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[1]
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 is 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 is collapsed into discrete values representing 15 minutes intervals. The findings are presented.

Key Words: Wait Time, ED, NHAMCS, Poisson Distribution, Poisson Regression, Negative Binomial Regression, Clustered Data, GEE

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 "Facts Sheets": http://newsroom.acep.org/index.php?s=20301&item=29937)

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 since 1992, NHAMCS collects data about sample visits made for medical care to EDs and outpatient departments (OPDs) of a national

[1] 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.

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 (till 2007), such as that released by NCHS for 2007 [3], or summary tables published on the NCHS web site (see http://www.cdc.gov/nchs/ahcd/ahcd_products.htm). 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 whom ED 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 the 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). A linear regression analysis was applied with logarithmically transformed wait time as a dependent variable and with adjustment for potential confounders, including hospital location, geographic region, and payer status.

The current study examines 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 assumed mutually independent. Wait time is collapsed into discrete values based on 15-minute intervals. The resulting frequency plots show a Poisson like distribution for the wait time. A Poisson regression model is initially fitted to the data with wait time as the dependent variable; hospital as the subject variable (all visits to one hospital as a cluster); patient’s arrival time, arrival mode, age, race/ethnicity, hospital location, etc. as covariates. Because of an over-dispersion problem associated with the Poisson model for the data set, a negative binomial regression model is subsequently used to fit the data instead. The Generalized Estimating Equation (GEE) method for generalized linear model is applied.

In a search of the literature on ED wait time, no studies are found in which either GEE for generalized linear models (to which both Poisson regression model and negative binomial regression model belong) or the Poisson (or NB) distribution were examined for use in modeling ED wait time. The objective of this study is to find a fitting model to reflect characteristics of ED wait time data, such as Poisson like distribution of wait time and clustered data properties of the visits.

2. Statement of the problem

Most of the scientific questions concerning wait time for ED patient visits 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, NCHS annual Emergency Department Summaries report frequently used statistics on wait time by time intervals.

Some research questions focus on identifying independent effects of explanatory variables. For example, according to many publications [4], [6], racial minority patients have longer wait times than their white counterparts. Also, 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]. Because it may explain differences in ED wait times between race groups, one should adjust for hospital location in his/her analyses 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 significance 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

3.1 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 was 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 the patient waited 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 minutes (= 24 hours). Due to the fact that the initial 15 minutes are critical (emergent visits) in the treatment of the patients with serious illness, wait time is collapsed into discrete values based on 15-minute intervals. WAITTIME values of -7 ('Not applicable') or -9 ('Blank') are treated as missing (.). Wait time is not reported for about one quarter of the sample ED visits. The current study analyzes only the reported data and do not impute missing wait times. An analysis of potential study bias due to the missing data is needed but is not done in the current study.

The purpose of this study is to examine all risk factors associated with ED wait time together, instead of focusing on one particular variable, such as arrival mode or race. Based on publications about ED visit wait time and their most frequently analyzed variables, the following variables in the NHAMCS public ED micro dataset are 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'.

3.2 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 selected individual hospitals) and Figure 2 (over all visits and over all hospitals), the assumption of Poisson distribution for wait time is suggested as a starting point of the analysis.

