Paper 077-31

Data and Text Mining the Electronic Medical Record to Improve Care and to Lower Costs

Patricia Cerrito, University of Louisville, Louisville, KY
John C. Cerrito, Kroger Pharmacy, Louisville, KY

ABSTRACT

This analysis uses data-mining techniques on an electronic medical record in the Emergency Department of a hospital to improve care while lowering costs. All patients' records for a 6-month period were examined, and records of those patients who had an initial complaint of shortness-of-breath were extracted. The data-mining techniques of transactional time series in the HPF procedure and the association rules in SAS® Enterprise Miner™ were used to examine the data. Patients' orders, medications, and complaints were also examined using SAS® Text Miner to investigate relationships among the variable categories.

The Association Node in SAS Enterprise Miner is applied to one target variable that uses a patient identifier to link orders, medications, and charges. Unfortunately, the Association Node is inadequate when there are too many choices for each target; it cannot relate different target values to each other. An alternative method is to change the observational unit to the patient by using the TRANSPOSE and CONCAT procedures. In this way, all patients' orders, medications, and changes are linked in text strings that can be examined and compared using SAS Text Miner. It was discovered that patients with similar complaints were treated very differently depending on the attending physician, and those differences can impact both costs and care in a hospital Emergency Department.

INTRODUCTION

The introduction of the electronic medical record allows for the examination of cost and quality in healthcare. Large numbers of patient records can now be examined in a hospital or private practice setting, and patients can be tracked longitudinally through the healthcare system. In the past, data mining techniques have rarely been used to examine the totality of patient records; matched cohorts have been preferred to examine specific hypotheses.

This paper will examine physician decision-making to determine whether there is a general consensus as to the treatment of patients in the Emergency Department. All electronic records for a six month period were examined using SAS Enterprise Miner and SAS Text Miner.

It is also the purpose of this paper to examine changes in diagnosis from initial complaint to final outcome in the hospital emergency department. In the Emergency Department (ED), those patients who are determined to have emergent conditions are seen before patients seen as having urgent or non-urgent conditions. Generally, the triage values of emergent, urgent, and non-urgent are assigned based upon the initial complaint. Patients identified as non-urgent can become emergent once the final diagnosis is made.

EXAMINATION OF COMMON PROBLEM IN THE HOSPITAL EMERGENCY DEPARTMENT

Chronic obstructive pulmonary disease (COPD) was chosen for examination. Over the six-month period, approximately 53 patients were diagnosed in the ED with COPD. Table 1 shows triage values, where patients were categorized as non-urgent, urgent, and emergent upon entry into the ED. Table 2 shows the relationship between the triage level and the patient disposition after treatment in the ED.

Table 1. Patients Diagnosed with COPD as Triaged Upon Entry into the ED

Level of Urgency	Number of Patients
Non-Urgent	12
Urgent	38
Emergent	3

Note that 72% of the patients were triaged as urgent while less than 10% were emergent.

Table 2. Relationship of Initial Triage to Final Disposition

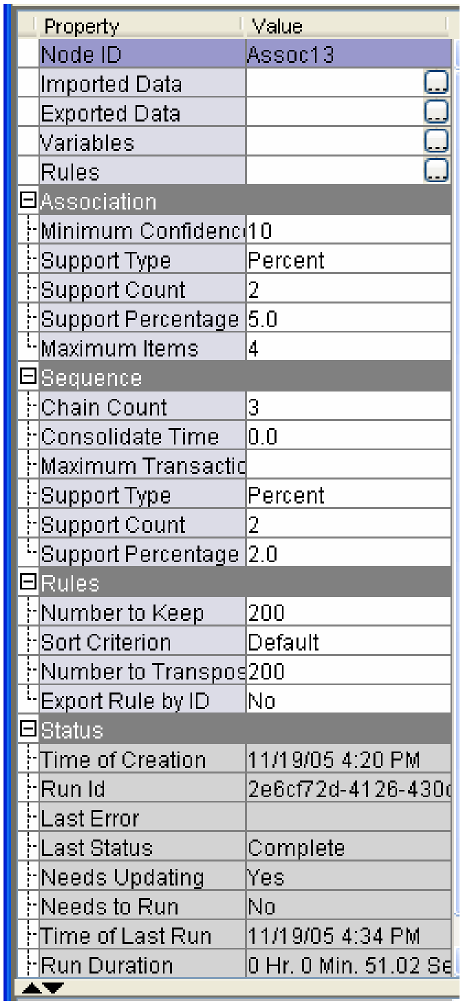
Patient Disposition	Non-Urgent	Urgent	Emergent
Discharged Home	6	4	0
Admit for 23-hour Observation	4	22	3
Admit as Inpatient	1	7	0
Admit to ICU	0	5	0

Since all 3 emergent patients were admitted for observation, it is clear that this must be a standard protocol. Note, however, that almost 50% of those initially classified as non-urgent were also admitted for observation.

Moreover, none of the patients admitted to ICU were initially identified as emergent. Their urgency level clearly changed during treatment in the ED. However, that so many were emergent but not initially identified as emergent suggests that drill down into the data should be done to see if there are any reasons that the emergent status was overlooked at triage.

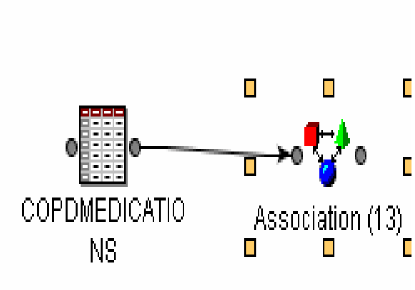
Another concern is with physician decision-making in the ED for patients with similar complaints. To examine the issue, association rules were used. An association rule is an if…then statement. There are two measures concerning the validity of the rule. The first is the support, which is the number of transactions that consists of the intersection of items from the “if” part of the rule and the “then” part. It is sometimes given as a percentage of the total number of items in the dataset. The confidence is the ratio of the support divided by the number of items from the “if” part of the statement. SAS Enterprise Miner can depict these association rules in the form of a link graph where the If part is connected to the then part by a line. The width of the line depicts the strength of the association. The size of the rectangle representing the item depicts the number of items. Figures 1 and 2 show how to set up the Association Node in SAS EM 5.2. Figure 3 gives the EM output. Figure 4 gives the link analysis for medications ordered (as contained within the pharmacy database) for patients diagnosed with COPD. In Figure 4, physician is given as the ID variable and medication is the Target variable.

Figure 1. Association Rules in SAS EM 5.2.



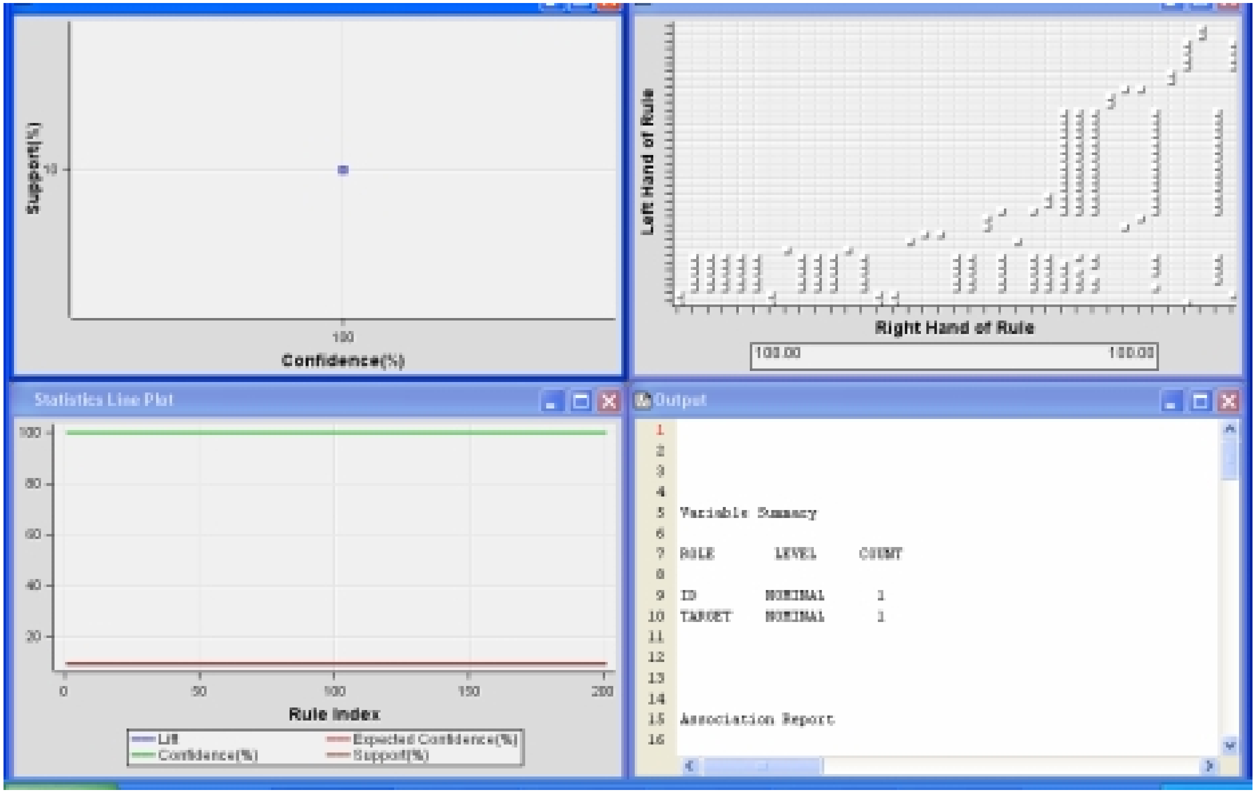
Property	Value
Node ID	Assoc13
Imported Data	...
Exported Data	...
Variables	...
Rules	...
⊟Association	
Minimum Confidenc	10
Support Type	Percent
Support Count	2
Support Percentage	5.0
Maximum Items	4
⊟Sequence	
Chain Count	3
Consolidate Time	0.0
Maximum Transactic	
Support Type	Percent
Support Count	2
Support Percentage	2.0
⊟Rules	
Number to Keep	200
Sort Criterion	Default
Number to Transpos	200
Export Rule by ID	No
⊟Status	
Time of Creation	11/19/05 4:20 PM
Run Id	2e6cf72d-4126-430
Last Error	
Last Status	Complete
Needs Updating	Yes
Needs to Run	No
Time of Last Run	11/19/05 4:34 PM
Run Duration	0 Hr. 0 Min. 51.02 Se

Figure 2. Association Rules Continued



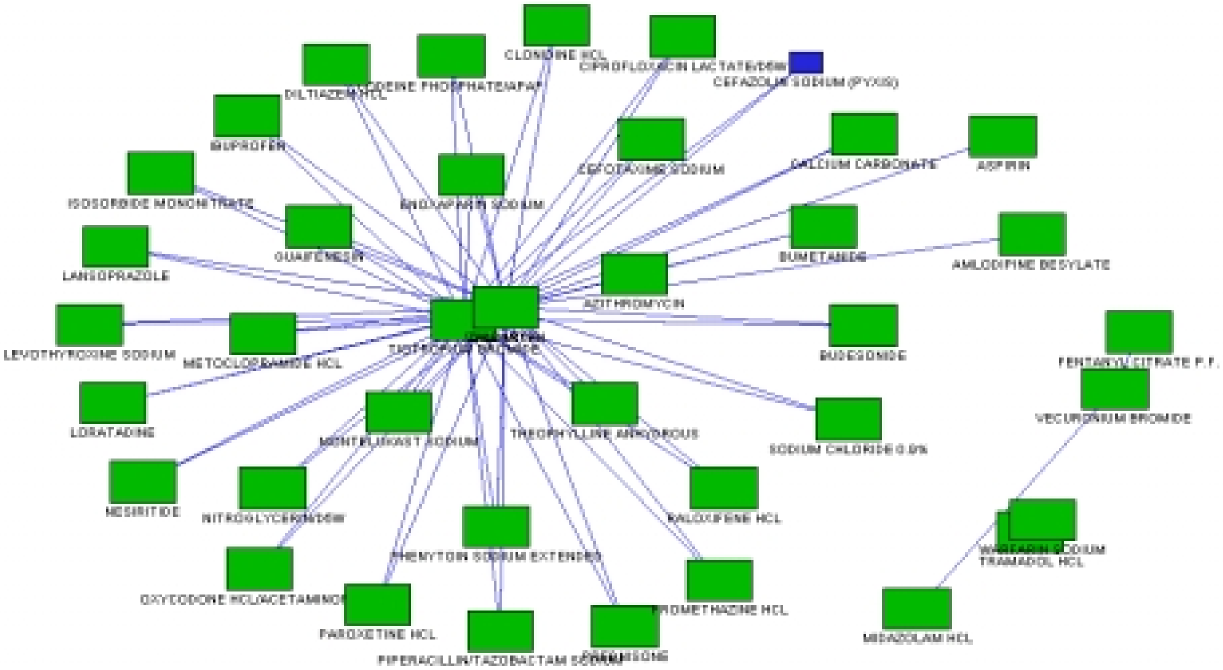
The default output gives the first 200 if…then rules that have the highest lift. In Figure 3, all 200 rules have a confidence of 100% and a support of 10%. Figure 4 also shows the homogeneity of the rules with identical nodes and connectors. Such homogeneity generally indicates that there are very few transactions for each combination of items.

