2018 IEEE International Conference on Information Reuse and Integration for Data Science

# Modeling CDC Data for Real Time Feature Estimation

Tannaz Rezaei Damavandi , Sampath Jayarathna, Yu Sun
Department of Computer Science
California State Polytechnic University
Pomona, CA 91768, USA
tannazr, ukjayarathna, yusun@cpp.edu

***Abstract*— Despite the fact that the main objective of an Emergency Department (ED) is to treat seriously ill patients, quality of care degrades due to prolonged waiting time. This paper investigates overcrowding as the main bottleneck in healthcare system performance and addresses patients' dissatisfaction and low quality of EDs' services on wait time. In this paper, we present empirical study, data analytics and application towards answering whether there is a significant correlation between medical immediacy and the length of wait time. This research is based on a series of feature estimation on 27 features extracted from publicly available dataset from Disease Control and Prevention (CDC) to find the extent of their correlation to prolonged wait time. The results suggest that in 2015 more than 63% of EDs visits were related to non-urgent medical issues. Our findings indicate that overcrowding occurs due to the presence of very large number of patients with non-urgent medical problems.**

***Keywords- data analysis; feature extraction; CDC data set***

## I. INTRODUCTION

Overcrowding is the major cause of patients' dissatisfaction and poor performance of emergency departments (EDs) on wait time [1]. This crucial health system problem happens due to the presence of large volume of admitted patients outnumbering medical staffs or treatment rooms' capacities which will force EDs to operate beyond their available resources. Overcrowding is a multi-elements problem; however, most of proposed solutions focus on increasing EDs' capacities especially with respect to equipment and medical staff.

In this research study, we investigate this problem based on statistics from the Centers for Disease Control and Prevention (CDC). The results of analyzing this data set show that most of patients who visit EDs are non-urgent patients; however, their presence at EDs can lead to the main bottleneck in degraded quality of care.

Our research is motivated by the fact that growing number of non-urgent patients, not only will lead to prolonged waiting time, but also will increase the cost of treatment due to unnecessary diagnostic medical exams.

This paper is studying two main features ***wait time*** and ***immediacy level*** and seeking to find their correlation.

Wait time is defined as the length of waiting in minutes for a patient upon her arrival at an ED till she is visited by a medical expert (MD/DO/PA/NP) and immediacy level is the urgency with which a patient should be seen by a medical expert [10].

This research utilizes year 2015 dataset which at this point is the most current one publicly available by CDC. In this data set as the urgency level decreases the triage number increases. Therefore, as the level of immediacy increases, patient has to undergo longer wait time. For instance, a patient with high medical urgency will receive immediacy number 1 or 2, while a patient with non-urgent medical issue will be assigned immediacy level ranged from 3 to 5.

After imputing missing data by utilizing three different approaches, we propose and analyze single variable linear regression models. Later, we show that prolonged wait time is mainly caused by the presence of non-urgent patients. Finally, we propose and implement our solution to overcrowding problem that can utilize UCs' potentials via a web application; it informs patients about UCs capabilities and provides them with the option to compare healthcare facilities' wait time and make an online schedule.

## II. RELATED WORKS

Although overcrowding problem in EDs has been widely studied, to the best of our knowledge, most studies merely are focused on the reason of overcrowding in hospitals without proposing practical solutions. Most of them conclude that prolonged waiting time and patients' dissatisfaction are mainly rooted in healthcare resource deficiency. For instance, [1] analyzes this prolonged wait time in EDs in Turkey and ascertains that overcrowding is mainly caused by the presence of non-urgent patients; it proposes that government should take proper action and develop an effective primary care system. Although this work shows a promising result, it doesn't offer practical solution on how it is possible to "keep non-urgent patients out of EDs". In [2], correlation between waiting time and hospital performance is addressed and its results show that only a small number of U.S. hospitals were able to fulfill the goal of admitting patients during the "recommended wait time" which is 15 minutes at maximum, and about half of the nationwide hospitals couldn't do it better than "4 to 6 hours". Researchers in [6, 7, 8, and 9] present study on a small group of population; solely focus on specific demographical features such as race, ethnicity, age, location or combination of them.

