International Statistical Review

International Statistical Review (2018), 86, 3, 379–402 doi:10.1111/insr.12269

Statistical Medical Fraud Assessment: Exposition to an Emerging Field

Tahir Ekin[1], Francesca Ieva[2], Fabrizio Ruggeri[3] and Refik Soyer[4]

[1]*McCoy College of Business, Texas State University, San Marcos, TX 78666, USA*

[2]*Department of Mathematics, Politecnico di Milano, Milan 20133, Italy*

[3]*CNR-IMATI, Milan 20133, Italy*

[4]*School of Business, The George Washington University, Washington, DC 20052, USA*
E-mail: t_e18@txstate.edu

Summary

Health care expenditures constitute a significant portion of governmental budgets. The percentage of fraud, waste and abuse within that spending has increased over years. This paper introduces the emerging area of statistical medical fraud assessment, which becomes crucial to handle the increasing size and complexity of the medical programmes. An overview of fraud types and detection is followed by the description of medical claims data. The utilisation of sampling, overpayment estimation and data mining methods in medical fraud assessment are presented. Recent unsupervised methods are illustrated with real world data. Finally, the paper introduces potential future research areas such as integrated decision making approaches and Bayesian methods and concludes with an overall discussion. The main goal of this exposition is to increase awareness about this important area among a broader audience of statisticians.

Key words: Fraud; medical audits; fraud detection; sampling; data mining; fraud analytics; decision analysis.

1 Introduction

Fraud has been around since the early days of commerce, continuously evolving and adapting to changing times. The current fraud instances can be seen in a wide range of domains such as money laundering, e-commerce and insurance. Among these, medical fraud has recently attracted more attention because of the increases in healthcare spending and overpayments especially in the developed countries. For instance, total health care expenditures in the United States reached $3.2 trillion which corresponds to $9,990 per person in 2015 CMS (2017b). US federal agencies estimate that 3% to 12.7% of this spending is lost to fraud, waste and abuse (Shin *et al.*, 2012; CMS, 2015a). The European Healthcare Fraud & Corruption Network reviewed fraud studies between 1997 and 2013 and found the range of percentage losses to be between 0.6% and 15.4% with average losses of 6.19% (Gee & Button, 2015). They report the increasing trend of the average losses and estimate the total annual global loss as $455 billion. In addition to the direct cost implications, overpayments also impact the effectiveness of the health care systems by preventing the delivery of efficient services to deserving patients (Anderson & Hussey, 2001). For instance, the negative impact of medical overpayments on the system efficiency is illustrated with Italian hospitals' data (Berta *et al.*, 2010). These medical overpayments can be in the form of fraud, waste and abuse. A variety of statistical methods are utilised in order to detect the suspicious activities prior to the audits. Then, medical auditors investigate both intentional and unintentional wrongdoings to reveal the overpayments. This paper serves as an expository venue to introduce this emerging crucial area to more statisticians, to discuss current statistical applications in medical fraud assessment and to provide directions for future research.

1.1 Fraud

The Merriam–Webster dictionary defines fraud as 'intentional perversion of truth in order to induce another to part with something of value or to surrender a legal right'[1]. Fraud occurs in many industries, including but not limited to finance, credit card, telecommunications, insurance and healthcare. Delamaire *et al.* (2009) provide a review of credit card fraud types. For instance, user credit card fraud includes falsifying information in a credit card/account application or bankruptcy declaration. The more common types involve identity theft and misuse of compromised/stolen credit card information. Whereas telecommunications fraud mostly occurs either in the form of subscription fraud that involves users falsifying information or superimposed fraud where fraudsters take over legitimate accounts for fraudulent use. Fraud also frequently takes place at various insurance programmes in the phases of application, eligibility, rating and billing of claims. For instance, Viaene *et al.* (2002) discuss the automobile insurance fraud cases of incorrect claim submissions. Health care insurance programmes are not exceptions. However, defining fraudulent behavior, detecting fraudulent cases and measuring fraud losses in health care industry are difficult tasks (Sparrow, 2000). Medical fraud broadly corresponds to the acts of intentional deception or misrepresentation which would result in unauthorised benefits within the health care system, while waste and abuse differ by the level of intention and knowledge. In the following, the term 'fraud' will be used while referring to overpayments, however, it should be noted that presented methods are applicable to detection of waste and abuse as well.

Medical fraud differs from the fraud types in other industries for a number of reasons. First, health care insurance programmes are provided by a variety of specialised organisations in most countries. The size and variety of programmes make administration and overseeing medical expenditures challenging. In the USA, a federal agency called *The Centers for Medicare & Medicaid Services* (CMS) administers the Medicare programme and works with state governments to manage Medicaid and other health insurance programmes. US Medicare is federally funded and mostly serves senior citizens. It provided health insurance to 55 million Americans in 2015 (CMS, 2015b). Whereas, US Medicaid is jointly funded by the federal government and states. It consists of a collection of insurance programmes serving people of all ages that do not have sufficient resources for health care. In other countries, there are similar programmes such as Australia Medicare and England National Health Service. The payment structures generally include fee for service and prospective payment systems. In fee for service systems, providers are reimbursed for the incurred costs after the visit. Whereas, prospective payment systems are based on providers receiving pre-determined payments for the admissions with respect to the diagnosis related groups.

Secondly, the objectives of healthcare industry and the billing systems as well as the involvement of the stakeholders in healthcare are different compared with the other industries. In general, fraud control and processing accuracy checks can slow down the claims processes and hinder the system productivity (Sparrow, 1998). However, in medical claim processing, providing timely payments to providers is crucial. Governmental procedures mostly require providers to be paid for legitimate claims in 30 days (CMS, 2005). Especially in the high volume systems, this may result in lack of thorough evaluation of payments at the expense of processing accuracy and fraud control. In addition, the stakeholders may lack incentives to report health care fraud and share feedback with the administrators. For example, if a customer's credit card is compromised; that customer would be exposed to the direct financial impact through the next monthly statement. However, in health care industry; statements are generally more complex. Even if the beneficiaries notice a fraudulent transaction, the incentives for reporting are not adequate because the insurance programme mostly pays the bill and the fraud reporting process may be found as tedious.

Medical fraud is generally classified into three categories based on who conducts the fraud; provider (such as hospitals and physicians) fraud, consumer (patient) fraud and insurer fraud. US law identifies the submission of false claims, the payment or receipt of kickbacks and self-referrals as provider fraud Kalb (1999). In addition, upcoding (overcharging for a more expensive service) and charging separately for procedures that are initially part of one procedure are examples of provider fraudulent activities. Consumer fraud includes patients who intentionally falsify documents or misuse their insurance cards to obtain prescription. Insurer fraud corresponds to the cases of insurers falsifying statements or not providing the insurance they have collected premiums for. Among these, provider fraud accounts for the largest proportion of the total medical fraud, and therefore, majority of research efforts have focused on that type of fraud (Li *et al.*, 2008). Sparrow (2000) presents a detailed discussion of real world cases of medical fraud. Asymmetry of information between providers and patients, inelastic demand for services and the presence of third party fees for service payments are the main drivers of the medical fraud. Studies of medical fraud are becoming more crucial because of the drastic monetary losses and the impact on the well-being of all citizens. Next, we present an overview of fraud assessment methods.

1.2 Fraud Assessment

The survey of Bolton & Hand (2002) lists major approaches used for fraud detection as artificial intelligence, distributed and parallel computing, econometrics, expert systems, fuzzy logic, genetic algorithms, machine learning, neural networks, pattern recognition and visualisation. The methods of practice are not generally discussed in detail, and the data sets generally are not shared with public for privacy concerns and preventing the criminals getting valuable information.

One of the widely used methods is based on comparing the transactions with the benchmark of expected occurrences and flagging the unexpected observations for further investigations. These types of outlier detection methods also help reveal emerging fraud schemes. For instance, in telecommunications networks, such benchmark-based rules are generally employed to detect unusual activities.

Another major set of methods involves computing risk (suspicion) scores for each transaction based on the outputs of classification or regression models. These are more successful in domains where the auditors have information about the fraud schemes. In particular, financial industry has made remarkable progress in tackling fraud with supervised methods. The review of Ngai *et al.* (2011) concentrates on the application of data mining methods for financial fraud detection. One of the major advantages for financial fraud methods is the abundance of labelled data. Credit card databases contain information about each transaction; including merchant code, account number, type of purchase, client name and size and date of transaction.

In addition to having access to this information real time, banks have a vast number of labelled fraudulent and legitimate past transactions. This enables the use of supervised methods such as logistic models, neural networks, Bayesian belief networks and decision trees for classifying financial transactions. One should still be aware of the unbalanced class sizes because fraudulent transactions are relatively scarce.

Despite the wide adoption of fraud detection methods in these domains, the level of attention given to medical fraud assessment has been relatively limited (Phua *et al.*, 2010). Some of the aforementioned methods are applicable for detection of fraudulent medical claims. Travaille *et al.* (2011) discuss the applicability of the fraud detection methods in other industries, such as credit card, telecommunications and computer security, to US Medicaid healthcare programmes. For instance, medical insurance claims can be argued to be similar to credit card transactions; in terms of recording the information of the provider and beneficiary and the details of the transaction.

The complexity and heterogeneity of the medical systems and data make the widespread use of fraud detection methods challenging in healthcare. The nature of medical data is different compared to other industries. First, the data in general are not reported in real time and accurately. Labelling data requires an audit, which makes its retrieval costly and time-consuming. In addition, the confidentiality of the data is crucial, and patient privacy is preserved by a number of federal acts such as Health Insurance Portability and Accountability Act. Meanwhile, fraud patterns are dynamic and adapt to changes in legislation, billing and medical procedures over time. This dynamic nature coupled with the lack of labelled data prevents supervised methods to be used as often as other industries' fraud detection frameworks (Furlan & Bajec, 2008). On the other hand, outlier detection methods are more difficult to employ because of heterogeneity of the data, pressure of timely processing of claims and the need for medical expertise in evaluation. For instance, auto insurance industry has tight controls and can stop the payments to providers that have different billing characteristics. However, in healthcare, timely payments are important, and the unusual provider behaviour may be the result of medical necessity.

