International Journal of Accounting Information Systems 21 (2016) 18–31

Contents lists available at ScienceDirect

International Journal of Accounting Information Systems

journal homepage: www.elsevier.com/locate/accinf

Outlier detection in healthcare fraud: A case study in the Medicaid dental domain

Guido van Capelleveen [a,1], Mannes Poel [b,1], Roland M. Mueller [a,c,2], Dallas Thornton [a,d,3], Jos van Hillegersberg [a,*]

[a] School of Behavior, Management and Social Sciences, University of Twente, The Netherlands
[b] Department of Computer Science, University of Twente, The Netherlands
[c] Department of Business and Economics, Berlin School of Economics and Law, Germany
[d] Clemson University, United States

ARTICLE INFO

Article history:
Received 2 August 2015
Received in revised form 31 March 2016
Accepted 4 April 2016
Available online 7 May 2016

Keywords:
Medical fraud detection
Decision support
Outlier detection

ABSTRACT

Health care insurance fraud is a pressing problem, causing substantial and increasing costs in medical insurance programs. Due to large amounts of claims submitted, estimated at 5 billion per day, review of individual claims or providers is a difficult task. This encourages the employment of automated pre-payment controls and better post-payment decision support tools to enable subject matter expert analysis. This paper presents how to apply unsupervised outlier techniques at post-payment stage to detect fraudulent patterns of received insurance claims. A special emphasis in this paper is put on the system architecture, the metrics designed for outlier detection and the flagging of suspicious providers which may support the fraud experts in evaluating providers and reveal fraud. The algorithms were tested on Medicaid data encompassing 650,000 health-care claims and 369 dentists of one state. Two health care fraud experts evaluated flagged cases and concluded that 12 of the top 17 providers (71%) submitted suspicious claim patterns and should be referred to officials for further investigation. The remaining 5 providers (29%) could be considered mis-classifications as their patterns could be explained by special characteristics of the provider. Selecting top flagged providers is demonstrated to be a valuable as an targeting method, and individual provider analysis revealed some cases of potential fraud. The study concludes that, through outlier detection, new patterns of potential fraud can be identified and possibly utilized in future automated detection mechanisms.

1. Introduction

Roughly $700 billion of the $2.7 trillion spent annually in the US healthcare system is attributable to fraud, waste, and abuse (Kelley, 2013). US health care costs represent 17.6% of Gross Domestic Product (GDP), compared to a 9.5% average among other nations (Organisation for Economic Co-operation and Development, 2012). Medicare and Medicaid, government-run health insurance programs designed for the elderly and those with low income and resources, supported over 72 million individuals and paid

* Corresponding author at: University of Twente, PO box 217, 7500 AE Enschede, The Netherlands.
E-mail addresses: g.c.vancapelleveen@utwente.nl (G. van Capelleveen), m.poel@utwente.nl (M. Poel), roland.mueller@hwr-berlin.de (R.M. Mueller), dallas@clemson.edu (D. Thornton), j.vanhillegersberg@utwente.nl (J. van Hillegersberg).
[1] University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands.
[2] Berlin School of Economics and Law, Badensche Strasse 52, 10825 Berlin, Germany.
[3] Clemson University, 120 McGinty Court, Clemson, South Carolina 29634, United States.

http://dx.doi.org/10.1016/j.accinf.2016.04.001
1467-0895/

for roughly 1/3 of the nation's health care costs in 2012 (U.S. Department of Health and Human Services, 2013). Any program of such size is a target for fraud. Not surprisingly, the Government Accountability Office (GAO) designated Medicare & Medicaid as high-risk programs due to their size and systemic complexity (U.S. Government Accountability Office, 2012). In combatting health care fraud, insurers and governments must deal with fraudulent practitioners, organized criminal schemes, and honest providers who make unintended mistakes. The relations involved in physician participation in government programs make it harder to exclude problematic providers than it is in privately managed provider networks. While much has been spent by governments in the Health Care Fraud and Abuse Control (HCFAC) program, the impact of these efforts could be considered marginal at best (Sparrow, 2000). Data analysis methods deployed in other sectors are not yet widely utilized in this domain. This has been blamed, in part, on the high level of subject matter expertise needed to adapt these techniques to the health care field and the idiosyncrasies of a third party payer system. However, with up-front engineering and ongoing adaptations, techniques such as outlier detection are suggested as effective predictors for fraud and offer a lifeline to programs struggling to rein in spiraling costs and remain solvent (Travaille et al., 2011; Bolton & David, 2002; Li et al., 2008). While a rich literature base exists on data mining and outlier detection techniques (Aggarwal, 2013; Chandola et al., 2009), less is published about the methodical application and evaluation of outlier detection to health care data sources. In this study we use a multi-dimensional data model for Medicaid claim data (Thornton et al., 2013) and apply a seven step methodology (Thornton et al., 2014) in a case study of outlier detection applied to one state's Medicaid dental claims. We thus provide a detailed case study application and evaluation of the methodology and multi-dimensional model presented in our previous work. This paper contributes to the literature by showing how outlier techniques can be used in health care to target potentially fraudulent activity. It shows that, through outlier detection, new patterns of potential fraud can be identified and potentially utilized in future automated detection mechanisms.

2. Research domain

The section starts with a description of the related literature of data mining for medical fraud detection. Second, medical fraud is put in the context of the Medicaid program, the claim processing is described and an outline of the current fraud detection mechanisms is given.

2.1. Related work

Advances in information technology, digitalization of health care information, and research on health insurance fraud have opened the door to use data mining and machine learning to fight fraud. While data mining is used by researchers as a tool to detect fraud (Aral et al., 2012), electronic fraud detection may also be used as a safeguard, securing the pre-processing of claims by identifying any irregularities or analyzing processed claims post processed to search for indicators of fraud (Aral et al., 2012; Forgionne et al., 2000; Bolton & David, 2002; Ortega et al., 2006). Unfortunately, health care is greatly lagging behind other industries such as banking and telecommunications in the use of statistical analysis and data mining methods (Travaille et al., 2011). The slow adoption of such techniques can be explained by the complexity of the health care industry, siloed claims processing systems and low political support and funding of fraud detection initiatives (Sparrow, 2000). Though previous research identified use cases of data mining to reveal fraud schemes (Forgionne et al., 2000; Major and Riedinger, 2002; Shin et al., 2012; Musal, 2010; Ng et al., 2010), application to a larger and broader medical domain for effective use by fraud experts remains a challenge.

A comprehensive survey of data mining-based fraud detection research includes provides a decade of data mining-based fraud detection research and proposes alternative data and solutions from related domains (Phua et al., 2010). Outliers were initially identified as a basic form of nonstandard observation, that could be used for validation of data quality as one may detect accidental errors, it could also reveal falsified data or data linked to a fraudulent pattern (Bolton and David, 2002). While outlier techniques do provide the opportunity of supervised learning, the prevalent argument for using hybrids or unsupervised methods are based on the low availability of fraudulent 'training' cases, the continuously changing program policies and the displacements of fraudulent activity to different fraud schemes.

Multiple uses of electronic fraud detection have been reported by various researchers in different fields. In the early 2000s, data warehousing for data mining purposes in health care became prevalent (Forgionne et al., 2000). Intelligent data mining systems to detect health care fraud were used in data warehousing, data mining, artificial intelligence and decision support systems to develop a proactive and effective health care fraud detection strategy. Larger scale application was researched by researchers who developed an electronic fraud detection application to review 20,000 providers on 27 behavioral heuristics and compare those to similar providers (Major and Riedinger, 2002). A provider score based on heuristics was calculated followed by a frontier identification method to select providers as candidates for investigation. Although the research alerted officials on almost 900 suspicious providers, only 91 (10%) were identified by experts as potentially fraudulent and warranting further investigation. Another example is the experimental application was found in a study which identified a number of rare cases in pathology insurance data from Australia's Health Insurance Commission using an online unsupervised outlier detection algorithm (Yamanishi et al., 2004). In Canada, researchers used Benford's law to detect anomalies in claim reimbursements (Lu and Boritz, 2005). Although the method did find some remarkable behavior, its potential for fraud identification was limited. One of the main reasons is that Benford's law looks for the frequency distribution of the first (and second) digits. This does not necessarily work for services with payer-fixed prices. In Taiwan within the National Health Insurance (NHI) program scientists developed a detection model based on a process mining framework that systematically identified practices derived from pathways to detect fraudulent claims (Yang and Hwang, 2006). The examples returned by the unstructured detection model captured 69% percent of the examples on