Figure 3. Default Output for the Association Node



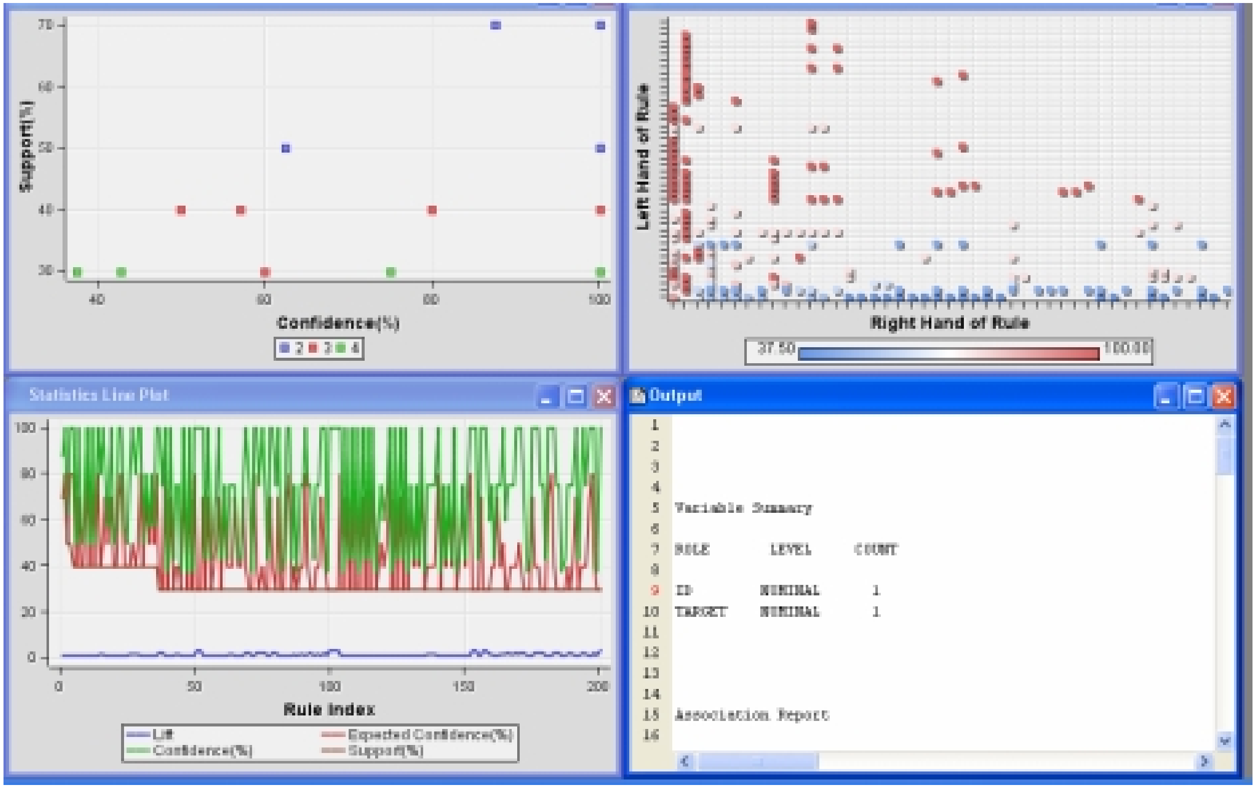
```
Variable Summary

ROLE        LEVEL      COUNT

ID          NOMINAL      1
TARGET      NOMINAL      1




Association Report

```

Figure 4. Link Analysis of Medication Orders Using Association Node Defaults



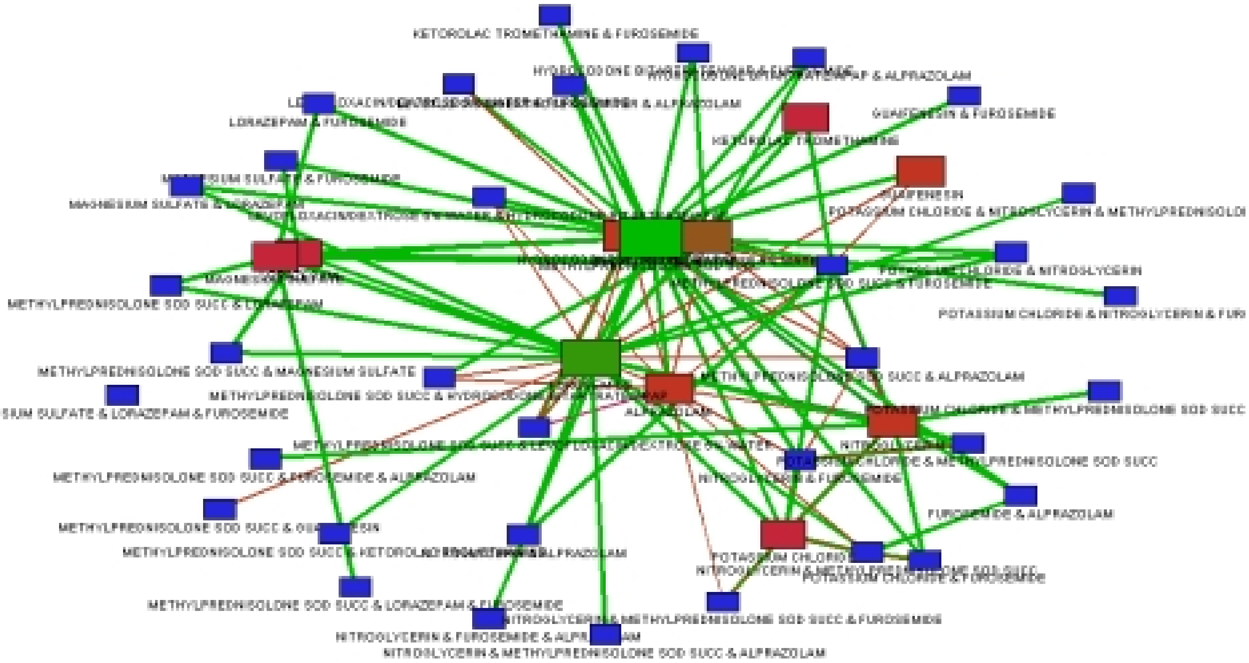
Note that the nodes and lines are virtually the same. It indicates that there are no dominant associations. Many of the nodes represent antibiotics (Azithromycin, Piperacillin/Tazobactam, defazolin, cefotaxime, ciprofloxacin). Others represent pain medications (oxycodone, ibuprofen, aspirin). This graph shows a lack of consistency in treatment from physician to physician, particularly for secondary problems since all patients represented here have COPD as the primary diagnosis. Figures 5 and 6 give the results when the default of highest confidence is changed to largest number of counts. High confidence occurs when there are almost no items in the intersection of antecedent and consequent. Of interest is why so many patients are prescribed antibiotics for a non-infectious diagnosis.

Figure 5. Output by Highest Transaction Counts



```




Variable Summary

ROLE       LEVEL      COUNT

ID         NOMINAL      1
TARGET     NOMINAL      1




Association Report

```

Figure 6. Corresponding Link Analysis


KETOROLAC TROMETHAMINE & FUROSEMIDE
HYDROCODONE BITARTRATE/ACETAMINOPHEN & ALPRAZOLAM
LORAZEPAM & FUROSEMIDE
GUAIFENESIN & FUROSEMIDE
KETOROLAC TROMETHAMINE
MAGNESIUM SULFATE & FUROSEMIDE
MAGNESIUM SULFATE & LORAZEPAM
GUAIFENESIN
POTASSIUM CHLORIDE & NITROGLYCERIN & METHYLPREDNISOLO
MAGNESIUM SULFATE
POTASSIUM CHLORIDE & NITROGLYCERIN
METHYLPREDNISOLONE SOD SUCC & FUROSEMIDE
METHYLPREDNISOLONE SOD SUCC & LORAZEPAM
POTASSIUM CHLORIDE & NITROGLYCERIN & FUR
METHYLPREDNISOLONE SOD SUCC & MAGNESIUM SULFATE
METHYLPREDNISOLONE SOD SUCC & ALPRAZOLAM
METHYLPREDNISOLONE SOD SUCC & HYDROCODONE BITARTRATE/APAP
SIUM SULFATE & LORAZEPAM & FUROSEMIDE
ALPRAZOLAM
POTASSIUM CHLORIDE & METHYLPREDNISOLONE SOD SUCC
NITROGLYCERIN
POTASSIUM CHLORIDE & METHYLPREDNISOLONE SOD SUCC
NITROGLYCERIN & FUROSEMIDE
METHYLPREDNISOLONE SOD SUCC & FUROSEMIDE & ALPRAZOLAM
FUROSEMIDE & ALPRAZOLAM
METHYLPREDNISOLONE SOD SUCC & GUAIFENESIN
METHYLPREDNISOLONE SOD SUCC & KETOROLAC TROMETHAMINE
POTASSIUM CHLORIDE
NITROGLYCERIN & METHYLPREDNISOLONE SOD SUCC
POTASSIUM CHLORIDE & FUROSEMIDE
METHYLPREDNISOLONE SOD SUCC & LORAZEPAM & FUROSEMIDE
NITROGLYCERIN & METHYLPREDNISOLONE SOD SUCC & FUROSEMIDE
NITROGLYCERIN & FUROSEMIDE & ALPRAZOLAM
NITROGLYCERIN & METHYLPREDNISOLONE SOD SUCC & ALPRAZOLAM

While difficult to read, there are some indications that methylprednisolone is the treatment chosen most often (a steroid used for emergency episodes of COPD and asthma). Note also that when the high counts are used, antibiotics are more rare in the association rules. It is possible to drill down into the data by using the Viewport (available in EM 5.2).

Figure 7. Drill Down into the Link Analysis

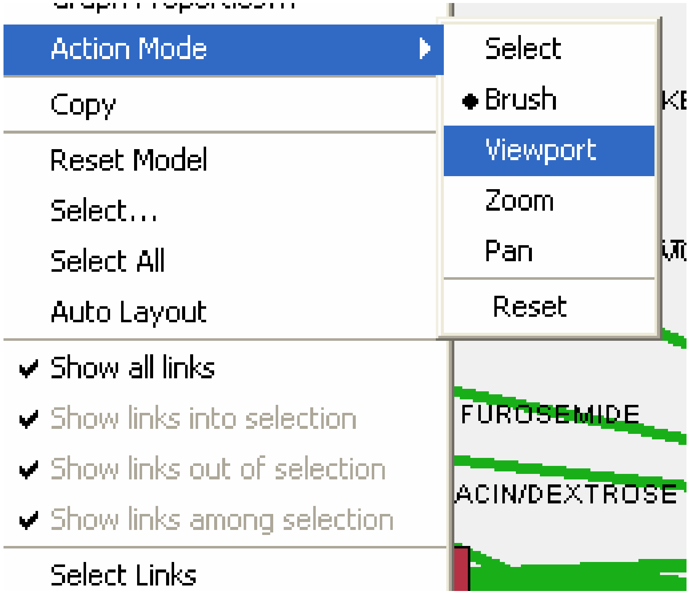
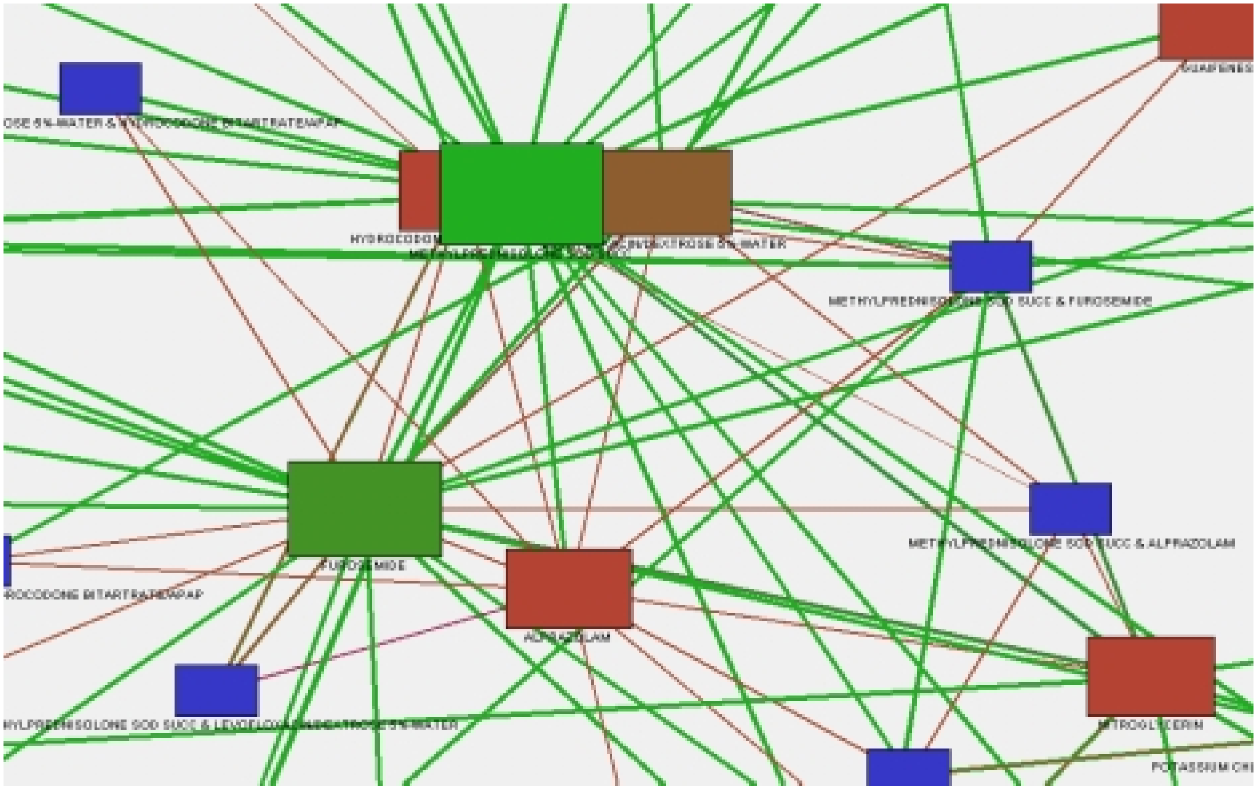