Overall, increasing budget deficits in developed countries has garnered the attention of public to health care expenditures. This has increased efforts to decrease the health care spending by cutting off the unnecessary payments via medical audits and fraud assessment. Pre-payment reviews are conducted more widely especially through identity checks that can filter fraudulent transactions (Suleiman *et al.*, 2014). Despite that, most of the medical audits are conducted after payment. The goal is to identify and recover improper payments through efficient detection procedures (CMS, 2016b). Each medical investigation requires the subject domain expertise of licensed professionals that manually audit claims. The investigation costs correspond to the time spent by the expert and the physical resources. The unnecessary audit of claims that have zero overpayment (false positives) results in loss of trust to the government and lost opportunity cost. Whereas the heterogeneous nature of medical claims data and existence of a number of fraud patterns can increase the estimation error. The trade-off between the audit costs and accuracy of the extrapolations is one of the main challenges in subsequent resource allocation decisions. Identification of the fraudulent activities should ideally be done by domain experts via comprehensive medical audits. However, that is generally impractical because of the complex nature and the big size of the medical claims data. Tools of descriptive data analysis can help detect obvious patterns and reveal the potential cases of overpayment. This is mainly used for cases where overpayment can easily be identified due to irrefutable evidence such as providers billing for beneficiaries for whom it is impossible to serve. These challenges make the systematic use of statistical approaches such as sampling and data mining necessary in medical fraud assessment.

1.3 Related Literature and Outline

In the last decade, there have been a number of attempts to provide overviews of different aspects of the emerging field of statistical medical fraud assessment. Li *et al.* (2008) provide a notable detailed review of the application of data mining methods until 2007. However, it should be noted that recent data sharing and transparency efforts of governmental health organisations give researchers more access to the medical claims data and resulted in more recent publications. This has resulted in a number of overviews in particular aspects of medical fraud assessment. Capelleveen (2013) presents an overview about the use of outlier detection methods in medical fraud assessment. Whereas, Joudaki *et al.* (2014) provide a brief review of data mining methods in healthcare fraud detection. Bauder *et al.* (2017) present a comprehensive survey on the state of healthcare upcoding fraud analysis and detection with an emphasis on data mining. All these surveys focus on application of data mining methods, and none of them discusses the statistical methods such as sampling and overpayment estimation.

This paper aims to provide a comprehensive up-to-date overview of statistical methods in medical fraud assessment; including sampling, overpayment estimation and data mining approaches. In addition, we present an illustration of recently proposed unsupervised methods using publicly available real world medical claims data. Such unsupervised methods are crucial because of limited availability of labelled data in the medical fraud context. Potential future research areas are also presented with a discussion of recent advances.

The rest of this paper is structured as follows. Next section describes the medical claims data. In Section 3, we present an overview of the use of sampling and overpayment estimation methods. Section 4 focuses on data mining applications, whereas Section 5 illustrates the use of recent methods using real world medical claims data. Section 6 presents directions for future research, and Section 7 concludes by discussing the respective challenges.

2 Medical Claims Data

The particular characteristics of medical data vary from programme to programme. The literature includes examples from the health insurance programmes of various countries. These include USA (Edwards *et al.*, 2003; Travaille *et al.*, 2011; Ekin *et al.*, 2015), Australia (He *et al.*, 1997; Shan *et al.*, 2009; Tang *et al.*, 2011), Taiwan (Chan & Lan, 2001; Yang & Hwang, 2006), Chile (Ortega *et al.*, 2006), South Korea (Shin *et al.*, 2012), Turkey (Aral *et al.*, 2012) and Brazil Carvalho *et al.* (2017). This section aims to provide a high level, not exhaustive, description.

In general, the medical data can be classified as practitioners data, clinical instance data and medical claims data (Liu & Vasarhelyi, 2013). Practitioners data provide a general description of service providers in a certain time period and summarise provider related aspects of service cost, usage and quality (Viveros *et al.*, 1996). Clinical-instance data consist of a set of activities performed by medical staff in a particular treatment (Yang & Hwang, 2006). This paper mainly focuses on the medical claims data, because most raw medical data are in the form of insurance claims. A medical claim involves the participation of a patient and a service provider and generally contains the attributes of patients, providers and the claim itself. Attributes of a patient can be gender, age and medical history, whereas the type and the location of facility are among the attributes of a provider. The prescription details, monetary amount and the paid amount are among the important characteristics of the claim.

Fraud data have many unique characteristics. First, fraud is a rare event because the legitimate claims almost always outnumber the fraudulent ones. For instance, more than 80% of the papers reviewed in Phua *et al.* (2010) have skewed data with less than 30% fraud. The sparsity of the fraud data can be addressed with methods such as non-negative matrix factorisation, singular value decomposition and principal component analysis (Zhu *et al.*, 2011). Secondly, both legitimate and fraudulent claims have dynamic patterns due to heavy competition in health care industry and updates in the policy and legal frameworks. In addition, because of the multiple styles of fraud happening around the same time, the fraudulent cases are not homogeneous (Fawcett, 2003).

In most industry practices, data pre-processing efforts, such as data cleaning and transformation, take most of the time of overall fraud detection procedure (see Lin & Haug, 2006, and Sokol *et al.*, 2001, for relevant discussions). For instance, being able to submit the same claim under the name of the hospital or provider may require the investigators to define new unique identifiers to analyse the medical data (Musal, 2010). Another crucial medical data issue is the abundance of missing values. Missing data can produce problems such as over/under sampling, non-representativeness and potential bias in inference. Despite that, many papers in literature do not explicitly discuss how they handle missing data before their data analysis. One of the widely used approaches is to remove the claim lines with missing information. For instance, Ortega *et al.* (2006) discard 35% of these medical claims of a given year because of poor quality in terms of missing values and low contribution. Yang & Hwang (2006) filter out noisy data by removing instances that have missing attribute values. Removing the instances can decrease the statistical power of an analysis, because potentially valuable information in the other fields is lost. In addition, the pattern of missing values may be systematic, and deleting records may create a biased subset. However, there are not any systematic guidelines to handle missing data. The benefits and drawbacks of handling missing values should be carefully evaluated by the domain experts. Grzymala-Busse & Hu (2000) present a comparison of a number of approaches. Such methods include replacing missing values with the mode or the mean of the data set, with random values sampled from the underlying distribution or treating missing values as user-defined constants. Imputation, substitution of a missing value using an estimate retrieved by a statistical analysis such as regression, can be proposed as another way of dealing with missing values (Li *et al.*, 2008). Little & Rubin (2014) provide a detailed discussion of such imputation techniques.

Choosing and transforming the attributes (features) is also a crucial step in data analysis. Attributes used in fraud studies can be numerical, categorical or binary type of variables. Different types of variables would require the use of different statistical techniques. The selected features are rarely revealed in publications due to agreements with the data sources. The motivation of such confidentiality agreements is to prevent the criminals from having access to the way how detection systems work. In practice, features are generally selected by medical domain experts. Domain experts are knowledgeable about frequent fraud occurrences or the fraud types with the most financial losses. Major & Riedinger (2002) list the relevant attributes of providers for fraud detection in five main categories; financial, medical logic, abuse, logistics and identification. Whereas money amount involved and the paid portion are deemed as important attributes of a claim from the investigation perspective. Manual feature selection can benefit from statistical checking of each selected feature's relevance and significance (see Dash & Liu, 1997, for an overview of feature selection methods). For instance, checking the relationship between attributes is generally overlooked in industry practices while selecting attributes for outlier detection methods. The study of Ortega *et al.* (2006) is one of the rare studies that utilises correlation checks to delete redundant features and to test discriminating power of each feature.

It is prohibitively time-consuming and costly to analyse all (or most) observations. Therefore, a number of statistical sampling and overpayment estimation methods are proposed. On the other hand, the heterogenous structure of data and the nature of fraud result in the widespread use of the data mining methods. Next two sections present overviews of sampling and data mining methods, respectively.

3 Sampling and Overpayment Estimation

Sampling and overpayment estimation methods help the medical auditors to retrieve the sample data and make extrapolations. Various techniques such as simple random sampling and stratified sampling are recommended to draw representative samples from the population of interest as efficiently as possible. In the USA, use of probability sampling methods for medical investigations has been accepted to be part of the legal framework since 1986. Yancey (2012) provides a comprehensive list about these legal sampling procedures and the parties involved in US governmental medical insurance programmes. In general, the variables of interest are the payment amounts to providers, the percentage of overpaid claims and the overpayment amount. There are governmental software packages which assist the auditors with sampling and analysis. For instance, in the USA, medical auditors can use RAT-STATS (OIG, 2010) that is offered by the Office of Inspector General, Office of Audit Services. RAT-STATS can perform functions such as determination of sample size, generating random numbers to select the sample and provide inference. Woodard (2015) presents a brief overview and a simple application to demonstrate the use of sampling by the US Medicaid to uncover and reclaim overpayments.

A payment amount associated with a claim can result in one of three outcomes when audited. A claim can be classified as completely legitimate, completely illegitimate or partially overpaid. A claims data set where each claim is either a legitimate payment or a completely illegitimate payment is referred to as 'all or nothing'. According to the current US sampling guidelines (CMS, 2001), in most situations, the lower limit of a one-sided 90% confidence interval for the total over payments should be used as the recovery amount from the provider under investigation. Using the lower bound allows for a reasonable and fair recovery without requiring the tight precision to support the point estimate, sample mean. In other words, the state is protected with a certain degree of confidence from recovering an amount greater than the true value of erroneous payments. However, this application of central limit theorem (CLT) is based on the assumption that overpayment population either follows the normal distribution or that the sample size of overpayments is reasonably large. Mohr (2005) shows that normality based models can work well for cases with one overpayment pattern.

It is common that medical claims data exhibit skewness and non-normal behavior requiring large sample sizes for the valid application of CLT. Edwards *et al.* (2003) show that methods based on the CLT may not perform well for certain kinds of overpayment populations with small sample sizes. They propose the 'minimum sum method', a non-parametric inferential method which makes use of the hyper-geometric distribution and computes the respective lower bound estimates. These estimates are shown to be efficient in settings where the claims are essentially 'all or nothing', the payment population is relatively homogeneous and well separated from zero. A number of extensions are proposed for the minimum sum method. Ignatova & Edwards (2008) propose a sequential sampling framework that aims to make inference on the proportion of claims with overpayments. Gilliland & Feng (2010) provide an adaptation in order to address cases of varying payments. Gilliland & Edwards (2010) improve its efficiency via randomised lower bounds in which payment amounts are audited in equal sized packets. Edwards *et al.* (2003) discuss a simple extension, so called q-adaptation minimum sum method, which is based on redefinition of illegitimate payments; so that the payments are defined as illegitimate if 'q' percent of the payment is in error. In order to deal with partial overpayments and 'all or nothing' claims, Ekin *et al.* (2015) propose a zero–one inflated mixture model that extends Mohr (2005). Musal & Ekin (2017) present a Bayesian mixture model that can be more efficient for claims with partial overpayments. In addition to these, standard stratified expansion and combined ratio estimators of the total are among proposed estimators.

