Outlier based Predictors for Health Insurance Fraud Detection within U.S. Medicaid

by

Guido Cornelis van Capelleveen

B.Sc., Utrecht University (2012)

Submitted to the School of Management and Governance
in partial fulfillment of the requirements for the degree of

Master of Science in Business Information Technology

at the

University of Twente

December 2013

"Trust me, I am a doctor."

Dr. Pepper

This Master thesis has been supervised and is examined by the following persons:

Prof. Dr. Jos van Hillegersberg .....
Professor, School of Management & Governance, University of Twente
Thesis Supervisor

Dr. Mannes Poel .....
Assistant Professor, Department of Computer Science, University of Twente
Thesis Supervisor

Prof. Dr. Roland Müller .....
Professor, Department of Business and Economics, Berlin School of Economics and Law
Thesis Supervisor

Dallas Thornton, M.Eng M.B.A .....
Division Director, Cyberinfrastructure Services for the San Diego Supercomputer Center at the University of California, San Diego
Thesis Supervisor

Outlier based Predictors for Health Insurance Fraud Detection within U.S. Medicaid

by

Guido Cornelis van Capelleveen

Submitted to the School of Management and Governance
on November 25th, 2013, in partial fulfillment of the
requirements for the degree of
Master of Science in Business Information Technology

Abstract

This paper describes an effective method of outlier based predictors for health insurance fraud detection that identifies suspicious behavior of health care providers. Fraud and abuse on medical claims became a major concern for health insurance companies last decades. Estimates made for the studied U.S. Medicaid health insurance program is that up to 10% of the claims are fraudulent. Unsupervised data mining techniques such as outlier detection are suggested to be an effective predictors for fraud detection and should be used to support the initiations of audits. A method, based on comparative research, fraud cases and literature study has been proposed. We evaluated the method, by applying the method in a real life case study, were behavioral metrics were designed and 14 analytic experiments were built using outlier detection. The analysis ran on dental claim data and showed promising results. The proposed methodology enabled successful identification of fraudulent activity in several cases; however linking these identified incidents with irrefutable de jure fraud proved to be a difficult process. From 17 top suspicions analyzed, we reported eventually 12 of those to officials, a precision rate of approximately 71%. In the two interviews conducted with Medicaid Fraud Experts, experiences were gained on requirements for the design of the analytics and an effective implementation of the method. We found that outlier based predictors are not likely to succeed as fraud classification technology, though it explored an important role as decision supportive technology for resource allocation of fraud audits.

keywords: Health Insurance, Outlier based Analytics, Fraud and Abuse, Electronic Fraud Detection, Outlier Detection

Acknowledgments

Grateful to those who helped, inspired and supported me on experiencing science, especially to those, who already did this for 23 years.

Contents

1 Introduction **10**

1.1 Problem Description 10

1.2 Related Work 13

1.3 Outline 17

2 Research Approach **18**

2.1 Research Goals & Objectives 18

2.2 Research Questions 19

2.3 Research Method and Validation 21

3 Health Insurance Fraud **24**

3.1 Fraud and the Detection Process 24

3.2 Medicaid and Fraud Schemes 26

3.3 Analyses & Applicable Detection Techniques 33

4 Metrics, Capturing Provider Behavior **38**

4.1 Provider Metrics, definition and types 38

4.2 Medicaid Data Model and Referential Data 40

4.3 Metric Identification 41

5 Method for Predictors of Health Insurance Fraud Detection **44**

5.1 The Method 44

5.2 Seven Iterative Phases 46

6 Case Study: Medicaid Dental Providers **57**

7 Experts Evaluation **58**

8 Discussion **59**

8.1 General observations 59

8.2 Limitations 61

8.3 Lessons Learned 61

8.4 Future Work 62

9 Conclusion **64**

A Tables **66**

B Figures **67**

C Interview Questions **72**

References **74**

List of Figures

4-1 Medicaid Multi-dimensional Data Model [52] 40

5-1 Method for Predictors of Health Insurance Fraud Detection 45

5-2 Box-plot and probability density function of a Normal N(0,σ^2) Population, derived of [57] 51

B-1 Medicaid Tooth Numbering System-Permanent Teeth. 68

B-2 Medicaid Tooth Numbering System-Deciduous Teeth 69

B-3 Elbow Figures K1-4 70

B-4 Elbow Figures K5, K6, U1 71

List of Tables

3.1 Fraud Audit Initiation Types. 30

3.2 The seven levels of health care fraud control by Sparrow, adapted by Thornton to emphasize their focus area [49] [52]. 31

Chapter 1

Introduction

1.1 Problem Description

During the fiscal year 2011, Medicare & Medicaid, United States (US) national social insurance programs, covered over 118 million people from which roughly 40 million aged 65 and older (Medicare), about 8 million disabled (Medicare) and 70 million individuals with low income and resources (Medicaid). Total expenditures cost the U.S. government around $549 billion for Medicare and $432 billion for Medicaid and are increasingly growing due to population aging, rising medical prices and changing program policies [15] [14].

With a total expenditure of 17.6% of the Gross Domestic Product (GDP) in 2010, the US are by far leader in citizen health care costs compared to other nations, that spend an average of 9.5% of their GDP, according to the Organization for Economic Co-operation and Development [34]. Such high expenses make it reasonable to assume the system is an interesting target for fraudulent activities. The Government Accountability Office (GAO) designated Medicare & Medicaid as high-risk programs, in part because its complexity that creates the vulnerability to fraud [40]. Government and private plaintiffs thus constantly try to condemn many of those fraudulent health care providers enrolled in Medicaid, however the have difficulties in identifying the fraud.

Due to the inconspicuous nature of fraud, it is hard to give concrete figures and percentages of it, however health care fraud estimates from different entities suggest estimations ranging from 3 to even 10% (3 to 10% according to the FBI [37], and 8% estimated by GAO [43]). Based on these estimates, the total yearly loss for the entire US health care system as a result of fraud could be calculated to be around $125 to $175 billion [26], of which about $80 billion in Medicare according to Coalition against Insurance Fraud [2] and comparable amounts in Medicaid [30]. This fraud exists in many forms and is becoming more covertly advanced on a daily basis. The most prevalent health care schemes include billing for services not rendered, up-coding of services or items, duplication of claims, un-bundling of claims and providing excessive or medically irrelevant services.

The huge amounts of economic losses demand for extra measures and prosecutions. The department of Health & Human Services (HHS) accounted in 2011 for an allocation of $1.7 billion, funded by the Health Care and Abuse Control (HCFAC) program, to prosecute and investigate health care fraud. Although the allocation in absolute figures is large, it is still small compared to the merits on restitution and fines, and is said to have more potential. Sparrow, for example, criticizes the high rate of return which he considers to be disproportionate to what could possibly be retrieved by fraud detection [49]. As can bee seen from budget expenses of the CMS integrity group, more dedication has been devoted to fraud detection over the past few years, and spending's are still rising [35]. The detection of abusive and fraudulent practice in health care is difficult because uncertainties inherent in medical practices result in variable care processes [21]. Therefore, medical experts must review each case, which can be time consuming and expensive. A large part of the budget is spent on auditing providers but because this is an extremely intensive and complex task, only a select group of providers can be screened. Detection and indication of fraud is therefore a crucial process to effectively allocate audit resources.

Health care fraud is mainly be detected using three types of strategies: audits, market signals and electronically fraud detection. Audits exist in the a random form as well as in a targeted initiated by fraud investigators. Market signals come from employees, better known as whistle blowers, or patients who visited the providers. The complaints by disgruntled interested parties may lead to qui tam cases. Electronically fraud detection, consisting largely of anomaly detection uses computational power to find fraud using pre-existing rule checks or statistics. Where auditing strategies require the use of trained personnel to evaluate the process and/or product, statistical methods rely on large data sets to identify potential anomalies [11]. These audits and qui tam cases are the common practices, while electronic fraud detection is a relatively new field within the health insurance branch. Rule based checks were implemented with the introduction of larger databases. Data mining became an possibility now supercomputers, data warehouses, and big data are becoming more common place.