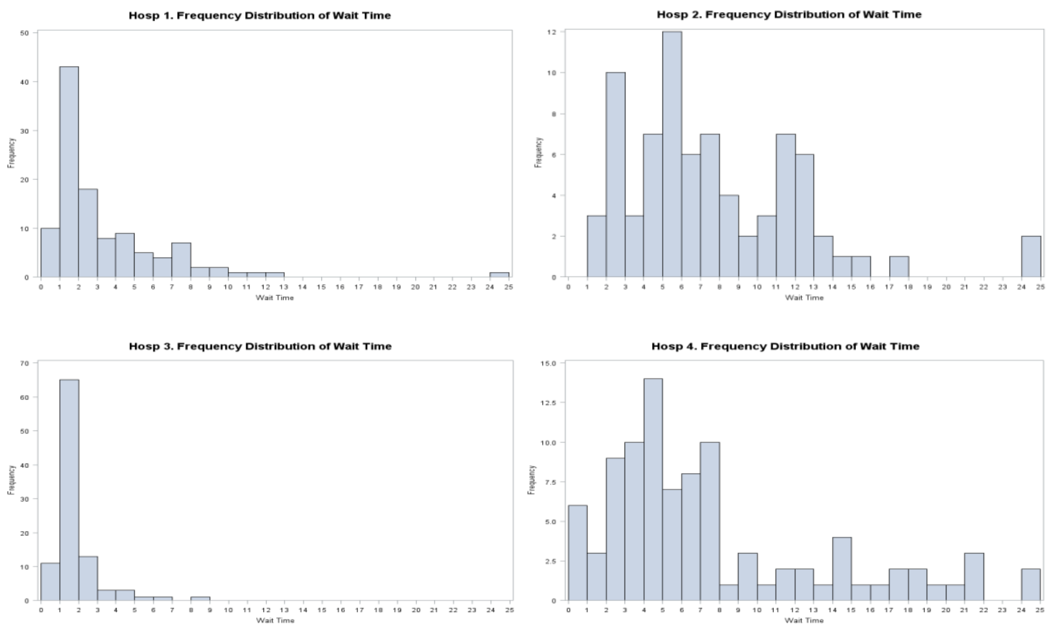

Figure 1: Wait Time Frequency Plot – Visits to Hospitals 1, 2, 3, 4

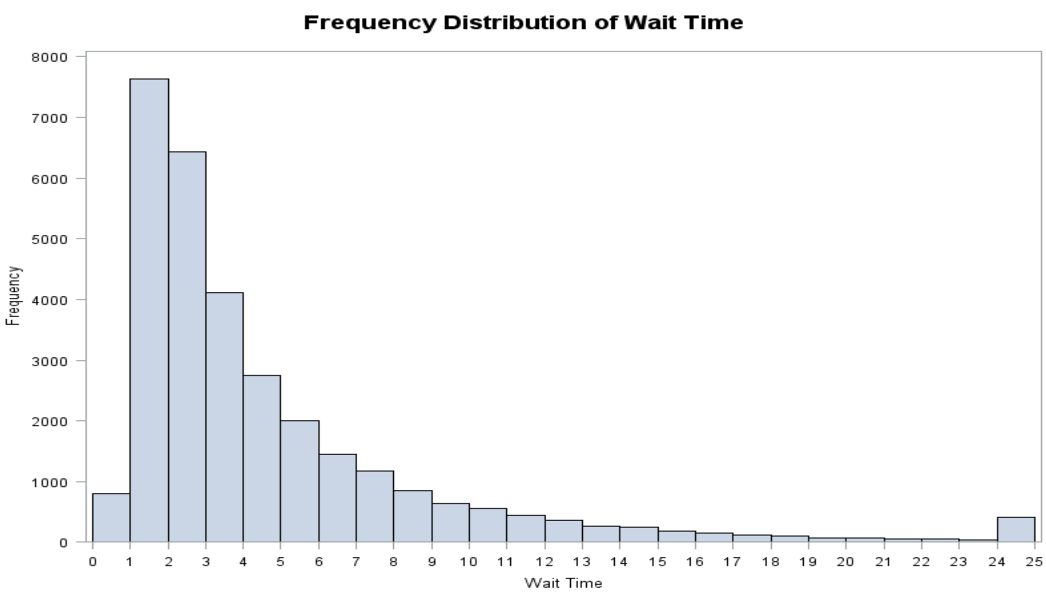

Figure 2: Wait Time Frequency Plot – Visits to All Hospitals

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 ED arrival time effects
x_{i2} represents 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, if Var(Y) > E(Y) , then over-dispersion 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,

$$\text{Var}(y) = 19.44 > 4.21 = \text{E}(y),$$

$$\text{where } y = \sum_i \sum_j y_{ij}$$

and,

$$\text{Var}(y) > \text{E}(y),$$

$$\text{where } y = \sum_j y_{ij}, \text{ for vast majority of } i = 1, 2, 3 \ldots$$

To address the over-dispersion issue, a negative binomial regression (NB) model is used instead. NB regression has the same mean structure as Poisson regression and it has an extra parameter to model the over-dispersion.