4 Data Mining Methods

Capabilities of generating, collecting and storing medical data have increased dramatically in the last two decades. In particular, medical databases increase in size with respect to both instances and variables. For instance, since 2006, US Medicare claims data are stored in continuously expanding Integrated Data Repository. The Integrated Data Repository can also be accessed through the web-based One Program Integrity Portal, which provides investigators a comprehensive view of data. This explosive growth in stored or transient data has generated an urgent need for methods that can intelligently transform the vast amounts of medical data into useful information and knowledge. Data mining, a step in the process of Knowledge Discovery in Databases, is a method of unearthing information from large data sets. Knowledge Discovery in Databases within the medical context can be used to map low level data into other forms that might be more compact and useful and reduces the high dimensionality. Built upon statistical analysis, it can analyse massive amounts of data and provide useful and interesting information about patterns and relationships that exist within the data that might otherwise be missed.

These developments urge the federal government to fund CMS for the objective of exploring the use of analytical methods for medical fraud detection. CMS has used their fraud prevention system to identify and prevent more than $1.5 billion USD in healthcare fraud, waste and abuse within the Medicare fee-for-service programme (Belliveau, 2016). Fraud prevention system consists of rule-based, predictive, anomaly-based and network-based methods. These analytical algorithms help CMS to reveal improper payments during both pre-payment and post-payment 'pay and chase' audits.

Commonly used medical fraud detection methods can be classified as supervised, unsupervised or hybrid, depending on the availability of labelled data (Li *et al.*, 2008). In medical fraud assessment context, labelled data mostly correspond to claims that are found to be fraudulent after investigation by domain experts and enable the use of supervised methods. In the absence of that information, unsupervised methods are proposed, mostly in order to detect potential deviations from the expected patterns (see the review of Capelleveen, 2013). Hybrid approaches are mostly based on using both unsupervised and supervised methods or ensembles of supervised methods for enhanced performance. The following subsections provide a literature overview of these supervised, unsupervised and hybrid data mining methods.

4.1 Supervised Algorithms

Supervised methods are based on using labelled fraudulent and non-fraudulent records in order to classify claims and make predictions. In terms of classification algorithms, Li *et al.* (2008) emphasise the extensive use of neural networks and decision trees for fraud detection. Liou *et al.* (2008) provide a comparison of decision trees, neural networks and logistic regression with respect to their correct identification rate of medical fraud.

Neural networks can handle complex, large data sets and non-linear variable relationships. However, application of neural networks generally requires statistical expertise, for instance, to tune the parameters. In addition, Padmaja *et al.* (2007) point out that classification may display poor performance and overfitting with skewed data sets. In order to avoid overfitting, Ortega *et al.* (2006) implement an early stopping technique in their neural network based medical fraud detection method. It is based on using one training data set to update the weights and biases and another data set to stop training when the network begins to overfit the data. They also address the issue of having a large prediction variance because of a small sample size with a large number of features.

In comparison, decision trees have generic rules which are easy to interpret especially with small number of categories. Decision trees can also handle sparse data but may result in overfitting and decreased interpretability of results with increasing size of data. For instance, Shin *et al.* (2012) propose a scoring model for likelihood of abuse and then classify providers using a decision tree. Ormerod *et al.* (2003) present a dynamic Bayesian network of fraud indicators, whose weights are determined by the fraud prediction power of each feature. Bayesian classifiers have shorter training times and are found to be effective in handling the sparsity of the data. He *et al.* (1998) use k-nearest neighbour algorithm to classify practitioners' practice profiles. As an alternative, a support vector machine-based approach is utilised by Kumar *et al.* (2010). A number of the medical detection efforts include combining different supervised methods. Chan & Lan (2001) combine fuzzy sets theory and a Bayesian classifier to detect suspicious claims in Taiwan National Health Insurance. Viveros *et al.* (1996) recommend a combination of association rules and a neural segmentation algorithm for fraud detection.

Overall, supervised methods are useful for detecting previously known patterns of fraud. Because they are based on classified past claims, one should be aware of potential overestimation issues (Liou *et al.*, 2008). The existence of unbalanced class sizes within the claims data can also lead to overfitting. These models should be regularly updated to deal with new fraud patterns and changes in the regulations. Inability of supervised methods to detect dynamic and adaptive fraud has increased the attention on unsupervised methods which will be discussed next.

4.2 Unsupervised Algorithms

Unsupervised methods are motivated by the unavailability of labelled medical data and deficiencies of supervised methods. They are mainly used to group the claims and detect the claims with potential deviations from the frequent patterns. Because they do not require pre-labelled data, unsupervised methods may serve as initial filters that list the potentially fraudulent claims before the actual audit. This can decrease personnel costs as less transactions are reviewed (Laleh & Azgomi, 2009). Another advantage of unsupervised methods is their independence from a particular classified data set, therefore, they can be used to detect changing fraud patterns. Even basic unsupervised approaches may prove to be beneficial when combined with the expertise regarding discriminating features (Copeland *et al.*, 2012). Even with the need of additional assessment of subject matter experts, unsupervised learning is still deemed as a valuable and promising tool given the nature of unlabelled medical data (Bauder *et al.*, 2017).

Clustering is first applied on medical data by Lin *et al.* (2008) to segment the practice patterns of general practitioners. Then, Musal (2010) and Liu & Vasarhelyi (2013) use geo-location data within a clustering-based approach. The Bayesian Bernoulli co-clustering algorithm of Ekin *et al.* (2013) models dyadic data focusing on the occurrence of visits among providers and beneficiaries. This can potentially reveal an emerging type of fraud called 'conspiracy fraud' that involves attributes of more than one party of the medical system. These clustering algorithms help the investigator group the billings and variable of interest.

Outlier detection methods are widely used in medical fraud detection. Outliers correspond to observations that lie outside the main grouping of the data and the unexpected observations. A simple way to detect outliers would be to rank observations with respect to variables of interest and designate outcomes that are lower or higher than a pre-determined threshold as outliers. Such threshold can be determined using the knowledge of the entire data set or as a certain deviation measure (standard deviation) away from a central measure (mean, median). Capelleveen (2013) provides an overview of outlier detection methods along with a number of experiments to assess their effectiveness. These outlier analysis methods include linear models, boxplots, peak analysis, multivariate clustering and expert evaluation. Shan *et al.* (2009) propose a local density based outlier detection method to identify inappropriate billing patterns in Australia Medicare. Ng *et al.* (2010) model Australia Medicare spatio-temporal data within an unsupervised anomaly detection framework. Lu & Boritz (2005) utilise Benford's law distributions to detect anomalies in claim reimbursements. Tang *et al.* (2011) present an integrated approach that combines feature selection, clustering, pattern recognition and outlier detection to detect fraud in Australia Medicare. Carvalho *et al.* (2017) propose a two-phase anomaly detection method to identify fraudulent hospitals in Brazilian public healthcare system. There are also outlier detection studies with prescription data. Aral *et al.* (2012) propose a distance based unsupervised algorithm to assess the fraudulent risk of prescriptions. Iyengar *et al.* (2014) develop a normalised baseline behavioural model to identify the anomalies for each prescription area. van Capelleveen *et al.* (2016) present a case study for Medicaid dental practice investigations. They discuss the application of a number of outlier detection methods using different metrics. Bauder & Khoshgoftaar (2016) present a Bayesian inference-based outlier detection model which uses probability distributions and credibility intervals to assess outliers. The multistage methodology of Johnson & Nagarur (2016) is also based on quantifying risk and distance from the thresholds. Lastly, Ekin *et al.* (2017b) present the use of concentration function as a pre-screening outlier detection tool to aid in medical fraud assessment.

In addition, industry tools based on graph analytics, association and link analysis may help the investigators to reveal relationships, links and hidden patterns of information sharing and interactions within potentially fraudulent groups of providers and patients. The number and quality of the links between businesses can be analysed using the similarities in their contact information, locations, service providers, assets and associates. Potential relationships with players that are found to be involved with fraud may provide red flags and lead for prospective investigations. These can especially be helpful to reveal organised, sophisticated and collusive networks of providers and patients.

Unsupervised approaches are generally used to flag potentially fraudulent activities before bringing the domain experts into the investigation. Therefore, a close cooperation between physicians, statisticians and people involved in decision making would be beneficial during the stages of defining and tuning the model and analysing and interpreting the results.

4.3 Hybrid Algorithms

Hybrid approaches are mostly based on using unsupervised and supervised approaches together to improve the performance of medical fraud detection. For instance, unsupervised methods can be used to choose the number of classes that are used in the classification process (He *et al.*, 1997). In their use of the k-nearest neighbour algorithm, the distance metric is optimised by a genetic algorithm in detecting different types of fraud. Williams & Huang (1997) apply clustering methods and then labelled these clusters along with a classification algorithm. Clustering is shown to overcome the deficiencies of decision trees with larger data sets and many categories.

Use of medical knowledge may improve the classification performance at the expense of the cost and the complexity of such models. In such an example, Yang & Hwang (2006) use a pattern discovery algorithm to define the normal behaviour before using an outlier detection method. Major & Riedinger (2002) utilise both rule extraction and outlier detection in order to compare provider characteristics.

5 Recent Developments

This section presents an illustration of recently proposed unsupervised data mining methods with real world CMS claims data. Supervised methods have been emphasised by a number of authors such as Li *et al.* (2008), therefore, we focus on relatively under-represented unsupervised methods. Particularly, we illustrate the use of concentration function (Ekin *et al.*, 2017b) and Bayesian co-clustering (Ekin *et al.*, 2013). We use the publicly available data set, *Provider Utilization and Payment Data Physician and Other Supplier Public Use File* of *The Centers for Medicare & Medicaid Services* (CMS, 2016a). It includes information related to providers, beneficiaries and the claim itself. The variables of focus are provider's national provider id, medical specialty of the provider, Healthcare Common Procedure Coding System code of medical procedure or service (CMS, 2017a), number of distinct Medicare beneficiary per day served by a particular provider, number of medical procedures per day that are billed by a particular provider and place of service. We provide the R scripts as part of the online supplementary material.

