Paper 1248-2017

USING SAS® TO ANALYZE EMERGENCY DEPARTMENT VISITS: MEDICAID PATIENTS COMPARED TO OTHER PAY SOURCES

Karen Guthrie, Brad Casselman, Jennifer Grice - Healthcare Business Solutions, LLC
Denise McManus, Taylor Larkin - The University of Alabama

Access to care for Medicaid beneficiaries is a topic of frequent study and debate. Section 1202 of the Affordable Care Act (ACA) requires states to raise Medicaid primary care payment rates to Medicare levels in 2013 and 2014. The federal government paid 100% of the increase. This program was designed to encourage primary care providers to participate in Medicaid, since this has long been a challenge for Medicaid. Whether this fee increase has increased access to primary care providers is still debated. Using SAS®, we evaluated whether Medicaid patients have a higher incidence of non-urgent visits to local emergency departments (ED) than do patients with other payment sources. The National Hospital Ambulatory Medical Care Survey (NHAMCS) data set, obtained from the Centers for Disease Control (CDC, 2015), was selected, since it contains data relating to hospital emergency departments. This emergency room data, for years 2003–2011, was analyzed by diagnosis, expected payment method, reason for the visit, region, and year. To evaluate whether the ED visits were considered urgent or non-urgent, we used the NYU Billings algorithm for classifying ED utilization (NYU-Wagner 2015). Three models were used for the analyses: Binary Classification, Multi-Classification, and Regression. In addition to finding no regional differences, decision trees and SAS® Visual Analytics revealed that Medicaid patients do not have a higher rate of non-emergent visits when compared to other payment types.

INTRODUCTION

The American College of Emergency Room Physicians (ACEP) cites that emergency care is the safety net of the nation's healthcare system (2014). Health care safety net providers are those that have a legal mandate and mission to offer medical care to all patients, regardless of their ability to pay, and have a substantial number of patients who are uninsured or on Medicaid (2015). Americans without insurance generally lack a regular source of medical care, and suffer from medical conditions that are either preventable or easily treated in the primary care setting. Consequently, the uninsured are four times more likely than the insured to forgo or postpone needed preventive care and three times more likely to skip recommended tests or treatments (Kaiser Family Foundation, 2000). Emergency departments have become a vital source of care for those without insurance who generally lack a source of primary care, since they are legally required to assess and treat patients regardless of their ability to provide payment. As a result, although care delivered through the emergency department is frequently for non-urgent problems, it is substantially more costly than comparable care delivered in other settings (Barnezai, Melnich, Nawathe, 2005). Additionally having a regular source of primary care provides continuity of care increases the likelihood of patients' receiving preventive screening, which are absent in the emergency department setting (Zheng, O'Leary, Sloss, Lopez, Dhanani & Melnick, 2006).

Low Medicaid payment rates are often cited as the reason for low provider participation, which consequently reduces access to care for beneficiaries (Paradise, 2015). The purpose of this study is to determine the historical incidence of non-urgent emergency department visits and consider whether the rate of these visits could potentially be reduced as more Medicaid patients have access to primary care providers. The researchers assumed, with consideration to the timeframe of the available data, that Medicaid beneficiaries had limited access to primary care physicians. The approach included analysis of public use data files obtained from the Centers for Disease Control, specifically data sets collected through the National Hospital Ambulatory Medical Care Survey (2015). Primary attributes for consideration include the diagnosis, expected payment method, reason for the visit, region, and calendar year. Diagnosis is defined by the associated ICD-9-CM code diagnosis data element, which will be analyzed using the Billings Algorithm (NYU-Wagner, 2015) for categorizing the severity of hospital emergency department visits. Therefore, three different modeling approaches were used for analysis: Binary Classification, Multi-Classification, and Regression, to determine if Medicaid patients have a higher incidence of non-urgent visits to local emergency rooms than do patients with other payment sources.

DATA ACQUISITION AND PREPARATION

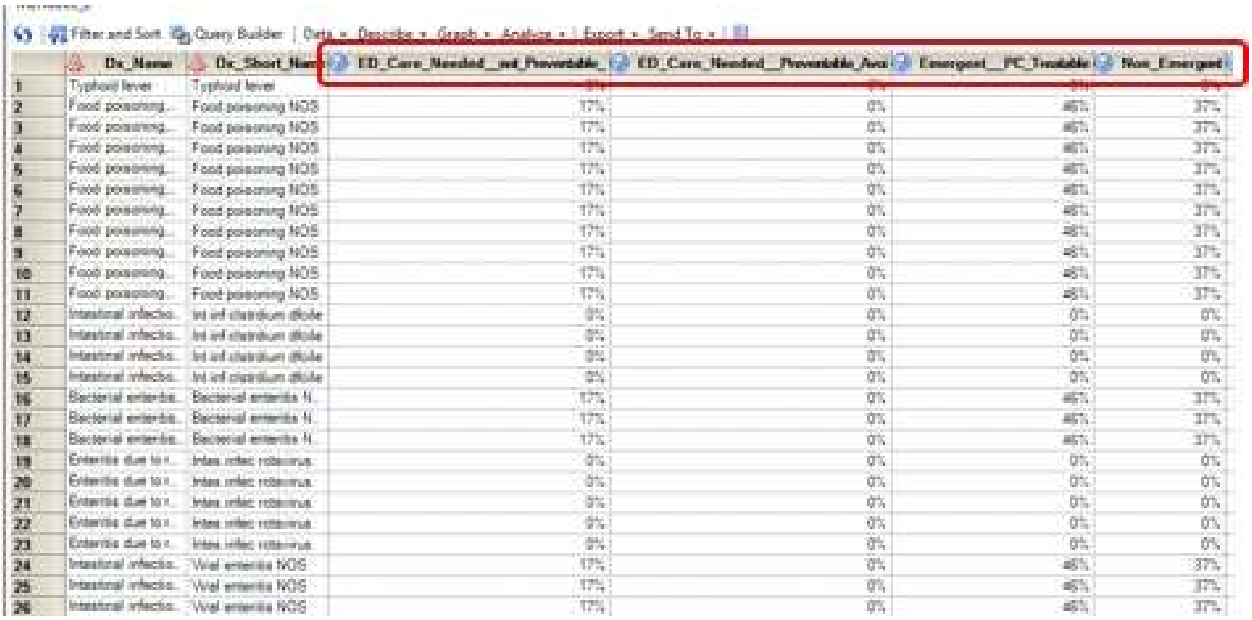
Data for this project was collected from two sources. First, data was downloaded yearly from the Centers for Disease Control (CDC) National Hospital Ambulatory Medical Care Survey (NHAMCS) (CDC,2015). The measures for emergent and non-emergent care as well as categorical data such as region where the care was provided, payment methods, and diagnosis information are included. In order to further describe the diagnosis codes, a downloadable ICD-9 Diagnosis Code list was accessed via the NYU-Wagner website (NYU Wagner, 2015) and merged into the data.

