Published in final edited form as:

IEEE Int Conf Healthc Inform. 2019 June ; 2019: . doi:10.1109/ICHI.2019.8904557.

Obstetric Patients with Repetitious Hospital Location Transfers Have Prolonged Stays

Cheng Gao,

Department of Biomedical Informatics Vanderbilt University

Abel N. Kho,

Institute for Public Health and Medicine Northwestern University

Sarah Osmundson,

Department of Obstetrics and Gynecology Vanderbilt University Medical Center

Bradley A. Malin,

Department of Biomedical Informatics Vanderbilt University

You Chen

Department of Biomedical Informatics Vanderbilt University

Abstract

There is a general belief that the workflow of surrounding location transfers between locations documented in electronic health record (EHR) during hospitalization is associated with a patient's length of stay (LOS). However, this belief has had little formal investigation in a data-driven manner. Location transfers in patients' hospitalization are hypothesized to be related to LOS. The objective of this study is to assess this relationship, using data derived from the EHR system of a large hospital system, with a focus on the obstetric setting - a clinical environment that exhibits wide swing in resource utilization. We designed a data-driven framework to infer patterns of location transfers and developed a zero-truncated negative binomial model, adjusting for demographics and billed diagnoses, to learn the association between patterns of location transfers and LOS. Indicative factors found to be of indicative of location transfer patterns were further investigated via their odds ratios, Pearson Correlation Coefficients, and Chi-squared test. We evaluated our approach with two years of data on from 5,774 obstetric inpatient encounters from the EHR system of Northwestern Memorial Hospital. The results indicated that the average LOS for patients with patterns of repetitious location transfers (RLTs) was 4.25 days (95% confidence interval [4.02, 4.47]) longer than patients with no RLT. This difference reduced to 3.62 days (95% confidence interval [3.61, 3.64]) after adjusting for age, race and billed diagnoses. We further discovered 21 indicative factors of RLT (statistically significant with a significance level of 0.05), in the form of billed diagnosis codes, each of which exhibited an odds ratio larger than 4. This study suggests that RLT patterns are associated with a prolonged LOS in the obstetric setting. As

cheng.gao@vumc.org.

DECLARATIONS

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to the protection of the reputation and interest of the included hospitals but are available from the corresponding author on reasonable request.

such, healthcare organizations may need to pay more attention to patients with RLTs to refine location transfers workflow and to boost efficiency in obstetric care.

Keywords

EHR; length of stay; location transfer; obstetrics

I. INTRODUCTION

Healthcare costs in the United States (U.S.) have steadily risen over time, reaching \$3.35 trillion (\$10,345 per person) in 2016 and \$3.52 trillion (\$10,870) in 2017. [1] This escalation in hospitalization costs is one of the driving factors behind higher healthcare expenditure. Hospital care accounts for the largest share of national health expenditure in 2016, at around 32% [1].

One of the most expensive types of hospitalizations in the U.S. stems from obstetric issues [2, 3]. This may be because obstetric morbidity derives from a wide range of complex complications. In particular, these problems are affiliated with various complications for soon-to-be mothers (e.g., amniotic fluid embolism, blood clots, congenital hypothyroidism, hemorrhage, maternal obesity and maternal cardiac disease), as well as for fetuses and newborns (e.g., abnormal position of the fetus, birth defects, birth injuries, disorders of the brain and nervous system, fetal distress, hypoglycemia, low birth weight, and prolapsed umbilical cord) [4–6]. The volume and variety of possible morbidities lead to complex obstetric care coordinating routines (e.g., transitions between ambulatory surgery and intensive care) [7, 8]. An appropriate optimization and management of these care coordination routines can could improve hospital resource allocation and reduce a patient's length of stay (LOS), along with further healthcare cost savings. However, to date, there has been a lack of systematic investigation into the relationship between LOS and location transfers within the hospital.

