Exploratory Study for Readmission in Cancer Patients

Completed Research

Luwen Huangfu
University of Arizona
huangfuluwen@email.arizona.edu

Stephen Hayne
Colorado State University
Stephen.Hayne@colostate.edu

Jian James Ma
University of Colorado, Colorado Springs
jma@uccs.edu

Nicholas Roberts
Colorado State University
nick.roberts@colostate.edu

Abstract

Cancer, a major threat to public health worldwide, causes substantial financial burdens. Cancer readmission time interval, or Out-of-Hospital Days (OHD) between two consecutive hospital admissions, has been widely adopted as an important measure of healthcare service quality. However, there is a paucity of models focusing on OHD and associated risk factors due to limited access to cancer patients' data and absence of important factors (e.g., geographic factors) in the data. We utilize OHD (>30) as the result of cancer patient's conditions (e.g., personal, medical) and treatment costs. We analyze a sample of 22,231 admissions extracted from 635,261 cancer patient Electronic Health Records (EHR) from 190 hospitals in China. We apply text mining on free-form address fields to extract patients' home address and hospitals' location information. Using hierarchical linear regression, we find various types of factors significantly influence OHD: age, marital status, number of admissions and whether the treating hospital is in the same province as the patient.

Keywords

Cancer, OHD, EHR, text mining, cost, geographic.

Introduction

Cancer is one of the major diseases threatening human life and can impose a huge financial burden on patients and their families (Warner et al. 2015). Cancer treatment can be a long-term process, during which cancer patients may be readmitted to the hospital multiple times (Chen et al. 2016; Stewart et al. 2015). Frequent readmission is shown to be associated with poorer long-term survival and higher financial burden (Orcutt et al. 2016; Slankamenac et al. 2017; Warner et al. 2015; Zhuang et al. 2015). As a result, a patients' readmission rate is an important research topic in the public health field (Ji et al. 2012; Merkow et al. 2015).

Cancer readmission time interval, or Out-of-Hospital Days (OHD), is defined as the time interval between when a patient has been discharged from hospital and when they are admitted again the next time. OHD has been widely adopted as an important measure of healthcare service quality. Multiple time intervals for OHD have been proposed, i.e., 30 days, 90 days, or 180 days between two consecutive hospital
(re)admissions. The majority of studies choose to examine OHD of less than 30 days because the expectation is that poor medical service was provided if a patient returns within that time window. However, in cancer treatment plans, OHD that is more than 30 days is very typical. Rather than an indicator of poor medical service quality, the OHD for cancer patients is usually related to the decisions made by the patient and her doctor(s) in order to maximize treatment outcome with minimal costs (Bayati et al. 2014). For cancer patients, the treatment cost, and the patient’s demographic, medical, and geographic factors may play important roles in OHD. In this study, we aim to analyze the relationship between OHD over 30 days and the above variables.

Even though some studies have analyzed the risk factors when OHD is within a certain threshold (e.g., 30-days, 90-days or 180-days), an exploratory model for exact OHD can provide more detailed information about healthcare quality and cost (Dorney et al. 2017; Orcutt et al. 2016; Zuckerman et al. 2016). A longer OHD provides dual benefits: it can improve the patient’s quality of life, and it can reduce medical expenses for all stakeholders. Therefore, it is important to understand the factors associated with OHD and predict OHD based on these factors.

We suggest there is a paucity of exploratory models that focus on exact OHD and associated risk factors for OHD. One reason may be that data specific to readmissions in cancers are limited, or many studies largely focus on one cancer type (making it impossible to generalize). Previous studies neglect critical factors that might be important in OHD prediction, such as geographic factors.

Our Electronic Health Records (EHR) dataset collected from a national public health reporting system in China, has several advantages for the study of OHD. We have access to cancer patient data from 190 hospitals located in 127 cities in China from 2010 to 2014. In total, we obtained 635,261 inpatients’ admissions EHR. Our dataset contains 18 various types of cancer, e.g., skin cancer, stomach cancer, and larynx cancer. The large numbers and wide coverage of cancer type, time series, and patients’ information allow us to conduct a very thorough analysis. Results of analysis from national-level, multiple-cancer-type data set may be more generalizable compared to studies conducted with small samples targeting one or two cancer types.

