Negative Binomials Regression Model in Analysis of Wait Time at Hospital Emergency Department

Bill Cai[1], **Iris Shimizu**[1]

[1]National Center for Health Statistic, 3311 Toledo Road, Hyattsville, MD 20782

Abstract

Wait time is the differences between the time a patient arrives in the emergency department (ED) and the time an ED provider examines that patient. This study focuses on the development of a negative binomial model to examine factors associated with ED wait time using the National Hospital Ambulatory Medical Care Survey (NHAMCS). Conducted by National Center for Health Statistics (NCHS), NHAMCS has been gathering, analyzing, and disseminating information annually about visits made for medical care to hospital outpatient department and EDs since 1992. To analyze ED wait times, a negative binomial model was fit to the ED visit data using publically released micro data from the 2009 NHAMCS. In this model, the wait time is the dependent variable while hospital, patient, and visit characteristics are the independent variables. Wait time was collapsed into discrete values representing 15 minutes intervals. The findings are presented.

Keywords

Wait Time; ED; NHAMCS; Poisson Distribution; Poisson Regression; Negative Binomial Regression; Clustered Data

1. Introduction

Wait time is the difference between the time a patient arrives in the emergency department (ED) and the time an ED provider examines that patient. Wait time is a problem faced by both patients and hospitals. The mean wait time in EDs increased 25%, from 46.5 minutes in 2003 to 58.1 minutes in 2009. [1] The time it takes a patient to see a doctor (wait time) can be critical to the patient's health and to a hospital's service quality. Clinically, prolonged ED wait times may result in protracted pain and suffering and in delays in diagnosis and treatment. Many factors, such as ED overcrowding, can affect the wait time. Overcrowding in EDs may also place patients at greater risk for medical errors. (refer to American College of Emergency Physicians web site "Fact Sheets")

Many studies examining ED wait times have used data from the National Hospital Ambulatory Medical Care Survey (NHAMCS, see http://www.cdc.gov/nchs/ahcd/about_ahcd.htm). Conducted by the National Center for Health Statistics (NCHS) annually

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Center for Health Statistics or the Centers for Disease Control and Prevention.

since 1992, NHAMCS collects data about sample visits made for medical care to EDs and outpatient departments (OPDs) of a national probability sample of non-Federal, general and short-stay hospitals. The survey is a component of the National Health Care Survey, which measures health care utilization across a variety of health care providers. There are two micro-data files produced from NHAMCS and released for public use, one for OPD visits and one for ED visits. [2]

Wait time is a topic of many publications by staff at NCHS and researchers from different organizations. A majority of the publications use descriptive statistics or quote results from annual Emergency Department Summaries using NHAMCS, such as that released by NCHS for 2007 [3]. Those summaries report total numbers of visits in the following intervals: fewer than 15 minutes, 15 to 59 minutes, 1 hour but less than 2 hours, 2 hours but less than 3 hours, 3 hours but less than 4 hours, 4 hours but less than 6 hours, 6 hours or more, not seen by a physician, and missing blank. They also give a median wait time (e.g., 33 minutes in 2009 [1]).

To further analyze wait time using NHAMCS data, some statistical model-based methodologies were also introduced. [4], [5], [6]

In [4], multivariate linear regression methodology was applied to the NHAMCS 1997 to 2004 ED data. The change in wait time to see ED physicians was evaluated for all adults, patients diagnosed with acute myocardial infarction (AMI), and patients whose triage personnel designated as needing "emergent" attention. SURVEYFREQ and SURVEYREG procedures were used in SAS to better adjust for NHAMCS's complex sample design.

In [5], multivariate logistic regression was used to estimate the association between wait time and the patient's age, sex, payment status, and geography of that visit using NHAMCS 1997 to 2006 data. Wait times were analyzed for visits with diagnoses of acute pancreatitis, appendicitis, cholecystitis, and upper gastrointestinal hemorrhage (UGIH). For these analyses, the association between the patient's race/ethnicity and frequency of delay relative to triage assignment was evaluated by year.

In [6], both descriptive statistics and linear regression analysis were conducted. Data for ED visits by patients 15 years of age or less from 1997 to 2000 were examined. Sample weights were applied to the identified patient records to yield national estimates. For the purposes of that study, ED wait time was analyzed for 3 major groups, i.e., non-Hispanic white (NHW), non-Hispanic black (NHB), and Hispanic white (HW). Using a linear regression analysis with logarithmically transformed wait time as a dependent variable and with adjustment for potential confounders, including hospital location, geographic region, and payer status, the study found that NHB and HW patients waited longer than NHW patients..

