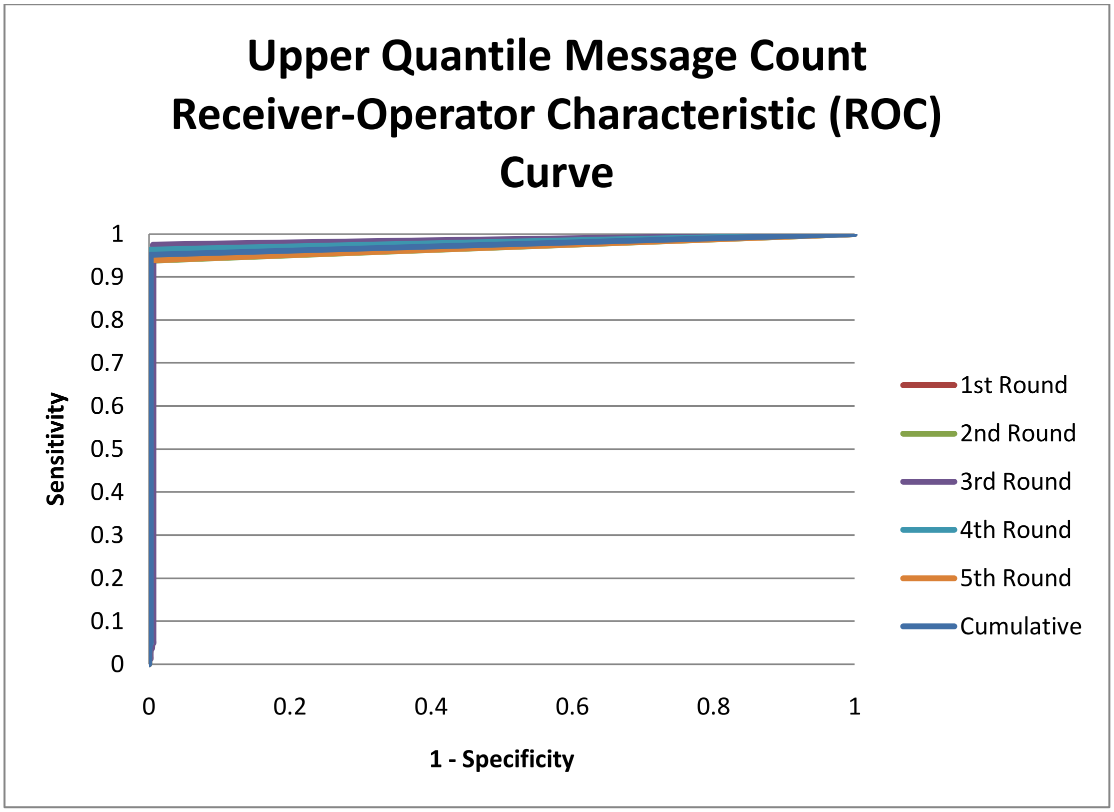


Figure 51: ROC Curve for Prediction of High Messaging Volume

Model Analysis

As the CPT for LOG_MSG_COUNT_SEGMENT in Appendix G: Conditional Probability Tables for Message Volume Prediction shows, estimated chart (EST_CHART_OBS_SEGMENT) and I/O observations (EST_IO_OBS) have a significant influence on all three of the values this variable can hold. According to its CPT, EST_CHART_OBS_SEGMENT appears to be more heavily influenced by the number of nurses (NUM_NURSE_SEGMENT) than EST_IO_OBS. EST_IO_OBS is more strongly influenced by NUM_CAREGIVERS than estimated additive observations (EST_ADDITIVE_OBS). So, it appears that NUM_CAREGIVERS has a strong indirect influence on transmission volume through multiple dependant factors.

The CPT for NUM_CAREGIVERS shows that, for a high number of caregivers, there is greater influence from number of nurses (NUM_NURSE_SEGMENT) than from number of respiratory technicians (NUM_RT_SEGMENT); however, the reverse is true for a low number of caregivers: the number of respiratory technicians is more influential.

NUM_RT_SEGMENT is influenced by both the SAPS I score and the number of nurses, albeit more strongly from the latter. The number of nurses is primarily affected by the number and type of care-unit locations. These locations are, in turn, influenced by certain patient conditions and ultimately the SAPS I score, the top-most parent in the network. The SAPS I score has a more direct effect on transmission volume through the number of respiratory technicians.

As discussed previously, clique analysis also shows that messaging volume is directly influenced by the record type (i.e., charts entered), and indirectly influenced by who is caring for the patient and in which care-unit he/she is visiting. Clinical conditions appear to only influence the amount of information recorded for patients through these indirect factors.

Transmission Duration

Based on univariate and correlation analyses, the target variable chosen for this metric is the trinomial variable corresponding to the logarithmically transformed, message time, also labeled LOG_MSG_TIME_SEGMENT in Figure 52.

Bayesian Belief Network

The network model in Figure 52 includes the other transmission metrics, LOG_MSG_COUNT_SEGMENT and MSG_LOAD_SEGMENT. As in the model for message count, these other metrics are included in order to investigate their level of influence in comparison to the other explanatory variables.

As illustrated in the model, the messaging duration is influenced by a similar set of variables as message count, including the estimated number of note (EST_NOTE_OBS_SEGMENT), chart (EST_CHART_OBS_SEGMENT), and I/O observations (EST_IO_OBS_SEGMENT). As messaging volume may be dependent upon the recording amount of each of the record types, this relationship may seem intuitively obvious.

The SAPS I score (SAPS_SEGMENT) is conditionally independent with messaging duration through number of caregivers (NUM_CAREGIVERS), which may indicate that messaging time is not directly associated with the acuity level, but may be affected indirectly by the number of caregivers involved in the patient case.

Although messaging time is partially derived from total time of the visit, the length of stay (TOTAL_ADM_TIME) is shown to be conditionally independent from messaging time through many of the estimated observation metrics, including EST_IO_OBS_SEGMENT, EST_NOTE_OBS_SEGMENT, EST_CHART_OBS_SEGMENT, and EST_TOTAL_OBS_SEGMENT.

The transmission load metrics, including message and observational load, are conditionally independent from messaging duration through the length of stay and transmission volume metrics. This suggests that volume of observations may be the key driver for all metrics.

Similar to the model for transmission volume, attributes in the transmission duration model cluster around the objects they are characterizing and form cliques. Like volume, duration appears to be directly influenced by clinical observations recorded by type and indirectly influenced by the caregiver. Unlike volume, clinical conditions appear to have a closer effect on transmission duration through the

SAPS I score. The care-unit location and specific clinical conditions appear to be conditionally independent of duration through this acuity score.

However, before any of these relationships can be established, the model's performance must be assessed.

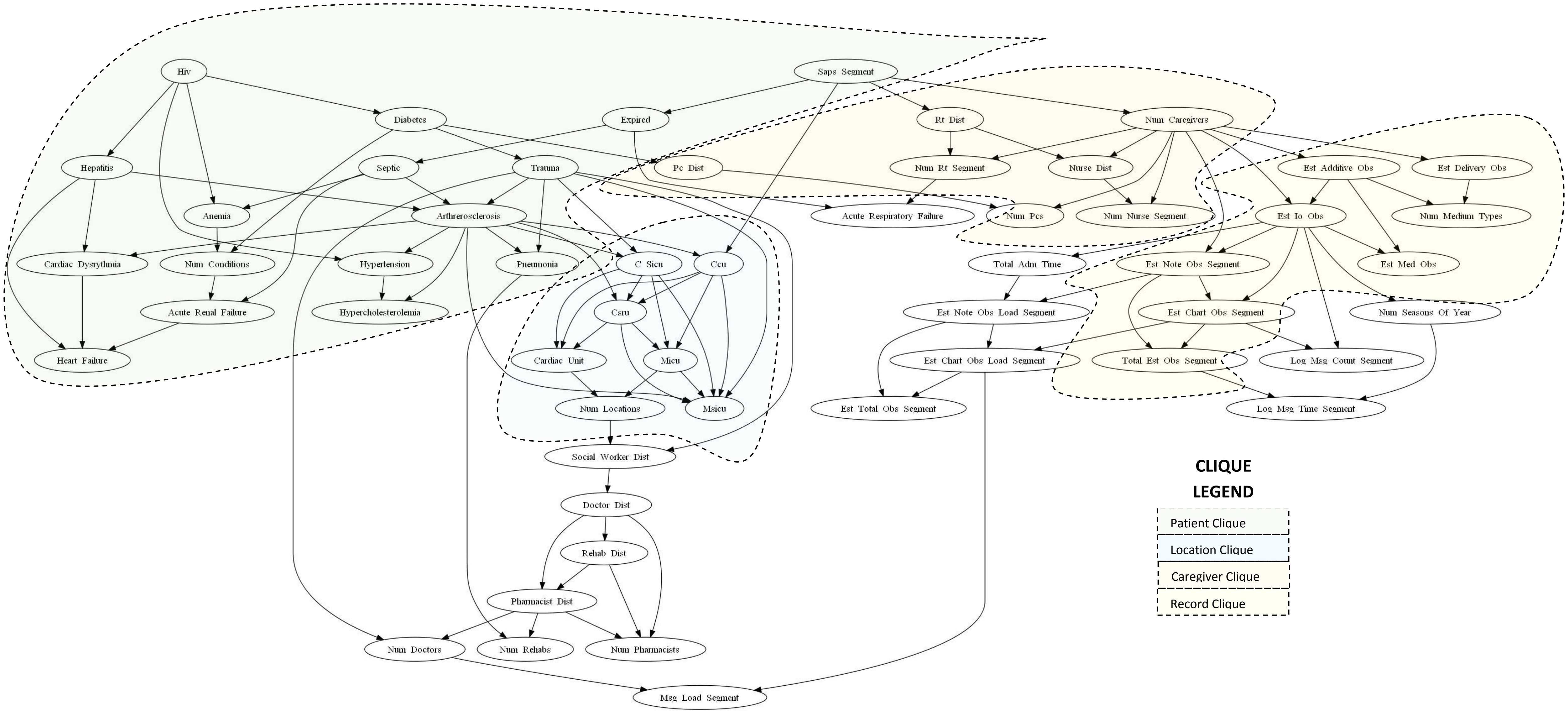


Figure 52: Bayesian Network Model to Predict Messaging Time

Prediction Performance

The prediction performance for the model in

Figure 48 on the test dataset is evaluated using ROC analysis from Figure 53 to Figure 55 and the performance metrics described in Table 13.

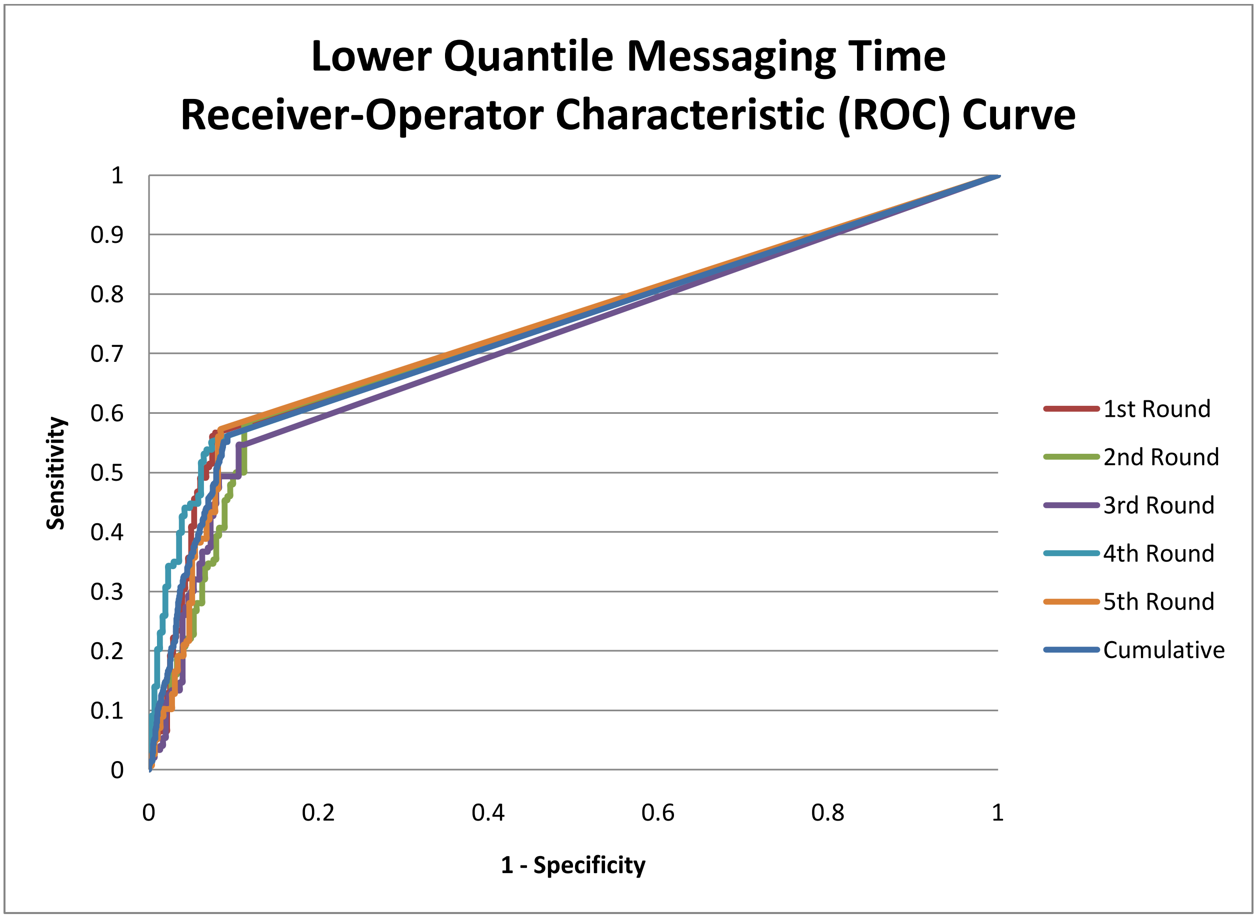


Figure 53: ROC Curve for Prediction of Low Messaging Time

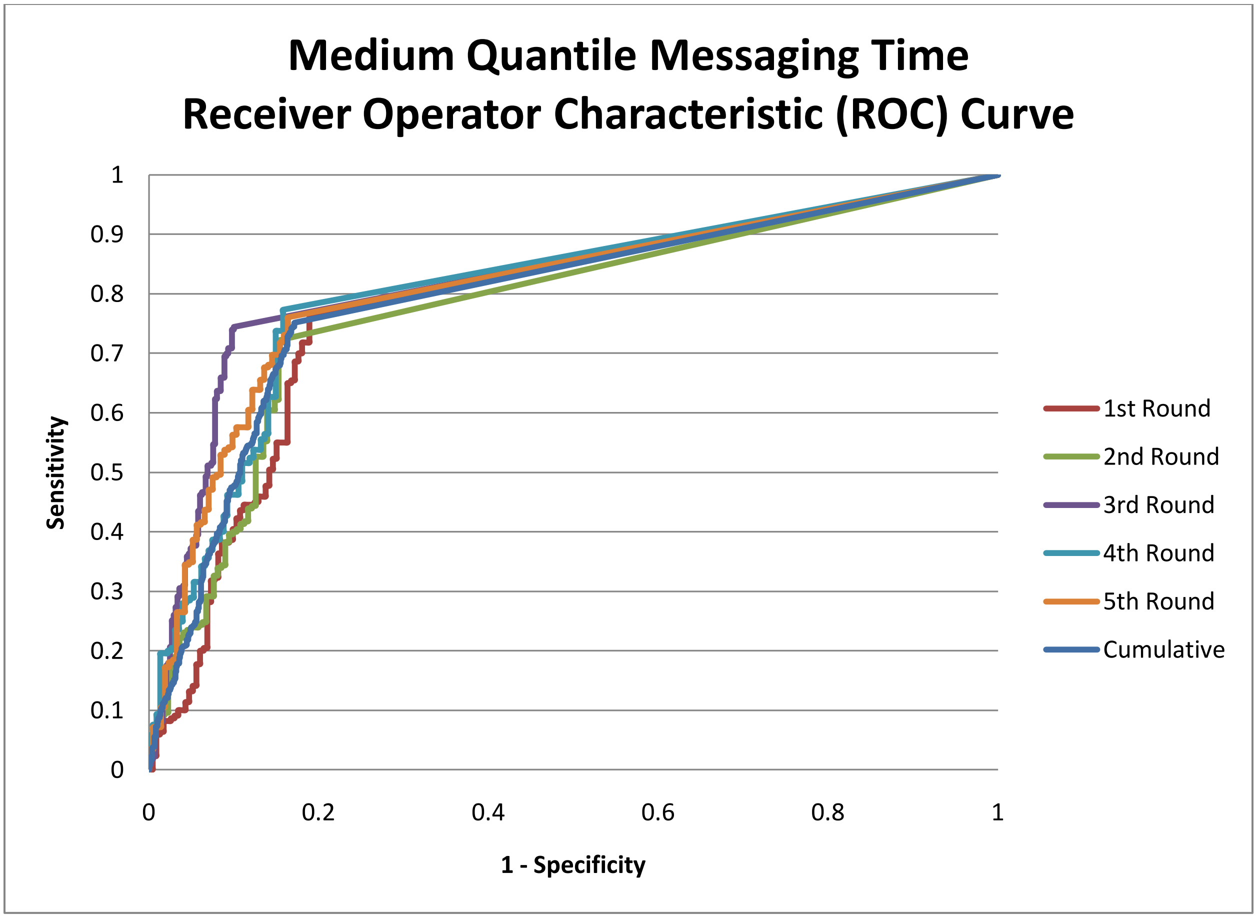


Figure 54: ROC Curve for Prediction of Medium Messaging Time

As the ROC Curves show greater tendency toward the right side of the plots (in comparison to the previous model for transmission volume), they indicate a lower performing model. The lower performance is more apparent for the extreme ends of messaging duration than for the intermediate level.

