Published in final edited form as:

Proc Int Conf Data Eng. 2017 April ; 2017: 1533–1540. doi:10.1109/ICDE.2017.224.

A Mortality Study for ICU Patients using Bursty Medical Events

Luca Bonomi and

University of California, San Diego, La Jolla, CA

Xiaoqian Jiang

University of California, San Diego, La Jolla, CA

Abstract

The study of patients in Intensive Care Units (ICUs) is a crucial task in critical care research which has significant implications both in identifying clinical risk factors and defining institutional guidances. The mortality study of ICU patients is of particular interest because it provides useful indications to healthcare institutions for improving patients experience, internal policies, and procedures (e.g. allocation of resources). To this end, many research works have been focused on the length of stay (LOS) for ICU patients as a feature for studying the mortality. In this work, we propose a novel mortality study based on the notion of burstiness, where the temporal information of patients longitudinal data is taken into consideration. The burstiness of temporal data is a popular measure in network analysis and time-series anomaly detection, where high values of burstiness indicate presence of rapidly occurring events in short time periods (i.e. burst). Our intuition is that these bursts may relate to possible complications in the patient's medical condition and hence provide indications on the mortality. Compared to the LOS, the burstiness parameter captures the temporality of the medical events providing information about the overall dynamic of the patients condition. To the best of our knowledge, we are the first to apply the burstiness measure in the clinical research domain. Our preliminary results on a real dataset show that patients with high values of burstiness tend to have higher mortality rate compared to patients with more regular medical events. Overall, our study shows promising results and provides useful insights for developing predictive models on temporal data and advancing modern critical care medicine.

I. Introduction

Intensive Care Units (ICUs) are responsible for admitting more than 5.7 million patients annually in the United States. While the patients represent an heterogeneous sample of the entire population, they are all characterized by the need for frequent care and greater medical assistance compare to patients hospitalized in non-ICU [1], [2], [3]. Therefore, the study of patients hospitalized in ICUs is an important field in critical care research which contributes to the definition of institutional guidances [4], [5] (i.e. resources utilization and economic impact) and identification of clinical risk factors for patients [6], [7]. Among these works, the studies in [8], [9], [10], [11] have assessed the effect of length of stay (*LOS*) on the mortality in ICU patients. The results in [8], [9] show a mortality of nearly 50% for the patients with *LOS* of 14 days or longer in ICU admissions.

In general, the *LOS* is a well known feature that is often taken into consideration during the mortality analysis of patients admitted to ICU. Nevertheless, this measure presents some limitations. A primary goal for the intensive care management is to reduce the length of stay when possible to consequently reduce cost and use of resources [12]. In fact, even though only a small fraction of the overall ICU admissions (11%) typically stays more than 7 days, these patients could use more than 50% of the overall ICU resources [13]. Furthermore, the *LOS* tends to be influenced by additional factors such as: type and severity of illness, geographical location and type of hospital, stress and depression among patients' family [12]. As a consequence, the solely use of the *LOS* in constructing predictive models may lead to inadequate guidances.

In this work, we propose a different approach for studying the mortality rate of patients in ICU that it is orthogonal to the traditional analysis based on the *LOS*. Specifically, we take into consideration the patients longitudinal data (i.e. sequence of visits to ICU) and using their temporal information we aim to understand if there exists a relationship between the mortality rate and the temporal distribution of the visits. Our approach is motivated by the fact that the temporal distribution of the medical events may carry information about the severity of the illness. For example, regular visits may suggest a chronic disease (e.g. diabetes), while saddle chances may indicate the presence of complications. To capture this temporal information, we adopt the concept of *burstiness* that is commonly used in network systems. Intuitively, high values of burstiness indicate the presence of rapidly occurring events in short time periods (i.e. burst). Our preliminary results on a real dataset show that patients with bursty sequences present higher mortality rate compared to patients with regular temporal sequences. Furthermore, we show that there is no or little correlation between the burstiness and the length of stay for the patients in ICU. Hence, burstiness can be used in addition to the length of stay to design more robust and accurate statistical models for patients admitted in ICU.

Contributions

Our major contributions are reported below.