As suggested in previous literature electronic fraud detection could make a huge difference in health care fraud as it could secure the claim input process, check on irregularities and afterwards could analyze claim data sets to search for (behavioral) indicators of potential fraud [4] [17] [7] [41]. Although of the late acknowledgement of fraud in health care, the complexity of the claim systems, the size and distributed storage of claim data and the late and relatively low funding on fraud detection, development of electronically fraud detection systems is lacking behind compared to the adoption of such technologies within comparative industries such as credit-loaning and telecommunication. As mentioned in the work of Travaille et al., there is a large base of statistical methods that are also used in other industries and could potentially be applied within the health care industry [53]. Some research reported already on specific fraud scheme detection using data mining approaches [17] [29] [46] [31] [32], however the outstanding challenge is to explore other health care fields for potential data-mining possibilities and develop a generic approach towards this problem.

Although statistics are a promising candidate, it should be noted that detection

still relies for a large part on market signals and audits. This is because behavior might be an indication of fraud, but it will never be a fault proof way to separate the good from bad, neither will it identify all fraudulent providers. Thus we encourage the use of data mining as predictors for fraud detection and suggest that the key for success lies within a mix of methods.

1.2 Related Work

Advances in information technology, digitalization of health care information and the research on health insurance fraud have opened an area for data mining and machine learning applications to fight fraud. Within the area of data mining and machine learning, technologies have been widely studied last years and with knowledge from similar areas, progress has been made. Especially data-mining (DM) is gaining more attention by researchers as a potential tool to find health insurance fraud more easily [4]. Literature differentiates data mining and machine learning into supervised, unsupervised, hybrids or semi-supervised methods. As supervised techniques require the data to be labeled for building a training set, unsupervised techniques will deal with data based on group or statistical outlying behavior. The unsupervised methods are a piece of technology to identify potentially fraudulent transactions, that additional require the use of expertise to determine the legitimacy of the claims. Although fraud detection research is a relatively large field, most of the studies consider outlier detection as the primary tool [56].

After introduction of many data mining technologies researchers have combined multiple methodologies such as fuzzy logic in medical claims assessment and neural networks for automatically classification [8]. In the early zeros, a first concepts of data warehousing for data mining purposes in health care arose, described inter alia by Forgionne, who reported on an intelligent data mining system to detect health care fraud [17]. The solution described was an early attempt on the utilization of

data warehousing, data mining, artificial intelligence and decision support systems to develop a proactive and effective health care fraud detection strategy.

Realization on a larger scale was researched by Major and Riedinger who developed an electronic fraud detection application to review twenty thousand providers on 27 behavioral heuristics and compare those to similar providers. A provider score was calculated based on these heuristics followed by a frontier identification method to select providers as candidates for investigation. Although the research alerted officials on almost 900 suspicious providers of which only 23 led to further investigations [29]. The consultants identified 91 providers with fraudulent activities. The performance measure, defined as the subset proportion of EFD identified candidates in candidates identified by investigative consultants, yields value in the performance of the task, although not yet in huge profits. Another example was the experimental application of Yamanashi who identified a number of meaningful rare cases in pathology insurance data from Australias Health Insurance Commission using an on-line discounting learning algorithm [59]. In Taiwan within the National Health Insurance (NHI) program scientists developed a detection model based on a process mining framework that systematically identified practices derived from pathways to detect fraudulent claims [60]. The examples returned by the non-structure detection model captured 69% percent of the examples on average. The empirical results showed that the proposed detection model was efficient and capable of identifying some of the fraudulent and abusive cases within clinical instances. In Canada researchers used Benford's Law Distributions to detect anomalies in claim reimbursements [28]. Although the method did find some remarkable behavior there were some requirements for the use of Benford's law and the potential fraud identification set seemed to be limited. One of the main reasons is that Benford's law searches for the relation between frequently used same number and fraud, however this does not always seem to be the case. Many services have fixed prices and applying Benford's law requires qualitative judgement by domain experts and lacks on validation. In Chili a private health insurance company has built applications of MLP neural networks used to

find medical abuse and fraud. The innovative aspects of the application concerned a method that could process the claims as fast as possible on a real time basis. The researchers reported a detection rate of approximately 75 fraudulent and abusive cases per month making the detection 6.6. months earlier than without the system [41]. The system was able to retrieve 73.4% of the fraudulent billings, while having a false positive rate of 6.9% at the operation value calculated by FP cost/FN cost of 0.2.

More recent examples on health insurance fraud detection can be found in the work of Shan et al., Musal, Ng et al., and Tang et al. In the research of Shan et al. applied association rule mining were studied to examine billing patterns within a particular specialist group to detect these suspicious claims and potential fraudulent individuals [45]. Domain experts in Medicaid Australia examine the identified associated results. The subject matter expert rated the rules into low, medium or high categories and providers were measured on the occasions they broke those rules. According to the fraud experts, the medium and higher rules, may be directly related to non-compliant practices and could be taken for measurement of effectiveness. The research reported an accuracy of 20.83% of providers with more than 5 violations, which is more effective for identifying suspicious billing patterns than random sampling. Musal described two models to investigate Medicare fraud that made use of clustering procedures as well as regression models for geographical analysis of possible fraud [31]. The authors believe in a system dynamic approach required to investigate the Medicare system in decisions involving the investigation of possible providers of fraud. Ng et al. experimented on detecting non-compliant consumers (prescription shoppers) in spatio-temporal health data of Medicare Australia using multiple metrics that flagged providers [32]. A modular framework that brings disparate data mining techniques together was adopted and showed high hit rates. Of the 12 people identified, 8 are believed to be prescription shoppers, 4 with high confidence and 4 potentials. Although beneficial experimental results were achieved and the authors consider spatial and temporal factors to be effective in metrics, significant benefits concerning the use of spatial-temporal factors instead of more traditional metrics

could not be verified. The more simple metrics such as multiple visits or prescription percentages of pharmacy visits for drugs of concern have proved valuable activity as well. Tang et al. described in their research the problem of prescription shopping, although it took a slightly different approach than Ng et al. [51]. Instead, the application of Tang integrated techniques like feature selection, clustering, pattern recognition and outlier detection. Using a threshold on the outlier score provider groups could be marked as potential fraudulent. The transaction records revealed that some extreme cases have multiple visits to different doctors on one day, however effectiveness was only qualitatively measured. The most recent study is that of Iyengar et al. who described a methodology for identifying and ranking candidate audit targets from prescription drugs. The researchers developed a normalized baseline behavioral model for each prescription area and searched for statistically significantly deviations from that model [24]. For some of the areas, up to 500 features were used to find anomalies. Validated was if known cases of fraud were detected by the model. For the narcotic analgesics drug class, all the known cases of fraud were correctly identified by the model as being very abnormal and excessive. The research of Thornton et al. builds upon Sparrows fraud type classifications and developed a Medicaid multidimensional data schema and elaborated on analysis techniques that help to predict the likelihood of finding fraudulent activities. [52]

Besides found case studies, the scope and extent of health care fraud was described by Travaille et al. that provided an overview of the electronic fraud detection from other industries that could be applied within the health care industry. The authors advocated the use of statistical methods for detection fraud and abuse for many of the health care areas, and gave insight in the multiple fraud schemes that are used by criminals in health care [53]. The work of Phua might be interesting as well concerning this topic, because he published a comprehensive survey of data mining-based fraud detection research [42]. He categorizes, compares and summarizes from almost all published technical and review articles in automated fraud detection of a decade research papers.

In general, the papers suggest and justify the applicability of data mining techniques for detecting health care fraud. Most describe the process of metric gathering, valuing and validation and how dynamics should be adopted within a continuously changing fraud environment. Most papers have a focus on specific health care area which seems to indicate a non-homogeneous field for application. In search for generalizability we look for a common approach that can be extended to multiple areas and applied flexibly on a large scale. Our personal goal is to apply these methodology on the total set of Medicaid data, which is concerning its 70 million beneficiaries one of worlds largest health insurance programs. Therefore there is a need for a generic approach to developing predictors for detection of health insurance fraud in multiple health areas.