In this article, we examine the factors that can influence the number of OHD between consecutive cancer treatments. Specifically, we have four types of independent variables: patient demographic (e.g., gender, age), medical (e.g., a patient’s number of cancers), financial (e.g., accumulated cost of previous treatments), and geographic factors (e.g., whether patient’s home and hospital are in the same province).

Our central research question is:

**What factors (e.g., demographic, medical, financial, geographic) interact with Out-of-Hospital Days?**

In the next sections, we present the methodology of studying OHD, results, discussion, and conclusions.
Methodology

Using (Hevner et al. 2004)’s design science framework, we position our work as an exploratory analysis of factors associated with the readmission of cancer patients based on EHR dataset. We adopt the framework proposed by (Peffers et al. 2007) as a schematic (see Figure 1) of the steps followed in building the explanatory models in this study.

![Figure 1. Research Framework](image-url)

**Problem Identification and Motivation**

We are highly motivated to study OHD of cancer patients because hospital readmissions are increasingly recognized as drivers of healthcare spending (Hu et al. 2014). A better understanding of OHD can be leveraged by public health researchers, practitioner, and policymakers to allocate medical resources more efficiently. Moreover, cancer patients and health insurance agencies can optimize payment and disease treatment plans according to OHD, which then reduces medical expenses and improve a patient’s quality of life.

Patient demographic information (e.g., age, sex, and marital status) could be one of the factors that influence the OHD. For example, studies have shown that age is an important factor that is associated with OHD (Doumouras et al. 2016; Luryi et al. 2016).

Treatment-related variables, including the number of cancers, number of admissions and number of hospitals before the next readmission have been demonstrated to correlate with OHD (Greenblatt et al. 2010). Costs of admission are not well studied, which can be an important factor influencing patients’ decisions (Luryi et al. 2016). In our EHR dataset, detailed information about each patient and each visit is available, including time spent in the hospital, and cost of each treatment, which allows us to examine cost in different forms (e.g., overall cost, daily cost).

Geographic locations might be an important consideration for patients when deciding whether to go back to the hospital. Many studies have shown that quality of treatment and OHD are highly correlated (Puri et al. 2015). The quality of treatment is very much determined by geographic location in China. Bigger cities tend to have better doctors and equipment than smaller cities, especially for severe diseases like cancer. Rural areas in China have the lowest quality or even lack healthcare facilities.

The cost of traveling from one location (patient’s home) to another location (hospital) can be a concern. For example, rural patients may have to travel long distances to large cities (which is time-consuming and expensive) to get their diseases treated due to the imbalance of medical facility capabilities. These patients might delay treatment. Previous studies appear to neglect the impact of geographic information (Chaudhary et al. 2017; Donzé et al. 2017) on OHD. In our EHR dataset, we have detailed information for each admission about each patient’s hospital address and her home address.
We explore the factors that influence the number of OHD between consecutive cancer treatments. Specifically, we analyze the relationship between OHD and patient’s cost of previous hospital treatment, patient’s demographic and geographic, medical, and financial characteristics.

**Define Objectives**

We define four types of independent variables: patient demographic (e.g., gender, age), medical (e.g., a patient’s number of cancers), financial (e.g., accumulated cost of previous treatments), and geographic (e.g., whether patient’s home and hospital are in the same province) factors as illustrated in Figure 2. While other factors were readily available in the dataset, geographic variables were not. We used text mining based on free text written in the address fields, and compared that text to extensive published lists of cities and provinces.