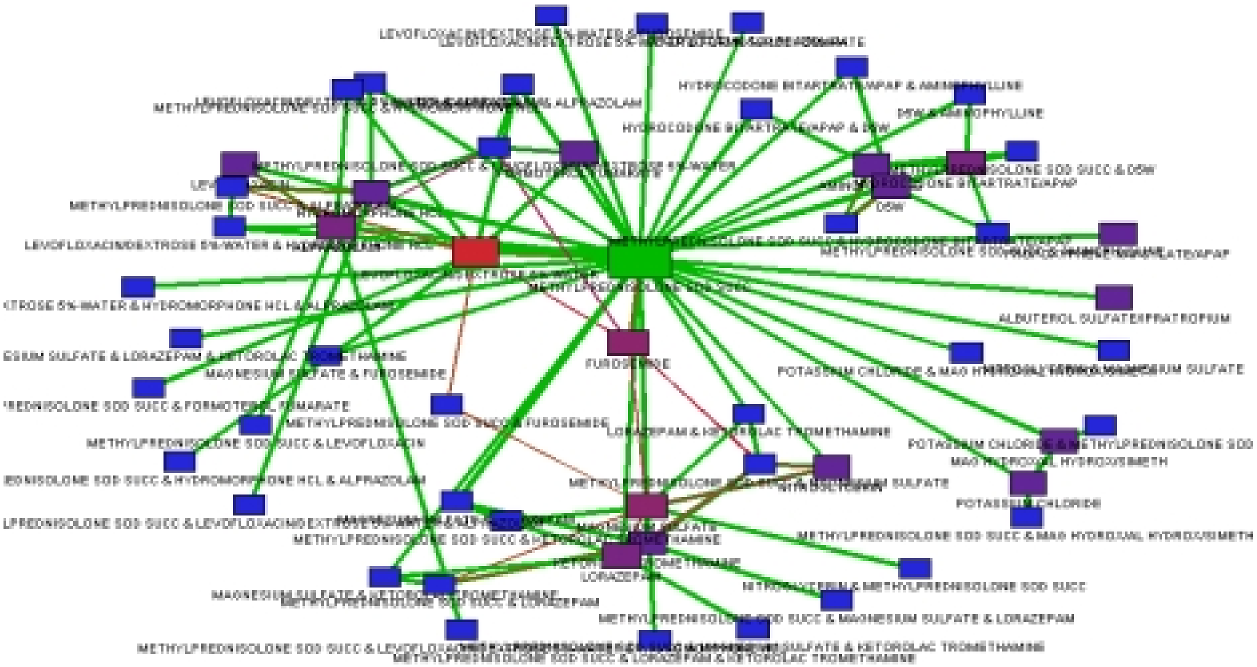
Figure 8 gives the resulting Viewport. The resulting window shows that hydrocodone (pain relief) is also prominent in terms of the number of transactions. Note the priminance of the drug, methylprednisolone.

Figure 8. Viewport Using Transaction Counts



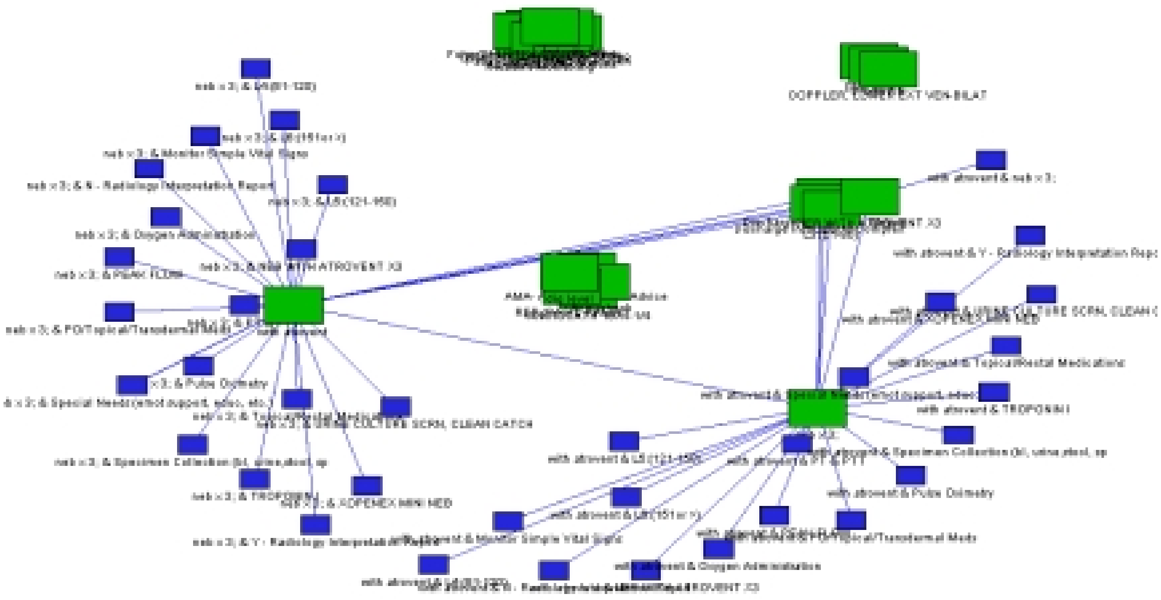
In Figure 9, the perspective is changed. Instead of using the physician as the ID variable, the patient becomes the ID to determine whether patients with a similar problem in the ED (COPD) are treated similarly. Again, the transaction counts are used for the analysis.

Figure 9. Link Analysis of Associations by Patient


HYDROCODONE BITARTRATE/APAP & AMINOPHYLLINE
D5W & AMINOPHYLLINE
HYDROCODONE BITARTRATE/APAP & D5W
METHYLPREDNISOLONE SOD SUCC & D5W
HYDROCODONE BITARTRATE/APAP
D5W
LEVOFLOXACIN/DEXTROSE 5%-WATER & HYDROMORPHONE HCL
XTROSE 5%-WATER & HYDROMORPHONE HCL & ALPRAZOLAM
MAGNESIUM SULFATE & FUROSEMIDE
REDNISOLONE SOD SUCC & FORMOTEROL FUMARATE
METHYLPREDNISOLONE SOD SUCC & FUROSEMIDE
METHYLPREDNISOLONE SOD SUCC & LEVOFLOXACIN
EDNISOLONE SOD SUCC & HYDROMORPHONE HCL & ALPRAZOLAM
ALBUTEROL SULFATE/IPRATROPIUM
FUROSEMIDE
POTASSIUM CHLORIDE & MAG HYDROXIDE
MAGNESIUM SULFATE
LORAZEPAM & KETOROLAC TROMETHAMINE
POTASSIUM CHLORIDE & METHYLPREDNISOLONE SOD
MAG HYDROXIDE HYDROXIMETH
POTASSIUM CHLORIDE
METHYLPREDNISOLONE SOD SUCC & MAG HYDROXAL HYDROXIMETH
NITROGLYCERIN & METHYLPREDNISOLONE SOD SUCC
LORAZEPAM
MAGNESIUM SULFATE & KETOROLAC TROMETHAMINE
METHYLPREDNISOLONE SOD SUCC & LORAZEPAM
METHYLPREDNISOLONE SOD SUCC & MAGNESIUM SULFATE & LORAZEPAM
SULFATE & KETOROLAC TROMETHAMINE
METHYLPREDNISOLONE SOD SUCC & LORAZEPAM & KETOROLAC TROMETHAMINE

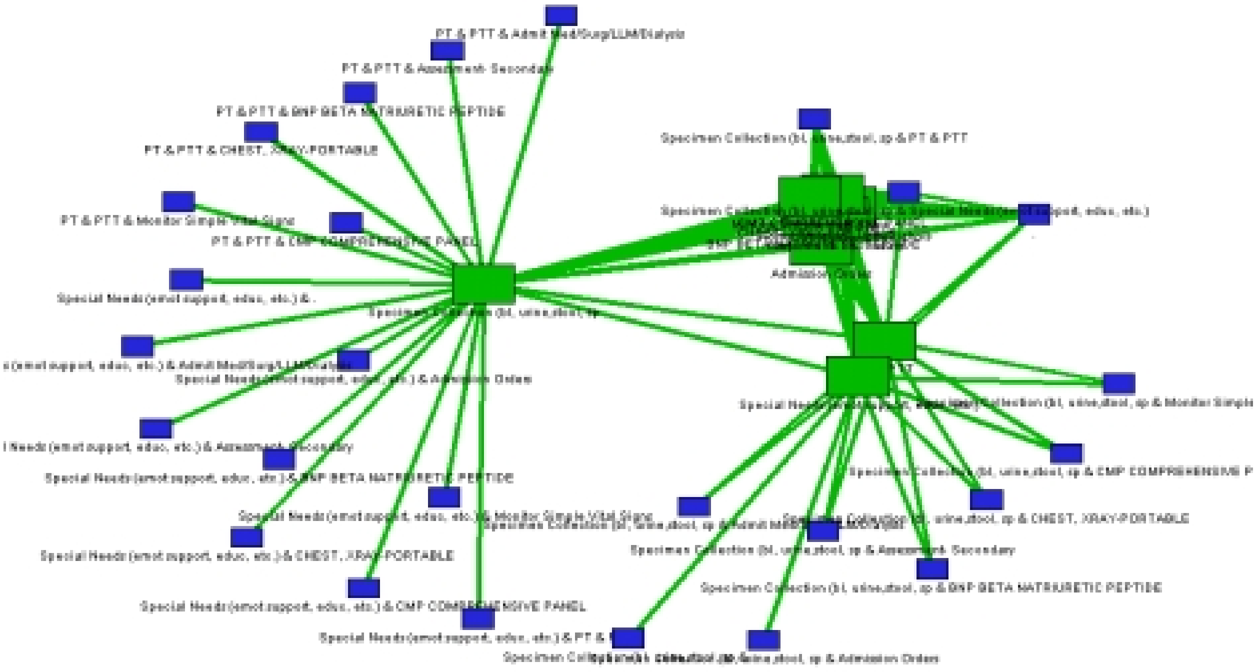
Methylprednisolone is at the center of the link analysis. The primary medications connected to this center node are Levofloxacin (an antibiotic), Lorazepam (anti-anxiety), Potassium chloride (potassium deficiency), and albuterol (asthma). In addition to medications, charges (including labs) were also recorded in the electronic database. Figure 10 gives the charges, using patient as the ID variable in the link analysis, using the default of highest confidence.

Figure10. Charges by Patient Identifier


DOPPLER, LOWER EXT VEN-BILAT
neb x 3; & Monitor Simple Vital Signs
neb x 3; & X - Radiology Interpretation Report
neb x 3; & Oxygen Administration
neb x 3; & PEAK FLOW
neb x 3; & NEB WITH ATROVENT X3
neb x 3; & PO/Topical/Transdermal Meds
x 3; & Pulse Oximetry
d x 3; & Special Need (emot support, educ, etc.)
neb x 3; & Specimen Collection (bl, urine, stool, sp
neb x 3; & TROPONIN I
neb x 3; & ADPEMEX MINI NEB
neb x 3; & Y - Radiology Interpretation Report
with atrovent & neb x 3;
with atrovent & Y - Radiology Interpretation Repc
with atrovent & Topical/Rectal Medications
with atrovent & TROPONIN I
with atrovent & Specimen Collection (bl, urine, stool, sp
with atrovent & Pulse Oximetry
with atrovent & Monitor Simple Vital Signs
with atrovent & Oxygen Administration

While some nodes remain unassociated, a consensus is reached as indicated by the symmetry in the model. Atrovent is used for COPD patients along with neb x 3 as shown by the nodes on the left-hand side of the diagram. Similarly, at the lower right, neb x 3 is used with atrovent. These are medications kept in the ED because they need to be administered immediately. Therefore, there is no documentation of their administration in the pharmacy database. This lack of documentation indicates a need for the ED electronic system to be able to send information to the pharmacy database. Figure 11 gives the result using the transaction counts.

Figure 11. Link Analysis of Patient Charges Using Transaction Counts


PT & PTT & Admit Med/Surg/LLM/Dialysis
PT & PTT & Assessment- Secondary
PT & PTT & BNP BETA NATRIURETIC PEPTIDE
PT & PTT & CHEST, XRAY-PORTABLE
PT & PTT & Monitor Simple Vital Signs
PT & PTT & CMP COMPREHENSIVE PANEL
Special Needs (emot support, educ, etc.) & .
Specimen Collection (bl, urine,stool, sp
s (emot support, educ, etc.) & Admit Med/Surg/LLM/Dialysis
Special Needs (emot support, educ, etc.) & Admission Orders
l Needs (emot support, educ, etc.) & Assessment- Secondary
Special Needs (emot support, educ, etc.) & BNP BETA NATRIURETIC PEPTIDE
Special Needs (emot support, educ, etc.) & Monitor Simple Vital Signs
Special Needs (emot support, educ, etc.) & CHEST, XRAY-PORTABLE
Special Needs (emot support, educ, etc.) & CMP COMPREHENSIVE PANEL
Special Needs (emot support, educ, etc.) & PT &
Specimen Collection (bl, urine,stool, sp & PT & PTT
Specimen Collection (bl, urine,stool, sp & Special Needs (emot support, educ, etc.)
Admission Orders
Special Needs (emot support, educ, etc.)
Specimen Collection (bl, urine,stool, sp & Monitor Simple
Specimen Collection (bl, urine,stool, sp & CMP COMPREHENSIVE P
Specimen Collection (bl, urine,stool, sp & Admit Med/Surg/LLM/Dialysis
Specimen Collection (bl, urine,stool, sp & CHEST, XRAY-PORTABLE
Specimen Collection (bl, urine,stool, sp & Assessment- Secondary
Specimen Collection (bl, urine,stool, sp & BNP BETA NATRIURETIC PEPTIDE
Specimen Collection (bl, urine,stool, sp & Admission Orders

Standard procedure is for a specimen collection. Most connections also lead to special needs (emotional support and education).