Data from the NHAMCS is annually produced by calendar year and are public-use data files. For this study, data was retrieved for the period of 2003 to the most recent year available which at the time was 2011. A subset of 60+ variables was selected from the downloaded data and prepared for analysis. The additional variables included such information as causal factors, wait times, if the patient was admitted to a facility, pain scale and other variables. An Inner Join was used to merge in the Diagnosis descriptions from the NYU-Wagner data and then the yearly data was concatenated together via PROC SQL utilizing the CORR option to force data with the same columns to append:

```
PROC SQL;
  CREATE TABLE WORK.APPEND_TABLE_0000 AS
  SELECT * FROM WORK.FILTER_FOR_2003_FINAL_DATA_SAS7B
    OUTER UNION CORR
  SELECT * FROM WORK.FILTER_FOR_2004_FINAL_DATA_SAS7B
    OUTER UNION CORR
  SELECT * FROM WORK.FILTER_FOR_2005_FINAL_DATA_SAS7B
    OUTER UNION CORR
  SELECT * FROM WORK.FILTER_FOR_2006_FINAL_DATA_SAS7B
    OUTER UNION CORR
  SELECT * FROM WORK.FILTER_FOR_2007_FINAL_DATA_SAS7B
    OUTER UNION CORR
  SELECT * FROM WORK.FILTER_FOR_2008_FINAL_DATA_SAS7B
    OUTER UNION CORR
  SELECT * FROM WORK.FILTER_FOR_2009_FINAL_DATA_SAS7B
    OUTER UNION CORR
  SELECT * FROM WORK.FILTER_FOR_2010_FINAL_DATA_SAS7B
    OUTER UNION CORR
  SELECT * FROM WORK.FILTER_FOR_2011_FINAL_DATA_SAS7B;
quit;
```

DEFINING EMERGENT STATUS

After merging the ICD-9 codes with the Medicaid data set, each observation contained four different variables that help assess the degree of which cases were emergent. Using these percentages, we created a new variable called “Emergent_Status” that defines a case as Emergent, non-Emergent or indeterminable based on the probabilities defined by the ICD-9 codes. The new variable selected Emergent if emergency department is needed and not preventable with a probability above 50%. The new variable selects Non-Emergent if the visits have a probability of needing emergency department care and preventability is below 50%. This will serve as our target variable as illustrated in Figure 1. This is in agreement with the NYU data study (Figure 2) on classification of emergent care.



	Dx_Name	Dx_Short_Nam	ED_Care_Needed__not_Preventable_	ED_Care_Needed__Preventable_Avoi	Emergent__PC_Treatable	Non_Emergent
1	Typhoid fever	Typhoid fever	0%	0%	0%	0%
2	Food poisoning...	Food poisoning NOS	17%	0%	46%	37%
3	Food poisoning...	Food poisoning NOS	17%	0%	46%	37%
4	Food poisoning...	Food poisoning NOS	17%	0%	46%	37%
5	Food poisoning...	Food poisoning NOS	17%	0%	46%	37%
6	Food poisoning...	Food poisoning NOS	17%	0%	46%	37%
7	Food poisoning...	Food poisoning NOS	17%	0%	46%	37%
8	Food poisoning...	Food poisoning NOS	17%	0%	46%	37%
9	Food poisoning...	Food poisoning NOS	17%	0%	46%	37%
10	Food poisoning...	Food poisoning NOS	17%	0%	46%	37%
11	Food poisoning...	Food poisoning NOS	17%	0%	46%	37%
12	Intestinal infectio...	Int inf clstrdium dfcile	0%	0%	0%	0%
13	Intestinal infectio...	Int inf clstrdium dfcile	0%	0%	0%	0%
14	Intestinal infectio...	Int inf clstrdium dfcile	0%	0%	0%	0%
15	Intestinal infectio...	Int inf clstrdium dfcile	0%	0%	0%	0%
16	Bacterial enteritis...	Bacterial enteritis N...	17%	0%	46%	37%
17	Bacterial enteritis...	Bacterial enteritis N...	17%	0%	46%	37%
18	Bacterial enteritis...	Bacterial enteritis N...	17%	0%	46%	37%
19	Enteritis due to r...	Intes infec rotavirus	0%	0%	0%	0%
20	Enteritis due to r...	Intes infec rotavirus	0%	0%	0%	0%
21	Enteritis due to r...	Intes infec rotavirus	0%	0%	0%	0%
22	Enteritis due to r...	Intes infec rotavirus	0%	0%	0%	0%
23	Enteritis due to r...	Intes infec rotavirus	0%	0%	0%	0%
24	Intestinal infectio...	Viral enteritis NOS	17%	0%	46%	37%
25	Intestinal infectio...	Viral enteritis NOS	17%	0%	46%	37%
26	Intestinal infectio...	Viral enteritis NOS	17%	0%	46%	37%

Figure 1: Target Variable Creation

```
data work.finalmedicaid;
set 'x:\Medicaid\finalmergeddata_nov18';
IF ED_Care_Needed__not_Preventable_ > .5 then Emergent_Status=1;
  else IF ED_Care_Needed__Preventable_Avoi>.5 then Emergent_Status=1;
  else IF sum(ED_Care_Needed__not_Preventable_,ED_Care_Needed
Preventable_Avoi) > sum(Emergent PC_treatable,Non_Emergent) then
Emergent_Status=1;
  else if Unclassified >.5 then Emergent_Status=1;
  else Emergent_Status=0;
run;
```