1.3 Outline

This paper is organized as follows. Chapter 2 states the research questions and elaborates on the methodology taken to design the method for health insurance fraud prediction, how the design was implemented in a case study after which multiple experiments were conducted that would be validated by fraud subject matter experts. Chapter 3 elaborates on the health insurance domain, the kind of fraud that is known to be committed, the current state of practice and available technology to fight it. In chapter 4 described how metrics are created, used in analyses and the relation to fraud cases found within the health industry. Chapter 5 provides an extensive description on the proposed method we developed. In chapter 6 results are reported concerning the 14 analytic experiments we performed within the Medicaid program. Chapter 7 examines the results and method by use of the fraud subject matters, and we finish by discussing the results, compare it to similar research methodologies within data mining to find fraud, and conclude our findings in chapter 8.

Chapter 2

Research Approach

2.1 Research Goals & Objectives

The goals (long-term) and objectives (short-term) of a research project summarize what is to be achieved within the study. This study is an attempt to gain knowledge on the applications of unsupervised data mining techniques to predict fraud within health care insurance domain. For this research the following goals have been set:

- The research will contribute to the health care fraud domain by providing insight and guidelines for designing fraud prediction/detection systems using unsupervised data mining techniques. (Scientific request)
- The research tries to contribute to the early detection of fraudulent billings in Medicaid while reducing the false positives in selection of providers for audit. (Business request)
- The research tries to contribute by defrauding the Medicaid health care insurance to reduce costs for this sector enabling better and accessible health care. (Environmental/Social request)

The objective in the research is to:

- Provide insight in the development of metrics and application of unsupervised data mining techniques that can be used for fraud detection within health care
- Describe a design for applying metrics as a generic approach for provider fraud detection within health care
- Validate the effectiveness of this design by use of a case study and expert validation

2.2 Research Questions

The research objectives are general statements of what should be pursued in the research. To translate the objectives they have been specified to the research questions to determine what to study.

Based on the problem area we concluded that an as-of-yet insufficiently addressed challenge exists concerning holistic and effective fraud prediction method that is flexible and configurable to be transferable to the entire domain of health insurance fraud detection. The request for a generic fraud detection approach that could be placed on top of existing data warehousing and business intelligence architectures is a currently problem that multiple national and private health insurance organizations have struggled with over the last couple of years. The design of such an approach, application of prediction techniques and metrics will be marked as overarching themes during this research, with an emphasis on a case study to test the applicability of the method. As may be read in the research methodology section, an approach of design science was chosen as a combination of Organizational Design & Information System Design activities should be addressed and disseminated. Both the fraud detection process as well as the role of information systems and technology should be addressed in an equal fashion and therefore Hevners design science theory is an appropriate method to implement in this case [23]. For the formulation of the research question we took the important elements of Hevner's design science theory to compose are main research

question:

- Artifact: Prediction Method for Fraud Detection
- Context: Health insurance
- Treatment: Unsupervised data mining techniques
- In order to: Identify suspicious providers claiming behavior
- Design Criteria: Effective Fraud Detection

The general research question will therefore be stated as:

Ẅhat is an effective method for predictive health insurance fraud detection that identifies suspicious provider claiming behavior using unsupervised data mining techniques?¨

The main question will be subdivided into three important section:

First should be identified which unsupervised techniques are applicable to identifying the suspicious provider behavior. To understand what are applicable techniques it is required to know the nature of fraud, the schemes that are being used and second data mining techniques available should be studied and though on how they could be used to find fraud based on unlabeled data.

Second, further investigation should occur in order to identify effective metrics that would identify the suspicious behavior of providers. In order to answer an overview of data set features is required in order to design the metrics, and to map those to provider behavior. Apart from the claim data, external sources could be used as well. Questions on the identification of metrics, the maintenance and life expectancy of metrics because of the changing insurance environment, and the optimization of metrics will rise accordingly.

The third area of interest consists of finding out how the analytic detection technology should be adopted within the process of fraud detection. Technological feasibility is a great accomplishment, however the knowledge and tools should be transferred to the domain were fraud experts can and will actually work the provided tools to support them in their professions. This relates to the capability of information technology towards the possibilities of intervening in the fraud detection process.

2.3 Research Method and Validation

Research methodology refers to the behavior and instruments used in research for selecting and constructing the research techniques to gather data. Operationalization is the development of specific research procedures that will result in empirical observations representing those concepts in the real world. We will thus discuss the selected approach and the procedures and research instruments involved.

Since the nature of the research is in an Information Systems (IS) area, it consists mainly of two disciplines, business administration (behavior science) and computer engineering (design science) [23]. While behavior science in IS is mainly concerned with the development of theories and prediction of events that occur when an artifact is used, called perceived usefulness and impact on individuals and/or organizations depending on system, service and information quality [12], design science focus on the creation and evaluation of IT artifact intended to solve organizational problems [23]. Hevner et al. describe the design science paradigm seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artifacts. The artifact in Hevner, can be defined as constructs (vocabulary and symbols), models (abstractions and representations) and instantiations (prototype and implemented systems). The research methodology of Hevner et al. is chosen because in the research we design a construct, represented by the method of fraud predictors for health insurance fraud detection, that should give the health fraud domain guidelines for de-

veloping applications based on these outlying techniques.

Our first step was to gain awareness of the problem area. The approach for researching the problem area, has taken the form of an extensive literature review to establish insight in previous experiments and the learned lessons from using data mining technology for this purpose. Secondly, to identify current fraud detection practices and processes, documentation of the Center of Medicare and Medicaid on Medicare & Medicaid Fraud & Abuse on Prevention, Detection, and Reporting has been studied, and thoughts have been exchanged with two fraud experts from Medicaid.

Second step was to identify the unsupervised techniques applicable for the use of identifying the suspicious provider behavior and in which way they should be used or implemented. This was again supported by literature review on available techniques and the study on comparable case studies from large health insurance providers, most found within scientific literature. The gathered set of candidate unsupervised techniques proposed were adopted in our knowledge base.

The third step was to gather a set of promising metrics that could be used as features for the provider analysis that would compare their behavior in both a time dimension as on group similarity. The metrics have been identified by remarkable outlying behavior reported in fraud cases by the FBI [38], as well as a large private lawyer office, operating in most U.S. states, dealing with Medicare and Medicaid fraud cases. The reported cases, together over 650 items spread over a period of two years, have been analyzed for outlying provider behavior. Together with considerable items suggested by fellow researchers in the field reporting on similar case studies we composed a set of metrics which consisted over a hundred metrics, of which fifteen interesting, feasible implementable metrics were chosen to elaborate and adopt in the case study.

Fourth, metrics and unsupervised techniques were matched together and configured to compare provider behavior to build suspicion of certain providers based on the knowledge of fraud schemes, as learned previously. The gathered set of metrics, the knowledge of fraud schemes and the availability of applicable unsupervised data mining techniques leaded to the creation of the method as presented in this thesis. Using the scoring of multiple predictors a suspicion on potential fraudulent providers could be build.

Fifth, a case study has been performed to gain experiences on provider behavior, metric effectiveness, but above all, to be able to discuss results with medical fraud experts in order to evaluate and validate the proposed method. The case study was performed on Arkansas dental claim data consisting records of a million beneficiaries, over 500 dental care providers and millions of transactions. Specific details on data preparation can be found in the case study. The analyses were designed in the R language using a custom developed R application that ran on the Medicaid Integrity Group Data Engine at the San Diego Super Computer Center.

Our last step was to evaluate the case study and and to fit the designed method into the counteract process of fraud. This includes a proposition on the implementation of the method on top of Medicaid analytic tools for claim analysis and the extension of the case study to a larger scale. To develop the predictors and analysis we conducted subject matter expert interviews with two fraud Medicaid Experts to evaluate and validate the method developed in this research.

Chapter 3

Health Insurance Fraud

3.1 Fraud and the Detection Process

In order to research health care fraud domain, it is important to have clear distinctive view on what fraud is, certainly because mostly it is named together with abuse and waste. In general fraud and abuse refer to the situation in which health care is paid for but not provided, or a situation in which reimbursements are paid to third party insurance companies or federal programs such as Medicare or Medicaid, though no such services were rendered. Other definitions define fraud and abuse as health care providers receiving kickbacks, patients seeking treatments that are potentially harmful to them (such as seeking drugs to satisfy addictions), and the prescription of services known to be unnecessary [26]. Waste on the other hand, refers more to the inefficiency of health care processes, which mostly is unrelated to fraud. Provision of medically unnecessary services is also a form of waste, but only considered fraud when provided intentionally. We will use the definitions of fraud and abuse given by the Center for Medicare and Medicaid Services [16]:

Fraud: Purposely billing for services that were never furnished and or supplies not provided, medically unnecessary services and altering claims to receive higher reimbursement than the service produced.

