Early Prediction of Antibiotics in Intensive Care Unit Patients

by

Donald Misquitta

BS Biomedical Informatics, Kent State University, 2004

MD, Northeast Ohio Medical University, 2008

SUBMITTED TO THE CENTER FOR BIOMEDICAL

INFORMATICS AT THE HARVARD MEDICAL SCHOOL IN

PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF MEDICAL SCIENCE

MAY 2013

Thesis Supervisor: Peter Szolovits, PhD

Title: Professor of Computer Science and Engineering & Health Sciences and Technology,

Massachusetts Institute of Technology/Harvard University

Thesis Committee Members:

David Gagnon, MD, PhD, MPH, Associate Professor of Biostatistics, Boston University School of Public Health

Leo Celi, MD, MSc, MPH, Assistant Clinical Professor of Medicine, Harvard Medical School

ABSTRACT

Introduction. Predictive models derived from electronic health records in the intensive care unit (ICU) have traditionally used data from the first 24 hours of admission, or up to 48-72 hours. While these may have high accuracy and have fewer limitations due to missing data, they are not useful for decision-making during the early hours of an admission. Infections are common in the ICU and international guidelines recommend that antibiotics be administered as soon as possible. Established goals for early administration of antibiotics range from 1-6 hours. Using structured admission data, we attempt to develop a predictive model to provide early identification of patients who warrant antibiotic administration, from a cohort of patients who were not identified by clinicians as having an infection. *Methods*. The Multi-parameter Intelligent Monitoring for Intensive Care II (MIMIC II) is a database of patients admitted to the Beth Israel Deaconess Medical Center ICU between 2001 and 2008. Using the MIMIC- II database and a combination of natural language processing and inpatient orders, we identified patients who did not receive antibiotics within 6 hours of admission. Sociodemographic, clinical, and process variables were extracted for each patient. The dataset was divided into a training and test set with an 80:20 split. Logistic regression models were built. *Results*. 9478 patients met inclusion criteria. Of these, 1403 (14.8%) did not receive antibiotics during the first 6 hours but were subsequently started on antibiotics within two days of hospital admission. The most common antibiotics started were vancomycin, levofloxacin, and metronidazole. A forward-selection logistic regression on the training set, based on a candidate list of variables comprised from theory and bivariate testing, was significant with a c-statistic of 0.67. A logistic regression model on the test set had a c-statistic of 0.65. Most of the variables could not be tested due to data not missing completely at random. The only variables that were significant were physicians' ordering of lactic acid and liver function tests (LFTs). *Conclusion*. It is possible to build a significant logistic regression model based on admission data. The importance of ordering behavior, a proxy for clinician decision-making, indicates that all relevant data is not captured in structured fields.

Introduction

Predictive models in the ICU such as the Acute Physiology and Chronic Health Evaluation (APACHE) score [1] and the Simplified Acute Physiology Score (SAPS) [2] typically use 24 hours or more of admission data. While these models have been validated in diverse patient populations and used in various settings, they cannot be applied to decisions that must be made within 24 hours without significant customization. For example, early administration of antibiotics have garnered significant international interest due to excess mortality and the Surviving Sepsis campaign [3]. Decisions

regarding antibiotics should be made shortly after admission and ideally within 4 hours [4]. As there is great clinical uncertainty surrounding infections, it is often difficult to determine whether a patient has an infection at the time of admission.

An infection in the ICU may be the primary cause of admission or may be present in addition to, or because of, another diagnosis. It is a cause of morbidity, mortality, and high healthcare costs. Sepsis, one form of infection, has a treated mortality between 20 and 50 percent and is the 10th leading cause of death in the United States [5]. The financial cost per hospital admission can be as much as $50,000 per patient which sums to $17 billion in annual costs in the United States. In addition, the incidence has been increasing for unclear reasons [5].

It has been well established that early administration of antibiotics reduces morbidity and mortality in patients with infection. A landmark study by Kumar found that after the onset of hypotension in patients with sepsis, each one-hour delay in initiation of antibiotics resulted in increased mortality, with 46% overall mortality if antibiotics were not started in the first six hours [6]. This finding has been confirmed in other studies [7], [8]. Delayed administration of antibiotics has been associated with acute lung injury in patients with pulmonary sepsis [9], increased medical complications [10], and increased rate of transfer to the ICU [10].

However, it is not always easy to determine if there is an infection at the time. The gold standard is growth of pathogenic bacteria from a culture, but such data is not available on admission and may take up to 48 hours to return. Procalcitonin is a relatively new marker but is expensive [11], can take 7-10 days to return, and has low positive predictive value [12]. In addition, emergency room physicians may be dealing with high volume and other critically ill patients, which may both contribute to a delay in antibiotics. If a patient needs emergency resuscitation or an immediate procedure, efforts will first be made to stabilize the patient and antibiotics may not be considered until later. The decision

to treat must also be balanced by possible side effects of antibiotics, contribution to antimicrobial resistance, and cost.

One recent avenue of research has been on sepsis bundle protocols [4], [13]. The concept is that there is a predefined checklist of criteria available to the physician, which when met, is tied to a bundle of standardized but institution-specific orders which includes diagnostic studies as well as treatment, including antibiotics. When patients meet certain criteria, an alert may appear in the electronic medical record. Because an order set is tied to the criteria, it is less likely that a particular necessary order will be omitted. However, all of these protocols require clinical suspicion of an infection in order to activate, which may not be clear.

In general, previous predictive models of infection and/or sepsis have focused on a particular infection, type of infection, or context [10], [14], [15], used data from structured and unstructured sources, used up to 48 hours of data from admission[16], and have investigated novel markers [16]. While customized models lead to higher accuracy, the tradeoff is a greater number of models and cognitive overload. A small number of validated models are clinically in use today. Another impediment to use is unstructured data, which is time-consuming for a physician or nurse to enter, assuming the data was collected. Models that use more than a few hours of data may be used for retrospective studies, analysis of treatment options, or for mortality studies, but has low clinical utility as a decision support tool given that the goal for antibiotic administration is less than 1 hour [4].

We propose development of a model trained on the general outcome of infection, using only structured data, and using commonly available variables that were present on or shortly after admission. By using admission data, we attempt to identify infections and/or sepsis at an earlier stage. We focus on commonly available and inexpensive blood tests that would be available in other ICUs [17]. In addition, our model differs from prior models in a second important way. Previous studies have

used an outcome of prediction of infection in an unselected cohort of patients. We hypothesize that patients with certain infections will be easy for clinicians to identify. For example, a patient with a fever, a markedly elevated white count, and a cough productive of sputum likely has pneumonia, and would be easy to distinguish from a patient who did not have these characteristics. The usefulness of a predictive model is in discriminating between patients that clinicians have a hard time separating [18]. For this reason our initial cohort is patients who are not started on antibiotics within the first 6 hours, indicating that there was no suspicion of infection. We exclude patients who were started on antibiotics within 6 hours as these are patients that clinicians are already able to identify.

Methods

The Multi-parameter Intelligent Monitoring for Intensive Care (MIMIC II) database consists of high-resolution data of all ICU patients admitted to the Beth Israel Deaconess Medical Center (BIDMC) from 2001 to 2008. It was created through a collaboration between the BIDMC, Philips Healthcare, and the Massachusetts Institute of Technology (MIT). As it is a de-identified database [19], institutional review board (IRB) approval for this study was not required. IRB approval was obtained from both MIT and BIDMC for the development, maintenance and public use of MIMIC-II.

The database consists of data from more than 25,000 patients, including pediatric and adult, and from the medical, surgical, and neurological ICUs, and the cardiac surgery unit. While data from outside the ICU at the BIDMC is generally not available, complete hospital course information is available for patients who were transferred to or from the ICU. Clinical data consists of vital sign information, laboratory data, high resolution waveform information, nursing notes, discharge summaries, and medication orders. Documentation of medication administration is not available. As the emergency department was using a different information system, emergency department notes and orders are not available, but laboratory data from the ED course is present.

The outcome of the study is prediction of infection in a cohort of patients in whom infection was not suspected. To operationalize this, we extracted a cohort of patients who did not receive antibiotics during the first 6 hours of admission, but were subsequently started on antibiotics within the first 2 days. As the incubation period of bacteria is 48 hours, the selection of this time window necessitates that all patients who were started on antibiotics must also have had the infection present upon admission to the hospital. Contrariwise, if an infection at any point during the hospitalization were an outcome, a patient could have developed an infection on hospital day 3 that became clinically apparent on hospital day 5. Looking at initial laboratory data would degrade performance given that no signs or symptoms of infection would have been present until day 3.

While we have access to culture and microbiologic data, which have been used as outcomes in previous studies, we selected a clinician-centered outcome due to the fact that cultures are often not able to be obtained from patients. In addition, when cultures are able to be obtained, results may be negative in the presence of infection due to slow-growing or difficult-to-grow bacteria, and in intra-abdominal sepsis [20]. The initial cohort further represents a group of patients that the admitting clinician or clinicians were not able to distinguish, shown by the fact that none received antibiotics during the first 12 hours. Further analysis of the cohort started on antibiotics late was performed to ascertain that this group had a high rate of infection and will be presented in the results section.

ICD-9 discharge diagnoses of infections were not used as the study outcome for two reasons. Several recent studies have found ICD-9 codes for sepsis to be inaccurate [5], [21], [22]. For example, a study by Martin found a positive predicted value of 88.9% and a negative predictive value of 80.0%. Sensitivity and specificity were not reported. A study by Ollendorf concluded that using ICD-9 codes for sepsis in research may be “prone to substantial error.”[22]. The second reason is that ICD-9 discharge diagnosis codes are time-insensitive. Because an infection could have occurred at any time

during the hospitalization, parameters of infection might not have been present upon admission (because infection was not present upon admission), which was used as the input time period for predictive variables.

The inclusion criteria for the study are adult patients, defined as > 15 years, who are admitted directly to the ICU, go from the emergency department straight to the ICU, and were not transferred to the BIDMC from another hospital. We excluded patients who were transferred from the wards to the ICU as we could not rule out hospital-acquired infection. Transfers from other hospitals were excluded as we could not determine if antibiotics had been administered at the previous setting. The table icustay_detail was used as a master list of hospitalizations. As MIMIC-II was constructed from real EHR data, it has the missing data consistent with live systems including missing hospital admission IDs. Hospitalizations from the icustay_detail that were missing hospital admission IDs were linked by using subject IDs and dates of admissions from the admissions table. A consort diagram is shown in Figure 1 [23].

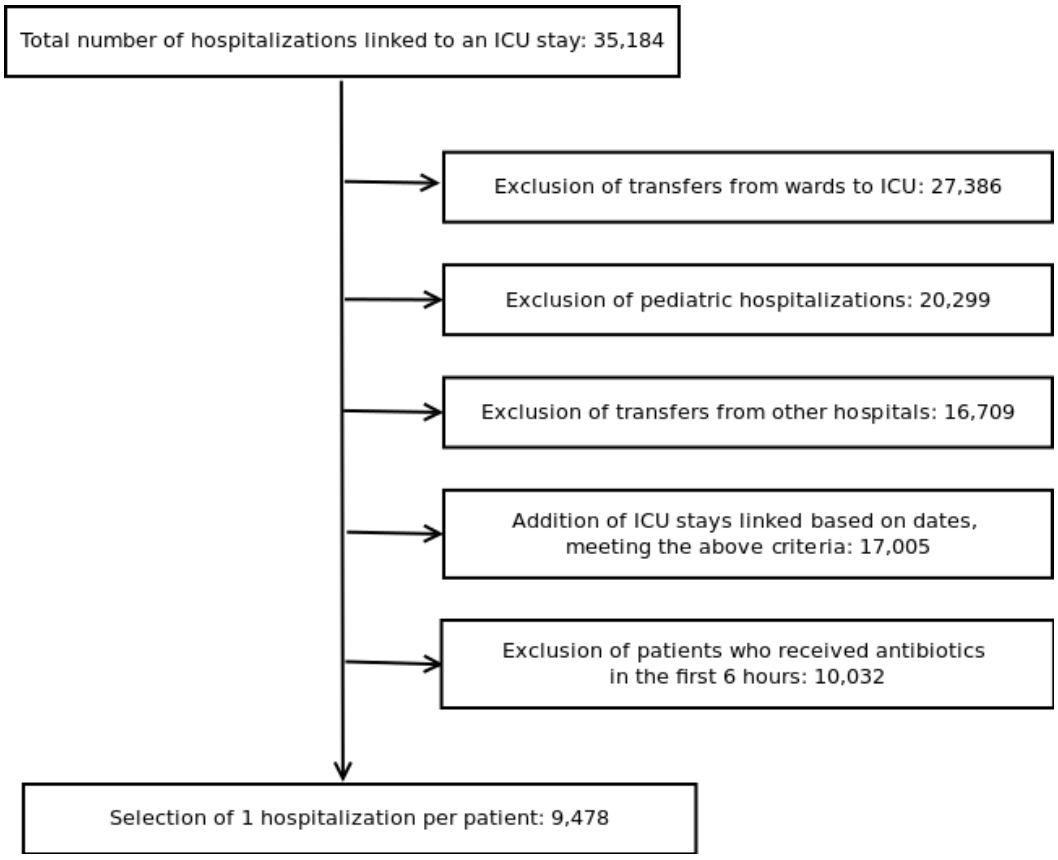

Figure 1 – Consort Diagram

Description of Natural Language Processing

As documentation of medication administration is not available in the database, inpatient orders were used to determine if and when patients received antibiotics. It is assumed that if a medication order was placed, the medication was received by the patient. The initial cohort was created based on absence of antibiotic orders during the first 6 hours. As we did not have access to ED orders, and some of these patients may have received antibiotics in the ED, we further used natural language processing (NLP) to identify these patients. Since no prior studies could be identified that used NLP for this particular task, we created a custom algorithm. A prior study used a commercial NLP tool, MedLEE, on oncology nursing notes[24].

A new cohort was extracted from the MIMIC-II database, which consisted of adult hospitalizations that were not transferred from another facility nor the wards. Out of 17,005 hospitalizations, 86.9% of patients had at least one nursing note during this time period. Nursing admission notes were selected by taking the first nursing note from each hospitalization. A random selection of notes were reviewed and found to be consistent with being admission notes. The caregiver id had to be assigned a label of "RN" and a time period of 24 hours was specified, such that all notes documented more than 24 hours after the time of ICU admission were excluded. Because nursing shifts are 8-12 hours in length, additional time was included so that a nurse who documented his/her findings after the shift would not be excluded. However, if multiple notes were present, only the first one was used, and other notes (by medical students, nursing students, respiratory therapy, and so on) were excluded. Each note is computerized but consists of free-text. While there are no requirements on the content of a nursing note, it is expected that all administrations of antibiotics will be documented. Through manual review of a random selection of notes, it was determined that in addition to excellent documentation of antibiotic administration in the emergency department, the notes also frequently contained information on past medical history, reason for admission, medications, allergies, and descriptions of the plan.

A list of relevant antibiotics is displayed in Table 1. The list was compiled from published drug references. Cefazolin, an antibiotic which is frequently used perioperatively, was excluded, as we are interested in antibiotics used for the treatment of infection and not prophylactic antibiotics. Because medication orders are stored as free-text fields in MIMIC-II, we could not identify antibiotics based on a unique identifier, and also needed to use variations of commonly misspelled words. For this reason the antibiotic list in Table 1 was designed to be as inclusive as possible. Topical antibiotics were excluded.

Amikacin	Amoxicillin	Ampicillin	Azithromycin
Aztreonam	Vancomycin	Cefaclor	Cefadroxil
Cephalexin	Cefamandole	Cefepime	Cefixime
Cefotaxime	Cefoxitin	Cefpodoxime	Cefprozil
Ceftazidime	Ceftriaxone	Cefuroxime	Chloramphenicol
Ciprofloxacin	Clarithromycin	Clindamycin	Colistin
Dapsone	Daptomycin	Dicloxacillin	Doripenem
Doxycycline	Ertapenem	Erythromycin	Ethambutol
Flucloxacillin	Fosfomycin	Gatifloxacin	Geldanamycin
Gentamicin	Imipenem	Isoniazid	Kanamycin
Levofloxacin	Linezolid	Meropenem	Methicillin
Metronidazole	Minocycline	Moxifloxacin	Trovafloxacin
Nafcillin	Nalidixic acid	Tobramycin	Netilmicin
Nitrofurantoin	Norfloxacin	Ofloxacin	Paromomycin
Penicillin	Piperacillin	Polymyxin	Pyrazinamide
Quinupristin	Rifabutin	Rifampicin	Rifampin
Spectinomycin	Streptomycin	Sulfadiazine	Sulfamethoxazole
Telithromycin	Tetracycline	Ticarcillin	Tigecycline
Trimethoprim			

Table 1

A list of approximately 50 words/expressions that were likely to indicate antibiotic administration was developed based on expert opinion. These mainly consisted of actual antibiotic names, both generic and trade names, and terms such as “abx,” “antibiotic,” and “antibiotics.” From the initial nursing notes, a training and test set were created with an 82:18 split.

In order to identify additional keywords that may indicate administration of antibiotics in the ED, all notes that corresponded with hospitalizations in which patients were ordered for antibiotics within the first 6 hours were extracted from the training set. As there has been a proliferation of antibiotic ordering in the emergency department and a trend toward giving early antibiotics, it is assumed that if patients were ordered for antibiotics on the inpatient side, they would have received the first doses of antibiotics in the ED, and documentation would be present in the initial nursing note. While we do not have ED nursing notes, the ICU nurse is required to receive signout from the emergency department nurse during which administration of antibiotics and other important considerations will be relayed. All notes in which patients received early inpatient antibiotics (defined as within 12 hours) were combined into a single document and tokenized using Unix text tools. Case and punctuation were removed and words were sorted by frequency. Words that were likely to indicate antibiotic use based on expert opinion were added to the previously constructed list of 50 words.

Once the list of words was complete, notes in the training set were searched for notes containing specific keywords to determine whether each word was in fact associated with the outcome of receiving antibiotics. During this process, some words were excluded from the list and other ones were added based on manual reading of notes in the training set. Given that there were more than 12,000 notes in the training set, only a small fraction of the total notes were examined. Once this process was complete, documents from the training set were then randomly selected based on an overall composite of search terms/phrases. Further changes were made to the list of terms. A list of the final selection of words/expressions may be found in the Appendix.

Once the above process was complete, the test set was used to determine accuracy. 80 nursing notes were randomly selected from the test set, which had not been seen by the NLP classifier nor a reader. Based on manual reading by a physician, each note was classified as patient receiving or not

receiving antibiotics in the ED. After an assignment was made for each note, the natural language processing classifier was run on these notes. Sensitivity, specificity, positive predictive value, and negative predictive value were calculated, with manual review being considered the gold standard.

The results of the consort diagram above include both inpatient orders and natural language processing. In other words, patients were categorized as not having received antibiotics within the first 6 hours if both inpatient orders during that time window were negative for any of the specified antibiotics, as well as if there was no indication of the patient receiving antibiotics based on the initial MICU nursing note. If the initial MICU nursing note was not available, inpatient orders were used.