average. The empirical results showed that the proposed detection model was efficient and capable of identifying fraudulent and abusive cases within clinical instances. In Chile, a private health insurance company built applications using neural networks to find medical abuse and fraud. The application used an innovative method that could process the claims on a real time basis. The researchers reported a detection rate of approximately 75 fraudulent and abusive cases per month making the detection 6.6 months earlier than without the system (Ortega et al., 2006). The system was able to retrieve 73.4% of the fraudulent billings, while having a minimal false positive rate of 6.9%. In Medicare Australia, association rule mining was used to examine billing patterns within a particular specialist group to detect these suspicious claims and potential fraudulent individuals (Shan et al., 2008). Through domain experts the identified associated results were tested. The subject matter experts rated the rules into low, medium or high categories and providers were measured on the occasions they broke those rules. According to the fraud experts, the medium and higher rules, may be directly related to non-compliant practices and could be used as a measurement of effectiveness. The research reported an accuracy of 20.8% of providers with more than 5 violations, which is more effective for identifying suspicious billing patterns than random sampling. For US Medicare, two models were developed to investigate fraud that used clustering procedures as well as regression models for geographical analysis of possible fraud (Musal, 2010). They argued for a dynamic system approach to analyze decisions involving the investigation of possible fraudulent providers. Another attempt in Medicare Australia was to detect prescription shoppers or non-compliant consumers in spatio-temporal health data using multiple metrics that could flag providers (Ng et al., 2010). A modular framework that brings disparate data mining techniques together was adopted and showed high success rates. Of the 12 people identified, 8 are believed to be prescription shoppers, 4 with high confidence and 4 potentials. Although beneficial experimental results were achieved and the authors consider spatial and temporal factors to be effective in metrics, significant benefits using spatial–temporal factors instead of more traditional metrics could not be verified. The more simple metrics such as multiple visits or prescription percentages of pharmacy visits for drugs of concern have proved valuable as well. In addition, in Medicare Australia, they addressed the problem of prescription shopping using a different approach (Tang et al., 2011). The researchers integrated techniques like feature selection, clustering, pattern recognition and outlier detection. Using a threshold on the outlier score provider groups could be marked as potentially fraudulent. Another study described a methodology for identifying and ranking candidate audit targets prescription drugs fraud (Iyengar et al., 2013). The researchers developed a normalized baseline behavioral model for each prescription area and searched for statistically significant deviations from that model. For some of the areas, up to 500 features were used to find anomalies. In one the narcotic analgesics drug class, one of the experiments, all the known cases of fraud were correctly identified by the model as being very abnormal and excessive. In Brazil, researchers proposed a model for assessing the behavior of providers engaged and used k-means clustering algorithm as framework for identifying outliers in order to detect excessive billing of medical visitation at claim data of a Brazilian health insurance provider (Hillerman et al., 2015). Lastly, we found a study that proposed a seven step process, based on a literature review, to evaluate the potential efficacy of mining health care data was proposed in (Joudaki et al., 2015).

In general, previous research has focused on applications using different techniques in multiple sub-domains of medical insurance, justifying the applicability of data mining techniques to detect health care fraud and has increased the awareness of the insurance industry about the possibilities of data mining in this field. In order to apply data mining in a general way that can support fraud experts in their task of allocating resources, we conclude from these papers that there are several difficulties that have to be overcome when applying these techniques. First, the continuous nature of fraud results in changing fraud schemes over time. Second, the complexity of the health insurance domain requires subject matter expertise on each topic to initially identify fraudulent behavior. Third, the structure of insurance policies and state regulations affect the behavior of providers and fraud schemes. Thus, other research has mainly focused on very specific medical fraud domains. Furthermore, outlier detection has hardly not been applied to health care fraud detection. Therefore in this paper we show how to apply unsupervised outlier techniques at post-payment stage in order to detect fraudulent health insurance claims.

2.2. Medical fraud

To detect fraud, one needs to understand the complex medical insurance industry, associated claim processing and potential fraud schemes. Fraud can be defined as “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” (U.S. Department of Health and Human Services, 2011). In Medicaid three groups can be involved in a fraud: patients, insurers and service providers (Li et al., 2008). The focus will be on the latter as providers initiate fraud schemes by billing insurers fraudulently. They are the nexus for fraud schemes, though others may defraud as well.

In order to address fraud, one should consider that fraud detection is difficult due to the uncertainties and inconsistencies inherent in medical care (Henderson, 2009). Moreover, fraud in healthcare is dynamic as detection methods continue to improve. As systems improve those participating in fraud typically move on to find new ways to exploit weaknesses in the system. The most prevalent health care fraud schemes include billing for services not rendered, upcoding, duplication of claims, unbundling of claims and providing excessive or medically irrelevant services (Sparrow, 2000). Billing mistakes, such as miscodings or charging the wrong patient, by honest providers need to be detected and recouped but excluded from the fraud definition.

2.3. Medicaid claim process

When a provider participates in Medicaid, the provider is reimbursed by the state and submits claims for payment directly to the state or managed care entity. If the provider is not participating in Medicaid, the provider sends the patient the bill which he

or she has to pay before requesting reimbursement for partial payment from Medicaid or the state Medicaid insurer. In both scenarios, the agency or insurer processes the claim and sends an explanation of benefits (EOB) to the patient that describes the services paid for along with their codes and costs. After submission, claims processing systems perform various prepayment checks and edits to inspect the claim's legitimacy (Sparrow, 2000). Examples are form validity, cross procedure checks, pricing range validity, re-submission or duplication prevention. The systems however do not verify whether the service was provided as claimed, the diagnosis were correct or if a patient is even aware of the claimed services. Without follow-up, re-submission is possible and eventually fraudulent claims will pass inspection.

EOBs, while well-intentioned, in their current form provide minimal protection against fraud (Sparrow, 2000). The beneficiary has no financial incentive to pay attention to a list of complex computer-generated forms and billing codes delivered to their mailbox with no balance due. In addition, many fraud schemes deliberately target vulnerable populations in Medicaid, such as the homeless, mental health patients, and those with disabilities that would be unable to understand the EOB or are paid kick-backs from the providers to remain quiet and not to complain (Kelley, 2013).

The prevention and fraud detection initiatives in Medicaid are organized at the State level within agencies and contractors. Fraud is typically detected and recovered through audits performed by one of the agencies or contractors. While most cases are revealed through audits, either randomly selected or found through submission inconsistencies or structural monitoring, the system still relies for a large part on filed cases under the false claim act (U.S. HHS and DOJ, 2014). Fraud analytic initiatives in Medicaid are starting to become more prevalent (Centers for Medicaid and Medicare Services, 2014). Many argue that the electronic fraud detection may be intensified and contributes by securing the claim input process, checking for irregularities and analyzing claims searching for indicators of potential fraud (Aral et al., 2012; Forgionne et al., 2000; Bolton and David, 2002; Ortega et al., 2006).

3. Research design

The research design used in the study consists of three steps. First, potential relevant metrics were identified through a literature study. Second, a representative data set was composed to assess the relevance of the metrics. Third, the case study was evaluated through expert interviews. Given the limited research on outlier techniques in health care fraud, we consider this a practice-based problem in which experiences of actors and context are important (Benbasat et al., 1987; Yin, 2008).

3.1. Metric identification

We refer to metrics as a collection of measurements of data attributes, features, aggregates or derivations that profile provider behavior. Two sources were used for this purpose. The FBI Federal fraud news reports (U.S. Federal Bureau of Investigation, 2013) retrieved from the available periods 2009–2013 described fraud cases at federal level. In addition, 53 editions of fraud reports from 2004–2012, published by the National Association of Medical Fraud Control Units (NAMFCU, 2013), covered state prosecutions. Both sources provide insight on how the fraud was discovered, what fraud schemes were used, which claim patterns were found, and how the providers were prosecuted.

An example of metric derivation was using a fraud case in New Jersey. Here a physician and owner of a home-based services for seniors business pleaded guilty for charging lengthy visits to elderly patients that they did not receive (U.S. Attorneys Office, District of New Jersey, 2013). The physician received at least $500,000 before he was eventually detected. He came to light after becoming the highest billing home care provider among more than 24,000 doctors in New Jersey. Intentionally up-coding for services, is a typical fraudulent behavior that can be detected by a peer comparison of the ratio of lengthy visits to all billings.