The typical approach to workflow implementation for location transfers is guided by expert intuition-based, which leverages the experience of knowledgeable hospital employees to design care transfer routines to improve the efficiency of hospital resource utilization (e.g., allocation of hospital beds and rooms used for radiographic imaging) [9–12]. However, this type of modeling is relatively static and, thus, it is insufficient to manage frequent situational changes, as is commonplace in hospitals, which tend to be dynamic in complex care environments. As an alternative, a data-driven model focuses to infer undocumented and dynamic transfers patterns that may or may not be defined as a formal policy, which can assist the administrators of healthcare organizations (HCOs) to make better resource allocation decisions [13–16]. Since this perspective is relatively new, most investigations have focused on the learning of transfers patterns (e.g., clinical workflows [14, 17–19]). Yet these approaches for doing so are limited in that they do not quantify the association between patterns of location transfers and LOS.

It is natural to expect that LOS is associated with transfers between locations in a hospital. Especially, unnecessary location transfers, such as repetitious transfers between hospital

locations, can lead to prolonged LOS. This might happen, for instance, when a patient is sent to the wrong place or is shuffled between various holding areas in a hospital, leading to an accumulation in the waiting time for a transfer from one location to the actual endpoint.

Thus, the study in this paper was based on the hypothesis that patterns of location transfers (PLT) incurred by patients during their hospitalization are associated with their LOS. To investigate this hypothesis, we developed a data-driven framework that infers this association to support hospital resource management optimization. First, in the framework, we inferred PLT from the data in EHR and then categorized these patterns into two types: 1) those with repetitious location transfers (RLT) and 2) those with no RLT (NRLT). Second, we compared the LOS between patients associated with RLT and NRLT. Third, the factors indicative of RLT were investigated in the form of billed diagnoses. To evaluate such a framework in a real system, we conducted this study with two years of obstetric inpatient encounter from Northwestern Memorial Hospital, a large academic medical center in downtown Chicago, Illinois.

II. MATERIALS AND METHODS

A. Data Collection

The data for this study was drawn from the Cerner Powerchart EHR system at Northwestern Memorial Hospital (NMH). We extracted two years (2010 – 2012) of data, which included 5,774 obstetric inpatient encounters. Each obstetric inpatient encounter included 1) patient demographics (e.g., age and self-reported race), 2) encounter administrative information (e.g., admission and discharge date), 3) diagnosis codes (e.g., ICD-9 codes) assigned to the patient affiliated with that encounter, and 4) the physical locations where a patient received care during the encounter.

Summary information for the inpatient encounters is provided in Table 1. To orient the audience to the information in this table, we define a location transfer as the transfer of a patient from one location in a hospital to another for care purposes. Each hospital location is represented in I_i (where the subscript i distinguishes between different locations) and location transfers in one hospitalization are represented as s , which can include several locations transfers. For instance, if a patient has a sequence of location transfers during an encounter of the form $s_1 = \{I_1 \rightarrow I_2 \rightarrow I_1\}$ (where I_1 and I_2 are two distinct locations), then this patient will have two location transfers: I_1 to I_2 ($I_1 \rightarrow I_2$) and I_2 back to I_1 ($I_2 \rightarrow I_1$). In this setting, a repetitious location transfer (RLT) is defined as a sequence of location transfers, where there is at least one location that appears more than once. Based on this definition, s_1 is an RLT.

The data in this study contains 37 locations, 99 types of location transfers and 723 distinct ICD-9 codes. There are 9249 location transfers and 9178 ICD-9 codes assigned to these obstetric inpatient encounters. The minimum, median and maximum number of assigned ICD-9 codes is 1, 2 and 17, respectively. For these encounters, the minimum number, median and maximum number of location transfers is 1, 3 and 48, respectively.

LOS is measured in days. Figure 1 depicts the distribution of LOS for inpatient encounters. In this figure, the x -axis corresponds to the LOS in days, while the y -axis corresponds to the number of patient encounters associated with a specific LOS. It can be seen that the patients exhibited a wide range in LOS - from 1 to 55 days. Over 9% of the patients exhibited a LOS longer than 5 days.