This paper will examine the 2009 NHAMCS micro data on ED visits that were publicly released by NCHS. The data include hospital, patient, and visit characteristics for each visit. The sample ED visits from a hospital are treated as a cluster and hospitals are mutually independent. Wait time was collapsed into discrete values based on 15-minute intervals. The resulting frequency plots showed a Poisson distribution for the wait time. A Poisson regression model was fitted to the data with wait time as the dependent variable;

hospitals as subject variables (all visits to one hospital as a cluster); patient's arrival time, arrival by ambulance, age, race/ethnicity, hospital location, etc. as covariates. Because of an overdispersion problem associated with the Poisson model for the data set, a negative binomial regression model was used to fit the data instead.

The authors have found that neither generalized linear models (to which both Poisson regression model and negative binomial regression model belong) nor the treatment of ED visits to each hospital as a cluster have been examined in the literature on ED wait time. The objective of this study is to find a fitting model to examine characteristics associated with ED wait time."

2. Statement of the problem

Scientific questions concerning wait time for ED patient visits are broad. Most of them can be answered with descriptive statistics, but some of them require statistical inference based on a statistical model.

Examples of descriptive statistics are: what's the average wait time, what's the median wait time, what's the average wait time for females, at what time of the day is the wait time longest, etc. Those questions could be answered by applying SAS, R or other statistical software directly to the ED visit data. For a complex multistage survey like NHAMCS, the PROC SURVEYFREQ procedure in SAS is quite useful to produce those results. Also, the wait times published in NCHS annual Emergency Department Summaries report frequently used statistics on wait time by time intervals.

When one investigates the attribution of some independent factors on ED wait time, one needs to adjust for confounding factors to avoid the possibility of spurious relationship. For example, according to many publications [4], [6], racial minority patients have longer wait times than their white counterparts. However, a majority of racial minority patients live in metropolitan areas, and hospitals in metropolitan areas have on average longer wait times than hospitals in non-metropolitan areas [1], [4]. One should adjust for hospital location when investigating the differences in mean wait times between racial groups. Statistical modeling and statistical inferences are required to answer those kinds of questions.

To compare differences in ED wait time between population subgroups, such as those defined by age, sex, race, etc., one needs to test for significance of the difference. If the p-value of the test statistic is less than the predetermined significant level α, then the null-hypothesis of no difference is rejected and one concludes that there is a difference in wait time between the groups. The probability of a Type I (α) error for a statistical test describes the probability of incorrectly rejecting the null-hypothesis and concluding there is a difference in ED visit wait time. The smaller the α, the less frequently one would reject the null hypothesis. It is common practice to set the probability of a Type I error at 5%.

The statistical models and the hypothesis testing described above are used to examine the wait time in hospital EDs among different population subgroups.

3. Methodologies

I. NHAMCS ED data

The sample for the 2009 NHAMCS consisted of 489 hospitals of which 389 were in scope and had eligible EDs. Of these, 356 EDs participated by providing data for a sample of their ED visits for a four-week period. The basic sampling unit for NHAMCS is the patient visit or encounter. During the 2009 NHAMCS, data were collected on 34,942 ED visits.

In the data file, the wait time variable is WAITTIME, which is minutes to be seen by a physician, physician assistant, or nurse practitioner after the patient's arrival at the ED, with values from 0 to 1440. Due to the fact that the initial 15 minutes are critical (emergent visits) in the treatment of the patients with serious illness, wait time was collapsed into discrete values based on 15-minute intervals. WAITTIME values of −7 ('Not applicable') or −9 ('Blank') were treated as missing (.). Wait time was not reported for about one quarter of the sample ED visits. The current study analyzed only the reported data and did not impute missing wait times. An analysis of potential study bias due to the missing data is needed but was not done in the current study.

The purpose of this study is to examine broadly risk factors associated with ED wait time, instead of focusing on one particular variable, such as arrival mode or race. Based on publications on ED visit wait time and their most frequently analyzed variables, the following variables were selected for this investigation:

○ ARRTIME:	arrival time (military time), converted to minutes.
○ ARREMS:	arrival by ambulance, with 1='Yes';
	2='No'; −8='Unknown'; −9='Blank'. For the current study, the categories for −8 and −9 were replaced by 3="unknown or blank'
○ HOSPCODE:	hospital number (identifier), starting from 1.
○ SEX:	patient sex, 1='Female'; 2='Male'.
○ RACER:	patient race.
○ ETHIM:	patient ethnicity. For the current study, the variables RACER and ETHIM were combined into
	1='non-Hispanic white (NHW) ',
	2='non-Hispanic black (NHB) ',
	3='Hispanic white (HW) ', and
	4='others'.
○ AGER:	patient age, 1='less than 15 years old'; 2='15 – 24 years old'; 3='25 – 44 years old'; 4='45 – 64 years old'; 5='65 – 74 years old'; 6='75 years or older'.