Table 13 also shows the large number of false positives and negatives generated by the model, especially for the lower and upper levels. Examining the CPT for LOG_MSG_TIME_SEGMENT, it appears that at higher values for number of seasons, the messaging duration is predictably high, but at lower number of seasons, the total volume of observations is more predictive of the duration. Because the performance is much more poor for higher level of messaging duration, NUM_SEASONS_OF_YEAR attribute is not a very strong predictor. The performance is much better at medium and higher levels, implying that the observational volume is a much better predictor. Following the CPTs for the conditional dependance chain starting bottom up from TOTAL_EST_OBS_SEGMENT, it appears that the joint probability of note and chart observational volume is high, and their conditional dependance on number of caregivers is also high. The weakest link in this chain is between number of caregivers and the SAPS score, which may be driving the overall performance of the predictor.

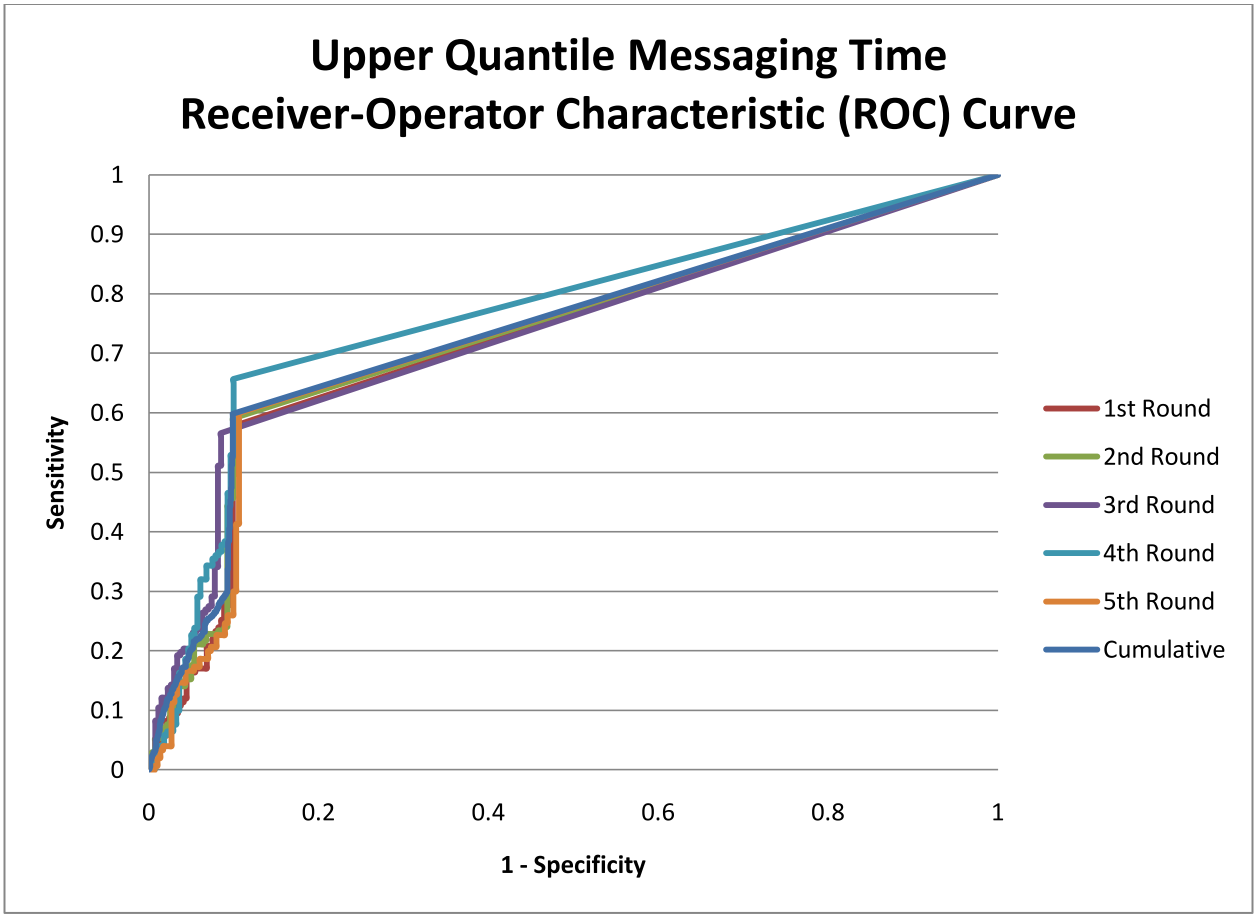


Figure 55: ROC Curve for Prediction of High Messaging Time

Quantile	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)	True Positive Rate (TPR - Sensitivity)	False Positive Rate (FPR)	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Area Under ROC Curve (AUROC)
low	436	141	1353	335	0.565499351	0.09437751	0.755632582	0.801540284	0.737929235
medium	854	194	935	282	0.751760563	0.171833481	0.814885496	0.768282662	0.794835109
high	499	141	1290	335	0.598321343	0.098532495	0.7796875	0.793846154	0.514305189

Table 13: Results of prediction of messaging duration on test dataset.

Model Analysis

The network in Figure 52 appears to be fragmented. Most of the patient conditions are at the left end of the network and are separated by the expiration and trauma attributes from the rest of the network. The location variables at the center are influencing attributes related to number and distribution of certain types of caregiver specialists. The right hand side of the network is dominated by the number of observations by type and number of total nurses and respiratory therapists, which also is the place for our transmission metric of interest, the messaging time.

The parent node, SAPS_SEGMENT, appears to drive most of the network on the center and right hand sides. Examining the CPTs, however, the associations are not as strong. The poor prediction performance of the model, in comparison to the other models, suggests that the relationships within the model are not conclusive.

Transmission Load

Based on the univariate and correlation analyses, the target variable chosen for this metric is the trinomial variable corresponding to the message load, also labeled MSG_LOAD_SEGMENT in **Error! Reference source not found.**.

Belief Network

Not included in the network model in **Error! Reference source not found.** are the other transmission metrics, LOG_MSG_COUNT_SEGMENT and MSG_LOAD_SEGMENT, as the messaging load was assessed on the basis of the explanatory variables alone. The previous two network models have included all three transmission metrics to try to elicit the relationship between these metrics. Since transmission duration and volume may have a dominant influence over transmission load (the metric is derived from these other two), they have been excluded from the construction of this model. In addition, univariate analysis shows different variables associated with messaging load and these other two metrics, inciting further investigation into the influence of the explanatory variables by themselves.

As in the other two models for transmission duration and volume, features are clustered around the object to which they are attributed, forming cliques. Like these other models, load is predominantly influenced by the type of clinical record, including notes and charts. Unlike the other two models, however, the care-unit location has an influence on the rate at which charts are entered. Caregivers have an indirect impact on load through the type of clinical record being used. The patient's condition has a marginal effect upon load through the caregiver.

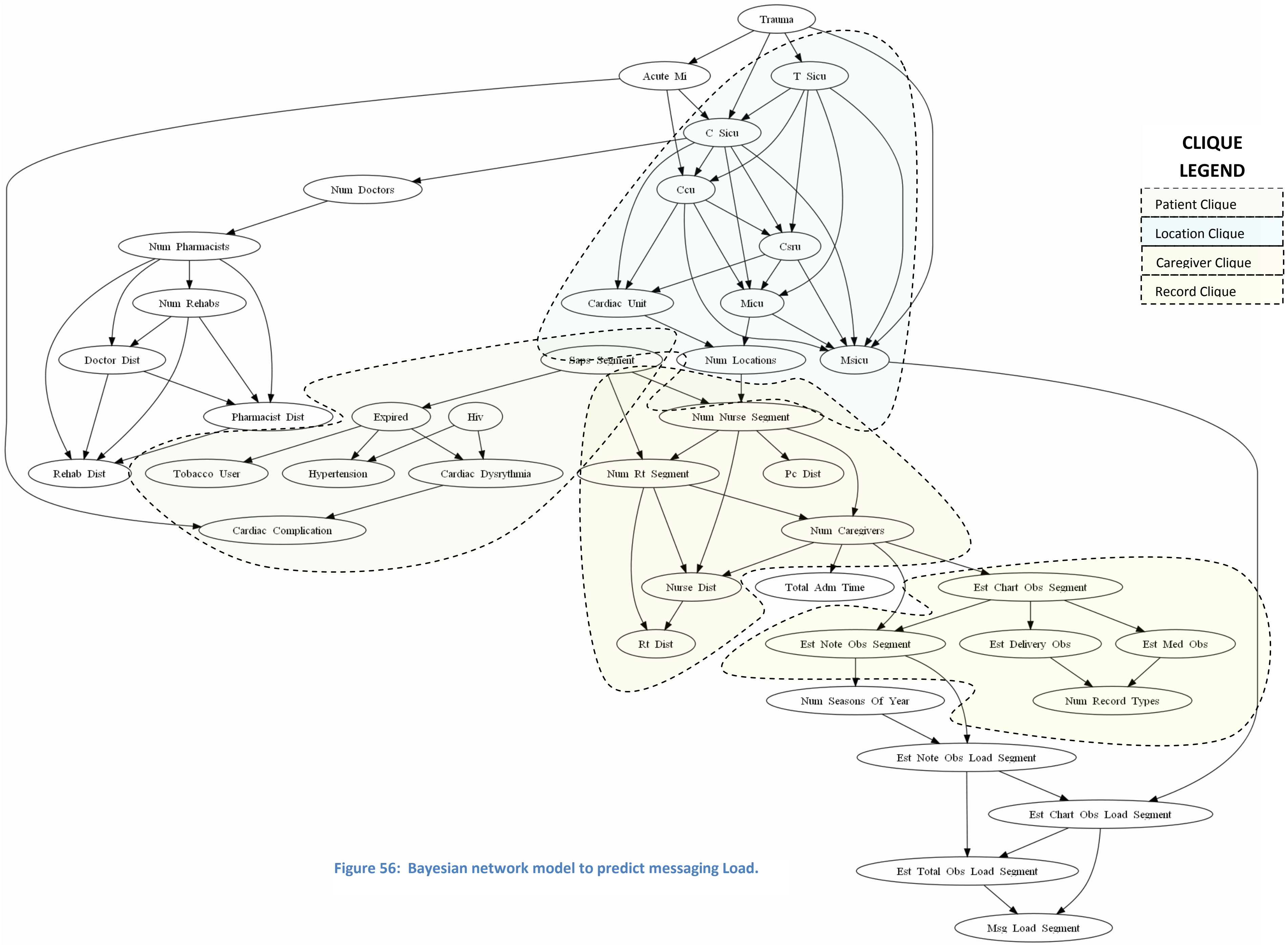


Figure 56: Bayesian network model to predict messaging Load.

Prediction Performance

The prediction performance for the model in **Error! Reference source not found.** on the test dataset is evaluated using ROC analysis from Figure 57 to Figure 59 and the performance metrics described in Table 14.