- We conduct a mortality study based on the temporal information of the patients longitudinal data. To capture such information, we adopt the burstiness notion from network systems analysis and time-series anomaly detection.
- To the best of our knowledge, we are the first to apply the notion of burstiness to the critical care research domain.
- We conduct an extensive evaluation on a real medical dataset. Our preliminary results show that the mortality for bursty patients is 59.8% compared to 52.9% observed in the overall dataset. Furthermore, we combine the length of stay and burstiness to study possible risk factors among ICU patients.

A. Related Work

A wide variety of research has been proposed in the domain of critical care medicine to study the medical, economical, and social aspects of patients hospitalized in ICUs. Among these works, those in [8], [9], [10], [11] investigate the relationship between the length of

stay for patients in ICU and their mortality rate; therefore, they are most related to the problem considered in this paper. In this work, we conduct a similar study, where we propose to apply the concept of burstiness on the temporal sequence of visits associated with ICU patients. Our goal is to determine possible connections between the burstiness of medical events and the mortality rate.

The concept of burstiness has been extensively studied in the computer science community with applications in networking and time-series anomaly detection [14], [15]. The burstiness aims to model the presence of high frequency events in temporal data which often has been used to detect features of interest. For example in the work proposed by Matchioudakis and Koudas [16], bursty keywords are detected in the Twitter stream to identify trends and perform analysis. In [17], the authors investigate the problem of extracting “hot” bursty events from text stream which are considered as potential features for constructing a text classifier.

II. Methods

In this section, we report some basic notions about the temporal representation of the medical data and how the burstiness parameter is defined.

A. Temporal Model

In this work, we represent patients longitudinal data as temporal sequences of medical events recorded over time. For example, in a case of a patient that has recurrent hospitalizations, each visit represents a medical event associated with a specific time stamp recorded by the healthcare institution.

Definition 1 (Temporal Sequence of Events)—We define a temporal sequence $X = \{(X_i, t_i)\}_{i=1}^n$ as an ordered sequence of n medical events X_i each associated with a positive number t_i that represents the time in which the event X_i has been recorded (e.g. admission time at the hospital). Furthermore, we assume $t_i < t_{i+1}$ to maintain the temporal order among the events.

This representation allows us to model medical events with different levels of granularity. In fact, in a typical medical setting a variety of features (e.g. diagnosis codes, lab tests, medications) are available at each patient’s visit and can be considered in our formulation. However, in this work, we are more interested in computing statistics on the temporal information of the events rather than considering their specific medical information. Our preliminary results demonstrate that even considering the temporal information associated with the patients’ visits suffices to provide significant indications on their mortality. Therefore, considering medical information at a finer level of granularity (e.g. diagnoses, medications) may further improve our results.

B. Interevent Time Distribution

We denote by $\mathbf{t}(X) = [t_1, t_2, \dots, t_n]$, the vector of time occurrences for the temporal sequence $X = \{(X_i, t_i)\}_{i=1}^n$. As we mentioned before, in our analysis we will focus on the interevent

times $T_i = t_i - t_{i-1}$, for $i = 2, \dots, n$. Then $T(X) = T_1, T_2, \dots, T_n$, where $T_0 = 0$, denotes the interevent time sequence associated with the events in the sequence X . Specifically, these sequences provide information about the time interval between two consecutive events in the original patients longitudinal data.

Under this model, we aim to analyze the interevent time sequences for the ICU patients to study possible relationships with their mortality. In fact, similarly to the length of stay (*LOS*), the temporal information in the interevent time sequences may provide indications on the severity of the patients illness. Consider for example a first patient that has regular visits and a second patient that presents saddle changes of frequency in his/her visits. Due to the regularity of the interevent time sequence, the first patient may be associated with a chronic disease (e.g. diabetes) that requires periodic check-ups. On the other hand, the irregularity in the second patient may reveal the presence of critical events in his/her medical condition or possible complications.

To capture such information from the interevent time distributions, we use the notion of *burstiness* [18]. This notion captures the presence of bursty behaviors in temporal sequences and for this reason it has been extensively applied in dynamic network analysis to detect the presence of rapidly occurring events in short time periods alternating with long inactive periods (i.e. burst) [14], [15], [16], [17]. Formally, the burstiness for interevent time sequences is defined as follows.

Definition 2 (Burstiness Parameter)—Let $T = T_1, T_2, \dots, T_n$ be a interevent time sequence with standard deviation σ and mean μ . Then, the burstiness parameter for T is computed as:

$$B = \frac{\sigma - \mu}{\sigma + \mu} \quad (1)$$

The burstiness parameter B has values ranging between -1 and 1 . Specifically, $B = -1$ indicates a periodic sequence, $B = 0$ indicates a Poisson distribution interevent sequence, and B approaches 1 for extremely bursty sequences [18].