With over a hundred metrics were derived, 14 metrics presented in Table 1 were selected for our case study. The metrics were selected in consultation with our experts based on their applicability to the dental domain (the case study) and likelihood for usefulness in fraud detection. Related metrics were categorized based on the types of fraud they were likely to uncover. The

Table 1
Overview of metrics and the outlier techniques used.

Metric	Method	Outlier detection
–Reimbursement per beneficiary –Number of reimbursed claims over time	Linear model outlier detection	Trend deviation above threshold
–Number of reimbursed claims over time –Amount of reimbursed claims over time	Linear model outlier detection	Deviating trend from peer group
–Proportion of weekend claims	Univariate outlier detection	Z-score above threshold
–Average number of reimbursed claims per beneficiary –Average amount reimbursed per beneficiary –Average number or reimbursed visits per beneficiary last 12 months –Amount of beneficiaries with high number of yearly visits –Average number of reimbursed procedures per claim –proportion number of reimbursed high cost claims	Multivariate outlier detection, cluster analysis	Mahalanobis distance above threshold, deviating cluster, deviation from nearest cluster
–Procedure code –High cost procedure - Tooth code	Box-plot outlier detection	Peak deviation above threshold

categories were labeled with the associated data mining method/outlier detection technology in Table 1. The metric, or combination of metrics were plotted using scatter plots in R (The R Foundation, 2015) to visualize the distribution of data. Next, using distribution algorithms, clusters, or linear models, boxplots were created for depicting groups of numerical data and outliers. Because the dental domain is much less varied than many other specialties, metric scores in relation to the organizational size or claim submission sizes were assumed to be normally distributed. Therefore, the outlier technique relied on a Gaussian distribution of data. Though data points did not follow this exact distribution, each metric was assigned with its own outlier criteria, set in standard deviations from the mean. Following a normal distribution, outliers were selected at one tail, 1.96 standard deviations from the mean. The tail with higher metric scores was selected, as fraud is typically characterized by over-utilization of claimed resources. Following this strategy, we would generally aim to select around 2.5% of the providers. In cases where the data was not normally distributed, we would increase the outlier criteria to 2.33 standard deviations to limit the number of providers and focus on what might be the more extreme cases more likely to be fraud.

3.2. Data collection

Next, we gathered dental claims data to be used in experimenting with our metrics. A prototype for fraud analysis and visualization was developed, cf. Section 4. A representative data set was composed from the Medicaid dental claims data from one state. Dental claims fraud has not been widely covered in literature despite over $100B in US dental spending annually. Dentistry is also known as a large and relatively homogeneous group of providers, which makes it particularly applicable for peer group analysis. Moreover, the availability of dental domain knowledge was present and metrics and results could therefore be evaluated with experts. 11 months of claims from a single state Medicaid program were analyzed. In consultation with experts, we set requirements on the data set to minimize external influencing factors that could skew the data. This included that,

- The set contained all claims submitted for the given period.
- It included all adjustments that were made on the provided claims.
- No severe changes were implemented in the Medicaid State policy during the time frame of the data set.

3.3. Interviews with experts

In the third step, the goal was to gain insight into the results of our metrics value and usability. The usability of the metrics was assessed by conducting interviews with fraud experts. The interviewer prepared a semi-structured interview protocol and recorded the session for analysis. We felt that an open-ended protocol was optimal, allowing the experts to comment both on the facts and provide their experience-based opinions on the results (Yin, 2008). We discussed the design of metrics and the patterns that were found, allowing the experts to opine on the meaning of the findings and the likelihood of fraud discoveries. We invited experts from two different organizations working with Medicaid. The interview was held with the two subject matter experts simultaneously, and they were encouraged to discuss the results during the interview. Both fraud experts had a role in fighting fraud at a national level and also had specific knowledge of the state analyzed.

4. The fraud detection architecture

An infrastructure was required in order to conduct the research on claims analysis and consolidate the results into a usable format. For this purpose, an initial version of a fraud detection architecture was developed. A cube contains our multidimensional data model while a second data store contains the fraud metrics results. The combination of two sources is used to integrate data and metrics into a fraud analysis and visualization tool. The tool uses a data warehouse infrastructure of the Centers for Medicaid and Medicaid Services (CMS) created at for this purpose. The architecture is presented in Fig. 1.

Three types of data files were loaded into this architecture; provider claims at individual state format (Medicaid Management Information System), federally determined state statistical extracts (Medicaid Statistical Information System), and reference files sourced from several states, CMS and other governmental agencies. Examples of reference data include birth registrations, death records, medical provider registrations and criminal records. The received files were frequently distributed, processed (extracted and transformed) over multiple nodes in parallel loaded in a single staging environment.

Data integrity and completeness had to be secured in this process. To do this, adjustments claims were processed and merged into the originals. Entries containing incorrect data such as null values, zero dollar payments, adjustments without original claims and claims with future servicing dates were removed. Duplicate detection, data reference checks and format checks procedures were ran over the data. Data files were validated for completeness by meta data check such as row counts and data structure validations. Finally, data was transformed according to the schema of the data warehouse and if successful, loaded in the production environment. This production environment could then be used for metric calculations and analysis by fraud experts.

Next, metric calculations were processed. Because of its computationally intensive consumption, results were stored in tables that could be queried and combined with claims from the multidimensional model. Furthermore, scripts were developed to calculate and compare providers based on their metric scores by fitting logical models, k-means algorithms, and boxplots used in analyses. A parameter file then was used as input for the algorithms, to select data filters, set outlier criteria, write back capabilities, and visualization types for each of the experiments. Eventually, a fraud analysis and visualization environment was created. Now, fraud experts can investigate flagged providers, drill down at claim level, browse and compare outlier scores of providers.

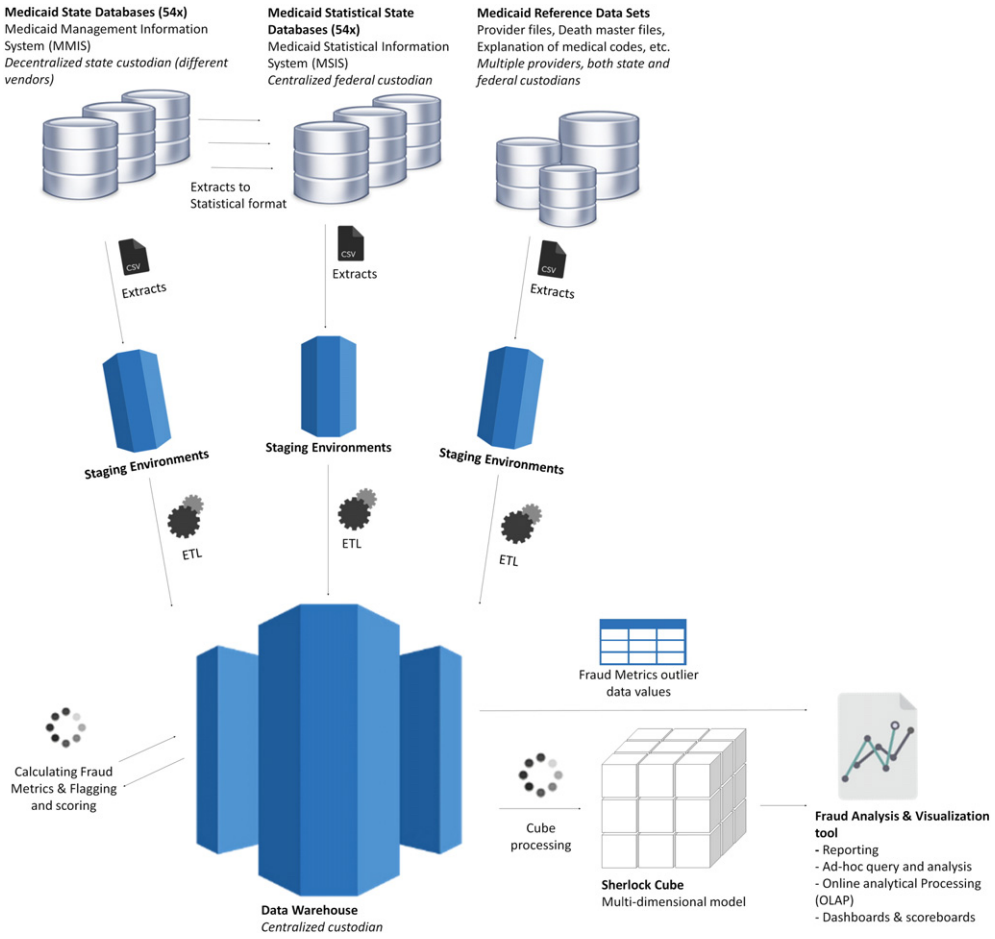

Fig. 1. Fraud detection architecture.

