

Spatio-temporal Structure of US Critical Care Transfer Network

K.P. Unnikrishnan
University of Michigan
Ann Arbor, MI 48109
kpuk@umich.edu

Debprakash Patnaik
Virginia Tech
Blacksburg, VA 24060
patnaik@vt.edu

Theodore J. Iwashyna
University of Michigan
Ann Arbor, MI 48109-5419
tiwashyn@umich.edu

Abstract

Most Americans are in Intensive Care Units (ICUs) at some point during their lives. There is wide variation in the outcome quality of ICUs and so, thousands of patients who die each year in ICUs may have survived if they were at the appropriate hospital. In spite of a policy agenda from IOM calling for effective transfer of patients to more capable hospitals to improve outcomes, there appear to be substantial inefficiencies in the existing system. In particular, patients recurrently transfer to secondary hospitals rather than to a most-preferred option. We present data mining schemes and significance tests to discover these inefficient cascades. We analyze critical care transfer data in Medicare across nearly 5,000 hospitals in the United States over 10 years and present evidence that these transfers to secondary hospitals repeatedly cascade across multiple transfers, and that some hospitals seem to be involved in many cascades.

Keywords: Critical care, cascades, data mining, alerts, transfer networks, administrative data, Medicare claims

1 Introduction

The intensive care unit is the apogee of modern technologically-intensive medical care, providing invasive support of many failing organs and offering potentially profound life-saving ability. It is also in relatively short supply, and inadequate availability of critical care beds is a salient policy concern. In order to distribute some of the load of critically ill patients, informal transfer networks have evolved over which patients are routinely transferred between hospitals. But this system is not managed or overseen at the present time, and there is anecdotal evidence that it may have important limitations. Potential overload of the critical care system becomes all the more pressing in the event of disasters – correlated stresses on multiple parts of the system that may result from acute events (such as terrorist attacks or extreme weather) or subacute stresses (such as pandemic flu). During routine functioning of the critical care transfer system, most hospitals have a primary transfer destination to which the plurality of their patients are sent [1]. These primary transfers seem to be favored by routine practice at the hospitals

– deviations from the primary transfer are likely to slow down patient care and represent substantial additional delays in care. In principle, a non-primary destination for one hospital might be a primary destination for another. Thus non-primary transfers might use up excess capacity, and result in other hospitals unable to transfer patients to their own primary hospitals. The goal of this paper is to develop a rigorous approach to detecting such cascades in re-oriented network activity.

2 Data Characteristics

In this study, we use the final action claims from the 1996 through 2005 Medicare Provider Analysis and Review (MedPAR) file. Medicare claims and enrollment data capture 96% of the American population aged 65 and older [2]. Critical care use is indicated directly in the hospital-filed claims. Although these data have some limitations [3], they have been used by our group [4] and others [5] to measure critical care at both the patient and national levels. We excluded beneficiaries covered under certain types of group health organizations with capitated premiums who are not required to file claims. We excluded from critical care “psychiatric critical care” and intermediate or step-down units [3], but included medical, surgical, cardiac and burn units. Transfers between hospitals are not directly indicated in the claims. We defined a transfer as occurring between two hospitals g and h when a patient was observed to be in Hospital g until a certain day, and then in Hospital h beginning on the same day or the next day. Critical care transfers were defined as transfers that occurred between two hospitalizations, both of which involved critical care use. Thus the data \mathcal{S} is structured as a date-stamped list of ICU transfers, with uniquely identified sending and receiving hospitals.

$$\mathcal{S} = \langle (g_1 \rightarrow h_1, t_1), (g_2 \rightarrow h_2, t_2), \dots, (g_n \rightarrow h_n, t_n) \rangle$$

where $g_i, h_i \in \mathcal{H}$, the set of all hospitals, $(g_i \rightarrow h_i)$ indicates an ICU transfer from hospital g_i to hospital h_i at time t_i . A snippet of the raw data is shown in Table 1. Some of the statistics related to the ICU transfers are listed in Table 2. Figure 1 shows the distribution of transfers in six month time windows.