5.1 Concentration Function Based Fraud Detection

First, we illustrate the use of concentration function (Cifarelli & Regazzini, 1987) as a pre-screening tool to aid in medical fraud assessment. The concentration function is a mathematical tool used to graphically describe the discrepancy between two probability measures defined on the same measurable space. Here, we will use it to compare two discrete probability measures. It helps us graphically summarise and compare the overall billing patterns of providers for all the prescribed services. In addition, the likelihood ratios help quantify the potential differences of each provider compared with the average charges by the population. In particular, the likelihood ratio for each procedure billed by a provider is computed as the ratio of the billing percentage by that provider for that procedure and the average billing percentage for the population of all providers. If the likelihood ratio is very large compared with a user-specified threshold, this indicates potential overcharging which may warrant an investigation. This paper presents a basic demonstration with real world claims data (see Ekin *et al.*, 2017b, for a more detailed discussion).

We rearrange the data and focus on a subset that includes billings from the optometrists in the state of Texas who provide services in a facility. We assume that the billing behaviour of all 141 optometrists in Texas is relatively homogenous. Three optometrists who billed more than 10 unique procedures are selected for demonstration. The billing of these three providers for 17 procedures is compared with the overall billing behaviour of optometrists in Texas. Table 1 lists the billing percentages for each i-th procedure, p_i^{MD1}, p_i^{MD2} and p_i^{MD3} where $i = 1, \ldots, 17$. The average billing percentage for the population for each i-th procedure is also listed as q_i. For each optometrist, the likelihood ratios of each i-th procedure billing are computed as $LR_i = p_i/q_i$. The likelihood ratios can reveal outlier billings performed by each provider with respect to a procedure. For instance, a likelihood ratio more than 5 indicates that the percentage of charges for the related service prescribed by an MD is at least five times larger than the average charge for that service. Such likelihood ratios, LR, that are greater than 5 are highlighted in bold in Table 1.

These three providers (MDs) are assumed to bill similarly to the population of optometrists. In order to check that assumption, using the billing percentages in Table 1, we compare the probability measures of each provider, $\boldsymbol{P}^{MD1}$, $\boldsymbol{P}^{MD2}$, $\boldsymbol{P}^{MD3}$ and the overall population, $\boldsymbol{Q}$. Particularly, we order each procedure with respect to the ascending order of their likelihood ratios, LR_i. In other words, the outcomes are ordered starting from the least billed procedures (when compared to the rest of the MDs population) towards the most frequently billed. The concentration function is constructed by cumulatively adding the probabilities of these ordered billings. The plot of these ordered probabilities and respective procedure codes between the points of $(0, 0)$ and $(1, 1)$ gives the concentration function of $\boldsymbol{P}^{MD1}$, $\boldsymbol{P}^{MD2}$ and $\boldsymbol{P}^{MD3}$ versus

Table 1. *Billing percentages of each procedure for MD1, MD2, MD3 and population q and likelihood ratios of each procedure for MD1, MD2 and MD3.*

Index	Procedure	p_i^{MD1}	p_i^{MD2}	p_i^{MD3}	q_i	LR_i^{MD1}	LR_i^{MD2}	LR_i^{MD3}
$i = 1$	67820	0	0.0124	0	0.0021	0	**5.840**	0
$i = 2$	92002	0.0134	0	0	0.0065	2.062	0	0
$i = 3$	92004	0.2002	0.2831	0.0722	0.1921	1.042	1.473	0.376
$i = 4$	92012	0.0403	0	0	0.0254	1.587	0	0
$i = 5$	92014	0.2918	0.0479	0.2586	0.1427	2.045	0.336	1.812
$i = 6$	92082	0.0171	0	0	0.0038	4.5	0	0
$i = 7$	92083	0.0781	0.0328	0.0268	0.0232	3.366	1.415	1.156
$i = 8$	92133	0.0452	0.0319	0.0351	0.0287	1.575	1.113	1.222
$i = 9$	92134	0.0134	0.047	0.0254	0.0121	1.107	3.887	2.103
$i = 10$	92250	0	0.0293	0	0.0140	0	2.096	0
$i = 11$	99202	0.0134	0.0319	0.0124	0.0054	2.481	**5.915**	2.293
$i = 12$	99203	0.0269	0.0289	0.0289	0.0137	1.964	2.108	2.108
$i = 13$	99204	0.0256	0.0834	0.0928	0.0274	0.934	3.044	3.389
$i = 14$	99212	0.1636	0	0.0103	0.0224	**7.304**	0	0.461
$i = 15$	99213	0.0366	0.1047	0.1018	0.0572	0.64	1.83	1.78
$i = 16$	99214	0.0342	0.2422	0.3356	0.1660	0.206	1.457	2.019
$i = 17$	99215	0	0.0231	0	0.0092	0	2.508	0

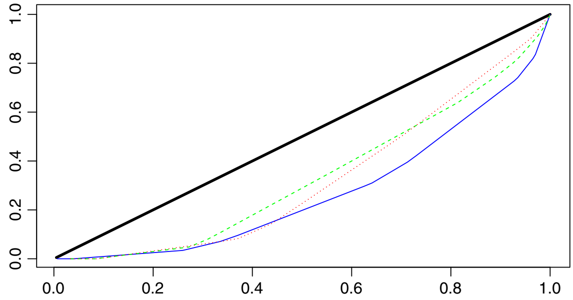

Figure 1. *Concentration function for MD 1 (blue straight line), MD 2 (green dashed) and MD 3 (red dotted) versus the population (bold* 45° *line). [Colour figure can be viewed at wileyonlinelibrary.com]*

Q. Figure 1 presents the concentration functions for all three medical doctors. It can be seen that MD1 differs more from the population because its corresponding line is further from the straight line. This indicates that the overall billing of MD1 is different from his/her peers.

Investigation of billing for specific procedures may reveal more specific outlier activities. For instance, Table 1 shows that the first medical doctor is found to charge more than 7 times than the average for the procedure code of *99212*. Whereas the second medical doctor overcharges for procedure codes of *67820* and *99202* compared to the population. Concentration function based fraud detection is simple and scalable to generate leads even in the presence of massive data. It does not require iterative computations or convergence to generate leads. However, these leads still require further investigations by medical auditors in order to assess the legitimitacy of such billings.

5.2 Bayesian Co-clustering for Fraud Detection

Next, the use of Bayesian co-clustering is demonstrated. Ekin *et al.* (2013) address the issue of conspiracy fraud by such a co-clustering model for the provider–patient pairs. In this paper, we present an illustration to describe the dyadic relationships among providers and procedure codes.

Bayesian co-clustering allows mixed membership for both K clusters of providers and L clusters of procedures; so called soft-clustering. Let us assume that we have I health-care providers who bill for J unique procedures. Let X_{ij} be binary representing if the provider i bill for procedure code j. $\mathbf{X} = \{X_{ij};\, i = 1,\ldots,I,\ j = 1,\ldots,J\}$ is a data matrix of size $I \times J$. Membership probabilities are denoted by $\pi_{1k};\, k = 1,\ldots,K$ for row clusters and by $\pi_{2l};\, l = 1,\ldots,L$ for column clusters such that

$$\sum_{k=1}^{K} \pi_{1k} = \sum_{l=1}^{L} \pi_{2l} = 1.$$

The latent variables Z_{1i} and $Z_{2j}, i = 1,\ldots,I,\ j = 1,\ldots,J$, denote membership to the row (provider) and column (procedure) clusters such that $Z_{1i} \in \{1,\ldots,K\}$ and $Z_{2j} \in \{1,\ldots,L\}$. Given $\boldsymbol{\pi}_1 = (\pi_{1k}; k = 1,\ldots,K)$ and $\boldsymbol{\pi}_2 = (\pi_{2l}; l = 1,\ldots,L)$, Z_{1i} and Z_{2j} are independent discrete random variables.

The generative function can be described as

$$(X_{ij}|Z_{1i} = k, Z_{2j} = l, \theta_{kl}) \sim Ber(\theta_{kl}), \tag{5.1}$$

where θ_{kl} denotes the probability of billing of a procedure from l-th cluster by a provider in k-th cluster. The co-clustering problem involves assignment of each X_{ij} to a co-cluster defined by the latent pair (Z_{1i}, Z_{2j}). The Bayesian model involves specification of priors for the unknown parameters $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$ and $\boldsymbol{\theta} = (\theta_{kl}; k = 1,\ldots,K,\ l = 1,\ldots,L)$. We can assume independent Dirichlet priors for $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ and independent beta priors for elements of $\boldsymbol{\theta}$. Particularly, we have

$$\begin{aligned}\boldsymbol{\pi}_1 &\sim Dir(\alpha_{1k}; k = 1,\ldots,K),\\ \boldsymbol{\pi}_2 &\sim Dir(\alpha_{2l}; l = 1,\ldots,L),\\ \theta_{kl} &\sim B(a_{kl}, b_{kl}),\ k = 1,\ldots,K,\ l = 1,\ldots,L.\end{aligned}$$

Given data matrix $\boldsymbol{X} = \{X_{ij}; i = 1,\ldots,I,\ j = 1,\ldots,J\}$, the posterior analysis can be developed by using a standard Gibbs sampler. The full conditionals for θ_{kl}'s, $k = 1,\ldots,K,\ l = 1,\ldots,L$, can be obtained as (conditionally) independent beta densities

$$\theta_{kl}|\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{X} \sim B\left(a_{kl} + \sum_{i,j} X_{ij}\mathbf{I}(Z_{1i} = k, Z_{2j} = l),\right.$$
$$\left. b_{kl} + \sum_{i,j}(1 - X_{ij})\mathbf{I}(Z_{1i} = k, Z_{2j} = l)\right),$$

where $\boldsymbol{Z}_1 = \{Z_{1i}; i = 1,\ldots,I\}$, $\boldsymbol{Z}_2 = \{Z_{2j};\, j = 1,\ldots,J\}$ and $\mathbf{I}(\bullet)$ is the indicator function. The full conditionals of $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ are (conditionally) independent Dirichlet distributions given by

$$\boldsymbol{\pi}_1|\boldsymbol{Z}_1 \sim Dir\left(\alpha_{1k} + \sum_{i,j} \boldsymbol{I}(Z_{1i} = k); k = 1,\ldots,K\right),$$

$$\boldsymbol{\pi}_2|\boldsymbol{Z}_2 \sim Dir\left(\alpha_{2l} + \sum_{i,j} \boldsymbol{I}(Z_{2j} = l); l = 1, \ldots, L\right).$$