Highlights could be given to providers with multiple flags and alert fraud experts proactively. Though the presentation of fraud results in this prototype was visualized statically and fraud experts used traditionally querying tools for their specific questions, the use of interactive dashboards or querying tools may extend the targeting support for fraud experts (Dilla and Raschke, 2015).

5. Results

In this section we present the metrics derived from literature applied to this study and classify them by method and outlier detection technique as reported in Table 1. Secondly, 14 experiments have been performed of which some examples are provided that illustrate the fraudulent behavior that was found with the help of individual experiments. Furthermore, an overview on the overall flagging results is inserted that was used to see if target selection using scoring could be effective for target selection. To conclude, a summary of the experts evaluation results is given.

The experiments all started using the same set of data which then was filtered by unique criteria. Two times filters were applied during this process. The first occurrence was previously described in Section 4, applied at the data loading stage, covering the integrity and completeness of data. The second time filters were applied, the goal was to prepare data for valid peer group analysis. Providers that held a small quantity of claims, low amounts of reimbursements or few unique patients, were excluded. Generally a minimum of $10.000 of reimbursed claims or at least 10 unique beneficiaries per month were chosen as requirements for analysis. The resulting data set consisted of 369 providers and was the basis for our analysis. In addition for some of the experiments, such as the procedure code analysis, a required minimum service amount per year was needed to exclude those providers with low quantities for these properties. Limiting the set of claims contributed this way to the validity of peer group analysis.

Multiple analysis techniques were used in the experiments. These included variant analysis, multi-variate analysis, time series analysis and boxplot analysis (For specifics see Table 1). In addition to the analysis techniques, a detection method was used in each experiment. The following outlier detection methods have been used: deviation from linear model, deviation clusters, single deviations from clusters, trend deviations, and peak deviations, making use of both non-parametric and parametric (Gaussian mixture models) deviations.

In each of the experiments, criteria were outlined to define the outliers. In the linear model analysis, a deviation of more than 2.33 standard deviations from the underlying general linear model was considered to be an outlier. In the variant analysis, because outliers were close to each other, it was regarded as a group and an outlying cluster algorithm was used. In the multivariate analysis, using k-means clustering, outliers were defined by the outlier criteria of single data points deviating more than 2.33 standard deviations in y direction from its belonging cluster, or if a cluster contains less than 5 items. The number of outliers in our experiments were usually not significant enough to perform an outlying cluster analysis. Clusters were formed using the k-means algorithm, set to 10 iterations. In the box plots, the interquartile ranges were used as outlier criteria and configured for each metric separately. Within most experiments, multiple outliers presented themselves. Outliers influence sample means and deviations and could therefore mask themselves. This masking effect could be reduced by robust estimation procedures (Rousseeuw and van Zomeren, 1990). However, we did not use any unmasking procedures in this study because the set of outliers we intended to select in our experiments seemed to deviate sufficiently. Any deviating specific criteria related to an individual metric, if used, are described when discussing the metric further on.

According to the followed fraud method, a scoring mechanism should be used to identify the targets to be selected for fraud experts to investigate. The reported scoring formula considers the use of metric importance and history. In this study, the history was ignored due to the limited length of the data set. As only one initial full cycle has been completed, flags for each of the metrics were equally weighted to evaluate their impact and relevance, because no previous knowledge existed to value the metric importance. The result of flags from all experiments is summed and determined the score a provider received.

5.1. Outliers based on linear model

Fig. 2 shows an outlier analysis based on deviations on a simple linear model. On the two axles the relationship is measured between the total dollar amount reimbursed and the number of reimbursed claims of a provider. The red line is the fitted general linear model (GLM) through the data points, achieved by applying the linear model function from R, designed for simple linear model analysis. No offset was set in the linear fitting, as there was no intention for corrective behavior of coefficient. Also, the GLM did not have to consider NULL values, as we removed them earlier. The blue lines represent the 2.33 standard deviation from the logical model. Provider 23,481, plotted in the left top corner, was one of the providers that attracted attention because of its severe outlying behavior and was an interesting candidate for further analysis to find the cause of deviating average of this providers reimbursements.

The provider submitted just over 200 claims that month of which 30 were considered high cost claims. The main high cost claims were mainly reimbursements for complex comprehensive orthodontic treatments (codes D8080, D8090 and D8999). Because all dentist with a specialty type were excluded in our study, claiming many specific high cost procedures attracted attention. However, arguments were given by the fraud experts that there are possibilities of dealing with non-fraudulent behavior. Sometimes provider enrollment registers are outdated or the specialty of a provider is classified as non-specialty under Medicaid program regulations. In the flagging results one may see, that specific provider received 6 flags eventually and was classified by the fraud experts to be a case for formal investigation.

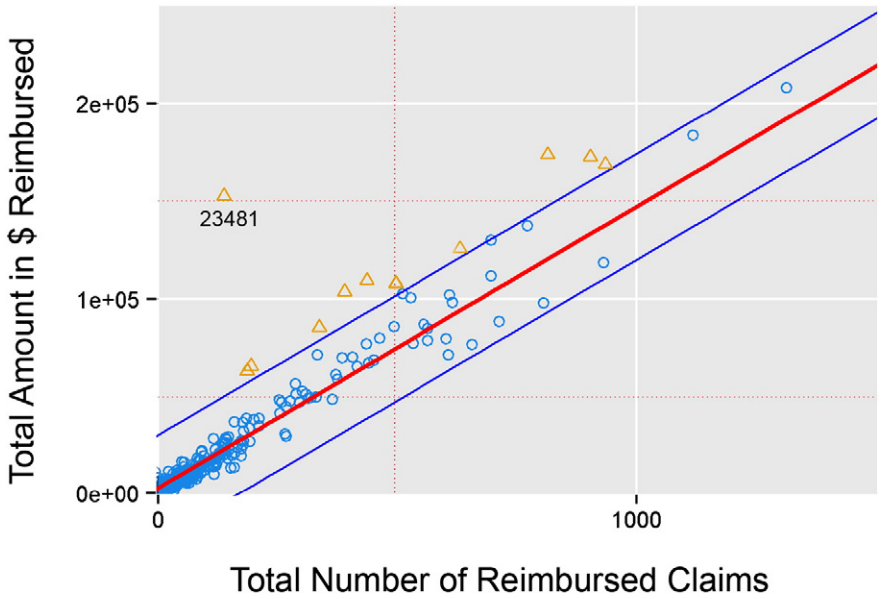

Fig. 2. Example outliers on a simple linear model.

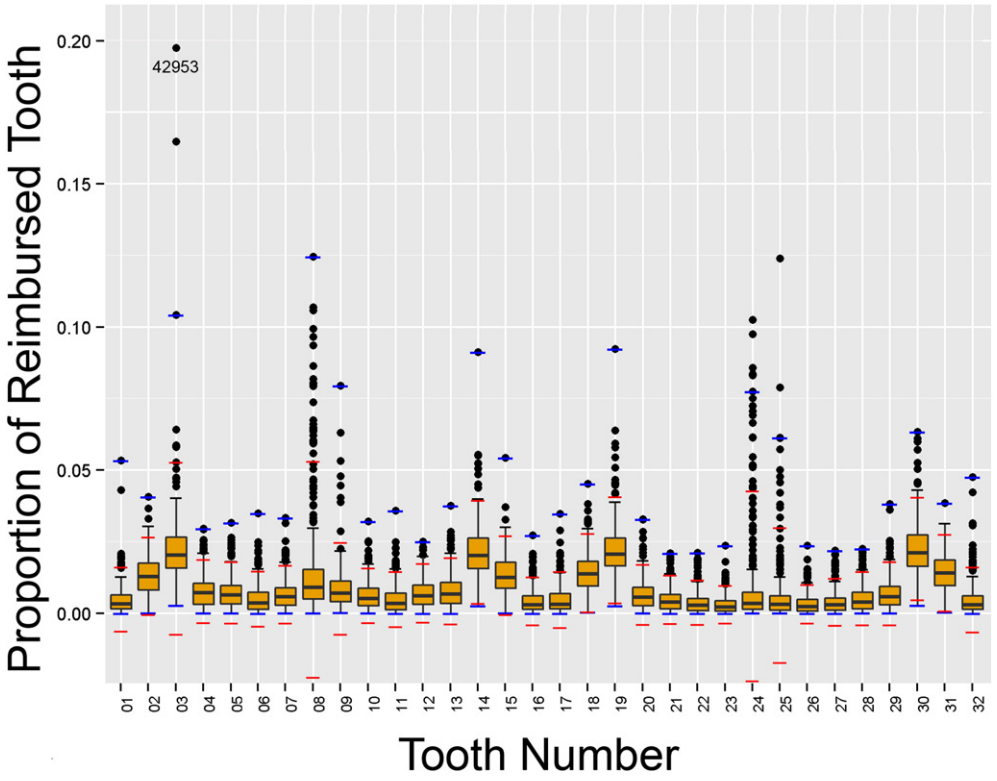

Fig. 3. Tooth code analysis (showing adultery teeth).