In our scenario, we compute the burstiness parameter of each interevent time sequence with the goal of understanding how medical events distributed in time may relate to the mortality rate. Our intuition is that the presence of bursty events in patients longitudinal data may indicate sudden complications. Hence, providing significant indications on the mortality.

III. Results

In this section, we conduct our experimental study on the mortality rate in ICU patients considering both the length of stay (*LOS*) and the burstiness parameter (B).

A. Setting

Dataset—We conduct our experiments on the real-world medical dataset MIMIC-III which is a de-identified dataset comprising over 58,000 hospital admissions for 38,645

adults and 7,875 neonates [19]. Each ICU admission is represented as an event recording the ICD9 codes associated with the patient at the moment of hospitalization. Then for each patient, we consider the interevent time sequence associated with the sequence of his/her hospitalizations. From the overall sequences, we discard those that comprise less than 2 observations. Then, the resulting dataset contains 6,577 sequences.

In our evaluations, we partition the ICD9 codes into 19 disjoint diagnostic groups and for each patient's visit the primary diagnosis code is associated to one of these groups. This stratification of the primary diagnosis codes allows us to understand how the distribution of each group changes according to our parameters (i.e. burstiness and length of stay) with respect to the overall data. Furthermore, in Table I, we report the ICD9 codes together with their description that often appear in our results. Figure 1 reports the top-10 most frequent diagnosis codes in the overall dataset. From the results obtained on the overall patients, we observe that Hypertension (401.9) is the most common diagnosis which is shared among more than 55% of the all ICU patients.

Variables in our analysis—In our evaluations, we consider two main variables: the length of stay (LOS) and burstiness parameter (B). We are interested in study the relationships between these variables and the mortality of patients in the MIMIC-III dataset. For each patient's interevent time sequence, the burstiness parameter is computed as in Definition 2. Since, in the original dataset a patient has multiple admissions to ICU, we compute the length of stay (LOS) as the maximum length of stay recorded among all his/her visits. While previous research has been conducted on understanding the relationship between the LOS and mortality rate for ICU patients [8], [9], [10], [11], here we are the first to use the notion of burstiness in this setting. To understand if this parameter carries additional information compared to the LOS , we first evaluate if there is any correlation between the LOS and B . The result of the Pearson's test on the population of patients associated with mortality cases leads to $\rho = 0.024$ which indicates little or no correlation between the LOS and B parameter. The dependency between these two variables is reported in Figure 2. From these results, we observe that the burstiness parameter provides orthogonal information about the mortality rate of the patients which is not directly correlated to the length of stay.

B. Burstiness Results

In this section, we report the results obtained by using the burstiness parameter B on the interevent time sequences obtained from the visits in MIMIC-III data. Specifically, we divide the patients according to their burstiness values into three groups. The first group contains patients with $B = -1$ which are associated with temporal sequences that are periodic or have only two visits. Since in MIMIC-III a considerable number of patients have only two visits, this group counts for 68.8% of the overall patients. In the second group, we consider patients with $-1 < B \leq 0$ and they represent 26.2% of the overall patients. These patients are associated with interevent sequences that approximately follow a Poisson distribution. Lastly, we consider the group of patients with $0 < B \leq 1$ that tend to have very bursty interevent time sequences. In this case, only 5% of the overall patients belong to this group. For these groups of patients, we report their cohort characteristics in Table II. From

these results, we observe that there are no significant changes across these groups in terms of gender distribution compared to the overall data. However, we notice an increment in the mortality rate as the burstiness parameter increases. In fact, for the last two groups we have a mortality of 59.1% and 59.8% compared to the 52.9% observed on the overall dataset. Furthermore, for the group with $B = -1$ we obtain a lower mortality rate (50.0%) compared to the entire population. This suggests that patients with more bursty interevent time sequences tend to have an higher mortality rate than those that have more regular/periodic visits.