II. Model

First, all the patient visits to one hospital could be treated as repeated measurements; therefore, each hospital could be treated as a cluster. Secondly, as seen in the frequency plots for wait time shown in Figure 1 (for visits in some individual hospitals) and Figure 2 (over all visits and over all hospitals), the assumption of Poisson distribution for wait time is justified as a starting point of the analysis.

In statistics, Poisson regression is a form of regression analysis commonly used to model count data in longitudinal and clustered data analysis. Poisson regression assumes the response variable has a Poisson distribution, and assumes the logarithm of its expected value can be modeled by a linear combination of unknown parameters.

For a fixed Poisson regression model for ED wait time, with linear covariates; and no interactions, let

$$y_{ij} \sim \text{Independent Poisson}\ (\lambda_i)$$
$$\log(\lambda_i) = x_i^T \beta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5}$$

where,

y_{ij}: wait time (1, 2, …,) of j^{th} patient visit to i^{th} hospital ED, with $E(y_{ij}) = \lambda_i$

x_{i1} represents patient visit to ED arrival time effects

x_{i2} represents patient visit to ED arrival mode effects

x_{i3} represents patient sex effects

x_{i4} represents patient race/ethnic effects

x_{i5} represents patient age effects

However, for a Poisson distributed variable Y_i, if $\mathrm{Var}(Y_i) > E(Y_i)$, then overdispersion is present, which is the presence of greater variability (statistical dispersion) in a data set than would be expected based on a given simple statistical model.

For the wait time data,

$$\mathrm{Var}(y) = 19.44 > 4.21 = E(y), \quad \text{with } y = y_1 + y_2 + y_3 + \ldots \text{and, } \mathrm{Var}(y_i) > E(y_i) \text{ for vast majority of } i = 1, 2, 3 \ldots$$

To address the overdispersion issue, a negative binomial regression (NB) model was used instead.

If Y_i is a negative binomial distributed count with mean μ_i and dispersion parameter α, a general form of the probability mass function (pmf) of $Y_i \sim NB(\mu_i, \alpha)$ is given by

$$f(y_i, \mu_i, \alpha) = \begin{cases} \dfrac{\Gamma\left(y_i + \alpha^{-1}\mu_i^{1-k}\right) \alpha^{y_i} \mu_i^{k y_i}}{y_i!\ \Gamma\left(\alpha^{-1}\mu_i^{1-k}\right)\left(1 + \alpha\mu_i^{k}\right)^{y_i + \alpha^{-1}\mu_i^{1-k}}}, & y_i = 0, 1, \ldots, \alpha > 0 \\ 0, & otherwise \end{cases}$$
$$\text{with } E(Y_i) = \mu_i \text{ and } \mathrm{Var}(Y_i) = \mu_i(1 + \alpha\mu_i^k)$$

Here α is assumed to be a constant. The index k identifies various forms of the NB distribution, but two well-known models are given by k = 0 and 1. For k = 0 we have a linear-variance NB regression, or NB1 model, with $\mathrm{Var}(Y_i) = \mu_i(1+\alpha)$ [this is often approximated by fitting the constant overdispersion quasi-likelihood model with $\mathrm{Var}(Y_i)$ =

φμ_i, where φ is a constant]. Taking k = 1 gives the more usual quadratic-variance NB, or NB2 model, with Var(Y_i) = $\mu_i(1+\alpha\mu_i)$. As α → 0, the NB model reduces to the Poisson model. For both models, some specific regression model is assumed for the mean, i.e. $\log(\mu_i) = x_i^T \beta$.