978-1-5386-2659-7/18/$31.00 ©2018 IEEE
DOI 10.1109/IRI.2018.00045

At application level, [5] implements a web-based program that merely shows the average wait time in EDs nationwide. It is not capable of showing wait time in nearby urgent care facilities nor does it provide patients with making an online appointment feature. In other related applications such as [3,4] online scheduling option is provided for limited number of EDs or UCs. These healthcare facilities are either affiliated or located in the same county or city.

We believe that the proposed solution is robust and grounded on the feature analysis compared to related works. It offers both 1) comparing/checking wait time in patients' nearby healthcare facilities and 2) online appointment option. As opposed to many related works where the only solution to overcrowding problem is increasing and equipping EDs, we believe that there is enough potential at UCs; these facilities can deliver the same quality of care at lower cost and with shorter wait time to *non-urgent* patients.

## III. METHODOLOGIES

We use a data set of a survey conducted in 2015 from the National Hospital Ambulatory Medical Care Survey (NHAMCS). This is a stratified random sampling survey that contains 21,061 electronic patient's records representing 136,943,000 visits to 377 U.S. hospital EDs, not including Federal, military, and Veterans Administration hospitals. This dataset of 1013 columns contains a broad range of patients' data such as sex, age, ethnicity, race medical issue, wait time, immediacy level, length of visit, reason for visit, transferred by ambulance, seen in past 72 hours, injury type, pain level, vital signs, prescription, diagnosis, residency, insurance type, source of payment, location/region of hospital and many more from the time she is admitted at an ED upon her discharge.

Something that is worth noting is that patients with higher level of medical urgency receive smaller triage number. Immediacy levels in 2015 survey fall under 7 categories and are shown in table I.

TABLE I. IMMEDIACY LEVELS

| | **Description** |
|---|---|
| **0** | No Triage- Patient was dead on arrival (DOA) |
| **1** | Immediate |
| **2** | Emergent |
| **3** | Urgent |
| **4** | Semi Urgent |
| **5** | Non Urgent |
| **7** | Visit occurred in ESA (Emergency Service Area) without nursing triage |

Defined by [12], "an urgent care center is a medical clinic with expanded hours that is specially equipped to diagnose and treat a broad spectrum of *non-life and limb threatening* illnesses and injuries." These facilities are enhanced by on-site radiology and laboratory services, and work under direct supervision of physicians; they also accept unscheduled walk-in patients. Given that and based on [11], most of medical problems with immediacy level 3 or higher can be treated at urgent care facilities.

This research hypothesizes that waiting time and immediacy levels are significantly correlated. To verify that, we use linear regression models. In order to handle missing values which based on our knowledge of data set fall under Missing Not At Random (MNAR) category, we take three modeling approaches.

In the first model, we take the approach used by [2]; therefore, we simply ignore wait time missing values and drive the model. In the second approach, wait time missing values are imputed by their medians. Finally, third approach exploits values predicted by the first model to impute wait time missing data. Table II shows all three linear regression models and their respected P-values.

TABLE II. DERIVED LINEAR REGRESSION MODELS

| | **Regression Model** | **P-Value** |
|---|---|---|
| **1** | $y= 1.895736\ x+37.0000975+ \varepsilon$ | 0.004 |
| **2** | $y= 1.4333\ x+35.8737+ \varepsilon$ | 0.004 |
| **3** | $y= 0.4417328\ x+38.27636+ \varepsilon$ | 0.000 |

As expected, all three models show that there is a significant positive correlation between wait time and immediacy levels. Besides immediacy, we use forward stepwise regression method to investigate the association of 26 features with wait time. Table III lists all 13 out of 26 extracted features which are significantly correlated to wait time.