Table 1: A segment of ICU transfer data. Each hospital is assigned an unique random identifier.

transfer-date	sending-hospital-id	receiving-hospital-id	type
18 May 99	1170	2468	CARDIAC
27 Nov 00	2911	2468	CVSURG
11 Mar 03	1170	2468	CARDIAC
02 Jun 04	3155	2468	CARDIAC

Table 2: Data statistics

Date range	01-Oct-95 - 31-Dec-06
Total number of hospitals	5,083
Total number of transfers	765,171
Total number of unique pairs	62,529
Distribution of transfers	(see Figure 1)

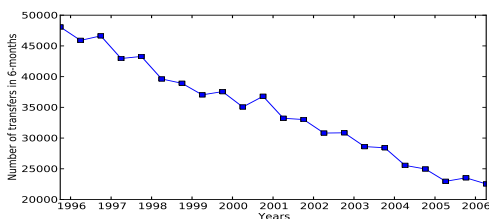


Figure 1: Plot of the number of ICU transfers from 1996 to 2006 shown in windows of six months. Note the substantial reduction in the overall number of transfers.

3 Data Mining Methodology

As mentioned in the Introduction section, our hypothesis is that the secondary transfers may cascade - that the use of a secondary destination by one hospital may make secondary transfers by other nearby hospitals more likely. In this section we propose an efficient data mining scheme and associated tests of significance for discovering such cascades.

Definition 1. A primary transfer pair consists of a pair of hospitals ($g \rightarrow h$) such that the number of transfers from g to h exceeds the number of transfers from g to any other hospital \hat{h} , $\forall \hat{h} \in \mathcal{H}$ and $\hat{h} \neq h$. The hospital h will be referred to as the primary recipient for g , denoted by $recv(g) = h$.

From the above definition, every hospital has a primary transfer partner that receives most of its ICU transfers. And there is a set of hospitals where each hospital receives transfers from one or more hospitals and acts as their primary recipient hospital. Let $send(h)$ denote the set of hospitals for which h is the primary receiving partner.

$$send(h) = \{g : recv(g) = h, g \in \mathcal{H}\}$$

Definition 2. A secondary transfer refers to an ICU transfer from hospital g to \hat{h} , where \hat{h} is not the primary recipient for g i.e. $\hat{h} \neq recv(g)$.

Definition 3. A cascade is defined in terms of an ordered list of hospitals. A k -size cascade is denoted by $\langle h_1, h_2, \dots, h_k \rangle$ where $h_i \in \mathcal{H}$, the set of all hospitals in the data. An occurrence of a cascade is characterized by a sequence of transfers in the data (but not necessarily adjacent):

$$(x \rightarrow h_1), t_1 \rightarrow (g_1 \rightarrow h_2), t_2 \rightarrow (g_2 \rightarrow h_3), t_3 \\ \dots (g_{k-1} \rightarrow h_k), t_k \rightarrow (g_k \rightarrow y), t_{k+1}$$

where $x, y, g_i, h_i \in \mathcal{H}$, h_i is the primary recipient hospital for g_i , i.e. $g_i \in send(h_i)$. In addition a gap constraint is enforced to ensure the temporal proximity of these transfers:

$$t_2 - t_1 \leq \delta; t_3 - t_2 \leq \delta; \dots; t_{k+1} - t_k \leq \delta$$

where δ is the allowed upper bound on the time gap between a pair of consecutive transfers.

Example 1. A cascade of size 3 that occurs in data is illustrated in Figure 2. The primary transfer hospitals that are part of this cascade are 1422, 2220 and 4500. Here a primary transfer from 4483 to 1422 (on 18-Oct-95) is followed by a secondary transfer from 1365 to 2220 (on 20-Oct-95); where 1422 is the primary transfer partner of 1365. This secondary transfer is followed by another secondary transfer from 113 to 4500 (on 23-Oct-95); where 113 and 2220 are primary partners. Finally there is a transfer from 2956 to 914 (on 26-Oct-95) where 4500 is the primary recipient for 2956.