B. Study Design

The framework designed for this study (depicted in Figure 2) quantifies the association between PLT and LOS. The framework is composed of three primary components: i) PLT inference (i.e., RLT and NRLT), ii) Association modeling between LOS and RLT and iii) Analysis of the factors indicative of RLT.

Before diving into the details of each model, we provide a guide for the common notation used in the study in Table 2. There are N inpatient encounters $E = \{e_1, e_2, \dots, e_n\}$ and m distinct locations $L = \{l_1, l_2, \dots, l_m\}$. Each encounter has a corresponding patient location sequence s_i . All of the location sequences are managed in the set $S = \{s_1, s_2, \dots, s_n\}$. Based on S , we infer i) Patterns of RLTs P_r , within which each individual pattern contains an RLT and ii) Patterns of NRLTs P_{nr} , within which each individual pattern is devoid of an RLT.

C. Repetitious and Non-Repetitious Location Transfer Patterns Inference

We infer each location transfer sequence from the location sequences in S and categorize into P_r and P_{nr} . The pseudocode for the inference algorithm (Algorithm 1), Pattern Inference in Transfer Between Locations (PIETRAIL), is depicted below. The input of the algorithm is the set of location sequences S , while the outputs are P_r and P_{nr} .

Steps 1 through 7 of PIETRAIL removes repetitious locations in location sequences. For instance, a sequence $s_1 = \{l_1 \rightarrow l_1 \rightarrow l_2\}$ contains a location l_1 , repetitious to itself. We remove such repetition of locations because the sequence of locations s_1 is based on the action events of healthcare professionals. If two healthcare professionals at different times act on the same patient who stays at the same location l_1 , then the location sequence will self-repeat. Since our study focuses on inter-location patterns of transfers, we remove self-repetitive transfers.

Steps 8 through 14 determine if a PLT has an RLT. Upon doing, we obtain two PLT groups: P_r and P_{nr} . For each group, we construct a corresponding patient encounter cohort. If the location sequence of a patient encounter contains the PLT in P_r , then this patient will be included in an RLT cohort, otherwise, the patient will be included in an NRLT cohort. The two groups are leveraged to quantify the relationships between RLT and LOS.

D. Association Modeling between LOS and RLT

Each patient encounter has a corresponding LOS, a set of demographics and billed diagnosis codes. Since LOS is measured as an integer and the count begins at 1 (day), the positive Poisson regression model could be used to model this type of data. However, the Poisson distribution assumes that the mean and variance are equal, which is not a valid assumption in this study. Rather, the data displays overdispersion, where the LOS variance is larger than its

mean. As such, we apply a zero-truncated negative binomial distribution to model LOS. We denote LOS as Y , such that the probability mass function of the zero-truncated negative binomial distribution is defined as:

$$P(Y = y) = \frac{P(Y = y)}{1 - P(Y = 0)}, y = 1, 2, \dots \quad (1)$$

where the numerator $P(Y = y; \mu, k)$ is the probability density function of the negative binomial.

$$P(Y = y; \mu, k) = \frac{\Gamma(y + k)}{\Gamma(k)\Gamma(y + 1)} \left(\frac{k}{\mu + k}\right)^k \left(1 - \frac{k}{\mu + k}\right)^y \quad (2)$$

where μ is the mean of Y , $\Gamma(\cdot)$ is the gamma function and k is a parameter to control the extent of overdispersion. Now, we have a sample of N LOS observations $y_1, y_2, \dots, y_n$ that follow a zero-truncated negative binomial distribution $y_i \sim P(Y = y_i)$. We denote the expectation of y_i as $\widehat{LOS}$, which depends on certain explanatory variables, each of which x_i corresponds to one of the patient-specific features noted earlier (e.g., age, self-reported race, repetitive transfer location indication, and ICD-9 codes). Thus, we model $\widehat{LOS}$ as a linear function of $\mathbf{x}_i$ as $w_1x_1 + w_2x_2 + \dots + w_mx_m$. For simplicity in notation, we let $\mathbf{w} = [w_1, \dots, w_m]$ and $\mathbf{x} = [x_1, \dots, x_m]$. Since $\widehat{LOS}$ can take on non-negative values only, we model the logarithm of $\widehat{LOS}$ as a linear function of $\mathbf{x}_i$:

