Big Data Research ••• (••••) •••–•••

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

Revealing Physicians Referrals from Health Insurance Claims Data ☆

Vagner Figueredo de Santana [a], Ana Paula Appel [a], Luis Gregorio Moyano [b], Marcia Ito [a], Claudio Santos Pinhanez [a]

[a] IBM Research, Brazil
[b] FCEN/CONICET, Mendoza, Argentina

ARTICLE INFO

Article history:
Received 13 October 2017
Received in revised form 23 January 2018
Accepted 7 March 2018
Available online xxxx

ABSTRACT

Health insurance companies in Brazil have their data about claims organized having the view only for service providers. In this way, they lose the view of physicians' activity and how physicians share patients. Partnership between physicians can be seen as fruitful, when they team up to help a patient, but could represent an issue as well, when a recommendation to visit another physician occurs only because they work in same clinic. This work took place during a short-term project involving a partnership between our lab and a large health insurance company in Brazil. The goal of the project was to provide insights (with business impact) about physicians' activity from the analysis of the claims database. This work presents one of the outcomes of the project, i.e., a way of modeling the underlying referrals in the social network of physicians resulting from health insurance claims data. The approach considers the flow of patients through the physician–physician network, highlighting connections where referrals between physicians potentially occurred. We present the results from the analysis of a claims database (detailing 18 months of activity) from the health insurance company we partnered with. The main contribution presented in this paper is the model to reveal *mutual referrals* between physicians. Results show the proposed model reveals underlying characteristics of physicians' activity from real health insurance claims data with multiple business applications.

1. Introduction

Health insurance costs are a main issue of concern in almost every country in the world as budget constraints impact directly on the quality of the service. As a result, health insurance companies have been extensively trying to reach a trade-off between offered services and costs as a way to meet budget constraints.

One way for health insurance companies to address those issues is to better understand the complex relationships among the diverse participants of the healthcare systems, including patients, physicians, hospitals, and other service providers. To support this quest, healthcare insurance companies and other health service providers have often a wealth of data from their own operations at their disposal, especially transactional data.

In the case of health insurance companies, an important piece of transactional data involves the claims presented by their ecosystem of providers. In the present work, a claim represents a report from a physician or a healthcare service provider to a health insurance company requesting some form of fee related to a patient's consultation with a physician, a clinical exam, or a medical procedure. Even though claims data may vary, it generally contains at least the ID of the healthcare professional involved in the procedure (it may also be a group of professionals), the ID of the patient, the type of procedure, and time information related to the event. It may include other types of information such as location of the service, pre-authorization codes, etc.

Traditionally the analysis of claims data is based on applying statistics and Data Mining methods to the individual elements of the system (physicians, service providers, patients) or to the set of claims. However, healthcare is often provided by collaborative teams of physicians, nurses, and technicians which are connected to each other by often strong professional relationships. Physicians that refer patients to other physicians have clear preferences about who they want to team up with for specific procedures and often are involved in master-apprentice structures. Physicians also have preferences for specific service providers such as hospitals and clinical analysis laboratories. Those recommendations could be good for building patient trust or indicate a fraud when this is not the patient's will. Similarly, patients establish bonds of trust and reliance with specific physicians or group of physicians.

In practice, mining claims is difficult because claims are paid to a wide variety of providers, such as hospitals, clinics, or even

☆ This article belongs to Special Issue: Medical Data Analytics.

E-mail address: santana.vagner@gmail.com (V. Figueredo de Santana).

https://doi.org/10.1016/j.bdr.2018.03.002
2214-5796/

physicians registered as small companies. A single patient can consult a physician through all those channels. It could be even difficult to know who exactly is taking care of a given patient, since that there are cases in which the physician ID used in a claim is from a professional registered in the health care provider system, but the one taking care of the patient is an unregistered physician. Moreover, claims contain no information about referrals and often the connections among service provider team members are not recorded explicitly in claims. It is also well known that claims data is riddled with errors and unreliable information. Despite all those difficulties, we show in this paper that meaningful and reliable insights about the flow of patients through the network of physicians and how physicians refer each other can be inferred from claims data.