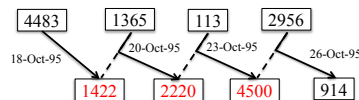


Figure 2: Illustration of a 3-size cascade consisting of the primary transfer hospitals 1422, 2220 and 4500.

3.1 Mining Scheme for Longer Cascades

The task of discovering cascades in the transfers is a combinatorially hard problem. For instance, finding all cascades of size 3 requires us to look through all possible combinations of primary recipient hospitals and count the occurrences of associated cascades in the data. As we start to look for longer cascades the combinatorial explosion of the search space makes this approach infeasible (see [6, 7, 8] for details).

Here we propose a level-wise procedure for discovering long cascades in data that overcomes the combinatorial explosion. The overall steps of this procedure are given below in Algorithm 1. Here we begin with a set of candidate cascades of size 2. We restrict our analysis to

secondary transfers between hospitals within a distance d miles of each other. The number of non-overlapping occurrences of each candidate is calculated using a finite-state automata-based counting scheme. This counting scheme presented in Algorithm 2 follows from our earlier work [6]. After the counts of all secondary transfer candidates of size 2 is determined, we retain only those cascades whose count exceeds a user-specified threshold. This set of cascades is called the frequent cascades of size-2. The frequent cascades of size-2 are used to generate a set of candidate cascades of size 3. The candidate generation scheme based on prefix-suffix matching (see [6]). The procedure then alternates between candidate generation, counting and pruning until there are no more frequent cascades. This level-wise procedure helps control the number of combinations one has to count in order to find all frequent long-cascades.

Algorithm 1 Level-wise procedure for cascade mining

- 1: Generate an initial list of candidate cascades of size-2, $\mathcal{C} = \{ \langle h_i, h_j \rangle : \text{geographical distance}(h_i, h_j) \leq d \text{ miles} \}$.
 - 2: Set $k \leftarrow 2$
 - 3: **repeat**
 - 4: Count the non-overlapped occurrences of candidate cascades in the data (See Algorithm 2).
 - 5: Retain only those cascades that occur more often than a user-specified threshold.
 - 6: Generate $k + 1$ -size candidate cascades, using the list of frequent n -size cascades
 - 7: Set $k \leftarrow k + 1$
 - 8: **until** no candidate cascades remain.
 - 9: Output all the frequent cascades discovered.
-

Counting cascades: Algorithm 2 outlines the pseudo-code for counting occurrences of a single cascade over primary transfer hospitals $\langle h_1, h_2, \dots, h_k \rangle$ where consecutive secondary transfers satisfy the gap-constraint δ . T_α list maintains the latest secondary transfer for each $h_i \in \alpha$ such that it is preceded by a secondary transfer with in δ time gap and are related as per the definition of the cascade. If the condition on Line 8 is met it implies that there exists a sequence of secondary transfers that together constitute an occurrence of the cascade α and also the consecutive pairs of transfers satisfy the gap constraint.

Note that the counting step of the mining procedure (Algorithm 1) on Line 4 requires counting occurrences of all cascades of the candidate list. In the actual implementation we count occurrences of all cascades in the candidate while making one pass of the data. We employ a data-structure built around hash-maps to efficiently access the T_α list for each cascade α in the candidate list only when a transfer $(g \rightarrow h, t)$ that can potentially update T_α is seen in \mathcal{S} .

Algorithm 2 Count cascade with primary transfer hospitals $\alpha = \langle h_1, h_2, \dots, h_k \rangle$ and a gap-constraint δ .