5.2. Boxplot outlier detection

Within the tooth code analysis providers are compared by the percentage of dental claims claimed for specific tooth codes. The idea for detection through high percentages of claims on certain teeth was inspired by one of the studied fraud cases. It reported that some dentists claimed over and over for the same set of procedures by only changing patient IDs in order to reimburse as much as possible with the least effort involved. For example, providers that substantially claim more exploiting fraud schemes among other phantom billing, duplicate billing or unbundling of claims, stand out when they do not adjust the properties of claims randomly, by higher proportions on specific tooth codes reimbursed. Another type of fraud that might be revealed from this analysis is the recursive treatment on a tooth. First the dentist fills the tooth with amalgam, follows up with a correctional procedure, after has to pull the tooth and places a replacement. Thought the procedure might be legal as it may result from misdiagnoses, treatment errors, or a medical reason, the level of occurrence of these situations are estimated to be low on the overall patient base, especially if one patient receives multiple in a short period.

Fig. 3 shows such analysis by presenting a boxplot for each tooth code. Within the analysis, only the permanent adultery teeth are shown, thus excluding child teeth and supernumerary teeth. The teeth are numbered started on the upper side of the mouth, counting from left to right. The outliers, represented by black dots, are those exceeding the fourth quartile. However, as still many providers report slightly above of this quartile, we used the k value of the boxplot formula that determines the upper fence to increase the value of the outlier criteria. Normally, the k value is set to 1.5 (black whiskers), however we increased it to 12, represented by the blue whiskers on top, or above the boxplot.

$$Outliercriteria = Q_3 + k \cdot InterQuartileRange$$

Consider provider 42,953 that claimed over 140 procedures, nearly 20% of its total dental claims, using tooth code number 03. After profound analysis at claim level, we found that the claims were submitted for multiple patients, in general about one or two of these procedure patterns per patient. Most of the procedures used code D0120, a periodic oral evaluation for an established patient. The pattern continued during the whole period, following likely the 'steal at little, all the time' fraud tactic. Though, the metric was found useful to form a conjecture of fraud, this particular provider did not receive any other flags in the experiments.

A similar kind of boxplot analysis was performed by using the procedure code attached to a claim. Fig. 4 shows a boxplot for each procedure code that was submitted. For this analysis an additional criterion of at least 300 claims (each claim has a procedure code) has been used. This would remove all providers not suitable to compare their percentages with because of a low

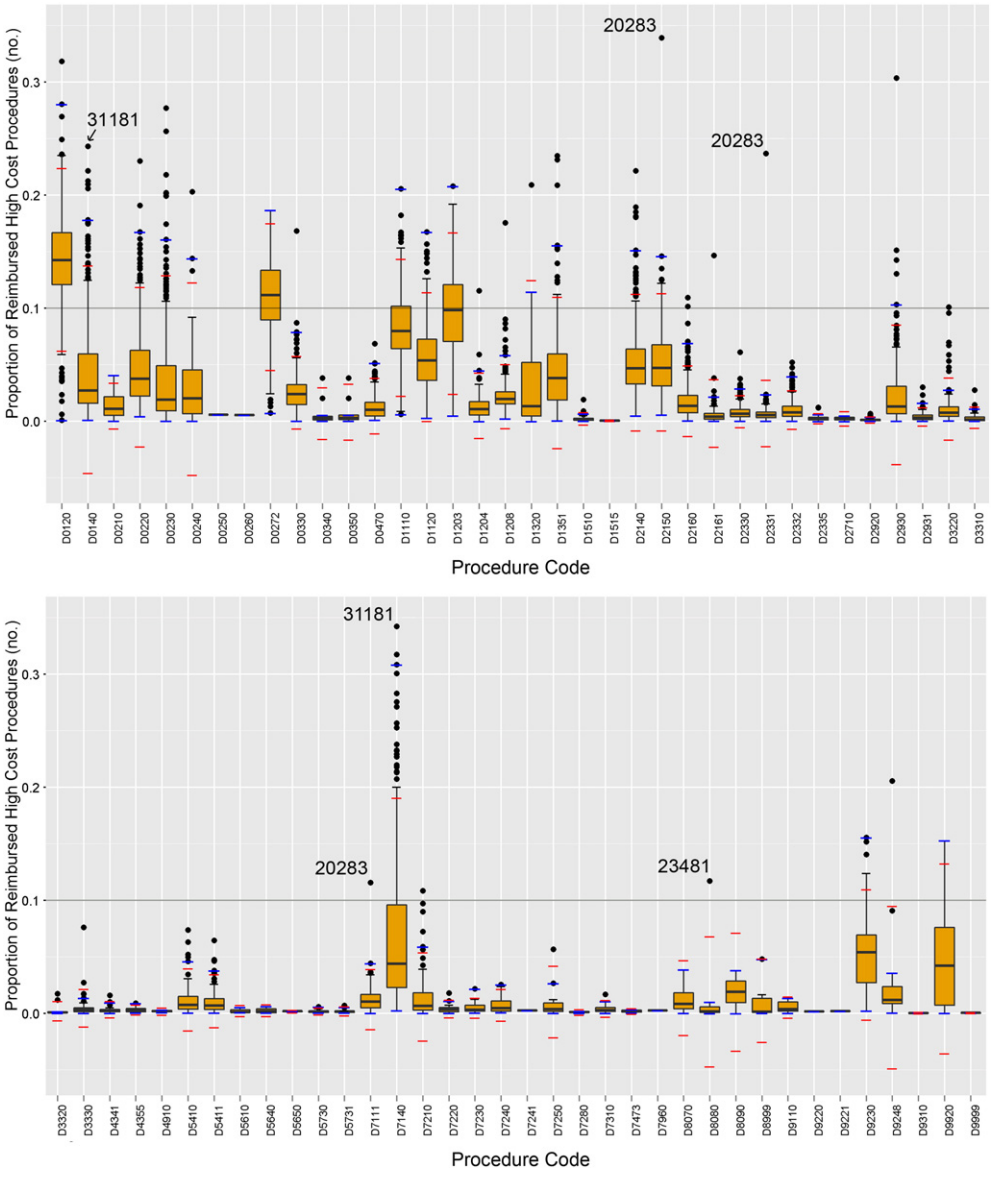

Fig. 4. Procedure code analysis.

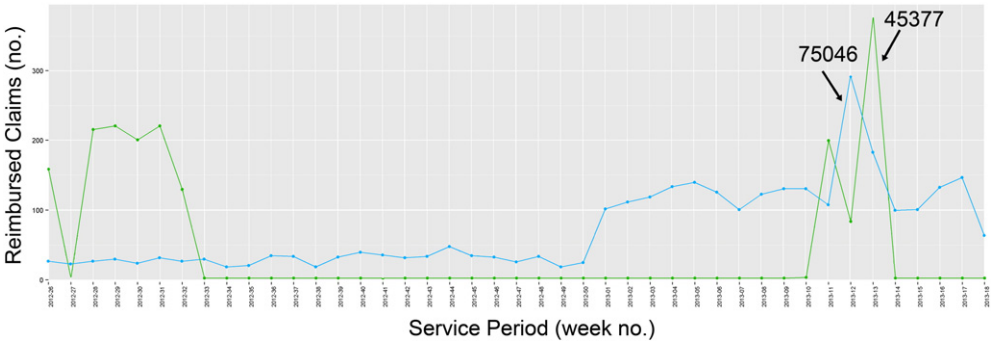

Fig. 5. Peak analysis: time series with outliers of reimbursed claims.