Variable Selection

Variables were selected based on a combination of prior theory and bivariate testing. While the specific question of this research project has not been tested, the literature was evaluated for predictive models on infections and sepsis, both in the ICU, non-ICU inpatient, and outpatient settings. Variables that were found to be significant (p-value <0.05) in previous multivariate studies are displayed below.

Variable	Source
Temperature 38.3 or higher	Bates 1990, Giuliano 2007, Shapiro 2008
presence of fatal disease	Bates 1990
presence of shaking chills	Bates 1990
IV drug abuse	Bates 1990
presence of acute abdomen	Bates 1990
major comorbidity	Bates 1990
eosinophil value	Abidi 2008
CRP	Abidi 2008, Wildi 2011
platelet count (low)	Cho 2012, Shapiro 2008
absolute band neutrophil count	Cho 2012
lymphocyte differential	Cho 2012
mean arterial pressure, hypotension	Giuliano 2007, Shapiro 2008
charlson score >= 2	Tudela 2010
Procalcitonin > 0.4	Tudela 2010
albumin	Wildi 2011
presence of SIRS	Wildi 2011

liver disease (cirrhosis or chronic hepatitis)	Bates 1997
hickman catheter or indwelling vascular catheter	Bates 1997, Shapiro 2008
altered mental status	Bates 1997
focal abdominal signs	Bates 1997
clinical suspicion of endocarditis	Shapiro 2008
Age > 65	Shapiro 2008
chills	Shapiro 2008
vomiting	Shapiro 2008
Neutrophil % > 80	Shapiro 2008
WBC > 18,000	Shapiro 2008
Bands > 5%	Shapiro 2008
Creatinine > 2	Shapiro 2008
gender	Martin 2003
ethnicity	Martin 2003

Table 2

To the above list were added the most common laboratory variables present in MIMIC to create a candidate list. This final criterion was used because none of the previous studies took an exploratory approach to variable selection, leading to the possibility that important predictive variables were missed. As an example, Gil and others first identified eosinopenia as a sensitive and specific predictor of bacterial infections in 2003, despite the fact that this commonly-obtained marker has been available for over 100 years [25]. A plethora of predictive models created before 2003 did not include eosinophils as a candidate variable. In addition, since systemic bacterial infections cause changes to numerous laboratory parameters, it is plausible that other common laboratory tests would have abnormal values. A list of the most common laboratory tests ordered is at the end of the Appendix.

The dataset was divided into a training and test data set with an 80/20 split.

All variables were inspected for missing data and erroneous values. In particular, a number of temperatures, respiratory rates, and oxygen saturations had physiologically impossible values and were removed. For laboratory values, the first value available within 4 hours of admission was used. For

vital signs, the high, low, mean, and initial was calculated based on 4 hours of data. For the statistical analysis, the training set was divided into patients who did not receive antibiotics and patients who received antibiotics after 6 hours but within two days. For continuous variables, t-tests were done with equal or unequal variance as appropriate, with the outcome variable. Fisher's exact test was used for categorical variables. All continuous variables were kept continuous. Variables with a p-value of < 0.15 were added to the candidate list that was previously formed. As there was a large amount of missing data, ordering of individual laboratory tests were coded as variables. From this list of variables, a forward selection logistic regression was performed using SAS 9.3 with a significance to enter of 0.05. A 20% test set was withheld.

To adjust for the effects of co-morbidities in predictive models, Charlson and Elixhauser scores have been used. Both of these rely on prior co-morbidity data that is present at the time of admission. Since prior data is not available in MIMIC-II (excluding patients that have had multiple hospitalizations), Elixhauser scores were not used, which would falsely improve model performance by incorporating data into the prediction that had not yet occurred in real life [26]. For the same reason, SAPS and SOFA (Sequential Organ Failure Assessment) scores were not used. While SOFA scores can be calculated based on initial data, the sofa_first computed in MIMIC-II was based on data from the first 24 hours of admission.

Results

9,478 patients met inclusion criteria. Of these, 1403 (14.8%) were not started on antibiotics within the first 6 hours but received antibiotics during the first two days. The most common antibiotics patients were started on are vancomycin, levofloxacin, and metronidazole. Additional antibiotics started and the number of times each antibiotic was administered are displayed in Table 3. If multiple orders were present for the same patient, these were added together.

Most Common Antibiotics Administered

Vancomycin	1838
Levofloxacin	1143
Metronidazole	631
Piperacillin-Tazobactam Na	401
Ciprofloxacin HCl	363
Azithromycin	167
Clindamycin	152
Gentamicin	152
Ceftriaxone	146
CeftriaXONE	96
Sulfameth/Trimethoprim DS	94
Ampicillin	90
Ceftazidime	86
Ampicillin-Sulbactam	71
Meropenem	69
CefePIME	60
Aztreonam	44
Oxacillin	42
Sulfameth/Trimethoprim SS	41
Linezolid	41
Erythromycin	35
Amoxicillin	29
Amoxicillin-Clavulanic Acid	27
Cefepime	25
Nafcillin	22
Vancomycin Oral Liquid	22
Clarithromycin	21
Doxycycline Hyclate	20
Cefpodoxime Proxetil	18
Imipenem-Cilastatin	14
Dicloxacillin	13
Sulfameth/Trimethoprim	12
Dapsone	10
Nitrofurantoin (Macrodantin)	9
Daptomycin	8
Penicillin G Potassium	7
Clindamycin HCl	7
Amikacin	6

Cefotetan	6
Tobramycin	5
Penicillin V Potassium	5
Sulfameth/Trimethoprim Suspension	4
Isoniazid	3
Amoxicillin Oral Susp.	3
Minocycline HCl	3
DiCLOXacillin	2
Ethambutol HCl	2
Pyrazinamide	2
Cefuroxime Sodium	2
Nitrofurantoin Monohyd (MacroBID)	2
SulfADIAzine	1
Tetracycline HCl	1
Rifampin	1
Minocycline	1

Table 3

Natural Language Processing

80 notes were randomly selected from the test set. The results of the validation are displayed in Figure 2. 95% confidence intervals are displayed in parentheses.

	Antibiotics received	No antibiotics received
NLP positive result	16	6

NLP negative result	1	57

Sensitivity: 94.1% (82.9-100)

Specificity: 90.5% (83.2-97.7)

PPV: 72.7% (54.1-91.3)

NPV: 98.3% (94.9-100)

Figure 2

Results of t-tests and Fisher exact tests are shown in Table 4. P-values of < .05 are shown in bold. In a cohort of patients who were not suspected of having an infection, patients who received antibiotics were more likely to be older, on Medicare, in the medical ICU, have greater derangement in laboratory values, and were more likely to have additional tests ordered on admission. All results are for the training data.

Characteristic	**No antibiotics (N = 6,477)**	**Antibiotics (N = 1105)**	**P Value**
Age – yr	60.9	62.9	**0.001**
Male sex – No (%)	3798 (58.7%)	617 (55.9%)	0.08
Insurance – No. Medicare (%)	2481 (38.3%	503 (45.5%)	**<.0001**
Careunit – No. MICU (%)	1426 (22.02%)	361 (32.67%)	**<.0001**
Temperature			
Maximum	36.7	36.7	0.09
Minimum	36.3	36.4	**0.05**
Mean	36.5	36.6	0.07
Respiratory Rate			
Maximum	20.7	21.9	**<.0001**
Mean	17.1	18.5	**<.0001**
Heart Rate			
Maximum	89.9	91.7	**0.01**
Mean	82.6	84.8	**0.002**
Blood Pressure			
Minimum systolic	110.6	110.7	0.93

Minimum diastolic	53.7	53.2	0.38
Average systolic	125.3	125.1	0.87
Average diastolic	63	62.6	0.4
Oxygen Saturation			
Maximum	0.991	99	0.22
Minimum	96.2	95.9	0.11
Mean	98.1	97.9	**0.03**
Initial	98.1	98	0.13
Laboratory Data			
Platelets	235.8	242.4	0.12
Creatinine	1.22	1.58	**<.0001**
BUN	22.3	28.1	**<.0001**
Leukocytes	11.9	12.4	0.13
MCHC	34.4	34	**<.0001**
MCH	30.6	30.5	0.18
MCV	89.4	90	**0.02**
Erythrocytes	3.99	3.94	0.11
RDW	14.1	14.8	**<.0001**
INR	1.5	1.5	0.3
PTT	34.3	34.6	0.71
Chloride	105	103.8	**<.0001**
Hemoglobin	12.3	11.9	**<.0001**
Hematocrit	36.3	35.3	**0.0001**
Bicarbonate	24.6	24.1	**0.0034**
Potassium	4.2	4.32	**0.0003**
Glucose	152	169.6	**<.0001**
Sodium	139	138.5	**0.0072**
Neutrophils	4.28	4.43	**0.0003**
Anion gap	16.2	16.6	**0.04**
pH	7.4	7.37	**<.0001**
Eosinophils	1.3	1.13	**0.04**
Lactate	2.99	3.12	0.24
Base Excess	-1.21	-1.92	**0.02**
Albumin	3.69	3.49	**<.0001**
Ordered Tests - % ordered			
Auto-Differential	0.376	0.537	**<.0001**
Urine Studies	0.333	0.426	**<.0001**
Lactic Acid	0.356	0.475	**<.0001**
ABG	0.522	0.518	0.81
Magnesium	0.451	0.542	**<.0001**
CPK	0.343	0.47	**<.0001**
Amylase	0.247	0.324	**<.0001**
Fibrinogen	0.246	0.192	**<.0001**
LFTs	0.188	0.339	**<.0001**

Albumin	0.134	0.249	**<.0001**

Table 4

The number of each variable present and the amount of missing data are displayed in Table 5. A small amount of data was likely data missing completely at random. For example, 10-20% of patients did not have a basic metabolic panel or complete blood count within the first 4 hours of admission. It is assumed that all patients received these tests and the data did not get imported into MIMIC-II. On the other hand, a large amount of data was not missing completely at random and reflected ordering behavior by the clinician. For example, a lipase was obtained in about 13% of the patients.

Missing Data

Variable	Number	% Missing
age	7121	0
gender	7115	0.0008426
bicarb	6527	0.0834153
hematocrit	6523	0.083977
potassium	6481	0.089875
hemoglobin	6388	0.102935
sodium	6363	0.1064457
glucose	6355	0.1075692
hr_max	6306	0.1144502
hr_mean	6306	0.1144502
minbpsys	6299	0.1154332
minbpdias	6299	0.1154332
avgsysbp	6299	0.1154332
avgdiasbp	6299	0.1154332
minsat	6297	0.1157141
avgsat	6297	0.1157141
maxsat	6297	0.1157141
max_temp	6256	0.1214717
min_temp	6256	0.1214717
mean_temp	6256	0.1214717
rr_max	6212	0.1276506
rr_mean	6212	0.1276506
platelet	5849	0.1786266
creatinine	5602	0.2133127

bun	5595	0.2142957
leukocytes	5546	0.2211768
mchc	5532	0.2231428
mch	5530	0.2234237
mcv	5530	0.2234237
erythrocytes	5530	0.2234237
rdw	5512	0.2259514
pt	5245	0.2634461
inr	5243	0.263727
ptt	5210	0.2683612
chloride	5048	0.2911108
hr_min	4841	0.3201798
neutrophils	4683	0.3423676
ag	4565	0.3589384
ph	3870	0.456537
baseexcess_abg	3675	0.4839208
magnesium	3292	0.5377054
eosinophils	2876	0.5961241
lactate	2641	0.6291251
cpk	2571	0.6389552
urine_ph	2474	0.6525769
amylase	1837	0.7420306
fibrinogen	1684	0.7635164
alt	1503	0.7889341
lipase	892	0.8747367

Table 5

Given the amount of missing data not missing completely at random, all the variables that achieved bivariate significance could not be entered into the model due to the introduction of bias. Only variables with a sufficient number of observations were used. In order to obtain accurate regression coefficients, we limited the number of variables included in the candidate list such that there were at least 10 events per variable [27]. This list was entered into a forward selection logistic regression model with a significance to enter of 0.05. The variable list includes both results of testing as well as ordering behavior by clinicians. Results of global hypothesis testing, stepwise selection, maximum likelihood testing, and odds ratios are displayed in Figure 3. The overall model was significant with p-

value <.0001 and a c-statistic of 0.67 was obtained.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	3808.912	3603.661
SC	3815.319	3661.319
-2 Log L	3806.912	3585.661

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	221.2514	8	<.0001
Score	230.7060	8	<.0001
Wald	213.6440	8	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
23.2702	20	0.2757

Summary of Stepwise Selection								
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
	Entered	Removed						
1	o_alt		1	1	90.8533		<.0001	
2	rr_mean		1	2	38.2019		<.0001	rr_mean
3	rdw		1	3	30.5736		<.0001	rdw
4	o_lactate		1	4	22.0258		<.0001	
5	o_autodiff		1	5	32.7836		<.0001	
6	mchc		1	6	7.9097		0.0049	mchc
7	glucose		1	7	5.6821		0.0171	glucose
8	age		1	8	4.0871		0.0432	age

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.4668	1.2818	3.7034	0.0543
age	1	0.00461	0.00228	4.0817	0.0433
rr_mean	1	0.0410	0.0101	16.5132	<.0001
mchc	1	-0.0670	0.0311	4.6501	0.0311
rdw	1	0.0800	0.0227	12.3833	0.0004
glucose	1	0.000924	0.000383	5.8163	0.0159
o_autodiff	1	0.4764	0.0969	24.1642	<.0001
o_lactate	1	0.5374	0.0897	35.8819	<.0001
o_alt	1	0.6059	0.0922	43.2000	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
age	1.005	1.000	1.009
rr_mean	1.042	1.021	1.063
mchc	0.935	0.880	0.994
rdw	1.083	1.036	1.133
glucose	1.001	1.000	1.002
o_autodiff	1.610	1.332	1.947
o_lactate	1.711	1.436	2.040
o_alt	1.833	1.530	2.196

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	66.7	Somers' D	0.342
Percent Discordant	32.6	Gamma	0.344
Percent Tied	0.7	Tau-a	0.088
Pairs	2575044	c	0.671

Figure 3

A two-stage procedure was used to test the model. In order to determine if individual variables

remained significant, they were entered into a second model using the test set, without a selection procedure. The individual coefficients and p-values are displayed in Figure 4. The c-statistic of this model was 0.65. It includes 1155 hospitalizations of which 181 were started on antibiotics after six hours.

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	43.2636	8	<.0001
Score	46.0124	8	<.0001
Wald	43.0363	8	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-2.0358	2.3381	0.7581	0.3839
age	1	0.00604	0.00449	1.8067	0.1789
rr_mean	1	-0.00526	0.0208	0.0638	0.8006
mchc	1	-0.0613	0.0572	1.1502	0.2835
rdw	1	0.1087	0.0417	6.7882	0.0092
glucose	1	0.000643	0.000890	0.5220	0.4700
o_autodiff	1	0.1458	0.1802	0.6547	0.4184
o_lactate	1	0.3985	0.1714	5.4075	0.0201
o_alt	1	0.6935	0.1798	14.8744	0.0001

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
age	1.006	0.997	1.015
rr_mean	0.995	0.955	1.036
mchc	0.941	0.841	1.052
rdw	1.115	1.027	1.210
glucose	1.001	0.999	1.002
o_autodiff	1.157	0.813	1.647

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
o_lactate	1.490	1.065	2.084
o_alt	2.001	1.406	2.846

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	64.2	Somers' D	0.292
Percent Discordant	34.9	Gamma	0.295
Percent Tied	0.9	Tau-a	0.077
Pairs	176294	c	0.646

Figure 4

Finally, the coefficients from the initial model were applied to the test set to determine overall model fit. The c-statistic was 0.63. All models were globally significant with p-values < .0001 for the likelihood ratio, the Score test, and the Wald test.

Discussion

We demonstrated that a model can be developed using only very early admission data to distinguish patients who have infection from those who do not. While a number of clinical and laboratory values appeared highly significant in bivariate testing, we were not able to test the majority of these due to nonrandom missing data, and consequently the final models only contain a small number of variables. This may be one of the reasons that the c-statistics were not higher. Nevertheless, a difficult task was selected for the algorithm in that the initial cohort of patients was not distinguishable by clinicians, in terms of infection. The only clinical variable that was significant was

the red cell distribution width (RDW). Previous studies have identified RDW as predictive of mortality in ICU patients with community-acquired pneumonia, with sepsis, and in unselected patients. Ours is the first study to identify RDW as a significant predictor of antibiotic use in a cohort of patients who did not initially receive antibiotics. There are multiple possible explanations. In patients without infection, sicker patients may be more likely to receive antibiotics due to the higher clinical consequences of not treating a patient who may have an infection. All other things being equal, patients with infections are also more likely to be sicker, which could increase the RDW. Future work will need to determine whether RDW is predictive of infection, as well as whether RDW is predictive of initial antibiotic use. Answers to these questions have important implications for improved antibiotic prescribing, including avoidance of antibiotics in patients without strong indications.

Of the three variables that were significant, two reflected ordering behavior. As virtually all of the ordering variables checked were bivariately highly significant in the training set, it is likely that the specific test ordered was not important and some of the same information was being captured in multiple laboratory tests. The importance of ordering behavior raises several issues. By modeling the behavior of clinicians, we are getting insight into their thought process, and in the current study, this demonstrates a level of concern for a subgroup of patients who were not treated differently during the first 6 hours, but were subsequently started on IV antibiotics. Since none of the blood tests ordered were sensitive or specific for infection, the ordering of additional tests may have represented nonspecific uncertainty in the diagnosis. We were not able to reproduce this suspicion in an algorithm on the basis of other clinical and laboratory data, as ordering behavior was significant while controlling for the RDW. As the algorithm had access to the same structured data that the clinician had access to, minus possible missing data in the construction of the MIMIC database, this indicates that there is likely important unstructured data in patient presentations. This includes the appearance of the patient and other intangibles, some or all of which might have been incorporated into histories and physicals

(H&Ps), which were not present in this database. Some of this information may have also been represented in vital sign data, but we had to discard a number of clearly erroneous values, which may indicate artifact in the probes used to collect data. Working in a live clinical setting, a physician can easily determine artifacts in data through visual inspection of waveforms and other data and by looking at the patient, gleaning vital sign information we did not have. However, because all data is captured electronically the same way, there is no way to distinguish artifact once this importing has occurred. As automated capturing of high-resolution data becomes more frequent, it may be useful to also record meta-data, such as information on the accuracy of the data captured.

Our study outcome was infection, and yet many of the clinical variables that we were interested in based on prior studies, were not present in sufficient numbers. For example, an auto-differential was only obtained in about 40% of the initial cohort. One of the components of the auto-differential is eosinophils, which have been shown in prior studies to have high sensitivity and specificity for infection. The low use may reflect differential ordering behavior by emergency medicine physicians, who obtain differentials at a lower rate compared to inpatient doctors. Even though the automated differential can be run from the same specimen as a complete blood count, and is done by a machine and therefore requires no additional labor, it is additional work for the inpatient physician to call the laboratory or add-on the test in the computerized physician ordering system. By looking at future orders during the same hospitalization, we found that an additional 20% of the initial cohort will have an automated differential obtained. A further difficulty in obtaining data in this cohort is that it was defined on the basis of lack of physician suspicion and consequently physicians are less likely to order tests indicative of infection. We suggest that automated differentials be obtained more often in patients admitted to the ICU because it provides additional information at no additional cost. Unlike other types of laboratory testing and radiology, it is unlikely that there will be adverse consequences to the patient based on actions taken by clinicians based on the results of an automated differential.