From the primary diagnostic groups reported in Table II, we observe that as the burstiness parameter increases the percentage of patients diagnosed with: Endocrine, Nutritional and Metabolic Diseases and Immunity disorders, Circulatory, Respiratory, and Genitourinary diseases tend to increase. For example, the percentage of patients diagnosed with codes associated to diseases of the respiratory system increases to 44.79% in the last group from 22.49% observed in the overall dataset. Furthermore, we observe that patients diagnosed with diseases related to Conditions originating in the perinatal period tend to decrease as the interevent sequences are more bursty. In fact, from 1.44% of overall patients with these primary diagnoses, no patients with these conditions are reported in the group with $B = -1$.

In Figure 3, we report the top-10 diagnosis codes across the three groups of patients. For the group of patients with $B = -1$ (Figure 3a), we observe that their most popular codes present minimal changes compared to those obtained from the overall data (Figure 1). As in the overall dataset, the most frequent code has a considerably higher support with respect to the other diagnoses. Similarly, to the results from the overall dataset, Hypertension (401.9) is the most popular diagnosis and it is shared across 52.5% of the patients. In the second group representing patients with $-1 < B \leq 0$, we observe an overall increment of the support for the top-10 codes. Even though Hypertension (401.9) is still the most popular diagnosis (58%), Congestive Heart Failure (428.0) and Acute Kidney Failure (584.9) are shared among 57.2% and 54.6% of the patients respectively. For the patients with very bursty sequences, $0 < B \leq 1$, the overall support for the top-10 codes further increases. In this group, the most common diagnosis code is associated with Congestive Heart Failure (428.0) which is shared among 68% of the patients. In this case, Kidney Failure (584.9) is reported as the second most common diagnosis shared across 64.7% of the patients, followed by Hypertension (401.9) which is reported in 62.2% of the patients.

These results point out some interesting trends across the diagnosis codes in these groups. First, we observe that from both results in Table II and Figure 3 as the burstiness increases a greater percentage of patients tends to be associated with respiratory system diagnosis codes. For example, acute respiratory failure (518.81) which is commonly shared across 32.6% of the overall patients, it reaches 60.1% of the patients in the last group. Similarly, Pneumonia (486) increases from 25.3% on the overall dataset to 48.2% in the last group of patients. Another interesting trend is related to diagnosis codes associated with the circulatory system. Even though, from Table II the percentage of diagnosed patients increases, by looking in Figure 3 some codes change more than others. For example, while the percentage of patients with Congestive Heart Failure (428.0) significantly increases, those with Atrial Fibrillation (427.31) show a different trend. In fact, Atrial Fibrillation (427.31) is ranked

third in the first group with a percentage of 34.7%, while for bursty patients this percentage only reaches 44.4%, and it is ranked last in the top-10 codes for this group.

C. Length of Stay Results

Here, we consider the impact of the length of stay (*LOS*) as a parameter in our mortality study. As the related works in [8], [9], [10], [11] pointed out, the *LOS* provides important indications on the mortality rate of patients in ICU. Similarly to the analysis conducted in [10], [11], we partition the patients in two groups: those with maximum *LOS* < 10 days and those with *LOS* ≥ 10 days. From the overall patients, the proportion of patients who stayed less than 10 days (76.4%) is considerably greater than those who had *LOS* ≥ 10 days (23.6%). The cohort characteristics for these groups are reported in Table III. By looking at the mortality rate across of these sets of patients, we observe that those who stayed less than 10 days have a relatively smaller mortality rate (49.8%) compared to the entire population (52.9%), while for those who stayed in ICU for more than 10 days their mortality rate increases up to 62.6%.

Figure 4 reports the top-10 diagnosis codes across the two groups of patients in terms of their support (i.e. percentage of patients). For the set of patients with *LOS* < 10, the top-10 diagnoses (Figure 4a) present small changes compared to the top-10 codes from the overall patients (Figure 1). As in the overall database, the majority of the patients in this group (55.7%) are diagnosed with Hypertension (401.9). We observe that for patients who stayed more than 10 days in ICU, the distribution of frequent diagnosis codes (Figure 4b) has several differences compared to the distribution of the codes in the previous group. All the top-10 codes tend to have higher support than those in the previous group indicating a more uniform distribution of the codes across the patients. Furthermore, in this case the majority of the patients (57.1%) are diagnosed with Acute Respiratory Failure (518.81), while Hypertension is only ranked third among the most popular diagnoses. Furthermore for patients with *LOS* ≥ 10, we observe the presence of Severe Sepsis (995.92) in the top-10. This increment of patients diagnosed with sepsis is in line with the results in Table III. In fact, for patients with *LOS* ≥ 10 days the percentage of patients with primary diagnostic group in Injury and Poisoning (ICD9 codes between 800 and 999) grows to 30.93% from 26.87% obtained considering the overall dataset.