$$\log(\widehat{LOS}) = \mathbf{w}^T \mathbf{x}, \quad (3)$$

where ω_i represents the expected change in the log of $\widehat{LOS}$ per unit change in the i^{th} variable $\mathbf{x}_i$. Given N observations we can derive the likelihood function from Equations (1) through (3) and then estimate $\mathbf{w}$.

• **Algorithm 1: Pattern Inference in Transfer Between Locations(PIETRAIL)**

• **Input:** $\mathcal{S}$, a set of individual location sequences

• **Output:** $\mathcal{P}_r$ a set of individual location transfer patterns containing an RLT

• $\mathcal{P}_{nr}$ a set of individual location transfer patterns devoid of an RLT

• **Steps:**

- 1: **for** each sequence s_i in $\mathcal{S}$ **do** • //process each sequence
- 2: initialize $p_i = \{\}$ •
- 3: **for** each loaction l_j in s_i **do** • //remove location transitions starting
- 4: **if** $l_j \neq l_{j-1}$ **do** • //and ending at the same location
- 5: $p_i = p_i \cup l_j$
- 6: **end if**
- 7: **end for** •
- 8: **if** p_i is repeated • //determine repeated location transition
- 9: $\mathcal{P}_r = \mathcal{P}_r \cup p_i$ •
- 10: frequency(p_i) + + •
- 11: **else** • //calculate the number of patient encounters
- 12: add p_i into $\mathcal{P}_{nr}$ •
- 13: frequency(p_i) + +
- 14: **end if**
- 15: **end for**

We derive adjusted LOS for patients in the NRLT and RLT groups based on the inferred zero-truncated negative binomial model (estimated as $\mathbf{w}$). To compare the difference in the LOS between the two groups, we apply the following statistical model:

$$\log(\widehat{LOS}) = \alpha + \beta \times t, \quad (4)$$

where t is an indicator for patterns of location transfer. If a patient is in $\mathcal{P}_{nr}$, then t will be set to 1 and 0 otherwise. Since LOS is not normally distributed (as is illustrated in Figure 1), we apply a logarithmic transformation. The differences in LOS for the two groups (RTL and NRLT) can then be measured through the β coefficient in Equation (4). This is because, if the observed encounters are in the RLT group, then t is 1 and the expected LOS is $\alpha + \beta$. If the observed encounters are in the NRLT group, then t is 0 and the expected LOS is α . The difference in LOS between these two groups corresponds to β . We invoke an analysis of variance (ANOVA) model to test for the significance of β at the 0.05 significance level.

E. Analysis of Factors Indicative of RLT

RLT could be influenced by various factors, such as patient demographics and the severity of their health status. We investigate if such confounders are indicative of RLT. We conduct our investigation at two levels: i) the model level to test for the differences in age and ICD-9 codes between the RLT and NRLT groups and ii) the code level to test the relationship between each ICD-9 code and RLT.

At the model level, we test the difference between the RLT and NRLT groups with respect to all of the ICD-9 codes observed in the inpatients' EHRs. For each code, we compute the

proportion of patients assigned the code in both groups. We manage the proportion of the codes across the two groups via two vectors. For each code, two proportions r_{RTL}^i and r_{NRTL}^i represent a point in a two-dimensional space (RTL and NRTL, respectively). The difference in proportions for all of the ICD-9 codes between the RTL and NRTL groups can be measured as a Pearson correlation coefficient (PCC) [20] using the two corresponding vectors. The PCC is in the range $[-1, 1]$, where 1 indicates a positive direct correlation, 0 indicates no correlation, and -1 indicates a negative direct correlation.