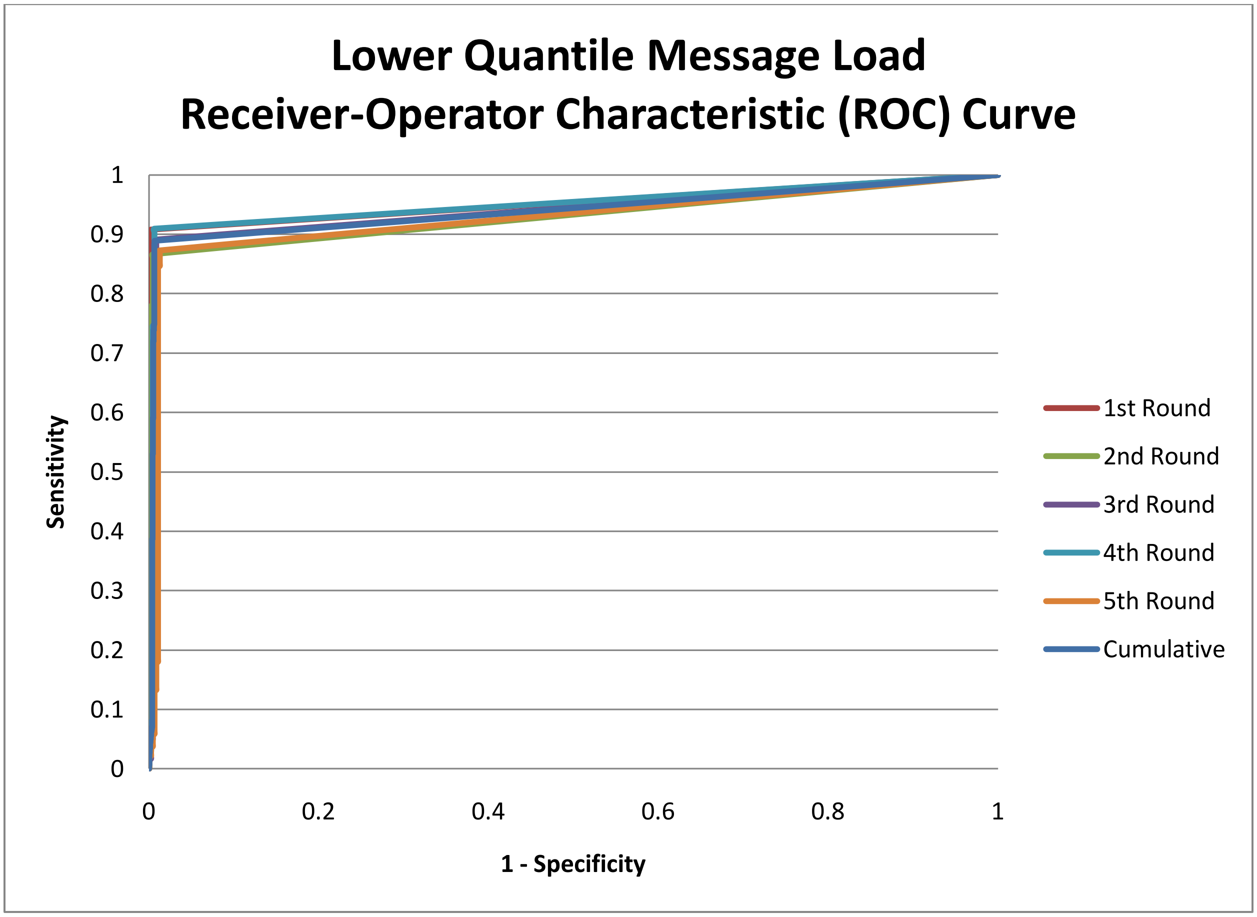


Figure 57: ROC Curve for Prediction of Low Messaging Load

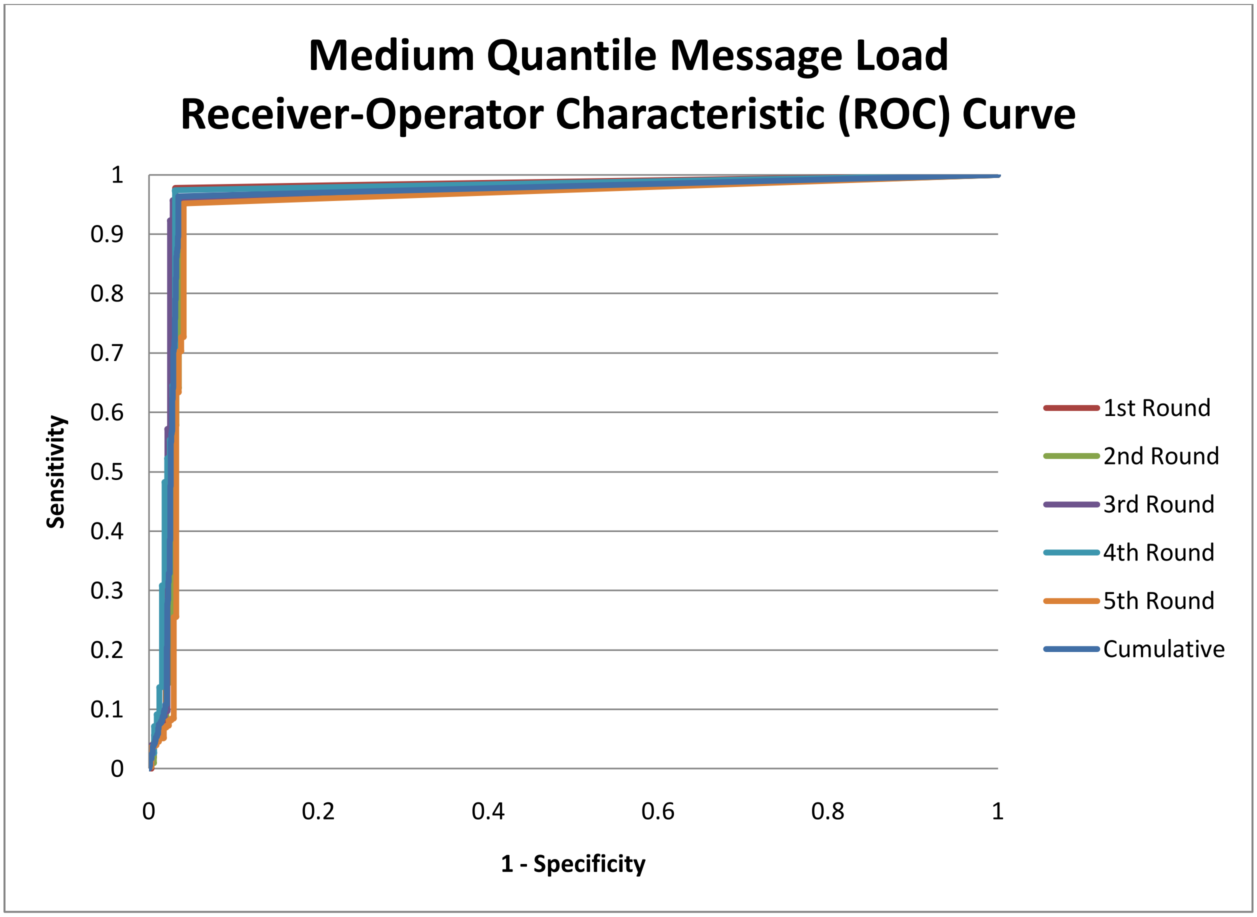


Figure 58: ROC Curve for Prediction of Medium Messaging Load

Like the other two transmission metrics, ROC curves are generated for five partitions of the test dataset, as well as all of the samples in the dataset combined. The curves indicate good performance for predicting the lowest quantile of messaging load, better performance for the upper quantile, and the best performance for the medium value. As in the case for messaging count, the slight improvement in performance for the medium level may be attributed to the larger bin size.

The results in Table 14 show overall good performance for all three levels, with very low false positive rates, but slightly higher false negative rates. Like the model for messaging volume, due to the class reference method chosen for this ROC analysis, the prediction performance may have some class bias.

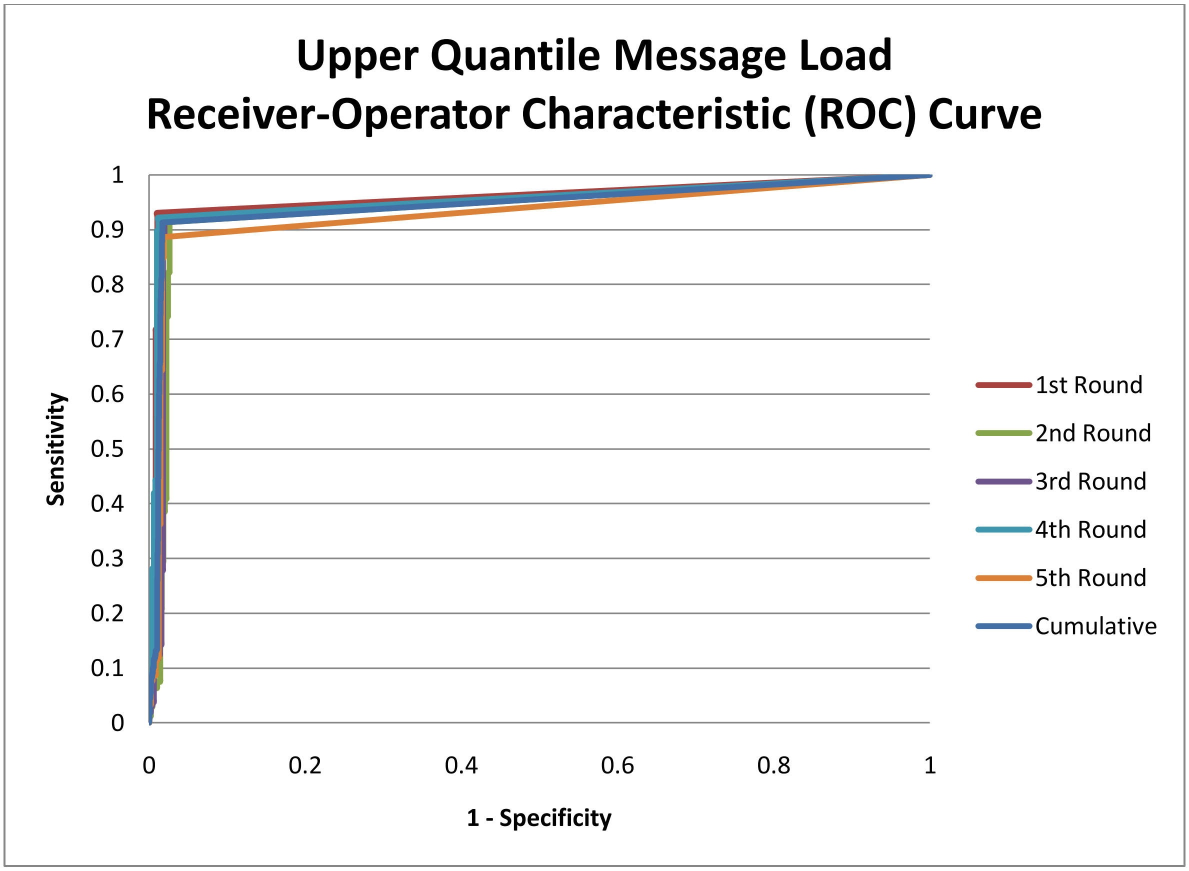


Figure 59: ROC Curve for Prediction of High Messaging Load

Quantile	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)	True Positive Rate (TPR - Sensitivity)	False Positive Rate (FPR)	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Area Under ROC Curve (AUROC)
low	827	16	2430	103	0.889247312	0.006541292	0.981020166	0.959336755	0.940494465
medium	1630	58	1627	61	0.963926671	0.034421365	0.96563981	0.963862559	0.957464987
high	800	43	2457	76	0.913242009	0.0172	0.948991696	0.969996052	0.945381279

Table 14: Results of prediction of messaging load on test dataset.

Model Analysis

As the model illustration shows, messaging load is strongly influenced by other transmission load metrics, such as estimated total (EST_TOTAL_OBS_LOAD_SEGMENT) and chart observations load (EST_CHART_OBS_LOAD_SEGMENT). The CPT for MSG_LOAD_SEGMENT shows stronger influence by the chart observations load. The estimated chart observation load is influenced by the estimated notes observation load (EST_NOTE_OBS_LOAD_SEGMENT), targeting the lower and medium values over the upper value for prediction. The note and chart loads are primarily influenced by the estimated number of note observations (EST_NOTE_OBS_SEGMENT). These results suggest that the message load is strongly influenced by the transmission of note and chart entries. Estimated note and chart observations are, in turn, directly influenced by the number of caregivers for all classes. This result also corroborates with the previous models.

Beyond the caregiver number and type nodes, the network is partitioned into a location clique and a clique involving certain patient conditions. The node for the SAPS I scores (SAPS_SEGMENT) is a junction point between these cliques. It also influences the number of nurses and respiratory therapists, as it does in the other models.

The MICU and MSICU nodes appear to play an interesting role in the network. Examining the CPTs, it appears that admission of patients in this location type results in a reduced chart observation load. This corroborates with the univariate analysis, which shows an inverse relationship between MICU and MSICU with messaging load.

The topmost parent in the network is the trauma patient condition, which influences the location clique and the patient condition clique through the acute MI node. As evident from their CPTs, these relationships are much weaker in the network than those toward the middle and bottom.

As discussed previously, clique analysis shows that messaging load is directly influenced not only by the type of clinical record used, but also by the location in which the patient is observed. The caregiver and clinical condition of the patient appear to have an indirect impact on the rate of information through these other two objects.

Integrated Model of Transmission

Visualization, univariate, correlation, and Bayesian network analyses have revealed a number of patterns in transmission of clinical information. Several of these patterns appear to reinforce each other and are summarized as follows.

- The number of chart entries by nurses and respiratory technicians appears to have the strongest influence on all transmission metrics, over any other type of caregiver (including physicians and clinical care associates) and record type (including notes and medication lists, which appear to be bounded). The importance of the chart is underscored by this finding.
- At higher acuity levels, the number of transmitting respiratory technicians relative to the number of transmitting nurses is much greater. At SAPS I scores above 18, the number of chart and note entries from the RTs is indicative of the overall transmission volume. Surprisingly, messaging by nurses and other caregivers is reduced for this class of patients. Future research may investigate potential causes for this pattern and determine its impact on patient outcome.
-
- The ICU location also has a strong bearing on the overall transmission rate of clinical information. The medical ICUs (MICU and MSICU) observe a reduced transmission load, whereas cardiac units (esp. CCU) have higher transmission rates. Since the location to which patients are assigned may be due to factors other than clinical condition, it would be interesting to investigate whether the variance in rates is also observed for these “misplaced” patients and whether it has a bearing on their outcome.
- The drivers of transmission duration are not readily apparent from the analyses conducted in this study. Although the admission time may have been the intuitive choice as the prime driver, there appear to be other mediating factors. As such, it is possible that factors other than those cited in this study may have a greater explanatory effect on this metric. Further research is necessary to explore this possibility.

These findings are only applicable to the MIMIC II clinical database and the BIDMC ICU, from where these records were collected.

Conclusions

This thesis proposes a model for the transmission of clinical information in an ICU clinical database. This section summarizes the conclusions it makes based on the results of data visualization and statistical analyses and the limitations of the methodology followed. A summary of the contributions it makes is presented, as well as recommendations for future work.

Summary of Contributions

In the first chapter, an overview of Mediated Agent social interaction theory is described, which provides the foundation for this thesis work. This chapter also describes clinical databases and proposes an object-process view of this technology to show its emergent value-added properties. A brief introduction to Bayesian networks is also included to describe the principles underlying the predictive models discussed in later chapters.

The second chapter proposes a framework for the study of clinical communications, based on the OSI model, used in telecommunications. A message-based interaction event model is also offered as a model for the lowest layers of this seven-layered architecture.

The third chapter describes the methods and tools used to implement the transmission aspect of the Interaction Event model. The transmission model is implemented on a corpus extracted from the MIMIC II clinical database. The process of extracting messaging data, summarizing it into tables, and preparing the data for subsequent analysis is also described. Techniques for data visualization and univariate logistic regression are presented as a descriptive approach to determine the key variables for transmission of clinical information. Correlation analysis and data modeling using Bayesian networks are also presented as a predictive approach to determine how the explanatory variables affect the transmission metrics.

The fourth chapter details the results of the methods applied from the previous chapter on the clinical datasets. The results of descriptive analysis show that volume, load, and duration are informative response variables in the transmission of clinical information. They also determine the primary set of explanatory variables associated with these transmission metrics. The results of the predictive analysis show that each metric is influenced by a common set of explanatory variables, including the number of charts entries and the distribution of caregivers. Explanatory variables specifically influencing transmission volume are number of nurses and respiratory therapists, which appear to have a distributional relationship for patients with higher acuity scores. Those variables specifically influencing the transmission load are the care-unit locations, which are lower in medical units in comparison to the cardiac units. The models for the transmission duration are found to be inconclusive, as the prediction models are not as accurate as those for the volume and load.

Limitations of Study

This study and the methodology it has followed are based on several assumptions, which are described as follows:

- Messaging data used for creating and analyzing the models of transmission has not been collected for the exact purpose of studying clinical communications. For instance, it does not contain direct evidence of transmission and reception of clinical information by caregivers. In the ideal dataset, each message would be associated with an acknowledgement of receipt by caregiver(s).
- Clinical information occurring only within the clinical database is the subject of this study. Other methods of communication including face-to-face, e-mail, telephone, and other such conversations are not supported by this thesis.
- Only the transmission aspect of clinical communications is investigated, as the chosen dataset is limited to only have evidence showing transmission events and does not contain data for reception of this information.
- Clinical records only provided within the dataset are used to model transmission. Records that are not distributed may be other sources of transmission, not accounted for in this research.
- Only the objects and attributes provided by the clinical database comprise the set of variables explored by this study. Availability of more types of information in the database may have expanded the scope of these features and resulted in different models.
- Adult patients, over the age of fifteen and under the age of ninety, have been selected as the cohort for this study. Models of transmission do not support those for newborns which comprise the other subpopulation in the ICU clinical database.
- Assumptions have been made of the titles specified in the caregiver dictionary, regarding the job function and experience of the role associated with the title. Caregivers for which no titles have been assigned in this dictionary have been removed from this study.
- Attributes are forcefully eliminated from the synthesized model based on univariate association. This process may have left out critical variables that influence the response variables only in concert with others. The benefits of this multivariate approach are not available to this study.
- Only the Bayesian network modeling technique, using Bayesware software package, was used to create a predictive model for multivariate analysis, although other techniques could have been utilized to perform similar analysis.
- Influence and association between variables are only predicted by the models used in this study. Causation or dependence relationships cannot be drawn.

Future Work

During the course of this work, several questions were raised which merit further investigation. Foremost, the findings of this study should be corroborated with evidence from primary source data collected for the purpose of studying clinical communications, including transmission of clinical information verified with evidence of reception of this information. In addition, it was found that the prediction models for transmission duration were not accurate enough to draw conclusions about this metric's association with the other variables. This gap may require a larger set of attributes, larger datasets to incorporate significant sample sizes for these attributes, and better strategies for prediction.

This study has determined statistical associations between the variables of interest. In order to determine causal and dependency relationships between these variables, temporal analyses may be performed on the datasets used in this study. To corroborate the prediction models proposed by the Bayesian techniques used in this work, other classification methods may be used including multivariate logistic regression, support vector machines, neural networks, and tree-based methods.