Similarly, c-reactive protein (CRP), a marker of inflammation, was rarely obtained. This intuitively makes sense because it is not considered a useful ICU test because it is likely to be elevated in all patients, and therefore provides no discriminatory ability. The patients in this cohort who had CRP tested may have not been earmarked for the ICU at the time of ordering. However, in the patients in whom CRP was checked, 48% had a normal value. This suggests there may be utility in using CRP as it is unlikely that a patient would have a serious infection in the presence of a normal CRP. Review of the literature demonstrates several recent published papers that have found CRP to be a significant predictor of various clinical outcomes in the ICU. Procalcitonin is a promising marker of infection but was not available at the Beth Israel Deaconess during the time this study was conducted. Future studies will need to combine information provided by procalcitonin with other clinical and laboratory data since procalcitonin lacks specificity to be used alone.

Due to limitations in the exchange of data between systems, we used natural language processing to identify patients who had received antibiotics, on the basis of nursing notes. While this was an intermediate step in order to create our cohort for modeling, there are a number of future directions for this component of the research. NLP is an active area of research in Biomedical Informatics and corpora of medications have been created for NLP engines, including antibiotics. To our knowledge there has been minimal work applying NLP to antibiotics in nursing notes.. The lack of interoperability between electronic medical record systems in hospitals and the large percentage of unstructured data is not unique to our institution, making it important to be able to glean information from available sources, including nursing documentation. Specific applications in using NLP for antibiotics in nursing notes can include improving the quality of medication reconciliation, obtaining historical antibiotic information for clinical purposes, and to identify antibiotic allergies. Dr. Goss, as part of his master's thesis at Harvard, found a high prevalence of medication allergies entered as unstructured data, which effectively bypasses all current and future computerized decision support or

drug-allergy checking [28], potentially leading to medical error and jeopardizing patient safety. Many medication allergies are to antibiotics. As inpatient notes become computerized at more institutions, it will be important to use NLP to improve the quality of care, for research purposes, and to improve documentation.

While our NLP classifier had excellent negative predictive value, the positive predictive value was only 72.7%. The mistakes fell into three general areas: negation, antibiotic allergies, and historical information. Our current classifier could not distinguish between an antibiotic that had been given before the hospitalization or that was an allergy, from an antibiotic administered. It also did not recognize expressions such as "no abx given." Future iterations will be focused at improving these three areas. Another future step may be to use emergency department or inpatient orders as a gold standard as opposed to manual chart review.

The terms used in the final NLP classifier demonstrated the importance of punctuation, including locations of commas, slashes, and periods, in order to disambiguate antibiotics from other abbreviations. Evaluating nursing notes also revealed differences in terminology and vocabulary when compared to physicians. For example, "CAP" is used by physicians as an abbreviation for community-acquired pneumonia but nurses universally use it to indicate "capillary," as in "CAP refill." There were other examples of nursing-specific terms and NLP that uses nursing notes as the substrate will need to take into account these differences in nomenclature, as well as frequent abbreviations and misspellings of antibiotics.

This study has a number of limitations. The initial cohort was created using a combination of natural language processing and inpatient orders. While we are fairly certain that patients in this cohort did not receive antibiotics on admission, we likely also excluded additional patients that were not on antibiotics but should have been part of the cohort because the positive predictive value of the NLP was

73%. This slightly reduced our sample size, limiting power to find differences in groups. About 13% of the patients did not have a nursing note documented in MIMIC-II within 24 hours of the ICU admission time, and some of these patients may have received antibiotics in the emergency department. In addition, it is possible that a small number of patients received antibiotics in the emergency department but there was no documentation in a nursing note, nor an inpatient order in the first 6 hours.

As we used data from electronic medical records, we were subject to erroneous data, especially for vital signs, which may have impacted our findings. A number of patients had a temperature of two degrees recorded, blood pressures of zero, and oxygen saturations of less than 30%. It is not known whether non-palpable blood pressures were coded as zeros. While this may have made intuitive sense to the nurse, it is physiologically impossible as a blood pressure of zero is incompatible with life. We chose to exclude these values as they may have reflected errors in data entry.

Although it is assumed that patients who received delayed antibiotics had clinically significant infections, it is possible that these antibiotics were given inappropriately, which would weaken our assertion that it is important to identify predictors of delayed antibiotics. However, this is unlikely for several reasons. Based on an analysis of inpatient antibiotic orders in MIMIC-II, 70% of patients who receive antibiotics during the first 48 hours of hospitalization received them during the first six hours, which represents just 12.5% of the 48-hour window. For patients to receive antibiotics during the remaining 42 hours is unusual and generally represents new data, new information, or a clinical change in the patient's status, all of which make it more likely that these antibiotic orders are more correct compared to initial antibiotic orders. As a preliminary analysis, we also compared the rate of positive cultures during the first 96 hours of each hospitalization for patients in the two groups. In the group receiving delayed antibiotics, 35.1% had a positive culture compared to 14.0% of patients not started on antibiotics. This absolute higher rate of infection gives credence to the fact that antibiotics were not

being administrered unnecessarily, or to patients who were sicker but without infections. We did not evaluate the rate of contaminated blood cultures in each group, nor did we exclude positive cultures in patients without clinical evidence of infection, which may indicate colonization instead of infection. A possible future direction is to categorize and understand the kinds of infections and reasons for presumed delayed antibiotics in this population. Nevertheless, the fact that a significant model can be created shows promise for the future, and there are many possible avenues for additional work in predictive modeling in order to improve patient care and reduce health care costs.

Acknowledgements

I would like to thank Drs. Leo Celi, Robert Friedman, Peter Szolovits, and David Gagnon for their invaluable help and support. Thank you to Dr. Daniel Schulman for assistance with initial postgreSQL.

References

[1] W. Knaus, E. Draper, D. Wagner, and J. Zimmerman, "APACHE II: A severity of disease classification system," *Critical Care Medicine*, vol. 13, no. 10, pp. 818–828, 1985.

[2] S. Hunziker, L. A. Celi, J. Lee, and M. D. Howell, "Red cell distribution width improves the simplified acute physiology score for risk prediction in unselected critically ill patients.," *Critical care (London, England)*, vol. 16, no. 3, p. R89, May 2012.

[3] R. P. Dellinger, M. M. Levy, J. M. Carlet, J. Bion, M. M. Parker, R. Jaeschke, K. Reinhart, D. C. Angus, C. Brun-Buisson, R. Beale, T. Calandra, J.-F. Dhainaut, H. Gerlach, M. Harvey, J. J.

Marini, J. Marshall, M. Ranieri, G. Ramsay, J. Sevransky, B. T. Thompson, S. Townsend, J. S. Vender, J. L. Zimmerman, and J.-L. Vincent, "Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008.," *Critical care medicine*, vol. 36, no. 1, pp. 296–327, Jan. 2008.

[4] H. L. Cronshaw, R. Daniels, A. Bleetman, E. Joynes, and M. Sheils, "Impact of the Surviving Sepsis Campaign on the recognition and management of severe sepsis in the emergency department: are we failing?," *Emergency medicine journal : EMJ*, vol. 28, no. 8, pp. 670–5, Aug. 2011.

[5] G. S. Martin, D. M. Mannino, S. Eaton, and M. Moss, "The epidemiology of sepsis in the United States from 1979 through 2000.," *The New England journal of medicine*, vol. 348, no. 16, pp. 1546–54, Apr. 2003.

[6] A. Kumar, D. Roberts, K. E. Wood, B. Light, J. E. Parrillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, D. Gurka, A. Kumar, and M. Cheang, "Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock.," *Critical care medicine*, vol. 34, no. 6, pp. 1589–96, Jun. 2006.

[7] A. L. H. Kwa, J. G. H. Low, T. P. Lim, P. C. Leow, A. Kurup, and V. H. Tam, "Independent predictors for mortality in patients with positive Stenotrophomonas maltophilia cultures.," *Annals of the Academy of Medicine, Singapore*, vol. 37, no. 10, pp. 826–30, Oct. 2008.

[8] M. Varpula, S. Karlsson, I. Parviainen, E. Ruokonen, and V. Pettilä, "Community-acquired septic shock: early management and outcome in a nationwide study in Finland.," *Acta anaesthesiologica Scandinavica*, vol. 51, no. 10, pp. 1320–6, Nov. 2007.

[9] R. Iscimen, R. Cartin-Ceba, M. Yilmaz, H. Khan, R. D. Hubmayr, B. Afessa, and O. Gajic, "Risk factors for the development of acute lung injury in patients with septic shock: an observational cohort study.," *Critical care medicine*, vol. 36, no. 5, pp. 1518–22, May 2008.

[10] B. J. Hamburg, G. A. Storch, S. T. Micek, and M. H. Kollef, "The importance of early treatment with doxycycline in human ehrlichiosis.," *Medicine*, vol. 87, no. 2, pp. 53–60, Mar. 2008.

[11] K. J. Smith, A. Wateska, M. P. Nowalk, M. Raymund, B. Y. Lee, R. K. Zimmerman, and M. J. Fine, "Cost-Effectiveness of Procalcitonin-Guided Antibiotic Use in Community Acquired Pneumonia.," *Journal of general internal medicine*, Mar. 2013.

[12] S. Riedel, J. H. Melendez, A. T. An, J. E. Rosenbaum, and J. M. Zenilman, "Procalcitonin as a marker for the detection of bacteremia and sepsis in the emergency department.," *American journal of clinical pathology*, vol. 135, no. 2, pp. 182–9, Feb. 2011.

[13] A. V Barochia, X. Cui, D. Vitberg, A. F. Suffredini, N. P. O'Grady, S. M. Banks, P. Minneci, S. J.

Kern, R. L. Danner, C. Natanson, and P. Q. Eichacker, “Bundled care for septic shock: an analysis of clinical trials.,” *Critical care medicine*, vol. 38, no. 2, pp. 668–78, Feb. 2010.

[14] D. W. Bates, E. F. Cook, L. Goldman, and T. H. Lee, “Predicting bacteremia in hospitalized patients. A prospectively validated model.,” *Annals of internal medicine*, vol. 113, no. 7, pp. 495–500, Oct. 1990.

[15] K. J. Davies, J. H. Walters, I. M. Kerslake, R. Greenwood, and M. J. C. Thomas, “Early antibiotics improve survival following out-of hospital cardiac arrest.,” *Resuscitation*, Nov. 2012.

[16] A. Nierhaus, J. Linssen, D. Wichmann, S. Braune, and S. Kluge, “Use of a weighted, automated analysis of the differential blood count to differentiate sepsis from non-infectious systemic inflammation: the intensive care infection score (ICIS).,” *Inflammation & allergy drug targets*, vol. 11, no. 2, pp. 109–15, Apr. 2012.

[17] T. F. Bishop, A. D. Federman, and J. S. Ross, “Laboratory test ordering at physician offices with and without on-site laboratories.,” *Journal of general internal medicine*, vol. 25, no. 10, pp. 1057–63, Oct. 2010.

[18] G. Singal and P. Currier, “How can we best use electronic data to find and treat the critically ill?*.,” *Critical care medicine*, vol. 40, no. 7, pp. 2242–3, Jul. 2012.

[19] L. A. Celi, L. C. Hinske, G. Alterovitz, and P. Szolovits, “An artificial intelligence tool to predict fluid requirement in the intensive care unit: a proof-of-concept study.,” *Critical care (London, England)*, vol. 12, no. 6, p. R151, Jan. 2008.

[20] D. W. Bates, K. Sands, E. Miller, P. N. Lanken, P. L. Hibberd, P. S. Graman, J. S. Schwartz, K. Kahn, D. R. Snydman, J. Parsonnet, R. Moore, E. Black, B. L. Johnson, and A. Jha, “Predicting Bacteremia in Patients with Sepsis Syndrome,” *The Journal of Infectious Diseases*, vol. 176, no. 6, pp. 1538–1551, Dec. 1997.

[21] R. Gedeborg, M. Furebring, and K. Michaëlsson, “Diagnosis-dependent misclassification of infections using administrative data variably affected incidence and mortality estimates in ICU patients.,” *Journal of clinical epidemiology*, vol. 60, no. 2, pp. 155–62, Mar. 2007.

[22] D. A. Ollendorf, A. M. Fendrick, K. Massey, G. R. Williams, and G. Oster, “Is sepsis accurately coded on hospital bills?,” *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, vol. 5, no. 2, pp. 79–81.

[23] “ScienceDirect.com - Contemporary Clinical Trials - A novel diagram and complement to the CONSORT chart for presenting multimodal clinical trials.” [Online]. Available: http://www.sciencedirect.com.ezp-prod1.hul.harvard.edu/science/article/pii/S1551714409000111. [Accessed: 20-May-2013].

[24] S. Hyun, S. B. Johnson, and S. Bakken, "Exploring the ability of natural language processing to extract data from nursing narratives.," *Computers, informatics, nursing : CIN*, vol. 27, no. 4, pp. 215–23; quiz 224–5.

[25] H. Gil, N. Magy, F. Mauny, and J.-L. Dupond, "[Value of eosinopenia in inflammatory disorders: an 'old' marker revisited].," *La Revue de médecine interne / fondée ... par la Société nationale francaise de médecine interne*, vol. 24, no. 7, pp. 431–5, Jul. 2003.

[26] C. Van Walraven, H. Quan, and A. J. Forster, "O RIGINAL A RTICLE A Modification of the Elixhauser Comorbidity Measures Into a Point System for Hospital Death Using," vol. 47, no. 6, pp. 626–633, 2009.

[27] "ScienceDirect.com - Journal of Clinical Epidemiology - A simulation study of the number of events per variable in logistic regression analysis." [Online]. Available: http://www.sciencedirect.com.ezp-prod1.hul.harvard.edu/science/article/pii/S0895435696002363. [Accessed: 20-May-2013].

[28] F. Goss, "Master's Thesis," Harvard Medical School, 2012.

Appendix A – SQL Code and Explanatory Comments

Total number of patients in database – 32535 (from d_patients, considered distinct sum from

icustay_detail, admissions, demographic_detail tables)

32074 of these pts have an admission listed in the admissions table

Total number of hospitalizations– 36094

(from admissions table. Master list, not possible to add additional hospitalizations)

Number of hospitalizations described in icustay_detail based on hadm_id: 35184

-Means 910 hospitalizations for which icu_stay information is not linked

- 3258 icu_stays in icu_stay_detail that does not have an associated hospital admission, admission date, or discharge date.

icustay_detail table: 40425 rows, 40424 subject_ids, 37617 hadm_ids, 40425 icustay_ids

```
CREATE TABLE tenor_ab AS

SELECT * FROM poe_order WHERE lower(medication) SIMILAR TO
'%amikacin%|%gentamicin%|%kanamycin%|%netilmicin%|%tobramycin%|%paromomycin%|%spect
inomycin%|%geldanamycin%|%ertapenem%|%doripenem%|%imipenem%|%meropenem%|%cefadroxil
%|%cefalexin%|%cefaclor%|%cefoxitin%|%cefprozil%|%cefamandole%|%cefuroxime%|%cefixi
me%|%cefotaxime%|%cefpodoxime%|%ceftazidime%|%ceftriaxone%|%cefepime%|%vancomycin%|
%vanc%|%clindamycin%|%daptomycin%|%azithromycin%|%clarithromycin%|%erythromycin%|%t
elithromycin%|%aztreonam%|%nitrofurantoin%|%linezolid%|%amoxicillin%|%ampicillin%|%
dicloxacillin%|%flucloxacillin%|%methicillin%|%nafcillin%|%oxacillin%|%penicillin%|
%piperacillin%|%cefotetan%|%ticarcillin%|%timentin%|%colistin%|%bactrim%|%polymyxin
%|%ciprofloxacin%|%gatifloxacin%|%levofloxacin%|%moxifloxacin%|%nalidixic
acid%|%norfloxacin%|%ofloxacin%|%trovafloxacin%|%sulfadiazine%|%sulfamethoxazole%|%
trimethoprim%|%TMP%|%doxycycline%|%minocycline%|%tetracycline%|%dapsone%|%ethambuto
l%|%isoniazid%|%pyrazinamide%|%rifampicin%|%rifampin%|%rifabutin%|%streptomycin%|%c
hloramphenicol%|%synercid%|%fosfomycin%|%metronidazole%|%quinupristin%|%tigecycline
%|%unasyn%'

AND route IN ('IV', 'PO');
```

list of antibiotics in all patients in all hospitalizations, containing 94832 rows. all contain a subject_id and hadm_id. 35701 are missing icustay_id

cohort3 – initial selection of adult hospitalizations excluding transfers from wards and other hospitals for which an ICU stay can be linked, either automatically or manually. 17,005 hospitalizations.