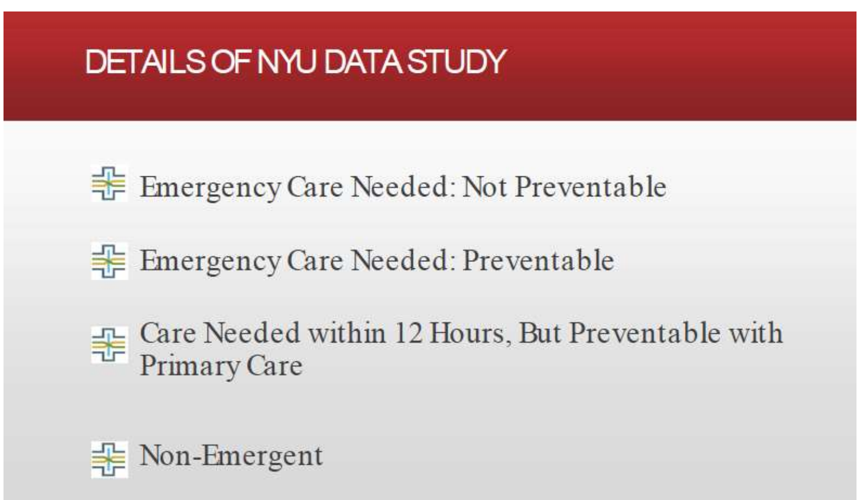


Figure 2: NYU Study Classification System

INITIAL HYPOTHESIS CONCLUSIONS

Once the final data set was completed and combined with the metrics of emergent vs non- emergent statuses, we investigate the hypothesis statement again: Were there higher instances of Medicaid patients coming to the Emergency Room with non-emergent or preventable illnesses than other pay types? The following code was used to answer this question and the generated output is displayed in Table 1.

```
proc freq data="x:\xxxxx";
  table PAYTYPE*Emergent_Status
run;
```

What is important and notable in this table is the first and third row within each paytype. The first row gives a total frequency for each emergent status. The third row shows what percentage of that paytype is non-emergent (0) and emergent (1). Figure 3 illustrates some example from the proc freq.

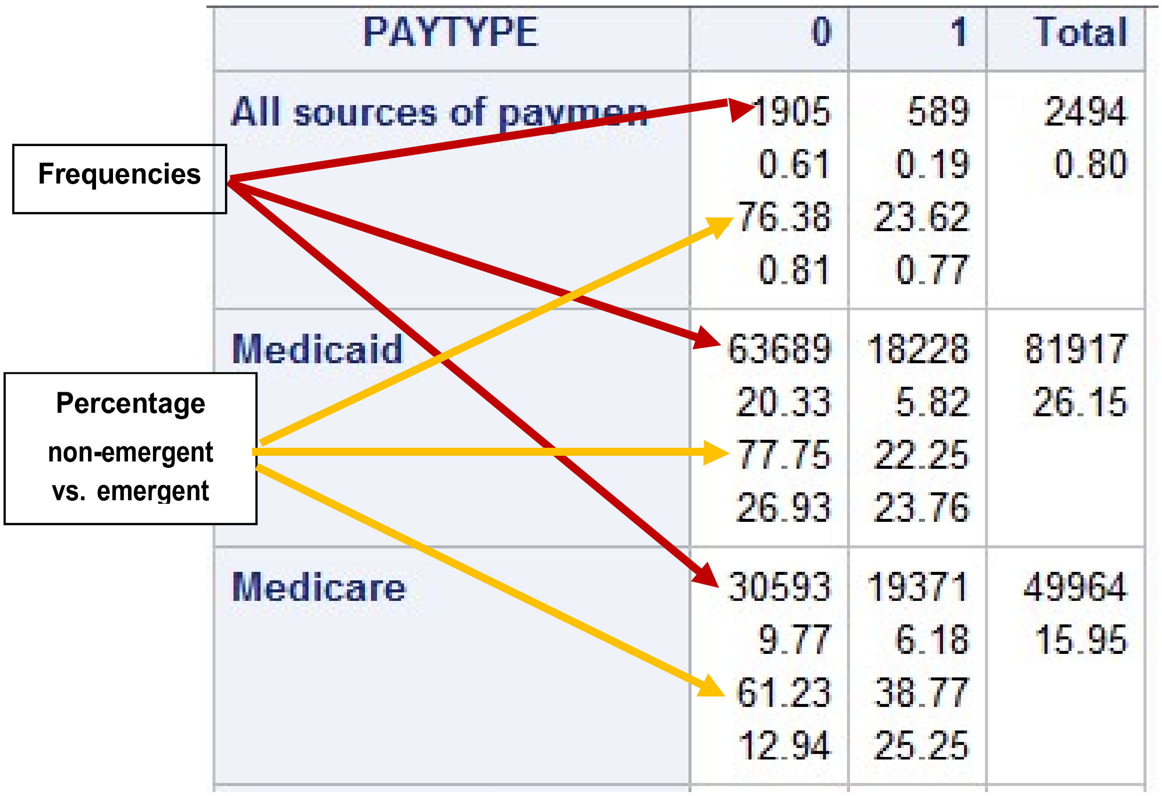
The SAS System

The FREQ Procedure

Table of PAYTYPE by Emergent_Status

Frequency Percent Row Pct Col Pct	Emergent_Status		
PAYTYPE	0	1	Total
All sources of paymen	1905 0.61 76.38 0.81	589 0.19 23.62 0.77	2494 0.80
Medicaid	63689 20.33 77.75 26.93	18228 5.82 22.25 23.76	81917 26.15
Medicare	30593 9.77 61.23 12.94	19371 6.18 38.77 25.25	49964 15.95
No charge	1334 0.43 78.56 0.56	364 0.12 21.44 0.47	1698 0.54
No charge/Charity	1087 0.35 74.91 0.46	364 0.12 25.09 0.47	1451 0.46
Other	6000 1.92 78.08 2.54	1684 0.54 21.92 2.20	7684 2.45
Private insurance	78774 25.15 77.07 33.31	23441 7.48 22.93 30.56	102215 32.63
Self-pay	36504 11.65 80.55 15.43	8817 2.81 19.45 11.49	45321 14.47
Unknown	12513 3.99 78.05 5.29	3520 1.12 21.95 4.59	16033 5.12
Worker's compensation	4113 1.31 92.45 1.74	336 0.11 7.55 0.44	4449 1.42
Total	236612 75.51	76714 24.49	313226 100.00

Table 1: PROC FREQ output of PAYTYPE by Emergent Status



PAYTYPE	0	1	Total
All sources of paymen	1905 0.61 76.38 0.81	589 0.19 23.62 0.77	2494 0.80
Medicaid	63689 20.33 77.75 26.93	18228 5.82 22.25 23.76	81917 26.15
Medicare	30593 9.77 61.23 12.94	19371 6.18 38.77 25.25	49964 15.95

Figure 3: Expanded view of PROC FREQ

The emergent and non-emergent data was loaded in MS Excel (see Figure 3).