III. Model selection

The final negative binomial regression model adopted for the wait time of patient ED visits in the 2009 NHAMCS is:

$$y_{ij} \sim \text{independent NB}(\mu_i, \alpha)$$
$$\log(\mu_i) = x_i^T \beta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$$

where,

α = 0.55, σ(standard error) = 0.0061;

β_0 = 0.73, σ =0.0657;

β_1 = 0.18, σ =0.0211;

β_2 = 0.05, σ =0.0139;

β_3 = 0.13, σ =0.0180.

y_{ij}: wait time (1, 2, …) of j^{th} patient visit to i^{th} hospital ED, with E(y_{ij}) = μ_i

x_{i1} represents patient ED arrival mode effects

x_{i2} represents patient sex effects

x_{i3} represents patient race/ethnic effects

4. Results

The negative binomial model applied in this research accounts for the properties of the NHAMCS data – properties such as the Poisson distribution of wait time and correlated visits to each ED, which have not been investigated before to examine ED visit wait time with NHAMCS data. Table 1 shows that collapsing wait time values into 15-minute units increased the mean minutes of wait time from 58.8 to 63 (=4.2*15) minutes. That is, collapsing wait times into 15 minute-based discrete values introduced imprecision into the data and, hence, the analysis results.

Table 2 presents results of comparisons between visit groups defined by patient race and ethnicity and groups defined by arrival mode. While the wait time difference between blacks and Hispanic whites is not significant, the ED wait time differences between non-Hispanic whites and Hispanic whites and between non-Hispanic whites and non-Hispanic blacks are both significant. Non-Hispanic whites have the shortest mean wait time compared to non-Hispanic blacks and Hispanic whites.

Table 2 also shows that the ED wait time difference between arrivals by ambulance and arrivals by non-ambulance methods is significant.

Also, from the model selection process, it cannot be concluded that patient age significantly effects wait times (not shown).

5. Summary

This study focuses on the development of a fitting model to examine factors associated with wait time experienced by patients in hospital EDs, while some previously well-known results are also investigated. The study analyzed publicly released micro data on ED visits from the 2009 NHAMCS, and the negative binomial model was chosen.

Previously known results that non-Hispanic black patients and Hispanic white patients waited longer than non-Hispanic white patients were independently confirmed in this analysis. It was also shown that wait time was less for patients who arrived by ambulance than for those who arrived by other means.

There are limitations to this study. First, ED wait time was not reported for about one quarter of the sample ED visits; therefore, some bias analysis is needed. Second, the main focus of this study is on the development of a fitting model that is used to examine factors associated with ED wait time. Not all relevant patient- and ED-related factors have been included in the study. Other research questions concerning ED wait time can be studied using the model derived in this study. For example, will the difference remain between whites and non-whites in ED wait time if the model adjusts for hospital location?

As discussed in this paper, the negative binomial regression model with associated clustered data methods is an appropriate statistical method for use in studying wait time for ED visits using NHAMCS data. The model reflects characteristics of ED wait time data, such as Poisson distribution of wait time and clustered data property of the visits while models used previously in literature, such as linear regression model and logistic model, did not.

References

[1]. Hing E, Bhuiya F, 2012. Wait Time for Treatment in Hospital Emergency Departments: 2009. NCHS Data Brief, No. 102, August 2012.

[2]. National Center for Health Statistics (NCHS). 2012. 2009 National Hospital Ambulatory Medical Care Survey Micro-data File Documentation. Hyattsville, Md.:NCHS.

[3]. Niska R, Bhuiya F, Xu J, 2010. National Hospital Ambulatory Medical Care Survey: 2007 Emergency Department Summary. National Health Statisticals Reports, No. 26.

[4]. Wilper A, Woolhandler S, Lasser K, Mccormick D, Cutrona S, Bor D, Himmelstein D, 2008. Wait to See an Emergency Department Physician: U.S. Trends and Predictions, 1997–2004. Health Affairs (Published online January 15, 2008).

[5]. Wu B, Banks P, Conwell D, 2009. Disparities in emergency department wait times for acute gastrointestinal illnesses: results from the National Hospital Ambulatory Medical Care Survey, 1997–2006. Am J Gastroenterol. 2009 May 12.

[6]. James CA, Bourgeois FT, Shannon MW. Association of race/ethnicity with emergency department wait times. Pediatrics. 2005 Mar.