We use the same strategy to measure the difference in the age of the patients between the two groups. Since age is a continuous variable we discretize it into five categories, 0–20, 21–30, 31–40, 41–50 and above 50.

At the code level, the odds ratio is calculated to evaluate the association between each ICD-9 code and type of PLT (RLT vs. NRLT). For each ICD-9 code, we conduct a Chi-squared test to determine if the billing code is associated with RLT in a statistically significant manner. Billing codes are considered to be significant when their p -value is less than 0.05 after a Bonferroni correction according to the Chi-squared test.

III. RESULT

In this section, we report on i) Patterns of RLT and NRLT, ii) the difference in LOS between the RLT and NRLT groups and iii) the factors indicative of an RLT.

A. RLT and NRLT

The framework inferred 176 PLTs, of which 91 and 85 satisfy the RLT and NRLT criteria, respectively. Table 4 depicts the 10 repeated and non-repeated PLTs with the largest number of affiliated patient encounters. Notably, the patient encounters affiliated with these 20 PLTs account for 94.3% of all patient encounters. A description of each location in a PLT is depicted in the location index legend at the bottom of the table. Out of 10 PLTs in NRLT, the 1st – 6th, 8th and 10th PLTs indicate that 5,000 (86.6%) patient encounters went through a location path of *Galter Ambulatory Surgery (33)* → [*Women's Health (28)*, *Recovery Room (29)* and *Postpartum (30)*]. For patterns with RLT, there is a frequent repetitious transfer between *Outpatient (34)* and *Galter Ambulatory Surgery (33)*, which appears in the 1st, 2nd, 3rd and 5th PLTs in RLT group.

B. Associations between Patterns of Location Transfer and LOS

The LOS difference between the RLT and NRLT groups are shown in Table 5. This depicts the differences before and after the adjustments that were made for age, race and ICD-9 codes. It can be seen that patients in the RLT group stayed in the hospital for 4.25 days longer before the adjustment and 3.62 days longer after adjustment than those in the NRLT group. The LOS difference between RLT and NRLT groups were statistically significant ($p < 0.0001$).

To provide intuition into the LOS difference, Figure 3 depicts boxplots for the LOS associated with the top 10 PLTs in the NRLT (left) and RLT group (right). The indices for

the LTPs in Figure 3 are the same as those in Table 4. It can be seen that the LOS for patients in the RLT group is larger than those in the NRLT group. Furthermore, patients who experienced repetitious location transfers between *Galter Ambulatory Surgery* (33) and *Outpatient* (34) (e.g., the 1st, 2nd, 3rd, and 5th RLT) have a prolonged LOS in comparison to those who do not have this repetitious location transfer.

C. Factors Indicative of RLT

The distribution of ICD-9 codes across the two groups is depicted in Figure 4 and the model difference for the ICD-9 codes between the RLT and NRLT groups. The distribution for the 723 ICD-9 codes is shown in the left of Figure 4, where each point is a specific code.

In Figure 4, the x - and y -axis represents the proportion of patients in the RLT and NRLT groups, respectively. The 45-degree line indicates where the two groups would have the same number of patients assigned to the code. It can be seen that there are several codes that are quite far from this line, which implies that the proportion of these codes are different between NRLT and RLT groups and potentially indicative of RLT for those deviating to the right of the 45-degree line. The PCC for the ICD-9 codes was 0.613, which indicates that the sets of patients, in terms of ICD-9 codes, in the two groups are not similar. For example, the code *V28.3 - Antenatal screening for malformation using ultrasonics* is affiliated with 20.1% of the encounters in the RLT group, but only 0.6% of the encounters in the NRLT group. This implies that patients with *V28.3* are likely to have RLT, which leads to a prolonged LOS.