	A	B	C	D	E
1	PAYTYPE	% Non Emergent	% Emergent	Count Non Emergent	Count Emergent
2	All sources of paymen	76.38	23.62	1905	589
3	Medicaid	77.75	22.25	63689	18228
4	Medicare	61.23	38.77	30593	19371
5	No Charge	78.56	21.44	1334	364
6	No Charge/Charity	74.91	25.09	1087	364
7	Other	78.08	21.92	6000	1684
8	Private Insurance	77.07	22.93	78774	23441
9	Self-Pay	80.55	19.45	36504	8817
10	Unknown	78.05	21.95	12513	3520
11	Worker's Compensation	92.45	7.55	4113	336
12					

Figure 4: Emergent vs. Non-Emergent Care by Payment Type

As illustrated in Figure 4 every paytype had much higher percentages of non-emergent than emergent. When looking at Medicaid specifically, it seems quite consistent with the majority of paytypes. There seem to be only two paytypes that stray from the average distribution. Medicare patients have significantly higher percentage of emergency room visits with emergent conditions rather than non-emergent. On the other hand, worker's compensation patients seem to visit the emergency room much more often with non-emergent conditions.

When looking at frequencies alone, however, worker's compensation paytype is a significantly smaller sample size than some of the others, so it may not be as safe to assume the distribution. However, Medicaid, Medicare and Private Insurance have the largest sample sizes and can easily be compared. As noted above, Medicaid seems to show no discrepancy from Private insurance. Medicare is the only paytype of all the paytypes that shows lower percentages of non- emergent conditions in the emergency room.

The final piece of demographic exploration performed related to the four regions defined in the data set: West, Midwest, South, and Northeast. A frequency procedure was generated to determine the number of emergent and non-emergent Medicaid cases for each of the four regions. The following results were yielded:

West: 61,893

Midwest: 66,647

South: 108,393

Northeast: 76,292

Each region was broken down by emergent (green) and non-emergent (red) cases. Figure 5 illustrates the corresponding percentages which were constant across all regions:

Figure 5: PROC GMAP of Emergent vs. Non-Emergent Care by Region

DATA MODELING

Three different modeling approaches were used in assessing our hypothesis. Each modeling technique allowed even more advanced exploration of the data and provided insight into additional relationships across various payment types and regions. The approach was to target any contributing factors to the use of emergent care, especially within the South region and specifically, the State of Alabama. The team used classical regression techniques for modeling which included binary classification models as well as multi-classification models (Figure 6):

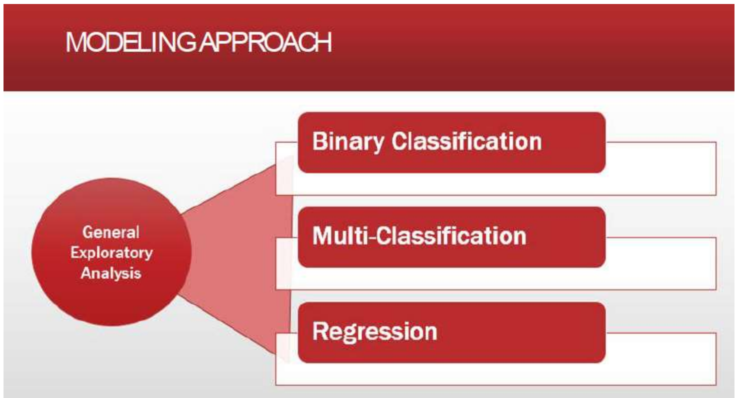


Figure 6: Modeling Approach

BINARY CLASSIFICATION/REGRESSION

First, the dataset was loaded into enterprise guide to a build a series of different models (regardless of pay type), so that we could determine the model that best fits this data for binary classification purposes. The following steps illustrate the process of starting a new project and performing binary classification modeling in Enterprise Miner. Our target dependent variable was the binary emergent status created earlier from the ICD-9 codes (Figure 7).

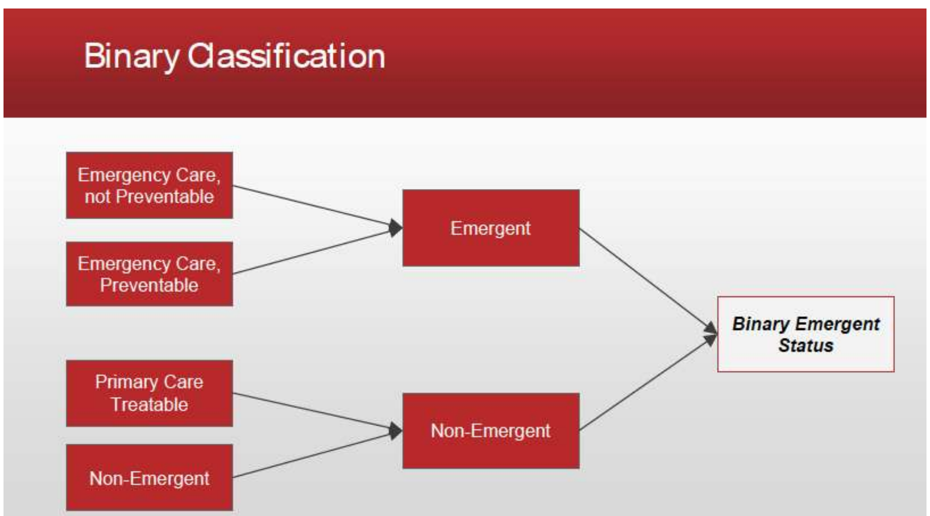


Figure 7: Binary Classification

To set up the binary classification models, the “Raw Data” Node was used in the diagram workspace. In the properties panel, under the “TRAIN” section click on the “… “beside the variables option. This will allow you to identify which variable to reject (such as out of scope triage details, etc.) and which to set as the target. The new emergent status variable is set to be the target and is classified as binary: (see Figures 8 and 9).