To expand on clinical communications research based on this study, the Interaction Event model must be assessed using both transmission and reception of messages, which requires time-stamped logs of when clinical records are read. Future work may also include messaging information on other channels, such as logs of telephone conversations, pager alerts, and e-mail.

Appendix A: Federal Guideline for Meaningful Use of EMRs related to Clinical Communication

Health outcomes policy priority	Care goals	Stage 1 objectives		Stage 1 measures
		Eligible professionals	Hospitals	
Improve care coordination.	Exchange meaningful clinical information among professional healthcare team.	Capability to exchange key clinical information (for example, problem list, medication list, allergies, diagnostic test results), among providers of care and patient authorized entities electronically.	Capability to exchange key clinical information (for example, discharge summary, procedures, problem list, medication list, allergies, diagnostic test results), among providers of care and patient authorized entities electronically.	Performed at least one test of certified EHR technology's capacity to electronically exchange key clinical information.
		Perform medication reconciliation at relevant encounters and each transition of care.	Perform medication reconciliation at relevant encounters and each transition of care.	Perform medication reconciliation for at least 80% of relevant encounters and transitions of care.
		Provide summary care record for each transition of care and referral.	Provide summary care record for each transition of care and referral.	Provide summary of care record for at least 80% of transitions of care and referrals.

Table 15: Federal Guidelines for Meaningful Use of EMRs Related to Clinical Communication (HHS, Proposed Rules for Electronic Health Record Incentive Program, 2010)

Appendix B: Glossary of Clinical Communication Terms

Term	Definition
agent	An object (either human or non-human) that is capable of interacting with another.
interaction	An exchange of messages between agents.
transmitting agent	the sender of a message
receiving agent	the receiver of a message
message	An object that encapsulates information that comprises of message content and a message header.
message content	Portion of the message that consists of information being shared by the transmitting and receiving agents (also called the payload).
message header	Portion of the message that consists of information required to pass the message between the transmitting and receiving agents across the channel.
communication	Process of sharing information and knowledge between agents.
channel	Medium upon which information is passed.
communication mode	Describes who are the receiving agent(s) of a message – must be either one of unicast, multicast, or broadcast.
unicast	Mode in which there is a single receiving agent.
multicast	Mode in which there are multiple receiving agents.
broadcast	Mode in which all agents are receiving agents.
information	Un-interpreted data.
background information	Contextual information required to interpret new information.
new information	Message content that has never been shared between agents.
data (plural of datum)	A set of observations.
model	Cognitive apparatus of an agent that interprets and stores information as knowledge.

knowledge	Structured information that is the result of interpretation of information.
common ground	Knowledge shared between agents.
grounding	Process of verifying common ground between agents.
grounding efficiency	Informational cost of grounding measured as the ratio between sizes of background information to new information.
ground-positive	Grounding efficiency when size of new information is reduced with inclusion of background information.
ground-negative	Grounding efficiency when size of new information is increased with inclusion of background information.
ground-neutral	Grounding efficiency when size of new information is unchanged with inclusion of background information.
Law of Mediated Centre	Agents seek the minimal grounding cost over a series of interactions.
clinical information	Clinical data in either structured or unstructured form.
clinical data (plural of clinical datum)	A set of clinical observations.
clinical observation	Recording of a value for an attribute of a patient at a specific instant in time.
clinical record	Clinical data corresponding to a common attribute.
clinical database	Collection of clinical records.
clinical knowledgebase	Clinical information that has been interpreted by clinicians for mutual understanding.
medical record system	A set of objects and processes intended to make clinical information accessible, interpretable (as clinical knowledge), and communicable to providers and consumers in a healthcare system.
paper-based medical record system	A medical record system in which the clinical database and knowledgebase is stored in paper form.
electronic Medical Record (EMR) System	A medical record system in which the clinical database and knowledgebase is stored and processed electronically (i.e., on a computing platform).

Appendix C: Caregiver Role Dictionary

TITLE	DESCRIPTION	EXPERIENCE	ROLE
	Unknown	Unknown	Unknown
NursUnk	Unknown Nurse		Nurse
DocUnk	Unknown Doctor		Doctor
TechUnk	Unknown Technician		Technician
PhaUnk	Unknown Pharmacist		Pharmacist
SWUnk	Unknown Social Worker		Social_Worker
AdmUnk	Unknown Administrator		Administrator
AL	Assisted Living??		Unknown
AR	Account Representative??		Unknown
Admin	Administrator		Administrator
BSMT	Bachelor of Science in Medical Technology	Junior	MedTech
CCP	Certified in Clinical Perfusion (??)	Experienced	Nurse
CCRN	Critical Care Registered Nurse	Experienced	Nurse
CPhT	Certified Pharmacology Technician (??)	Senior	Pharmacist
CRA	Clinical Research Associate	Junior	Associate
CRS	Clinical Research Social Worker	Senior	Social_Worker
CRT	Certified Respiratory Therapist	Experienced	RT
CO-Op	Cooperative Student	Student	Nurse
CoOPSt	Cooperative Student	Student	Nurse
CoOPst	Cooperative Student	Student	Nurse
CoOpSt	Cooperative Student	Student	Nurse
CoOpst	Cooperative Student	Student	Nurse
Coopst	Cooperative Student	Student	Nurse
CoopStu	Cooperative Student	Student	Nurse
Co-Wk	Cooperative Worker	Junior	PC
Co-Wkr	Cooperative Worker	Junior	PC
Co-Wor	Cooperative Worker	Junior	PC
Co-wkr	Cooperative Worker	Junior	PC
Co-wor	Cooperative Worker	Junior	PC
CoWker	Cooperative Worker	Junior	PC
Cowkr	Cooperative Worker	Junior	PC
CoWkr	Cooperative Worker	Junior	PC
CoWork	Cooperative Worker	Junior	PC
Cowork	Cooperative Worker	Junior	PC
coWker	Cooperative Worker	Junior	PC
cowker	Cooperative Worker	Junior	PC
Co-Ord	Care Coordinator	Junior	PC

Coor	Care Coordinator	Junior	PC
Coord	Care Coordinator	Junior	PC
Coordi	Care Coordinator	Junior	PC
CsMngm	Customer Service Management	Experienced	Administrator
D	Doctor	Junior	Doctor
DI	Doctor (intern)	Junior	Doctor
DInter	Doctor (Intern)	Junior	Doctor
DML		Senior	Unknown
DO	Doctor of Osteopathy	Senior	Doctor
DRM	Medicine Doctor	Junior	Doctor
DietIn	Dietician Intern	Junior	Dietician
DirHCQ	Director Health Care Quality	Experienced	Administrator
DR	Physician or Scientist	Senior	Doctor
Dr	Physician or Scientist	Senior	Doctor
H	Medical Technologist in Hematology(??)	Senior	MedTech
HMS	Harvard Medical Student	Student	Doctor
HMSIV	Harvard Medical Student	Student	Doctor
I.S.		Senior	Technician
IS		Senior	Technician
ISCS		Senior	Technician
ISOPS		Senior	Technician
Isops		Senior	Technician
ISSUPP		Senior	Technician
ISSupp		Senior	Technician
Intern	Intern Resident	Junior	Doctor
LCP	Laboratory for Computational Phsyio	Experienced	Researcher
LICSW	Licensed Clinical Social Worker	Experienced	Social_Worker
LPN	Licensed Practical Nurse	Experienced	Nurse
LS		Experienced	Unknown
MD	Doctor of Medicine	Senior	Doctor
MD,PhD	Doctor of Medicine, Doctor of Philosophy	Experienced	Doctor
md	Doctor of Medicine	Senior	Doctor
Md	Doctor of Medicine	Senior	Doctor
MDS	Master of Dental Surgery	Senior	Doctor
MDs	Master of Dental Surgery	Senior	Doctor
MFD		Senior	Unknown
Ms	Medical Student	Student	Doctor
ms	Medical Student	Student	Doctor
MS	Medical Student	Student	Doctor
MS,RD	Master of Science, Registered Dietician	Experienced	Dietician
MSII	2nd year medical student	Student	Doctor

MSIV	4th year medical student	Student	Doctor
MSV	5th year medical student	Student	Doctor
MSWint	Master of Social Work	Junior	Social_Worker
MTASCP	Medical Technologist (ASCP specialty)	Experienced	MedTech
Med		Junior	Unknown
Med.		Junior	Unknown
MedSt	Medical Student	Student	Doctor
MedSt.	Medical Student	Student	Doctor
MedStu	Medical Student	Student	Doctor
MEDST	Medical student	Student	Doctor
Medst	Medical Student	Student	Doctor
MedSty		Student	Doctor
Medica		Student	Doctor
NA	Nursing Assistant	Junior	Associate
na	Nursing Assistant	Junior	Associate
NEOB	Neonatal/Obstetrician	Experienced	Doctor
NNP	Neonatal Nurse Practitionner	Senior	Nurse
nnp	Neonatal Nurse Practitionner	Senior	Nurse
NPS	Nurse Practionner	Senior	Nurse
NSV		Senior	Nurse
Np	Nurse Practionner	Senior	Nurse
NP	Nurse Practionner	Senior	Nurse
NsgSt	Nursing Student	Student	Nurse
Nurs	Nurse	Junior	Nurse
OTR/L	Registered Occupational Therapist/Licensed	Experienced	Rehab
PA	Physician Assistant	Junior	Associate
PA-C	Physician Assistant-Certified	Junior	Associate
PC	Patient Care	Junior	PC
PCA	Patient Care Assistant	Junior	PC
PCT	Patient Care Technician (certified)	Senior	PC
PHD	Scientist	Experienced	Unknown
PHaD		Experienced	Pharmacist
PS		Senior	Unknown
PT	Physical Therapist	Senior	Rehab
PTA	Physical Therapist Assistant/Attending	Junior	Rehab
Par		Senior	Unknown
Ph	Pharmacy	Junior	Pharmacist
Ph.Stu	Pharmacy Student	Student	Pharmacist
PhD	Pharmacy Doctorate	Experienced	Pharmacist
PhaD	Pharmacy Doctorate	Experienced	Pharmacist
PharmD	Pharmacy Doctorate	Experienced	Pharmacist

PhStud	Pharmacy Student	Student	Pharmacist
PhaStu	Pharmacy Student	Student	Pharmacist
PrADM		Experienced	Administrator
PracSt		Student	Nurse
Prog		Junior	Administrator
RA		Junior	Nurse
RD		Senior	Nurse
RD,LDN		Senior	Nurse
RD/LDN		Senior	Nurse
RHP		Senior	Nurse
RN,RPh	Registered Nurse, Pharmacist	Experienced	Nurse
RNBA		Experienced	Nurse
RNC	Registered Nurse Certified	Experienced	Nurse
RNCM		Experienced	Nurse
RNStu	Registered Nurse Student	Student	Nurse
RNs	Registered Nurse	Student	Nurse
RPHS	Registered Pharmacist	Senior	Pharmacist
RRT	Registered Respiratory Therapist	Experienced	RT
RRt	Registered Respiratory Therapist	Experienced	RT
rrt	Registered Respiratory Therapist	Experienced	RT
RRTs	Registered Respiratory Therapist	Student	RT
RRts	Registered Respiratory Therapist	Student	RT
RT	Respiratory Therapist	Senior	RT
RTS	Respiratory Therapist	Student	RT
rts	Respiratory Therapist	Student	RT
RTSt	Respiratory Therapist	Student	RT
RTStu	Respiratory Therapist Student	Student	RT
ReAssi		Junior	Doctor
Res	Resident	Junior	Doctor
res	Resident	Junior	Doctor
Reside	Resident	Junior	Doctor
ReschA		Junior	Doctor
RN	Registered Nurse	Senior	Nurse
Rn	Registered Nurse	Senior	Nurse
rn	Registered Nurse	Senior	Nurse
Rph	Registered Pharmacist	Senior	Pharmacist
RPh	Registered Pharmacist	Senior	Pharmacist
RPH	Registered Pharmacist	Senior	Pharmacist
R.Ph	Registered Pharmacist	Senior	Pharmacist
R.Ph.	Registered Pharmacist	Senior	Pharmacist
SN	Skilled Nurse	Experienced	Nurse

SNNP	Skilled Neonatal Nurse Practitionner	Experienced	Nurse
SNP	Skilled/Specialized Nurse Practitionner	Experienced	Nurse
SPG		Senior	Unknown
SRN	Student Registered Nurse	Student	Nurse
SRT	Student Respiratory Therapist	Student	RT
STN		Senior	Nurse
SW	Social Worker	Senior	Social_Worker
SWInt	Social Worker Intern	Junior	Social_Worker
StPHa	Pharmacy Student	Student	Pharmacist
StPh	Pharmacy Student	Student	Pharmacist
StResp	Respiratory Student	Student	RT
St	Student	Student	Nurse
Stu	Student	Student	Nurse
StNRS	Nursing Student	Student	Nurse
StNuIV	IV Nursing Student	Student	Nurse
StNur	Nursing Student	Student	Nurse
StNurs	Nursing Student	Student	Nurse
Stn	Nursing Student	Student	Nurse
StuNur	Nursing Student	Student	Nurse
StuRN	Student RN	Student	Nurse
StRN	Student RN	Student	Nurse
Studen	Student	Student	Nurse
studen	Student	Student	Nurse
U	Unit	Junior	Unknown
UA	Unit A	Junior	PC
UC	Unit C	Junior	PC
UCO		Junior	PC
UCO/PC	/Patient Care	Senior	PC
Unit A	Unit A	Junior	Unknown
VS		Junior	Unknown
ajm		Junior	Unknown
cda		Junior	Unknown
taf		Junior	Unknown
tap		Junior	Unknown