Finally, the full conditionals of (Z_{1i}, Z_{2j}) can be obtained as

$$p(Z_{1i} = k, Z_{2j} = l|\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\theta}, X_{ij}) = \frac{\theta_{kl}^{X_{ij}} (1 - \theta_{kl})^{1-X_{ij}} \pi_{1k}\, \pi_{2l}}{\sum_{r=1}^{K} \sum_{c=1}^{L} \theta_{rc}^{X_{ij}} (1 - \theta_{rc})^{1-X_{ij}} \pi_{1r}\, \pi_{2c}}. \tag{5.2}$$

This algorithm provides a co-cluster of providers and procedures to reveal the common billing patterns. The inference carried out observing the posterior conditional distribution of $\boldsymbol{\theta}$ enables the user to describe the expected billing pattern of a given provider. Next, the discrepancies between the expected behaviour and the actual behaviour of a given provider can provide investigative leads. For a given billing; if the provider does not behave similar to his co-cluster; this may reveal a potential fraudulent behaviour.

For demonstration, we rearrange the publicly available CMS data and focus on billings from the anaesthesiologists in the state of Texas who provide services in a facility. We examine the providers that have billed for at least 10 unique procedures and the procedures that are billed by at least 20 unique providers. This results in a binary billing matrix that lists whether each of 94 procedures are billed by 376 providers.

The number of clusters are set as $K = 3$ and $L = 2$. The hyper-parameters are determined assuming uniform prior uncertainty. The algorithm is run for 20 000 iterations, and 2000 samples are used for posterior analysis within a Markov chain Monte Carlo simulation framework. The most frequent occurrences between this provider–procedure pair are found to be in co-cluster (3, 1). Let us assume we assess the billing of procedure 64941 that corresponds to facet joint injection by provider with ID 100. The posterior distributions of their memberships are shown in Figure 2. The posterior modes are $Z_{1100} = 3$ and $Z_{264\,941} = 1$. This points out to the co-cluster that has the highest association with a billing. Therefore, we can argue that this is more likely to be a legitimate billing. However, if the procedure 64941 was billed by provider i with the posterior mode $Z_{1,i} = 2$, that could have been argued to be less likely and a potential candidate for investigation.

The assumptions of this model can be relaxed straightforwardly. For instance, the numbers of clusters K and L are assumed to be known. Estimation of K and L can be considered as a model selection problem and comparison of the marginal likelihoods, $p(\boldsymbol{X}|K, L)$ for different models where each model is defined by the specific values (K, L) can help with the choice of K and L. However, in many problems, marginal likelihood is not available in an analytical form,

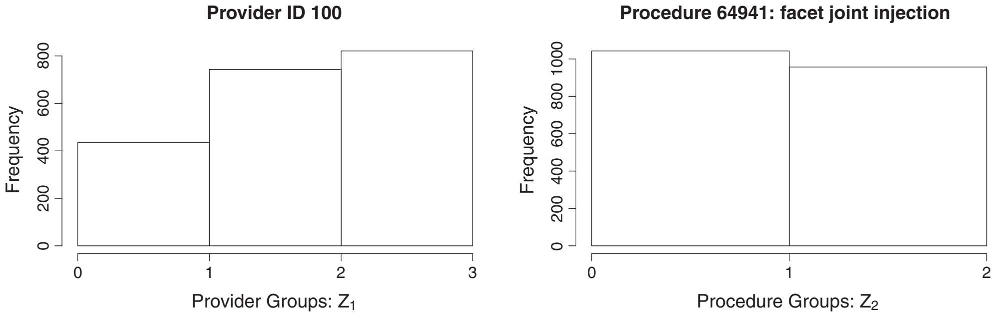

Figure 2. *Posterior distributions of the memberships for provider 100 and procedure 64941.*

and its evaluation using posterior Monte Carlo samples is not a trivial task. In cases where all full conditional distributions are known, it is possible to approximate the marginal likelihoods from the posterior samples as outlined in Chib (1995). Ekin *et al.* (2017a) illustrate such an extension for the Bayesian co-clustering model.

Another natural extension is to introduce covariates to model the probabilities $\boldsymbol{\theta}$, which can be accomplished by using a logit or probit transformation on components of $\boldsymbol{\theta}$. More specifically, let $\boldsymbol{Y}_{ij}$ denote the m-dimensional covariate vector associated with the characteristics of the provider i, procedure j and their common characteristics. Because elements of $\boldsymbol{\theta}$ are now functions of $\boldsymbol{Y}_{ij}$'s, we denote them by $\theta_{kl}(\boldsymbol{Y}_{ij}) = \theta_{kl}^{ij}$. Ekin *et al.* (2017a) discuss the probit case as follows:

$$\theta_{kl}^{ij} = \Phi(\boldsymbol{Y}_{ij}\,\boldsymbol{\beta}_{kl}),$$

where Φ is the standard normal distribution function. This probit model also enables to use a full Gibbs sampler without a Metropolis step (Albert & Chib, 1993) to draw the parameter vector, $\boldsymbol{\beta}_{kl}$.

In terms of computational requirements, it should be noted that Bayesian co-clustering generally requires data pre-processing. The medical claims data are mostly provided in the form of rows of claims that include the Healthcare Common Procedure Coding System and national provider id codes. Therefore, the user needs to check if there is a billing between each provider–procedure pair in the attempt of constructing a binary provider–procedure matrix. Finally, practical Markov chain Monte Carlo simulation convergence needs to be attained in order to draw from the distributions that are not analytically available. This can especially be computationally challenging in heterogenous and sparse data sets. As a potential remedy, methods based on variational inference can be utilised for approximation of integrals. Overall, Bayesian co-clustering may be helpful to reveal relationships between dyadic data pairs which would otherwise remain hidden. However, in the presence of big data, one should be aware of potential issues such as identifiability of parameters and computational time for convergence.

6 Directions for Future Research

The use of statistical methods such as sampling and data mining has been extensively explored in the domain of medical fraud assessment. However, the complex nature of claims data and heterogenity of medical systems would still benefit from new methods. The integration of statistical methods within decision making frameworks is one such area. We also believe Bayesian approaches have the potential to be beneficial in a number of ways. Causal models and models that consider dyadic data are also understudied compared with the fraud detection efforts in other fields. In the following, we provide a discussion of the limited existing work in addition to the needs and potential ideas in each domain.

6.1 Integrated Medical Audit Decision Making

Fraud detection and the respective audit activities are typically costly, and they can benefit from integrated approaches to utilise the limited resources efficiently. US Governmental Accountability Office (GAO, 2012) highlights the need for measures to determine the effectiveness of the medical assessment methods. However, the integration of statistical methods and actual decision making by experts is a challenging and understudied step in the overall fraud assessment procedure.

In the domain of medical fraud detection, there are a number of performance evaluation methods available despite the limited number of integrated studies. Ortega *et al.* (2006) provide a discussion about incorporation of their method to the decision flow. Their method assigns fraud probabilities to each form acting like a pre-screen filter. Then, they utilise a Receiver Operating Characteristic (ROC) curve based approach to consider cost structures such as personnel salaries and false alarms. Shin *et al.* (2012) propose a scoring and segmentation model. Capelleveen (2013) recommends the use of outlier-based methods as decision supportive technologies for resource allocation of medical audits. Cost performance, percentage of false positives and false negatives (sensitivity and specificity of the method), historical predictive power of the model and deviation from the average are recommended to be considered while incorporating the statistical outputs to the decision support system (Capelleveen, 2013). The use of clinical pathways in an attempt to standardise the medical decision making (Yang & Hwang, 2006) can be important especially for well-established domains. Other than these studies, evaluation of fraud detection algorithms and their use in decision making procedures have not been given enough consideration. As pointed out by Rashidian *et al.* (2012), there is a lack of evidence on the effectiveness of the health care fraud intervention strategies. Especially in the case of unsupervised methods, the requirement of another level of investigation and domain expertise to determine the legitimacy of the claims is not formalised. A recent relevant method is the Analytical Hierchary Process based method of Hillerman *et al.* (2017) which can be used to rank the providers and determine outliers. This can help prioritise the leads that are retrieved from the unsupervised method.

In the general fraud detection literature, a number of decision theoretic approaches are considered especially financial and auto insurance industries. For instance, fraud detection tools can be compared with respect to their expected utilities and costs. Phua *et al.* (2010) provide a discussion of performance measures for different fraud detection algorithms using examples from auto insurance fraud. In addition to the evaluation of accuracy with error-based methods, cost-based metrics such as ROC analysis are also considered in performance evaluation. ROC analysis plots the costs of different true positive (correct identification of fraud) and false positive (incorrect identification of fraud) rates. In a review of financial fraud detection methods, Ngai *et al.* (2011) point out that the cost of false negatives (misclassifying a fraudulent case as normal) can be higher than the cost of false positives. These false negatives result in opportunity costs and construction of more complex policies against fraud. Dionne *et al.* (2009) propose a scoring-based auditing approach for in auto insurance fraud cases. Decision analysis tools have also been considered in Ulvila & Gaffney (2004) for evaluating computer intrusion detection systems. The authors present an integration of ROC analysis and cost analysis to develop an expected cost metric. In so doing, they also demonstrate how decision trees can be used to combine these two tools. Another decision theoretic approach by Torgo & Lopes (2011) addresses the prioritisation of investigation leads to audit first for potential fraud given limited resources. A utility-based fraud detection model is proposed, providing rankings ordered by decreasing expected outcome of inspecting the potentially fraudulent cases. Their outcome is affected by the likelihood of fraud, inspection costs and expected payoff.