If Y is a negative binomial distributed count with mean μ and dispersion parameter α, a general form of the probability mass function (pmf) of Y ~ NB(μ, α) is given by

$$f(y, \mu, \alpha) = \begin{cases} \dfrac{\Gamma(y + \alpha^{-1}u^{1-k})\, \alpha^y u^{ky}}{y!\, \Gamma(\alpha^{-1}u^{1-k})(1 + \alpha u^k)^{y + \alpha^{-1}u^{1-k}}}, & y = 0, 1, \ldots, \\ 0, & otherwise \end{cases}$$

$$\text{with } E(Y) = \mu \text{ and } \text{Var}(Y) = \mu(1 + \alpha u^k)$$

Here α is assumed to be a positive 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 Var(Y = μ(1+α) [this is often approximated by fitting the constant over-dispersion quasi-likelihood model with Var(Y) = φμ, where φ is a constant]. Taking k = 1 gives the more commonly used quadratic-variance NB regression, or NB2 model, with Var(Y) = μ(1+αμ). 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(μ)= $x^T\beta$.

3.3 Model Selection

To determine which effects should be included in the final model, an analysis in SAS of the initial NB regression model with all the effects selected is performed. The patient arrival time effect is removed, because its estimate of 0.0001 is small compared with other effect estimates. The final negative binomial regression model fitted for the wait time of patient ED visits in the 2009 NHAMCS is:

$$y_{ij} \sim \text{independent NB}(\mu_i, \hat{\alpha})$$

$$\log(\mu_i) = x_i^T\ \hat{\beta} = \widehat{\beta_0} + \widehat{\beta_2}x_{i2} + \widehat{\beta_3}x_{i3} + \widehat{\beta_4}x_{i4} + \widehat{\beta_5}x_{i2}$$

where,

$\hat{\alpha}$ = 0.62 with standard error σ = 0.0064;
y_{ij}: wait time (1, 2, …) of j^{th} patient visit to i^{th} hospital ED, with E(y_{ij}) = μ_i
x_{i2} represents patient ED arrival mode effects
x_{i3} represents patient sex effects
x_{i4} represents patient race/ethnic effects
x_{i5} represents patient age effects

To help assess the fit of the model, we use the goodness-of-fit-chi-squared test. This assumes the deviance follows a chi-square distribution with degrees of freedom equal to the model residual. The resulting p-value of 0.5640 shows the goodness-of-fit-chi-squared test is not statistically significant; we conclude that the model fits reasonably well.

4. Results

The negative binomial model applied in this research accounts for the properties of the NHAMCS data – properties such as the Poisson like distribution of wait time and correlated characteristics of ED visits and hospitals, which have not been investigated before in analyses of 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 66 (=4.4*15) minutes. That is, collapsing wait times into 15 minute-based discrete values introduce imprecision into the data and, hence, the analysis results.

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	83.0
ED Wait Time (unit: 15 minutes)	4.4 (=66min)	5.5 (=82.5min)

Table 2 presents results of comparisons between visit groups defined by patient race/ 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.

Table 2: Test of mean differences in ED wait times between selected groups of patients

Label	Estimated Mean Difference	Mean Confidence Limits		Chi-Square	P-value
NHB - NHW	0.7	0.6334	0.7712	50.93	<.0001
HW - NHW	0.74	0.6623	0.8329	25.87	<.0001
HW - NHB	1.06	0.94	1.2013	0.94	0.3313
(arrival by non-ambulance) - (by ambulance)	0.81	0.7553	0.863	39.6	<.0001

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 analyzes publicly released micro data on ED visits from the 2009 NHAMCS, and the negative binomial model is chosen.

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

There are limitations to this study. First, ED wait time is not reported for about one quarter of the sample ED visits; therefore, some non-response bias analysis is needed. Second, the model fitted in the current study is not survey design based, which should lead to a new study when time permits. Third, the main focus of this study is on the development of a fitting model that could be used to examine factors associated with ED wait time. Not all relevant patient and ED related factors are included in the current study; and not all questions of interest to health workers and organizations are considered. An example of questions the study does not address is whether the difference between weekday and weekend mean ED wait time is significant? And last, wait times are collapsed into 15 minute-based discrete values. That collapsing results in increased mean wait times, hence, potentially introduces imprecision into the analysis results.

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 like distribution of wait time and clustered data properties of the visits while models used previously in literature, such as linear regression model and logistic model, do 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.