TEXT ANALYSIS FOR CLASSIFICATION

There are many different diagnoses related to shortness of air; COPD is just one of those diagnoses. If each diagnosis is used separately for analysis, the dataset becomes too diluted. Therefore, the diagnosis should be grouped. Out of almost 11,000 patients collected over a 6-month period, a total of 1329 patients were seen in the ED with initial complaints related to shortness of air, but only 50 were specifically diagnosed with COPD. It is one of the main occurrences of shortness of air. It is possible to aggregate patients with similar diagnoses and complaints through the use of SAS Text Miner.

There were a total of 1,112 different terms used to define the patient's initial complaint. In contrast, there were a total of only 335 terms used to define the final diagnosis, and a total of 280 different medications used. The reduction in the number of terms from initial complaint to final diagnosis does indicate that some standardization of language has taken place, but the physicians have full discretion to write in non-standard terms. Since these are all nominal values, text analysis using SAS Text Miner was performed to reduce the number of terms into a meaningful number of patient clusters. Because of so many different, but related diagnoses, it is very difficult to separate and investigate each diagnosis separately.

Both initial complaints and final diagnoses were grouped using text analysis into 10 different clusters. The clusters for initial complaints are given in Table 3.

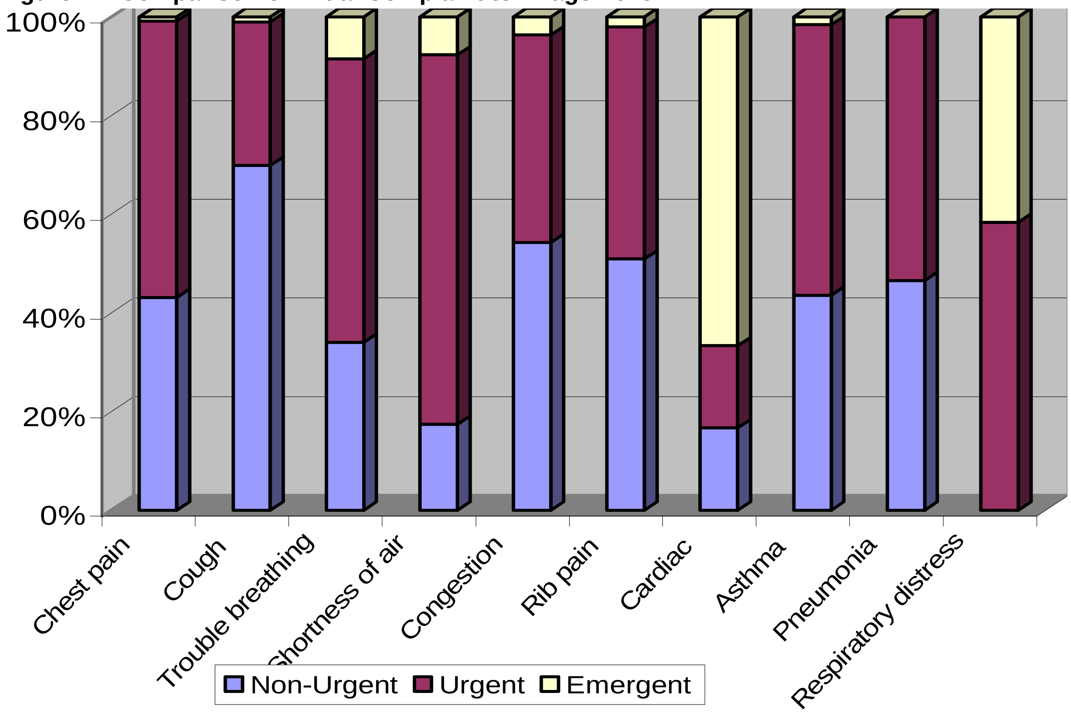
Table 3. Complaint Clusters for Shortness of Air

Cluster Label	Cluster Number	Descriptive Terms	Frequency	Percent
Chest pain	1	Chest pain, deep breath, throat	126	9%
Cough	2	Cough, sore, productive cough, sore throat	192	14%
Trouble breathing	3	Trouble breathing, dizziness, weakness	50	4%
Shortness of air	4	Shortness of breath	269	21%
Congestion	5	Congestion, cold, cough	171	13%
Rib pain	6	Rib, ear, back pain	359	27%

Cluster Label	Cluster Number	Descriptive Terms	Frequency	Percent
Cardiac	7	Arrest, cardiac complaint, multiple complaints, chest pain	14	1%
Asthma	8	Asthma, asthma attack, cough	72	5%
Pneumonia	9	Pneumonia, possible pneumonia, fever, nausea	63	5%
Respiratory distress	10	Respiratory distress	13	1%

The descriptive terms are provided by the text analysis as the ones most important in defining the clusters. Figure 12 shows the relationship between initial complaint and triage level. Complaints centered on cardiac problems, or respiratory distress have a high proportion of emergent patients compared to the others. In fact, respiratory distress has zero patients identified as non-urgent. Patients with no other complaint except shortness of air have a non-urgent level that is similar to those with cardiac complaints. Similarly, Figure 13 shows the relationship between initial complaint and final patient disposition.

Figure 12. Comparison of Initial Complaint to Triage Level



In Figure 13, most patients in the respiratory distress cluster are discharged to ICU (TCU is considered one step down from ICU), a higher proportion than those complaining of cardiac problems. Those with cough or congestion are likely to be discharged home and only a small proportion are discharged to ICU. The final diagnoses clusters are given in Table 4. A stacked graph comparing complaints to final diagnoses is given in Figure 14.

It demonstrates that there is considerable shifting from the initial complaint to the final cluster. For example, only 27% of those patients with an initial complaint of asthma had a final diagnosis of asthma; 4% with an initial complaint of asthma were diagnosed with Chronic obstructive pulmonary disease (COPD). In contrast, nearly 71% with an initial cardiac complaint had a final diagnosis of cardiac arrest.

Figure 15 gives the final diagnosis by triage level; Figure 16 gives the final patient disposition by diagnosis cluster. The clusters with the greatest proportion of non-urgent patients are COPD and pleuritic pain. All of the cardiac arrest patients were either urgent or emergent based upon their initial complaints. Next in order of severity is the cluster of pulmonary edema.

Figure 13. Comparison of Initial Complaint to Final Patient Disposition

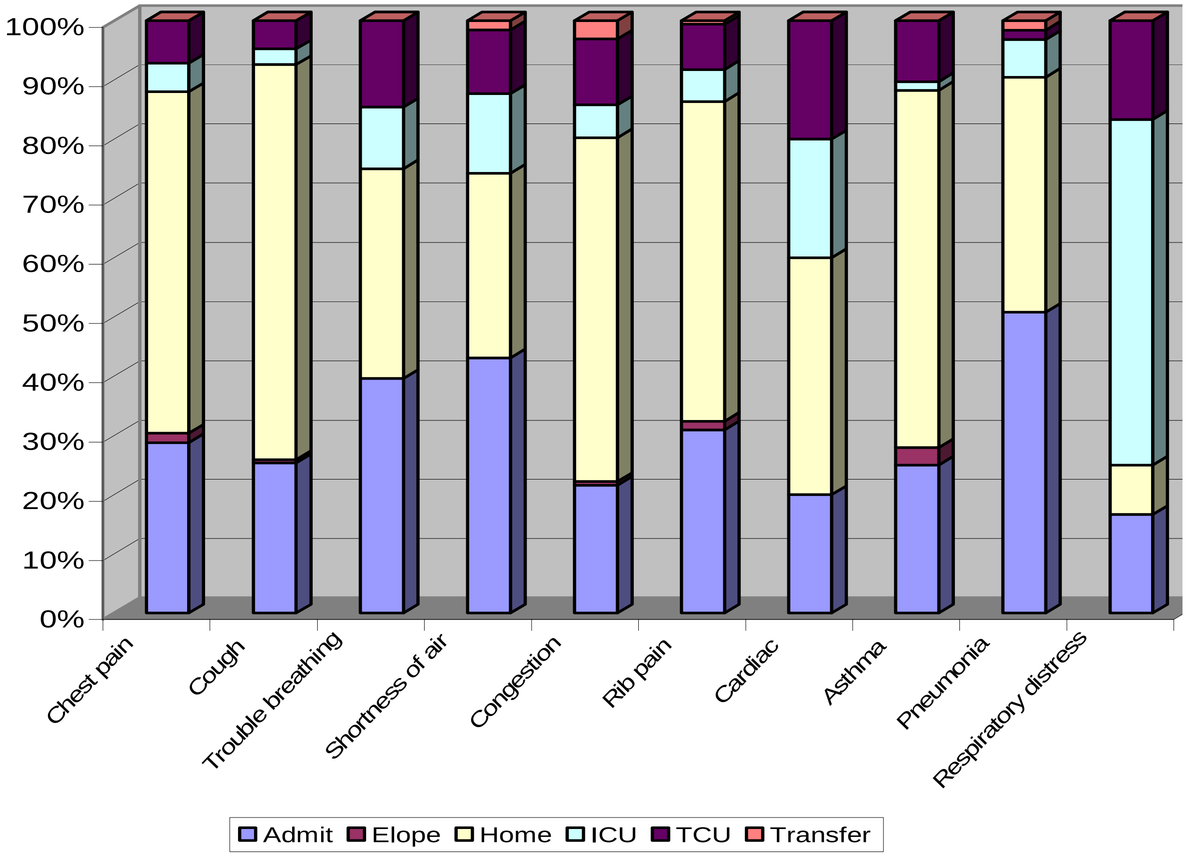


Table 4. Diagnosis Clusters for Shortness of Air

Cluster Label	Cluster Number	Descriptive Terms	Frequency	Percent
Bronchitis	1	Bronchitis, acute bronchitis	224	17%
Infection in Lung	2	Right, lobe, infection	155	12%
COPD	3	COPD, acute COPD, acute exacerbation	142	11%
Pneumonia	4	Pneumonia	166	12%
Pleuritic chest pain	5	Chest pain, pleuritic chest pain	33	2%
Sinusitis	6	Sinusitis, acute sinusitis, chronic sinusitis	90	7%
Pulmonary edema	7	Bilateral, acute pulmonary edema	122	9%
Cardiac arrest	8	Arrest, cardiopulmonary arrest	19	1%
Respiratory failure	9	Respiratory failure, congestive heart failure, acute congestive heart failure	126	9%
Asthma	10	Asthma, asthma exacerbation, acute asthma exacerbation	252	19%

Figure 14. Graph of Initial Complaint by Final Diagnosis (X-axis=Complaints)

