CENTERIS 2013 - Conference on ENTERprise Information Systems / ProjMAN 2013 - International Conference on Project MANagement / HCIST 2013 - International Conference on Health and Social Care Information Systems and Technologies

Predicting Healthcare Fraud in Medicaid: A Multidimensional Data Model and Analysis Techniques for Fraud Detection

Dallas Thornton[a]*, Roland M. Mueller[b], Paulus Schoutsen[a], Jos van Hillegersberg[c]

[a] *San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0505, USA*
[b] *Department of Business and Economics, Berlin School of Economics and Law, Badensche Straße 52, 10825 Berlin, Germany*
[c] *Department of Information Systems and Change Management, University of Twente, Drienerlolaan 5, 7522 NB Enschede, Netherlands*

Abstract

It is estimated that approximately $700 billion is lost due to fraud, waste, and abuse in the US healthcare system. Medicaid has been particularly susceptible target for fraud in recent years, with a distributed management model, limited cross-program communications, and a difficult-to-track patient population of low-income adults, their children, and people with certain disabilities. For effective fraud detection, one has to look at the data beyond the transaction-level. This paper builds upon Sparrow's fraud type classifications and the Medicaid environment and to develop a Medicaid multidimensional schema and provide a set of multidimensional data models and analysis techniques that help to predict the likelihood of fraudulent activities. These data views address the most prevalent known fraud types and should prove useful in discovering the unknown unknowns. The model is evaluated by functionally testing against known fraud cases.

© 2013 The Authors Published by Elsevier Ltd.
Selection and/or peer-review under responsibility of SCIKA – Association for Promotion and Dissemination of Scientific Knowledge

Keywords: Fraud Detection; Predictive Analytics; Healthcare; Medicaid; Data Modeling

* Corresponding author. Tel.: +1-858-534-8364; fax: +1-858-225-3661.
E-mail address: dallas@sdsc.edu.

1. Introduction

Roughly one-third of the $2.7T spent on healthcare in the US is attributable to fraud, waste, and abuse [1]. Payers for healthcare services must deal with fraudulent practitioners, organized criminal schemes, and honest providers who make unintended mistakes while billing for their legitimate services. The US Medicaid system is particularly susceptible fraud and abuse, since it is harder to exclude problematic providers and is managed separately and with limited coordination across the states. Each state has program sovereignty and maintains its own eligibility and benefits criterion, making nationwide initiatives challenging.

While there is undoubtedly an abundance of fraud in the system, states and the federal government lack sophisticated fraud control systems. Current systems are static, lack real time detection, and focus on detection in specific claim transactions without looking for patterns of suspicious behavior over time and the relationships and interactions between relevant entities.

Fraud control is a risk management activity akin to others but with its unique challenges. In June 2002, Donald Rumsfeld, then United States Secretary of Defense, succinctly addressed the challenge [2]: "The message is that there are no 'knowns'. There are thing we know that we know. There are known unknowns. That is to say there are things that we now know we don't know. But there are also unknown unknowns. There are things we don't know we don't know. So when we do the best we can and we pull all this information together, and we then say well that's basically what we see as the situation, that is really only the known knowns and the known unknowns. And each year, we discover a few more of those unknown unknowns."

This challenge of unknown unknowns faces program administrators looking to root out fraud in healthcare. New and existing data must be integrated, mastered, and utilized in new ways to discover the unknown unknowns. While we can be assured that this cat and mouse game will continue as fraudsters adapt to our new knowns and their countermeasures, we can undoubtedly make significant strides in detecting potential patterns of fraud, waste, and abuse in the system, ending these vulnerabilities and removing known bad actors.

In this paper, we applied Hevner et al. [3] to help us develop a framework for fraud detection in Medicaid that provides specific data models and techniques that identify the most prevalent fraud schemes and should help identify the unknown unknowns. The environment, discussed in the first section, includes the payers, providers, and patients. The knowledge base, covered in section two, is represented by fraud detection literature and the state of the industry. Based on this analysis, in the third section we develop a multidimensional schema based on Medicaid data and describe a set of multidimensional models and techniques to detect fraud in large sets of claim transactions. In the fourth section, these artifacts are evaluated through functionally testing against known fraud schemes. As in the domain of risk management as a whole, healthcare fraud control must address the unknown unknowns. This paper offers a set of multidimensional data models and analysis techniques that can be used to detect the most prevalent known fraud types and should prove useful in detecting the unknown unknowns.

2. Environment

The following definition of fraud from the US Department of Health and Human Services [4] will be used for the purposes of this paper: "Fraud is the intentional deception or misrepresentation that an individual knows to be false or does not believe to be true and makes, knowing that the deception could result in some unauthorized benefit to himself/herself or some other person." Within the healthcare system three main parties commit fraud [5]: healthcare providers, beneficiaries (patients), and insurance carriers. Providers are the initiating actor for billing insurers, and, as such, quickly become the nexus for fraud schemes. When a provider participates in Medicaid, the provider agrees to the reimbursement rates set by the state and submits claims for payment directly to the state or managed care entity. If the provider is not participating in Medicaid,