Appendix D: MySQL Commands to Construct Interaction Event Table

```
DROP TABLE IF EXISTS mimic2_interaction_event_table;

CREATE TABLE mimic2_interaction_event_table (
  msg_id   serial,
  tx_agent bigint NOT NULL,
  rx_agent bigint NOT NULL,
  location bigint NOT NULL,
  patient_id bigint NOT NULL,
  stime    datetime NOT NULL,
  etime    datetime NOT NULL,
  record_type ENUM ('note','report','totalbal','med','io','delivery','chart','additive') NOT NULL,
  record_id   bigint NOT NULL);

DROP TABLE IF EXISTS patient_00002_interaction_events;

CREATE TABLE patient_00002_interaction_events (
  msg_id   serial,
  tx_agent bigint NOT NULL,
  rx_agent bigint NOT NULL,
  location bigint NOT NULL,
  patient_id bigint NOT NULL,
  stime    datetime NOT NULL,
  etime    datetime NOT NULL,
  record_type ENUM ('note','report','totalbal','med','io','delivery','chart','additive') NOT NULL,
  record_id   bigint NOT NULL);

DROP TABLE IF EXISTS patient_00002_ADDITIVES_interaction_events;

CREATE TABLE patient_00002_ADDITIVES_interaction_events (
  msg_id   bigint,
  tx_agent bigint NOT NULL,
  rx_agent bigint NOT NULL,
  location bigint NOT NULL,
  patient_id bigint NOT NULL,
  stime    datetime NOT NULL,
  etime    datetime NOT NULL,
  record_type ENUM ('note','report','totalbal','med','io','delivery','chart','additive') NOT NULL,
  record_id   bigint NOT NULL)

SELECT

cgid AS tx_agent,
-65535 AS rx_agent,
```

```
cuid AS location,
subject_id AS patient_id,
chart_time AS stime,
chart_time AS etime,
'additive' AS record_type,
item_id AS record_id
FROM patient_00002_ADDITIVES ORDER BY stime;

INSERT INTO mimic2_interaction_event_table SELECT * FROM
patient_00002_ADDITIVES_interaction_events WHERE (patient_id != 0);

INSERT INTO patient_00002_interaction_events SELECT * FROM
patient_00002_ADDITIVES_interaction_events WHERE (patient_id != 0);

DROP TABLE IF EXISTS patient_00002_CHARTEVENTS_interaction_events;

CREATE TABLE patient_00002_CHARTEVENTS_interaction_events (
  msg_id   bigint,
  tx_agent bigint NOT NULL,
  rx_agent bigint NOT NULL,
  location bigint NOT NULL,
  patient_id bigint NOT NULL,
  stime    datetime NOT NULL,
  etime    datetime NOT NULL,
  record_type ENUM ('note','report','totalbal','med','io','delivery','chart','additive') NOT NULL,
  record_id   bigint NOT NULL)
  SELECT
cgid AS tx_agent,
-65535 AS rx_agent,
cuid AS location,
subject_id AS patient_id,
chart_time AS stime,
chart_time AS etime,
'chart' AS record_type,
item_id AS record_id
FROM patient_00002_CHARTEVENTS ORDER BY stime;

INSERT INTO mimic2_interaction_event_table SELECT * FROM
patient_00002_CHARTEVENTS_interaction_events WHERE (patient_id != 0);

INSERT INTO patient_00002_interaction_events SELECT * FROM
patient_00002_CHARTEVENTS_interaction_events WHERE (patient_id != 0);

DROP TABLE IF EXISTS patient_00002_DELIVERIES_interaction_events;
```

```
CREATE TABLE patient_00002_DELIVERIES_interaction_events (
  msg_id   bigint,
  tx_agent bigint NOT NULL,
  rx_agent bigint NOT NULL,
  location bigint NOT NULL,
  patient_id bigint NOT NULL,
  stime    datetime NOT NULL,
  etime    datetime NOT NULL,
  record_type ENUM ('note','report','totalbal','med','io','delivery','chart','additive') NOT NULL,
  record_id   bigint NOT NULL)
  SELECT
cgid AS tx_agent,
-65535 AS rx_agent,
cuid AS location,
subject_id AS patient_id,
chart_time AS stime,
chart_time AS etime,
'delivery' AS record_type,
io_item_id AS record_id
FROM patient_00002_DELIVERIES ORDER BY stime;

INSERT INTO mimic2_interaction_event_table SELECT * FROM
patient_00002_DELIVERIES_interaction_events WHERE (patient_id != 0);

INSERT INTO patient_00002_interaction_events SELECT * FROM
patient_00002_DELIVERIES_interaction_events WHERE (patient_id != 0);

DROP TABLE IF EXISTS patient_00002_IOEVENTS_interaction_events;

CREATE TABLE patient_00002_IOEVENTS_interaction_events (
  msg_id   bigint,
  tx_agent bigint NOT NULL,
  rx_agent bigint NOT NULL,
  location bigint NOT NULL,
  patient_id bigint NOT NULL,
  stime    datetime NOT NULL,
  etime    datetime NOT NULL,
  record_type ENUM ('note','report','totalbal','med','io','delivery','chart','additive') NOT NULL,
  record_id   bigint NOT NULL)
  SELECT
cgid AS tx_agent,
-65535 AS rx_agent,
cuid AS location,
```

```
subject_id AS patient_id,
chart_time AS stime,
chart_time AS etime,
'io' AS record_type,
item_id AS record_id
FROM patient_00002_IOEVENTS ORDER BY stime;

INSERT INTO mimic2_interaction_event_table SELECT * FROM
patient_00002_IOEVENTS_interaction_events WHERE (patient_id != 0);

INSERT INTO patient_00002_interaction_events SELECT * FROM
patient_00002_IOEVENTS_interaction_events WHERE (patient_id != 0);

DROP TABLE IF EXISTS patient_00002_MEDEVENTS_interaction_events;

CREATE TABLE patient_00002_MEDEVENTS_interaction_events (
  msg_id   bigint,
  tx_agent bigint NOT NULL,
  rx_agent bigint NOT NULL,
  location bigint NOT NULL,
  patient_id bigint NOT NULL,
  stime    datetime NOT NULL,
  etime    datetime NOT NULL,
  record_type ENUM ('note','report','totalbal','med','io','delivery','chart','additive') NOT NULL,
  record_id   bigint NOT NULL)
  SELECT
cgid AS tx_agent,
-65535 AS rx_agent,
cuid AS location,
subject_id AS patient_id,
chart_time AS stime,
chart_time AS etime,
'med' AS record_type,
item_id AS record_id
FROM patient_00002_MEDEVENTS ORDER BY stime;

INSERT INTO mimic2_interaction_event_table SELECT * FROM
patient_00002_MEDEVENTS_interaction_events WHERE (patient_id != 0);

INSERT INTO patient_00002_interaction_events SELECT * FROM
patient_00002_MEDEVENTS_interaction_events WHERE (patient_id != 0);

DROP TABLE IF EXISTS patient_00002_TOTALBALEVENTS_interaction_events;
```

```
CREATE TABLE patient_00002_TOTALBALEVENTS_interaction_events (
  msg_id   bigint,
  tx_agent bigint NOT NULL,
  rx_agent bigint NOT NULL,
  location bigint NOT NULL,
  patient_id bigint NOT NULL,
  stime    datetime NOT NULL,
  etime    datetime NOT NULL,
  record_type ENUM ('note','report','totalbal','med','io','delivery','chart','additive') NOT NULL,
  record_id   bigint NOT NULL)
  SELECT
cgid AS tx_agent,
-65535 AS rx_agent,
cuid AS location,
subject_id AS patient_id,
chart_time AS stime,
chart_time AS etime,
'totalbal' AS record_type,
item_id AS record_id
FROM patient_00002_TOTALBALEVENTS ORDER BY stime;

INSERT INTO mimic2_interaction_event_table SELECT * FROM
patient_00002_TOTALBALEVENTS_interaction_events WHERE (patient_id != 0);

INSERT INTO patient_00002_interaction_events SELECT * FROM
patient_00002_TOTALBALEVENTS_interaction_events WHERE (patient_id != 0);

DROP TABLE IF EXISTS patient_00002_NOTEEVENTS_interaction_events;

CREATE TABLE patient_00002_NOTEEVENTS_interaction_events (
  msg_id   bigint,
  tx_agent bigint NOT NULL,
  rx_agent bigint NOT NULL,
  location bigint NOT NULL,
  patient_id bigint NOT NULL,
  stime    datetime NOT NULL,
  etime    datetime NOT NULL,
  record_type ENUM ('note','report','totalbal','med','io','delivery','chart','additive') NOT NULL,
  record_id   bigint NOT NULL)
  SELECT
cgid AS tx_agent,
-65535 AS rx_agent,
cuid AS location,
```

```
subject_id AS patient_id,
chart_time AS stime,
chart_time AS etime,
'note' AS record_type,
note_id AS record_id
FROM patient_00002_NOTEEVENTS ORDER BY stime;

INSERT INTO mimic2_interaction_event_table SELECT * FROM
patient_00002_NOTEEVENTS_interaction_events WHERE (patient_id != 0);

INSERT INTO patient_00002_interaction_events SELECT * FROM
patient_00002_NOTEEVENTS_interaction_events WHERE (patient_id != 0);

DROP TABLE IF EXISTS patient_00002_REPORTEVENTS_interaction_events;

CREATE TABLE patient_00002_REPORTEVENTS_interaction_events (
  msg_id   bigint,
  tx_agent bigint NOT NULL,
  rx_agent bigint NOT NULL,
  location bigint NOT NULL,
  patient_id bigint NOT NULL,
  stime    datetime NOT NULL,
  etime    datetime NOT NULL,
  record_type ENUM ('note','report','totalbal','med','io','delivery','chart','additive') NOT NULL,
  record_id   bigint NOT NULL)
  SELECT
-1 AS tx_agent,
-65535 AS rx_agent,
-1 AS location,
subject_id AS patient_id,
report_dt AS stime,
report_dt AS etime,
'report' AS record_type,
report_id AS record_id
FROM patient_00002_REPORTEVENTS ORDER BY stime;

INSERT INTO mimic2_interaction_event_table SELECT * FROM
patient_00002_REPORTEVENTS_interaction_events WHERE (patient_id != 0);

INSERT INTO patient_00002_interaction_events SELECT * FROM
patient_00002_REPORTEVENTS_interaction_events WHERE (patient_id != 0);

DROP TABLE IF EXISTS patient_00002_NOTEINTERACTIONS_interaction_events;
```

```
CREATE TABLE patient_00002_NOTEINTERACTIONS_interaction_events (
  msg_id   bigint,
  tx_agent bigint NOT NULL,
  rx_agent bigint NOT NULL,
  location bigint NOT NULL,
  patient_id bigint NOT NULL,
  stime    datetime NOT NULL,
  etime    datetime NOT NULL,
  record_type ENUM ('note','report','totalbal','med','io','delivery','chart','additive') NOT NULL,
  record_id   bigint NOT NULL)
  SELECT
tx_agent AS tx_agent,
rx_agent AS rx_agent,
location AS location,
patient_id AS patient_id,
stime AS stime,
etime AS etime,
record_type AS record_type,
med_id AS record_id
FROM patient_00002_NOTEINTERACTIONS ORDER BY stime;

INSERT INTO mimic2_interaction_event_table SELECT * FROM
patient_00002_NOTEINTERACTIONS_interaction_events WHERE (patient_id != 0);

INSERT INTO patient_00002_interaction_events SELECT * FROM
patient_00002_NOTEINTERACTIONS_interaction_events WHERE (patient_id != 0);

DROP TABLE IF EXISTS patient_00003_interaction_events;

CREATE TABLE patient_00003_interaction_events (
  msg_id   serial,
  tx_agent bigint NOT NULL,
  rx_agent bigint NOT NULL,
  location bigint NOT NULL,
  patient_id bigint NOT NULL,
  stime    datetime NOT NULL,
  etime    datetime NOT NULL,
  record_type ENUM ('note','report','totalbal','med','io','delivery','chart','additive') NOT NULL,
  record_id   bigint NOT NULL);
```

Appendix E: ICD-9 Categories with Codes and Predominant Conditions

Code	ICD-9 range	Predominant Conditions in training/validation data-set	Description of Conditions in Range
Y1	001-139	Sepsis, HIV, Hepatitis	Infectious and Parasitic Diseases
Y2	140-239	-	Neoplasms
Y3	240-279	Diabetes, Hypercholesterolemia	Endocrine, Nutritional, and Metabolic Diseases and Immunity Disorders
Y4	280-289	Anemia	Diseases of the Blood and Blood-forming organs
Y5	290-319	Alcoholism, Tobacco Use, Depression	Mental Diseases and Conditions
Y6	320-389	Alzheimer’s	Diseases of the Nervous System and Sense Organs
Y7	401-405	Chronic Hypertension	Hypertensive Disease
Y8	410-414	Acute Myocardial Infarction, Atherosclerosis	Ischemic Heart Disease
Y9	428	Congestive Heart Failure	Heart Failure
Y10	420-429	Cardiac dysrythmia	Other forms of Heart Disease
Y11	430-438	-	Cerebrovascular Disease
Y12	440-449	-	Diseases of Arteries, Arterioles, and Capillaries
Y13	390-459	-	Other diseases of the Circulatory System
Y14	460-519	Pneumonia, Acute Respiratory Failure	Pulmonary disorders and conditions
Y15	520-579		Digestive System
Y16	580-629	Acute Renal Failure, Urinary Tract Infection	Genitourinary system
Y17	630-679	-	Pregnancy and Birth
Y18	680-709	-	Skin disorders and conditions
Y19	710-739	-	Musculoskeletal system
Y20	740-759	-	Congenital Anomalies
Y21	760-779	-	Certain Conditions originating in the Perinatal Period
Y22	780-799	-	Symptoms, Signs, and ill-defined conditions
Y23	800-960	Trauma	Injuries including fractures, burns, wounds, contusions, traumatic conditions
Y24	960-966	-	Other Injuries and Poisons
Y25	996-999	Sepsis, Cardiac Complication	Other Complications, Causes of Injuries and Poisons
Y26	E800-E999	-	External Causes of Injury and Poisoning
Y27	V01-V89	-	Supplementary Classification of Factors Influencing Health Status and Contact with Health Services

Appendix F: Selection of Features for Messaging Prediction Models

Object	Feature Attribute	Attribute Values	Description
Patient	Expired	True, False	Mortality status at time of discharge.
	Num_conditions	4 equal bins from 0 to 50	Number of ICD-9 codes assigned to patient.
	Alcoholic	True, False	ICD-9 code: 305.1
	Hypertension	True, False	ICD-9 code: 401.9
	Diabetes	True, False	ICD-9 codes: 250.00 – 250.92
	Hypercholesterolemia	True, False	ICD-9 code: 272.0
	Atherosclerosis	True, False	ICD-9 code: 414.01
	Anemia	True, False	ICD-9 codes: 280.0 – 286.0
	HIV	True, False	ICD-9 code: 042
	Trauma	True, False	ICD-9 codes: 800.09 – 959.09
	Hepatitis	True, False	ICD-9 codes: 070.1 – 070.71
	Acute_MI	True, False	ICD-9 codes: 410.01 – 411.0
	Acute_renal_failure	True, False	ICD-9 codes: 583.9 – 585
	Acute_respiratory_failure	True, False	ICD-9 codes: 518.81
	Heart_failure	True, False	ICD-9 code: 428.0
	Cardiac_dysrythmia	True, False	ICD-9 codes: 427.0 – 427.9
	Pneumonia	True, False	ICD-9 code: 486
	Septic	True, False	ICD-9 code: 038.0 – 038.9, 995.92
	Urinary_tract_infection	True, False	ICD-9 code: 599.0
	Trauma	True, False	ICD-9 codes: 800.09 – 959.09
	SAPS_segment	Low, Medium, High	Binned SAPS I scores: 0-9, 10-13, 14-17, 18-35
Caregiver	Num_caregivers	4 bins: 1-6, 7-10, 11-21, 22-202	Number of caregivers per patient
	Num_social_workers	4 equal bins from 0 to 10	Number of social workers
	Num_nurse_segment	Low, Medium, High	Number of nurses
	Num_rehabs	4 equal bins from 0 to 10	Number of rehabilitation specialists (physical and occupational therapists).
	Num_RT_segment	Low, Medium, High	Binned number of respiratory therapists
	Num_pharmacists	4 equal bins from 0 to 10	Number of pharmacists

	Num_doctors	4 equal bins from 0 to 10	Number of doctors
	Num_associates	4 equal bins from 0.0 to 1.0	Number of associates
	Num_pcs	4 equal bins from 0.0 to 1.0	Number of patient care associates.
	Nurse_dist	4 equal bins from 0.0 to 1.0	Fraction of caregivers as nurses
	RT_dist	4 equal bins from 0.0 to 1.0	Fraction of caregivers as respiratory therapists.
	Social_worker_dist	4 equal bins from 0.0 to 1.0	Fraction of caregivers as social workers.
	Doctor_dist	4 equal bins from 0.0 to 1.0	Fraction of caregivers as doctors.
	PC_dist	4 equal bins from 0.0 to 1.0	Fraction of caregivers as patient care associates.
	Pharmacist_dist	4 equal bins from 0.0 to 1.0	Fraction of caregivers as pharmacists.
	Rehab_dist	4 equal bins from 0.0 to 1.0	Fraction of caregivers as rehabilitation specialists.
	Associate_dist	4 equal bins from 0.0 to 1.0	Fraction of caregivers as associates.
Care-unit	T_SICU	True, False	Trauma Surgical Care Unit
	C_SICU	True, False	Cardiac Surgical Care Unit
	MICU	True, False	Medical Intensive Care Unit
	MSICU	True, False	Medical Surgical Intensive Care Unit
	CCU	True, False	Coronary Care Unit
	CSRU	True, False	Cardiac Surgery Recovery Unit
	Cardiac_unit	True, False	Either CCU, CSRU, or CSICU.
	Num_locations	4 equal bins from 0 to 15	Number of care-units visited.
Care-time	Total_Adm_time	4 bins: 24 – 96, 97-192, 193-312, 313-2184	Total Length of Stay during admission visit.
	Num_seasons_of_year	1-4	Number of seasons of year.
	Num_days_of_week	1-7	Number of days of week.
	Log_msg_time_segment	Low, Medium, High	Binned Logarithm of total messaging time.
Clinical Record	Est_Note_obs_segment	Low, Medium, High	Binned number of estimated note observations
	Est_Chart_obs_segment	Low, Medium, High	Binned number of estimated chart observations

	Est_Delivery_obs	4 bins: 0-1, 2-3, 4-5, 6-282	Number of estimated delivery observations
	Est_Med_obs	4 bins: 0-7, 8-45, 46-166, 167-18498	Number of estimated medication observations
	Est_Additive_obs	4 bins: 0-1, 2-3, 4-8, 9-264	Number of estimated additive observations
	Est_IO_obs	4 bins: 0-62, 63-164, 165-472, 473-4196	Number of estimated Input/Output observations.
	Est_Total_obs_segment	Low, Medium, High	Total number of estimated observations.
	Num_record_types	4 equal bins from 0 to 10	Number of record types.
	Log_msg_count_segment	Low, Medium, High	Binned logarithm of number of messages.
	Est_note_obs_load_segment	Low, Medium, High	Binned estimated number of note observations per hour.
	Est_chart_obs_load_segment	Low, Medium, High	Binned number of chart observations per hour.
	Est_total_obs_load_segment	Low, Medium, High	Binned total number of observations per hour.
	msg_load_segment	Low, Medium, High	Binned total number of messages per hour.