TABLE III. FEATURES CORRELATED TO WAIT TIME EXTRACTED BY FORWARD STEPWISE REGRESSION METHOD

| | **Feature** |
|---|---|
| **1** | Arrive Time |
| **2** | Transferred by Ambulance |
| **3** | Length of visit |
| **4** | Pain scale |
| **5** | If patient has been seen in past 72 hours |
| **6** | Injury |
| **7** | Method of Payment |
| **8** | Paid by patient |
| **9** | Paid by Medicaid |
| **10** | Reason for visit |
| **11** | Ethnicity |
| **12** | Region of Hospital |
| **13** | Living in metropolitan area |

## IV. DATA ANALYSIS AND EVALUATION

So far we could prove that wait time is significantly associated with the levels of immediacy. Hence, patients with non-chronic medical issues have to wait longer at EDs; this problem ultimately can lead to patients' dissatisfaction and EDs degraded performance.

Later, ANOVA test is exploited to determine whether means of wait time change in accordance with levels of immediacy. Since the resulted P-value is less than 0.05, we infer that the answer is yes. Although obtained results validate our assumptions, they don't indicate either these changes occur merely between two or amongst all immediacy levels. To have a better result/understanding, four main immediacy groups are defined. Group one contains all missing and unknown values. Second group includes records of patients who are announced dead on arrival. Group three covers all patients' information with high level of immediacy (1 or 2) and finally, fourth group contains records of non-urgent patients. Later, by utilizing Stata (statistical software) as shown in figure 1, we conclude that means of wait time change amongst all four groups.

```
. tabulate immediacy, summarize(waittime)

                         |        Summary of WAITTIME
               immediacy |      Mean   Std. Dev.       Freq.
-------------------------+------------------------------------
                     DOA | 13.834206   47.887858         573
 Missing / Unknown Value | 24.199743   70.843107        4671
              Non-Urgent | 34.568026   72.908628       14024
                  Urgent | 28.731177   60.249715        1793
-------------------------+------------------------------------
                   Total | 31.207493   71.068233       21061
```

Figure 1.Changes in means of wait time in four immediacy groups

Thus far, our methodology supports the hypothesis that "main bottleneck in EDs is caused by the presence of large number of non-urgent patients". This conclusion also can be drawn from table IV which shows weighted total of all visits to 377 U.S. hospital EDs in 2015, not including Federal, military, and Veterans Administration. Based on table IV data, about 87 million of EDs' visits represent patients with no need to immediate medical care.

TABLE IV. NUMBER AND PERCENTAGE OF WEIGHTED VISITS BASED ON IMMEDIACY LEVEL

| | Records | Weighted Visits | Percentage |
|---|---|---|---|
| **Total** | 21,061 | 136,943,181 | 100.000 |
| **-9 – Blank** | 488 | 4,799,967 | 3.505 |
| **-8 – Unknown** | 4,183 | 31,659,347 | 23.119 |
| **0- Dead on Arrival** | 573 | 2,678,029 | 1.956 |
| **1 – Immediate** | 210 | 951,193 | 0.695 |
| **2 – Emergent** | 1,583 | 10,103,506 | 7.378 |
| **3 – Urgent** | 6,605 | 40,870,483 | 29.845 |
| **4 - Semi-urgent** | 5,652 | 35,786,580 | 26.132 |
| **5- Non-Urgent** | 1,043 | 7,466,234 | 5.452 |
| **7- Visit in ESA. No Nursing Triage.** | 724 | 2,627,842 | 1.919 |

Figure 2 is derived from table IV and illustrates the distribution of four main immediacy levels. It clearly shows that 63% of EDs visits fall under non-urgent category. This number may even get larger if we consider that some of those patients with missing immediacy level (about 27%) are non-urgent patients.