The age distribution between the NRLT and RLT groups is shown on the right of Figure 4. It can be seen that the five points are aligned around the 45-degree line. The PCC was 0.995 with $p < 0.0001$, which statistically confirms that there is no significant difference in terms of patients' age between two groups.

In Figure 5, we report the 21 ICD-9 codes with $p < 0.05$ along with their odds ratio (OR) and 95% confidence interval (CI). It can be seen that conditions for soon-to-be mothers, as well as fetuses, are highly associated with an RLT. For example, patients with *642.03 - benign essential hypertension* (OR 23.1, CI [12.1, 44.2]) and *642.93 unspecified hypertension* (OR 21.9, CI [10.7, 44.7]) is highly associated with RLT. Problems associated with fetal growth (e.g., *V28.3 - Antenatal screening for malformation using ultrasonics* (OR 37.2, CI [24.1, 57.4]) are also associated with RLT.

IV. DISCUSSION

In the framework proposed in this work, the association between PLTs and LOS can be used to support hospital resource management optimization and reduce LOS. This paper introduces a data-driven framework in which to study the relationship between patterns of location transfers (PLTs) and length of stay (LOS) are studied. In this framework, the patients from the repetitious location transfers (RLT) group stay in the hospital 4.25 days more than those from the NRLT group. This difference reduces to 3.62 days after adjustment to for demographics and health condition. The application of this framework may provide intuition into factors associated with a prolonged LOS. Still, we realized that this is a pilot

study and there are several limitations we would like to point out and consider them as guidelines for further investigations.

First, although we verified that the LOS for the RLT and NRLT groups was significantly different after adjusting for age, race and health conditions, we did not ascertain the reason that the two groups exhibit such large differences. LOS could be influenced by many factors, such as a patient's health history, nutrition status, social determinants, and their emotional status. Such information may potentially be derived from less structured EMR data, such as clinical notes, and assessed in a future study.

Second, although patterns of repetitious location transfers were associated with prolonged LOS, the reasons for these transfers were unknown due to the limitation of current dataset. It seems that the transfer between outpatient and ambulatory surgery appeared multiple times in patients with prolonged LOS. It may be due to that patients need to go back to ambulatory surgery for additional treatment after the first one.

Third, the patterns of location transfers were studied at the level of location in a hospital, which does not necessarily represent the details of transfers in care. More detailed information regarding patients' location transfer, such as the reason that a patient proceeds through a transfer, the care providers in the locations and the procedures performed in each location may be factors that influence the patterns.

V. CONCLUSION

This study introduced a data-driven framework to study the relationship between patterns of location transfers and length of stay. Obstetric data was studied in a large academic medical center. The patients in the repetitious location transfers (RLT) group had a prolonged LOS during their hospitalization with and without adjustment for demographics and health conditions. Additionally, 21 certain diagnosis codes were found to be indicative of RLT. We believe that the framework provides healthcare organizations with an opportunity to analyze their own populations. In doing so, organizations have the potential to establish and refine care coordination for these patients with RLT, which may lead to care improvement and reduction in LOS.

Acknowledgments

FUNDING

This research was supported, in part, by the following grants from the National Library of Medicine at the National Institutes of Health R00LM011933 and R01LM010207.

REFERENCES

- [1]. Center for Medicare and Medicaid Service (US), National Health Expenditures 2015 Highlights Available from: <https://www.cms.gov>.
- [2]. Wikstrom AK, Gunnarsdottir J, and Cnattingius S, "The paternal role in pre-eclampsia and giving birth to a small for gestational age infant; a population-based cohort study," *BMJ Open*, vol. 2, no. 4, 2012.