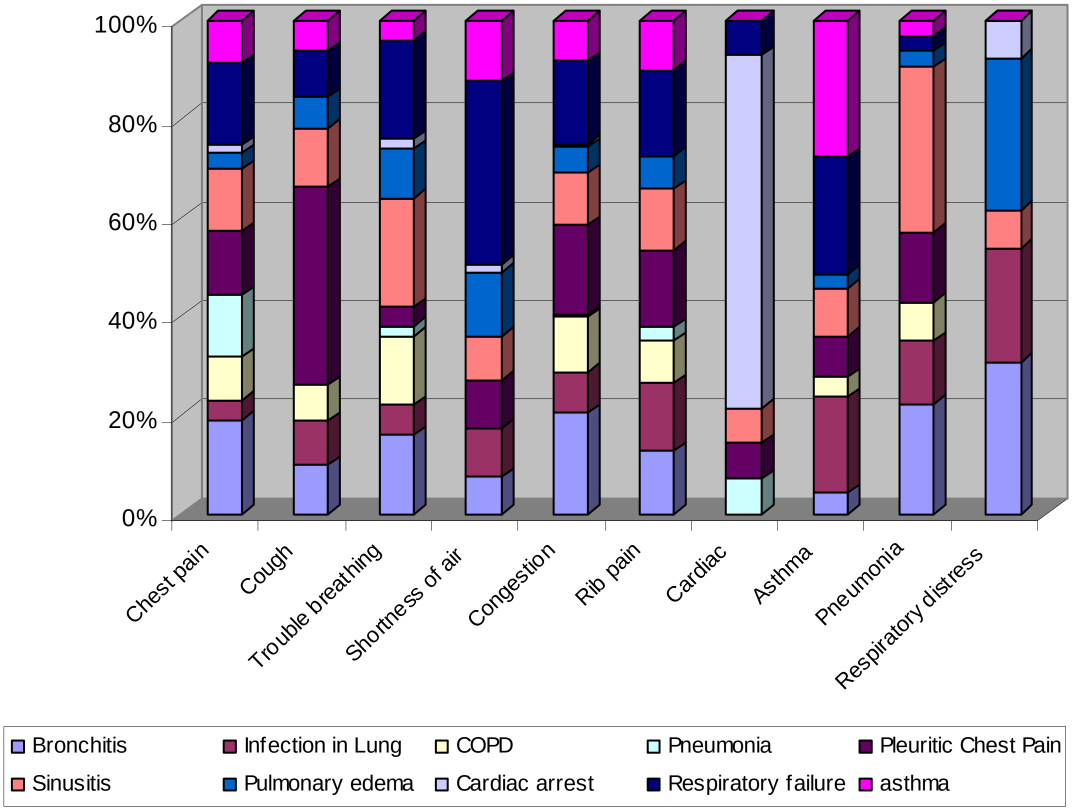
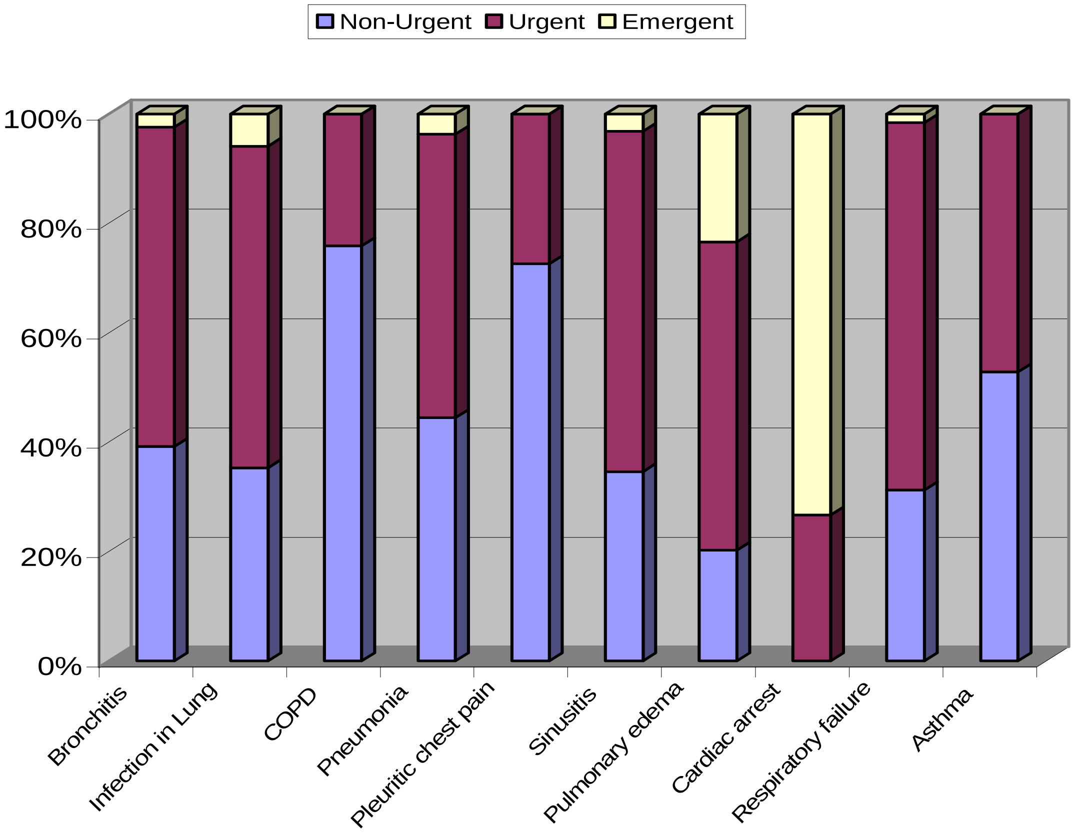
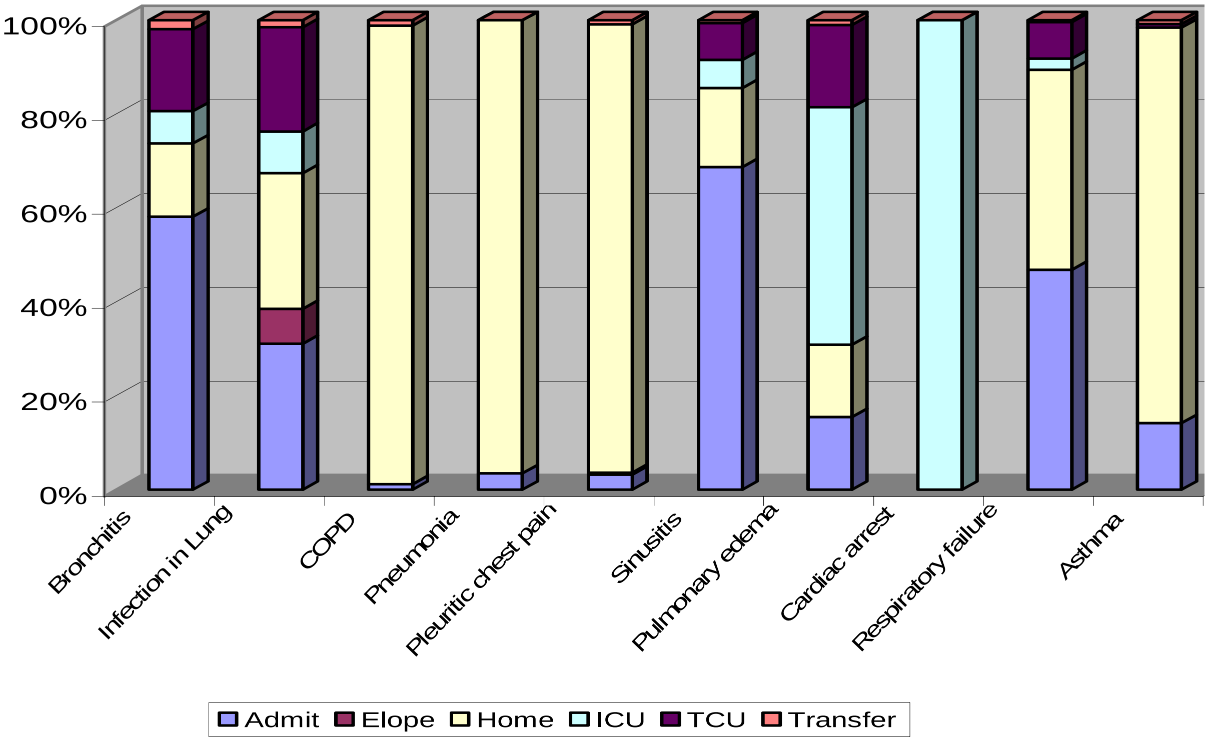


Figure 15. Comparison of Final Diagnosis by Triage



Every patient in the cardiac arrest cluster was discharged to ICU. Virtually every patient in the COPD, pneumonia, and pleuritic chest pain clusters were discharged home. 10% of asthma patients were admitted, the remainder were also discharged home. Interestingly enough, over 60% of patients in the sinusitis cluster were admitted to the hospital. Only 18% of those with pulmonary edema were not admitted, over 50% went to ICU. There are the beginnings of a consensus for several of the categories of diagnoses as to patient disposition. However, respiratory failure and lung infections do not yet have such consensus.

Figure 16. Comparison of Final Diagnosis to Final Disposition



DEFINITION OF MEDICATION CLUSTERS

To examine the issue of treatment, medications were also examined. Since patients often received multiple medications while in the ED, all medications for one patient were transposed and concatenated into one text string. SAS Text Miner was then used to define clusters of medication treatments (Table 5). The optimal result consisted of 13 different clusters, including one cluster of patients receiving no medications while in the ED.

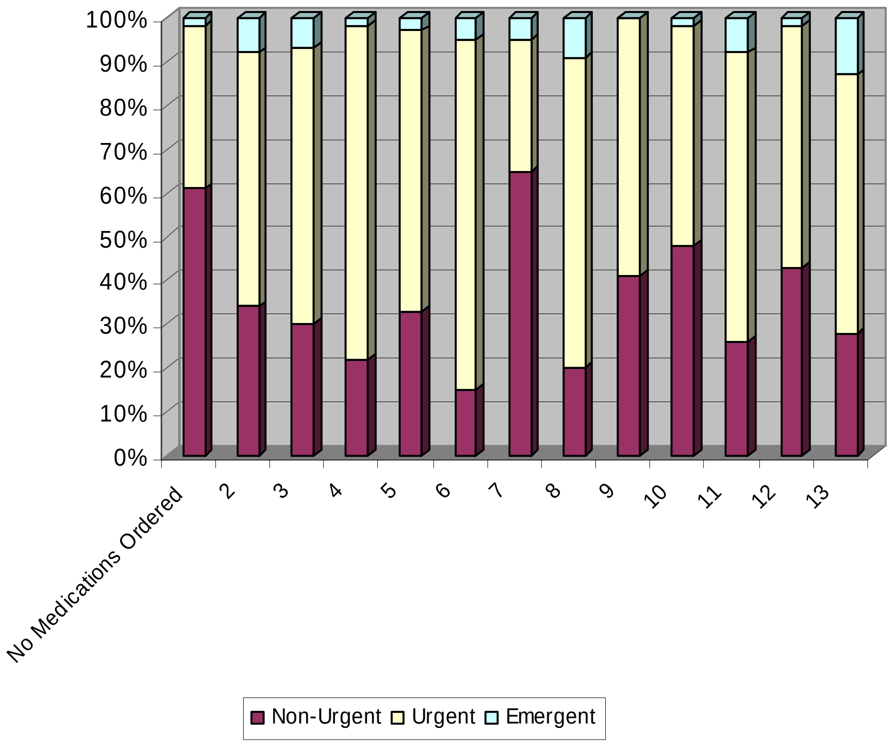
Table 5. Medication Clusters

Cluster Number	Descriptive Terms	Frequency
1	No medications listed	541
2	lorazepam, ketorolac_tromethamine, levofloxacin, prednisone, midazolam_hcl, hydrocodone_bitartrate/apap, magnesium_sulfate, oxycodone_hcl/acetaminophen, lidocaine_hcl, famotidine	293
3	ceftriaxone_sodium, furosemide, azithromycin, simvastatin, clonidine_hcl, oxycodone_hcl/acetaminophen, bumetanide, acetaminophen, aspirin, hydrocodone_bitartrate/apap	57
4	nitroglycerin, furosemide, bumetanide, budesonide, insul_nph_hu_rec/ins_rg_hu_rec, aspirin, tiotropium_bromide, hydrocodone_bitartrate/apap, methylprednisolone_sod_succ, diltiazem_hcl	66
5	methylprednisolone_sod_succ	69
6	porcine, porcine heparin_sodium, heparin_sodium, diphenoxylate_hcl/atrop_sulf, eptifibatide, warfarin_sodium, hydrocodone_bitartrate/apap, enoxaparin_sodium, ondansetron_hcl, morphine_sulfate	21

Cluster Number	Descriptive Terms	Frequency
7	ibuprofen, acetaminophen, cefotaxime_sodium, methylprednisolone_sod_succ	20
8	piperacillin/tazobactam_sodium, tacrolimus_anhydrous, thiamine_hcl, vancomycin_hcl_(iv_room), famotidine, ondansetron_hcl, lansoprazole, insulin_reg_human_rec, azithromycin, enoxaparin_sodium	35
9	ondansetron_hcl, ketoprofen, morphine_sulfate, promethazine_hcl, metoclopramide_hcl, ketorolac_tromethamine, lorazepam, ceftriaxone_sodium, azithromycin, methylprednisolone_sod_succ	35
10	promethazine_hcl, meperidine_hcl, guaifenesin/codeine_phosphate, hydromorphone_hcl, lidocaine_hcl, heparin_sodium, magnesium_sulfate, porcine, ondansetron_hcl, methylprednisolone_sod_succ	45
11	epoetin_alfa, calcium_acetate, enoxaparin_sodium, oxycodone_hcl, nesiritide, folic_acid, hydralazine_hcl, famotidine, metoprolol_succinate, aspirin	37
12	potassium_chloride, lidocaine_hcl, vancomycin_hcl_(iv_room), magnesium_sulfate, mag_hydrox/al_hydrox/simeth, guaifenesin, ketorolac_tromethamine, ondansetron_hcl, methylprednisolone_sod_succ, predniso	54
13	cefotaxime_sodium, sodium_bicarbonate, epinephrine, calcium_chloride, diphther_toxoid_adult, tetanus, sodium_polystyrene_sulfonate, azithromycin, atropine_sulfate, fosphenytoin_sodium	56

Figure 17 gives the relationship between medications given and initial triage value. 60% of those not given any medications were identified as non-urgent; a slightly higher proportion of non-urgent are in Cluster 7, requiring primarily a mild pain medication. Clusters 6 and 8 had the highest proportion of urgent and emergent combined. Cluster 8 largely contained antibiotics; cluster 6, heart medications. Similarly, Figure 18 shows the relationship between final disposition and medications.

Figure 17. Comparison of Triage and Medications



Cluster 7 has the highest proportion of non-urgent and the highest proportion discharged home. Note that some of those receiving no medications eloped before being seen by a physician. Clusters 3,4,8,11, and 13 had very few patients discharged home. Both 3 and 4 were prescribed medications for the heart, and intensive pain medications. Cluster 8 focused on antibiotics indicating infections. No patient in cluster 6 was discharged home; all were admitted.

This cluster also concentrates on heart medications. The relationship between medication cluster and final diagnosis is given in Figure 19.