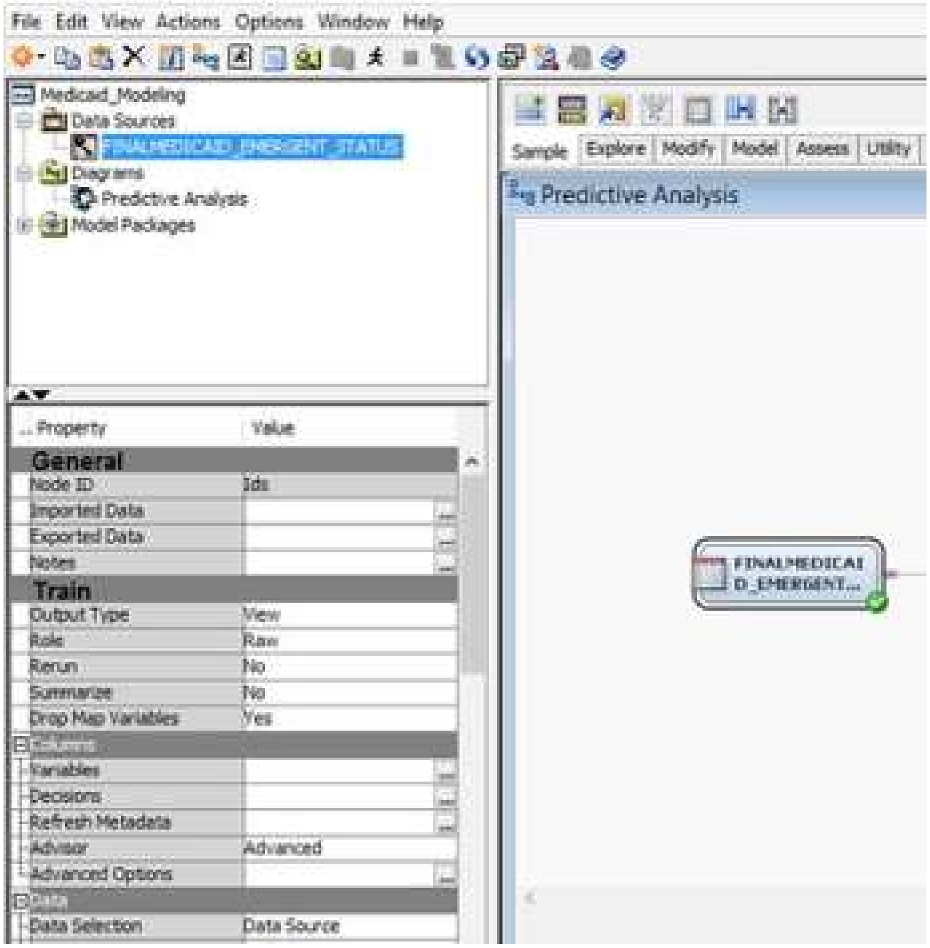


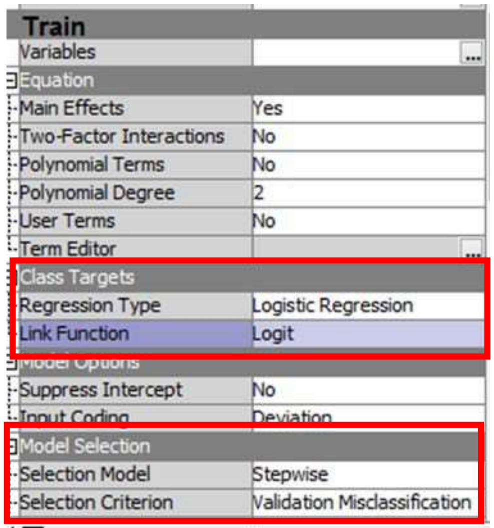
Figure 8: SAS Enterprise Miner Raw Data Node Setup

Name	Role	Level	Report	Order	Drop
DIAG3R	Rejected	Interval	No		No
DIAGSCRN	Input	Nominal	No		No
DOA	Input	Nominal	No		No
Drug	Input	Interval	No		No
Dx_Name	Text	Nominal	No		No
Dx_Short_Name	Input	Nominal	No		No
ED_Care_Neede	Input	Interval	No		No
ED_Care_Neede	Input	Interval	No		No
EKG	Input	Nominal	No		No
ELECTROL	Input	Nominal	No		No
EMT	Input	Nominal	No		No
ENDOINT	Input	Nominal	No		No
ETHNICPL	Input	Nominal	No		No
Emergent_Statu	Target	Binary	No		No
Emergent__PC_	Input	Interval	No		No

Figure 9: Variable Role Assignment

Then a Data partition node was added to divide the total data set into training and validation data sets for predictive modeling. A Replacement node was also added to convert arrival times for the Emergency Department visits to “Primary” during primary care hours, and to “After” for after primary care hours. For missing values, the mean was utilized as the imputation method (via the Impute node). Other preparatory nodes included utilizing a Stat Explore node, then a Transform node to perform log transforms on variables with positive skewness.

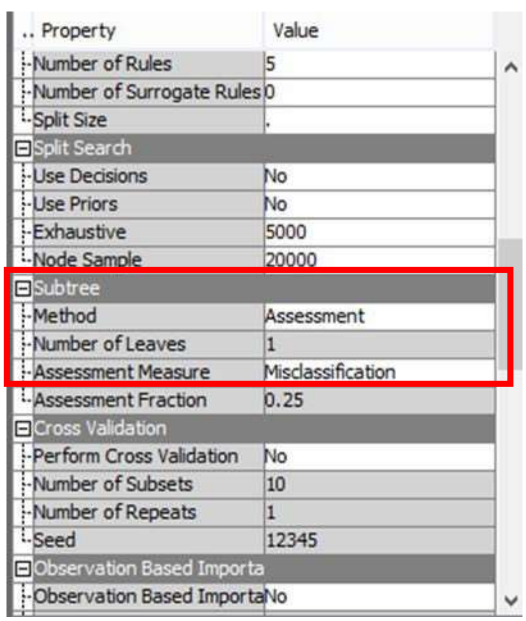
To begin modeling, a logistic regression was selected using Stepwise selection and Validation Misclassification as the selection criterion as shown in Figures 9 and 10:



Train	
Variables	...
Equation	
Main Effects	Yes
Two-Factor Interactions	No
Polynomial Terms	No
Polynomial Degree	2
User Terms	No
Term Editor	...
Class Targets	
Regression Type	Logistic Regression
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	Stepwise
Selection Criterion	Validation Misclassification

Figure 9: Regression Node Selections

A Neural Network node was also connected to the regression node, and a decision tree node was added to the initial data partitioning node (as it does not require imputation and transformations):



Property	Value
Number of Rules	5
Number of Surrogate Rules	0
Split Size	.
Split Search	
Use Decisions	No
Use Priors	No
Exhaustive	5000
Node Sample	20000
Subtree	
Method	Assessment
Number of Leaves	1
Assessment Measure	Misclassification
Assessment Fraction	0.25
Cross Validation	
Perform Cross Validation	No
Number of Subsets	10
Number of Repeats	1
Seed	12345
Observation Based Importa	
Observation Based Importa	No

Figure 10: Neural Network Node Selections

The Enterprise Miner linear regression model in Figure 11 illustrates the workflow.