These decision-theoretic approaches or their extensions are applicable to medical fraud assessment. Even a simple decision analysis setup can be useful for formal evaluation of tools and utilisation of their output. For instance, a typical fraud detection algorithm may provide the probability of fraud or risk score of a claim. Let $P(D_F)$ denote the probability of the event D_F that the detection tool predicts fraud. Its complement $\overline{D}_F$ denotes the event that the tool predicts no fraud and has probability $(1 - P(D_F))$. Prior to the prediction by the detection tool, the decision maker has prior probability $P(F)$ for the case being fraudulent, F. The decision maker's probabilities of the case being fraudulent (F) or not $(\overline{F})$ are conditional on the detection tool's prediction. Thus, once prediction is provided by the tool, this probability can be revised accordingly to the posterior probabilities via the Bayes' rule

$$P(F|D_F) = \frac{P(D_F|F)P(F)}{P(D_F)}.$$

In the aforementioned discussion, $P(D_F)$ can be simply written as

$$P(D_F) = P(D_F|F)P(F) + P(D_F|\overline{F})P(\overline{F}),$$

where $P(D_F|F)$ is referred to as the *sensitivity* of the tool, and $P(D_F|\overline{F})$ is referred to as the probability of *false positive*. Similarly, in evaluating $P(\overline{D}_F)$, we have

$$P(\overline{D}_F) = P(\overline{D}_F|F)P(F) + P(\overline{D}_F|\overline{F})P(\overline{F}),$$

where $P(\overline{D}_F|\overline{F})$ is called the *specificity* of the tool, and $P(\overline{D}_F|F)$ is the probability of *false negative*. These error probabilities $P(D_F|\overline{F})$ and $P(\overline{D}_F|F)$ are also used in ROC analysis.

Given the prediction of the detection tool, the action chosen by the decision maker and outcome of the case collectively define the consequence. Let us denote the actions of audit and no audit by A and $\overline{A}$, respectively. For instance, if the detection tool prediction is fraud and the decision maker chooses to audit, then the consequence depends on whether the case is fraudulent or not. We can denote these consequences with the utility terms $u(D_F, A, F)$ and $u(D_F, A, \overline{F})$. The first term $u(D_F, A, F)$ reflects the benefits of correct prediction by the tool, correct action by the decision maker and the cost of audit, whereas $u(D_F, A, \overline{F})$ reflects the costs of audit and the false positive.

Such Bayesian updating and simple decision analysis tools can be used to compare the performance of fraud detection tools. It has been successfully implemented in supporting management decisions in healthcare organisations, in evaluation of healthcare providers and in helping physicians in identifying effective treatments (see the examples in Spiegelhalter *et al.*, 2004, and Faltin *et al.*, 2012). Bayesian decision analysis can help to incorporate subjective knowledge especially for borderline cases that require expertise (Berger, 2013). However, it should be noted that the investigators should be careful about the legal framework such as assuming all providers are initially assumed to be innocent.

Li *et al.* (2008) argue that the research efforts to detect conspiratorial fraud, for which insurer, patient and provider may conduct collobarative fraud, can be rewarding. This still has not been adressed. Another wide-spread concern in industry is the ability of fraudsters to adapt to algorithms and policy changes. This makes especially supervised algorithms obsolete after a while, because training data sets have an expiration date because of changing fraud patterns. For these purposes, we believe game theoretic approaches can be useful. For instance, adversarial risk analysis (Rios Insua *et al.*, 2009) can be suitable for more sophisticated fraud schemes. These methods are adaptive to dynamic fraud patterns and adjustments of fraudsters. Another consideration to combat changing fraud schemes can be social network-based methods, as argued by an official from a state Medicaid programme (GAO, 2012).

6.2 Bayesian Approaches

In addition to their potential advantages in decision analysis, Bayesian approaches can also be beneficial in other ways as part of medical fraud assessment frameworks. Bayesian estimation and inference has a number of advantages in statistical modelling and data analysis (Congdon, 2007). It can provide probability interpretations on quantities of interest such as hypotheses, intervals of parameters, membership of a subject and model selection (Jackman, 2009). For instance, probabilistic assessment of fraud can be retrieved and revised based on new data and expert knowledge. Probabilistic inference can be beneficial while profiling the providers and assessing their differences from their benchmark group for potential outlier detection.

Bayesian methods provide tools to work within hierarchical settings such as random effect models that account for variability of a given parameter across a number of groups (Gelman *et al.*, 2014). In that direction, the multinomial Bayesian latent variable model of Bayerstadler *et al.* (2016) uses covariate information to predict fraud probabilities within a supervised framework. An unsupervised alternative would be to extend latent Dirichlet allocation (Blei *et al.*, 2003) to identify hidden patterns among providers and medical procedures. Bayesian non-parametrics could also be useful in identifying homogenous groups of providers such as exploiting the clusterisation naturally implied by Dirichlet process mixtures.

Another potential advantage is the learning aspect of Bayesian approaches. For instance, despite the extensive use of sampling in medical audits, there is not a framework that enables the auditors to utilise all the information from the available samples. Learning about the general population by using already audited samples within an iterative sampling framework can be beneficial. Measuring the information gain of a prospective sample can help the auditor to allocate resources more efficiently. One such approach would be to employ information theoretic iterative sampling frameworks to evaluate the expected amount of information from the next sample. Learning about patterns can also be considered in the context of overpayment estimation and modelling. Novel overpayment models that use Bayesian inference can be proposed to quantify learning about the parameters of the overpayment distribution. The natural updating mechanism enables the auditor to combine sources of information and analyse adaptive fraudulent behaviour. Lastly, Bayesian approaches also handle missing values in a straightforward manner.

6.3 Modelling Extensions

Most of the statistical medical assessment literature focus on data mining methods which aim to identify potential fraudulent claims. Li *et al.* (2008) argue that there has been a lack of research in causal models that aim to identify the drivers of fraud, and the literature has not improved a lot in the last decade. Among limited models, Musal (2010) use regression models with dummy variables for geographic analysis of potential fraud. Although logistic regression has been employed as a classification approach within supervised methods, the causal relationship is not generally reported. Liou *et al.* (2008) use step-wise logistic regression to identify the most effective factors for the claims of patients with diabetes. They also suggest that sensitivity analyses can be conducted for neural networks to understand the drivers of fraud despite the fact that the significance of individual variables to fraud cannot be specified.

On the other hand, the literature of causal models against other types of fraud is more established. For instance, Viaene *et al.* (2004) use logistic regression in classification of automobile insurance claims. Logistic regression model has also been used to detect factors associated with fraudulent financial statements (Yue *et al.*, 2009). El Bachir Belhadji & Tarkhani (2000) identify indicators of insurance fraud using a regression model. Bermúdez *et al.* (2008) propose a Bayesian skewed logit model for fraud data. In addition to these regression models, hidden Markov models are proposed to detect credit card fraud (Srivastava *et al.*, 2008). We believe these approaches are worthwhile investigating for the medical fraud data.

Models that utilise dyadic representation of medical data can be useful to capture pairwise relationships and provide novel insights. Dyadic medical data such as the number of claims processed and the monetary charges associated with provider–patient pairs may be of interest. One potential model is to focus on the relationship between provider types and the counts and types of services they bill for. Such a Poisson-based co-clustering model can be used to signal fraudulent collective activities among providers such as kickback payments and abnormal referral rates. Federal Bureau of Investigation has suggested that high or very low referral rates may signal fraud (FBI, 2013). Therefore, the referral rates among providers can be further analysed to understand the processes. It can also be utilised to model monetary amount data which can reveal the patterns of charging for excessive or unnecessary services.

7 Conclusion

This paper provides a comprehensive survey of statistical medical fraud assessment. After providing a description of the medical data, statistical methods are discussed with a focus on sampling, overpayment estimation and data mining. We present illustrations of recently proposed unsupervised methods using real world medical claims data. In addition, we provide a discussion of a number of potential directions for future research, including decision analytic and Bayesian approaches.

Data-related issues such as restricted funding, concerns about privacy of the data, incomplete data and poor integration of available data are among the biggest challenges of the widespread application of these statistical methods. Access to medical data sets has been relatively limited due to legal, privacy and competitive reasons. Governmental health organisations and private insurance companies aim to provide more opportunities for researchers to access the medical data they possess. The Research Data Assistance Center (CMS, 2016c) is a CMS contractor that provides assistance in using Medicare and Medicaid data for research purposes under certain conditions. It should be noted that Medicare and Medicaid programmes are limited to certain population groups such as people who are over 65 or people who are below a certain income level. Therefore, only one-third of US citizens have access to these governmental programmes. With these limitations in mind, Health Care Cost Institute was founded by researchers and some private insurers to understand the drivers of health care costs and utilisation using private insurance data (HCCI, 2016). Another institution which may provide researchers information about disease specific fraud patterns is Centers for Disease Control and Prevention. In order to overcome data availability issues, researchers can choose to work with synthetic data. Data fusion methods may also help integrate publicly available data sources. Another issue is the challenge of forming teams that have both the required medical knowledge and the statistical expertise. Data privacy also becomes a concern within the integrated teams. Traceable medical data should always be secured and accessed with respect to the appropriate protocols.

Ability to detect new fraud patterns and misclassification remain important statistical issues in the supervised methods. There should be more generic rules in integration of different methods for better classification. After a fraud type is detected by investigators, it will be less likely to be used again as fraudsters will try other means. Legitimate patterns also can shift as insurance plans are updated in the competitive health care market. As new data become available, parameters need to be retrained or tuned accordingly. Therefore, a fraud system is preferred to have self learning and evolving capabilities to adapt to changing patterns. Unsupervised data mining methods should be considered more because of their ability to model dynamic changing fraud patterns with smaller costs. It is also pointed out that there is also lack of research in identifying potential drivers of fraud and prediction of fraud using these identifiers.

The increase in computational power and speed of accessing databases and performing data analysis make real-time pre-payment monitoring and analysis more feasible. Learning-based methods can be preferred because of the power of natural updating and adaption despite the expense of computational challenges. However, even when fully automating a decision process is possible, legal and ethical concerns may still warrant auditors to act as the active responsibles.

Post-payment fraud analysis should be done with caution because of the heterogeneous nature of medical claims data. Both false negatives and false positives are crucial and costly. False negatives result in misclassifying fraudulent cases as legitimate and lead to undeserved payments.