Appendix G: Conditional Probability Tables for Message Volume Prediction

probability (_Log_Msg_Count_Segment_ | _Est_Io_Obs_, _Est_Chart_Obs_Segment_)
{
(0, 0): 0.200, 0.200, 0.200;
(0, 1): 1.000, 0.000, 0.000;
(0, 2): 0.154, 0.845, 0.000;
(0, 3): 0.200, 0.200, 0.200;
(0, 4): 0.200, 0.200, 0.200;
(1, 0): 0.200, 0.200, 0.200;
(1, 1): 0.930, 0.069, 0.000;
(1, 2): 0.000, 1.000, 0.000;
(1, 3): 0.200, 0.200, 0.200;
(1, 4): 0.200, 0.200, 0.200;
(2, 0): 0.200, 0.200, 0.200;
(2, 1): 0.010, 0.962, 0.010;
(2, 2): 0.000, 1.000, 0.000;
(2, 3): 0.001, 0.091, 0.906;
(2, 4): 0.200, 0.200, 0.200;
(3, 0): 0.200, 0.200, 0.200;
(3, 1): 0.200, 0.200, 0.200;
(3, 2): 0.000, 0.845, 0.154;
(3, 3): 0.000, 0.006, 0.994;
(3, 4): 0.200, 0.200, 0.200;
}

probability (_Est_Io_Obs_ | _Est_Additive_Obs_, _Num_Caregivers_)
{
(0, 0): 0.861, 0.139, 0.000, 0.000;
(0, 1): 0.545, 0.394, 0.061, 0.000;
(0, 2): 0.312, 0.250, 0.437, 0.001;
(0, 3): 0.005, 0.005, 0.985, 0.005;
(1, 0): 0.766, 0.233, 0.000, 0.000;
(1, 1): 0.360, 0.520, 0.120, 0.000;
(1, 2): 0.000, 0.475, 0.525, 0.000;
(1, 3): 0.001, 0.001, 0.634, 0.363;
(2, 0): 0.605, 0.394, 0.000, 0.000;
(2, 1): 0.085, 0.633, 0.282, 0.000;
(2, 2): 0.013, 0.210, 0.617, 0.161;
(2, 3): 0.000, 0.057, 0.245, 0.698;

```
   (3, 0):  0.003, 0.662, 0.332, 0.003;
   (3, 1):  0.001, 0.200, 0.748, 0.051;
   (3, 2):  0.000, 0.065, 0.630, 0.304;
   (3, 3):  0.000, 0.000, 0.069, 0.930;
}


probability (_Est_Chart_Obs_Segment_ | _Num_Nurse_Segment_, _Est_Io_Obs_)
{
   (0, 0):  0.006, 0.976, 0.006, 0.006, 0.006;
   (0, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 0):  0.000, 0.953, 0.046, 0.000, 0.000;
   (1, 1):  0.000, 0.359, 0.640, 0.000, 0.000;
   (1, 2):  0.002, 0.002, 0.992, 0.002, 0.002;
   (1, 3):  0.012, 0.012, 0.953, 0.012, 0.012;
   (2, 0):  0.000, 0.622, 0.377, 0.000, 0.000;
   (2, 1):  0.000, 0.049, 0.951, 0.000, 0.000;
   (2, 2):  0.000, 0.007, 0.945, 0.048, 0.000;
   (2, 3):  0.000, 0.000, 0.548, 0.451, 0.000;
   (3, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 1):  0.012, 0.012, 0.953, 0.012, 0.012;
   (3, 2):  0.000, 0.000, 0.881, 0.118, 0.000;
   (3, 3):  0.000, 0.000, 0.051, 0.949, 0.000;
}

probability (_Est_Additive_Obs_ | _Csru_, _Num_Caregivers_)
{
   (0, 0):  0.459, 0.393, 0.139, 0.008;
   (0, 1):  0.200, 0.521, 0.270, 0.009;
   (0, 2):  0.122, 0.290, 0.495, 0.094;
   (0, 3):  0.021, 0.072, 0.319, 0.587;
   (1, 0):  0.326, 0.245, 0.326, 0.102;
   (1, 1):  0.119, 0.179, 0.476, 0.226;
   (1, 2):  0.040, 0.119, 0.368, 0.473;
   (1, 3):  0.010, 0.040, 0.220, 0.729;
}

probability (_Num_Caregivers_ | _Num_Rt_Segment_, _Num_Nurse_Segment_)
{
   (0, 0):  0.956, 0.015, 0.015, 0.015;
```

```
   (0, 1):  0.905, 0.095, 0.000, 0.000;
   (0, 2):  0.131, 0.628, 0.241, 0.000;
   (0, 3):  0.001, 0.001, 0.461, 0.537;
   (1, 0):  0.250, 0.250, 0.250, 0.250;
   (1, 1):  0.250, 0.250, 0.250, 0.250;
   (1, 2):  0.250, 0.250, 0.250, 0.250;
   (1, 3):  0.250, 0.250, 0.250, 0.250;
   (2, 0):  0.956, 0.015, 0.015, 0.015;
   (2, 1):  0.521, 0.435, 0.044, 0.000;
   (2, 2):  0.000, 0.401, 0.581, 0.018;
   (2, 3):  0.000, 0.000, 0.295, 0.704;
   (3, 0):  0.250, 0.250, 0.250, 0.250;
   (3, 1):  0.003, 0.398, 0.596, 0.003;
   (3, 2):  0.000, 0.000, 0.555, 0.444;
   (3, 3):  0.000, 0.000, 0.000, 1.000;
}

probability (_Num_Rt_Segment_ | _Saps_Segment_, _Num_Nurse_Segment_)
{
   (0, 0):  0.250, 0.250, 0.250, 0.250;
   (0, 1):  0.250, 0.250, 0.250, 0.250;
   (0, 2):  0.250, 0.250, 0.250, 0.250;
   (0, 3):  0.250, 0.250, 0.250, 0.250;
   (1, 0):  0.956, 0.015, 0.015, 0.015;
   (1, 1):  0.873, 0.000, 0.127, 0.000;
   (1, 2):  0.747, 0.000, 0.217, 0.036;
   (1, 3):  0.217, 0.001, 0.434, 0.348;
   (2, 0):  0.250, 0.250, 0.250, 0.250;
   (2, 1):  0.544, 0.000, 0.422, 0.033;
   (2, 2):  0.366, 0.000, 0.530, 0.104;
   (2, 3):  0.092, 0.000, 0.264, 0.643;
   (3, 0):  0.015, 0.015, 0.956, 0.015;
   (3, 1):  0.281, 0.000, 0.655, 0.063;
   (3, 2):  0.091, 0.000, 0.647, 0.261;
   (3, 3):  0.000, 0.000, 0.133, 0.867;
}

probability (_Num_Nurse_Segment_ | _Num_Locations_)
{
   (0):  0.003, 0.336, 0.513, 0.148;
   (1):  0.000, 0.023, 0.372, 0.604;
   (2):  0.001, 0.001, 0.149, 0.850;
```

```
   (3):  0.012, 0.012, 0.012, 0.964;
   (4):  0.012, 0.012, 0.012, 0.964;
   (5):  0.023, 0.023, 0.023, 0.932;
   (6):  0.250, 0.250, 0.250, 0.250;
   (7):  0.250, 0.250, 0.250, 0.250;
   (8):  0.250, 0.250, 0.250, 0.250;
   (9):  0.250, 0.250, 0.250, 0.250;
}

probability (_Num_Locations_ | _C_Sicu_, _Ccu_, _Csru_, _Cardiac_Unit_, _Micu_)
{
   (0, 0, 0, 0, 0):  1.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000;
   (0, 0, 0, 0, 1):  0.892, 0.099, 0.008, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000;
   (0, 0, 0, 1, 0):  0.973, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003;
   (0, 0, 0, 1, 1):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (0, 0, 1, 0, 0):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (0, 0, 1, 0, 1):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (0, 0, 1, 1, 0):  0.930, 0.070, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000;
   (0, 0, 1, 1, 1):  0.000, 0.874, 0.083, 0.042, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000;
   (0, 1, 0, 0, 0):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (0, 1, 0, 0, 1):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (0, 1, 0, 1, 0):  0.976, 0.024, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000;
   (0, 1, 0, 1, 1):  0.000, 0.713, 0.285, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000;
   (0, 1, 1, 0, 0):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (0, 1, 1, 0, 1):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (0, 1, 1, 1, 0):  0.000, 0.886, 0.114, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000;
   (0, 1, 1, 1, 1):  0.001, 0.001, 0.796, 0.001, 0.199, 0.001, 0.001, 0.001, 0.001, 0.001;
   (1, 0, 0, 0, 0):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (1, 0, 0, 0, 1):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (1, 0, 0, 1, 0):  0.957, 0.042, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000;
   (1, 0, 0, 1, 1):  0.000, 0.898, 0.100, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000;
   (1, 0, 1, 0, 0):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (1, 0, 1, 0, 1):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (1, 0, 1, 1, 0):  0.000, 0.907, 0.091, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000;
   (1, 0, 1, 1, 1):  0.000, 0.000, 0.747, 0.125, 0.125, 0.000, 0.000, 0.000, 0.000, 0.000;
   (1, 1, 0, 0, 0):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (1, 1, 0, 0, 1):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (1, 1, 0, 1, 0):  0.000, 0.996, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000;
   (1, 1, 0, 1, 1):  0.003, 0.003, 0.973, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003;
   (1, 1, 1, 0, 0):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (1, 1, 1, 0, 1):  0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100;
   (1, 1, 1, 1, 0):  0.002, 0.002, 0.986, 0.002, 0.002, 0.002, 0.002, 0.002, 0.002, 0.002;
```

```
   (1, 1, 1, 1, 1):  0.003, 0.003, 0.003, 0.003, 0.003, 0.973, 0.003, 0.003, 0.003, 0.003;
}

probability (_Cardiac_Unit_ | _C_Sicu_, _Ccu_, _Csru_)
{
   (0, 0, 0):  0.995, 0.005;
   (0, 0, 1):  0.000, 1.000;
   (0, 1, 0):  0.000, 1.000;
   (0, 1, 1):  0.001, 0.999;
   (1, 0, 0):  0.001, 0.999;
   (1, 0, 1):  0.003, 0.997;
   (1, 1, 0):  0.008, 0.992;
   (1, 1, 1):  0.020, 0.980;
}

probability (_Micu_ | _Septic_)
{
   (0):  0.782, 0.218;
   (1):  0.567, 0.433;
}

probability (_Csru_ | _Ccu_, _C_Sicu_, _Micu_)
{
   (0, 0, 0):  0.296, 0.704;
   (0, 0, 1):  0.834, 0.166;
   (0, 1, 0):  0.865, 0.135;
   (0, 1, 1):  0.555, 0.445;
   (1, 0, 0):  0.743, 0.257;
   (1, 0, 1):  0.735, 0.265;
   (1, 1, 0):  0.774, 0.226;
   (1, 1, 1):  0.500, 0.500;
}

probability (_Ccu_ | _Acute_Renal_Failure_, _C_Sicu_, _Micu_)
{
   (0, 0, 0):  0.684, 0.316;
   (0, 0, 1):  0.864, 0.136;
   (0, 1, 0):  0.890, 0.110;
   (0, 1, 1):  0.996, 0.004;
   (1, 0, 0):  0.408, 0.592;
   (1, 0, 1):  0.947, 0.053;
   (1, 1, 0):  0.993, 0.007;
```

```
   (1, 1, 1):  0.598, 0.402;
}

probability (_C_Sicu_ | _Heart_Failure_)
{
   (0):  0.834, 0.166;
   (1):  0.908, 0.092;
}


probability (_Acute_Renal_Failure_ | _Heart_Failure_, _Septic_)
{
   (0, 0):  0.910, 0.090;
   (0, 1):  0.619, 0.381;
   (1, 0):  0.734, 0.266;
   (1, 1):  0.383, 0.617;
}
probability (_Heart_Failure_ | _Septic_)
{
   (0):  0.781, 0.219;
   (1):  0.649, 0.351;
}
probability (_Septic_ | _Saps_Segment_)
{
   (0):  0.500, 0.500;
   (1):  0.913, 0.087;
   (2):  0.894, 0.106;
   (3):  0.789, 0.211;
}
probability (_Saps_Segment_)
{
   5.902285291912783E-10, 0.2480000000047218, 0.47999999945698973, 0.2719999999480599;
}
```