Abuse: The billings of practices that, either directly or indirectly, are not consistent with the goals of providing patients with services that are medically necessary, meet professionally recognized standards, and are fairly priced.

In the war against health care fraud, we identified six process stages ranging from identification up to prosecution. The six stages will not necessarily subsequent for each particular fraud case; only one of the second or third stage often occurs, or are completely absent in some cases. Although some of the terms in literature are interchangeably used and overlap by the following descriptions was tried to distinguish the six stages:

1. ***Fraud Prevention:*** is the attempt to reduce crime and deter criminals to commit fraud in the health care system. This can be done by preventive measures such as education, morality, law enforcement and maintenance of criminal justice.

2. ***Fraud Suspicion:*** is the cognition of mistrust, there might be potential fraud because we doubt a person, his behavior, however without any proof. An example is of such suspicion is when we never see any clients at a health care provider or when you have received services you are not sure about the reason why they have been provided to you.

3. ***Fraud Detection Predictors:*** is a statement that some outcome is expected. We are not predicting providers that are going to commit fraud, but detect a suspicion of on providers based on claiming behavior which may lead to the detection of fraud or further fraud investigations. Fraud Detection predictors, analyses and forensic analytic tools may be used. This type is referred as indicative fraud potential.

4. ***Inconsistency or Fraud Detection:*** refers to the discovery of a fraud crime, the catch of fraud by inconsistencies in a submitted claim. The fraud can be ascertained without cooperation from or synchronization with the provider. Rule

based systems that check for un-bundling of service codes or check for treatment on dead patients are examples of fraud detection systems. The condition for detection is that there is a clear and direct relationship between the submission and unlawful acts, referred as observational fraud.

5. ***Fraud Audit or Investigation:*** An audit or investigation is a human process that goes through interviews, medical documentation or on site reviews to gather all proof for fraud which will could lead to prosecution or settlement. We can now speak of an internal fraud accusation.

6. ***Fraud Prosecution:*** After Investigation when enough proof of committed fraud has been gathered one could go for conduct of a legal proceeding, making a public accusation. If convicted of, we can refer as committed fraud.

As may be seen from the fraud detection process, one can only speak of fraud when it has been proven to be fraud. As well, it is a grey area when speaking of lawful inconsistencies, as the provider intention that should be revealed from the audit will indicate if the case will be judged as fraudulent or not. For example, a provider might send in a claim for a deceased patient. Although this might seem obvious case of fraud, still one should proof that it wasn't just an accident, where the wrong patient was claimed for the services provided. Until proved otherwise, in many of the cases one should when stating fraud, actually refer to as indication of fraud, potential fraud or fraud accusations.

3.2 Medicaid and Fraud Schemes

Three areas are important for the design of a system that may effectively predict where to detect fraud. First, the health insurance system and associated claim processes, because of the properties of the transaction data as well as the semantics of the stored data within a broader context of health insurance. Second, the variety of fraud that is committed, the schemes to look for and the prospect of measurement

of targeted predictions. Last, one should know about the current initiatives on fraud diminution, their associated fields, the available resources and the progressed results so far to be able fit in or replace new practices for fraud detection.

The health insurance program Medicaid, the United States federal health program for families and individuals with low income and resources, was taken for the development of our case study. With roughly 70 million people enrolled in the program, it belongs to the worlds largest insurance programs. People eligible to the program may request for health care services. Once received, the performing service providers may claim directly for patients costs made to Medicaid. An example is a doctor visit, prescribed medications or home health service, usually a fixed reimbursement defined in the beneficiaries program policy. A lot of the providers have an agreement on prices charged for services of Medicaid patients. Those providers sends their claims in case of full reimbursement directly to Medicaid. Beneficiaries will not to be bothered with the claim processing details and only receives the Explanation of Benefits (EOB) afterwards. Providers without such agreements, for instance laboratories, will send a bill and EOB to the beneficiary, who sends it subsequently to Medicaid for reimbursement. Every state will process the Medicaid claims and validates the legitimate of a claim under the state specific policies before continuing to payment. Because of the size of incoming claims, automated checks are performed that search for claim code inconsistencies, data incorrectness, service pricing's and claim duplication's. These edits and audits are designed to verify the information with honest providers in mind, the system lacks of effective fraud-referral mechanisms [49]. More or less, we can conclude that the payment integrity is sacrificed for processing efficiency. A claim may pass the audit creating reimbursements for service providers while services have not been performed, are not billed correctly or are medically irrelevant at all. Sparrow identified and described the forms of fraud and what to look for. In his book he reported on seven types of fraud schemes that form the problem in the American health insurance programs [49]:

- Billing for services not rendered (Identity theft & Phantom billing)
- Up-coding of services and items (Up-coding)
- Duplicate billing
- Un-bundling of claims (Un-bundling / Creative Billing)
- Medically unnecessary services (Bill Padding)
- Excessive services (Bill Padding)
- Kickbacks

While some of the schemes can be quite easily detected during the automatic claim processing, for most of the schemes it quite difficult to detect potential fraud. This is mainly caused by the invisibility of service provision to beneficiaries, the complexity of claims and the medical expertise necessary to review and judge the claims. Complementary, we have to deal with the invisible undefined nature fraud which makes it difficult to look for something when not knowing what to look for. There is for example no defined distinction between fraud and abuse when it comes to medical necessity. Understanding of motivations of a physician can not be computed. There are no tools yet to create a pattern of his reasoning logic comparing that to medical expedient behavior, in order to infer the criminal motivation. The equal applies for instance to physician referrals, which are a daily practice of the physicians job. The referrals may though be incentified by kickbacks, incentivizing the act of referral forgery. While most of the schemes are conducted in a slow approach, *steal a little all the time*, using schemes such as services not rendered, is also exists in the more aggressive form, known as the *hit and run* practice. A form of this hit and run practice is as false companies are registered to bill as much as possible followed by a sudden disappearance of all the staff once authorities start to investigate these companies.

The U.S. Government created new laws and changed policies and practices to fight fraud, as can be seen by the initiatives in the anti fraud affordable care act [48].

Current initiative to fight fraud were mainly found in programs on national and state level. The Medicaid Integrity Program is the federal attempt to reduce fraud and abuse in the Medicaid program since the states have difficulties to control the situation. The Center of Medicare and Medicaid Services (CMS), part of the department of Health and Human Services, runs Medicaids Integrity programs enacting both prevention and investigation [16]. The institute closely work together with the Federal Bureau of Investigation (FBI) for national investigation of fraud and the department of Justice (DOJ) to prosecute criminals. Within the Medicaid Integrity Program, fraud investigation teams are requested to find fraud on state basis. Medicaid Fraud Control Units (MFCU), single identifiable entities of state government, annually certified by the Secretary of the U.S. Department of Health and Human Services, conduct a statewide program for the investigation and prosecution of health care providers to defraud the Medicaid program [36]. At last the U.S. provides the Civil False Claim Act (U.S. Code; 3729 False claims) [39] allowing citizens to file actions against federal contractors on the accusing of fraud and gaining part of the recovered national expenses. This resulted last decade already in a huge gain in expenses for government as well as the complaint filers and representatives. Within the Integrity programs, both nation as state wide fraud has been identified and investigated in multiple ways and even some attempts regarding the application of data mining technologies have been initiated. Preventive initiatives may be found in policy changes within the programs of Medicaid and Medicare.

Based on the study of program initiatives concerning fraud identification a separation of initiations on fraud audits or investigations could be derived. Presented in table 3.1, five common ways may be seen in which the fraud investigations normally are initiated. Although other Electronic fraud detection techniques exists, only most frequent occurred ones were listed.