Fields: subject_id, hadm_id, icustay_id, admission_date, icustay_intime, discharge_date, dod

no missing values for subject_id, hadm_id, or icustay_id

```
SELECT hadm_id FROM firstnote1 WHERE lower(text) SIMILAR TO
'%amikacin%|%genta%|%kanamycin%|%netilmicin%|%tobramycin%|%paromomycin%|%spectinomy
cin%|%geldanamycin%|%ertapenem%|%doripenem%|%imipenem%|%meropenem%|%merepenum%|%cef
adroxil%|%cefalexin%|%cefaclor%|%cefoxitin%|%cefprozil%|%cefamandole%|%cefuroxime%|
```

```
%cefixime%|%cefotaxime%|%cefpodoxime%|%ceftazidime%|%ceftriaxone%|%cefepime%|%cefam
pine%|% vanco %|% vanc
%|%vancomycin%|%clindamycin%|%clinda%|%daptomycin%|%azithromycin%|%clarithromycin%|
%erythromycin%|%telithromycin%|%aztreonam%|%nitrofurantoin%|%linezolid%|%amoxicilli
n%|%ampicillin%|%ampcillin%|%dicloxacillin%|%flucloxacillin%|%methicillin%|%nafcill
in%|%oxacillin%|%penicillin%|%piperacillin%|%pipercillin%|%cefotetan%|%ticarcillin%
|%timentin%|%colistin%|%bactrim%|%polymyxin%|%ciprofloxacin%|% cipro
%|%gatifloxacin%|%levofloxacin%|%levoflox%|%levoquin%|%levaquin%|%moxifloxacin%|%na
lidixic
acid%|%norfloxacin%|%ofloxacin%|%trovafloxacin%|%sulfadiazine%|%sulfamethoxazole%|%
trimethoprim%|%TMP%|%doxycycline%|%minocycline%|%tetracycline%|%dapsone%|%ethambuto
l%|%isoniazid%|%pyrazinamide%|%rifampicin%|%rifampin%|%rifabutin%|%streptomycin%|%c
hloramphenicol%|%synercid%|%fosfomycin%|%metronidazole%|%linezoid%|%quinupristin%|%
tigecycline%|%bacteremia%|%abcess%|%ceftrioaxone%|%abscess%|%meropenium%|% gent
%|%antibiotic%|%antibx%|%abx%|%anbx%|%cdiff%|%septic%|%sepsis%|%cellulitis%|%ancef%
|% ampi
%|%ampi,%|%ampacillin%|%rocephin%|%/gent%|%urosepsis%|%zosyn%|%infection%|%pneumoni
a%|%pnuemonia%|% uti %|%flagyl%|%falgyl%|%unasyn%'

AND hadm_id NOT IN (SELECT C.hadm_id FROM tenor_ab T, cohort3 C WHERE C.hadm_id =
T.hadm_id

AND start_dt <= icustay_intime + interval '6 hours');



SELECT COUNT(DISTINCT C.hadm_id) FROM tenor_ab T, cohort3 C WHERE C.hadm_id =
T.hadm_id

AND start_dt <= icustay_intime + interval '6 hours';



SELECT DISTINCT ON (C.hadm_id) C.subject_id, C.hadm_id, C.icustay_id,
admission_date, icustay_intime, discharge_date, dod

FROM cohort3 C WHERE hadm_id NOT IN (SELECT DISTINCT C.hadm_id

FROM tenor_ab T, cohort3 C WHERE C.hadm_id = T.hadm_id AND start_dt <=
icustay_intime + interval '6 hours');
```

firstnote1 – the initial nursing note (must be within 24 hours) of each of the 17005 hospitalizations

```
CREATE VIEW tenor_nlp AS

SELECT hadm_id FROM firstnote1 WHERE lower(text)

SIMILAR TO
'%amikacin%|%genta%|%kanamycin%|%netilmicin%|%tobramycin%|%paromomycin%|%spectinomy
cin%|%geldanamycin%|%ertapenem%|%doripenem%|%imipenem%|%meropenem%|%merepenum%|%cef
adroxil%|%cefalexin%|%cefaclor%|%cefoxitin%|%cefprozil%|%cefamandole%|%cefuroxime%|
%cefixime%|%cefotaxime%|%cefpodoxime%|%ceftazidime%|%ceftriaxone%|%cefepime%|%cefam
pine%|% vanco %|% vanc
%|%vancomycin%|%clindamycin%|%clinda%|%daptomycin%|%azithromycin%|%clarithromycin%|
%erythromycin%|%telithromycin%|%aztreonam%|%nitrofurantoin%|%linezolid%|%amoxicilli
```

```
n%|%ampicillin%|%ampcillin%|%dicloxacillin%|%flucloxacillin%|%methicillin%|%nafcill
in%|%oxacillin%|%penicillin%|%piperacillin%|%pipercillin%|%cefotetan%|%ticarcillin%
|%timentin%|%colistin%|%bactrim%|%polymyxin%|%ciprofloxacin%|% cipro
%|%gatifloxacin%|%levofloxacin%|%levoflox%|%levoquin%|%levaquin%|%moxifloxacin%|%na
lidixic
acid%|%norfloxacin%|%ofloxacin%|%trovafloxacin%|%sulfadiazine%|%sulfamethoxazole%|%
trimethoprim%|%TMP%|%doxycycline%|%minocycline%|%tetracycline%|%dapsone%|%ethambuto
l%|%isoniazid%|%pyrazinamide%|%rifampicin%|%rifampin%|%rifabutin%|%streptomycin%|%c
hloramphenicol%|%synercid%|%fosfomycin%|%metronidazole%|%linezoid%|%quinupristin%|%
tigecycline%|%bacteremia%|%abcess%|%ceftrioaxone%|%abscess%|%meropenium%|% gent
%|%antibiotic%|%antibx%|%abx%|%anbx%|%cdiff%|%septic%|%sepsis%|%cellulitis%|%ancef%
|% ampi
%|%ampi,%|%ampacillin%|%rocephin%|%/gent%|%urosepsis%|%zosyn%|%infection%|%pneumoni
a%|%pnuemonia%|% uti %|%flagyl%|%falgyl%|%unasyn%';
```

4751 results who might have gotten abx during the first 6 hours based on NLP

10032 hospitalizations of pts not started on abx during the first 6 hours based on a combination of initial nursing note + inpatient orders:

```
CREATE VIEW no_abx AS

SELECT * FROM (

SELECT DISTINCT ON (C.hadm_id) C.subject_id, C.hadm_id, C.icustay_id,
admission_date, icustay_intime, discharge_date, dod

FROM cohort3 C WHERE hadm_id NOT IN (SELECT DISTINCT C.hadm_id

FROM tenor_ab T, cohort3 C WHERE C.hadm_id = T.hadm_id AND start_dt <=
icustay_intime + interval '6 hours')) AS foo

WHERE hadm_id NOT IN (SELECT * FROM tenor_nlp);
```

Removal of duplicate hospitalizations while maintaining pts started late on abx:

```
CREATE TABLE tenor_cohort AS

SELECT DISTINCT ON (C.subject_id) C.subject_id, C.hadm_id, C.icustay_id,
admission_date, icustay_intime, discharge_date, dod

FROM no_abx C, tenor_ab A WHERE C.hadm_id = A.hadm_id AND start_dt <=
icustay_intime + interval '48 hours'

UNION

SELECT DISTINCT ON (subject_id) * FROM no_abx WHERE subject_id NOT IN (SELECT
subject_id FROM tenor_late);
```

Results in 9478 unique pts that were not started on abx within the first 6 hours

```
CREATE TABLE tenor_late AS

SELECT DISTINCT ON (C.subject_id) C.subject_id, C.hadm_id, C.icustay_id,
admission_date, icustay_intime, discharge_date, dod

FROM no_abx C, tenor_ab A WHERE C.hadm_id = A.hadm_id AND start_dt <=
icustay_intime + interval '48 hours';
```

outcome – pts started on abx late

Creation of training and test sets

random selection of 80% of the 9478 pts above:

```
CREATE TABLE tenor_train AS

SELECT * FROM tenor_cohort ORDER BY random() LIMIT 7582;
```

no missing values for subject_id, hadm_id, or icustay_id

20%:

```
CREATE TABLE lucky_test AS

SELECT * FROM tenor_cohort WHERE subject_id NOT IN (SELECT subject_id FROM
tenor_train);
```

in training set:

```
SELECT DISTINCT ON (C.subject_id) C.subject_id, C.hadm_id, C.icustay_id,
admission_date, icustay_intime, discharge_date, dod

FROM tenor_train C, tenor_ab A WHERE C.hadm_id = A.hadm_id AND start_dt <=
icustay_intime + interval '48 hours';
```

1105 pts started late on abx. Out of 7582, this is 14.6%.

Extraction of labs

```
CREATE TABLE tenor_labs AS

SELECT C.subject_id, C.hadm_id, L.itemid, icustay_intime, charttime, value,
valuenum, test_name, fluid, loinc_description

FROM labevents L, d_labitems D, tenor_train C WHERE L.itemid = D.itemid AND
C.subject_id = L.subject_id
```

```
AND charttime <= icustay_intime + interval '4 hours' AND charttime >=
icustay_intime - interval '4 hours';
```

all labs in the training set obtained 4 hours before or after the icu start time

```
SELECT COUNT(DISTINCT subject_id) FROM tenor_labs;

SELECT COUNT(DISTINCT hadm_id) FROM tenor_labs;
```

both return 7121, confirmation that no duplicate subject_ids or hadm_ids

```
CREATE TABLE tenor_pivot AS

SELECT subject_id,

MAX(CASE WHEN itemid = 50018 THEN value ELSE NULL END) AS "ph",

MAX(CASE WHEN itemid = 50019 THEN value ELSE NULL END) AS "po2",

MAX(CASE WHEN itemid = 50002 THEN value ELSE NULL END) AS "baseexcess_abg",

MAX(CASE WHEN itemid = 50016 THEN value ELSE NULL END) AS "pco2_abg",

MAX(CASE WHEN itemid = 50025 THEN value ELSE NULL END) AS "bicarb_abg1",

MAX(CASE WHEN itemid = 50009 THEN value ELSE NULL END) AS "potassium_abg",

MAX(CASE WHEN itemid = 50006 THEN value ELSE NULL END) AS "glucose_abg",

MAX(CASE WHEN itemid = 50383 THEN value ELSE NULL END) AS "hematocrit",

MAX(CASE WHEN itemid = 50428 THEN value ELSE NULL END) AS "platelet",

MAX(CASE WHEN itemid = 50029 THEN value ELSE NULL END) AS "hematocrit_abg",

MAX(CASE WHEN itemid = 50007 THEN value ELSE NULL END) AS "hemoglobin_abg",

MAX(CASE WHEN itemid = 50386 THEN value ELSE NULL END) AS "hemoglobin",

MAX(CASE WHEN itemid = 50468 THEN value ELSE NULL END) AS "leukocytes",

MAX(CASE WHEN itemid = 50412 THEN value ELSE NULL END) AS "mchc",

MAX(CASE WHEN itemid = 50411 THEN value ELSE NULL END) AS "mch",

MAX(CASE WHEN itemid = 50413 THEN value ELSE NULL END) AS "mcv",

MAX(CASE WHEN itemid = 50442 THEN value ELSE NULL END) AS "erythrocytes",
```

```
MAX(CASE WHEN itemid = 50090 THEN value ELSE NULL END) AS "creatinine",

MAX(CASE WHEN itemid = 50177 THEN value ELSE NULL END) AS "bun",

MAX(CASE WHEN itemid = 50444 THEN value ELSE NULL END) AS "rdw",

MAX(CASE WHEN itemid = 50399 THEN value ELSE NULL END) AS "inr",

MAX(CASE WHEN itemid = 50439 THEN value ELSE NULL END) AS "pt",

MAX(CASE WHEN itemid = 50440 THEN value ELSE NULL END) AS "ptt",

MAX(CASE WHEN itemid = 50172 THEN value ELSE NULL END) AS "bicarb",

MAX(CASE WHEN itemid = 50083 THEN value ELSE NULL END) AS "chloride",

MAX(CASE WHEN itemid = 50030 THEN value ELSE NULL END) AS "ionca_abg",

MAX(CASE WHEN itemid = 50149 THEN value ELSE NULL END) AS "potassium",

MAX(CASE WHEN itemid = 50159 THEN value ELSE NULL END) AS "sodium",

MAX(CASE WHEN itemid = 50112 THEN value ELSE NULL END) AS "glucose",

MAX(CASE WHEN itemid = 50012 THEN value ELSE NULL END) AS "sodium_abg",

MAX(CASE WHEN itemid = 50068 THEN value ELSE NULL END) AS "ag",

MAX(CASE WHEN itemid = 50010 THEN value ELSE NULL END) AS "lactate",

MAX(CASE WHEN itemid = 50140 THEN value ELSE NULL END) AS "magnesium",

MAX(CASE WHEN itemid = 50079 THEN value ELSE NULL END) AS "calcium",

MAX(CASE WHEN itemid = 50148 THEN value ELSE NULL END) AS "phosphate",

MAX(CASE WHEN itemid = 50086 THEN value ELSE NULL END) AS "cpk",

MAX(CASE WHEN itemid = 50149 THEN value ELSE NULL END) AS "neutrophils",

MAX(CASE WHEN itemid = 50333 THEN value ELSE NULL END) AS "basophils",

MAX(CASE WHEN itemid = 50417 THEN value ELSE NULL END) AS "monocytes",

MAX(CASE WHEN itemid = 50408 THEN value ELSE NULL END) AS "lymphocytes",

MAX(CASE WHEN itemid = 50373 THEN value ELSE NULL END) AS "eosinophils",

MAX(CASE WHEN itemid = 50087 THEN value ELSE NULL END) AS "ckmb",

MAX(CASE WHEN itemid = 50623 THEN value ELSE NULL END) AS "urine_app",
```

```
MAX(CASE WHEN itemid = 50626 THEN value ELSE NULL END) AS "urine_bili",

MAX(CASE WHEN itemid = 50627 THEN value ELSE NULL END) AS "urine_hemoglobin",

MAX(CASE WHEN itemid = 50653 THEN value ELSE NULL END) AS "urine_ph",

MAX(CASE WHEN itemid = 50655 THEN value ELSE NULL END) AS "urine_protein",

MAX(CASE WHEN itemid = 50633 THEN value ELSE NULL END) AS "urine_color",

MAX(CASE WHEN itemid = 50650 THEN value ELSE NULL END) AS "urine_nitrite",

MAX(CASE WHEN itemid = 50661 THEN value ELSE NULL END) AS "urine_grav",

MAX(CASE WHEN itemid = 50647 THEN value ELSE NULL END) AS "urine_ketone",

MAX(CASE WHEN itemid = 50641 THEN value ELSE NULL END) AS "urine_glucose",

MAX(CASE WHEN itemid = 50671 THEN value ELSE NULL END) AS "urine_urobilinogen",

MAX(CASE WHEN itemid = 50648 THEN value ELSE NULL END) AS "urine_leukocytes",

MAX(CASE WHEN itemid = 50189 THEN value ELSE NULL END) AS "troponin",

MAX(CASE WHEN itemid = 50065 THEN value ELSE NULL END) AS "amylase",

MAX(CASE WHEN itemid = 50015 THEN value ELSE NULL END) AS "o2sat_abg",

MAX(CASE WHEN itemid = 50656 THEN value ELSE NULL END) AS "urinesed_eryth",

MAX(CASE WHEN itemid = 50677 THEN value ELSE NULL END) AS "urinesed_yeast",

MAX(CASE WHEN itemid = 50674 THEN value ELSE NULL END) AS "urinesed_leuko",

MAX(CASE WHEN itemid = 50624 THEN value ELSE NULL END) AS "urinesed_bact",

MAX(CASE WHEN itemid = 50637 THEN value ELSE NULL END) AS "urinesed_epi",

MAX(CASE WHEN itemid = 50378 THEN value ELSE NULL END) AS "fibrinogen",

MAX(CASE WHEN itemid = 50013 THEN value ELSE NULL END) AS "o2satinsp_abg",

MAX(CASE WHEN itemid = 50024 THEN value ELSE NULL END) AS "tv_abg",

MAX(CASE WHEN itemid = 50056 THEN value ELSE NULL END) AS "acetaminophen",

MAX(CASE WHEN itemid = 50072 THEN value ELSE NULL END) AS "asa",

MAX(CASE WHEN itemid = 50099 THEN value ELSE NULL END) AS "ethanol",

MAX(CASE WHEN itemid = 50187 THEN value ELSE NULL END) AS "benzo_serum",
```

```
MAX(CASE WHEN itemid = 50198 THEN value ELSE NULL END) AS "tca_serum",

MAX(CASE WHEN itemid = 50186 THEN value ELSE NULL END) AS "barb_serum",

MAX(CASE WHEN itemid = 50062 THEN value ELSE NULL END) AS "alt",

MAX(CASE WHEN itemid = 50073 THEN value ELSE NULL END) AS "ast",

MAX(CASE WHEN itemid = 50170 THEN value ELSE NULL END) AS "bili_total",

MAX(CASE WHEN itemid = 50061 THEN value ELSE NULL END) AS "alk_phos",

MAX(CASE WHEN itemid = 50193 THEN value ELSE NULL END) AS "gfr",

MAX(CASE WHEN itemid = 50291 THEN value ELSE NULL END) AS "benzo_urine",

MAX(CASE WHEN itemid = 50290 THEN value ELSE NULL END) AS "barb_urine",

MAX(CASE WHEN itemid = 50289 THEN value ELSE NULL END) AS "amphet_urine",

MAX(CASE WHEN itemid = 50296 THEN value ELSE NULL END) AS "opiate_urine",

MAX(CASE WHEN itemid = 50292 THEN value ELSE NULL END) AS "cocaine_urine",

MAX(CASE WHEN itemid = 50295 THEN value ELSE NULL END) AS "methadone_urine",

MAX(CASE WHEN itemid = 50021 THEN value ELSE NULL END) AS "reqo2_abg",

MAX(CASE WHEN itemid = 50001 THEN value ELSE NULL END) AS "aagrad_abg",

MAX(CASE WHEN itemid = 50396 THEN value ELSE NULL END) AS "hypochromia",

MAX(CASE WHEN itemid = 50490 THEN value ELSE NULL END) AS "macrocytes",

MAX(CASE WHEN itemid = 50415 THEN value ELSE NULL END) AS "microcytes",

MAX(CASE WHEN itemid = 50326 THEN value ELSE NULL END) AS "anisocytosis",

MAX(CASE WHEN itemid = 50332 THEN value ELSE NULL END) AS "bands",

MAX(CASE WHEN itemid = 50060 THEN value ELSE NULL END) AS "albumin",

MAX(CASE WHEN itemid = 50431 THEN value ELSE NULL END) AS "poikilocytosis",

MAX(CASE WHEN itemid = 50022 THEN value ELSE NULL END) AS "bicarb_abg2",

MAX(CASE WHEN itemid = 50138 THEN value ELSE NULL END) AS "lipase",

MAX(CASE WHEN itemid = 50429 THEN value ELSE NULL END) AS "platelet_manual",

MAX(CASE WHEN itemid = 50432 THEN value ELSE NULL END) AS "polychromasia",
```

```
MAX(CASE WHEN itemid = 50023 THEN value ELSE NULL END) AS "temp_abg"

FROM tenor_labs GROUP BY subject_id;
```

First version of code which does not select first lab for each patient

```
CREATE TABLE tenor_labs1 AS

SELECT DISTINCT ON (subject_id, itemid, charttime) C.subject_id, C.hadm_id,
L.itemid, icustay_intime, charttime, value, valuenum, test_name, fluid,
loinc_description

FROM labevents L, d_labitems D, tenor_train C WHERE L.itemid = D.itemid AND
C.subject_id = L.subject_id

AND charttime <= icustay_intime + interval '4 hours' AND charttime >=
icustay_intime - interval '4 hours'

ORDER BY subject_id, itemid, charttime;
```

fixes the problem of not selecting the first lab when multiple of the same lab is available,

will be used in the next step

Query returned successfully: 409564 rows affected, 36668 ms execution time.