Appendix H: Conditional Probability Tables for Message Duration Prediction

```
probability (_Saps_Segment_)
{
   0.24266682742288284, 0.47733307801463715, 0.2800000945624801;
}

probability (_Expired_ | _Saps_Segment_)
{
   (0):  0.967, 0.033;
   (1):  0.919, 0.081;
   (2):  0.738, 0.262;
}

probability (_Num_Nurse_Segment_ | _Nurse_Dist_, _Num_Caregivers_)
{
   (0, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 1):  0.001, 0.997, 0.001, 0.001, 0.001;
   (0, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 3):  0.001, 0.001, 0.997, 0.001, 0.001;
   (1, 0):  0.000, 1.000, 0.000, 0.000, 0.000;
   (1, 1):  0.000, 1.000, 0.000, 0.000, 0.000;
   (1, 2):  0.000, 0.238, 0.762, 0.000, 0.000;
   (1, 3):  0.000, 0.000, 0.250, 0.750, 0.000;
   (2, 0):  0.000, 1.000, 0.000, 0.000, 0.000;
   (2, 1):  0.000, 0.440, 0.560, 0.000, 0.000;
   (2, 2):  0.000, 0.000, 0.927, 0.073, 0.000;
   (2, 3):  0.000, 0.000, 0.040, 0.960, 0.000;
   (3, 0):  0.000, 0.843, 0.157, 0.000, 0.000;
   (3, 1):  0.000, 0.000, 1.000, 0.000, 0.000;
   (3, 2):  0.000, 0.000, 0.757, 0.243, 0.000;
   (3, 3):  0.000, 0.000, 0.000, 1.000, 0.000;
}

probability (_Num_Rt_Segment_ | _Rt_Dist_, _Num_Caregivers_)
{
   (0, 0):  0.953, 0.000, 0.047, 0.000, 0.000;
   (0, 1):  0.710, 0.000, 0.290, 0.000, 0.000;
   (0, 2):  0.319, 0.000, 0.681, 0.000, 0.000;
   (0, 3):  0.059, 0.000, 0.456, 0.485, 0.000;
   (1, 0):  0.000, 0.000, 1.000, 0.000, 0.000;
   (1, 1):  0.000, 0.000, 1.000, 0.000, 0.000;
   (1, 2):  0.000, 0.000, 0.560, 0.440, 0.000;
   (1, 3):  0.000, 0.000, 0.000, 1.000, 0.000;
   (2, 0):  0.000, 0.000, 1.000, 0.000, 0.000;
   (2, 1):  0.000, 0.000, 0.800, 0.200, 0.000;
   (2, 2):  0.000, 0.000, 0.000, 1.000, 0.000;
```

```
   (2, 3):  0.000, 0.000, 0.000, 1.000, 0.000;
   (3, 0):  0.001, 0.001, 0.997, 0.001, 0.001;
   (3, 1):  0.001, 0.001, 0.001, 0.997, 0.001;
   (3, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 3):  0.001, 0.001, 0.001, 0.997, 0.001;
}

probability (_Num_Caregivers_ | _Saps_Segment_)
{
   (0):  0.401, 0.319, 0.159, 0.121;
   (1):  0.190, 0.299, 0.285, 0.226;
   (2):  0.076, 0.186, 0.329, 0.410;
}

probability (_Nurse_Dist_ | _Num_Caregivers_, _Rt_Dist_)
{
   (0, 0):  0.000, 0.000, 0.079, 0.921;
   (0, 1):  0.000, 0.056, 0.389, 0.556;
   (0, 2):  0.000, 0.364, 0.636, 0.000;
   (0, 3):  0.001, 0.997, 0.001, 0.001;
   (1, 0):  0.000, 0.000, 0.145, 0.855;
   (1, 1):  0.000, 0.055, 0.891, 0.055;
   (1, 2):  0.000, 0.400, 0.600, 0.000;
   (1, 3):  0.997, 0.001, 0.001, 0.001;
   (2, 0):  0.000, 0.000, 0.416, 0.584;
   (2, 1):  0.000, 0.173, 0.773, 0.053;
   (2, 2):  0.000, 0.667, 0.333, 0.000;
   (2, 3):  0.250, 0.250, 0.250, 0.250;
   (3, 0):  0.000, 0.000, 0.574, 0.426;
   (3, 1):  0.000, 0.214, 0.777, 0.009;
   (3, 2):  0.000, 1.000, 0.000, 0.000;
   (3, 3):  0.997, 0.001, 0.001, 0.001;
}

probability (_Rt_Dist_ | _Saps_Segment_)
{
   (0):  0.841, 0.132, 0.027, 0.000;
   (1):  0.587, 0.360, 0.050, 0.003;
   (2):  0.395, 0.510, 0.086, 0.010;
}

probability (_Est_Additive_Obs_ | _Num_Caregivers_)
{
   (0):  0.382, 0.433, 0.166, 0.019;
   (1):  0.172, 0.377, 0.373, 0.078;
   (2):  0.100, 0.215, 0.435, 0.250;
   (3):  0.021, 0.063, 0.233, 0.683;
}
```

```
probability (_Est_Chart_Obs_Segment_ | _Est_Note_Obs_Segment_, _Est_Io_Obs_)
{
   (0, 0):  0.000, 1.000, 0.000, 0.000, 0.000;
   (0, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 0):  0.000, 0.947, 0.053, 0.000, 0.000;
   (1, 1):  0.000, 0.283, 0.717, 0.000, 0.000;
   (1, 2):  0.000, 0.000, 1.000, 0.000, 0.000;
   (1, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 0):  0.000, 0.574, 0.426, 0.000, 0.000;
   (2, 1):  0.000, 0.127, 0.873, 0.000, 0.000;
   (2, 2):  0.000, 0.000, 0.961, 0.039, 0.000;
   (2, 3):  0.000, 0.000, 0.280, 0.720, 0.000;
   (3, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 2):  0.000, 0.000, 0.778, 0.222, 0.000;
   (3, 3):  0.000, 0.000, 0.030, 0.970, 0.000;
   (4, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
}

probability (_Est_Delivery_Obs_ | _Num_Caregivers_)
{
   (0):  0.255, 0.446, 0.274, 0.025;
   (1):  0.255, 0.279, 0.382, 0.083;
   (2):  0.080, 0.185, 0.460, 0.275;
   (3):  0.005, 0.042, 0.143, 0.810;
}

probability (_Est_Io_Obs_ | _Est_Additive_Obs_, _Num_Caregivers_)
{
   (0, 0):  0.867, 0.133, 0.000, 0.000;
   (0, 1):  0.571, 0.400, 0.029, 0.000;
   (0, 2):  0.200, 0.450, 0.300, 0.050;
   (0, 3):  0.250, 0.000, 0.250, 0.500;
   (1, 0):  0.735, 0.265, 0.000, 0.000;
   (1, 1):  0.390, 0.519, 0.091, 0.000;
   (1, 2):  0.047, 0.419, 0.488, 0.047;
   (1, 3):  0.000, 0.000, 0.750, 0.250;
   (2, 0):  0.692, 0.308, 0.000, 0.000;
   (2, 1):  0.118, 0.618, 0.263, 0.000;
   (2, 2):  0.011, 0.218, 0.701, 0.069;
   (2, 3):  0.000, 0.000, 0.182, 0.818;
   (3, 0):  0.000, 1.000, 0.000, 0.000;
```

```
    (3, 1):  0.000, 0.125, 0.875, 0.000;
    (3, 2):  0.000, 0.020, 0.700, 0.280;
    (3, 3):  0.000, 0.000, 0.031, 0.969;
}

probability (_Est_Med_Obs_ | _Est_Additive_Obs_, _Est_Io_Obs_)
{
    (0, 0):  0.896, 0.104, 0.000, 0.000;
    (0, 1):  0.806, 0.194, 0.000, 0.000;
    (0, 2):  0.875, 0.000, 0.125, 0.000;
    (0, 3):  0.666, 0.000, 0.000, 0.333;
    (1, 0):  0.415, 0.585, 0.000, 0.000;
    (1, 1):  0.316, 0.579, 0.105, 0.000;
    (1, 2):  0.243, 0.270, 0.459, 0.027;
    (1, 3):  0.200, 0.000, 0.600, 0.200;
    (2, 0):  0.179, 0.714, 0.107, 0.000;
    (2, 1):  0.068, 0.446, 0.473, 0.014;
    (2, 2):  0.022, 0.157, 0.697, 0.124;
    (2, 3):  0.024, 0.000, 0.381, 0.595;
    (3, 0):  0.250, 0.250, 0.250, 0.250;
    (3, 1):  0.000, 0.333, 0.667, 0.000;
    (3, 2):  0.000, 0.038, 0.566, 0.396;
    (3, 3):  0.007, 0.000, 0.065, 0.928;
}

probability (_Est_Note_Obs_Segment_ | _Est_Io_Obs_, _Num_Caregivers_)
{
    (0, 0):  0.042, 0.725, 0.233, 0.000, 0.000;
    (0, 1):  0.000, 0.458, 0.542, 0.000, 0.000;
    (0, 2):  0.000, 0.000, 1.000, 0.000, 0.000;
    (0, 3):  0.001, 0.001, 0.997, 0.001, 0.001;
    (1, 0):  0.000, 0.432, 0.568, 0.000, 0.000;
    (1, 1):  0.000, 0.330, 0.670, 0.000, 0.000;
    (1, 2):  0.000, 0.064, 0.936, 0.000, 0.000;
    (1, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
    (2, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
    (2, 1):  0.000, 0.119, 0.857, 0.024, 0.000;
    (2, 2):  0.000, 0.016, 0.854, 0.130, 0.000;
    (2, 3):  0.000, 0.000, 0.545, 0.455, 0.000;
    (3, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
    (3, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
    (3, 2):  0.000, 0.000, 0.565, 0.435, 0.000;
    (3, 3):  0.000, 0.000, 0.072, 0.928, 0.000;
}

probability (_Total_Est_Obs_Segment_ | _Est_Note_Obs_Segment_, _Est_Chart_Obs_Segment_)
{
    (0, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
```

```
   (0, 1):  0.000, 1.000, 0.000, 0.000, 0.000;
   (0, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 1):  0.000, 1.000, 0.000, 0.000, 0.000;
   (1, 2):  0.000, 0.137, 0.863, 0.000, 0.000;
   (1, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 1):  0.000, 0.768, 0.232, 0.000, 0.000;
   (2, 2):  0.000, 0.003, 0.997, 0.000, 0.000;
   (2, 3):  0.000, 0.000, 0.333, 0.667, 0.000;
   (2, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 2):  0.000, 0.000, 0.692, 0.308, 0.000;
   (3, 3):  0.000, 0.000, 0.000, 1.000, 0.000;
   (3, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
}

probability (_Log_Msg_Count_Segment_ | _Est_Io_Obs_, _Est_Chart_Obs_Segment_)
{
   (0, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 1):  0.000, 1.000, 0.000, 0.000, 0.000;
   (0, 2):  0.000, 0.057, 0.943, 0.000, 0.000;
   (0, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 1):  0.000, 0.844, 0.156, 0.000, 0.000;
   (1, 2):  0.000, 0.000, 1.000, 0.000, 0.000;
   (1, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 2):  0.000, 0.000, 1.000, 0.000, 0.000;
   (2, 3):  0.000, 0.000, 0.250, 0.750, 0.000;
   (2, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 2):  0.000, 0.000, 0.917, 0.083, 0.000;
   (3, 3):  0.000, 0.000, 0.000, 1.000, 0.000;
   (3, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
```

```
}

probability (_Num_Seasons_Of_Year_ | _Est_Io_Obs_)
{
   (0):  0.000, 0.989, 0.011, 0.000, 0.000;
   (1):  0.000, 0.952, 0.048, 0.000, 0.000;
   (2):  0.000, 0.866, 0.134, 0.000, 0.000;
   (3):  0.000, 0.550, 0.370, 0.058, 0.021;
}

probability (_Total_Adm_Time_ | _Est_Io_Obs_)
{
   (0):  0.471, 0.316, 0.155, 0.059;
   (1):  0.187, 0.476, 0.225, 0.112;
   (2):  0.080, 0.305, 0.337, 0.278;
   (3):  0.011, 0.127, 0.233, 0.630;
}

probability (_Log_Msg_Time_Segment_ | _Num_Seasons_Of_Year_, _Total_Est_Obs_Segment_)
{
   (0, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 1):  0.000, 0.780, 0.220, 0.000, 0.000;
   (1, 2):  0.000, 0.114, 0.806, 0.080, 0.000;
   (1, 3):  0.000, 0.000, 0.382, 0.618, 0.000;
   (1, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 1):  0.000, 0.500, 0.500, 0.000, 0.000;
   (2, 2):  0.000, 0.031, 0.406, 0.562, 0.000;
   (2, 3):  0.000, 0.000, 0.042, 0.958, 0.000;
   (2, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 3):  0.000, 0.000, 0.000, 1.000, 0.000;
   (3, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 3):  0.000, 0.000, 0.000, 1.000, 0.000;
   (4, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
}
```

Appendix I: Conditional Probability Tables for Message Load Prediction

```
probability (_Msg_Load_Segment_ | _Est_Total_Obs_Load_Segment_,
_Est_Chart_Obs_Load_Segment_)
{
   (0, 0):  0.008, 0.969, 0.008, 0.008, 0.008;
   (0, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 0):  0.003, 0.989, 0.003, 0.003, 0.003;
   (1, 1):  0.001, 0.991, 0.008, 0.000, 0.000;
   (1, 2):  0.000, 0.133, 0.867, 0.000, 0.000;
   (1, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 1):  0.000, 0.880, 0.120, 0.000, 0.000;
   (2, 2):  0.000, 0.003, 0.984, 0.012, 0.000;
   (2, 3):  0.000, 0.000, 0.178, 0.822, 0.000;
   (2, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 2):  0.000, 0.000, 0.829, 0.171, 0.000;
   (3, 3):  0.000, 0.000, 0.024, 0.976, 0.000;
   (3, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (4, 4):  0.008, 0.008, 0.008, 0.008, 0.969;
}
probability (_Est_Total_Obs_Load_Segment_ | _Est_Note_Obs_Load_Segment_,
_Est_Chart_Obs_Load_Segment_)
{
   (0, 0):  0.332, 0.661, 0.003, 0.003, 0.003;
   (0, 1):  0.000, 1.000, 0.000, 0.000, 0.000;
   (0, 2):  0.000, 0.411, 0.587, 0.000, 0.000;
   (0, 3):  0.000, 0.000, 0.545, 0.455, 0.000;
   (0, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 0):  0.008, 0.969, 0.008, 0.008, 0.008;
   (1, 1):  0.000, 1.000, 0.000, 0.000, 0.000;
   (1, 2):  0.000, 0.276, 0.724, 0.000, 0.000;
   (1, 3):  0.000, 0.000, 0.734, 0.266, 0.000;
   (1, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 1):  0.000, 0.743, 0.257, 0.000, 0.000;
   (2, 2):  0.000, 0.014, 0.984, 0.002, 0.000;
   (2, 3):  0.000, 0.000, 0.295, 0.705, 0.000;
```

```
    (2, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
    (3, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
    (3, 1):  0.000, 0.048, 0.952, 0.000, 0.000;
    (3, 2):  0.000, 0.000, 0.675, 0.325, 0.000;
    (3, 3):  0.000, 0.000, 0.000, 1.000, 0.000;
    (3, 4):  0.008, 0.008, 0.008, 0.008, 0.969;
    (4, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
    (4, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
    (4, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
    (4, 3):  0.008, 0.008, 0.008, 0.969, 0.008;
    (4, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
}
probability (_Est_Chart_Obs_Load_Segment_ | _Msicu_, _Est_Note_Obs_Load_Segment_)
{
    (0, 0):  0.053, 0.475, 0.280, 0.193, 0.000;
    (0, 1):  0.000, 0.700, 0.276, 0.023, 0.000;
    (0, 2):  0.000, 0.117, 0.602, 0.281, 0.000;
    (0, 3):  0.000, 0.023, 0.515, 0.461, 0.001;
    (0, 4):  0.018, 0.018, 0.018, 0.927, 0.018;
    (1, 0):  0.000, 0.885, 0.110, 0.005, 0.000;
    (1, 1):  0.007, 0.888, 0.104, 0.000, 0.000;
    (1, 2):  0.000, 0.234, 0.752, 0.015, 0.000;
    (1, 3):  0.000, 0.053, 0.667, 0.281, 0.000;
    (1, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
}
probability (_Est_Note_Obs_Load_Segment_ | _Num_Seasons_Of_Year_, _Est_Note_Obs_Segment_)
{
    (0, 0):  1.000, 0.000, 0.000, 0.000, 0.000;
    (0, 1):  0.000, 0.342, 0.525, 0.132, 0.001;
    (0, 2):  0.000, 0.090, 0.627, 0.283, 0.000;
    (0, 3):  0.000, 0.106, 0.407, 0.486, 0.000;
    (0, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
    (1, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
    (1, 1):  0.001, 0.544, 0.363, 0.091, 0.001;
    (1, 2):  0.000, 0.682, 0.262, 0.056, 0.000;
    (1, 3):  0.000, 0.571, 0.236, 0.193, 0.000;
    (1, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
    (2, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
    (2, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
    (2, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
    (2, 3):  0.000, 0.923, 0.058, 0.019, 0.000;
    (2, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
    (3, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
    (3, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
    (3, 2):  0.200, 0.200, 0.200, 0.200, 0.200;
    (3, 3):  0.000, 0.937, 0.000, 0.063, 0.000;
    (3, 4):  0.010, 0.010, 0.962, 0.010, 0.010;
}
```

```
probability (_Num_Seasons_Of_Year_ | _Est_Note_Obs_Segment_)
{
   (0):  1.000, 0.000, 0.000, 0.000;
   (1):  0.986, 0.014, 0.000, 0.000;
   (2):  0.925, 0.075, 0.000, 0.000;
   (3):  0.568, 0.351, 0.062, 0.019;
   (4):  0.042, 0.042, 0.042, 0.875;
}
probability (_Est_Note_Obs_Segment_ | _Num_Caregivers_, _Est_Chart_Obs_Segment_)
{
   (0, 0):  0.743, 0.249, 0.002, 0.002, 0.002;
   (0, 1):  0.113, 0.745, 0.142, 0.000, 0.000;
   (0, 2):  0.000, 0.424, 0.576, 0.000, 0.000;
   (0, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (0, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 1):  0.004, 0.582, 0.414, 0.000, 0.000;
   (1, 2):  0.000, 0.225, 0.773, 0.002, 0.000;
   (1, 3):  0.200, 0.200, 0.200, 0.200, 0.200;
   (1, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (2, 1):  0.001, 0.182, 0.815, 0.001, 0.001;
   (2, 2):  0.000, 0.022, 0.907, 0.070, 0.000;
   (2, 3):  0.000, 0.000, 0.388, 0.611, 0.000;
   (2, 4):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 0):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 1):  0.200, 0.200, 0.200, 0.200, 0.200;
   (3, 2):  0.000, 0.000, 0.676, 0.324, 0.000;
   (3, 3):  0.000, 0.000, 0.076, 0.924, 0.000;
   (3, 4):  0.010, 0.010, 0.010, 0.010, 0.962;
}
probability (_Est_Chart_Obs_Segment_ | _Num_Caregivers_)
{
   (0):  0.006, 0.818, 0.176, 0.000, 0.000;
   (1):  0.000, 0.315, 0.685, 0.000, 0.000;
   (2):  0.000, 0.011, 0.882, 0.107, 0.000;
   (3):  0.000, 0.000, 0.127, 0.871, 0.001;
}

probability (_Num_Caregivers_ | _Num_Rt_Segment_, _Num_Nurse_Segment_)
{
   (0, 0):  0.250, 0.250, 0.250, 0.250;
   (1, 0):  0.250, 0.250, 0.250, 0.250;
   (2, 0):  0.199, 0.263, 0.286, 0.251;
   (3, 0):  0.250, 0.250, 0.250, 0.250;
}
probability (_Num_Rt_Segment_ | _Saps_Segment_, _Num_Nurse_Segment_)
{
```

```
   (0, 0):  0.000, 0.000, 1.000, 0.000;
   (1, 0):  0.000, 0.000, 1.000, 0.000;
   (2, 0):  0.000, 0.000, 1.000, 0.000;
}
probability (_Num_Nurse_Segment_ | _Saps_Segment_, _Num_Locations_)
{
   (0, 0):  1.000;
   (0, 1):  1.000;
   (0, 2):  1.000;
   (0, 3):  1.000;
   (1, 0):  1.000;
   (1, 1):  1.000;
   (1, 2):  1.000;
   (1, 3):  1.000;
   (2, 0):  1.000;
   (2, 1):  1.000;
   (2, 2):  1.000;
   (2, 3):  1.000;
}
probability (_Num_Locations_ | _Cardiac_Unit_, _Micu_)
{
   (0, 0):  0.999, 0.000, 0.000, 0.000;
   (0, 1):  1.000, 0.000, 0.000, 0.000;
   (1, 0):  0.999, 0.000, 0.000, 0.000;
   (1, 1):  0.947, 0.052, 0.000, 0.000;
}
probability (_Cardiac_Unit_ | _C_Sicu_, _Ccu_, _Csru_)
{
   (0, 0, 0):  0.992, 0.008;
   (0, 0, 1):  0.000, 1.000;
   (0, 1, 0):  0.000, 1.000;
   (0, 1, 1):  0.000, 1.000;
   (1, 0, 0):  0.000, 1.000;
   (1, 0, 1):  0.000, 1.000;
   (1, 1, 0):  0.000, 1.000;
   (1, 1, 1):  0.000, 1.000;
}
probability (_Msicu_ | _T_Sicu_, _C_Sicu_, _Trauma_, _Micu_, _Ccu_, _Csru_)
{
   (0, 0, 0, 0, 0, 0):  0.034, 0.966;
   (0, 0, 0, 0, 0, 1):  0.984, 0.016;
   (0, 0, 0, 0, 1, 0):  0.967, 0.033;
   (0, 0, 0, 0, 1, 1):  0.972, 0.028;
   (0, 0, 0, 1, 0, 0):  0.948, 0.052;
   (0, 0, 0, 1, 0, 1):  0.913, 0.087;
   (0, 0, 0, 1, 1, 0):  0.914, 0.086;
   (0, 0, 0, 1, 1, 1):  0.880, 0.120;
   (0, 0, 1, 0, 0, 0):  0.001, 0.999;
```

