

## Methods

The primary goals of this analysis are to 1) offer a better understanding of the characteristics and use patterns of patients at urgent care centers and 2) to examine empirically the differences between these individuals and those who continue to use traditional means of healthcare. Given these dual interests, the analysis was performed in three stages. First, an exploratory cluster analysis was performed on an abundance of patient characteristics for those who were coded as having gone to urgent care centers. Second, the results of the cluster analysis were used to inform a logistic regression analysis which examines the significant indicators which have an affect on whether a patient will make the initial decision to go to urgent care over traditional primary care. Third, a secondary regression analysis was performed to determine differences between patients using urgent care as primary care and those who do not.

## Data

An empirical exploration of the hypotheses proposed in chapter two requires data that provides an abundance of variables which may or may not be statistically important but which we cannot initially rule out, as well as a large sample size since the phenomenon is still comparatively rare when thinking about how patients access primary care in the US. Given these initial requirements, I chose the National Ambulatory Medical Care Survey (NAMCS), which is a national survey designed and distributed by the American Center for Disease Control (CDC) to provide researchers in the medical and social science fields “accurate and reliable information about the provision and use of ambulatory medical care services in the United States” [①namcs].

For those unfamiliar with healthcare lingo, ambulatory care is defined by the survey as health services or acute care services provided to patients on an outpatient basis (without an overnight stay). Once a year NAMCS surveys randomly selected visits to non-federal employed office-based physicians, which are collected from a representative sample of the United States’ medical care facilities. These surveys record over 500 variables on information about how the patients utilize physician services, the conditions most often treated, and the diagnostic and therapeutic services rendered, including medications prescribed. Because it is both representative of the larger trends in the United States and includes specific information regarding urgent care centers which can be used to explore that health care trend in particular, the data set serves the purposes of the current study quite well.

Some limitations to the data should be noted. There may be related errors given that as the popularity of

urgent care centers have risen, so too have the number that participated in the NAMCS. In 2008, there were 842 visits surveyed compared to 1168 surveyed in 2010. There may also be a compounding factor of selection: many urgent care centers are classified as retail clinics by the CDC, and are thus not included in the NAMCS random selection pool. Even with these limitations, the CDC specifically offers the NAMCS as a tool for social scientists, and this analysis depends on their collection methods being statistically expandable.

## Variable Selection and Summary Statistics

The NAMCS is a large data set: compiling observations for the years 2008-2012 generated a data set of over 100,000 observations with 547 variables. In order to explore such a large amount of data, I began the analysis by utilizing unsupervised statistical learning methods in order to develop descriptive categories of urgent care center patients. The advantages of performing an initial analysis of this nature are its ability to efficiently examine a large amount of variables and highlight those that should be included in later analyses. This choice was motivated by the current lack of statistical analysis and social theory surrounding urgent care centers, which makes it difficult to develop a model from scratch. The results of the clustering act as a guide for the regression techniques by identifying variables with some relationship to urgent care.

For the clustering, I examined the group of visits which were coded as having been at “Urgent Care Centers/Freestanding Clinics” by the NAMCS. While the combined years produced a data set of 123,120 observations, only 3,863 of those occurred at urgent care centers (about 3 percent). Of these visits, I limited the analysis to observations without missing information, bringing the sample to 3,695 visits to urgent care centers. *insert table 3.1*

## *k*-Modes Cluster Analysis

This method allows for grouping patients that have similar characteristics across a set of variables by dividing a set of cases into ever more numerous and specific subsets, thus leading to homogeneous empirical types (Rapkin and Luke, 1993). One of the most powerful exploratory aspects of cluster analysis is that you do not need to have a response variable in order to better understand your data. For this project, this is extremely useful since we initially only know who is going to urgent care and who is not, but would like to understand them as a group better before drawing comparisons between patients who visited a traditional primary care clinic. Another advantage for cluster analysis is that since such inductive methodologies are based only on quantitative similarities among cases, only two factors may be responsible for trends in the data: the actual structure of the observed phenomenon and the methodological decisions I made concerning choosing the

cases and variables (including the statistical method used to identify subsets). A further explanation of the  $k$ -modes clustering performed on the data is available in Appendix 1.