- 1: Initialize $count \leftarrow 0$
 - 2: Initialize $T_\alpha[h_i] = \phi, \forall i = \{1, \dots, k\}$
 - 3: **for** $(g \rightarrow h, t) \in \mathcal{S}$ **do**
 - 4: **for** $i \in \{1, \dots, k\}$ **do**
 - 5: **if** $h = h_i$ **then**
 - 6: **if** $(i = 1)$ or $(g \in \text{send}(h_{i-1})$ and $t - T_\alpha[h_{i-1}] \leq \delta)$ **then**
 - 7: $T_\alpha[h_i] = t$
 - 8: **if** $(g \in \text{send}(h_k)$ and $h \notin \alpha)$ and $(t - T_\alpha[h_k] \leq \delta)$ **then**
 - 9: $count \leftarrow count + 1$
 - 10: $T_\alpha[h_i] = \phi \forall i = \{1, \dots, k\}$
 - 11: **return** $count$
-

4 Statistical Significance

In this section, we present a number of null models against which the significance of the cascades discovered in the data can be established. Significant cascades have a temporal structure that depends on the exact order and timing of the constituent transfers. A null model that removes such structure will help ascertain the p-value of a cascade.

4.1 Temporal shuffling

Here we generate surrogate datasets by randomly shuffling the time of occurrence of transfers in consecutive chunks of 100 transfers. Since we do not add or remove transfers from the original data, the first order statistics are preserved. For each cascade discovered in the level-wise mining procedure we determine its count over n surrogate datasets. This allows us to estimate the distribution of the number of occurrences of a cascade under the null model and determine its p-value. In the results section, we report only those cascades that have p-values below a threshold.

4.2 Spatial shuffling

Here we redistribute the receivers of the secondary transfers. The primary transfers are left undisturbed. For every secondary transfer $g \rightarrow \hat{h}$, we replace \hat{h} with h' , where h' is randomly chosen from the set of hospitals known to receive transfers from g excluding its primary recipient. This null model ensures that the spatial structure of the secondary transfers are sufficiently removed. Again in the results section, we report only those cascades that have p-values below a threshold.

5 Results

In the ICU transfer records over the entire period of 10 years (1996-2006), (described in Section 2), we discovered 163 statistically significant ($p \leq 0.001$ with respect to both the null models) cascades of size 3. We used a maximum distance d of 250 miles between the hospitals, a maximum delay δ of 3 days between transfers, and a minimum count of 15 for mining. Over all there were

3,204 cascade occurrences that were significant and many cascades shared transfers. In total the cascades accounted for 10,208 transfers, or 1.33% of the total number of ICU transfers. The top five cascades found in the ICU transfers data are shown in Table 3 along with p-values of statistical significance under two different noise models.

Table 3: Top five cascades found in the ICU transfers data. (p-value-1: p-value of a cascade under the null model of temporal shuffling; p-value-2: p-value of a cascade under the null model of spatial shuffling).

Cascade	Count	p-value-1	p-value-2
1422-2220-4500	57	0.001	0.001
2419-1099-552	55	0.001	0.001
4661-1204-225	48	0.001	0.001
552-1099-839	47	0.001	0.001
4661-1204-4531	45	0.001	0.001

5.1 Geographical Distribution

Figure 3 shows all the 163 significant cascades of size 3. All the secondary transfers ($g \rightarrow \hat{h}$) are indicated in the plot with blue arrows. The primary pairs in the cascades are indicated by red arrows. These edges indicate the potential ICU transfers that got diverted to secondary locations according to our hypothesis.