D. Combining Burstiness and Length of Stay

Here, we use both burstiness and length of stay to conduct our study by grouping the patients in four distinct groups as follow. Group 1 consists in patients with *LOS* < 10 days and $-1 < B \leq 0$ (18.4%), patients with *LOS* < 10 days and $0 < B \leq 1$ are in Group 2 (3.0%) while those with *LOS* ≥ 10 days and $-1 < B \leq 0$ are in Group 3 (7.8%). Finally, Group 4 contains the patients with *LOS* ≥ 10 days and $0 < B \leq 1$ (2.0%). Patients with $B = -1$ are omitted in this analysis because they have similar cohort characteristics with the overall dataset, hence they do not present specific characteristics.

Table IV reports the cohort characteristics for these four groups. From the mortality rate, we observe that Group 3 and Group 4 have the highest mortality rates, respectively 67% and 65.9% compared to 52.9% measured in the overall database. Furthermore, we observe that

bursty sequences are associated with patients whose primary diagnosis codes are related to the circulatory and respiratory system. For example, for the two groups of patients with *LOS* 10 days, patients in Group 4 (i.e. high burstiness values) tend to have a higher percentage of primary codes associated with the circulatory system (63.57%) compared to the patients in Group 3 (48.64%). Considering the gender distribution, we observe that Group 4 contains a considerably larger fraction of male patients compared to the other groups. This change in the gender distribution is also noticeable across the primary diagnosis codes. In fact, in this group no patients are associated with Pregnancy, Childbirth and Puerperium, and Condition originated in the perinatal period diseases, which are typically associated to female patients.

From the single diagnosis code distribution in Figure 5, we observe a general trend that leads to an increment of the support value for each code in the top-10 diagnosis codes similarly to the results obtained considering *LOS* and *B* separately. For patients with *LOS* < 10 days and time sequences with regular behavior (Figure 5a), the common diagnosis codes are Hypertension (401.9), Congestive heart failure (428.0), and Acute Kidney Failure (584.9). The support for these codes tend to be considerable higher compared to the rest of the codes in the top-10. For patients with *LOS* < 10 days and bursty interevent sequences (Figure 5b), we observe that the support distribution for the top-10 codes tends to become smoother indicating a weaker separation among the leading codes. In the groups of patients with *LOS* 10 days, we observe a considerable increment of support for the code associated with Acute Respiratory Failure (518.81) both in the case of patients with regular (Figure 5c) and bursty interevent sequences (Figure 5d). Among the top codes, their supports in the case of bursty sequences tend to be higher compared to more regular sequences suggesting that more patients share a combinations of multiple diseases.

Finally, we consider a summary of six important diagnostic groups: Neoplasms, Endocrine, Nutritional and Metabolic Diseases and Immunity Disorders, Nervous Systems, Circulatory System, Respiratory System and Genitourinary System. For these groups, we report in Figure 6 the percentage of patients recorded with any primary code related to these diagnoses across all the four groups. Among these diagnosis codes, those related to circular and respiratory system dominate the population of patients in each group. Specifically, patients with bursty interevent sequences (Group 2 and Group 4) tend to have more cases of diseases related to circulatory system compared to the groups with more regular interevent sequences (Group 1 and Group 3). Regarding diseases related to the respiratory system, we observe a steady increment across the groups as the length of stay and burstiness parameter increase. In Group 2 (*LOS* < 10 days and regular interevent sequences), we observe a large number of patients who are recorded with diagnosis codes related to Nutritional and Metabolic Diseases and Immunity Disorders. Among all the four groups, there are no significant changes among patients diagnoses with Neoplasms and Nervous system related diseases. Finally, we notice an increment of patient associated with Genitourinary system diseases among the patients with bursty interevent sequences (Group 2 and Group 4) which is also reflected in higher Acute Kidney Failure (584.9) cases as reported in Figure 5.