Figure 18. Comparison of Final Disposition and Medications

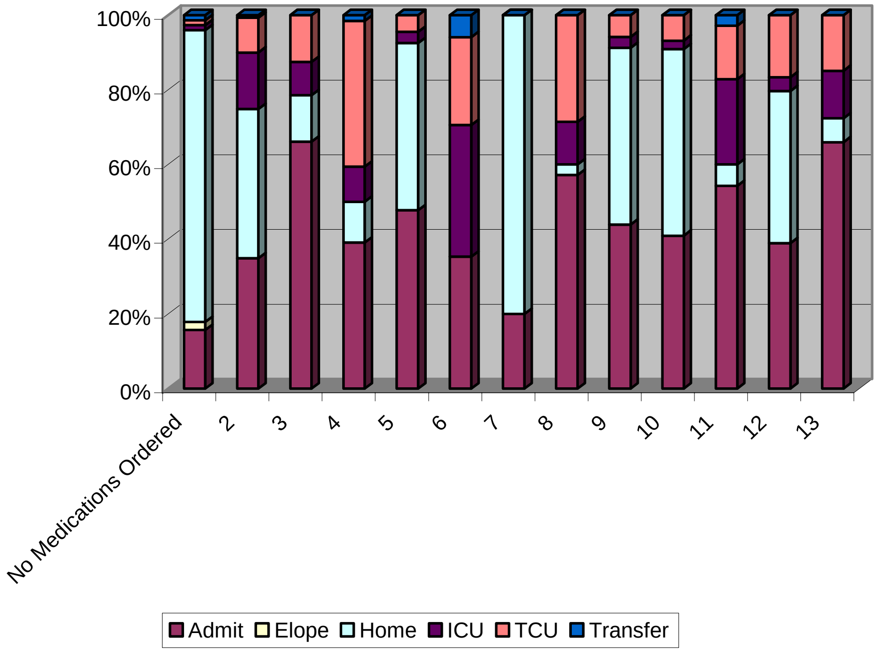
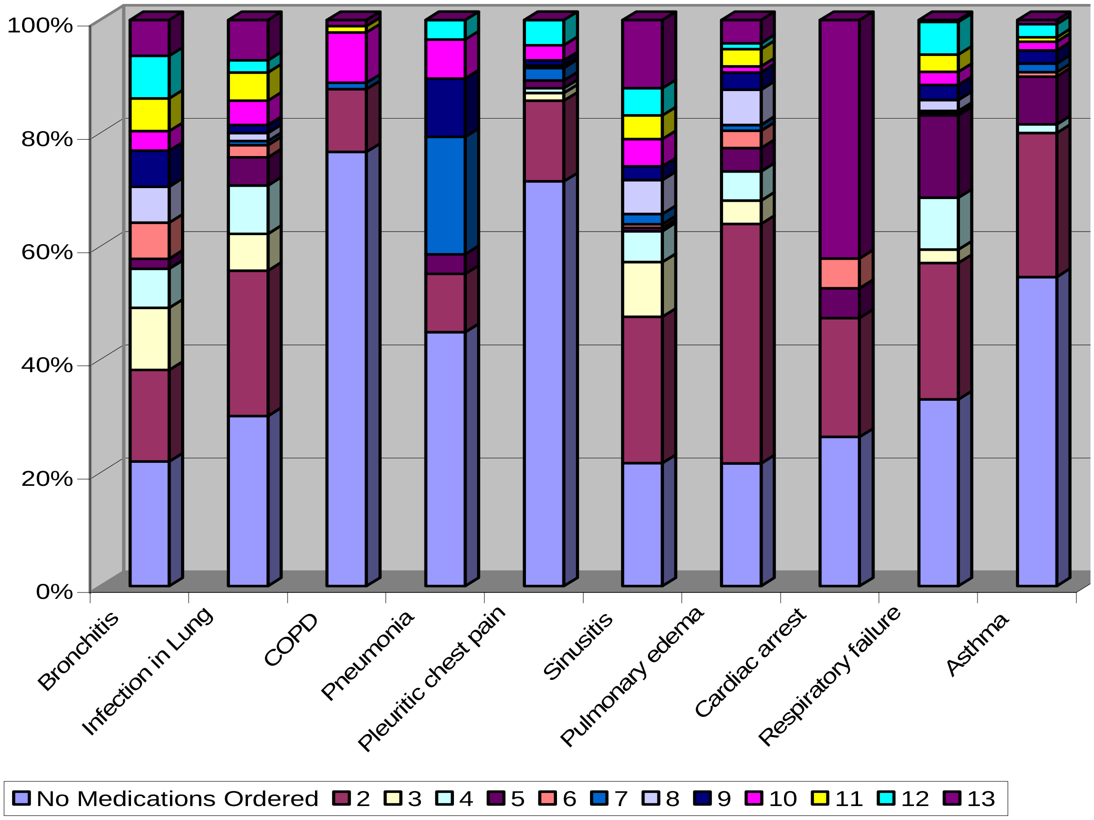


Figure 19. Comparison of Medication Cluster and Final Diagnosis



Of interest is that the diagnoses with the highest proportion of no medications ordered are COPD and Asthma.

Contrast this with the Bronchitis diagnosis with medications from just about every cluster. Upon investigation, it was discovered that patients in those categories received medications from a source separate from the pharmacy. This result indicates that electronic documentation may be incomplete because of a lack of integration in the record system.

To determine the relationship between complaints, diagnoses, and medications and the patient length of stay in the ED, an analysis of variance was performed. Interactions were included in the model. While initial complaint clusters were not statistically significant as primary effects in the model, final diagnosis was statistically significant (p=0.0021) as was medication cluster (p<0.0001). In addition, the interaction effects of initial complaint with medication (p=0.0031) and final diagnosis with medication cluster (p=0.0485) were also statistically significant. Restricting attention to Pneumonia, a link analysis of charges is given in Figure 20. When the keyword, "Sacrum" is used with the link analysis, the left-hand side of the link analysis is highlighted (Figures 21,22). However, when the keyword "lumbar" is used, the right-hand side of the link analysis is highlighted (Figure 23).

Figure 20. Link Analysis of Charges

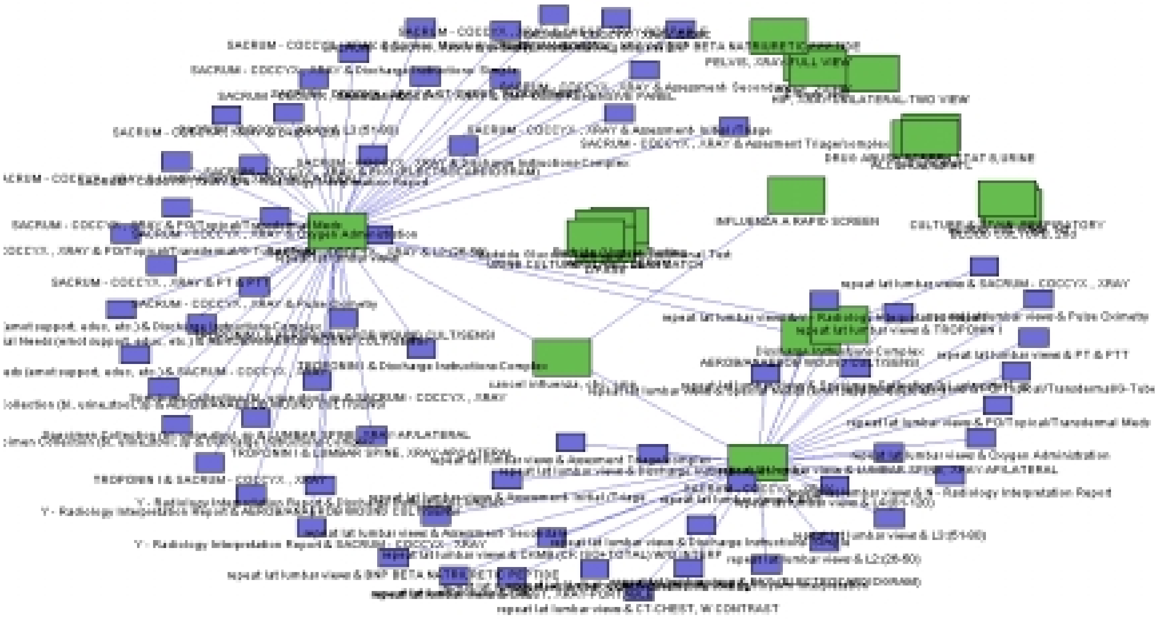
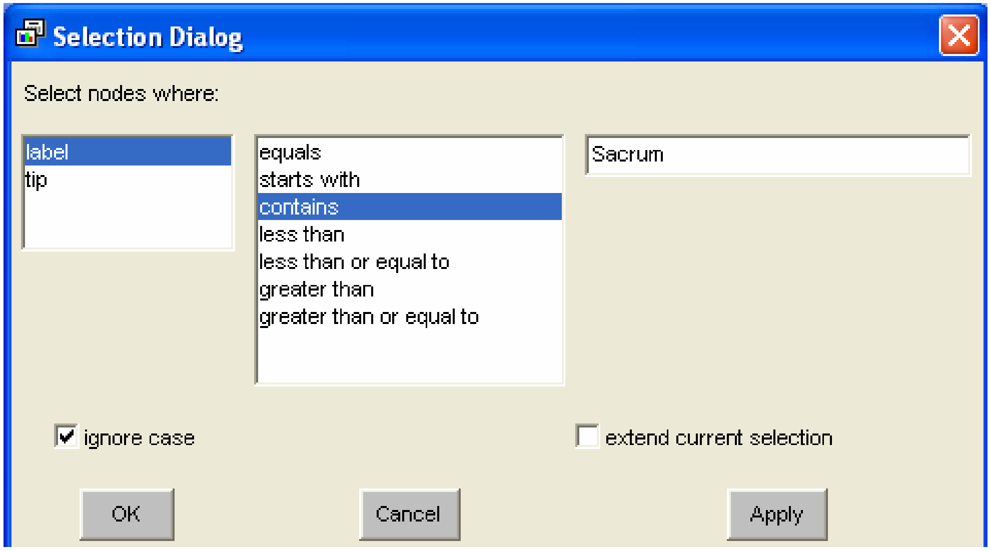


Figure 21. Keyword Selection for Link Analysis


Selection Dialog

Select nodes where:

label

equals
starts with
contains
less than
less than or equal to
greater than
greater than or equal to

Sacrum

☑ ignore case

☐ extend current selection

Figure 22. Selection of Sacrum-Lumbar X-Ray for Pneumonia

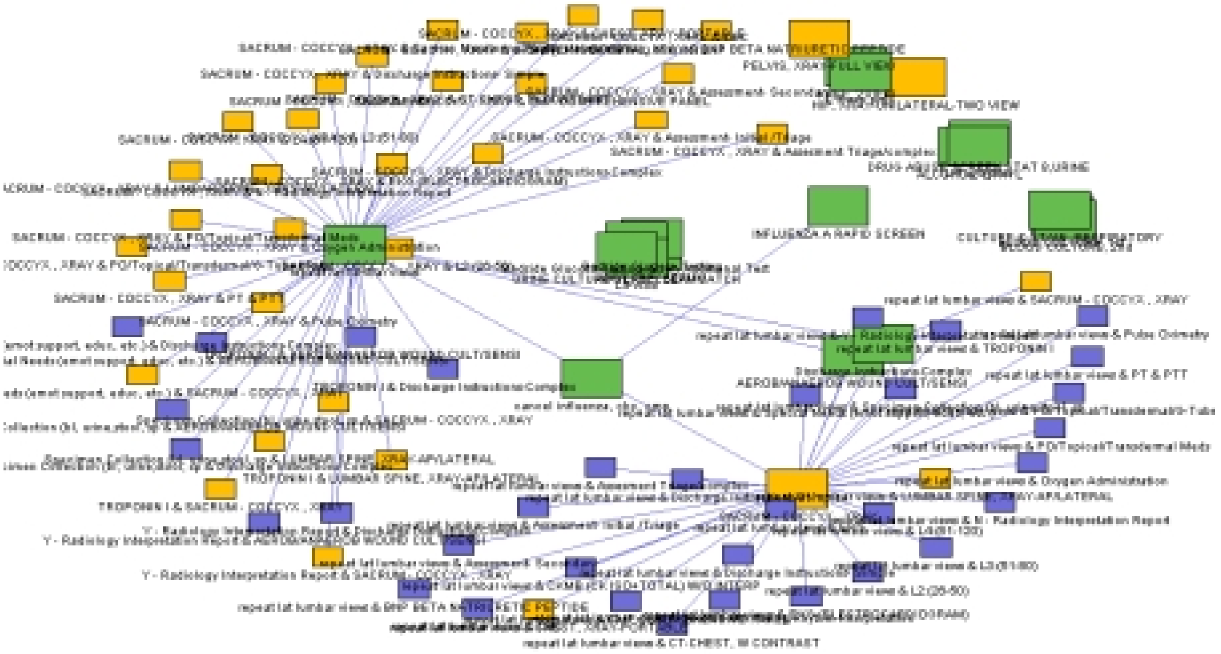
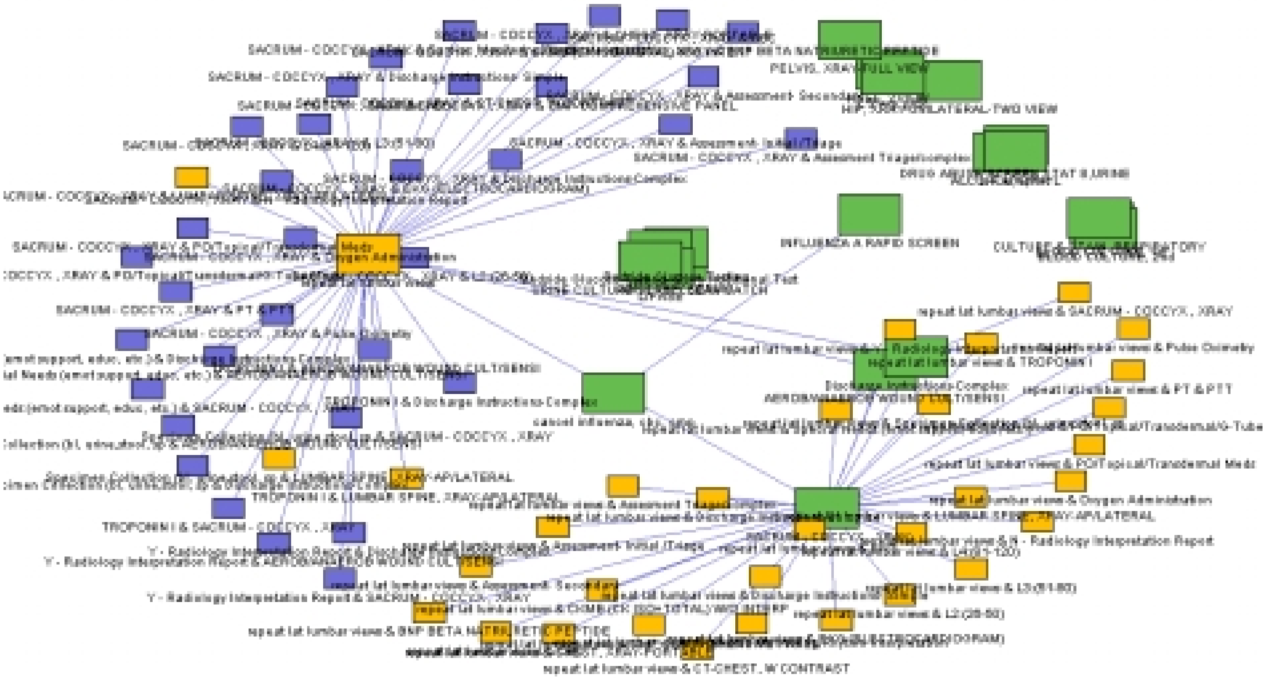

PELVIS, XRAY FULL VIEW
HIP, XRAY UNILATERAL-TWO VIEW
INFLUENZA A RAPID SCREEN
TROPONIN I & LUMBAR SPINE, XRAY AP/LATERAL
repeat lat lumbar views & CT-CHEST, W CONTRAST

Figure 23. Selection of Lumbar X-Ray for Pneumonia


PELVIS, XRAY FULL VIEW
HIP, XRAY UNILATERAL-TWO VIEW
INFLUENZA A RAPID SCREEN
TROPONIN I & LUMBAR SPINE, XRAY AP/LATERAL
repeat lat lumbar views & CT-CHEST, W CONTRAST

Here then is a real difference in treatment discovered through link analysis and association rules. The Sacrum-lumbar x-ray covers a larger portion of the lung, the lumbar x-ray tends to cover the lower lung, where pneumonia occurs. Figure 24 shifts the focus from the default of confidence to transaction counts. Similar to the result in Figure 11, specimen collection is connected to just about any treatment. There is considerable crowding of the diagram in the upper right. Using the Viewport, those nodes can be separated (Figure 25).