New tenor pivot:

```
CREATE VIEW tenor_pivot1 AS

SELECT subject_id,

MAX(CASE WHEN itemid = 50018 THEN value ELSE NULL END) AS "ph",

MAX(CASE WHEN itemid = 50019 THEN value ELSE NULL END) AS "po2",

MAX(CASE WHEN itemid = 50002 THEN value ELSE NULL END) AS "baseexcess_abg",

MAX(CASE WHEN itemid = 50016 THEN value ELSE NULL END) AS "pco2_abg",

MAX(CASE WHEN itemid = 50428 THEN value ELSE NULL END) AS "platelet",

MAX(CASE WHEN itemid = 50468 THEN value ELSE NULL END) AS "leukocytes",

MAX(CASE WHEN itemid = 50412 THEN value ELSE NULL END) AS "mchc",

MAX(CASE WHEN itemid = 50411 THEN value ELSE NULL END) AS "mch",
```

```
MAX(CASE WHEN itemid = 50413 THEN value ELSE NULL END) AS "mcv",

MAX(CASE WHEN itemid = 50442 THEN value ELSE NULL END) AS "erythrocytes",

MAX(CASE WHEN itemid = 50090 THEN value ELSE NULL END) AS "creatinine",

MAX(CASE WHEN itemid = 50177 THEN value ELSE NULL END) AS "bun",

MAX(CASE WHEN itemid = 50444 THEN value ELSE NULL END) AS "rdw",

MAX(CASE WHEN itemid = 50399 THEN value ELSE NULL END) AS "inr",

MAX(CASE WHEN itemid = 50439 THEN value ELSE NULL END) AS "pt",

MAX(CASE WHEN itemid = 50440 THEN value ELSE NULL END) AS "ptt",

MAX(CASE WHEN itemid = 50083 THEN value ELSE NULL END) AS "chloride",

MAX(CASE WHEN itemid = 50068 THEN value ELSE NULL END) AS "ag",

MAX(CASE WHEN itemid = 50010 THEN value ELSE NULL END) AS "lactate",

MAX(CASE WHEN itemid = 50140 THEN value ELSE NULL END) AS "magnesium",

MAX(CASE WHEN itemid = 50079 THEN value ELSE NULL END) AS "calcium",

MAX(CASE WHEN itemid = 50148 THEN value ELSE NULL END) AS "phosphate",

MAX(CASE WHEN itemid = 50086 THEN value ELSE NULL END) AS "cpk",

MAX(CASE WHEN itemid = 50149 THEN value ELSE NULL END) AS "neutrophils",

MAX(CASE WHEN itemid = 50408 THEN value ELSE NULL END) AS "lymphocytes",

MAX(CASE WHEN itemid = 50373 THEN value ELSE NULL END) AS "eosinophils",

MAX(CASE WHEN itemid = 50653 THEN value ELSE NULL END) AS "urine_ph",

MAX(CASE WHEN itemid = 50189 THEN value ELSE NULL END) AS "troponin",

MAX(CASE WHEN itemid = 50065 THEN value ELSE NULL END) AS "amylase",

MAX(CASE WHEN itemid = 50378 THEN value ELSE NULL END) AS "fibrinogen",

MAX(CASE WHEN itemid = 50056 THEN value ELSE NULL END) AS "acetaminophen",

MAX(CASE WHEN itemid = 50072 THEN value ELSE NULL END) AS "asa",

MAX(CASE WHEN itemid = 50099 THEN value ELSE NULL END) AS "ethanol",

MAX(CASE WHEN itemid = 50187 THEN value ELSE NULL END) AS "benzo_serum",
```

```
MAX(CASE WHEN itemid = 50198 THEN value ELSE NULL END) AS "tca_serum",

MAX(CASE WHEN itemid = 50186 THEN value ELSE NULL END) AS "barb_serum",

MAX(CASE WHEN itemid = 50062 THEN value ELSE NULL END) AS "alt",

MAX(CASE WHEN itemid = 50073 THEN value ELSE NULL END) AS "ast",

MAX(CASE WHEN itemid = 50170 THEN value ELSE NULL END) AS "bili_total",

MAX(CASE WHEN itemid = 50291 THEN value ELSE NULL END) AS "benzo_urine",

MAX(CASE WHEN itemid = 50290 THEN value ELSE NULL END) AS "barb_urine",

MAX(CASE WHEN itemid = 50289 THEN value ELSE NULL END) AS "amphet_urine",

MAX(CASE WHEN itemid = 50296 THEN value ELSE NULL END) AS "opiate_urine",

MAX(CASE WHEN itemid = 50292 THEN value ELSE NULL END) AS "cocaine_urine",

MAX(CASE WHEN itemid = 50295 THEN value ELSE NULL END) AS "methadone_urine",

MAX(CASE WHEN itemid = 50396 THEN value ELSE NULL END) AS "hypochromia",

MAX(CASE WHEN itemid = 50332 THEN value ELSE NULL END) AS "bands",

MAX(CASE WHEN itemid = 50060 THEN value ELSE NULL END) AS "albumin",

MAX(CASE WHEN itemid = 50138 THEN value ELSE NULL END) AS "lipase"

FROM tenor_labs1 GROUP BY subject_id;
```

Extraction of vital signs

```
CREATE TABLE tenor_temp AS

SELECT C.subject_id, C.icustay_id, itemid, charttime, value1num AS value5num FROM
chartevents C, tenor_train T WHERE itemid IN (676,677)

AND c.subject_id = T.subject_id AND charttime < icustay_intime + interval '4 hours'
AND charttime > icustay_intime - interval '4 hours'

UNION

SELECT C.subject_id, C.icustay_id, itemid, charttime, (value1num-32)*5/9 AS
value5num FROM chartevents C, tenor_train T

WHERE itemid IN (678,679) AND C.subject_id = T.subject_id AND charttime <
icustay_intime + interval '4 hours'

AND charttime > icustay_intime - interval '4 hours';
```

Query returned successfully: 29861 rows affected, 845718 ms execution time.

```
CREATE VIEW tenor_temp1 AS

SELECT subject_id, value5num FROM tenor_temp WHERE value5num > 20

UNION

SELECT subject_id, valuenum FROM tenor_labs1 WHERE itemid = 50023 AND valuenum IS
NOT NULL;

subject_id, value



CREATE VIEW tenor_maxtemp AS

SELECT DISTINCT ON (subject_id) * FROM

(SELECT subject_id, value5num FROM tenor_temp1

ORDER BY subject_id, value5num DESC) AS foo;



CREATE VIEW tenor_mintemp AS

SELECT DISTINCT ON (subject_id) * FROM

(SELECT subject_id, value5num FROM tenor_temp1

ORDER BY subject_id, value5num ASC) AS foo;



CREATE VIEW tenor_meantemp AS

SELECT subject_id, AVG(value5num) FROM tenor_temp1

GROUP BY subject_id;



CREATE VIEW tenor_rrmax AS

SELECT DISTINCT ON (subject_id) * FROM

(SELECT C.subject_id, C.icustay_id, itemid, icustay_intime, charttime, value1num
FROM chartevents C, tenor_train T
```

```
WHERE itemid = 618 AND T.subject_id = C.subject_id AND charttime < icustay_intime +
interval '4 hours'

AND charttime > icustay_intime - interval '4 hours' AND value1num != 0 ORDER BY
subject_id, value1num DESC) AS foo;

subject_id, value



CREATE VIEW tenor_rrmean AS

SELECT T.subject_id, AVG(value1num) AS average FROM chartevents C, tenor_train T

WHERE itemid = 618 AND C.subject_id = T.subject_id AND charttime < icustay_intime +
interval '4 hours'

AND charttime > icustay_intime - interval '4 hours' AND value1num != 0 GROUP BY
T.subject_id;




CREATE VIEW tenor_hrmax AS

SELECT T.subject_id, MAX(value1num) AS HR FROM chartevents c, tenor_train T

WHERE itemid = 211 AND T.subject_id = C.subject_id AND charttime < icustay_intime +
interval '4 hours'

AND charttime > icustay_intime - interval '4 hours' AND value1num != 0 GROUP BY
T.subject_id;



CREATE VIEW tenor_hrmin AS

SELECT T.subject_id, MIN(value1num) AS HR FROM chartevents T, tenor_train C

WHERE itemid = 211 AND T.subject_id = C.subject_id AND charttime < icustay_intime +
interval '4 hours'

AND charttime > icustay_intime - interval '4 hours' AND value1num != 0 GROUP BY
T.subject_id;



CREATE VIEW tenor_hrmean AS

SELECT T.subject_id, AVG(value1num) AS HR FROM chartevents C, tenor_train T
```

```
WHERE itemid = 211 AND T.subject_id = C.subject_id AND charttime < icustay_intime +
interval '4 hours'

AND charttime > icustay_intime - interval '4 hours' AND value1num != 0 GROUP BY
T.subject_id;



CREATE VIEW tenor_BP AS

SELECT C.subject_id, min(value1num) AS minbpsys, min(value2num) AS minbpdias,
avg(value1num) AS avgsysbp, avg(value2num) AS avgdiasbp

FROM chartevents C, tenor_train T WHERE C.subject_id = T.subject_id AND

itemid IN (51,455) AND value1num >= 40 AND value2num >= 20 AND charttime >
icustay_intime - interval '4 hours'

AND charttime < icustay_intime + interval '4 hours' GROUP BY C.subject_id;



CREATE VIEW tenor_saturation AS

SELECT C.subject_id, min(value1num) AS minsat, avg(value1num) AS avgsat,
max(value1num) AS maxsat

FROM chartevents C, tenor_train T WHERE C.subject_id = T.subject_id AND

itemid IN (1148, 646, 834) AND charttime > icustay_intime - interval '4 hours' AND
charttime < icustay_intime + interval '4 hours'

AND value1num >= 30

GROUP BY C.subject_id;



CREATE TABLE initialsat AS

SELECT DISTINCT ON (C.subject_id) C.subject_id, realtime, value1num AS sat

FROM chartevents C, tenor_train T WHERE C.subject_id = T.subject_id AND

itemid IN (1148, 646, 834) AND charttime > icustay_intime - interval '4 hours' AND
charttime < icustay_intime + interval '4 hours'

AND value1num >= 30 ORDER BY subject_id, realtime;
```

Extraction of demographic and other variables

```
CREATE VIEW tenor_age AS

SELECT T.subject_id, MAX(CASE WHEN icustay_admit_age >103 THEN 103 ELSE
icustay_admit_age END) AS age

FROM icustay_detail I, tenor_train T WHERE I.subject_id = T.subject_id AND
I.icustay_intime = T.icustay_intime GROUP BY T.subject_id;

CREATE VIEW tenor_gender AS

SELECT T.subject_id, MAX(CASE WHEN gender='M' THEN 0 WHEN gender='F' THEN 1 ELSE
NULL END) AS gender

FROM icustay_detail I, tenor_train T WHERE I.subject_id = T.subject_id AND
I.icustay_intime = T.icustay_intime GROUP BY T.subject_id;

CREATE VIEW tenor_demographic AS

SELECT subject_id, hadm_id, ethnicity_descr, overall_payor_group_descr FROM
demographic_detail WHERE hadm_id IN

(SELECT hadm_id FROM tenor_train);

CREATE VIEW tenor_icu AS

SELECT subject_id, hadm_id, icustay_id, icustay_first_careunit,
icustay_first_service, sofa_first FROM icustay_detail

WHERE icustay_id IN (SELECT icustay_id FROM tenor_train);
```

Combining labs with multiple mappings

```
CREATE VIEW tenor_hematocrit AS

SELECT DISTINCT ON (subject_id) * FROM tenor_labs WHERE itemid IN (50383,50029)

ORDER BY subject_id, charttime;

CREATE VIEW tenor_bicarb AS

SELECT DISTINCT ON (subject_id) * FROM tenor_labs WHERE itemid IN (50025, 50172,
50022)

ORDER BY subject_id, charttime;

CREATE VIEW tenor_potassium AS

SELECT DISTINCT ON (subject_id) * FROM tenor_labs WHERE itemid IN (50009, 50149)

ORDER BY subject_id, charttime;

CREATE VIEW tenor_glucose AS
```

```
SELECT DISTINCT ON (subject_id) * FROM tenor_labs WHERE itemid IN (50006, 50112)

ORDER BY subject_id, charttime;

CREATE VIEW tenor_hemoglobin AS

SELECT DISTINCT ON (subject_id) * FROM tenor_labs WHERE itemid IN (50007, 50386)

ORDER BY subject_id, charttime;

CREATE VIEW tenor_sodium AS

SELECT DISTINCT ON (subject_id) * FROM tenor_labs WHERE itemid IN (50159, 50012)

ORDER BY subject_id, charttime;
```

Combining all data into one table

```
CREATE TABLE tenor1 AS

SELECT tenor_train.*,

age,

gender,

ethnicity_descr,

overall_payor_group_descr,

icustay_first_careunit,

icustay_first_service,

sofa_first,

tenor_maxtemp.value5num AS max_temp,

tenor_mintemp.value5num AS min_temp,

tenor_meantemp.avg AS mean_temp,

tenor_rrmax.value1num AS rr_max,

tenor_rrmean.average AS rr_mean,

tenor_hrmax.HR AS hr_max,

tenor_hrmin.HR AS hr_min,
```

```
tenor_hrmean.HR AS hr_mean,

minbpsys,

minbpdias,

avgsysbp,

avgdiasbp,

minsat,

avgsat,

maxsat,

platelet,

creatinine,

bun,

leukocytes,

mchc,

mch,

mcv,

erythrocytes,

rdw,

pt,

inr,

ptt,

chloride,

tenor_hematocrit.valuenum AS hematocrit,

tenor_bicarb.valuenum AS bicarb,

tenor_potassium.valuenum AS potassium,

tenor_glucose.valuenum AS glucose,

tenor_hemoglobin.valuenum AS hemoglobin,
```

```
tenor_sodium.valuenum AS sodium,

neutrophils,

ag,

ph,

eosinophils,

lactate,

baseexcess_abg,

magnesium,

cpk,

urine_ph,

fibrinogen,

amylase,

lipase,

alt,

FROM tenor_train

LEFT JOIN tenor_maxtemp ON tenor_train.subject_id = tenor_maxtemp.subject_id

LEFT JOIN tenor_mintemp ON tenor_train.subject_id = tenor_mintemp.subject_id

LEFT JOIN tenor_meantemp ON tenor_train.subject_id = tenor_meantemp.subject_id

LEFT JOIN tenor_rrmax ON tenor_train.subject_id = tenor_rrmax.subject_id

LEFT JOIN tenor_rrmean ON tenor_train.subject_id = tenor_rrmean.subject_id

LEFT JOIN tenor_hrmax ON tenor_train.subject_id = tenor_hrmax.subject_id

LEFT JOIN tenor_hrmin ON tenor_train.subject_id = tenor_hrmin.subject_id

LEFT JOIN tenor_hrmean ON tenor_train.subject_id = tenor_hrmean.subject_id

LEFT JOIN tenor_BP ON tenor_train.subject_id = tenor_BP.subject_id

LEFT JOIN tenor_saturation ON tenor_train.subject_id = tenor_saturation.subject_id

LEFT JOIN tenor_age ON tenor_train.subject_id = tenor_age .subject_id
```

```
LEFT JOIN tenor_gender ON tenor_train.subject_id = tenor_gender.subject_id

LEFT JOIN tenor_demographic ON tenor_train.subject_id =
tenor_demographic.subject_id

LEFT JOIN tenor_icu ON tenor_train.subject_id = tenor_icu.subject_id

LEFT JOIN tenor_hematocrit ON tenor_train.subject_id = tenor_hematocrit.subject_id

LEFT JOIN tenor_bicarb ON tenor_train.subject_id = tenor_bicarb.subject_id

LEFT JOIN tenor_potassium ON tenor_train.subject_id = tenor_potassium.subject_id

LEFT JOIN tenor_glucose ON tenor_train.subject_id = tenor_glucose.subject_id

LEFT JOIN tenor_hemoglobin ON tenor_train.subject_id = tenor_hemoglobin.subject_id

LEFT JOIN tenor_sodium ON tenor_train.subject_id = tenor_sodium.subject_id

LEFT JOIN tenor_pivot1 ON tenor_train.subject_id = tenor_pivot1.subject_id

ORDER BY tenor_train.subject_id;
```

list of tables that were inserted above:

```
tenor_maxtemp

tenor_mintemp

tenor_meantemp

tenor_rrmax

tenor_rrmean

tenor_hrmax

tenor_hrmin

tenor_hrmean

tenor_BP

tenor_saturation

tenor_age

tenor_gender

tenor_demographic
```

```
tenor_icu

tenor_hematocrit

tenor_bicarb

tenor_potassium

tenor_glucose

tenor_hemoglobin

tenor_sodium

tenor_pivot1
```

with outcome variable:

```
CREATE VIEW tenor_cohort1 AS

SELECT subject_id, hadm_id, icustay_id, admission_date, icustay_intime,
discharge_date, dod,

CASE WHEN hadm_id IN (SELECT DISTINCT hadm_id FROM tenor_late) THEN 1 ELSE 0 END AS
outcome

FROM tenor_cohort;
```

includes both train and test sets

```
CREATE VIEW tenor2 AS

SELECT tenor1.*, outcome FROM tenor1 LEFT JOIN tenor_cohort1 ON tenor1.subject_id =
tenor_cohort1.subject_id;

the training set with the outcome

CREATE VIEW tenor3 AS

SELECT tenor2.*, albumin, sat FROM tenor2

LEFT JOIN tenor_pivot1 ON tenor2.subject_id = tenor_pivot1.subject_id

LEFT JOIN initialsat ON tenor2.subject_id = initialsat.subject_id;



CREATE TABLE havelabs1 AS
```

```
SELECT * FROM tenor3 WHERE hadm_id IN (

SELECT DISTINCT L.hadm_id FROM labevents L, tenor_train C WHERE C.subject_id =
L.subject_id

AND charttime <= icustay_intime + interval '4 hours' AND charttime >=
icustay_intime - interval '4 hours');
```

Results

```
SELECT DISTINCT ON (subject_id) * FROM tenor_labs WHERE itemid = 50091 AND valuenum
<= 10.0 ORDER BY subject_id, charttime;
```

pts ordered for CRP. Only 21 pts out of 7121 in the training set. 0.3% of pts during the first 4 hrs, but 48% of values normal.

```
SELECT DISTINCT hadm_id FROM labevents WHERE itemid = 50373 and hadm_id IN (SELECT
hadm_id FROM tenor_train);
```

--4508 rows (eosinophils during the hospitalization)

```
SELECT DISTINCT hadm_id FROM labevents WHERE hadm_id IN (SELECT hadm_id FROM
tenor_train);
```

--7540 have labs during the hospitalization

-- means 61% get it during the hospitalization

```
SELECT DISTINCT L.hadm_id FROM labevents L, tenor_train C WHERE C.subject_id =
L.subject_id

AND charttime <= icustay_intime + interval '4 hours' AND charttime >=
icustay_intime - interval '4 hours';
```

--7026 pts with a lab during the first 4 hours

so 41% got eosinophils during the first 4 hours

```
SELECT COUNT(DISTINCT hadm_id) FROM microbiologyevents WHERE org_itemid != 80001
AND hadm_id IN (SELECT hadm_id FROM tenor2
```

```
WHERE outcome = 0);
```

--1420/6477 positive cxs in negative group, 21.9%

```
SELECT COUNT(DISTINCT hadm_id) FROM microbiologyevents WHERE org_itemid != 80001
AND hadm_id IN (SELECT hadm_id FROM tenor2

WHERE outcome = 1);
```

--510/1105 in positive group, 46.2%

mortality calculation:

```
SELECT COUNT(*) FROM tenor2 WHERE outcome = 0 AND discharge_date = dod;
```

550/6477 = 8.5%

```
SELECT COUNT(*) FROM tenor2 WHERE outcome = 1 AND discharge_date = dod;

138/1105 = 12.5%
```

so about 50% higher mortality

Creation of initial cohorts:

```
CREATE cohort1 AS

SELECT DISTINCT ON (hadm_id) hadm_id, hospital_admit_dt, icustay_intime, difference
FROM admissiontimes

ORDER BY hadm_id, icustay_intime;


CREATE VIEW cohort2 AS

SELECT * FROM cohort1 WHERE difference <'30:00:00';


SELECT C.hadm_id, C.hospital_admit_dt, C.icustay_intime, difference FROM cohort2 C,
icustay_detail I

WHERE C.hadm_id = I.hadm_id AND C.icustay_intime = I.icustay_intime AND
icustay_age_group = 'adult';
```

```
      (above code must match on both hospital ID AND icustay_start time in order to
avoid creating    duplicate patients)


SELECT C.hadm_id, C.hospital_admit_dt, C.icustay_intime, difference FROM cohort2 C,
icustay_detail I, demographic_detail D

WHERE C.hadm_id = I.hadm_id AND C.icustay_intime = I.icustay_intime AND
icustay_age_group = 'adult' AND C.hadm_id = D.hadm_id

AND admission_source_itemid != 200074;
```

Addition of pts not linked from icustay_detail, code written to link:

```
SELECT M.hadm_id, M.subject_id, admit_dt, icustay_intime, icustay_intime - admit_dt
AS difference, icustay_age_group

FROM missingadmits M, icustay_detail I, demographic_detail D WHERE M.subject_id =
I.subject_id

AND icustay_intime - admit_dt <'30:00:00' AND icustay_intime >= admit_dt AND
icustay_age_group = 'adult'

AND M.hadm_id = D.hadm_id AND admission_source_itemid != 200074;



CREATE VIEW cohort3 AS

SELECT A.subject_id, C.hadm_id, icustay_id, C.hospital_admit_dt AS admission_date,
C.icustay_intime, disch_dt AS discharge_date, dod

FROM cohort2 C, icustay_detail I, demographic_detail D, admissions A

WHERE C.hadm_id = I.hadm_id AND A.hadm_id = C.hadm_id AND C.icustay_intime =
I.icustay_intime AND icustay_age_group = 'adult'

AND C.hadm_id = D.hadm_id AND admission_source_itemid != 200074 AND icustay_seq = 1

UNION

SELECT M.subject_id, M.hadm_id, icustay_id, admit_dt AS admission_date,
icustay_intime, disch_dt AS discharge_date, dod

FROM missingadmits M, icustay_detail I, demographic_detail D WHERE M.subject_id =
I.subject_id AND icustay_intime - admit_dt <'30:00:00'

AND icustay_intime >= admit_dt AND icustay_age_group = 'adult' AND M.hadm_id =
D.hadm_id AND admission_source_itemid != 200074

AND icustay_seq = 1;
```

Code to combine previous cohort with manually linked patients.