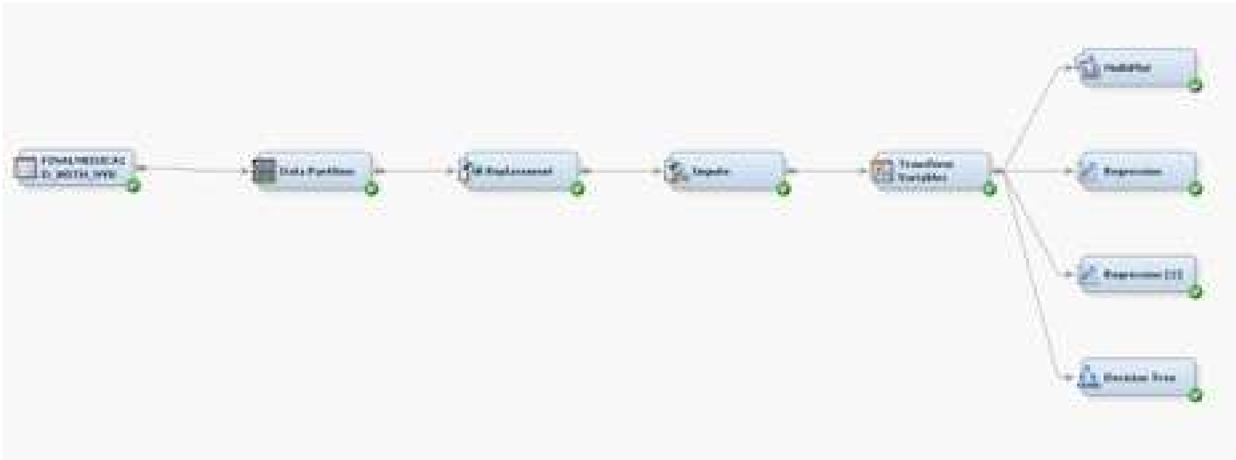


Figure 11: Enterprise Miner Regression Workflow

For this model, the variables were limited even further to information only known by an insurance provider or primary care physician (Table 2).

	Variable	Definition
1	ADMITHOS	Was the patient admitted to the hospital?
2	AGE	Patient age
3	AGER	Patient age (range)
4	DIAGSCRN	Was a diagnostic screen given?
5	EKG	Was an EKG used?
6	BPDIAS	Diastolic blood pressure
7	BPSYS	Systolic blood pressure
8	LOV	Length of visit
9	PULSE	Pulse
10	TEMPF	Temperature in Fahrenheit
11	WAITTIME	Wait time in emergency room
12	IV FLUIDS	Were IV fluids given?
13	MED	Medication administered
14	PATWT	Patient weight
15	PAYTYPE	Pay type
16	REGION	Region
17	SEX	Sex
18	VDAYR	Day of the month
19	VMONTH	Month of the year
20	YEAR	Year

Table 2: Listing of Variables for Regression

In the model, the data partition node was used to create a 50/50 split between training and validation data. The replacement node was used to remove outlying data that was determined to be present in the data set erroneously. The outlying data for the variables BPDIAS, BPSYS, NUMMED, TEMPF, and WAITTIME were replaced with missing values, to be corrected in the following imputation node. The imputation node was used to impute reasonable data for values that are missing for certain variables within the dataset. The missing data for the variables WAITTIME, BPSYS, LOV, TEMPF, BPDIAS, and PULSE were imputed with the median value for each of the respective variables. Finally, a transformation node was used to reduce the bias for variables with skewed data. LOV, BPDIAS, BPSYS, PULSE, and PATWT were put through a logarithmic transformation in order to rectify the data.

Once all of the data was organized and cleaned, six predictive models were created, three using an interval target and three using a binary target.

Using an interval target, a predictive model using regression was created. This model had 48 weights and a relatively low average squared error of 0.05. A predictive model using stepwise regression was also created, with comparable results (40 weights; average squared error of 0.05). Finally, a decision tree was created, which provided more useful analysis. For this decision tree, EKG was found to be the most influential variable, followed by IV FLUIDS and ADMITHOS.

Given the nature of the data and after doing predictive modeling using an interval target, it seems to be more useful to use a binary target. Using a binary target, the same three types of models were created. The model using regression was found to have 48 weights and a misclassification rate of 0.22. The model using stepwise regression was found to have 38 weights and a misclassification rate of 0.22. Once again, the decision tree seems to be the most useful for analysis. As with the interval target, the most influential variable was EKG. According to the model, 81% of patients that were administered an EKG were found to fall into the emergent category. The next two most influential variables in the binary target decision tree model were the same as with the interval target, except ADMITHOS was found to be slightly more influential than IV Fluids.

MULTI-CLASSIFICATION

Similar to the “Emergent Status” variable creation for binary classification, we created a new variable called Emergent_Level to assess the extent of emergency for each observation beyond just two outcomes. This allows us to get an even more accurate estimation of which cases are more emergent than others, helping us to determine the specific characteristics that cause the most emergent visits. Thus, after merging the ICD-9 codes with the Medicaid data, we had the option to select among four different variables (ED_Care_Needed_not_Preventable, ED_Care_Needed_Preventable_Avoi, Emergent_PC_Treatable, Non_Emergent) to assess the emergent level of each observation.

After review of the variables, the team used the percentages of two different variables (out of the four listed above) – Emergent__PC_Treatable & Non_Emergent. These two variables enabled us to group Emergent_Level (our new variable) into five different categories, ranging from most to least emergent/urgent cases plus a category of undetermined based upon the sum of the percentages of the two variables. The categories are listed below:

Division of Categories:

0 – Undetermined

4 – 0% - 25% = Most Emergent/Urgent

3 – 26% - 50%

2 – 51% - 75%

1 – 76% - 100% = Least Emergent/Urgent

For the multiclass modeling, we began by filtering out patients which had a PAYTYPE other than Medicaid. This left a population sample of only Medicaid pay type patients. Medicaid patients make up the second largest type of payments from this data set (Private insurance was the most common type of payment, with Medicare and Self-Payment also making up a large portion of the dataset). Only Medicaid pay types were selected for modeling.

The multiclass model focused on trying to predict the Emergent Level, which was determined as a scale between 1 and 4, with 1 being least urgent and 4 being most urgent. However, since much of the information in this dataset is required information, which is available before the patient's visit, it makes it difficult to predict these emergent cases before they actually occur. Based upon this information, the focus shifted to determining variable importance, which will help guide future research.