IV. Conclusion

In this paper, we presented some preliminary results that showed an interesting relationship between the mortality and the burstiness of interevent time sequences for ICU patients. Specifically, our results showed that the mortality for bursty patients reaches 59.8% compared to 52.9% observed the overall patients. Furthermore, this parameter is orthogonal to the feature currently used in critical care mortality studies (i.e. length of stay). Therefore, the burstiness parameter provides new information that can be effectively used to improve state of the art prediction models. Future research directions consist in conducting a more in depth medical analysis of specific diseases assisted by clinicians and incorporating temporal features for developing new predictive risk models.

Acknowledgments

This work was supported by NIH awards R00LM011392 and R21LM012060.

References

1. Wunsch H, Angus DC, Harrison DA, Linde-Zwirble WT, Rowan KM. Comparison of medical admissions to intensive care units in the united states and united kingdom. *American journal of respiratory and critical care medicine*. 2011; 183(12):1666–1673. [PubMed: 21471089]
2. Barrett M, Smith M, Elixhauser A, Honigman L, Pines J. Utilization of intensive care services, 2011. HCUP statistical brief. 2014; 185
3. Society of critical care medicine: Critical care statistics. [Accessed 2017-01-21]. [Online]. Available: <http://www.sccm.org/Communications/Pages/CriticalCareStats.aspx>
4. DiGiovine B, Chenoweth C, Watts C, Higgins M. The attributable mortality and costs of primary nosocomial bloodstream infections in the intensive care unit. 1999; 160(3):976–981. [Online]. Available: <http://dx.doi.org/10.1164/ajrccm.160.3.9808145>.
5. Dasta JF, McLaughlin TP, Mody SH, Piech CT. Daily cost of an intensive care unit day: the contribution of mechanical ventilation. *Critical care medicine*. 2005; 33(6):1266–1271. [PubMed: 15942342]
6. Blumberg HM, Jarvis WR, Soucie JM, Edwards JE, Patterson JE, Pfaller MA, Rangel-Frausto MS, Rinaldi MG, Saiman L, Wiblin RT, et al. Risk factors for candidal bloodstream infections in surgical intensive care unit patients: the nemis prospective multicenter study. *Clinical Infectious Diseases*. 2001; 33(2):177–186. [PubMed: 11418877]
7. Barr J, Fraser GL, Puntillo K, Ely EW, Gélinas C, Dasta JF, Davidson JE, Devlin JW, Kress JP, Joffe AM, et al. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. *Critical care medicine*. 2013; 41(1):263–306. [PubMed: 23269131]
8. Ryan TA, Rady MY, Bashour CA, Leventhal M, Lytle B, Starr NJ. Predictors of outcome in cardiac surgical patients with prolonged intensive care stay. *Chest*. 1997; 112(4):1035–1042. [Online]. Available: + <http://dx.doi.org/10.1378/chest.112.4.1035>. [PubMed: 9377914]
9. Wong DT, Gomez M, McGuire GP, Kavanagh B. Utilization of intensive care unit days in a canadian medical-surgical intensive care unit. *Critical care medicine*. 1999; 27(7):1319–1324. [PubMed: 10446826]
10. Williams T, Ho K, Dobb G, Finn J, Knuiman M, Webb S. Effect of length of stay in intensive care unit on hospital and long-term mortality of critically ill adult patients. *British journal of anaesthesia*. 2010; 104(4):459–464. [PubMed: 20185517]
11. Philbin EF, Rogers VA, Sheesley KA, Lynch LJ, Andreou C, Rocco TA. The relationship between hospital length of stay and rate of death in heart failure. *Heart & Lung: The Journal of Acute and Critical Care*. 1997; 26(3):177–186. [PubMed: 9176685]