- [3]. Leger J, dos Santos S, Larroque B, and Ecosse E, "Pregnancy outcomes and relationship to treatment adequacy in women treated early for congenital hypothyroidism: a longitudinal population-based study," *J Clin Endocrinol Metab*, vol. 100, no. 3, pp. 860–9, 3 2015. [PubMed: 25423569]
- [4]. Bauer CH, "Care of the high-risk neonate," *Pediatr Ann*, vol. 4, no. 5, pp. 114–6, 5 1 1975. [PubMed: 24849564]
- [5]. Clement MS, "Perinatal care in Arizona 1950–2002: a study of the positive impact of technology, regionalization and the Arizona Perinatal Trust," *Journal of perinatology*, vol. 25, no. 8, p. 503, 2005. [PubMed: 15889131]
- [6]. Skjaerven R et al., "Cardiovascular mortality after pre-eclampsia in one child mothers: prospective, population based cohort study," *BMJ*, vol. 345, p. e7677, 11 27 2012. [PubMed: 23186909]
- [7]. Hofmeyr GJ, Lawrie TA, Atallah AN, and Duley L, "Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems," *Cochrane Database Syst Rev*, no. 8, p. CD001059, 8 4 2010. [PubMed: 20687064]
- [8]. Kyser KL et al., "The association between hospital obstetrical volume and maternal postpartum complications," *American journal of obstetrics and gynecology*, vol. 207, no. 1, pp. 42.e1–42.e17, 2012. [PubMed: 22727347]
- [9]. Ingersoll S, Valente SM, and Roper J, "Nurse care coordination for diabetes: a literature review and synthesis," *Journal of nursing care quality*, vol. 20, no. 3, pp. 208–214, 2005. [PubMed: 15965384]
- [10]. McAllister JW, Presler E, and Cooley WC, "Practice-based care coordination: a medical home essential," *Pediatrics*, vol. 120, no. 3, pp. e723–e733, 2007. [PubMed: 17766512]
- [11]. Darkins A et al., "Care Coordination/Home Telehealth: the systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions," *Telemedicine and e-Health*, vol. 14, no. 10, pp. 1118–1126, 2008. [PubMed: 19119835]
- [12]. Katz EB, Carrier ER, Umscheid CA, and Pines JM, "Comparative effectiveness of care coordination interventions in the emergency department: a systematic review," *Annals of emergency medicine*, vol. 60, no. 1, pp. 12–23. e1, 2012. [PubMed: 22542309]
- [13]. Zhang H, Mehotra S, Liebovitz D, Gunter CA, and Malin B, "Mining deviations from patient care pathways via electronic medical record system audits," *ACM Transactions on Management Information Systems (TMIS)*, vol. 4, no. 4, p. 17, 2013.
- [14]. Chen Y et al., "Inferring clinical workflow efficiency via electronic medical record utilization," in *AMIA annual symposium proceedings*, 2015, vol. 2015, p. 416: American Medical Informatics Association. [PubMed: 26958173]
- [15]. Bennett PJ, Fosbinder D, and Williams M, "Care coordination in an academic medical center," *Nursing case management: managing the process of patient care*, vol. 2, no. 2, pp. 75–82, 1997. [PubMed: 9171683]
- [16]. Carter EM and Potts HW, "Predicting length of stay from an electronic patient record system: a primary total knee replacement example," *BMC medical informatics and decision making*, vol. 14, no. 1, p. 26, 2014. [PubMed: 24708853]
- [17]. Chen Y, Lorenzi N, Nyemba S, Schildcrout JS, and Malin B, "We work with them? Healthcare workers interpretation of organizational relations mined from electronic health records," *International journal of medical informatics*, vol. 83, no. 7, pp. 495–506, 2014. [PubMed: 24845147]
- [18]. Yan C, Chen Y, Li B, Liebovitz D, and Malin B, "Learning clinical workflows to identify subgroups of heart failure patients," in *AMIA Annual Symposium Proceedings*, 2016, vol. 2016, p. 1248: American Medical Informatics Association. [PubMed: 28269922]
- [19]. Chen Y, Lorenzi NM, Sandberg WS, Wolgast K, and Malin BA, "Identifying collaborative care teams through electronic medical record utilization patterns," *Journal of the American Medical Informatics Association*, vol. 24, no. e1, pp. e111–e120, 2017. [PubMed: 27570217]
- [20]. Benesty J, Chen J, Huang Y, and Cohen I, "Pearson correlation coefficient," in *Noise reduction in speech processing*: Springer, 2009, pp. 1–4.