The selected modeling techniques allowed for increased interpretability. Specifically, for this project, a decision tree model and a random forest model were used as shown in Figure 12:

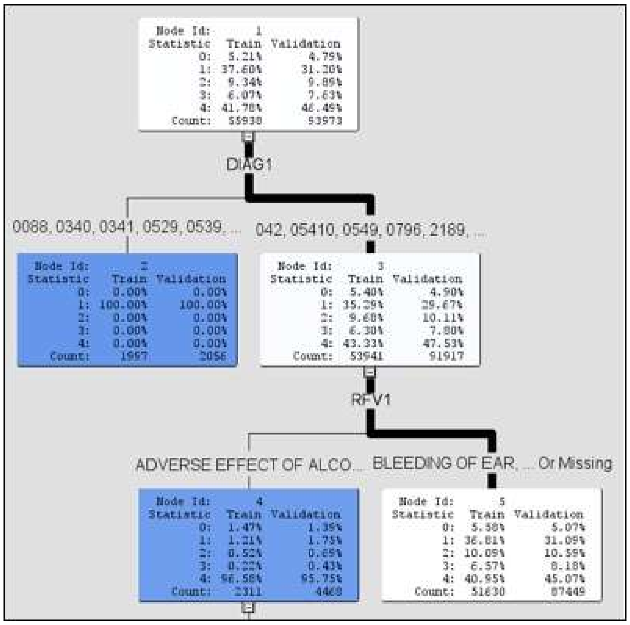


Figure 12: Multi-Classification - Decision Tree

The figure on the previous page shows the first split of the decision tree model. This split occurs on the variable DIAG1, which is the primary diagnosis from this patient's current visit to the emergency room. This makes sense because the reason for the visit would be highly correlated with level of urgency associated with that visit. However, due to the high number of diagnosis codes, this split suffers from high cardinality, meaning that it is easy for the model to make an impactful split on this variable due to the extremely high number of levels that are present within the data.

To reconcile this issue, future analysis may perform better if diagnosis codes are grouped, resulting in much fewer levels, (i.e., group codes by prefix as illustrated in Figure 13).

The following is a list of codes for International Statistical Classification of Diseases and Related Health Problems.

- List of ICD-9 codes 001–139: infectious and parasitic diseases
- List of ICD-9 codes 140–239: neoplasms
- List of ICD-9 codes 240–279: endocrine, nutritional and metabolic diseases, and immunity disorders
- List of ICD-9 codes 280–289: diseases of the blood and blood-forming organs
- List of ICD-9 codes 290–319: mental disorders
- List of ICD-9 codes 320–359: diseases of the nervous system
- List of ICD-9 codes 360–389: diseases of the sense organs
- List of ICD-9 codes 390–459: diseases of the circulatory system
- List of ICD-9 codes 460–519: diseases of the respiratory system
- List of ICD-9 codes 520–579: diseases of the digestive system
- List of ICD-9 codes 580–629: diseases of the genitourinary system
- List of ICD-9 codes 630–679: complications of pregnancy, childbirth, and the puerperium
- List of ICD-9 codes 680–709: diseases of the skin and subcutaneous tissue
- List of ICD-9 codes 710–739: diseases of the musculoskeletal system and connective tissue
- List of ICD-9 codes 740–759: congenital anomalies
- List of ICD-9 codes 760–779: certain conditions originating in the perinatal period
- List of ICD-9 codes 780–799: symptoms, signs, and ill-defined conditions
- List of ICD-9 codes 800–999: injury and poisoning
- List of ICD-9 codes E and V codes: external causes of injury and supplemental classification

Figure 13: ICD-9 Large Group Classification Index

The list above shows how ICD-9 codes are broken down further. This might be a good way to consolidate the diagnosis codes into a smaller list and create more meaning splits in the decision tree. Next, a neural network was run to help with variable selection and DIAG1 again had the highest Gini Reduction which reinforced the results seen from the decision tree. Looking at the model comparison (Figure 14), the random forest predicts marginally better than the decision tree; hence, both models are very beneficial.

Fit Statistics

Selected Model	Model Node	Valid: Average Squared Error	Valid: Misclassification Rate	Valid: Roc Index
Y	HPDMFore...	0.094486	0.303768	0.875
	Tree2	0.108994	0.43438	0.751

Figure 14: Fit Statistics

A review of the ICD-9 codes broken into groups revealed an abnormal spike for the data from Alabama in terms of Respiratory Systems, which was a strong category after the usual Injury and Poisoning and Ill-Defined conditions grouping. Figure 15 shows the breakdown by ICD-9 groupings as shown above for the Medicaid-only data for Alabama from the dataset:

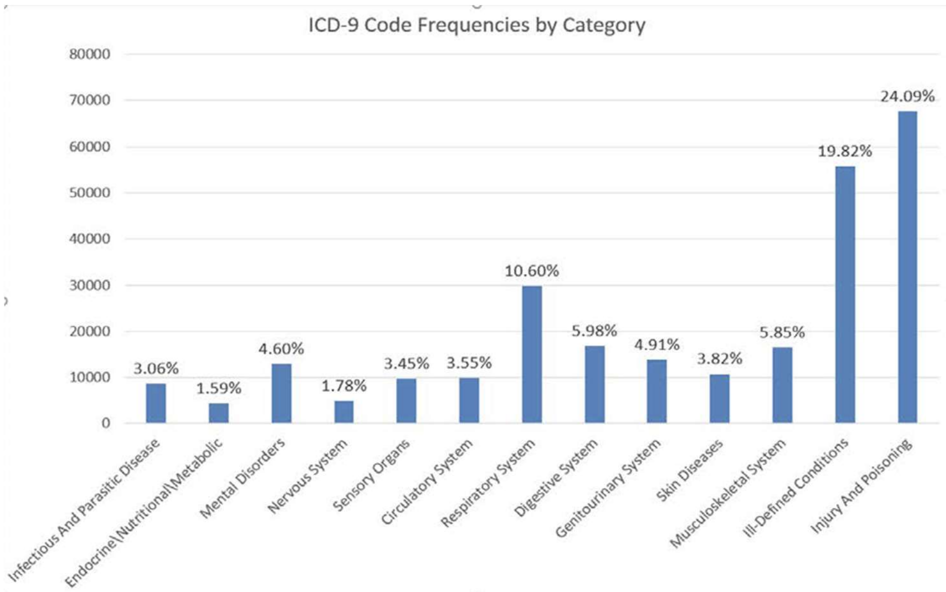


Figure 15: Breakdown of ICD-9 Categories (AL)