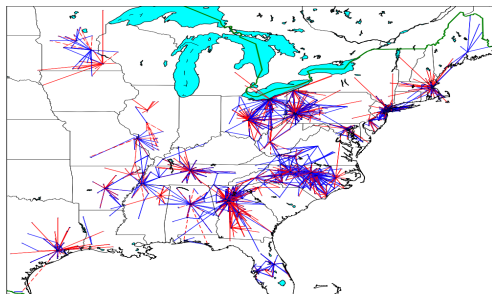


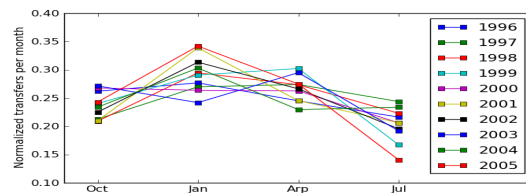
Figure 3: Plot of the cascade occurrence on the US map. In the figure, red arrows indicate primary transfer pairs and blue arrows show the actual secondary transfers.

Figure 3 also shows the locations of the cascades we discovered in the data. It is fairly obvious that there is a high number of cascades occurring across the east coast of United States as is the density of population in this region. The high density of hospitals in this region provide true alternatives in bigger cities and hence the impact of cascades is potentially higher here compared to other regions of the country.

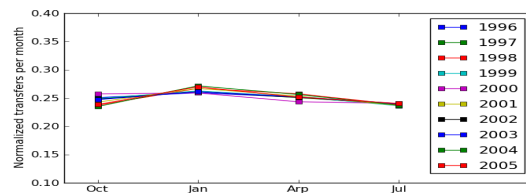
5.2 Seasonal Variations

Figure 4(a) shows the seasonal variations in the number of transfers involved in significant cascades of size 3 for four

quarters in each year of data. The data is normalized with respect to the total number of transfers in occurrences of these cascades. We can clearly see the increase in cascades over the winter quarter. To illustrate the sensitivity provided by the cascade analysis, in Figure 4(b) we show the same seasonal variations reflected in the total number of ICU transfers. The seasonal variation in this data is much smaller than that in Figure 4(a). Hence the cascades can be used as early predictors of seasonal effects such as the winter flu with higher sensitivity. This can in turn be used for capacity planning of ICUs.



(a) ICU transfers participating in cascades



(b) All ICU transfers

Figure 4: Seasonal variation in the occurrence of cascades. The transfers are presented in buckets of three months. The data is normalized with respect to the total number of transfers in occurrences of cascades in (a) and all transfers in (b).

5.3 Network Hot-spots

A hospital can be considered to be a hot-spot in the ICU transfer network if there are many cascades involving it. In Figures 5, we show one such hot-spot (hospital 39 that is part of 29 different significant cascades). Hospital 39 is present in 717 distinct cascade occurrences. There are several other such nodes in the discovered cascades that can be considered as bottle-necks.

6 Discussion and Future Work

Secondary transfers are a common part of the care of critically ill patients. It appears that these secondary transfers frequently cascade. These cascades present a prima facie case that there are important and currently unacknowledged and unmanaged interdependencies among ICUs. Examining the reasons for these specific cascades may offer substantial insights into the functioning of the critical care transfer system and suggest ways to improve the efficiency and quality of care provided to patients. It may be possible to detect these cascades in real time, and use

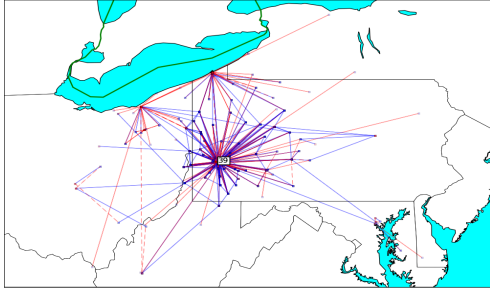


Figure 5: Plot of cascades involving Hospital 39.

them as an early warning sign for impending strain on the ICU system. Further, the hotspots that have been identified offer particular targets for public health planning interventions. Those interventions may be to improve capacity at the hotspots (if they are high functioning hospitals) or to encourage nearby hospitals to seek other transfer destinations (if patients would be better served by reducing the dependence on the hotspot hospitals).