Figure 24. Pneumonia Cluster With Transaction Counts

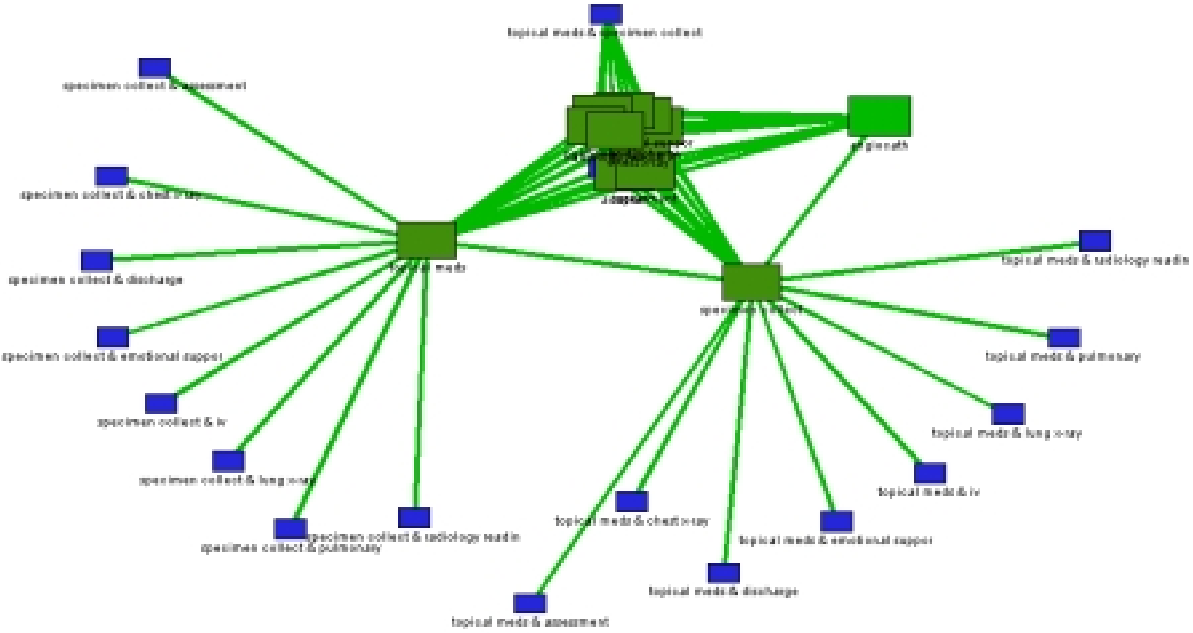
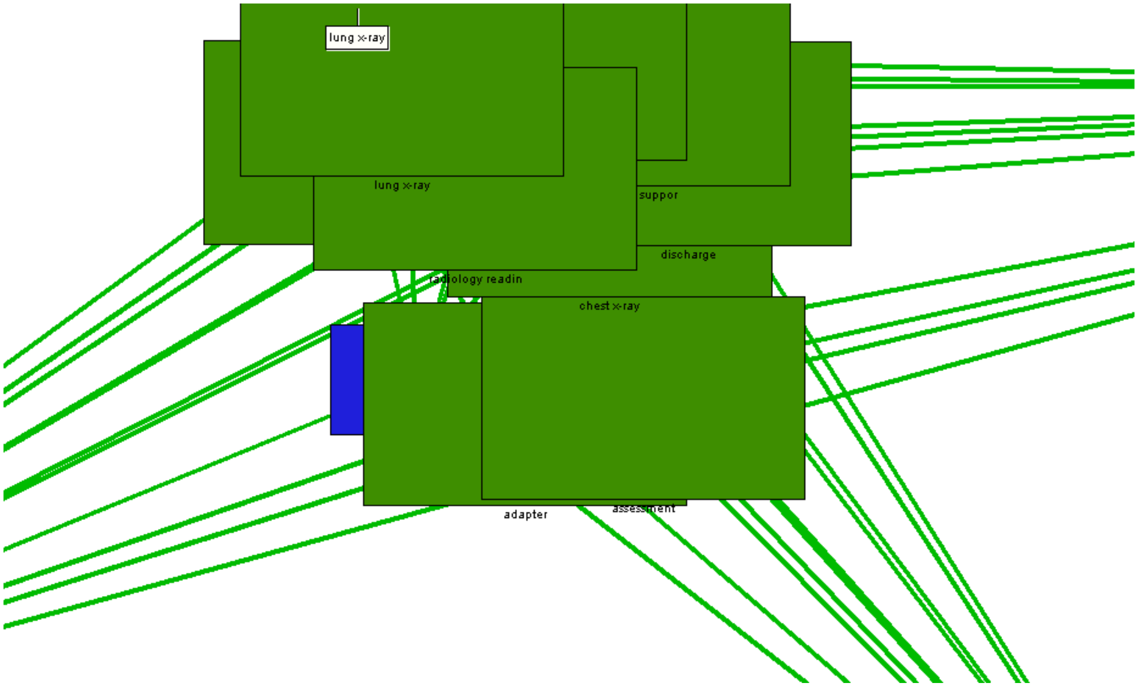
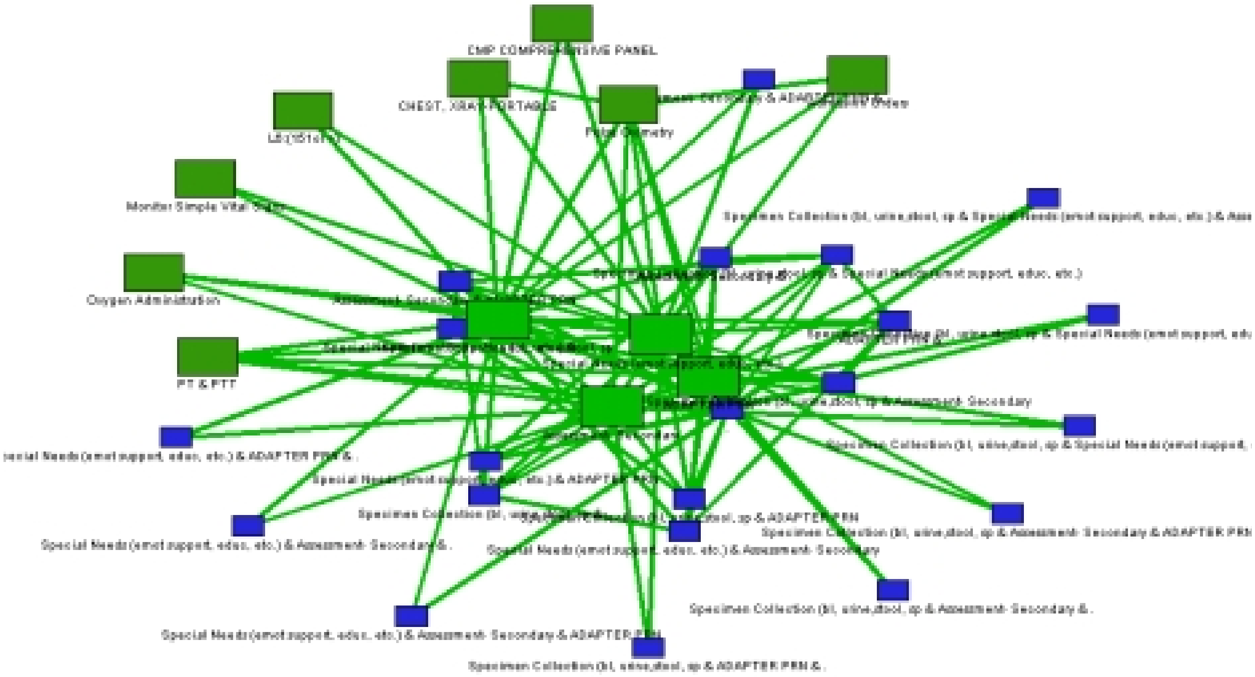

topical meds & specimen collect
specimen collect & assessment
specimen collect & chest x-ray
specimen collect & discharge
specimen collect & emotional suppor
specimen collect & iv
specimen collect & lung x-ray
specimen collect & pulmonary
specimen collect & radiology readin
topical meds
specimen collect
topical meds & radiology readin
topical meds & pulmonary
topical meds & lung x-ray
topical meds & iv
topical meds & chest x-ray
topical meds & emotional suppor
topical meds & discharge
topical meds & assessment

Figure 25. Viewport of Figure 24.



Again, there seems to be a division in physician orders in that some call for a lung x-ray and others for a chest x-ray. A lung x-ray is more tightly focused on the lung compared to a chest x-ray. The question is, which view is optimal for a suspected case of pneumonia? Figure 25 gives the link analysis with Patient as the ID rather than physician. Treatment from a patient perspective also includes oxygen in addition to specimen collection and special needs.

Figure 25. Link Analysis of Transaction Counts with Patient as ID.


CMP COMPREHENSIVE PANEL
CHEST, XRAY PORTABLE
Pulse Oximetry
Monitor Simple Vital Signs
Oxygen Administration
PT & PTT
Specimen Collection (bl, urine,stool, sp & Special Needs (emot support, educ., etc.) & Asse
Special Needs (emot support, educ. etc.) & ADAPTER PRN &.
Special Needs (emot support, educ., etc.) & ADAPTER PRN
Specimen Collection (bl, urine,stool, sp & ADAPTER PRN
Special Needs (emot support, educ. etc.) & Assessment- Secondary &.
Special Needs (emot support, educ. etc.) & Assessment- Secondary
Specimen Collection (bl, urine,stool, sp & Assessment- Secondary & ADAPTER PRN
Specimen Collection (bl, urine,stool, sp & Assessment- Secondary &.
Special Needs (emot support, educ. etc.) & Assessment- Secondary & ADAPTER PRN
Specimen Collection (bl, urine,stool, sp & ADAPTER PRN &.
Specimen Collection (bl, urine,stool, sp & Special Needs (emot support,
Specimen Collection (bl, urine,stool, sp & Assessment- Secondary

TRANSACTIONAL TIME SERIES

Another aspect of data mining the emergency department database is the existence of time stamps. These stamps can be used to examine the availability of personnel to treat patients in the ED. Traditionally, scheduling of personnel is done without an examination of arrivals and treatment times as gathered from patient records[1] The Emergency Department (ED) can use its collected information to create a model of patient arrivals and treatment times. Patients have the option of arriving in a random pattern. However, time series methods can be used to estimate the pattern of arrivals. In the past, because of the difficulty in collecting arrival data, only broad patterns could be discerned on a weekly or daily basis, but not on an hourly basis. In addition, only the number of visits was tracked without regard to the length of time each patient spent in the ED. The introduction of an electronic medical record in the ED includes the automatic collection of times of patient entry into and exit from the ED. The data can be used in time series models that automatically optimize the forecast.