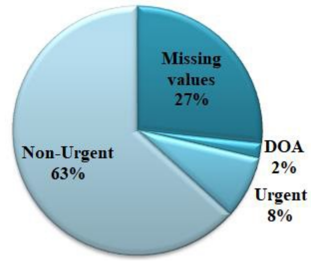


Figure 2. Distribution of four main groups of immediacy

Although patients have their own justifications to visit and prefer EDs, in this effort, we solely address one reason: "patient's inability to pay". Based on "Emergency Medical and Treatment Labor Act (EMTLA)", passed by congress in 1986, no hospitals - private and public - is allowed to refuse a patient care in an emergency due to "lack of ability to pay" [13]. Therefore, we hypothesize that patients prefer EDs due to their inability to pay for medical expenses. To verify this postulate, the correlation between wait time and patient's inability to pay should be tested. Since the frequency of some groups is 5 or less, we exploit Fisher's exact test instead of Chi square test. P-value = 0.073 > 0.05 indicates that inability to pay is not correlated to this preference. Further studies back up this conclusion as they show that 99% of all non-urgent patients who visited EDs in 2015 had at least one source of payment. Hence, we assume that a *potential* reason for overcrowded EDs is patients' lack of knowledge on the level of medical issue urgency and UCs' services and capabilities.

## V. PROPOSED SOLUTION

Thus far, this research concludes that patients prefer EDs to UCs and endure longer wait time / higher expenses due to their lack of knowledge. Hence, if we can help/inform patients to understand their medical problem urgency and provide them with the option to compare estimated wait time in UCs and EDs, then we *may* be able to keep non-urgent patients out of EDs. We bring our work to an end by implementing a web application capable of following options:

1- It lists and shows patients some common medical issues that can be treated at UCs.

2- It provides patients with searching options among all nearby healthcare facilities - both EDs and UCs - and comparing their *estimated* wait time.

3- It enables patients to schedule an online appointment at desired healthcare center.

As a patient gets into the website, an introductory page opens. This page briefly informs user by listing some common medical issues that can be treated at UCs. It also allows user to proceed with choosing either UC or ED. Then introductory page redirects patient to the main page of this application. Figure 3 demonstrates the main page which lists all nearby healthcare facilities - either UCs or EDs - and their *estimated* wait time (EDs' estimated wait time values are obtained from [5]; however, no such data is available for UCs , *therefore, wait time values are generated numbers for demonstration purpose only*). After choosing desired health care facility by clicking on "CHECK ME IN", patient will be sent to another form in which she can make an online appointment.

## VI. CONCLUSION AND FUTURE WORKS

In this research, we studied overcrowding and prolonged waiting time in U.S. emergency departments. To achieve this goal, year 2015 CDC survey data set is utilized. Data was obtained from a stratified random sampling survey that contains 21,061 electronic patients' records representing 136,943,000 visits to 377 U.S. hospital EDs, not including Federal, military, and Veterans Administration hospitals. We ground our approaches to impute missing values and proposed linear regression models. Using proposed models and ANOVA test, it was proven that wait time is significantly associated with immediacy levels. Fisher's exact test is exploited to prove that patients' inability to pay is not a main reason behind EDs' overcrowding problem. Then, we assumed that a *possible* explanation for patients' preference is their lack of knowledge on their medical urgencies and UCs capabilities. Finally, we proposed our application framework that provides patients not only with basic information on UCs potentials but also with an option to make an online appointment ahead of time.

Several questions are still left unanswered by this research. Foremost is this question that whether it is possible to utilize machine learning methods to predict EDs and UCs wait time. Currently, due to HIPAA policies, hospitals' names are masked and it is impractical to predict individual hospital wait time. In addition, our application is not employed and tested with real user/data; therefore, currently we have no means to evaluate how beneficial this solution is. In the future, we are aiming to deploy this application with real time predicted wait time values. In addition, we plan to probe patients' preference closely with respect to age, sex, ethnicity and location.