This notation led the team to infer that respiratory ailments (specifically, asthma) were helping to drive the significance of the ICD-9 splits seen in the multi-classification modeling. Looking at the cohort of 29,790 patients, this group made up about 10.6% of all emergent care. When broken further down into looking at just respiratory ailments between emergent and non-emergent care, there is evidence to show that the phenomenon here reverses where instance rates for Medicaid patients are as high or higher than the private pay counterparts (Figure 16):

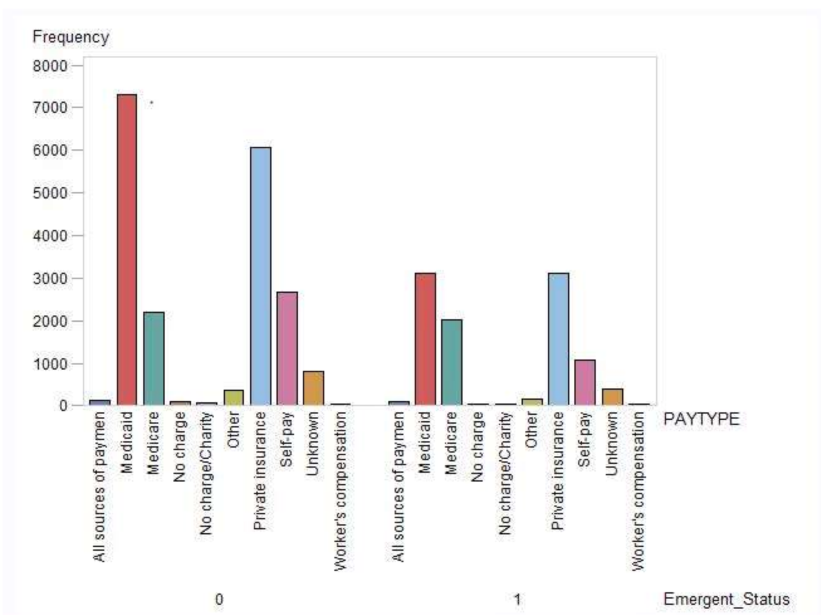


Figure 16: Respiratory Ailments - Emergent vs. Non-Emergent by Pay Type

CONCLUSION

Emergency departments are the only health care entities with a legal mandate to provide health care due to the Emergency Medical Treatment and Labor Act (EMTALA). This law ensures that anyone who comes to an emergency department, regardless of their insurance status or ability to pay, must receive a medical screening examination and be stabilized. In this study, we used three statistical models to analyze Center of Disease Control Medicaid data. In addition to finding no regional differences, analysis indicates from this data that Medicaid patients do not have a higher rate of non-emergent visits when compared to other payment types. However, the problem remains that overall Medicaid payments continue to rise. Access to care is only one necessary element for improving health outcomes. Some of what may seem to be misuse of emergency services may be a natural response to an underdeveloped primary care delivery system that is not meeting patient's needs. Providing better management for patients with or at risk for chronic disease is another strategy for not only reducing emergency department utilization and related costs, but increasing health related quality of life. Management strategies could include devoting more time to educating patients about how to manage chronic conditions. Additional data analysis and study could help accomplish the identification of other high risk disease states, while better managing Medicaid cost.

REFERENCES

Centers for Disease Control and Prevention. 2015. "National Center for Health Statistics." Accessed August 24, 2015. https://www.cdc.gov/nchs/ahcd/ and http://www.cdc.gov/nchs/ahcd/ahcd_questionaire.htm.

NYU-Wagner: The Center for Health and Public Service Research. 2015. "Faculty and Research." Accessed August 26, 2015. http://wagner.nyu.edu/faculty/billings/nyued-background.

American College of Emergency Physicians. 2014. "Emergency Department Planning and Resource Guidelines [policy statement]; Approved April, 2014.

American College of Emergency Physicians. 2015. "The Uninsured: Access to Medical Care". Accessed February 5, 2016. http://www.acep.org/News-Media-top-banner/The-Uninsured--Access-To-Medical-Care/

The Kaiser Commission on Medicaid and the Uninsured. 2000. *Uninsured in America, a chart book*. Menlo Park, CA: The Henry J. Kaiser Family Foundation.

Barnezai A, Melnick G, Nawathe A. 2005. "The cost of an emergency department visit and its relationship to emergency department volume." *Annals of Emergency Medicine,* **45(5)**:483-90.

Zeng F., O'Leary J.F., Sloss E.M., Lopez, M.S., Dhanani, N., Melnick, G. 2006. "The effect of medicare health maintenance organizations on hospitalization rates for ambulatory care-sensitive conditions". *Med Care,* **44(10)**:900-7.

Paradise, J. "Medicaid Moving Forward." 2015. The Henry J. Kaiser Family Foundation. Accessed September 5, 2015. http://kff.org/health-reform/issue-brief/medicaid-moving-forward/

Althaus F, Paroz S, Hugli O, Ghali, W.A., Daeppen, J.B., Peytremann,-Bridevaux, I., Bodemann, P. 2011. "Effectiveness of interventions targeting frequent users of emergency departments: a systematic review." *Annals of Emergency Medicine,* **58**:41-52.

ACKNOWLEDGMENTS

We would like to acknowledge and thank Healthcare Business Solutions (HBS) for the collaborative opportunity to work on this Medicaid project. Thank you to Sherri LeMier, Karen Guthrie, Matt Smith, Jennifer Grice, and Randy Heal at HBS. HBS is a member of the Institute of Business Analytics at The University of Alabama (UA) and this corporate project was conducted with a group of graduate students. Thank you to the UA students that participated in the project: Benjamin Bailey, Andrew Brown, Matt Collins, Kelsey Gibson, Jake Gurkin, Erik Halaj, John Ikard, Wesley Moss, Lindsey Stackhouse, Joshua Teel, Chris Wilhelm, and Joey Weed. SAS® tools were used for all phases of the project.

RECOMMENDED READING

- *Base SAS® Procedures Guide*
- *SAS® Visual Analytics: User's Guide*

CONTACT INFORMATION <HEADING 1>

Your comments and questions are valued and encouraged. Contact the author at:

Bradley M. Casselman, M.S.
Clinical Instructor, Information Systems, Statistics, and Management Science
The University of Alabama
Phone: 205-348-8909
E-mail: bcasselman@cba.ua.edu
Web: http:Culverhouse.ua.edu