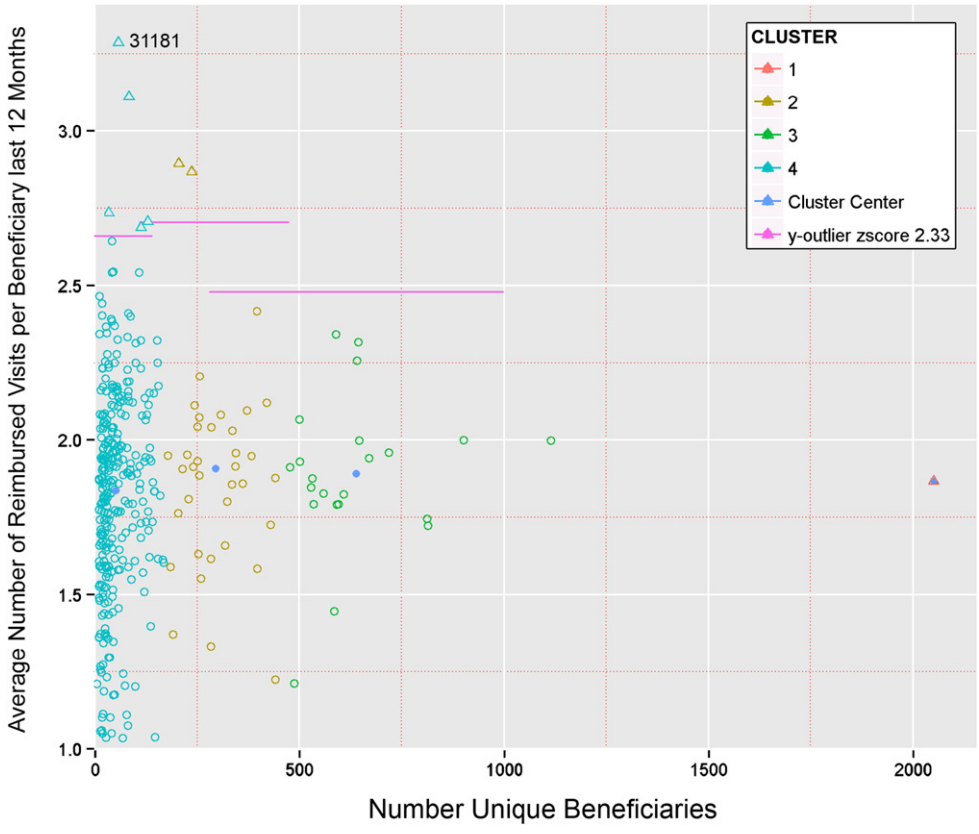

Fig. 6. Multivariate clustering and outlier analysis.

procedure percentage certainty. The outlier formula that was used in the tooth analysis was again used in this experiment, however a lower k value was necessary. We determined to assign k the value 3 in order to select those single dots above the still dense regions of dots that exceeded the first black whisker in majority of plots. Fig. 4 shows three boxplots where provider 20,283 is highlighted. No other provider used procedure code D2150 (Amalgam), D2331 (Resin-based composite) and D7111 (Extraction, coronal remnants) this excessive. The provider eventually received 5 flags in total, and was classified by the fraud experts to be a case for investigation.

Another provider that received our attention is provider 38,606, who claimed over 40% of his procedures on examinations (D0140). Many patients, in the months before and after their tooth adjustment, were examined six to seven times. There was even one patient that received 15 examinations. A different example was a tooth needed to be pulled, followed by 6 examinations and a second tooth pulled eventually. Though the series of claim occurrences is not fraudulent by itself and this particular time it could be explained well, it does form suspicion when multiple of such cases present in a short time frame, representing the greater part of claims of a provider. The provider in question only received one flag in our analysis.

5.3. Outlier detection based on peak analysis

Fig. 5 shows the time series of claim behavior of the 2 of the providers receiving a flag in the peak analysis. In the peak analysis, we searched for sudden increases or decreases in the number of claims submitted per week. We selected the peaks when the number of claims of a provider at least doubled or halved compared to the previous week. These outliers are marked using a thick black dot in the picture.

One of these providers, number 45,377 did not submit claims for a while until week 10, and in week 13 it submitted claims for over 300 patients. An explanation for such strong difference could be that a provider bills under multiple provider IDs or had problems with his claim registration. More likely however, is that the provider corresponds to a mobile dental practice. The provider received one flag in total eventually.

Another interesting provider in this experiment is provider 75,046. Questionable is his sudden peak in week 12 of 2013. The number of claims increased from around 100 claims per week, to almost 300. A search for service code patterns in the claims raised suspicion. Many children visited the clinic that week and received exactly the same treatments only on different teeth. The combination of procedure codes reappeared so frequently is was worth investigating. The first pattern found, exists of an oral examination of a child patient, followed by a two bitewing film, two periapical films, prophylaxis and a fluoride treatment.

The second pattern used the same set of procedures, only meant for adultery and added exact three amalgam claims for each patient. The medical necessity for multiple films every time and the recurrence of three amalgams was found to be odd. The provider received 6 flags in total and was classified by the fraud experts to be a case for investigation.

5.4. Outlier detection based on multivariate clustering

Fig. 6 shows one of the experiments that used a combination of multivariate clustering and outlier detection. Smaller providers, based on the number of unique beneficiaries, present themselves usually with more distributed y values. Therefore we considered multivariate clustering. Two attributes were used each time, one metric assigned to multivariate analysis mentioned in Table 1, the other attribute was the number of unique beneficiaries. Clusters were determined before applying a normal distribution outlier analysis at the y-axle metric. Clusters should be more or less equally sized to keep sufficient peers in all groups. The k-means algorithm, one of the most simple algorithms which uses unsupervised learning method to solve known, evenly sized clustering issues, by clustering data around centroids, was used for this purpose. By finding the elbow in a function of the sum of squared error (SSE), the optimal number of clusters was determined. The strongest bend was found at 4 clusters. One cluster consisted of a single provider. This unusually large provider had no peers to be compared to. The provider is known as a large provider and therefore needs a different kind of fraud analysis. If the provider would have been relatively unknown, we should have searched for explanations for such large patient base. Because the provider has been assigned its own cluster, no flags could be given. Other options for similar situations are to remove such providers, or change the clustering that the provider will be added to the second 'largest provider' cluster. The pink line represent the 2.33 standard deviation in the upper y direction from the clusters mean. By exceeding this threshold providers were marked as outliers and received a flag.

Recurring visits can be analyzed with the help Fig. 6. A visit is defined by all claims a patient received within a day. One of the providers, number 31,181 with 4 flags total, received a score of 3.29, an outlying average of recurring visit rate compared to the rest of the cluster with a mean of 1.84 times a year. The score was caused by a high number of patients visits that received a follow-up oral evaluation. Another finding was that the provider submitted most of the claims for extracting teeth. This could also be noted from the flag received for a high percentage of extraction codes in the procedure code analysis. For some of the patients the provider extracted all of the patients teeth which together adds up as a rather expensive insurance claim. While dental extractions are usually necessary in order to install a denture it remains suspicious when dentures are provided frequently, not being a specialist. Though it may be part of a provider specialty, it is also known as one of the more practiced fraud schemes in Medicaid (U.S. Office Inspector General, 2015). Experts suggested when multiple patient receive such treatments, documentation should be requested to validate the legitimacy of the claims.

5.5. Evaluation by experts

A target percentage should normally be chosen based on a strategy concerning the availability of fraud expert resources and the time allocated for investigation. As no heuristics for the target size exist so far, we selected 5% as we found it to be a reasonable target to process. Table 2) presents information on the flags that were given during the experiments. The results are divided in two groups, those with zero, one or two flags, and those with three or more. The group with three flags or more, made up for around 5% of the total providers. Though we would have liked to review all flagged cases, the extensiveness of this task would have been too time consuming. Therefore in the other group only a couple of interesting cases based on extreme outliers were selected to see if fraud could be possibly detected here as well. Due to the selective sample, no target success here is given. Within the experiments, 369 providers were able to receive flags. 106 providers (28.7%) received at least 1 flag, 35 (9.5%) providers more than two and 17 (4.6%) 3 or more. To evaluate the potential efficacy of our approach we focused on these 17 providers.