Code to determine inpatient antibiotic orders:

```
SELECT COUNT(DISTINCT C.hadm_id) FROM poe_order P, cohort6 C

WHERE C.hadm_id = P.hadm_id AND start_dt <= icustay_intime + interval '24 hours'
AND

lower(medication) SIMILAR TO
'%amikacin%|%gentamicin%|%kanamycin%|%netilmicin%|%tobramycin%|%paromomycin%|%spect
inomycin%|%geldanamycin%|%ertapenem%|%doripenem%|%imipenem%|%meropenem%|%cefadroxil
%|%cefazolin%|%cefalexin%|%cefaclor%|%cefoxitin%|%cefprozil%|%cefamandole%|%cefurox
ime%|%cefixime%|%cefotaxime%|%cefpodoxime%|%ceftazidime%|%ceftriaxone%|%cefepime%|%
vancomycin%|%vanc%|%clindamycin%|%daptomycin%|%azithromycin%|%clarithromycin%|%eryt
hromycin%|%telithromycin%|%aztreonam%|%nitrofurantoin%|%linezolid%|%amoxicillin%|%a
mpicillin%|%dicloxacillin%|%flucloxacillin%|%methicillin%|%nafcillin%|%oxacillin%|%
penicillin%|%piperacillin%|%cefotetan%|%ticarcillin%|%timentin%|%colistin%|%bactrim
%|%polymyxin%|%ciprofloxacin%|%gatifloxacin%|%levofloxacin%|%moxifloxacin%|%nalidix
ic
acid%|%norfloxacin%|%ofloxacin%|%trovafloxacin%|%sulfadiazine%|%sulfamethoxazole%|%
trimethoprim%|%TMP%|%doxycycline%|%minocycline%|%tetracycline%|%dapsone%|%ethambuto
l%|%isoniazid%|%pyrazinamide%|%rifampicin%|%rifampin%|%rifabutin%|%streptomycin%|%c
hloramphenicol%|%synercid%|%fosfomycin%|%metronidazole%|%mupirocin%|%quinupristin%|
%tigecycline%|%unasyn%'

AND route IN ('IV', 'PO');
```

Patients from the above cohort started on abx within 24 hours of icustay_intime: 5223

```
SELECT DISTINCT ON (C.subject_id) C.hadm_id FROM poe_order P, cohort6 C

WHERE C.hadm_id = P.hadm_id AND start_dt <= icustay_intime + interval '24 hours'
AND

lower(medication) SIMILAR TO
'%amikacin%|%gentamicin%|%kanamycin%|%netilmicin%|%tobramycin%|%paromomycin%|%spect
inomycin%|%geldanamycin%|%ertapenem%|%doripenem%|%imipenem%|%meropenem%|%cefadroxil
%|%cefazolin%|%cefalexin%|%cefaclor%|%cefoxitin%|%cefprozil%|%cefamandole%|%cefurox
ime%|%cefixime%|%cefotaxime%|%cefpodoxime%|%ceftazidime%|%ceftriaxone%|%cefepime%|%
vancomycin%|%vanc%|%clindamycin%|%daptomycin%|%azithromycin%|%clarithromycin%|%eryt
hromycin%|%telithromycin%|%aztreonam%|%nitrofurantoin%|%linezolid%|%amoxicillin%|%a
mpicillin%|%dicloxacillin%|%flucloxacillin%|%methicillin%|%nafcillin%|%oxacillin%|%
penicillin%|%piperacillin%|%cefotetan%|%ticarcillin%|%timentin%|%colistin%|%bactrim
%|%polymyxin%|%ciprofloxacin%|%gatifloxacin%|%levofloxacin%|%moxifloxacin%|%nalidix
ic
acid%|%norfloxacin%|%ofloxacin%|%trovafloxacin%|%sulfadiazine%|%sulfamethoxazole%|%
trimethoprim%|%TMP%|%doxycycline%|%minocycline%|%tetracycline%|%dapsone%|%ethambuto
l%|%isoniazid%|%pyrazinamide%|%rifampicin%|%rifampin%|%rifabutin%|%streptomycin%|%c
hloramphenicol%|%synercid%|%fosfomycin%|%metronidazole%|%mupirocin%|%quinupristin%|
%tigecycline%|%unasyn%'

AND route IN ('IV', 'PO');
```

Exclusion of duplicate hospitalizations through use of subject_id. 4862 hospitalizations.

Appendix B

Code for Natural Language Processing

Final result from multiple iterations of testing, as described in thesis. Code written in postgreSQL:

```
SIMILAR TO
'%amikacin%|%gentamicin%|%kanamycin%|%netilmicin%|%tobramycin%|%paromomycin%|%spect
inomycin%|%geldanamycin%|%ertapenem%|%doripenem%|%imipenem%|%meropenem%|%cefadroxil
%|%cefalexin%|%cefaclor%|%cefoxitin%|%cefprozil%|%cefamandole%|%cefuroxime%|%cefixi
me%|%cefotaxime%|%cefpodoxime%|%ceftazidime%|%ceftriaxone%|%cefepime%|%vancomycin%|
%vanc%|%clindamycin%|%daptomycin%|%azithromycin%|%clarithromycin%|%erythromycin%|%t
elithromycin%|%aztreonam%|%nitrofurantoin%|%linezolid%|%amoxicillin%|%ampicillin%|%
dicloxacillin%|%flucloxacillin%|%methicillin%|%nafcillin%|%oxacillin%|%penicillin%|
%piperacillin%|%cefotetan%|%ticarcillin%|%timentin%|%colistin%|%bactrim%|%polymyxin
%|%ciprofloxacin%|%gatifloxacin%|%levofloxacin%|%moxifloxacin%|%nalidixic
acid%|%norfloxacin%|%ofloxacin%|%trovafloxacin%|%sulfadiazine%|%sulfamethoxazole%|%
trimethoprim%|%TMP%|%doxycycline%|%minocycline%|%tetracycline%|%dapsone%|%ethambuto
l%|%isoniazid%|%pyrazinamide%|%rifampicin%|%rifampin%|%rifabutin%|%streptomycin%|%c
hloramphenicol%|%synercid%|%fosfomycin%|%metronidazole%|%mupirocin%|%quinupristin%|
%tigecycline%|%unasyn%'

AND route IN ('IV', 'PO');
```

Code used for Validation for Natural Language Processing:

```
SELECT hadm_id FROM cohort14 WHERE hadm_id IN (5609,
11168,20226,28023,25333,18296,19243,24387,10387,4871,576,12731,1473,23324,8118,

21914,28979,20210,22866,22303,8352,24920,19368,33812,19555,13101,22678,12869,15975,
7711,4101,31627,34063,31925,13598,7192,27549,29077,

18331,27540) AND lower(text)

NOT SIMILAR TO
'%amikacin%|%genta%|%kanamycin%|%netilmicin%|%tobramycin%|%paromomycin%|%spectinomy
cin%|%geldanamycin%|%ertapenem%|%doripenem%|%imipenem%|%meropenem%|%merepenum%|%cef
adroxil%|%cefalexin%|%cefaclor%|%cefoxitin%|%cefprozil%|%cefamandole%|%cefuroxime%|
%cefixime%|%cefotaxime%|%cefpodoxime%|%ceftazidime%|%ceftriaxone%|%cefepime%|%cefam
pine%|% vanco %|% vanc
%|%vancomycin%|%clindamycin%|%clinda%|%daptomycin%|%azithromycin%|%clarithromycin%|
%erythromycin%|%telithromycin%|%aztreonam%|%nitrofurantoin%|%linezolid%|%amoxicilli
n%|%ampicillin%|%ampcillin%|%dicloxacillin%|%flucloxacillin%|%methicillin%|%nafcill
in%|%oxacillin%|%penicillin%|%piperacillin%|%pipercillin%|%cefotetan%|%ticarcillin%
```

```
|%timentin%|%colistin%|%bactrim%|%polymyxin%|%ciprofloxacin%|% cipro
%|%gatifloxacin%|%levofloxacin%|%levoflox%|%levoquin%|%levaquin%|%moxifloxacin%|%na
lidixic
acid%|%norfloxacin%|%ofloxacin%|%trovafloxacin%|%sulfadiazine%|%sulfamethoxazole%|%
trimethoprim%|%TMP%|%doxycycline%|%minocycline%|%tetracycline%|%dapsone%|%ethambuto
l%|%isoniazid%|%pyrazinamide%|%rifampicin%|%rifampin%|%rifabutin%|%streptomycin%|%c
hloramphenicol%|%synercid%|%fosfomycin%|%metronidazole%|%linezoid%|%mupirocin%|%qui
nupristin%|%tigecycline%|%bacteremia%|%abcess%|%ceftrioaxone%|%abscess%|%meropenium
%|% gent
%|%antibiotic%|%antibx%|%abx%|%anbx%|%cdiff%|%septic%|%sepsis%|%cellulitis%|%ancef%
|% ampi
%|%ampi,%|%ampacillin%|%rocephin%|%/gent%|%urosepsis%|%zosyn%|%infection%|%pneumoni
a%|%pnuemonia%|% uti %|%flagyl%|%falgyl%|%unasyn%';
```

Appendix C – Previous Versions of Code and Descriptions of Tables and Views

Description of Created Views

admissiontimes – association of all possible hospital admissions and icu stays for the same patient where the icu stay must fall within that hospitalization, but multiple icu stays of the same hospitalization are included

admissiontimes1 – revised

organisms – list of all possible positive culture results, by organism

patient_cohort – ids of the 603 hospitalizations in which abx not ordered in the first 24 hours but ordered in the next 24 hours

cohort_meds – all hospitalizations where an antibiotic was ordered, based on Jenna's list of abx. Contains the time of the first medication order from the poe_order table, each antibiotic, and the start and stop times of each antibiotic. Will not use this since it is Jenna's code and I have rewritten it

creatinine – initial creatinine value from the 603 pts in patient_cohort above. Contains the hospitalization_id, the charttime of the creatinine, and the creatinine value (valuenum).

missingadmits – the 910 pts in the admissions table that does not have associated icustays based on hadm_id. Mimics the format of the admissions table

cohort3 – initial selection of adult hospitalizations excluding transfers from wards and other hospitals for which an ICU stay can be linked, either automatically or manually. 17,005 hospitalizations. Fields: subject_id, hadm_id, icustay_id, admission_date, icustay_intime, discharge_date, dod

cohort8 – hospitalizations of patients who did not get abx in the first 12 hours based on both inpatient

orders + NLP

9545 rows. Just contains hadm_id.

cohort13 – from the new set of notes, all notes that were in cohort10 to make sure there is no contamination

cohort14 – the remaining 20% of notes

cohort15 - 5864 hospitalizations out of 17005 started on one of the defined abx in the first 12 hours

based on inpatient orders (34.5%).

cohort16 - 4751 hospitalizations in which antibiotics received based on NLP +. 14772 is the

denominator of NLP task (32.2%). Since 75% PPV, means that actually 24.1% of pts receive abx in first 12 hours.

cohort17 – all the unique pt/hospitalizations started on antibiotics late. 934 total

cohort18 – the new initial cohort of 9043 hospitalizations of unique patients, which includes maximum number of patients started later on antibiotics.

frequency – the most common laboratory tests ordered in the first 4 hours of each icu admission in the

training data. Does not have labels but has itemid.

exportsas – initial code for all of the vital signs and labs for all the patients in the training data, as well

as the outcome.

outcome – the outcome in the training group, 1 indicates patient started on antibiotics late.

<u>User-Defined Tables</u>

antibiotic – all patients on an antibiotic I am interested in, po or iv, during any hospitalization, and including all patients in the MIMIC-2 database

cohort10 – a random sample of 80% of the 15968 hospitalizations from firstnote (so 12774 results).

contains hadm_id, subject_id, admission_date, icustay_intime

cohort11 – hospitalizations from previous where pts started on one of the defined abx within 24 hours.

Multiple rows representing 4943 hospitalizations. Only contains hadm_ids, no other fields. Changed to a table for processing speed reasons, however the processing time is actually due to

the regular expressions, not due to the fact that cohort11 was a view

cohort12- the 20% of first notes for the test set for the NLP

Firstnote – the first note of each hospitalization

5 minutes 47 seconds to create it

15568 is the total number in firstnote. training set is 82%

firstnote1 – corrected version. Includes the first note written by an RN of each of the hospitalizations

from cohort3, that note must be written within 24 hours of icustay_intime. Results in 14772 notes. Without the 24 hour window, there are 15188 notes (not in this table).

cohort9 – same info as cohort8 the view, except has all of the basic fields. 9545 rows. 9043 distinct pts,

meaning that 5% of the hospitalizations were of duplicates

cohort19 – 80% of the pts not started on abx in the first 12 hours. 7234

cohort20 – 20% of the pts not started on abx in the first 12 hours. 1809

Confirmed that ratio of outcomes is similar in cohort19 and cohort20.

all_lab_results – all lab results from all patients in the training cohort, that occurred 4 hours before or

after icustay_intime, have mapped test results to the name of the test.

labs – first value of the variables I am interested in for all labs of all pts in the training data, organized for use in SAS. One value per pt on admission.

Code to Create Views

```
CREATE VIEW creatinine AS SELECT DISTINCT ON (P.hadm_id) P.hadm_id, charttime,
valuenum FROM patient_cohort P, labevents L WHERE P.hadm_id = L.hadm_id

AND itemid = 50090 ORDER BY P.hadm_id, charttime;

      SELECT * FROM creatinine WHERE valuenum < 1.2;


CREATE VIEW admissiontimes AS

SELECT A.hadm_id,

hospital_admit_dt,

icustay_intime,

icustay_intime - hospital_admit_dt AS difference

FROM admissions A, icustay_detail I

WHERE A.hadm_id = I.hadm_id;
```