Each of the fraud indication types has pro and cons. Random sampling will usually result in low findings, however it may reveal new kinds of fraud. Searching for

Table 3.1: Fraud Audit Initiation Types.

Fraud Audit Initiation	**Description**
Submission problems	While submitting a claim, there are already problems identified by the first auditors. Example: frequent problematic claim submissions are subjected to investigation. (rule based transaction)
Random sampling	Randomly selected claim submissions are investigated.
Whistle blowers or indications from the sector	Co-workers report suspicious behavior or abuse of their colleagues or cooperation providers.
Searching for similar known fraud schemes	Fraud investigators profile a provider based on claim data on a specific fraud scheme and start investigation based similar profile (profiling, rule-based data mining)
Unusual claim behavior/ Predictive Fraud Detection	Fraud investigators analyze claim data and looking for unusual behavior (outlier detection/ anomaly detection / clustering outlier)

similar fraud schemes normally will be a short term approach, as it will optimize for the search of one type of fraud, leaving out the others, but might be more effective. Criminals will discover and undermine the detection method eventually, shifting their fraudulent behavior to other more sophisticated schemes. We argue that fraud should be fought and initiated at each level to have a balanced fraud investigation process. It will both be effective on short term for high recovering as well as sustaining contingency for keeping track of fraud movements and development. This study focused mainly on the unusual claim behavior as indicator for the initiation of an intensive fraud investigation due to the relatively new research area and potential for finding new fraud as well as known fraud with high detection rates.

Measuring behavior can be done on each level of sparrow health care fraud control stack, going beyond the single transaction level. Where the predictors within this research were mainly targeted for the provider (3b) and the patient-practice level (4b), one could create similar metrics for the other levels. It should be taken into consideration that developing fraud metrics for the higher levels are associated with higher

Table 3.2: The seven levels of health care fraud control by Sparrow, adapted by Thornton to emphasize their focus area [49] [52].

		Phantom Billing	**Duplicate Billing**	**Upcoding**	**Unbundling**	**Excessive or Unnecessary Services**	**kickbacks**
Level 1	Single Claim, or Transaction			*	*	*	
Level 2	Patient / Provider		*		*	*	
Level 3	a. Patient	*	***	*	***	*	
	b. Provider	**		***	*	***	
Level 4	a. Insurer Policy / Provider	**		*	**	**	*
	b. Patient / Provider Group	*	*	*	*	*	
Level 5	Insurer Policy / Provider Group	**		**	**	**	*
Level 6	a. Defined Patient Group	**		*	*	**	**
	b. Provider Group	**		***	**	***	*
Level 7	Multiparty, Criminal Conspiracies	**		**	*	**	***

Usefullness: * Low ** Medium *** High

complexity and requisite more resources. Behavioral metrics usually are developed to reveal suspicious behavior related to certain schemes types. According to Thornton et al. each level can also be related to different usefulness for the detection of each of the fraud types, as can be seen in Table 3.2 [52].

According to the usefulness and focus on provider behavior, the fraud schemes to look for should mainly be found in phantom billing, up-coding, un-bundling and the excessive or unnecessary services. Another approach that will provide an indication on the fraud to be found within the case study is that of Major and Riedinger who identified five categories of behavioral heuristics where fraud can be found or take place [29].

- **Financial:** The flow of dollars

- **Medical Logic:** Whether a medical situation would normally happen
- **Abuse:** Frequency of treatments
- **Logistics:** The place, time and sequence of activities
- **Identification:** How providers present themselves to the insurer

The metric design can be classified using these five behavioral heuristics. It would be desirable to find metrics for detection of anomaly on all five of behavioral fraud types, although we have to note that some are harder to design metrics for.

To conclude, in order to design a generic method for predictors of fraud detection we need to take into account:

- Health insurance is a large, complex environment. The detection of fraud requires experts with domain knowledge and is for a large part dependent on inside recognition of fraud that will give leads to fraud expert or evolve into qui tam cases.
- Fraud investigation spreads over multiple departments and organizations, each of which will target on a different geographic level. Resources are not abundant and obscurity of fraud remains to be an issue.
- Different fraud schemes exists of which each can be viewed from different perspectives. The perspectives enable us to determine the usefulness of finding fraud at each level or perspective and oversee the coverage of all kinds of fraud within the health insurance branches.
- The health insurance area is of such size and complexity that division of care areas for fraud detection is required. Comparisons will only be feasible when we would deal with a quite homogeneous group. Because regulations and reimbursement are dealt stately, making cross state comparisons are bound by limitations that should be set by fraud experts.

- Fraud is an evolving subject. The play between fraud investigators and criminals will shift fraud from areas, as well between schemes. The lesson here is to search the fraud that will be committed tomorrow while monitor the known.

3.3 Analyses & Applicable Detection Techniques

Outlier technology, our predictors for fraud detection can be used to identify high-risk fraud candidates in business or the public sector. While outlier detection and classification techniques are used on the individual metrics, the metrics are based on heuristics of fraudulent behavior and therefore form the predictors for detecting fraudulent activity. The ability to apply data mining techniques to the health care fraud domain is dependent on the individual measurement semantics, the assumed relations between the behavior of a metric and fraud, and on the availability, size, types and forms of claim registrations. In order to choose the right metrics and and associated mining techniques one should take into account to deal with a set of data characteristics of health insurance claims:

- **Data volume:** Over billions of claims in the warehouse. Calculations are subject to severe computational power in order to provide data mining possibilities.
- **Data complexity:** Due to the size of health care, diversity of services and providers, the medical knowledge and the jargon necessary to understand claims.
- **Data diversity:** Each state has its own registration format for Medicaid claims.
- **Data reliability:** Limited data Reliability, claim registration errors occur, willfully or unwillingly.
- **Data labels:** The percentage of fraud constitutes a very small percentage of the total number of transactions and has therefore fundamental limits for the classifier performance named class imbalance, likely to rule out the supervised techniques [7].

Keeping these characteristics in mind, a closer look was taken to the types of data mining techniques that could be applied. Profiling technologies are algorithms or mathematical techniques that will characterize entities to discover patterns or correlations in large quantities of data, aggregated in databases. The notion of profiling practices is not just about the construction of profiles, but also concerns the application of group profiles to individuals [13]. Profiling is a form of supervised data mining and possible when we know which providers are fraudulent, in order to search for common characteristics for those who were fraudulent. A fraud investigation database might enable provider profiling but brings concerns on privacy, security that limits the use exchanging this data. To begin with, only limited set of fraud cases for each of the homogeneous groups are available, which makes it difficult to construct reliable prior sets. Furthermore, fraud is changing by its nature and therefore as provider behavior will evolve, both because of domain changes and fraud scheme development, the created fraud model have to be adjusted [11]. One of the fundamental assumptions of the classification paradigm is that the various distributions involved will not change over time in order to classify correctly. In the insurance fraud however, they do. The criminals play a hide and seek game to bypass detection systems, which will result in the population drift [20]. At last unrevealed fraud can never be classified as non-fraudulent as long as the contrary not has been proven. The combination of the small to absent training set and the population drift results that supervised technologies are not considered a solution fit.

Supervised data mining seem to have drawbacks that makes it in most cases impossible to achieve big success with, therfore focus is on the unsupervised techniques. In unsupervised methods, there is no prior class label of legitimate or fraudulent behavior required [27]. Within unsupervised data mining, different techniques could be used for fraud detection. The chosen techniques however, are all a form of outlying or anomalous behaviour detection, categorized as outlier analysis techniques. Techniques like principal components a and association ruling were left out because the outlier technique was referred as to be the most practical. Association ruling

could be interesting though to find common billing combinations, which could reveal for example unbundling schemes. We listed four most promising outlier techniques below which we were used in the case study:

- **Peer Group Analysis** is a technique for monitoring behavior over time in data mining situations [7]. In particular, the technique detects individual objects that begin to behave in a way distinct from objects to which they had previously been similar. Example would be a larger growing claim amount than those of other entities. In such problem settings, the assumption of temporal continuity plays a critical role in defining and determining outliers [1]. In time-series, a strong relationship exists between points in time, whilst multi-dimensional data has a weaker relationship and refers to aggregated trends. Medical claim data mostly relies on the second definition.