Whereas, false positives correspond to improper identification of suspect providers that are not engaged in fraud. This can change the public opinion of a physician and can have adverse effects. Models should not be applied nationwide without considering local conditions. The level of aggregation may turn out to be important. The decision maker also should be aware of the legal and ethical aspects while using metrics. Issar (2015) discusses such concerns and analyse the admissibility of statistical proof for medical reimbursement.

Statistical medical fraud assessment approaches should complement medical prevention, detection and response efforts. While fraud detection involves identifying fraud as quickly as it has occurred, fraud prevention describes the measures to stop fraud from occurring in the first place. Therefore, creating an anti-fraud culture and improving internal compliance systems have long term effects against fraud. Response efforts include improving the system or law enforcement initiatives to reduce the chances of future fraud. Fraud assessment both affects and is affected by the changes in policy. Despite that, most of the literature has not reported the practical implications of their findings for health care managers and decision makers. There needs to be more collaboration among the relevant parties of medical insurance programmes.

Notes

[1] https://www.merriam-webster.com/dictionary/fraud

References

Albert, J. H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc.*, **88**(422), 669–679.

Anderson, G. & Hussey, P. (2001). Comparing health system performance in OECD countries. *Health Aff.*, **20**(3), 219–232.

Aral, K. D., Güvenir, H. A., Sabuncuoğlu, İ. & Akar, A. R. (2012). A prescription fraud detection model. *Comput. Methods Programs Biomed.*, **106**(1), 37–46.

Bauder, R., Khoshgoftaar, T. M. & Seliya, N. (2017). A survey on the state of healthcare upcoding fraud analysis and detection. *Health Serv. Outcomes Res. Method.*, **17**(1), 31–55.

Bauder, R. A & Khoshgoftaar, T. M. (2016). A probabilistic programming approach for outlier detection in healthcare claims. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 347–354. Anaheim, California, USA: IEEE.

Bayerstadler, A., van Dijk, L. & Winter, F. (2016). Bayesian multinomial latent variable modeling for fraud and abuse detection in health insurance. *Insurance: Math. Econ.*, **71**, 244–252.

Belliveau, J. (2016). *Big Data Tool Saves CMS $1.5b by Preventing Medicare Fraud*. http://revcycleintelligence.com/news/big-data-tool-saves-cms-1.5b-by-preventing-medicare-fraud. Accessed: 2/25/2017.

Berger, J. O. (2013). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer Science & Business Media.

Bermúdez, L., Pérez, J., Ayuso, M., Gómez, E. & Vázquez, F. (2008). A Bayesian dichotomous model with asymmetric link for fraud in insurance. *Insurance: Math. Econ.*, **42**(2), 779–786.

Berta, P., Callea, G., Martini, G. & Vittadini, G. (2010). The effects of upcoding, cream skimming and readmissions on the Italian hospitals efficiency: A population-based investigation. *Econ. Modell.*, **27**(4), 812–821.

Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.*, **3**, 993–1022.

Bolton, R. & Hand, D. (2002). Statistical fraud detection: A review. *Stat. Sci.*, **17**(3), 235–249.

Capelleveen, G. C. (2013). *Outlier based predictors for health insurance fraud detection within US Medicaid*, Master's Thesis, University of Twente.

Carvalho, L. F., Teixeira, C. H., Meira, W., Ester, M., Carvalho, O. & Brandao, M. H. (2017). Provider-consumer anomaly detection for healthcare systems. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 229–238. Park City, Utah, USA: IEEE.

Chan, CL & Lan, CH. (2001). A data mining technique combining fuzzy sets theory and Bayesian classifier application of auditing the health insurance fee. In *Proceedings of the International Conference on Artificial Intelligence*, pp. 402–408. Seattle, Washington, USA.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *J. Am. Stat. Assoc.*, **90**(432), 1313–1321.

Cifarelli, D. & Regazzini, E. (1987). On a general definition of concentration function. *Sankhyā: The Indian J. Stat. Ser. B*, **49**, 307–319.

CMS. (2001). *Program Memorandum Carriers Transmittal B-01-01*. The Centers for Medicare & Medicaid Services. http://www.cms.gov/Regulations-and-Guidance/Guidance/Transmittals/downloads/B0101.pdf. Accessed: 09/01/2016.

CMS. (2005). *Cms Medicare Learning Network*. The Centers for Medicare & Medicaid Services. https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNMattersArticles/downloads/MM3557.pdf/. Accessed: 1/12/2018.

CMS. (2015a). *Medicare Fee for Service 2014 Improper Payments Report*. The Centers for Medicare & Medicaid Services. https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-FFS-Compliance-Programs/CERT/Downloads/MedicareFeeforService2014ImproperPaymentsReport.pdf. Accessed: 09/01/2016.

CMS. (2015b). *On Its 50th Anniversary, More Than 55 Million Americans Covered by Medicare*. The Centers for Medicare & Medicaid Services. https://www.cms.gov/Newsroom/MediaReleaseDatabase/Press-releases/2015-Press-releases-items/2015-07-28.html. Accessed: 09/01/2016.

CMS. (2016a). *Dataset Downloads*. The Centers for Medicare & Medicaid Services. https://www.cms.gov/openpayments/explore-the-data/dataset-downloads.html. Accessed: 2/25/2017.

CMS. (2016b). *Recovery Audit Program*. The Centers for Medicare & Medicaid Services. https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-FFS-Compliance-Programs/Recovery-Audit-Program/index.html. Accessed: 09/01/2016.

CMS. (2016c). *Research Data Assistance Center*. The Centers for Medicare & Medicaid Services.http://www.resdac.org/cms-data. Accessed: 09/01/2016.

CMS. (2017a). *HCPCS - General Information*. The Centers for Medicare & Medicaid Services. https://www.cms.gov/MedHCPCSGenInfo/. Accessed: 4/25/2017.

CMS. (2017b). *NHE Fact Sheet*. The Centers for Medicare & Medicaid Services. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nhe-fact-sheet.html. Accessed: 07/01/2017.

Congdon, P. (2007). *Bayesian Statistical Modelling*, Vol. 704. New York: John Wiley & Sons.

Copeland, L., Edberg, D., Panorska, A. & Wendel, J. (2012). Applying business intelligence concepts to Medicaid claim fraud detection. *J. Inf. Syst. Appl. Res.*, **5**(1), 51.

Dash, M. & Liu, H. (1997). Feature selection for classification. *Intell. Data Anal.*, **1**(1-4), 131–156.

Delamaire, L., Abdou, H. & Pointon, J. (2009). Credit card fraud and detection techniques: A review. *Banks Bank Syst.*, **4**(2), 57–68.

Dionne, G., Giuliano, F. & Picard, P. (2009). Optimal auditing with scoring: Theory and application to insurance fraud. *Manage. Sci.*, **55**(1), 58–70.

Edwards, D., Ward-Besser, G., Lasecki, J., Parker, B., Wieduwilt, K., Wu, F. & Moorhead, P. (2003). The minimum sum method: A distribution-free sampling procedure for Medicare fraud investigations. *Health Serv. Outcomes Res. Method.*, **4**(4), 241–263.

Ekin, T., Ieva, F., Ruggeri, F. & Soyer, R. (2013). Application of Bayesian methods in detection of healthcare fraud. *Chem. Eng. Trans.*, **33**, 151–156.

Ekin, T., Ieva, F., Ruggeri, F. & Soyer, R. (2017a). Bayesian co-clustering methods for assessment of healthcare fraud. Technical report TR-2017-1 I^{2}SDS, The George Washington University The Institute of Integrating Statistics in Decision Sciences, in preparation. Washington D.C.

Ekin, T., Ieva, F., Ruggeri, F. & Soyer, R. (2017b). On the use of the concentration function in medical fraud assessment. *The Am. Stat.*, **71**(3), 236–241.

Ekin, T., Musal, R. M. & Fulton, L. V. (2015). Overpayment models for medical audits: Multiple scenarios. *J. Appl. Stat.*, **42**(11), 2391–2405.

El Bachir Belhadji, G. & Tarkhani, F. (2000). A model for the detection of insurance fraud. *The Geneva Pap. Risk Insurance*, **25**, 517–538.

Faltin, F., Kenett, R. S & Ruggeri, F. (2012). *Statistical Methods in Healthcare.* West Sussex, UK: John Wiley & Sons.

Fawcett, T. (2003). In vivo spam filtering: A challenge problem for kdd. *ACM SIGKDD Explor. Newslett.*, **5**(2), 140–148.

FBI. (2013). *Physician Pleads Guilty to Role in Health Care Fraud Conspiracy*. Federal Bureau of Investigation. https://archives.fbi.gov/archives/dallas/press-releases/2013/physician-pleads-guilty-to-role-in-health-care-fraud-conspiracy. Accessed: 09/01/2016.

Furlan, Š. & Bajec, M. (2008). Holistic approach to fraud management in health insurance. *J. Inf. Organiz. Sci.*, **32**(2), 99–114.

GAO. (2012). *Medicare Fraud Prevention: CMS Has Implemented a Predictive Analytics System, but Needs to Define Measures to Determine Its Effectiveness*. United States Governmental Accountability Office. http://www.gao.gov/products/GAO-13-104. Accessed: 09/01/2016.

Gee, J. & Button, M. (2015). The financial cost of healthcare fraud: What data from around the world shows. Technical report, PKF Littlejohn LLP.

Gelman, A., Carlin, J. B, Stern, H. S & Rubin, D. B. (2014). *Bayesian Data Analysis*, Vol. 2. Boca Raton, FL, USA: Chapman & Hall/CRC.

Gilliland, D. & Edwards, D. (2010). Using randomized confidence limits to balance risk: An Application to Medicare fraud investigations.

Gilliland, D. & Feng, W. (2010). An adaptation of the minimum sum method. *Health Serv. Outcomes Res. Method.*, **10**(3-4), 154–164.

Grzymala-Busse, J. W & Hu, M. (2000). A comparison of several approaches to missing attribute values in data mining. In *International Conference on Rough Sets and Current Trends in Computing*, pp. 378–385. Berlin, Germany: Springer.

HCCI. (2016). *Health Care Cost Institute*. http://www.healthcostinstitute.org/. Accessed: 09/01/2016.

He, H., Graco, W. & Yao, X. (1998). Application of genetic algorithm and k-nearest neighbour method in medical fraud detection. In *Asia-Pacific Conference on Simulated Evolution and Learning*, pp. 74–81. Canberra, Australia: Springer.