Table 2
Flagging results.

A 352 of 369 providers received 2 or less flags, a sample of extreme outliers was analyzed.

Analyzed	Flags received	Number of providers	Discussed in text
	0 flags	263	
Sample	1 flag	71	42953, 38606, 45377
	2 flags	18	
Total		352	

B 17 of 369 providers received 3 or more flags, all were analyzed.

Analyzed	Flags received	Number of providers	Reported	Discussed in text
	3 flags	8	4	
	4 flags	3	3	31181
All	5 flags	2	2	20283
	6 flags	3	3	23481, 75046
	7 flags	1	0	
Total		17	12	

We interviewed health care fraud subject matter experts to discuss the experiments and the extreme outliers in the individual experiments and the selected top 17 providers flagged. While some of the flags could be understood as acceptable given the types of services rendered or due to the providers operating environment, there was a preponderance of evidence suggesting that at least 12 of these 17 providers (71%) with three or more flags deemed appropriate for formal investigation. This meant, these met the criteria for a fraud expert to start an audit, retrieving documentation on the providers claims and to devote serious time on the investigation. The remaining 5 providers (29%) were mis-classifications. These outliers could be explained by some special characteristics of the provider or were not found strong enough to pursue the formal investigation. While some of providers with only one flag showed a potential for fraud detection, we felt there was a decreasing result in potential fraud.

Fraud experts noted, that outlier detection has advantages in revealing fraud and may be more effective than expensive periodical review, but there are some limitations. Firstly, outlier technology has not proven itself in the long run and is still in an experimental stage. Secondly, outlier detection is rather a more complex method than expensive review. Collaboration of technology and domain experts is required to design metrics and especially to interpret results. Thirdly, validation of effectiveness remains difficult. Condemnations form the most reliable source to validate a suspicion of fraud, however, before it comes to a condemnation, usually a long time goes by. Therefore experts rely more on heuristics and market expertise, but are searching for ways technology may support the work processes of fraud investigators. Where outlier detection may not yet be used for fraud detection as classification method, it does provide potential for indicative leads. Technology should therefore be seen as the enabling factor for interactive visualized technology, supporting the fraud analysis for program integrity units, that work together with attorney general and have abilities for formal investigation or on site audits. Visual interactivity allows the investigator to navigate large data sets, change the representation of data, filter transactions for further investigation and has therefore potential for making the detection of fraudulent transactions more effective (Dilla and Raschke, 2015). According to our experts, most promising results were found using boxplots analysis. Boxplots are relatively easy to develop with only one dimension application for all metrics are equal and may be repeated. They can also, usually with only small guidance, be used by fraud experts. Lastly, discussing the experiments, the task of interpretation of boxplot outliers was considered the one with least effort involved.

6. Discussion

Fraud detection in the U.S. medical insurance industry is a prevalent and costly problem. Outlier detection was shown as useful approach to reveal fraudulent providers, was found to be useful for targeting potential fraudulent cases, and was specifically promising as an interactive technology to guide fraud investigators.

Extreme outliers clearly revealed irregular provider billing activities which led experts to recommend further formal investigation. Based on the expert evaluation we learned that tooth and procedure code analysis showed the most promising results. Box plot outliers not only revealed many of the promising potential fraud cases, it was also the easiest methodology to utilize and simplest in terms of interpreting the results. The results indicated that flagging the results clearly exposes a pattern of potential fraud and is a key indicator of potential fraud, there is a correlation of the number of flags on providers and their fraud. Flagging fraudulent patterns may be an effective way of targeting potential fraudulent offenders. However, regardless of the results, reservations have to be made for the effectiveness of target selection, as true effectiveness only can be calculated when cases monitored over time and re-assessed based on fraud convictions. No such ability was present within our constrains, but can be subject for future research.

Some limitations were identified in this study. Firstly, given the small group of experts, the study had to base its evaluation on only two experts. Secondly, Medicaid program policies that effect the data source vary from state to state and based upon health insurance programs. Therefore the data partly influenced us choosing metrics, thresholds, clustering and even the detection methods themselves. These should be considered when interpreting or applying the study's results into different health programs, though we believe that most of our results, with no or small adaptations, may be easily transferred. The same holds for selecting the size of providers to target in relation to the detection rates to be found. We feel that the success of such experiments is subordinated in adjusting the settings that may be learned over time. While heuristics and domain expertise provide a great contribution at bootstrap, the research on this topic may be extended. The last identified limitation is the homogeneity of the dental domain in health care, that represented because of that a good candidate for outlier techniques. To evaluate effectiveness better, other domains with more complex claiming structure should be explored as well.

As we may have seen from the related literature, multiple data mining approaches to detect health care fraud exist and many of those reported were challenged with the detection of fraud. Were some researchers reported from only a few cases found (Major and Riedinger, 2002; Yamanishi et al., 2004; Shan et al., 2008), others presented higher detection rates above two third (Yang and Hwang, 2006; Ortega et al., 2006; Ng et al., 2010). It should be noted though, that detection rates reached may vary and depend on multiple factors used in studies, including the detection of 'potential fraud' instead of convicted cases. However, there seems to be a shared notion for the appliance of data mining methods, such as the outlier techniques we used, to target and detect fraud. From our point of view, outlier detection is a practical methodology to be used in peer group analysis that can be performed and automated when combining claims history. Using dashboards and visualization tools, problematic providers quickly stand out and raise flags for targeting.

7. Conclusions and future work

Fraud is a relevant and rampant problem in health care. The interest in data mining tools for the use of fraud detection in health insurance industry has gathered increased attention in the business world because of the bottom line impacts. Outlier

detection, as one of the promising fitting technologies for fraud detection, has not yet been widely researched in the health care domain.

This research presents a case study of applying outlier detection in practice to real data in the Medicaid dental insurance domain and utilized two experts to review the results of the analysis. The paper reports on an architectural design for the use of a fraud identification in health care. In addition, We listed 14 relevant metrics that have been identified from fraud case reports and relevant literature. Furthermore, using these metrics we conducted multiple experiments on the application of the outlier detection in a state-wide database of actual dental health care claims analyzing 369 providers. Some of the prominent patterns revealed during the analysis were discussed with fraud experts and illustrated within this paper.

Through this research, we learned many lessons about how to improve fraud prevention efforts. Significant health care subject matter expertise is required to design analysis techniques and interpret data mining results. Outlier detection, has been found a supportive tool for fraud investigators to detect potential fraud. 17 out of 369 (5%) primary dental providers were identified as warranting further analysis. Of these, 12 of 17 (71%) have been evaluated by experts and deemed appropriate for formal investigation. The experiments demonstrated that visualizations and outlier detection can support the identification of providers with unusual, potentially fraudulent claim behaviors. This approach could be used for the implementation of a decision support tool for investigators to more effectively target fraudulent providers.

As compared with prior comparative success rates of roughly 10% (Major and Riedinger, 2002), we see great opportunity in building upon this model in various ways. Future research will dive deeper, including evaluating specific outlier techniques relevant to types of health care fraud, and look more broadly at methods and models for storing and preserving metadata to allow for more automated scoring, model adaptability, and reconstruction. Longer term research should target success factors, potentially using a supervised outlier detection method. Transferring the methodology to less homogeneous health domains could be investigated to learn more on the adaptations needed for outlier techniques in more dynamic provider types.

The paper contributes to literature by providing a case study analysis that can be used in future applications of outlier detection in health care and potentially other corollary domains. We used the domain context of Medicaid and discussed considerations for its use in different data contexts. With this research we hope to both advance the state of the art in health care fraud detection and prevention as well as materially assist agencies responsible for paying health care costs and law enforcement agencies that confront this important and costly problem.

Acknowledgments

We are grateful to the state which provided the Medicaid dental data analyzed. We thank the subject matter experts that helped in evaluation of our results. Finally, our acknowledgement goes to the reviewers of our paper.