- **Break Point Analysis**, unlike peer group analysis, operates on the single entity level. A break point is an observation where anomalous behavior for a particular entity is detected based on its own behavior. Example is a steep change percentage in claim amounts. Noted should be that because we deal with the multi-dimensional data steep changes may as well occur on regular basis based on statistical deviations [1].

- **Cluster Analysis** goes to additional lengths further by forming groups of entities with the same behavior, those objects that are more similar to group objects than others. Clustering is also known as segmentation or partioning regarded a variant of unsupervised classification [50]. Based on the clustering algorithm an entity will receive its label of a clustered group, which could be highly dependent on the algorithm used. In terms of fraud detection we look at unusual or outlying clusters that cannot be explained by "normal" claim behavior.

- **Single Anomalies or Outliers** are patterns in data that do not conform to a well-defined notion of normal behavior [9]. In a data set mostly the gross

of items can formerly be grouped within regions where items have similar behavior on their dimensions, although some items differ significantly from those normal regions. These items are considered outliers, based on deviations of the group characteristics. Beside individual outlying observations, outlying group formations can exist as well.

To understand the development of our proposed metrics and analyses, it requires some background on statistics and working with outlier techniques. In statistics, an outlier is an observation that is numerically distant from the rest of the data [18]. Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior and are often also referred as outliers, discordant observations, exceptions, aberrations, surprises, peculiarities or contaminants in different application domains [9]. Outlier detection techniques are developed over the last decades and each of them has characteristics that will match certain purposes in respect to the data spread. Therefore there is no single correct method to select, previously from seeing the data characteristics, to detect the right outlying providers as much as possible. In fact, this rather is a process of learning by evaluating of results. In order to bootstrap from this problem, arguably it is likely to adopt a mix of algorithms. Each of those are assigned to metrics with data scatter that matches the strengths of these outlying detection methods and monitored for their outcomes. Outlier detection techniques can be divided mainly between univariate methods and multivariate methods. The difference is found in the number of dimensions; Uni-variate methods deal with only the first dimension, multivariate analyses will consist of at least two dimensions. Another taxonomy classification is the difference between parametric (statistical) methods and non-parametric methods (model-free) [58]. Parametric methods assume there is an underlying statistical model for the distribution of the data, or the model is an approach to predict the distribution [6] [19]. Distance based methods are an example of this non-parametric outlier detection methods. Another class of outlier detection techniques is those of outlier clustering: small clusters are classified as outlying clusters [33]. Clustering can be particular interesting when we would expect sets of providers that are us-

ing the same fraud scheme and will therefore reveal similar outlying group behavior compared to the rest. Interest exist in clustering when multiple time-flexible centers could be identified in a data set and there is a need to search for the outliers based on the centers those clusters. The underlying assumption would be that usually all data instances belong to a cluster in the data, while only a few anomalies will not belong to any cluster. Spatial outliers are a class of related methods that search for extreme observations or local instabilities concerning neighboring values such as Voronoi neighborhood formulations, but might have disadvantages in significance [44]. In statistics outliers are often indicative, either of measurement error or because of population heavy-tailed distribution. For health insurance data we experience that the distribution is usually not completely normally distributed since it is influenced by many factors who are frequently unknown. Because mostly the distributions are non-normal, there is a need in many of our metrics to adjust the parameters of outlier detection algorithms or detect an underlying model to set outlying behavior. In most cases we do not have such an underlying model, or it is difficult to establish as of the changing nature of health insurance. Therefore the general approach is to seek for for non-parametric outlier detection techniques to apply. In the case we dealt with both fraud metrics that are uni-variate as multivariate.

Although the methods might sound promising, it should be taken into account that these statistical analyses are no guarantee for correct fraud labeling, but solely indicators for initiating fraud investigations. In the evaluation of metrics, for the short term, false positives rates on initiated investigations can be measured and, for the long term we could measure false positive rates on successful prosecutions. The detection algorithms may be optimized iteratively.

Chapter 4

Metrics, Capturing Provider Behavior

4.1 Provider Metrics, definition and types

Metrics or more specific, provider metrics are used as features in data mining algorithms to compare providers with each other, or with themselves in a historical dimension. A metric can refer to an aggregate data attribute of an provider in a claim record set, such as the amount of an certain procedure code, as it can refer to an calculated value by means of proportions such as the proportion between low and high cost treatments serviced by a provider. In order to avoid confusion we defined the term metric as follows, within this context of this research:

A **metric** is a derived aggregate, or calculated relation between data -features, -attributes or -measurements that characterize the behavior of an entity over a given time period enabling comparative analysis both in time, between peers or among other dimensions.

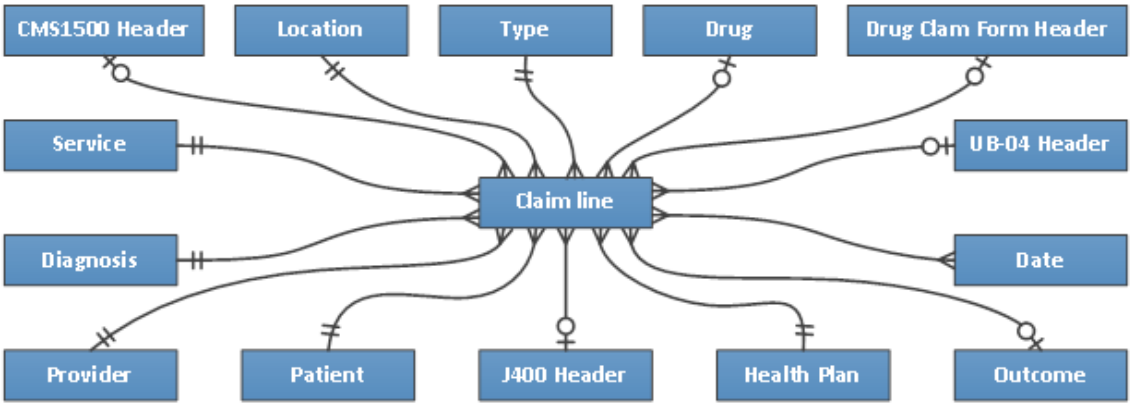
Metrics, by themselves, are loose indicators as they lack one to one relationships to fraud. Assumed is that entities that commit fraud will show aberrant behavior on multiple metrics while honest providers do not. Therefore we use the indicators to vi-

sualize the provider behavior which will help in the process of fraud detection. While some of the metrics are quite generic and may be applied to multiple domains, such as the average of claim or service rates, other metrics can only be used within specific areas. It are usually the metrics designed to investigate specific fraud schemes that tend to be more domain specific while the financial and aggregates of claims counts are less domain specific. An example is the metric that looks at the proportion between 90 minutes versus 60 or 30 minutes physician treatments, which is specifically looking for an up-coding fraud scheme. This analysis can only be applied on areas where we do have service codes that differ in time spend on a patient. This is for instance absent in the pharmacy area were the focus is on prescriptions instead of services. Another metric is the amounts of prescribed drugs per doctor, that searches for extreme prescription rates indicating a phantom billing scheme. Such an analysis could only be applied in the areas were doctors are licensed to prescribe drugs. Thus, metrics can mainly be divided in two categories, of which both can and should be adapted in the search for fraud. We define:

- **General metrics:** Metrics to indicate fraud that are applicable independent of the services or products type a provider can claim.
- **Domain specific metrics:** Metrics to indicate fraud that are applicable only to certain providers based on the services or products they can claim.

Medical claim records contain the performed medical procedures, patient diagnoses, charges, payment details and the relationship between beneficiary and providers. This information is the main source to design metrics from. While in theory there is much more provider behavior to be measured, there is a dependency on the information provided from the transaction records in Medical Insurance Systems, the referential data that is available and can legally be used, and at last, the sample size of the transactions and enrolled providers in order to make significant statistical analysis. Although there might be many ways to design, create or gather metrics, it

Figure 4-1: Medicaid Multi-dimensional Data Model [52]



is expected the most effective metrics could be designed by learning from provider behavior in earlier found fraud cases.