This cascading behavior and its concentration in certain hot-spots is strongly suggestive that there are areas of binding capacity constraint in critical care transfers. While it is true that these patients were able to be transferred to some other hospital, our data raise the concern that such transfers may have been delayed by the time it took to identify a secondary transfer location. Such delays may have important consequences for critically ill patients, whose care it often very time-sensitive; however, it is important to acknowledge that our data suggest but do not prove that this is true. These patterns may provide an important screening tool by which to target quality improvement initiatives to optimize the availability of high quality ICU referral capacity, and may offer a fruitful partnership between system-level informatics and hospital-specific quality improvement.

The approach we have developed here is a quite general tool for detecting and identifying unacknowledged dependencies and tight coupling in complex systems. There has been substantial interest in the potential of networks to shape cascades, particularly as they could relate to cascading failures due to overload – power grid failures may be the most prominent example in the public mind. However, many tightly coupled systems exist, from traffic on highways to work flows in organizations. But while the potential disruptiveness of large scale cascades has been explored, we are aware of little work that seeks to detect sub-catastrophic cascades that may be indicators of inefficiencies in routine practice. The couplings in many such systems are often not easily detected – the methods we propose may be quite general for detecting key nodes in coupled systems that lead to cascading deviations from usual practice.

The results presented in the paper are quite prelimi-

nary. We plan to quantify the definition of hotspots and extend the analysis to variation of hotspots over time and geographical regions. We also plan to look at the nature of hospitals initiating significant cascades and the nature of cascades for different types of transfers (cardiac, non-cardiac, etc.). These will be presented in a future publication.

The formalism presented in this paper has many limitations. For speedy discovery of significant cascades, we need to model the counting process and build equivalent statistical models (see [7, 8] for details). The data we have extracted from MedPAR does not have enough details to answer the most interesting question: are the secondary transfers made because the primary transfer hospital could not be used? We plan to address this in future work. We plan to extract capacity utilization data for a limited set of hospitals in Michigan and correlate it with the cascades. We also plan to set up a simulation platform to find the critical hospital-level and network-level parameters that are important for causing cascades. We are in the process of proposing a pilot study that involves a regional chain of hospitals in Michigan and Ohio for this purpose.

Acknowledgements

This work was supported in part by the NIH grants U54DA021519(KPU), UL1RR024986(KPU) and K08HL091249(TJI).

References

- [1] T. J. Iwashyna *et al.*, “Uncharted paths: hospital networks in critical care,” *Chest*, vol. 135, pp. 827–833, Mar 2009.
- [2] J. Hatten, “Medicare’s common denominator: the covered population,” *Health Care Financ Rev*, vol. 2, no. 2, pp. 53–64, 1980.
- [3] N. A. Halpern *et al.*, “Critical care medicine use and cost among medicare beneficiaries 1995-2000: major discrepancies between two united states federal medicare databases,” *Crit Care Med*, vol. 35, pp. 692–699, Mar 2007.
- [4] T. J. Iwashyna, “Critical care use during the course of serious illness,” *Am J Respir Crit Care Med*, vol. 170, pp. 981–986, Nov 2004.
- [5] N. A. Halpern *et al.*, “Critical care medicine in the united states 1985-2000: an analysis of bed numbers, use, and costs,” *Crit Care Med*, vol. 32, pp. 1254–1259, Jun 2004.
- [6] D. Patnaik *et al.*, “Inferring neuronal network connectivity from spike data: A temporal data mining approach,” *Scientific Prog.*, vol. 16, pp. 49–77, January 2007.
- [7] P. S. Sastry and K. P. Unnikrishnan, “Conditional probability-based significance tests for sequential patterns in multineuronal spike trains,” *Neural Comput.*, vol. 22, no. 2, pp. 1025–1059, 2010.
- [8] S. Laxman *et al.*, “Discovering frequent episodes and learning hidden markov models: A formal connection,” *IEEE Trans. on Knowl. and Data Eng.*, vol. 17, no. 11, pp. 1505–1517, 2005.