12. Gruenberg DA, Shelton W, Rose SL, Rutter AE, Socaris S, McGee G. Factors influencing length of stay in the intensive care unit. *American Journal of critical care*. 2006; 15(5):502–509. [PubMed: 16926372]
13. Stricker K, Rothen HU, Takala J. Resource use in the icu: short- vs. long-term patients*. *Acta Anaesthesiologica Scandinavica*. 2003; 47(5):508–515. [Online]. Available: <http://dx.doi.org/10.1034/j.1399-6576.2003.00083.x>. [PubMed: 12699506]
14. Kleinberg J. Bursty and hierarchical structure in streams. in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining ACM*. 2002:91–101.
15. Lazarevic A, Ertöz L, Kumar V, Ozgur A, Srivastava J. A comparative study of anomaly detection schemes in network intrusion detection. *Proceedings of the 2003 SIAM International Conference on Data Mining SIAM*. 2003:25–36.
16. Mathioudakis, M., Koudas, N. *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, ser SIGMOD '10*. New York, NY, USA: ACM; 2010. Twittermonitor: Trend detection over the twitter stream; p. 1155–1158.[Online]. Available: <http://doi.acm.org/10.1145/1807167.1807306>
17. Fung, GPC., Yu, JX., Yu, PS., Lu, H. Parameter free bursty events detection in text streams; *Proceedings of the 31st International Conference on Very Large Data Bases, ser VLDB '05 VLDB Endowment*. 2005. p. 181–192.[Online]. Available: <http://dl.acm.org/citation.cfm?id=1083592.1083616>
18. Goh K-I, Barabási A-L. Burstiness and memory in complex systems. *EPL (Europhysics Letters)*. 2008; 81(4):48002.
19. Johnson AE, Pollard TJ, Shen L, Lehman L-wH, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. Mimic-iii, a freely accessible critical care database. *Sci Data*. 2016; 3

Fig. 1.
Top-10 diagnosis codes in MIMIC-III.

Fig. 2.
Scatter plot between the length of stay (LOS) and burstiness parameter (B).

Fig. 3. Top-10 diagnosis codes ranked with respect their support across groups of patients in MIMIC-III with different values of burstiness B .

Fig. 4.

Top-10 diagnosis codes ranked with respect to their support across groups of patients in MIMIC-III with different values of length of stay LOS .

Fig. 5.

Top-10 diagnosis codes ranked with respect their support across groups of patients in MIMIC-III with different values of length of stay LOS and burstiness B .

Fig. 6.
Distribution of the patients among five important diagnostic groups across the partitions.

TABLE I

ICD9 Codes and description of the main diagnoses used in our experimental analysis.

ICD9 Code	Description
401.9	Hypertension NOS
428.0	Congestive heart failure
584.9	Acute kidney failure
414.01	Coronary atherosclerosis
427.31	Atrial fibrillation
599.0	Urinary tract infection
518.81	Acute respiratory failure
272.4	Hyperlipidemia
250.00	Diabetes without complication, type II or unspecified type
486	Pneumonia
285.9	Anemia
276.2	Acidosis
995.92	Severe sepsis
389	Hearing loss

TABLE II

Cohort characteristics for the three subgroups of patients according with different burstiness (B) parameters: $B = -1$, $-1 < B \leq 0$, and $0 < B \leq 1$.

BURSTINESS	ALL (N=6577)	$B = -1$ (N=4525)	$-1 < B \leq 0$ (N=1726)	$0 < B \leq 1$ (N=326)
Gender (%)				
Male	55.7	55.7	54.3	54.6
Female	44.3	44.3	45.6	45.4
Mortality (%)	52.9	50.0	59.1	59.8
Primary Diagnostic group (%)				
001–139: Infectious and Parasitic	21.77	16.44	32.38	39.57
140–239: Neoplasms	11.81	12.33	11.18	7.98
240–279: Endocrine, Nutritional, Metabolic and Immunity	6.66	4.33	10.89	16.56
280–289: Blood and Blood-forming Organs	1.52	1.17	2.20	2.76
290–319: Mental Disorders	1.69	1.41	2.38	1.84
320–359: Nervous System	3.18	2.34	4.29	8.90
360–389: Sense Organs	0.08	0.07	0.06	0.31
390–459: Circulatory System	49.32	47.8	80.98	61.66
460–519: Respiratory System	22.49	16.53	33.89	44.79
520–579: Digestive System	23.25	18.92	32.21	35.89
580–629: Genitourinary System	6.75	4.80	10.14	15.95
630–679: Pregnancy, Childbirth, and Puerperium	0.12	0.11	0.12	0.31
680–709: Skin and Subcutaneous Tissue	1.09	0.86	1.51	2.15
710–739: Musculoskeletal System and Connective Tissue	3.03	2.36	4.17	6.13
740–759: Congenital Anomalies	0.70	0.71	0.64	0.92
760–779: Conditions originating in the perinatal period	1.44	1.99	0.29	0
780–799: Symptoms, Signs, and ill-defined conditions	3.30	2.43	4.58	8.59
800–999: Injury and Poisoning	26.87	22.87	34.18	43.56
E–V: External Causes of Injury	2.69	3.07	1.74	2.45

TABLE III

Cohort characteristics for the two subgroups of patients according with different length of stay (*LOS*): *LOS* < 10, and *LOS* ≥ 10.