Because I am interested in two somewhat distinct aspects of the patients of Urgent Care – both their demographics and their patterns of use – The clustering was performed in two batches of parameters. Variables for *Age, Sex, Race, Urban Type, % Neighborhood Poverty, % Neighborhood college degree attainment, and Payment Type*, what I will refer to as the *demographic variables* from this point forward, were first analyzed for subgroups. Secondly, some of the same variables were again analyzed with the behavior parameters of *Injury related visit (y/n), Primary Caregiver, Seen Before?, Past Visits, Major Reason, and the day of the week*, what I will refer to as the *behavior parameters*. The summary statistics for these variables can be found in Table 1 of Appendix 2.

#### \* Patient and Visit Characteristics

Many of the variables chosen to include in the exploratory cluster analysis were selected with both an awareness of the historical and sociological understandings of the decision factors which affect how Americans choose their healthcare in mind and through an elimination process. Initially, 40 parameters on patient demographics and visit characteristics, along with regional data for each observation, were included in the clustering. To identify the subsets of characteristics, I began by clustering the urgent care observations using all of the initial parameters. Variables were then eliminated which failed to effect an observation’s placement within a cluster. Through this process of eliminating variables which did not cluster systematically, the parameters which are used in the following logistic regression analysis were chosen. Table 3.2 shows the proportions of each variable.

Table 1: Observed Proportions of Independant Variables

Variable	Category	Percentage
Sex	Female	57.1
	Male	42.9
AgeGroup	15-24 years	8.68
	25-44 years	22.89
	45-64 years	29.63
	65-74 years	13.93
	75 years and over	12.58
	Under 15 years	12.3
Race	Black	9.15
	Other	3.87

Variable	Category	Percentage
PaymentType	White	86.98
	All sources of payment are blank	0.44
	Medicaid	10.67
	Medicare	26.37
	No charge	0.58
	Other	3.07
	Private insurance	46.9
	Self-pay	4.86
	Unknown	1.93
UrbanCategory	Worker's compensation	5.17
	Large central metro	24.6
	Large fringe metro	14.62
	Medium metro	34.55
	Micropolitan/noncore (nonmetro)	15.98
	Missing data	0
PercentPoverty	Small metro	10.25
	Missing data	0
	Quartile 1 (Less than 5.00 percent)	18.16
	Quartile 2 (5.00-9.99 percent)	30.24
	Quartile 3 (10.00-19.99 percent)	39.22
	Quartile 4 (20.00 percent or more)	12.38

## Regression: Logistic Modeling of Urgent Care Visits

Following the cluster analysis, a logistic regression analysis is used in this study to examine the determinants of the odds of using urgent care as primary care. Often, sociologists and other social scientists performing quantitative analysis on a single dependant variable use a form of analysis called ordinary least squares (OLS) regression, which attempts to fit a line across a set of data points in order to best determine which independent variable has the greatest effect on determining the outcome for the dependant variable. This method works well for outcome variables which are quantifiable, however in the present study the variable of interest is binary and categorical: a person either uses the urgent care center as their primary care physician or they do not.

This creates many problems which violate the assumptions of linear models like OLS, and are the reason logistic regression was chosen instead. Logistic regression avoids the pitfalls of linear models by fitting a curved rather than a straight line to the data. This results in a better estimation of the effects of independent variables upon probabilities because the curvature of the line is able to fit the clustering of the answers at the two poles.

## Operationalization and Recoding of Variables

### \* Dependant Variables

The dependent variable analyzed in the logistic model in this study is whether or not a patient went to urgent care. Furthermore, a secondary analysis is conducted on how the patient demographics and visit characteristics relate to whether or not the patient uses urgent care as their primary care physician. Both of these outcome variables are analyzed in light of many visit level characteristics, up to 16 in the full logistic model.

### \* Independant Variables

*Injury Related Visit*

*Weekend Visit*

*Established Patient*

*Number Past Visits*

*Visit Reason*

*Age*

*Male*

*Race*

*High ZIP Income*

*Rural*