grabbing multiple icu stays per hospitalization but will be fixed during the next step

```
CREATE VIEW admissiontimes1 AS

SELECT A.hadm_id, A.hospital_admit_dt, A.icustay_intime, icustay_age_group FROM
admissiontimes A, icustay_detail I, demographic_detail D

WHERE difference <'30:00:00' AND A.icustay_intime = I.icustay_intime AND
icustay_age_group = 'adult' AND A.hadm_id = D.hadm_id AND D.admission_source_itemid
!= 200074;



CREATE VIEW starttime AS

SELECT A.hadm_id, MIN(enter_dt) AS first_medorder FROM poe_order P, admissiontimes1
A

WHERE P.hadm_id = A.hadm_id

GROUP BY A.hadm_id

ORDER BY A.hadm_id;



CREATE VIEW organisms AS

SELECT * FROM d_codeditems WHERE category = 'ORGANISM';



CREATE VIEW cohort4 AS

SELECT DISTINCT ON (C.hadm_id) C.hadm_id, C.subject_id, admission_date,
icustay_intime, charttime, valuenum FROM cohort3 C, labevents L

WHERE C.hadm_id = L.hadm_id AND itemid = 50090 ORDER BY C.hadm_id, charttime;



CREATE VIEW cohort5 AS

SELECT * FROM cohort3 WHERE (dod IS NULL OR NOT dod <= icustay_intime + interval '5
days')

AND discharge_date >= icustay_intime + interval '5 days';



CREATE VIEW cohort6 AS

SELECT C.hadm_id, C.subject_id, admission_date, icustay_intime, start_dt, stop_dt,
medication FROM cohort5 C, antibiotic A

WHERE C.hadm_id = A.hadm_id AND start_dt <= icustay_intime + interval '24 hours';



CREATE VIEW cohort8 AS
```

```
(SELECT hadm_id FROM cohort3

EXCEPT

SELECT * FROM cohort15)

EXCEPT

SELECT * FROM cohort16


CREATE VIEW cohort13 AS

SELECT * FROM firstnote1 WHERE hadm_id IN (SELECT hadm_id FROM cohort10);


CREATE VIEW cohort14 AS

SELECT * FROM firstnote1 WHERE hadm_id NOT IN (SELECT hadm_id FROM cohort10);


CREATE VIEW cohort15 AS

SELECT DISTINCT C.hadm_id FROM cohort3 C, antibiotic A

WHERE C.hadm_id = A.hadm_id AND start_dt <= icustay_intime + interval '12 hours'
ORDER BY hadm_id;


CREATE VIEW cohort16 AS

SELECT hadm_id FROM firstnote1 WHERE lower(text)

SIMILAR TO
'%amikacin%|%genta%|%kanamycin%|%netilmicin%|%tobramycin%|%paromomycin%|%spectinomy
cin%|%geldanamycin%|%ertapenem%|%doripenem%|%imipenem%|%meropenem%|%merepenum%|%cef
adroxil%|%cefalexin%|%cefaclor%|%cefoxitin%|%cefprozil%|%cefamandole%|%cefuroxime%|
%cefixime%|%cefotaxime%|%cefpodoxime%|%ceftazidime%|%ceftriaxone%|%cefepime%|%cefam
pine%|% vanco %|% vanc
%|%vancomycin%|%clindamycin%|%clinda%|%daptomycin%|%azithromycin%|%clarithromycin%|
%erythromycin%|%telithromycin%|%aztreonam%|%nitrofurantoin%|%linezolid%|%amoxicilli
n%|%ampicillin%|%ampcillin%|%dicloxacillin%|%flucloxacillin%|%methicillin%|%nafcill
in%|%oxacillin%|%penicillin%|%piperacillin%|%pipercillin%|%cefotetan%|%ticarcillin%
|%timentin%|%colistin%|%bactrim%|%polymyxin%|%ciprofloxacin%|% cipro
%|%gatifloxacin%|%levofloxacin%|%levoflox%|%levoquin%|%levaquin%|%moxifloxacin%|%na
lidixic
acid%|%norfloxacin%|%ofloxacin%|%trovafloxacin%|%sulfadiazine%|%sulfamethoxazole%|%
trimethoprim%|%TMP%|%doxycycline%|%minocycline%|%tetracycline%|%dapsone%|%ethambuto
l%|%isoniazid%|%pyrazinamide%|%rifampicin%|%rifampin%|%rifabutin%|%streptomycin%|%c
hloramphenicol%|%synercid%|%fosfomycin%|%metronidazole%|%linezoid%|%mupirocin%|%qui
nupristin%|%tigecycline%|%bacteremia%|%abcess%|%ceftrioaxone%|%abscess%|%meropenium
%|% gent
%|%antibiotic%|%antibx%|%abx%|%anbx%|%cdiff%|%septic%|%sepsis%|%cellulitis%|%ancef%
|% ampi
%|%ampi,%|%ampacillin%|%rocephin%|%/gent%|%urosepsis%|%zosyn%|%infection%|%pneumoni
a%|%pnuemonia%|% uti %|%flagyl%|%falgyl%|%unasyn%';
```

```
CREATE VIEW cohort17 AS

SELECT DISTINCT ON (C.subject_id) C.subject_id, C.hadm_id, C.icustay_id,
admission_date, icustay_intime, discharge_date, dod

FROM cohort9 C, antibiotic A WHERE C.hadm_id = A.hadm_id AND start_dt <=
icustay_intime + interval '48 hours';



CREATE VIEW cohort18 AS

SELECT DISTINCT ON (C.subject_id) C.subject_id, C.hadm_id, C.icustay_id,
admission_date, icustay_intime, discharge_date, dod

FROM cohort9 C, antibiotic A WHERE C.hadm_id = A.hadm_id AND start_dt <=
icustay_intime + interval '48 hours'

UNION

SELECT DISTINCT ON (subject_id) * FROM cohort9 WHERE subject_id NOT IN (SELECT
subject_id FROM cohort17);



CREATE VIEW frequency AS

SELECT itemid, COUNT(itemid) AS count FROM all_lab_results GROUP BY itemid ORDER BY
COUNT(itemid) DESC;




CREATE VIEW exportsas AS

SELECT cohort19.*,

max_temp.value1num AS tmax,

rr_max.value1num AS max_resp,

rr_mean.average AS mean_rr,

hr_max.hr AS max_hr,

hr_min.hr AS min_hr,

hr_mean.hr AS mean_hr,

minbpsys, minbpdias, avgsysbp, avgdiasbp,

minsat, avgsat, maxsat, temp_abg,

polychromasia,

platelet_manual,

lipase,
```

bicarb_abg2,

poikilocytosis,

albumin,

bands,

anisocytosis,

microcytes,

macrocytes,

hypochromia,

aagrad_abg,

reqo2_abg,

methadone_urine,

cocaine_urine,

opiate_urine,

amphet_urine,

barb_urine,

benzo_urine,

gfr,

alk_phos,

bili_total,

ast,

alt,

barb_serum,

tca_serum,

benzo_serum,

ethanol,

asa,

acetaminophen,

tv_abg,

o2satinsp_abg,

fibrinogen,

urinesed_epi,

urinesed_bact,

urinesed_leuko,

urinesed_yeast,

urinesed_eryth,

o2sat_abg,

amylase,

troponin,

urine_leukocytes,

urine_urobilinogen,

urine_glucose,

urine_ketone,

urine_grav,

urine_nitrite,

urine_color,

urine_protein,

urine_ph,

urine_hemoglobin,

urine_bili,

urine_app,

ckmb,

eosinophils,

lymphocytes,

monocytes,

basophils,

neutrophils,

cpk,

phosphate,

calcium,

magnesium,

lactate,

ag,

sodium_abg,

glucose,

sodium,

potassium,

ionca_abg,

chloride,

bicarb,

ptt,

pt,

inr,

rdw,

bun,

creatinine,

erythrocytes,

mcv,

mch,

mchc,

leukocytes,

hemoglobin,

hemoglobin_abg,

hematocrit_abg,

platelet,

hematocrit,

glucose_abg,

potassium_abg,

bicarb_abg1,

pco2_abg,

baseexcess_abg,

po2,

ph,

outcome

```
FROM cohort19

LEFT JOIN max_temp ON cohort19.subject_id = max_temp.subject_id

LEFT JOIN rr_max ON cohort19.subject_id = rr_max.subject_id

LEFT JOIN rr_mean ON cohort19.subject_id = rr_mean.subject_id

LEFT JOIN hr_max ON cohort19.subject_id = hr_max.subject_id

LEFT JOIN hr_min ON cohort19.subject_id = hr_min.subject_id

LEFT JOIN hr_mean ON cohort19.subject_id = hr_mean.subject_id

LEFT JOIN bp ON cohort19.subject_id = bp.subject_id

LEFT JOIN saturation ON cohort19.subject_id = saturation.subject_id

LEFT JOIN labs ON cohort19.subject_id = labs.subject_id

LEFT JOIN outcome ON cohort19.subject_id = outcome.subject_id

ORDER BY cohort19.subject_id;


CREATE VIEW bands AS

SELECT subject_id, to_number(bands, '99.9') FROM exportsas;


CREATE VIEW outcome AS

SELECT subject_id, MAX(CASE WHEN subject_id IN (SELECT subject_id FROM cohort17)
THEN 1 ELSE 0 END) AS outcome FROM cohort19

GROUP BY subject_id ORDER BY subject_id;
```

Creation of Tables

```
CREATE TABLE antibiotic AS

SELECT * FROM poe_order WHERE lower(medication) SIMILAR TO
'%amikacin%|%gentamicin%|%kanamycin%|%netilmicin%|%tobramycin%|%paromomycin%|%spect
inomycin%|%geldanamycin%|%ertapenem%|%doripenem%|%imipenem%|%meropenem%|%cefadroxil
%|%cefalexin%|%cefaclor%|%cefoxitin%|%cefprozil%|%cefamandole%|%cefuroxime%|%cefixi
me%|%cefotaxime%|%cefpodoxime%|%ceftazidime%|%ceftriaxone%|%cefepime%|%vancomycin%|
%vanc%|%clindamycin%|%daptomycin%|%azithromycin%|%clarithromycin%|%erythromycin%|%t
elithromycin%|%aztreonam%|%nitrofurantoin%|%linezolid%|%amoxicillin%|%ampicillin%|%
dicloxacillin%|%flucloxacillin%|%methicillin%|%nafcillin%|%oxacillin%|%penicillin%|
%piperacillin%|%cefotetan%|%ticarcillin%|%timentin%|%colistin%|%bactrim%|%polymyxin
%|%ciprofloxacin%|%gatifloxacin%|%levofloxacin%|%moxifloxacin%|%nalidixic
acid%|%norfloxacin%|%ofloxacin%|%trovafloxacin%|%sulfadiazine%|%sulfamethoxazole%|%
trimethoprim%|%TMP%|%doxycycline%|%minocycline%|%tetracycline%|%dapsone%|%ethambuto
l%|%isoniazid%|%pyrazinamide%|%rifampicin%|%rifampin%|%rifabutin%|%streptomycin%|%c
hloramphenicol%|%synercid%|%fosfomycin%|%metronidazole%|%mupirocin%|%quinupristin%|
```

```
%tigecycline%|%unasyn%'

AND route IN ('IV', 'PO');
```

94832 results

```
CREATE TABLE cohort10 AS

SELECT F.hadm_id, subject_id, admission_date, C.icustay_intime FROM firstnote F,
cohort3 C WHERE F.hadm_id = C.hadm_id

ORDER BY random() LIMIT 12774;
```

random sample of 80%

```
CREATE TABLE cohort11 AS

SELECT DISTINCT C.hadm_id FROM cohort10 C, antibiotic A

WHERE C.hadm_id = A.hadm_id AND start_dt <= icustay_intime + interval '24 hours';


CREATE TABLE firstnote AS

SELECT DISTINCT ON (hadm_id) hadm_id, icustay_intime, charttime, realtime,
category, title, text FROM notes ORDER BY hadm_id, charttime;


CREATE TABLE firstnote1 AS

SELECT * FROM

(SELECT DISTINCT ON (C.hadm_id) C.hadm_id, icustay_intime, charttime, realtime,
category, title, text FROM noteevents N, cohort3 C WHERE C.hadm_id = N.hadm_id

AND cgid IN (SELECT cgid FROM d_caregivers WHERE label = 'RN') ORDER BY hadm_id,
charttime) AS foo

WHERE realtime < icustay_intime + interval '24 hours';


CREATE TABLE cohort12 AS

SELECT * FROM firstnote WHERE hadm_id NOT IN (SELECT hadm_id FROM cohort10);


CREATE TABLE cohort9 AS

SELECT * FROM cohort3 WHERE hadm_id IN (SELECT * FROM cohort8);


CREATE TABLE cohort19 AS
```

```
SELECT * FROM cohort18 ORDER BY random() LIMIT 7234;


CREATE TABLE cohort20 AS

SELECT * FROM cohort18 WHERE subject_id NOT IN (SELECT subject_id FROM cohort19);


CREATE TABLE all_lab_results AS

SELECT C.subject_id, C.hadm_id, L.itemid, icustay_intime, charttime, value,
valuenum, test_name, fluid, loinc_description

FROM labevents L, d_labitems D, cohort19 C WHERE L.itemid = D.itemid AND
C.subject_id = L.subject_id

AND charttime <= icustay_intime + interval '4 hours' AND charttime >=
icustay_intime - interval '4 hours';


CREATE TABLE labs AS

SELECT subject_id,

         MAX(CASE WHEN itemid = 50018 THEN value ELSE NULL END) AS "ph",

         MAX(CASE WHEN itemid = 50019 THEN value ELSE NULL END) AS "po2",

         MAX(CASE WHEN itemid = 50002 THEN value ELSE NULL END) AS
"baseexcess_abg",

      MAX(CASE WHEN itemid = 50016 THEN value ELSE NULL END) AS "pco2_abg",

      MAX(CASE WHEN itemid = 50025 THEN value ELSE NULL END) AS "bicarb_abg1",

MAX(CASE WHEN itemid = 50009 THEN value ELSE NULL END) AS "potassium_abg",

MAX(CASE WHEN itemid = 50006 THEN value ELSE NULL END) AS "glucose_abg",

MAX(CASE WHEN itemid = 50383 THEN value ELSE NULL END) AS "hematocrit",

      MAX(CASE WHEN itemid = 50428 THEN value ELSE NULL END) AS "platelet",

      MAX(CASE WHEN itemid = 50029 THEN value ELSE NULL END) AS "hematocrit_abg",

MAX(CASE WHEN itemid = 50007 THEN value ELSE NULL END) AS "hemoglobin_abg",

MAX(CASE WHEN itemid = 50386 THEN value ELSE NULL END) AS "hemoglobin",

MAX(CASE WHEN itemid = 50468 THEN value ELSE NULL END) AS "leukocytes",

MAX(CASE WHEN itemid = 50412 THEN value ELSE NULL END) AS "mchc",

      MAX(CASE WHEN itemid = 50411 THEN value ELSE NULL END) AS "mch",

MAX(CASE WHEN itemid = 50413 THEN value ELSE NULL END) AS "mcv",

MAX(CASE WHEN itemid = 50442 THEN value ELSE NULL END) AS "erythrocytes",
```

```
MAX(CASE WHEN itemid = 50090 THEN value ELSE NULL END) AS "creatinine",
      MAX(CASE WHEN itemid = 50177 THEN value ELSE NULL END) AS "bun",
      MAX(CASE WHEN itemid = 50444 THEN value ELSE NULL END) AS "rdw",
MAX(CASE WHEN itemid = 50399 THEN value ELSE NULL END) AS "inr",
MAX(CASE WHEN itemid = 50439 THEN value ELSE NULL END) AS "pt",
MAX(CASE WHEN itemid = 50440 THEN value ELSE NULL END) AS "ptt",
      MAX(CASE WHEN itemid = 50172 THEN value ELSE NULL END) AS "bicarb",
      MAX(CASE WHEN itemid = 50083 THEN value ELSE NULL END) AS "chloride",
MAX(CASE WHEN itemid = 50030 THEN value ELSE NULL END) AS "ionca_abg",
MAX(CASE WHEN itemid = 50149 THEN value ELSE NULL END) AS "potassium",
MAX(CASE WHEN itemid = 50159 THEN value ELSE NULL END) AS "sodium",
      MAX(CASE WHEN itemid = 50112 THEN value ELSE NULL END) AS "glucose",
      MAX(CASE WHEN itemid = 50012 THEN value ELSE NULL END) AS "sodium_abg",
MAX(CASE WHEN itemid = 50068 THEN value ELSE NULL END) AS "ag",
MAX(CASE WHEN itemid = 50010 THEN value ELSE NULL END) AS "lactate",
MAX(CASE WHEN itemid = 50140 THEN value ELSE NULL END) AS "magnesium",
MAX(CASE WHEN itemid = 50079 THEN value ELSE NULL END) AS "calcium",
      MAX(CASE WHEN itemid = 50148 THEN value ELSE NULL END) AS "phosphate",
MAX(CASE WHEN itemid = 50086 THEN value ELSE NULL END) AS "cpk",
MAX(CASE WHEN itemid = 50149 THEN value ELSE NULL END) AS "neutrophils",
MAX(CASE WHEN itemid = 50333 THEN value ELSE NULL END) AS "basophils",
      MAX(CASE WHEN itemid = 50417 THEN value ELSE NULL END) AS "monocytes",
      MAX(CASE WHEN itemid = 50408 THEN value ELSE NULL END) AS "lymphocytes",
MAX(CASE WHEN itemid = 50373 THEN value ELSE NULL END) AS "eosinophils",
MAX(CASE WHEN itemid = 50087 THEN value ELSE NULL END) AS "ckmb",
MAX(CASE WHEN itemid = 50623 THEN value ELSE NULL END) AS "urine_app",
MAX(CASE WHEN itemid = 50626 THEN value ELSE NULL END) AS "urine_bili",
MAX(CASE WHEN itemid = 50627 THEN value ELSE NULL END) AS "urine_hemoglobin",
MAX(CASE WHEN itemid = 50653 THEN value ELSE NULL END) AS "urine_ph",
MAX(CASE WHEN itemid = 50655 THEN value ELSE NULL END) AS "urine_protein",
MAX(CASE WHEN itemid = 50633 THEN value ELSE NULL END) AS "urine_color",
```

```
MAX(CASE WHEN itemid = 50650 THEN value ELSE NULL END) AS "urine_nitrite",
MAX(CASE WHEN itemid = 50661 THEN value ELSE NULL END) AS "urine_grav",
MAX(CASE WHEN itemid = 50647 THEN value ELSE NULL END) AS "urine_ketone",
MAX(CASE WHEN itemid = 50641 THEN value ELSE NULL END) AS "urine_glucose",
MAX(CASE WHEN itemid = 50671 THEN value ELSE NULL END) AS "urine_urobilinogen",
MAX(CASE WHEN itemid = 50648 THEN value ELSE NULL END) AS "urine_leukocytes",
MAX(CASE WHEN itemid = 50189 THEN value ELSE NULL END) AS "troponin",
MAX(CASE WHEN itemid = 50065 THEN value ELSE NULL END) AS "amylase",
MAX(CASE WHEN itemid = 50015 THEN value ELSE NULL END) AS "o2sat_abg",
MAX(CASE WHEN itemid = 50656 THEN value ELSE NULL END) AS "urinesed_eryth",
MAX(CASE WHEN itemid = 50677 THEN value ELSE NULL END) AS "urinesed_yeast",
MAX(CASE WHEN itemid = 50674 THEN value ELSE NULL END) AS "urinesed_leuko",
MAX(CASE WHEN itemid = 50624 THEN value ELSE NULL END) AS "urinesed_bact",
MAX(CASE WHEN itemid = 50637 THEN value ELSE NULL END) AS "urinesed_epi",
MAX(CASE WHEN itemid = 50378 THEN value ELSE NULL END) AS "fibrinogen",
MAX(CASE WHEN itemid = 50013 THEN value ELSE NULL END) AS "o2satinsp_abg",
MAX(CASE WHEN itemid = 50024 THEN value ELSE NULL END) AS "tv_abg",
MAX(CASE WHEN itemid = 50056 THEN value ELSE NULL END) AS "acetaminophen",
MAX(CASE WHEN itemid = 50072 THEN value ELSE NULL END) AS "asa",
MAX(CASE WHEN itemid = 50099 THEN value ELSE NULL END) AS "ethanol",
MAX(CASE WHEN itemid = 50187 THEN value ELSE NULL END) AS "benzo_serum",
MAX(CASE WHEN itemid = 50198 THEN value ELSE NULL END) AS "tca_serum",
MAX(CASE WHEN itemid = 50186 THEN value ELSE NULL END) AS "barb_serum",
MAX(CASE WHEN itemid = 50062 THEN value ELSE NULL END) AS "alt",
MAX(CASE WHEN itemid = 50073 THEN value ELSE NULL END) AS "ast",
MAX(CASE WHEN itemid = 50170 THEN value ELSE NULL END) AS "bili_total",
MAX(CASE WHEN itemid = 50061 THEN value ELSE NULL END) AS "alk_phos",
MAX(CASE WHEN itemid = 50193 THEN value ELSE NULL END) AS "gfr",
MAX(CASE WHEN itemid = 50291 THEN value ELSE NULL END) AS "benzo_urine",
MAX(CASE WHEN itemid = 50290 THEN value ELSE NULL END) AS "barb_urine",
MAX(CASE WHEN itemid = 50289 THEN value ELSE NULL END) AS "amphet_urine",
```

```
MAX(CASE WHEN itemid = 50296 THEN value ELSE NULL END) AS "opiate_urine",
MAX(CASE WHEN itemid = 50292 THEN value ELSE NULL END) AS "cocaine_urine",
MAX(CASE WHEN itemid = 50295 THEN value ELSE NULL END) AS "methadone_urine",
MAX(CASE WHEN itemid = 50021 THEN value ELSE NULL END) AS "reqo2_abg",
MAX(CASE WHEN itemid = 50001 THEN value ELSE NULL END) AS "aagrad_abg",
MAX(CASE WHEN itemid = 50396 THEN value ELSE NULL END) AS "hypochromia",
MAX(CASE WHEN itemid = 50490 THEN value ELSE NULL END) AS "macrocytes",
MAX(CASE WHEN itemid = 50415 THEN value ELSE NULL END) AS "microcytes",
MAX(CASE WHEN itemid = 50326 THEN value ELSE NULL END) AS "anisocytosis",
MAX(CASE WHEN itemid = 50332 THEN value ELSE NULL END) AS "bands",
MAX(CASE WHEN itemid = 50060 THEN value ELSE NULL END) AS "albumin",
MAX(CASE WHEN itemid = 50431 THEN value ELSE NULL END) AS "poikilocytosis",
MAX(CASE WHEN itemid = 50022 THEN value ELSE NULL END) AS "bicarb_abg2",
MAX(CASE WHEN itemid = 50138 THEN value ELSE NULL END) AS "lipase",
MAX(CASE WHEN itemid = 50429 THEN value ELSE NULL END) AS "platelet_manual",
MAX(CASE WHEN itemid = 50432 THEN value ELSE NULL END) AS "polychromasia",
MAX(CASE WHEN itemid = 50023 THEN value ELSE NULL END) AS "temp_abg"
    FROM all_lab_results GROUP BY subject_id;
```