## REFERENCES


[1] ERENLER, Ali Kemal et al. "Reasons for overcrowding in the Emergency Department: Experiences and Suggestions of an Education and Research Hospital." *Turkish Journal of Emergency Medicine* 14.2 (2014): 59–63. *PMC*. Web. 05 Dec. 2017

[2] Horwitz, Leora I., Jeremy Green, and Elizabeth H. Bradley. "United States Emergency Department Performance on Wait Time and Length of Visit." *Annals of emergency medicine* 55.2 (2010): 133–141. *PMC*. Web. 20 Dec. 2017.

[3] "FastMed Urgent Care Centers | Walk-In Medical Clinics." *FastMed*. N.p., n.d. Web. 23 Jan. 2018.

[4] "Urgent Care | Online Reservations." *Carolinas HealthCare System*. N.p., n.d. Web. 23 Jan. 2018.

[5] Lena Groeger, Mike Tigas and Sisi Wei, ProPublica. Data Updated December 2017. "ER Wait Watcher." *ProPublica*. N.p., 19 Dec. 2013. Web. 10 Dec. 2017.

[6] Pines JM, Russell Localio A, Hollander JE. Racial Disparities in Emergency Department Length of Stay for Admitted Patients in the United States. *Acad Emerg Med*. 2009 Feb 24

[7] Wilper AP, Woolhandler S, Lasser KE, et al. Waits to see an emergency department physician: U.S. trends and predictors, 1997–2004. *Health Aff (Millwood)* 2008 Mar-Apr 27(2):w84–95.

[8] Gardner RL, Sarkar U, Maselli JH, Gonzales R. Factors associated with longer ED lengths of stay. *The American Journal of Emergency Medicine*. 2007 Jul;25(6):643–650.

[9] Park, Christine Y, Mary Alice Lee, and Andrew J Epstein. "Variation in Emergency Department Wait Times for Children by Race/Ethnicity and Payment Source." *Health Services Research* 44.6 (2009): 2022–2039. *PMC*. Web. 23 Feb. 2018.

[10] "National Center for Health Statistics." *Centers for Disease Control and Prevention*. Centers for Disease Control and Prevention, 31 Aug. 2017. Web. 20 Feb. 2018.

[11] *American Academy of Urgent Care Medicine (AAUCM).*N.p., n d, Web. 20 Feb. 2018.

[12] "Urgent Care Association of America (UCAOA)." Urgent Care Association of America (UCAOA). Accessed February 25, 2018. http://www.ucaoa.org/.

[13] "Overview."CMS.gov Centers for Medicare & Medicaid Services .N.p., 26 Mar. 2012. Web. 28 Feb. 2018.


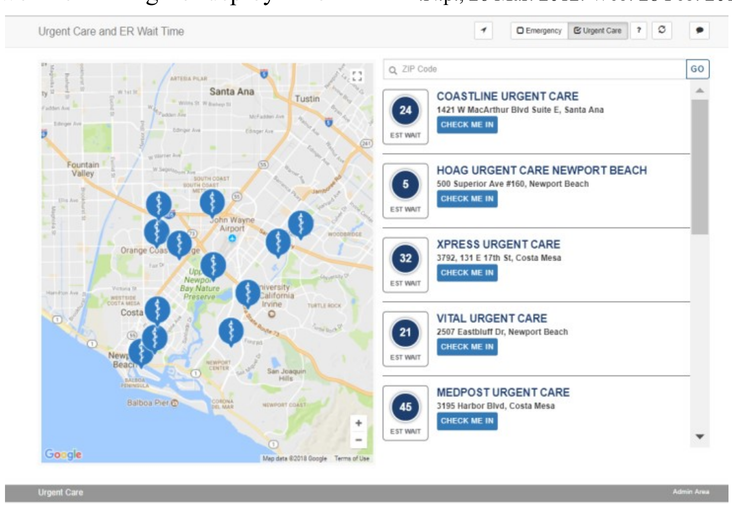


Figure 3. Demonstration of the main page - It enables patient to compare estimated wait time at nearby healthcare facilities and make an online appointment (*estimated wait time values are generated numbers for demonstration purpose only*).