![Figure 2. Research Objective and Plan](image)

**Design and Development**

Our dataset contains rich information about demographics, diseases, and cost (see Tables 1 and 2). The average cost (over all years) for each treatment was 24,839 Chinese Yuan (about 3,604 US dollars), and the maximum cost for one treatment was 99,997 Chinese Yuan (about 14,509 US dollars). To put this in
perspective, the average annual income in China was $10,220 in 2010¹. Clearly, these costs can represent a huge economic burden on patients and their families.

During the text-mining process, we extracted city and province. Then we coded these to identify whether a patient was living in a rural or urban area and whether they lived in the same province as the hospital they attended.

Since we focused on readmission time intervals, we filtered out the patients who only have one admission record and the OHD records that are less than 30 days.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHD</td>
<td>Number of days between this admission date and previous discharge date</td>
</tr>
<tr>
<td>AGE</td>
<td>Age</td>
</tr>
<tr>
<td>S</td>
<td>Sex. Binary, Male = 0, Female = 1</td>
</tr>
<tr>
<td>MAR</td>
<td>Marital status. Binary, Married = 0, Non-married (Single/Widowed/Divorced) = 1</td>
</tr>
<tr>
<td>CAN</td>
<td>Number of cancers diagnosed until this admission</td>
</tr>
<tr>
<td>ADM</td>
<td>Number of admissions until this admission</td>
</tr>
<tr>
<td>LOGCIN</td>
<td>Log of cost per day of all previous admissions</td>
</tr>
<tr>
<td>IP</td>
<td>Whether home and hospital are in the same province (No = 0, Yes = 1)</td>
</tr>
<tr>
<td>IC</td>
<td>Whether home and hospital are in the same city (No = 0, Yes = 1)</td>
</tr>
</tbody>
</table>

Table 1: Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHD</td>
<td>126.50</td>
<td>129.35</td>
</tr>
<tr>
<td>AGE</td>
<td>55.38</td>
<td>15.58</td>
</tr>
<tr>
<td>CAN</td>
<td>1.11</td>
<td>0.33</td>
</tr>
<tr>
<td>ADM</td>
<td>3.09</td>
<td>1.86</td>
</tr>
<tr>
<td>LOGCIN</td>
<td>7.52</td>
<td>1.17</td>
</tr>
</tbody>
</table>

Note: These means and standard deviations are based on data at the visit level with 22,231 records excluding ADM = 1 and OHD < 30.

Table 2. Descriptive Statistics for Continuous Variables

Individual patients have multiple hospital visits; thus, we have nested data. We used hierarchical linear modeling (HLM) (Raudenbush et al. 2004) to analyze our data because HLM is particularly well suited for nested data. We ran a null model for the dependent variable with no predictor to determine whether there was sufficient variance between individuals. The ICC1 value for OHD was 0.33, indicating that the variability between individuals was large and using HLM was appropriate (Raudenbush et al. 2004).

**Demonstration and Evaluation**

After building our model on the dataset, we can see that several of our proposed four types of independent variables influence the OHD significantly (see Tables 3 and 4). A patient’s age, marital status, number of cancers, number of admissions and whether the hospital they are treated at is in a different province than where they live, are all significant influences on OHD. A patient’s gender, how much their treatment cost and whether the hospital was in the same city as where they live were not significant influences.

<table>
<thead>
<tr>
<th>Source</th>
<th>Numerator df</th>
<th>Denominator df</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>22219</td>
<td>57.503</td>
<td>.000</td>
</tr>
<tr>
<td>AGE</td>
<td>1</td>
<td>22219</td>
<td>4.025</td>
<td>.045 *</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
<td>22219</td>
<td>.931</td>
<td>.335</td>
</tr>
<tr>
<td>MAR</td>
<td>1</td>
<td>22219</td>
<td>12.525</td>
<td>.000 ***</td>
</tr>
<tr>
<td>CAN</td>
<td>1</td>
<td>22219</td>
<td>7.905</td>
<td>.005 **</td>
</tr>
<tr>
<td>ADM</td>
<td>1</td>
<td>22219</td>
<td>526.801</td>
<td>.000 ***</td>
</tr>
<tr>
<td>LOGCIN</td>
<td>1</td>
<td>22219</td>
<td>.396</td>
<td>.529</td>
</tr>
<tr>
<td>IP</td>
<td>1</td>
<td>22219</td>
<td>4.728</td>
<td>.030 *</td>
</tr>
<tr>
<td>IC</td>
<td>1</td>
<td>22219</td>
<td>.454</td>
<td>.500</td>
</tr>
</tbody>
</table>

*Dependent Variable: OHD.