Appendix D – General SQL Code:

```
SELECT * FROM icustay_days;
SELECT * FROM deliveries;
SELECT * FROM d_careunits;
SELECT * FROM demographic_detail;
SELECT * FROM icustay_detail;

SELECT admission_source_descr FROM demographic_detail
GROUP BY admission_source_descr;
      types of admission_sources
```

```
SELECT label, COUNT(label) FROM d_caregivers
GROUP BY label
ORDER BY COUNT(label) DESC;
```

all the types of caregivers, including duplicates

```
SELECT * FROM icustay_detail WHERE hadm_id = 34404;
```

a pt with long hospital stay admitted from the ED who was in the MICU only 1 day and in the middle. This shows that admission source refers to the hospitalization, not the ICU stay

```
SELECT * FROM noteevents WHERE subject_id = 32785 and hadm_id = 34404;
```

```
SELECT * FROM icustay_detail
WHERE hospital_total_num > 1
ORDER BY subject_id, icustay_intime;
```

to clarify the relationship between icu_stay and hospital_admission. Shows that ICU flag first day refers to the first ICU stay of each hospitalization.

```
SELECT * FROM noteevents
LIMIT 100;
```

to limit the number of returning items

```
SELECT subject_id, hadm_id, admission_source_descr FROM demographic_detail
WHERE admission_source_itemid = 200029;
```

number of hospital admissions that came in through the ER based on hadm_id from the demographic detail file (16018)

```
SELECT D.hadm_id, hospital_admit_dt, icustay_intime
FROM icustay_detail I, demographic_detail D
WHERE D.admission_source_itemid = 200029 AND I.hadm_id = D.hadm_id;
```

number of hospital admits that came in through the ER when using hadm_id from the admissions file (16618)

```
SELECT A.hadm_id, admit_dt, hospital_admit_dt

FROM admissions A, icustay_detail I

WHERE A.hadm_id = I.hadm_id AND admit_dt != hospital_admit_dt;
```

checking if the admit dates are the same in the admissions and icustay_detail tables since they have different names, but they are the same

```
SELECT subject_id, COUNT(hadm_id)

FROM admissions

GROUP BY subject_id

HAVING COUNT(hadm_id)>1;
```

subject_ids that have more than one hospital admission (2717 in this query)

```
SELECT text FROM noteevents WHERE subject_id = 32785 and hadm_id = 34404 and
category like 'DISCHARGE_SUMMARY';
```

to get discharge summary

```
SELECT COUNT(DISTINCT subject_id) FROM demographic_detail;

SELECT * FROM demographic_detail;
```

number of patients with at least 1 hospitalization, based on demographic_detail (36094)

```
SELECT medication FROM poe_order

GROUP BY medication

ORDER BY medication;
```

distinct medications ordered (will need to use something like lower() to deal with caps and

lots of different names used for the same meds

```
SELECT * FROM poe_order

LIMIT 10;

GROUP BY medication

ORDER BY medication;
```

sample of poe_orders

To determine if hospitalization date is when the pt is admitted to the hospital (after ED stay) vs. when pt first enters:

```
SELECT P.hadm_id, A.hospital_admit_dt, enter_dt, medication, start_dt, stop_dt FROM
poe_order P, admissiontimes1 A

WHERE P.hadm_id = A.hadm_id and enter_dt < A.hospital_admit_dt;
```

5 results, 3 are clearly errors. Therefore, no orders are entered before the hospitalization date. hadm_ids 32975 and 35963 were likely added to the system late.

```
SELECT A.hadm_id, MIN(enter_dt) AS first_medorder FROM poe_order P, admissiontimes1
A

WHERE P.hadm_id = A.hadm_id

GROUP BY A.hadm_id

ORDER BY A.hadm_id;
```

Code to select 1 hospitalization per patient in cohort:

```
SELECT DISTINCT ON (C.subject_id) C.subject_id, C.hadm_id, C.icustay_id,
admission_date, icustay_intime, discharge_date, dod

FROM cohort9 C, antibiotic A WHERE C.hadm_id = A.hadm_id AND start_dt <=
icustay_intime + interval '48 hours'
```

– 934/9043 = 10.3% rate of outcome

Laboratory:

Most common tests ordered in the training set:

```
SELECT F.itemid, count, test_name, fluid, category, loinc_description FROM
frequency F, d_labitems D WHERE F.itemid = D.itemid;
```

```
50018;10941;"PH";"BLOOD";"BLOOD GAS";"pH of Blood"

50019;10557;"PO2";"BLOOD";"BLOOD GAS";"Oxygen [Partial pressure] in Blood"
```

50002;10556;"BASE XS";"BLOOD";"BLOOD GAS";"Base excess in Blood"

50016;10556;"PCO2";"BLOOD";"BLOOD GAS";"Carbon dioxide [Partial pressure] in Blood"

50025;10556;"TOTAL CO2";"BLOOD";"BLOOD GAS";"Bicarbonate [Moles/volume] in Blood"

50026;10239;"TYPE";"BLOOD";"BLOOD GAS";""

50009;9057;"K+";"BLOOD";"BLOOD GAS";"Potassium [Moles/volume] in Blood"

50006;8879;"GLUCOSE";"BLOOD";"BLOOD GAS";"Glucose [Mass/volume] in Blood"

50383;7630;"HCT";"BLOOD";"HEMATOLOGY";"Hematocrit [Volume Fraction] of Blood"

50428;7241;"PLT COUNT";"BLOOD";"HEMATOLOGY";"Platelets [#/volume] in Blood"

50029;6835;"calcHCT";"BLOOD";"BLOOD GAS";"Hematocrit [Volume Fraction] of Blood"

50007;6834;"HGB";"BLOOD";"BLOOD GAS";"Hemoglobin [Mass/volume] in Blood"

50386;6754;"HGB";"BLOOD";"HEMATOLOGY";"Hemoglobin [Mass/volume] in Blood"

50468;6734;"WBC";"BLOOD";"HEMATOLOGY";"Leukocytes [#/volume] in Blood"

50412;6698;"MCHC";"BLOOD";"HEMATOLOGY";"Erythrocyte mean corpuscular hemoglobin concentration [Mass/volume]"

50411;6696;"MCH";"BLOOD";"HEMATOLOGY";"Erythrocyte mean corpuscular hemoglobin [Entitic mass]"

50413;6696;"MCV";"BLOOD";"HEMATOLOGY";"Erythrocyte mean corpuscular volume [Entitic volume]"

50442;6696;"RBC";"BLOOD";"HEMATOLOGY";"Erythrocytes [#/volume] in Blood"

50090;6694;"CREAT";"BLOOD";"CHEMISTRY";"Creatinine [Mass/volume] in Serum or Plasma"

50177;6679;"UREA N";"BLOOD";"CHEMISTRY";"Urea nitrogen [Mass/volume] in Serum or Plasma"

50444;6670;"RDW";"BLOOD";"HEMATOLOGY";"Erythrocyte distribution width [Ratio]"

50399;6362;"INR(PT)";"BLOOD";"HEMATOLOGY";"INR in Blood by Coagulation assay"

50439;6340;"PT";"BLOOD";"HEMATOLOGY";"Prothrombin time (PT) in Blood by Coagulation assay"

50440;6305;"PTT";"BLOOD";"HEMATOLOGY";"Activated partial thrombplastin time (aPTT) in Blood by Coagulation assay"

50172;5779;"TOTAL CO2";"BLOOD";"CHEMISTRY";"Bicarbonate [Moles/volume] in Serum"

50083;5775;"CHLORIDE";"BLOOD";"CHEMISTRY";"Chloride [Moles/volume] in Blood"

50030;5661;"freeCa";"BLOOD";"BLOOD GAS";"Calcium.ionized [Moles/volume] in Blood"

50149;5526;"POTASSIUM";"BLOOD";"CHEMISTRY";"Potassium [Moles/volume] in Serum or Plasma"

50159;5372;"SODIUM";"BLOOD";"CHEMISTRY";"Sodium [Moles/volume] in Serum or Plasma"

50112;5309;"GLUCOSE";"BLOOD";"CHEMISTRY";"Glucose [Mass/volume] in Serum or Plasma"

50012;5306;"NA+";"BLOOD";"BLOOD GAS";"Sodium [Moles/volume] in Blood"

50068;5266;"ANION GAP";"BLOOD";"CHEMISTRY";"Anion gap in Blood"

50008;4723;"INTUBATED";"BLOOD";"BLOOD GAS";""

50010;3928;"LACTATE";"BLOOD";"BLOOD GAS";"Lactate [Moles/volume] in Blood"

50140;3476;"MAGNESIUM";"BLOOD";"CHEMISTRY";"Magnesium [Mass/volume] in Serum or Plasma"

50079;3114;"CALCIUM";"BLOOD";"CHEMISTRY";"Calcium [Mass/volume] in Serum or Plasma"

50148;3057;"PHOSPHATE";"BLOOD";"CHEMISTRY";"Phosphate [Mass/volume] in Serum or Plasma"

50027;2904;"VENT";"BLOOD";"BLOOD GAS";""

50086;2817;"CK(CPK)";"BLOOD";"CHEMISTRY";"Creatine kinase [Enzymatic activity/volume] in Serum or Plasma"

50419;2799;"NEUTS";"BLOOD";"HEMATOLOGY";"Neutrophils.segmented/100 leukocytes in Blood"

50333;2799;"BASOS";"BLOOD";"HEMATOLOGY";"Basophils/100 leukocytes in Blood"

50417;2799;"MONOS";"BLOOD";"HEMATOLOGY";"Monocytes/100 leukocytes in Blood"

50408;2799;"LYMPHS";"BLOOD";"HEMATOLOGY";"Lymphocytes/100 leukocytes in Blood"

50373;2799;"EOS";"BLOOD";"HEMATOLOGY";"Eosinophils/100 leukocytes in Blood"

50004;2742;"CL-";"BLOOD";"BLOOD GAS";"Chloride [Moles/volume] in Blood"

50087;2715;"CK-MB";"BLOOD";"CHEMISTRY";"Creatine kinase.MB [Mass/volume] in Blood"

50623;2412;"APPEAR";"URINE";"HEMATOLOGY";"Appearance of Urine"

50626;2412;"BILIRUBIN";"URINE";"HEMATOLOGY";"Bilirubin [Presence] in Urine"

50627;2412;"BLOOD";"URINE";"HEMATOLOGY";"Hemoglobin [Presence] in Urine by Test strip"

50653;2412;"PH";"URINE";"HEMATOLOGY";"pH of Urine"

50655;2412;"PROTEIN";"URINE";"HEMATOLOGY";"Protein [Mass/volume] in Urine by Test strip"

50633;2412;"COLOR";"URINE";"HEMATOLOGY";"Color of Urine"

50650;2412;"NITRITE";"URINE";"HEMATOLOGY";"Nitrite [Presence] in Urine by Test strip"

50661;2412;"SP GRAV";"URINE";"HEMATOLOGY";"Specific gravity of Urine by Test strip"

50647;2412;"KETONE";"URINE";"HEMATOLOGY";"Ketones [Mass/volume] in Urine"

50641;2412;"GLUCOSE";"URINE";"HEMATOLOGY";"Glucose [Mass/volume] in Urine"

50671;2412;"UROBILNGN";"URINE";"HEMATOLOGY";"Urobilinogen [Mass/volume] in Urine"

50268;2324;"HOURS";"URINE";"CHEMISTRY";"Collection duration of Urine"

50648;2318;"LEUK";"URINE";"HEMATOLOGY";"Leukocytes [Presence] in Urine"

50189;1946;"cTropnT";"BLOOD";"CHEMISTRY";"Troponin T.cardiac [Mass/volume] in Blood"

50065;1843;"AMYLASE";"BLOOD";"CHEMISTRY";"Amylase [Enzymatic activity/volume] in Serum or Plasma"

50113;1828;"GREEN HLD";"BLOOD";"CHEMISTRY";""

50015;1781;"O2 SAT";"BLOOD";"BLOOD GAS";"Oxygen saturation in Blood"

50656;1775;"RBC";"URINE";"HEMATOLOGY";"Erythrocytes [#/area] in Urine sediment by Microscopy high power field"

50677;1775;"YEAST";"URINE";"HEMATOLOGY";"Yeast [Presence] in Urine sediment by Light microscopy"

50674;1775;"WBC";"URINE";"HEMATOLOGY";"Leukocytes [#/area] in Urine sediment by Microscopy high power field"

50624;1775;"BACTERIA";"URINE";"HEMATOLOGY";"Bacteria [Presence] in Urine sediment by Light microscopy"

50637;1775;"EPI";"URINE";"HEMATOLOGY";"Epithelial cells [#/area] in Urine sediment by Microscopy high power field"

50378;1774;"FIBRINOGE";"BLOOD";"HEMATOLOGY";"Fibrinogen [Mass/volume] in Platelet poor plasma by Coagulation assay"

50013;1719;"O2";"BLOOD";"BLOOD GAS";"Oxygen/Inspired gas setting [Volume Fraction] Ventilator"

50024;1605;"TIDAL VOL";"BLOOD";"BLOOD GAS";"Tidal volume setting Ventilator"

50056;1565;"ACETMNPHN";"BLOOD";"CHEMISTRY";"Acetaminophen [Mass/volume] in Serum or Plasma"

50072;1564;"ASA";"BLOOD";"CHEMISTRY";"Acetylsalicylate [Mass/volume] in Serum or Plasma"

50099;1564;"ETHANOL";"BLOOD";"CHEMISTRY";"Ethanol [Mass/volume] in Blood"

50187;1542;"bnzodzpn";"BLOOD";"CHEMISTRY";"Benzodiazepines [Presence] in Blood"

50198;1538;"tricyclic";"BLOOD";"CHEMISTRY";"Tricyclic antidepressants [Presence] in Serum or Plasma"

50186;1538;"barbitrt";"BLOOD";"CHEMISTRY";"Barbiturates [Presence] in Serum, Plasma or Blood"

50062;1488;"ALT(SGPT)";"BLOOD";"CHEMISTRY";"Alanine aminotransferase [Enzymatic activity/volume] in Serum or Plasma"

50073;1482;"AST(SGOT)";"BLOOD";"CHEMISTRY";"Aspartate aminotransferase [Enzymatic activity/volume] in Serum or Plasma"

50170;1436;"TOT BILI";"BLOOD";"CHEMISTRY";"Bilirubin [Mass/volume] in Serum or Plasma"

50020;1426;"RATES";"BLOOD";"BLOOD GAS";""

50061;1409;"ALK PHOS";"BLOOD";"CHEMISTRY";"Alkaline phosphatase [Enzymatic activity/volume] in Blood"

50193;1395;"estGFR";"BLOOD";"CHEMISTRY";"Glomerular filtration rate/1.73 sq M.predicted by Creatinine-based formula (MDRD)"

50291;1179;"bnzodzpn";"URINE";"CHEMISTRY";"Benzodiazepines [Presence] in Urine"

50290;1179;"barbitrt";"URINE";"CHEMISTRY";"Barbiturates [Presence] in Urine"

50289;1178;"amphetmn";"URINE";"CHEMISTRY";"Amphetamines [Presence] in Urine"

50296;1178;"opiates";"URINE";"CHEMISTRY";"Opiates [Presence] in Urine"

50292;1178;"cocaine";"URINE";"CHEMISTRY";"Cocaine [Presence] in Urine"

50295;1176;"mthdone";"URINE";"CHEMISTRY";"Methadone [Presence] in Urine"

50021;1158;"REQ O2";"BLOOD";"BLOOD GAS";""

50001;1157;"AADO2";"BLOOD";"BLOOD GAS";"Oxygen.alveolar - arterial [Partial pressure] Respiratory system"

50396;1122;"HYPOCHROM";"BLOOD";"HEMATOLOGY";"Hypochromia [Presence] in Blood by Light microscopy"

50409;1099;"MACROCYT";"BLOOD";"HEMATOLOGY";"Macrocytes [Presence] in Blood"

50415;1080;"MICROCYT";"BLOOD";"HEMATOLOGY";"Microcytes [Presence] in Blood"

50326;1080;"ANISOCYT";"BLOOD";"HEMATOLOGY";"Anisocytosis [Presence] in Blood"

50157;1060;"RED HOLD";"BLOOD";"CHEMISTRY";""

50332;1056;"BANDS";"BLOOD";"HEMATOLOGY";"Neutrophils.band form/100 leukocytes in Blood"

50060;1052;"ALBUMIN";"BLOOD";"CHEMISTRY";"Albumin [Mass/volume] in Serum or Plasma"

50431;1010;"POIKILOCY";"BLOOD";"HEMATOLOGY";"Poikilocytosis [Presence] in Blood by Light microscopy"

50022;925;"TCO2";"BLOOD";"BLOOD GAS";"Bicarbonate [Moles/volume] in Blood"

50138;860;"LIPASE";"BLOOD";"CHEMISTRY";"Triacylglycerol lipase [Enzymatic activity/volume] in Serum or Plasma"

50429;822;"PLT SMR";"BLOOD";"HEMATOLOGY";"Platelets [Presence] in Blood by Manual count"

50267;789;"GR HOLD";"URINE";"CHEMISTRY";""

50432;782;"POLYCHROM";"BLOOD";"HEMATOLOGY";"Polychromasia [Presence] in Blood by Light microscopy"

50023;775;"TEMP";"BLOOD";"BLOOD GAS";""