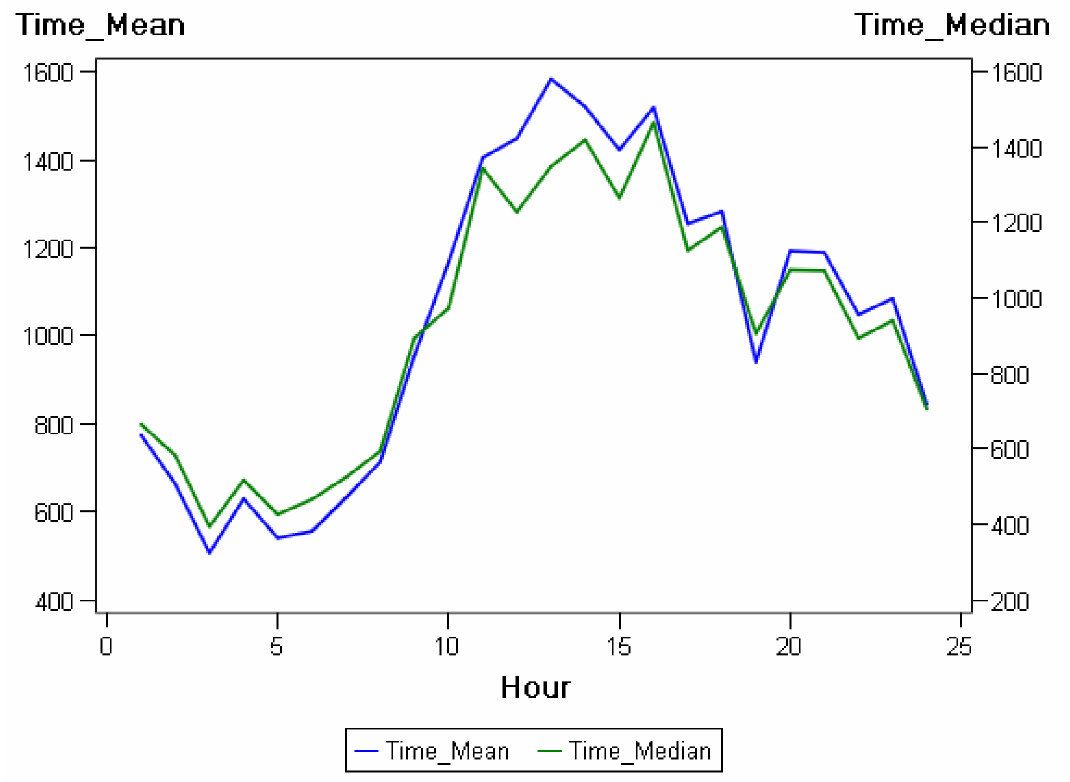
Usually, modeling of time series or econometric data is done in three parts: econometric modeling, simulation, and forecasting, primarily sequentially. Time series models have been used to predict the stock market, customer arrival for service, and customer waiting time. The Hospital Emergency Department (ED) is in a similar position with patients arriving at random times with complaints that require differing amounts of time for treatment, and who often need to queue up before service is available. While ED arrivals have occasionally been modeled using time series, it has been done on a limited basis of investigating the number of visits per day rather than on the timing of the visit or the length of time spent in the ED. It is the purpose of this paper to demonstrate how both time and length can be investigated using transactional time series models.

There are two important ways in which this information can be important to healthcare management. First, personnel can be scheduled in accordance with peak demand in the ED. If patient arrivals tend to increase at a specific time of the day and remain high until decreasing later in the day, additional personnel can be regularly scheduled to handle this demand. Second, regular demand can be predicted days, weeks, or months in advance. The differential between predicted and actual demand can be monitored. If a trend starts to develop for increased demand beyond that initially predicted, changes in personnel can be made for the length of the trend. In other words, management of demand can become proactive rather than retroactive.

The SAS PROC HPF can be used to examine time series data with random time periods. While the methodology presented here is given for a hospital emergency department, it can be applied to other issues in healthcare management. In particular, admissions, discharges, and patient holding time can be analyzed via time series with the goal of reducing the holding time.

Data from July-September, 2004 were examined. All patient visits were examined in that 3-month time interval, a total of 3300 visits. Data were accumulated by hour, and then averaged to determine change over a 24-hour period (Figure 26). The most noticeable spike in time occurs at 10:00 am when the average treatment time more than doubles. Attempts were made to explain the reason for the 10:00 spike in treatment time.

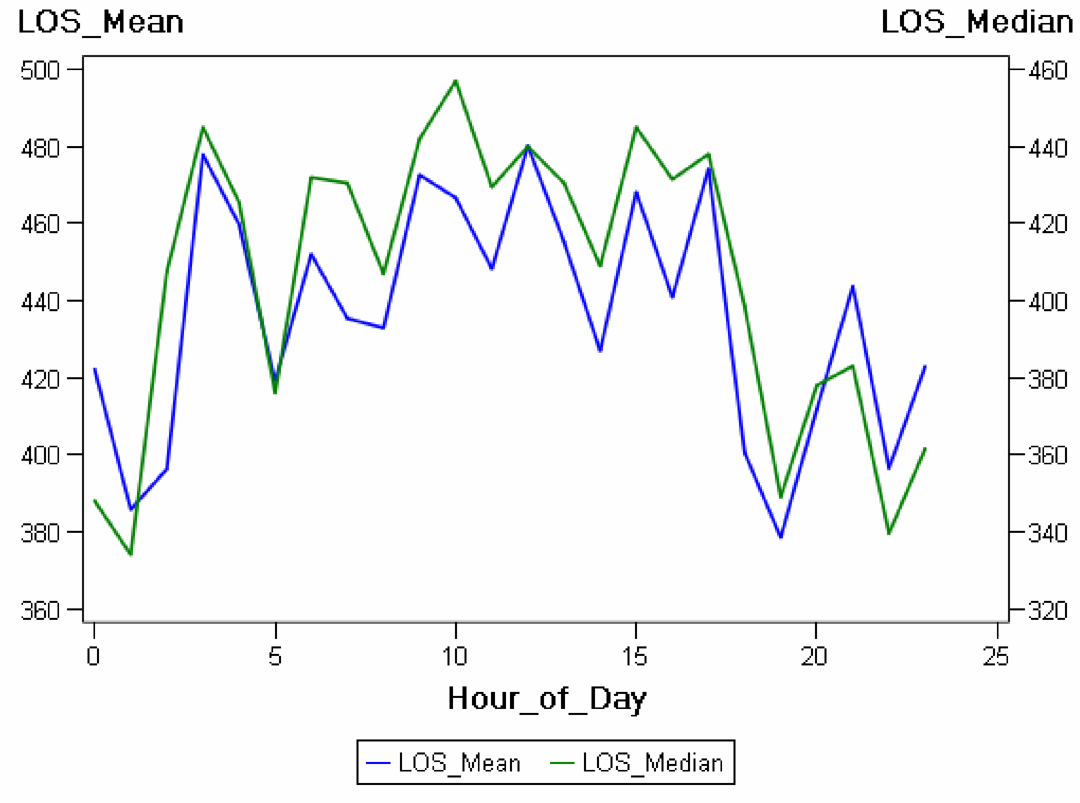
Figure 26. Average Treatment Time (LOS) Over 24 Hours (Military Time)



One reasonable explanation provided was that at that point in time, beds started becoming scarce and patients were waiting in the ED to be admitted to the hospital. If that were the case, patients admitted would have more time in the ED compared to patients discharged home.

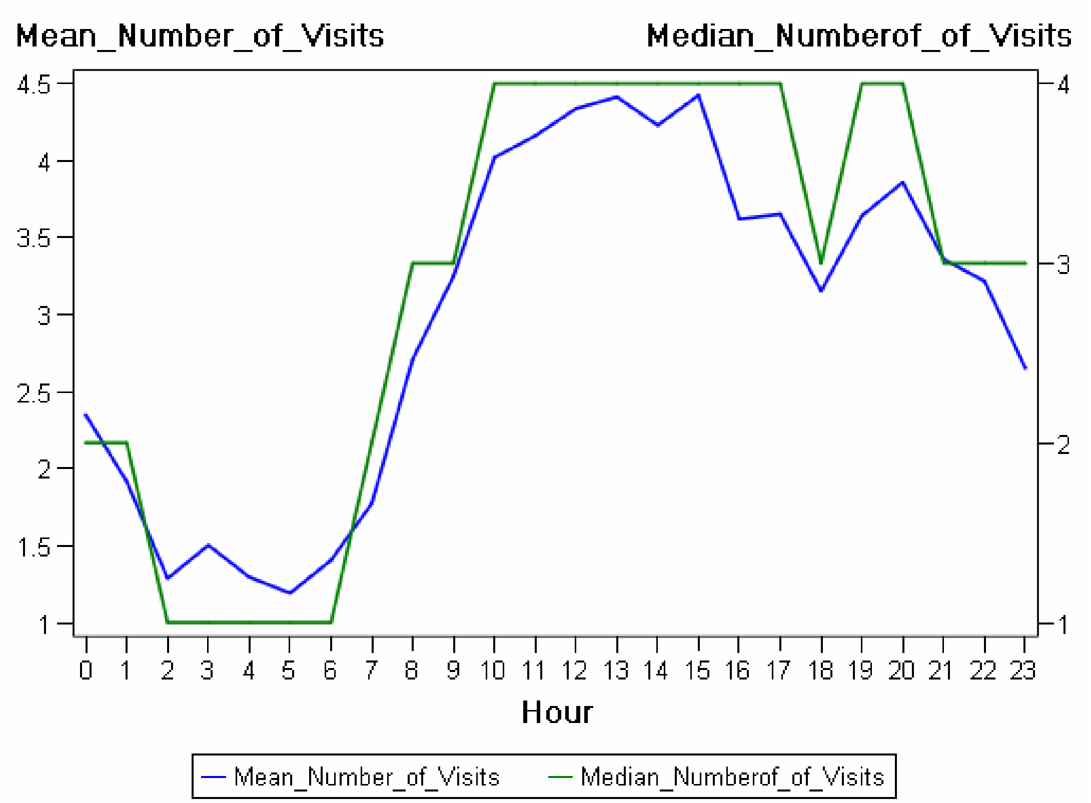
This was found not to be the case. To investigate that explanation, treatment times were examined by patient disposition (Figures 27). What becomes noticeable in that analysis is that there is no real spike in time for admitted patients.

Figure 27. Average Treatment Time (LOS) Over 24 Hours for Patients Admitted to the Hospital



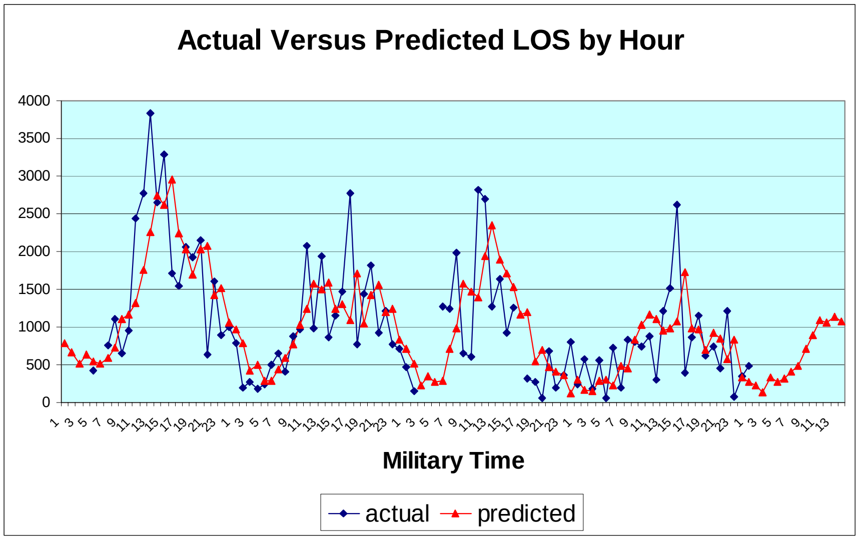
Another possible explanation given was that the increased service time resulted from an increase in the number of visits. This explanation has justification in the data. There was an increase in the number of visits starting at 10:00 am as shown in Figure 28. The number of patients starts to decline at hour 16, at which point the service time correspondingly decreases for patients discharged home. The relationship between number of visits and length of time for each visit is statistically significant with an r^2 of 74%.

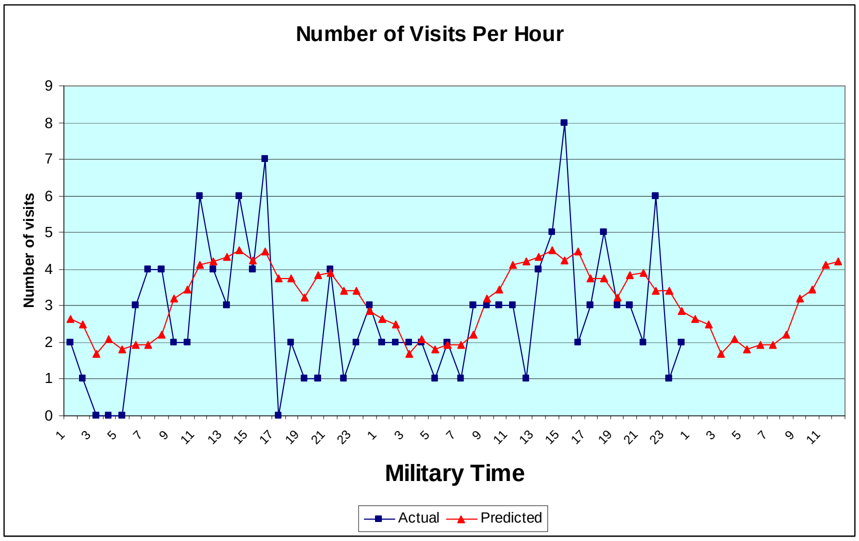
Figure 28. Average Number of Visits by Hour



The High Performance Forecasting System (HPF) was also used to determine whether the pattern in patient visits and length of stay is regular and predictable (Figure 29). The optimal forecast was seasonal with exponential smoothing. The seasonality indicates that the peak usage from 10:00 am through 4:00 pm is fairly regular, and should be considered when scheduling personnel.

Figure 29. High Performance Forecasting Results





Although there is some randomness in the arrival of patients to the Emergency Department, the randomness can be tracked to determine the existence of a regular pattern. In this particular ED, the pattern is seasonal (with seasonality defined as one 24-hour period). The periodicity can be taken into consideration in determining scheduling of personnel into the ED. Similar analyses can be performed to track demand for hospital beds, lab services, and other hospital units to optimize the use of these facilities and to reduce patient waiting time. The techniques can also be used in a physician’s office to improve scheduling and reduce waiting time.

CONCLUSION

SAS Enterprise Miner and SAS Text Miner can be used to explore the electronic medical record to determine differences in physician decision-making. Once the differences are found, they can be presented to the physician decision-makers so that treatments can be optimized, protocols can be developed, and patient care can be enhanced. Still other techniques, such as HPF, can be used to investigate other aspects of patient care. The improved documentation in the electronic medical records needs to be examined to improve patient care and to decrease costs.

CONTACT INFORMATION

Patricia Cerrito
University of Louisville
Department of Mathematics
Louisville, KY 40292
502-852-6826
502-852-7132 (fax)
pcerrito@louisville.edu