References

Aggarwal, A., 2013. Outlier Analysis. Springer, Heidelberg.
Aral, K.D., Güvenir, H.A., Sabuncuoğlu, I., Akar, A.R., 2012. A prescription fraud detection model. Comput. Methods Prog. Biomed. 106 (1), 37–46 (April).
Benbasat, I., Goldstein, D.K., Mead, M., 1987. The case research strategy in studies of information systems. MIS Q. 11 (3), 369–386.
Bolton, R.J., David, J.H., 2002. Statistical fraud detection: a review. Stat. Sci. 17 (3), 235–249.
Centers for Medicaid and Medicare Services, 2014. Report to congress fraud prevention system second implementation year. (Accessed: 2015-12-14. URL https://www.stopmedicarefraud.gov/fraud-rtc06242014.pdf).
Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection: a survey. ACM Comput. Surv. 41 (3), 15:1–15:58 (Jul.).
Dilla, W.N., Raschke, R.L., 2015. Data visualization for fraud detection: practice implications and a call for future research. Int. J. Account. Inf. Syst. 16, 1–22 (URL http://www.sciencedirect.com/science/article/pii/S1467089515000020).
Forgionne, G.A., Gangopadhyay, A., Adya, M., 2000. An Intelligent Data Mining System to Detect Healthcare Fraud. IGI Global, Hershey PA, p. 148169 (Ch. Chapter VII).
Henderson, J.W., 2009. Health Economics and Policy. South-Western Cengage Learning, Mason, OH.
Hillerman, T.P., Carvalho, R.N., Reis, A.B., 2015. Analyzing suspicious medical visit claims from individual healthcare service providers using k-means clustering. Electronic Government and the Information Systems PerspectiveVol. 9265 of Lecture Notes in Computer Science. Springer International Publishing, pp. 191–205.
Iyengar, V., Hermiz, K., Natarajan, R., 2013. Computer-aided auditing of prescription drug claims. Health Care Manag. Sci. 1–12.
Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., Arab, M., 2015. Using data mining to detect health care fraud and abuse: a review of literature. Glob. J. Health Sci. 7 (1), 194–202.
Kelley, R., 2013. Where can $700 billion dollar in waste be cut annually from the U.S. healthcare system? (Accessed: 2013–04-18. URL http://www.larson.house.gov/images/pdf/700billioninwaste.pdf)
Li, J., Huang, K.-Y., Jin, J., Shi, J., 2008. A survey on statistical methods for health care fraud detection. Health Care Manag. Sci. 11 (3), 275–287.
Lu, F., Boritz, J.E., 2005. Detecting fraud in health insurance data: learning to model incomplete Benfords law distributions. In: Gama, J., Camacho, R., Brazdil, P., Jorge, A., Torgo, L. (Eds.), Machine Learning: ECML 2005Vol. 3720 of Lecture Notes in Computer Science. Springer, Berlin Heidelberg, pp. 633–640.
Major, J.A., Riedinger, D.R., 2002. EFD: a hybrid knowledge/statistical-based system for the detection of fraud. J. Risk Insur. 69 (3), 309–324.
Musal, R.M., 2010. Two models to investigate Medicare fraud within unsupervised databases. Expert Syst. Appl. 37 (12), 8628–8633 (Dec.).
NAMFCU, 2013. Medicaid fraud reports. (Accessed: 2013–04-15. URL http://www.namfcu.net/resources/medicaid-fraud-reports-newsletters/).
Ng, K.S., Shan, Y., Murray, D.W., Sutinen, A., Schwarz, B., Jeacocke, D., Farrugia, J., 2010. Detecting non-compliant consumers in spatio-temporal health data: a case study from Medicare Australia. In: Fan, W., Hsu, W., Webb, G.I., Liu, B., Zhang, C., Gunopulos, D., Wu, X. (Eds.), Data Mining Workshops (ICDMW), 2010 IEEE International Conference on. IEEE Computer Society, pp. 613–622.
Organisation for Economic Co-operation and Development, 2012. OECD health data October 2012. (Accessed: 2013-3-4. URL http://www.oecd.org/els/health-systems/OECDHealthData2012FrequentlyRequestedData_Updated_October.xls).
Ortega, P.A., Figueroa, C.J., Ruz, G.A., 2006. A Medical Claim Fraud/Abuse Detection System Based on Data Mining: A Case Study in Chile. In: Crone, S.F., Lessmann, S., Stahlbock, R. (Eds.), DMIN. CSREA Press, pp. 224–231.
Phua, C., Lee, V., Smith-Miles, K., Gayler, R., 2010. A comprehensive survey of data mining-based fraud detection research. arXiv Preprint arXiv:1009.6119 1, pp. 50–53.
Rousseeuw, P.J., van Zomeren, B.C., 1990. Unmasking multivariate outliers and leverage points. J. Am. Stat. Assoc. 85 (411), 633–639.
Shan, Y., Jeacocke, D., Murray, D.W., Sutinen, A., 2008. Mining medical specialist billing patterns for health service management. In: Roddick, J.F., Li, J., Christen, P., Kennedy, P.J. (Eds.), AusDM'08 Proceedings of the 7th Australasian Data Mining ConferenceVol. 87 of CRPIT. Australian Computer Society, pp. 105–110.

Shin, H., Park, H., Lee, J., Jhee, W.C., 2012. A scoring model to detect abusive billing patterns in health insurance claims. Expert Syst. Appl. 39 (8), 7441–7450 (June).
Sparrow, M.K., 2000. License to Steal: How Fraud Bleeds America's Health Care System. Westview press.
Tang, M., Mendis, B.S.U., Murray, D.W., Hu, Y., Sutinen, A., 2011. Unsupervised fraud detection in Medicare Australia. Proceedings of the Ninth Australasian Data Mining Conference — Volume 121AusDM '11. Australian Computer Society, Inc., Darlinghurst, Australia, pp. 103–110.
The R Foundation, 2015. The R project for statistical computing. (Accessed: 2015-04-21. URL http://www.r-project.org/).
Thornton, D., Müller, R.M., Schoutsen, P., van Hillegersberg, J., 2013. Prediction healthcare fraud in medicaid: a multidimensional data model and analysis technique for fraud detection. Procedia Technol. 9, 1252–1264 (URL http://www.sciencedirect.com/science/article/pii/S2212017313002946).
Thornton, D., van Capelleveen, G., Poel, M., van Hillegersberg, J., Mueller, R.M., 2014. Outlier-based health insurance fraud detection for u.s. medicaid data. Proceedings of the 16th International Conference on Enterprise Information Systems, pp. 684–694.
Travaille, P., Müller, R.M., Thornton, D., van Hillegersberg, J., 2011. Electronic fraud detection in the U.S. medicaid healthcare program: lessons learned from other industries. 17th Americas Conference on Information Systems, AMCIS 2011 (URL http://doc.utwente.nl/78000/).
U.S. Attorneys Office, District of New Jersey, 2013. South Jersey doctor admits making half-a-million dollars in fraud scheme involving home health care for elderly patients. (Accessed: 2013-03-28. URL http://www.fbi.gov/newark/press-releases/2013/south-jersey-doctor-admits-making-half-a-million-dollarsin-fraud-scheme-involving-home-health-care-for-elderly-patients).
U.S. Federal Bureau of Investigation, 2013. FBI News Blog. (Accessed: 2013–04-18. URL http://www.fbi.gov/news/news_blog).
U.S. Government Accountability Office, 2012. Medicare Fraud Prevention: CMS Has Implemented a Predictive Analytics System, But Needs to Define Measures to Determine Its Effectiveness.
U.S. Office Inspector General, 2015. Questionable billing for Medicaid pediatric dental services in California. (URL http://oig.hhs.gov/oei/reports/oei-02-14-00480.pdf).
U.S. Department of Health and Human Services, 2011. Advancing the health, safety, and well-being of our people: HHS budget prospective. (Accessed: 2013–04-18. URL http://www.hhs.gov/budget/budget-brief-fy2013.pdf).
U.S. Department of Health and Human Services, 2013. 2013 actuarial report on the financial outlook for Medicaid. (URL http://medicaid.gov/medicaid-chip-program-information/by-topics/financing-and-reimbursement/downloads/medicaid-actuarial-report-2013.pdf).
U.S. HHS and DOJ, 2014. Health care fraud and abuse control program annual report for fiscal year 2013. (Accessed: 2015–12-14. URL http://oig.hhs.gov/publications/docs/hcfac/FY2013-hcfac.pdf).
Yamanishi, K., Ichi, T.,,J., Williams, G., Milne, P., 2004. On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. Data Min. Knowl. Disc. 8, 275–300.
Yang, W., Hwang, S., 2006. A process-mining framework for the detection of healthcare fraud and abuse. Expert Syst. Appl. 31 (1), 56–68.
Yin, R.K., 2008. Case Study Research: Design and Methods: Design and Methods. 5. Sage Publications.