He, H., Wang, J., Graco, W. & Hawkins, S. (1997). Application of neural networks to detection of medical fraud. *Expert Syst. Appl.*, **13**(4), 329–336.

Hillerman, T., Souza, J. C. F, Reis, A. C. B & Carvalho, R. N. (2017). Applying clustering and AHP methods for evaluating suspect healthcare claims. *J. Comput. Sci.*, **19**, 97–111.

Ignatova, I. & Edwards, D. (2008). Probe samples and the minimum sum method for medicare fraud investigations. *Health Serv. Outcomes Res. Method.*, **8**(4), 209–221.

Issar, N. (2015). More data mining for medical misrepresentation: Admissibility of statistical proof derived from predictive methods of detecting medical reimbursement fraud. *N. Ky. L. Rev.*, **42**(2), 341–374.

Iyengar, V. S, Hermiz, K. B & Natarajan, R. (2014). Computer-aided auditing of prescription drug claims. *Health Care Manage. Sci.*, **17**(3), 203–214.

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*, Vol. 846. Chichester: John Wiley & Sons.

Johnson, M. E. & Nagarur, N. (2016). Multi-stage methodology to detect health insurance claim fraud. *Health Care Manage. Sci.*, **19**(3), 249–260.

Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M. & Arab, M. (2014). Using data mining to detect health care fraud and abuse: A review of literature. *Global J. Health Sci.*, **7**(1), 194–202.

Kalb, P. (1999). Health care fraud and abuse. *The J. Am. Med. Assoc.*, **282**(12), 1163–1168.

Kumar, M., Ghani, R. & Mei, Z. (2010). Data mining to predict and prevent errors in health insurance claims processing. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 65–74. Washington, DC, USA: ACM.

Laleh, N. & Azgomi, M. A. (2009). A taxonomy of frauds and fraud detection techniques. In *International Conference on Information Systems, Technology and Management*, pp. 256–267. Ghaziabad, India: Springer.

Li, J., Huang, K.-Y., Jin, J. & Shi, J. (2008). A survey on statistical methods for health care fraud detection. *Health Care Manage. Sci.*, **11**, 275–287.

Lin, C., Lin, C.-M., Li, S.-T. & Kuo, S.-C. (2008). Intelligent physician segmentation and management based on KDD approach. *Expert Syst. Appl.*, **34**(3), 1963–1973.

Lin, J. & Haug, P. (2006). Data preparation framework for preprocessing clinical data in data mining. In *AMIA Annual Symposium Proceedings*, Vol. 2006: American Medical Informatics Association, Washington, DC, pp. 489.

Liou, F., Tang, Y. & Chen, J. (2008). Detecting hospital fraud and claim abuse through diabetic outpatient services. *Health Care Manage. Sci.*, **11**(4), 353–358.

Little, R. J. & Rubin, D. B. (2014). *Statistical Analysis with Missing Data.* Hoboken, NJ: John Wiley & Sons.

Liu, Q. & Vasarhelyi, M. (2013). Healthcare fraud detection: A survey and a clustering model incorporating geolocation information. In *29th World Continuous Auditing and Reporting Symposium*, pp. 1–10. Brisbane, Australia.

Lu, F. & Boritz, J. E. (2005). Detecting fraud in health insurance data: Learning to model incomplete benford law distributions. In *European Conference on Machine Learning*, pp. 633–640. Porto, Portugal: Springer.

Major, J. A & Riedinger, D. R. (2002). Efd: A hybrid knowledge/statistical-based system for the detection of fraud. *J. Risk Insurance*, **69**(3), 309–324.

Mohr, D. L. (2005). Confidence limits for estimates of totals from stratified samples, with application to Medicare Part B overpayment audits. *J. Appl. Stat.*, **32**(7), 757–769.

Musal, R. (2010). Two models to investigate Medicare fraud within unsupervised databases. *Expert Syst. Appl.*, **37**(12), 8628–8633.

Musal, R. M & Ekin, T. (2017). Medical overpayment estimation: A Bayesian approach. *Stat. Model.*, **17**(3), 196–222.

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 *IEEE International Conference on Data Mining Workshops*, pp. 613–622. Sydney, Australia: IEEE.

Ngai, EWT, Hu, Y., Wong, YH, Chen, Y. & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decis. Support Syst.*, **50**(3), 559–569.

OIG. (2010). *Ratstats statistical software*. Office of Inspector General, U.S. Department of Health and Human Services. http://oig.hhs.gov/compliance/rat-stats/index.asp. Accessed: 09/01/2016.

Ormerod, T., Morley, N., Ball, L., Langley, C. & Spenser, C. (2003). Using ethnography to design a mass detection tool (mdt) for the early discovery of insurance fraud. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems*, pp. 650–651. New York: ACM.

Ortega, P., Figueroa, C. & Ruz, G. (2006). A medical claim fraud/abuse detection system based on data mining: A case study in Chile. *In Conf. Data Mining*, **6**, 26–29.

Padmaja, T., Dhulipalla, N., Bapi, R. & Krishna, P. (2007). Unbalanced data classification using extreme outlier elimination and sampling techniques for fraud detection. In *International Conference on Advanced Computing and Communications, ADCOM 2007*, pp. 511–516. Guwahati, India: IEEE.

Phua, C., Lee, V., Smith, K. & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119*.

Rashidian, A., Joudaki, H. & Vian, T. (2012). No evidence of the effect of the interventions to combat health care fraud and abuse: A systematic review of literature. *Plos One*, **7**(8), e41988.

Rios Insua, D., Rios, J. & Banks, D. (2009). Adversarial risk analysis. *J. Am. Stat. Assoc.*, **104**(486), 841–854.

Shan, Y., Murray, D. W. & Sutinen, A. (2009). Discovering inappropriate billings with local density based outlier detection method. In *Proceedings of the Eighth Australasian Data Mining Conference*, Vol. 101, pp. 93–98. Melbourne, Australia.

Shin, H., Park, H., Lee, J. & Jhee, W. (2012). A scoring model to detect abusive billing patterns in health insurance claims. *Expert Syst. Appl.*, **39**(8), 7441–7450.

Sokol, L., Garcia, B., West, M., Rodriguez, J. & Johnson, K. (2001). Precursory steps to mining hcfa health care claims. In *IEEE Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, pp. 10. Wailea, Hawaii, USA.

Sparrow, M. (1998). *Fraud Control in the Health Care Industry: Assessing the State of Art*. U.S. Department of Justice. https://www.ncjrs.gov/pdffiles1/172841.pdf. Accessed: 1/12/2018.

Sparrow, M. K. (2000). *License to Steal: How Fraud Bleeds America's Health Care System.* Denver, Colorado, USA: Basic Books.

Spiegelhalter, D., Abrams, K. & Myles, J. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Vol. 13. New York: Wiley.

Srivastava, A., Kundu, A., Sural, S. & Majumdar, A.K. (2008). Credit card fraud detection using hidden markov model. *IEEE Trans. Dependable Secure Comput.*, **5**(1), 37–48.

Suleiman, M., Agrawal, R., Seay, C. & Grosky, W. (2014). Data driven implementation to filter fraudulent medicaid applications. In *SouthEastCon 2014*, pp. 1–8. Lexington, KY, USA: IEEE.

Tang, M., Mendis, B. S. U, Murray, D W., Hu, Y. & Sutinen, A. (2011). Unsupervised fraud detection in Medicare Australia. In *Proceedings of the Ninth Australasian Data Mining Conference-Volume 121*, pp. 103–110. Ballarat, Australia: Australian Computer Society, Inc.

Torgo, L. & Lopes, E. (2011). Utility-based fraud detection. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, Vol. 2, pp. 1517–1522. Barcelona, Spain: AAAI Press.

Travaille, P., Müller, R. M, Thornton, D. & Hillegersberg, J. (2011). Electronic fraud detection in the us Medicaid healthcare program: Lessons learned from other industries. In *17th Americas Conference on Information Systems, AMCIS 2011*, pp. 1–11. Detroit, United States.

Ulvila, J. W & Gaffney, J. E. Jr. (2004). A decision analysis method for evaluating computer intrusion detection systems. *Decis. Anal.*, **1**(1), 35–50.

van Capelleveen, G., Poel, M., Mueller, R. M, Thornton, D. & van Hillegersberg, J. (2016). Outlier detection in healthcare fraud: A case study in the Medicaid dental domain. *Int. J. Accounting Inf. Syst.*, **21**, 18–31.

Viaene, S., Derrig, R. & Dedene, G. (2004). A case study of applying boosting Naive Bayes to claim fraud diagnosis. *IEEE Trans. Knowledge Data Eng.*, **16**(5), 612–620.

Viaene, S., Derrig, R. A, Baesens, B. & Dedene, G. (2002). A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection. *J. Risk Insurance*, **69**(3), 373–421.

Viveros, M. S., Nearhos, J. P. & Rothman, M. J. (1996). Applying data mining techniques to a health insurance information system. In *Proceedings of the International Conference on Very Large Data Bases*, pp. 286–294. Mumbai, India: IEEE.

Williams, G. J. & Huang, Z. (1997). Mining the knowledge mine. In *Advanced Topics in Artificial Intelligence*, pp. 340–348. Berlin, Heidelberg: Springer.

Woodard, B. (2015). Fighting healthcare fraud with statistics. *Significance*, **12**(3), 22–25.

Yancey, W. (2012). *Sampling for Medicare and Other Claims*. http://www.willyancey.com/sampling-claims.html. Accessed: 09/01/2016.

Yang, W. & Hwang, S. (2006). A process-mining framework for the detection of healthcare fraud and abuse. *Expert Syst. Appl.*, **31**(1), 56–68.

Yue, D., Wu, X., Shen, N. & Chu, C. (2009). Logistic regression for detecting fraudulent financial statement of listed companies in China. In *International Conference on Artificial Intelligence and Computational Intelligence*, Vol. 2, pp. 104–108. Shanghai, China: IEEE.

Zhu, S., Wang, Y. & Wu, Y. (2011). Health care fraud detection using nonnegative matrix factorization. In *6th International Conference on Computer Science & Education (ICCSE)*, pp. 499–503. Singapore: IEEE.

[*Received August 2017, accepted March 2018*]

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.