LENGTH OF STAY	ALL (N=6577)	LOS < 10 (N=5022)	LOS ≥ 10 (N=1555)
Gender (%)			
Male	55.7	55.4	56.8
Female	44.3	44.6	43.2
Mortality (%)	52.9	49.8	62.6
Primary Diagnostic group (%)			
001–139: Infectious and Parasitic	21.77	19.37	29.52
140–239: Neoplasms	11.81	12.68	9.00
240–279: Endocrine, Nutritional, Metabolic and Immunity	6.66	7.18	4.95
280–289: Blood and Blood-forming Organs	1.52	1.47	1.67
290–319: Mental Disorders	1.69	2.03	0.58
320–359: Nervous System	3.18	2.99	3.79
360–389: Sense Organs	0.08	0.08	0.06
390–459: Circulatory System	49.32	50.42	45.79
460–519: Respiratory System	22.49	19.01	33.44
520–579: Digestive System	23.25	23.12	23.67
580–629: Genitourinary System	6.75	6.73	6.82
630–679: Pregnancy, Childbirth, and Puerperium	0.12	0.12	0.13
680–709: Skin and Subcutaneous Tissue	1.09	1.08	1.16
710–739: Musculoskeletal System and Connective Tissue	3.03	2.93	3.34
740–759: Congenital Anomalies	0.70	0.56	1.16
760–779: Conditions originating in the perinatal period	1.44	0.62	4.12
780–799: Symptoms, Signs, and ill-defined conditions	3.30	3.33	3.22
800–999: Injury and Poisoning	26.87	25.61	30.93
E–V: External Causes of Injury	2.69	1.61	6.17

TABLE IV

Cohort characteristics for the four subgroups of patients according with different burstiness (B) and length of stay (LOS).

GROUPS	ALL (N=6577)	-1 < B Group 1 (N=1210)	0 < B Group 2 (N=197)	-1 < B Group 3 (N=516)	0 < B Group 4 (N=129)
General Statistics					
Average LOS		4.54	5.25	22.71	22.71
Average B		-0.30	0.11	-0.29	0.11
Gender (%)					
Male	55.7	55.1	50.8	52.5	60.5
Female	44.3	44.9	49.2	47.5	39.5
Mortality (%)					
	52.9	54.5	55.8	67.0	65.9
Primary Diagnostic group (%)					
001-139: Infectious and Parasitic	21.77	28.35	38.07	41.86	41.86
140-239: Neoplasms	11.81	12.15	7.11	8.91	9.30
240-279: Endocrine, Nutritional, Metabolic and Immunity	6.66	12.31	20.30	7.56	10.85
280-289: Blood and Blood-forming Organs	1.52	2.48	2.03	1.55	3.88
290-319: Mental Disorders	1.69	3.22	2.03	0.39	1.55
320-359: Nervous System	3.18	3.88	9.14	5.23	8.53
360-389: Sense Organs	0.08	0.08	0.51	0	0
390-459: Circulatory System	49.32	51.98	60.41	48.64	63.57
460-519: Respiratory System	22.49	28.68	38.58	46.12	54.26
520-579: Digestive System	23.25	33.47	33.50	29.26	39.53
580-629: Genitourinary System	6.75	9.83	17.26	10.85	13.95
630-679: Pregnancy, Childbirth, and Puerperium	0.12	0.17	0.51	0	0
680-709: Skin and Subcutaneous Tissue	1.09	1.40	1.52	1.74	3.10
710-739: Musculoskeletal System and Connective Tissue	3.03	3.64	6.60	5.43	5.43
740-759: Congenital Anomalies	0.70	0.50	1.02	0.97	0.78
760-779: Conditions originating in the perinatal period	1.44	0	0	0.97	0
780-799: Symptoms, Signs, and ill-defined conditions	3.30	4.96	8.63	3.68	8.53
800-999: Injury and Poisoning	26.87	31.32	41.12	40.89	47.29

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

GROUPS	E-V: External Causes of Injury			
	ALL (N=6577)	-1 < B 0 AND LOS < 10 Group 1 (N=1210)	0 < B 1 AND LOS < 10 Group 2 (N=197)	-1 < B 0 AND LOS Group 3 (N=516)
	2.69	1.24	2.54	2.91
				2.33