Figure 1.
The frequency distribution of LOS ($N = 5,774$ inpatient encounters).

Figure 2.

A conceptualization of the framework for the inference of associations between RLT and LOS.

Figure 3.
Length of stay (LOS) for the top 10 PLTs in terms of the number of patients on the left) NRTL and right) RLT groups.

Figure 4.
Distribution of ICD-9 codes (left) and age group (right) between the RLT and NRLT groups.

Figure 5.

Odds ratio and 95% confidence intervals of the ICD-9 codes that were statistically significantly associated with repetitious location transfer (RLT).

Table 1.

Summary information for the inpatient encounters ($N = 5774$) in terms of patient demographics, billed health conditions, and location transfers.

Item	Value
Inpatient encounters	5,774
Unique ICD-9 codes	723
ICD-9 codes assigned to encounters	9,178
Unique locations	37
Unique location transfers	99
Location transfers	9,249
Age (Mean $\pm$ Standard Deviation)	31.7 $\pm$ 5.9

Table 2.

Common notation used in this study (RLT: repetitious location transfer)

Notation	Definition
$L = \{l_1, l_2, \dots, l_d\}$	A set of distinct hospital locations
$E = \{e_1, e_2, \dots, e_n\}$	A set of n inpatient encounters
$s_i = \{\dots \rightarrow l_i \rightarrow l_j \rightarrow \dots\}$	A location sequence for an individual patient
$S = \{s_1, s_2, \dots, s_n\}$	A set of n location sequences
$p_l = \{\dots l_k \rightarrow l_i \dots\}$	A location transfer
$P_l = \{\dots, p_i, p_j, \dots\}$	A set of location transfer patterns containing RLT
$P_w = \{\dots, p_k, p_i, \dots\}$	A set of location transfer patterns devoid of an RLT

Table 4.

A comparison of the NRLT and RLT groups according to their top 10 PLTs.

NRLT group				RLT group		
No.	Location Pattern	# of Patient Encounter (%)		No.	Location Pattern	# of Patient Encounter (%)
1	33→28	989 (17.1)		1	33→34→33→34	68 (1.2)
2	33→30	967 (16.7)		2	33→34→33→34→33→34	23 (0.4)
3	33→29	924 (16)		3	34→33→34	16 (0.3)
4	26→33→29	685 (11.9)		4	33→30→33→30	15 (0.3)
5	26→33→30	679 (11.8)		5	33→34→33→34→33→34→33→34	8 (0.1)
6	26→33→28	659 (11.4)		6	33→29→33→29	8 (0.1)
7	33→34	222 (3.8)		7	26→34→33→34	8 (0.1)
8	21→33→30	54 (0.9)		8	33→28→33→28	8 (0.1)
9	26→33→34	54 (0.9)		9	26→33→29→33→29	6 (0.1)
10	21→33→28	43 (0.7)	10	26→33→28→33→28	6 (0.1)	
Total		5276 (91.4)			166 (2.9)	
Location index legend						
Index	Location				Index	Location
21	Prentice Ambulatory Surgery				30	Postpartum
26	Admitting Prentice				33	Galter Ambulatory Surgery
28	Women’s Health				34	Outpatient
29	Recovery Room					

Table 5.

Comparison of the RLT and NRLT groups in terms of differences in LOS.

Group	Unadjusted LOS Difference (95% Confidence Interval)	<i>p</i>	Adjusted LOS Difference (95% Confidence Interval)	<i>p</i>
RLT vs. NRLT (directional, LOS for RLT group is greater than LOS for NRLT group)	4.25 days (4.02 – 4.47)	< 0.0001	3.62 days (3.61 – 3.64)	< 0.0001