(0, 0, 1, 0, 0, 1): 1.000, 0.000;
(0, 0, 1, 0, 1, 0): 1.000, 0.000;
(0, 0, 1, 0, 1, 1): 0.999, 0.001;
(0, 0, 1, 1, 0, 0): 1.000, 0.000;
(0, 0, 1, 1, 0, 1): 0.909, 0.091;
(0, 0, 1, 1, 1, 0): 1.000, 0.000;
(0, 0, 1, 1, 1, 1): 1.000, 0.000;
(0, 1, 0, 0, 0, 0): 0.978, 0.022;
(0, 1, 0, 0, 0, 1): 0.946, 0.054;
(0, 1, 0, 0, 1, 0): 1.000, 0.000;
(0, 1, 0, 0, 1, 1): 1.000, 0.000;
(0, 1, 0, 1, 0, 0): 0.971, 0.029;
(0, 1, 0, 1, 0, 1): 0.867, 0.133;
(0, 1, 0, 1, 1, 0): 0.875, 0.125;
(0, 1, 0, 1, 1, 1): 0.769, 0.231;
(0, 1, 1, 0, 0, 0): 1.000, 0.000;
(0, 1, 1, 0, 0, 1): 1.000, 0.000;
(0, 1, 1, 0, 1, 0): 0.856, 0.144;
(0, 1, 1, 0, 1, 1): 1.000, 0.000;
(0, 1, 1, 1, 0, 0): 1.000, 0.000;
(0, 1, 1, 1, 0, 1): 1.000, 0.000;
(0, 1, 1, 1, 1, 0): 0.996, 0.004;
(0, 1, 1, 1, 1, 1): 0.992, 0.008;
(1, 0, 0, 0, 0, 0): 0.969, 0.031;
(1, 0, 0, 0, 0, 1): 1.000, 0.000;
(1, 0, 0, 0, 1, 0): 0.999, 0.001;
(1, 0, 0, 0, 1, 1): 0.996, 0.004;
(1, 0, 0, 1, 0, 0): 0.999, 0.001;
(1, 0, 0, 1, 0, 1): 1.000, 0.000;
(1, 0, 0, 1, 1, 0): 0.500, 0.500;
(1, 0, 0, 1, 1, 1): 0.500, 0.500;
(1, 0, 1, 0, 0, 0): 1.000, 0.000;
(1, 0, 1, 0, 0, 1): 0.992, 0.008;
(1, 0, 1, 0, 1, 0): 0.992, 0.008;
(1, 0, 1, 0, 1, 1): 0.500, 0.500;
(1, 0, 1, 1, 0, 0): 1.000, 0.000;
(1, 0, 1, 1, 0, 1): 0.500, 0.500;
(1, 0, 1, 1, 1, 0): 0.500, 0.500;
(1, 0, 1, 1, 1, 1): 0.500, 0.500;
(1, 1, 0, 0, 0, 0): 0.998, 0.002;
(1, 1, 0, 0, 0, 1): 0.996, 0.004;
(1, 1, 0, 0, 1, 0): 0.997, 0.003;
(1, 1, 0, 0, 1, 1): 0.500, 0.500;
(1, 1, 0, 1, 0, 0): 0.500, 0.500;
(1, 1, 0, 1, 0, 1): 0.008, 0.992;
(1, 1, 0, 1, 1, 0): 0.992, 0.008;
(1, 1, 0, 1, 1, 1): 0.992, 0.008;
(1, 1, 1, 0, 0, 0): 0.500, 0.500;

```
   (1, 1, 1, 0, 0, 1):  0.500, 0.500;
   (1, 1, 1, 0, 1, 0):  0.500, 0.500;
   (1, 1, 1, 0, 1, 1):  0.992, 0.008;
   (1, 1, 1, 1, 0, 0):  0.500, 0.500;
   (1, 1, 1, 1, 0, 1):  0.500, 0.500;
   (1, 1, 1, 1, 1, 0):  0.500, 0.500;
   (1, 1, 1, 1, 1, 1):  0.500, 0.500;
}
probability (_Micu_ | _T_Sicu_, _C_Sicu_, _Ccu_, _Csru_)
{
   (0, 0, 0, 0):  0.297, 0.703;
   (0, 0, 0, 1):  0.926, 0.074;
   (0, 0, 1, 0):  0.866, 0.134;
   (0, 0, 1, 1):  0.843, 0.157;
   (0, 1, 0, 0):  0.865, 0.135;
   (0, 1, 0, 1):  0.730, 0.270;
   (0, 1, 1, 0):  0.783, 0.217;
   (0, 1, 1, 1):  0.563, 0.437;
   (1, 0, 0, 0):  0.929, 0.071;
   (1, 0, 0, 1):  0.908, 0.092;
   (1, 0, 1, 0):  1.000, 0.000;
   (1, 0, 1, 1):  0.985, 0.015;
   (1, 1, 0, 0):  0.992, 0.008;
   (1, 1, 0, 1):  0.663, 0.337;
   (1, 1, 1, 0):  0.746, 0.254;
   (1, 1, 1, 1):  0.500, 0.500;
}
probability (_Csru_ | _T_Sicu_, _C_Sicu_, _Ccu_)
{
   (0, 0, 0):  0.437, 0.563;
   (0, 0, 1):  0.776, 0.224;
   (0, 1, 0):  0.847, 0.153;
   (0, 1, 1):  0.590, 0.410;
   (1, 0, 0):  0.911, 0.089;
   (1, 0, 1):  0.780, 0.220;
   (1, 1, 0):  0.570, 0.430;
   (1, 1, 1):  0.663, 0.337;
}
probability (_Ccu_ | _T_Sicu_, _C_Sicu_, _Acute_Mi_)
{
   (0, 0, 0):  0.757, 0.243;
   (0, 0, 1):  0.398, 0.602;
   (0, 1, 0):  0.860, 0.140;
   (0, 1, 1):  0.502, 0.498;
   (1, 0, 0):  0.944, 0.056;
   (1, 0, 1):  0.753, 0.247;
   (1, 1, 0):  0.500, 0.500;
   (1, 1, 1):  0.944, 0.056;
```

```
}
probability (_C_Sicu_ | _Acute_Mi_, _T_Sicu_, _Trauma_)
{
   (0, 0, 0):  0.868, 0.132;
   (0, 0, 1):  0.605, 0.395;
   (0, 1, 0):  0.820, 0.180;
   (0, 1, 1):  0.987, 0.013;
   (1, 0, 0):  0.949, 0.051;
   (1, 0, 1):  0.834, 0.166;
   (1, 1, 0):  0.890, 0.110;
   (1, 1, 1):  0.500, 0.500;
}
probability (_T_Sicu_ | _Trauma_)
{
   (0):  0.976, 0.024;
   (1):  0.841, 0.159;
}
probability (_Acute_Mi_ | _Trauma_)
{
   (0):  0.840, 0.160;
   (1):  0.962, 0.038;
}
probability (_Trauma_)
{
   0.8598933965057743, 0.14010660349422563;
}
probability (_Expired_ | _Saps_Segment_)
{
   (0):  0.974, 0.026;
   (1):  0.896, 0.104;
   (2):  0.806, 0.194;
}
probability (_Saps_Segment_)
{
   0.2574024281907018, 0.48193070772875335, 0.2606668640805449;
}
```

Bibliography

AnAj. (2006, September). *Simple Bayes Network*. Retrieved April 2010, from Wikipedia: http://en.wikipedia.org/wiki/File:SimpleBayesNet.svg

Bhattacherjee, A., & Hikmet, N. (2007). Physician's Resistance toward Healthcare Information Technology: A Theoritical Model and an Empirical Test. *European Journal of Information Systems , 16*, 725-737.

Christensen, C., Grossman, J., & Hwang, J. (2009). *The Innovator's Prescription: A Disruptive Solution for Healthcare.* New York: McGraw Hill.

Clifford, G. D., Scott, D. J., & Villarroel, M. (2009, December 2). User Guide and Documentation for the MIMIC II Database. Cambridge, MA.

Coiera, E. (2002). Interaction design theory. *International Journal of Medical Informatics* , 1-18.

Coiera, E. (2001). Mediated Agent Interaction. *Proceedings of the Eighth Conference on Artificial Intelligence in Medicine Europe.* Berlin: Springer Lecture Notes.

Coiera, E. (2000). When Conversation Is Better Than Computation. *Journal of American Medical Informatics Association (JAMIA) , 7*, 277-286.

Coiera, E., Jayasuriya, R. A., Hardy, J., Bannan, A., & Thorpe, M. E. (2002). Communication loads on clinical staff in the emergency department. *MJA* , 415-418.

Edwards, A., Fitzpatrick, L., Augustine, S., Trzebuck, A., Cheng, S., Presseau, C., et al. (2009). Synchronous communication facilitates interruptive workflow for attending physicians and nurses in clinical settings. *International Journal of Medical Informatics , 78*, 629-637.

Everitt, B., & Hothorn, T. (2009). *A Handbook of Statistical Analyses Using R.* Boca Raton, FL: Chapman & Hall.

Germov, J., & Freij, M. (2009). The Doctor/Nurse Game. In J. Germov, & M. Freij, *Second Opinion: An Introduction to Health Sociology.* Oxford University Press.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation , 101* (23), e215-e220.

Groopman, J. (2008). *How Doctors Think.* New York: Houghton Mifflin.

Halamka, J. D. (2008, April 23). *Designing the Ideal Healthcare Record.* Retrieved April 30, 2010, from Life as a Healthcare CIO: http://geekdoctor.blogspot.com/search?q=advantages+of+EHR

Hall, M., Frank, E., Holmes, G., Mayo, M., Pfahringer, B., Smith, T., et al. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations , 11* (1).

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* New York: Springer.

HHS. (2009, August 24). Breach Notification for Unsecured Protected Health Information. *Federal Register* . Washington, DC, USA.

HHS. (2010, January 13). Proposed Rules for Electronic Health Record Incentive Program. *Federal Register* . Washington, DC, USA: National Archives and Records Administration.

Hornik, K. (1998). *Comprehensive R Archive Network (CRAN)*. Retrieved January-April 2010, from The R Project for Statistical Computing: http://www.r-project.org/

Hug, C. (2009). *MIT PhD Thesis: Detecting Hazardous Intensive Care Patient Episodes Using Real-time Mortality Models.* Cambridge, MA: MIT.

Hug, C. (2006). *MIT S.M. Thesis: Predicting the Risk and Trajectory of Intensive Care Patients using Survival Models.* Cambridge, MA: MIT.

JCAHO, J. C. (2005). *Root Cause of Sentinel Events.* JCAHO.

Mark, R. (2009). *MIMIC*. Retrieved 2009-2010, from PhysioNet: http://mimic.physionet.org

Miller, A., Scheinkestel, C., & Michele, J. (2009). Coordination and Continuity of Intensive Care Unit Patient Care. *Human Factors* , 354-367.

MySQL. (2009, November). *MySQL Documentation*. Retrieved from MySQL: http://dev.mysql.com/

NCHS, N. C. (2009). *Tabular Index of ICD-9 Codes*. Retrieved January-April 2010, from The International Classification of Diseases, 9th Revision, Clinical Modification: http://icd9cm.chrisendres.com/index.php?action=contents

P.C., T., & McDonald, C. J. (2006). Electronic Health Record Systems. In E. H. Shortliffe, *Biomedical Informatics: Computer Applications in Health Care and Biomedicine* (pp. 450-451). New York: Springer Science + Business Media, LLC.

Ramoni, M., & Sebastini, P. (2003). Bayesian Methods. In M. Berthold, & D. Hand, *Intelligent Data Analysis: An Introduction* (pp. 128-166). New York: Springer.

Sands, D. Z. (2008). *Challenges in Healthcare Communications: How Technology Can Increase Efficiency, Safety, and Satisfaction.* San Jose, CA: Cisco White Paper.

Sebastiani, P., & Ramoni, M. (1999-2000). *Bayesware*. Retrieved 2009-2010, from Bayesware: http://bayesware.com

Sebastiani, P., K.D., M., Szolovits, P., Kohane, I., & Ramoni, M. (2006, June). A Bayesian dynamic model for influenza surveillance. *Statistics in Medicine* , 1803-25.

Sebastini, P., & Ramoni, M. (1999-2000). Bayesware Discoverer.

Shortliffe, E. H., & Barnett, G. O. (2006). Biomedical Informatics. In E. H. Shortliffe, *Biomedical Informatics - Computer Applications in Health Care and Biomedicine* (pp. 46-69). New York, NY: Springer Science+Business Media, LLC.

Stein, L. I., Watts, D., & Howell, T. (1967). The Doctor-Nurse Game. *Archives of General Psychiatry* , 699-703.

Zhang, Y., & Szolovits, P. (2008). Patient-specific learning in real-time for adaptive monitoring in critical care. *Journal of Biomedical Informatics* , 452-460.