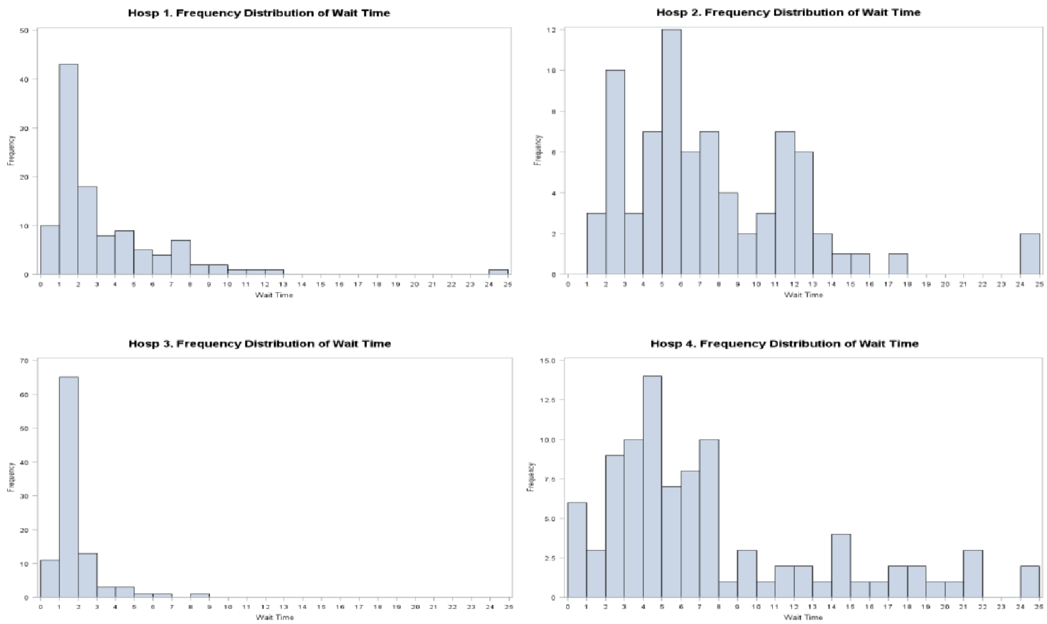

Figure 1.
Wait Time Frequency Plot – Visits to Hospital 1,2,3,4

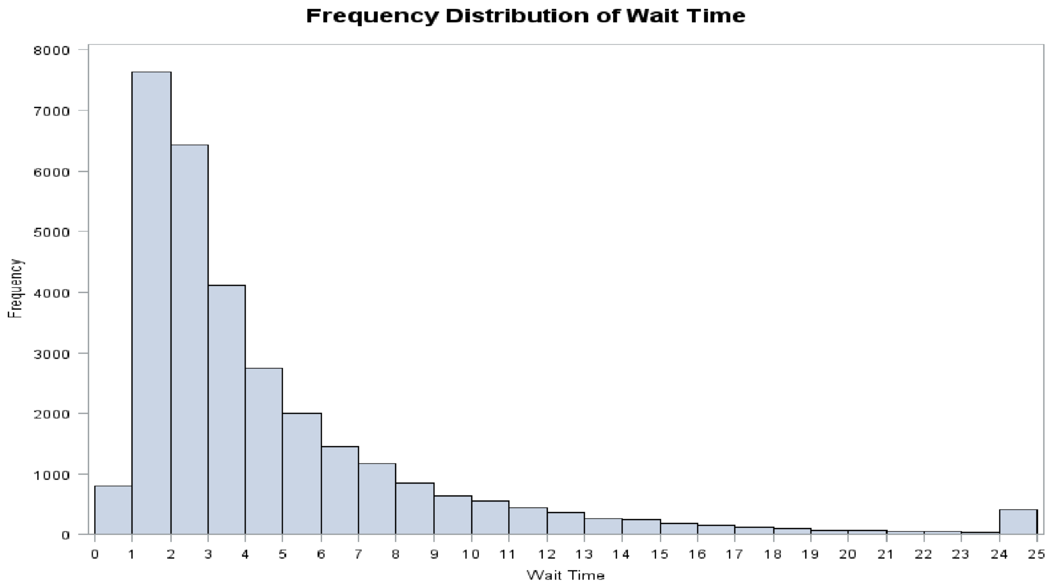

Figure 2.
Wait Time Frequency Plot – Visits to All Hospitals

Table 1.

Means and standard deviation of ED wait time before and after collapsing

Variable	Mean	Std Dev
ED Wait Time (unit: 1 minute)	58.8	83
ED Wait Time (unit: 15 minutes)	4.2 (=63min)	4.4 (=66min)

Table 2.

Test of mean differences in ED wait times between selected groups of patients

Label	EstimatedMean Difference	Mean Confidence Limits		Chi-Square	P-value
NHB - NHW	0.7165	0.6537	0.7853	50.77	<.0001
HW - NHW	0.7727	0.6941	0.8602	22.2	<.0001
HW - NHB	1.0784	0.9579	1.2141	1.56	0.2116
(arrival by non-ambulance) - (by ambulance)	0.8039	0.7562	0.8545	48.97	<.0001