**Table 3. Type III Tests of Fixed Effects**

(* = p<.05, ** = p<.01, *** = p<.001)
Communication

One interesting finding is that geographic factors can significantly influence OHD. People whose home and hospital are not in the same province have higher OHD than people whose home and hospital are in the same province. Perhaps this is not surprising, because the distance between home and hospital that are not in the same province is usually longer than the distance between home and hospital that are in the same province. It is possible that crossing provincial boundaries may keep patients from getting treatment again due to a psychological reason or a higher traveling expense.

Besides geographic factors mentioned above, demographic and medical factors can also significantly influence OHD. We find that age older patients have higher OHD. This might be due to the fact that older citizens are less likely to seek treatment (Hou et al. 2016; Prince et al. 2015). In addition, married people have higher OHD. Studies have shown that family members play important roles in cancer treatment decision-making (Laidsaar-Powell et al. 2016; Öhlén et al. 2006). On the other hand, we find that the number of admissions negatively influences OHD.

Surprisingly, the cost was not a significant factor. We note that this data contains the overall treatment cost, which might be covered by patients’ insurance. Perhaps if we had access to patients’ out-of-pocket payments, we might see a different result.

Discussion and Conclusion

As a chronic and often terminal disease, cancer’s treatment can be extremely lengthy and costly (Soerjomataram et al. 2012). The goal of this investigation was the development and validation of variables and model that are applicable to explain OHD (> 30 days). We used factors from four different perspectives, that is, demographic, medical, financial, and geographic factors.
This investigation is an important step toward the development and evaluation of new factors, geographic in particular, which are important for explaining OHD. Even though there are some studies on OHD, to the best of our knowledge, we are the first to combine four different types of factors together, enabled by text mining and geographic lookup knowledge. This study might help healthcare providers and policymakers identify new ways to enhance cancer management, particularly as it relates to where patients live.

We can see that whether the location is rural or urban, traveling between provinces can significantly influence OHD. If a patient lives in a different province than the hospital they are being treated at, she will have a longer OHD when compared to a patient who lives in the same province.

Our analysis can help hospital practitioners and administration to better allocate medical resources and improve treatment plans according to patient’s medical, demographic, and geographic factors. In this study, we found that OHD is significantly impacted by the patient’s age and previous medical admissions; providers may need to pay extra attention to them. Our analysis also suggests doctors should take patient’s location into consideration. Especially for patients who live in a different province and who may have to travel longer distances to where the hospital is located, doctors need to be aware that those patients tend to have longer OHD. Special attention and awareness are necessary for these patients since they may not come back for another treatment soon.

Another interesting finding that may provide guidelines for hospital administration is to consider the patient’s marital status. Non-married patients tend to have lower OHD than married patients do. This implies that spouses may play a role in patients’ decision-making. As a future extension of our work, we are interested in discovering how marital status may affect the number of days that a patient stays hospitalized. We will investigate how long a married patient stays in the hospital during treatment, and how that will influence OHD before the next admission. We will explore a spouse’ role in comforting the patient, communicating with doctors, etc., and how those efforts affect the patient’s treatment plan and OHD.

This research has several limitations. First, due to the complex nature of OHD, we do not know exactly how the amount of time between admissions is decided by the patient. Second, hospitals are only identified by their location, which means the hospitals are anonymous. We currently know nothing about their quality, facilities or technological capability.
REFERENCES


Twenty-fourth Americas Conference on Information Systems, New Orleans, 2018