This study took place during a partnership, in a short-term project, between our lab and a major Brazilian health insurance company involving the analysis of their claims database. The main challenge the insurance company brought to our team was to identify physicians that excel at medicine by using the claims database. After decomposing this challenge considering the health insurance workflow, the following components were identified as key factors for outstanding professionals, defined by the health insurance company: physicians referred by peers, relative importance in the network of physicians, and returning behavior of patients. The need for modeling referrals emerged during interactions with the health insurance company staff as a way of identifying physicians that excel in a certain specialty and are referred by their peers. These interactions occurred weekly, fomenting discussions between our team and subject matter experts, IT specialists, analytics team, and the superintendent of the health insurance company. Hence, this work aims at presenting a way of modeling mutual referrals in a physician–physician network, which connects to the hypothesis studied in this work: *H1) It is possible to identify underlying physicians' referrals from claims data.*

The main contribution of this work is a way of modeling mutual referral patterns in the physician–physician network. The method can therefore be used by health insurance companies to better manage the physicians they have businesses with, nurturing the experience of registered physicians and inviting unregistered physicians that collaborate with registered ones. It can also be used to support patients to receive more integrated care from a group of physicians and service providers.

This paper is organized as follows: section 2 describes the related work, section 3 details the database analyzed, section 4 presents the proposed model for highlighting mutual referrals, section 5 discusses the obtained results, and section 6 concludes and points to future works.

2. Related work

Healthcare data is heralded as the key element in the quest to improve efficiency and reduce costs in healthcare systems [1]. This trend is becoming more pronounced as multi-scale data generated from individuals is continuously increasing, particularly due to new high-throughput sequencing platforms, real-time imaging, and point-of-care devices, as well as wearable computing and mobile health technologies [2].

In healthcare, data heterogeneity and variety arise as a result of linking the diverse range of biomedical data sources available. Sources of quantitative data (e.g., sensor data, images, gene arrays, laboratory tests) and qualitative data (e.g., diagnostics, free text, demographics) usually include both structured and unstructured data, normally under the name of Electronic Health Records. Additionally, the possibility to process large volumes of both structured and unstructured medical data allows for large-scale longitudinal studies, useful to capture trends and to propose predictive models [3].

One of the most useful and commonly used datasets are claims databases. Claims data records are often rich in details, as they describe important elements of the events taking place around the healthcare professional and the patient, e.g., timestamps, geographical location, diagnosis codes, associated expenses, among others. The use of claims data in healthcare studies has been scrutinized in [4] and [5], providing a set of good practices and outlining the shortcomings of claims-based research.

Social network analysis has proven to be a useful analysis tool in this context, allowing for insights difficult to reach by traditional descriptive statistics as presented in [5]. For instance, social network analysis has been used to study comorbidity, the simultaneous presence of two chronic diseases or conditions in a patient. By structuring diseases as a network, it is possible to quantify some of the aspects of the complex interactions between conditions in the different patient populations. A number of studies have focused on extensive claims datasets to examine and understand comorbidity networks. In [6], the authors study a diffusion process on a comorbidity network to model the progressive spreading of diseases on a population depending on demographic data. In [7], the authors study how a given chronic disease (diabetes) correlates with age and gender, spanning almost 2 million patients from an entire European country. Such comorbidity networks have also been proposed as models to understand the connection between genetic and environmental risk factors for diseases [8].

Beyond clinical purposes, claims data have also been studied to understand the complex interactions of different organizational structures and management relationships involved in patient care processes. For instance, in [9], temporal patterns in electronic health records were modeled in order to present useful information for decision-making. The authors developed a data representation for knowledge discovery so as to extract useful insights on latent factors of the different processes involving a patient, aiming at improving workflows.

Another important trend is the understanding of the relationship among healthcare professionals, in particular the physicians. In [10], the authors apply social network analysis to mine networks of physicians which might be used to improve the designation of middle-sized administrative units (accountable care organizations). Sauter et al. [11] use social network analysis to understand networks of healthcare providers which share patients, providing insights in the interplay between general practitioners, internal specialists, and pediatricians. Also, the network structure of different healthcare providers taking care of a given individual can show important variability of the healthcare system [12].

Social networks have also been used to understand the state of coordination of healthcare actors. In [13], the authors describe a complex network approach applied to health insurance claims to understand the nature of collaboration among physicians treating in-hospital patients and to explore the impact of collaboration on cost and quality of care. Also, in [14], the authors study the social network structure in hospitals among healthcare professionals to understand which variables affect patient care efficiency measures. The idea is further developed in [15] from a statistical point of view in a medium-sized number of hospitals, through the analysis of temporal patterns and costs.

The medical referral system in the Canadian healthcare system is studied in [16], where the authors map and analyze the network between general practitioners and specialists. In [17], the authors describe the condition of the basic medical insurance for urban and rural residents in China, then they demonstrate that social network analysis can be used in the health insurance claims data to support better understanding of patients transfers among hospitals.

Please cite this article in press as: V. Figueredo de Santana et al., Revealing Physicians Referrals from Health Insurance Claims Data, Big Data Res. (2018), https://doi.org/10.1016/j.bdr.2018.03.002