4.2 Medicaid Data Model and Referential Data

This research build upon the research of Thornton et al. and follows the proposed multi-dimensional data model developed for Medicaid claim data which is presented in Figure 4-1. The metrics presented further on, were build on the data dimensions the model provided. Extending the design with metrics that require the use of referential data is possible, but out of scope.

The conglomeration of medical claim records with referential data may extend the set of possible metrics and could improve the detection of fraudulent providers by having more behavioral data. However, there is change it may also decrease the level of reliability and increase complexity. Good arguments are required to decide including reference data. Sources like death records, medical license databases or DEA pricing tables should support the rule based anomaly detection. For behavioral predictors, by example loan credibility scores, tax payments information, patient social network analysis or NICS criminal records could be used. Although technically feasible, legal and ethical aspects issues will arise. Data couplings entail complexities

by the separation of data ownership and data constrictions based on the licenses that may prohibit or limit of data exchange with third parties. It may also result in ethical risk concerns by the public, enrolled beneficiaries or providers in Medicaid. Size and market domination of state programs play an important role, especially concerning the privacy degradation effect many people will have to take in, because many of them are in this program because they didn't had real alternatives at private insurance companies. Second, Data duplication and spread has always impact on the risk of data exposure, which is especially in case of medical data, a severe security risk. Thus, referential data may enrich the set of metrics, although not in all cases is the best choice.

4.3 Metric Identification

Metrics can be derived and designed in multiple ways, through case analysis, by literature review, by study of attributes in the data model or by participation with business within the same field. Although case study may be an instrument that helps to create a set of metrics, evaluation of the metrics by means of experts and flagging results is no frivolous luxury; it is an absolute necessity. The set of metrics chose for our experiments consist of case identified metrics analyzed at the FBI news blog [38], as well as an large private lawyer office dealing with Medicare and Medicaid fraud cases operating in most U.S. states and some of the metric found in literature, mainly by [40] [31] [32] [46] [51]. To understand the process of fraud metric extraction we illustrated two examples of fraud cases that leaded to design of universal metrics.

First example is a recent fraud case in New Jersey, where a physician and owner of a home-based physician services for seniors business pleaded guilty for charging lengthy visits to elderly patients that they did not receive [54]. The physician in dispute received at least half a million dollars in criminal and was eventually detected because it became the highest billing home care provider among more than 24.000 doctors in New Jersey from January 1, 2008 through October 14, 2011. The inten-

tionally over billing for services, also known as up-coding is a typical behavior that can be detected using metrics. Derived from this case was that he used lengthy visits. A metric that compares peers on the proportion of each visit length could identify such a provider. The assumption is that criminal providers because of their up-coding behavior will have a higher proportion of lengthy patient visits claimed.

Second example is a fraud case on conspiracy in Texas, where a doctor, who owned a community medical center, falsely represented office visits and diagnostic test that were medically unnecessary from February 2010 until February 2011 [55]. In exchange for submitting themselves to diagnostic tests, patients at the clinic were prescribed controlled substances, which gave the doctor confidence that the patient would return for follow-up visits. The combination of fictitious symptoms gave the doctor not only reason to prescribe the narcotics, but covered him as well for ordering more tests. The indication for this type of fraud however can be found in the referral rate. Patients are rarely referred to specialist because this would reveal that they would be healthy and could start raise questions, beside that the doctor now should need to find new patients for his fraud scheme. It might also be found in the amount of specific tests he would prescribe in general to this patients, meaning having a higher rate on out-sourced lab works. A third way he would be remarked is on its returning customers. Although every provider would have returning customers, when a large proportion of clients return too often, many times per year, this would seem remarkable at least. A provider could be a specialist or a phenomenal diagnostician that would only treat the real sick or rare cases, but in general it would a bit suspicious.

Metric identification is a complex task were is a need of knowledge of both the health care domain, as of statistical theory. In the designing process of metrics, it requires more than analysis of fraud cases to find fraud indicators. A group of outliers will normally consist of some outliers based on statistical deviation, just by chance, which cannot be filtered within a single metric. Only when fraudulent providers will take a more deviant position in the group of outliers, normal providers will probably

shift to the non-outlying group, leaving the bad guys separated. There are however always providers that will be classified as fraudulent, although they aren't, probably because their practice actually differs to much from the homogeneous groups. Filtering this non-fraudulent providers or replacing them to other groups could be done, but is difficult without full understanding of the domain.

Although the initial set of behavioral metrics was quite large, and consisted of over 100 metrics, it was refined to fifteen that could be applied to the dental domain and were feasible for implementation within our research case constrains. The set of metrics does not necessarily has to be really large, in contrary, data mining prescribes 25 to maximal 30 features or item sets are used as predictors as of exponential computation expenses. Secondly, if hundreds of metrics will be designed, we increase the absolute amount of outliers as well, which eventually will result in that each providers will have outlying behavior in some of the metrics because of the statistical change. The finite set could be found through cycles of evaluation by changing over metrics. Metric identification is dependent on fraud experts and is iterative process to find a satisfying set of metrics that works effectively.

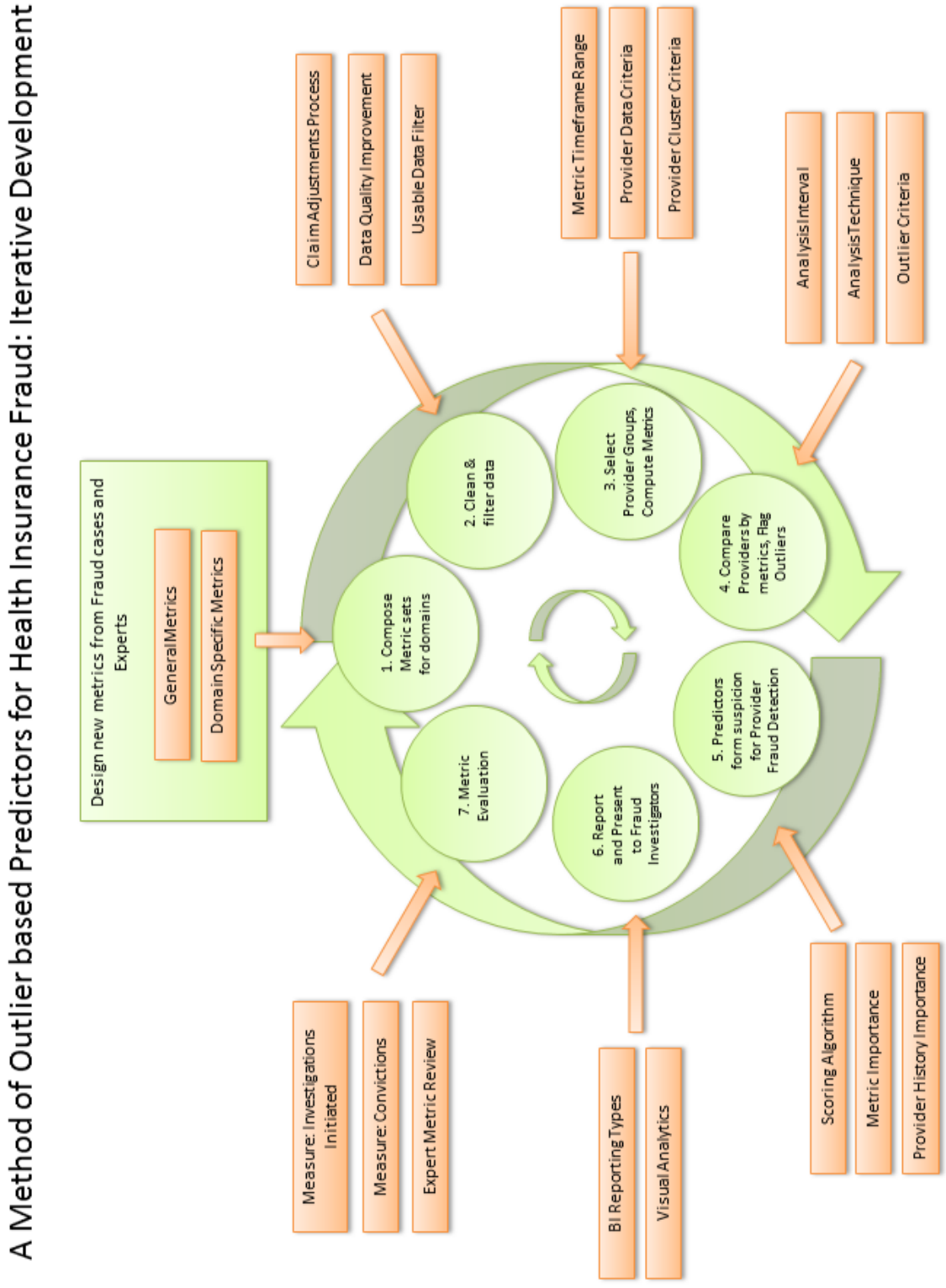
Chapter 5

Method for Predictors of Health Insurance Fraud Detection

5.1 The Method

As the initial steps for designing the metrics were described in the previous section we may continue to the process of implementing the metrics and going through the life cycle of development, testing and iterative improvement. The Method for Predictors of Health Insurance Fraud Detection which is presented in Figure 5-1 shows the stages of this cycle and the current aspects involved in those stages. Seven phases will be iterated in the method, which are in depth described in each of seven the subsections:

Figure 5-1: Method for Predictors of Health Insurance Fraud Detection



5.2 Seven Iterative Phases

5.2.1 Compose Metric sets for Domains

As described in the previous chapter, metrics are derived or calculated data -feature, -attribute or -measurement that characterizes the behavior of an entity over a given time period enabling comparative behavioral analysis using data mining algorithms. In the first iteration, we expect these metrics to be inferred from remarkable provider behavior supported by fraud cases and developed in cooperation with fraud experts. For each of the health insurance domains, analysis should be performed, fraud metrics have to be developed, meeting both domain specific as general metrics. Example of domains that could be taken are dental, physicians, home health or lab and x-ray. Some of them can even further be split to sub-domains to differentiate between specialists. Health insurance counts over 40 of such domains and may vary within the Medicaid programs. In the subsequent cycles, the composition of the metrics will consist of the creation of the latest designed metrics, to update existing metrics that need slightly different configurations, adjustments on confidence level, to optimize their hit rates, or the deletion of metrics that seem to counteract in finding fraud by only marking honest providers. Once agreed on a defined set of metrics supported for potential to predict fraud detections, the cycle is started by going through the phases from preparing data, calculating the metrics, compare the providers, analyze and evaluate results.

5.2.2 Clean & Filter Data

The Second phase concerns selecting a workable set of data for the analysis. This consist of two tasks, minimize measurement uncertainties by cleaning out the data set, second select only the relevant data of those providers to be analyzed.

The first task relates to the problem of data quality, which has to be estimated in

order to determine the precision of your provider behavior computations. Where data quality may be reduced by multiple influences, three main concerns are addressed concerning the health insurance data. First is the problem of merging multiple databases of information about common entities is frequently encountered in Knowledge Discovery in Databases (KDD) and decision support systems (DSS) in large commercial and government organizations [22]. Second is the problem of entered data quality. Health insurance data is subject to loss of quality in various ways. At first, data entry is done by hand, which is shown to inaccurate in about 4.4% in cases on personal information, and even higher percentages when abstracting data [10]. Third is the use of accurate data. Claims are often wrongly submitted and adjusted afterwards. Reasons are for example mistakes made in the process of creating the more complex claims that consist of multiple services, or services prescribed, billed, but never consumed [5]. These claims as far as possible should be removed. Data cleansing, data cleaning or scrubbing, is the quite common practice to improve the quality of the data and is highly suggested before performing analysis. Data cleansing will process the data in order to detect and correct (or remove) corrupt or inaccurate records from the record set, table, or database.

After cleaning, filtering is required, the task of selecting only that data which can be used for analysis. All data containing missing values causing the inability to calculate metrics should be removed. Claims that are voided from the system will be filtered out from the data set used for analysis. The result set of claim transaction data should meet as far as possible the ISO 8000 data quality criteria, before continuing the analysis.

5.2.3 Select Provider Groups, Compute Metrics

A common problem in data mining is to compare apples to oranges. Providers should be similar, in such sense that it is meaningful to compare them on certain behavior. The principal problem is that the more homogeneous a providers group is, comparison

may be better through, however the sample size of providers will decrease equivalent. Three questions arise: First what would be the minimum data criteria of a provider in order to perform a predictive metric. For example a provider with only 2 claims per month shouldn't be candidate for comparisons, because there is not much to compare. Second, what is the sample size that is acceptable to do a analysis, in order to classify outliers with enough certainty. If a group only consist of 5 providers, comparison results will not be very trustworthy. This is the problem of setting the right Provider Data Criteria. Third issue is selecting the characteristics criteria of a provider to fit within the provider group to be similar enough to its group members. Apart from operating within the same domain, or sub-domain in health care there are other provider characteristics that may influence the analysis, such as the provider size. Practices that claim 10 times the number of claims may be different in behavior as well because of other equipment, multiple practices or costs associated. These characteristics have to be considered before analysis. If a cluster analysis is done to detect such differences, it are the Provider Cluster Criteria that would identify the different groups. Using density or distance based clustering algorithms such as k-means or pam groups can be separated and only compared to their cluster members.

To apply metrics in analysis, calculations of those are performed and stored. The data time-frame over which the metric is calculated has to be defined. Our approach is to take a snapshot in time of provider behavior, that is calculated over the preceding time, which is captured in a time frame of which the length will be defined per metric. In general, this means taking a snapshot on first of July 2013, with a time frame length of 6 month, the metric is calculated over all the provider claims with servicing dates between first of January 2013 and first of July 2013. If the interval is bimonthly, the next analysis takes data from March 2013 until September 2013.

5.2.4 Compare Providers by Metric, Flag Outliers

The analysis interval should be defined, which basically means to define the frequency to compute the metrics and perform the analysis. A real-time or daily interval is somehow optimistic, because of the size of data and the computational resources it would require. A reasonable approach would be to calculate the metrics and performing the analyses on a monthly, quarterly, or yearly basis. Requirements on the computational resources as well as on the analysis of fraud experts should be estimated. On the other hand, consider if all the comparisons should be done at all, or if there is only interest in the largest problem areas and can analysis of other areas be passed over.

Second is to chose the right analysis technique(s) and the outlier detection method for each of the metrics. Examples of analysis techniques used in our case study were uni-variate analysis, multivariate analysis, time-series analysis and box-plot analysis. Outlier detection methods used were deviation from regression model, deviation clusters, single deviations from clusters, trend deviations, and peak deviations. The deviations in these methods made use of two types. First type is non-parametric, having no underlying statiscal distribution. Boxplots use inter quartile ranges to detect outliers. The second type is parametric, and makes use of Gaussian mixture models to find outliers, were no strong requirements were set to find the perfect model of the data distribution , rather a smart fit.

One of the analysis made use of a boxplot, which is composed of three quartiles, as can be seen in Figure 5-2. The first quartile (Q_1) is defined as the middle number between the smallest number and the median of the data set. The second quartile (Q_2) is the median of the data. The third quartile (Q_3) is the middle value between the median and the highest value of the data set. The inter quartile range is the difference between Q_1 - Q_3. Based on the inter quartile range, the upper and lower fences, also referred sometimes as the whiskers, may be calculated. P was parameterized as a variable enabling us tho define the outlier criteria, which usually is set to

1.5, however in our case study higher.

$$Upperfence = Q_3 + P \cdot IQR$$
$$Lowerfence = Q_1 - P \cdot IQR$$

The other analysis methods made use of Gaussian mixture models. Although these might not always perfectly fit the data spread, parameters of the model may skew the model slightly to get reasonable results. Building a perfect model to minimize the misfit is over complicated for this purpose. First because it is time intensive to create it, second as the data spread may change over time. Argued is it shouldn't made more complex than necessary, following the heuristic theory of Occam's Razor, who argues that one should proceed to simpler theories until simplicity can be traded for greater explanatory power. The z-value test is used in many of the analyses, using the parameterized Z-score as the criteria for anomaly. In a set of data were we have observations denoted by $X_1..X_n$ a mean of μ and standard deviation of σ. The z-value for data points may be calculated using the formula of the z-test:

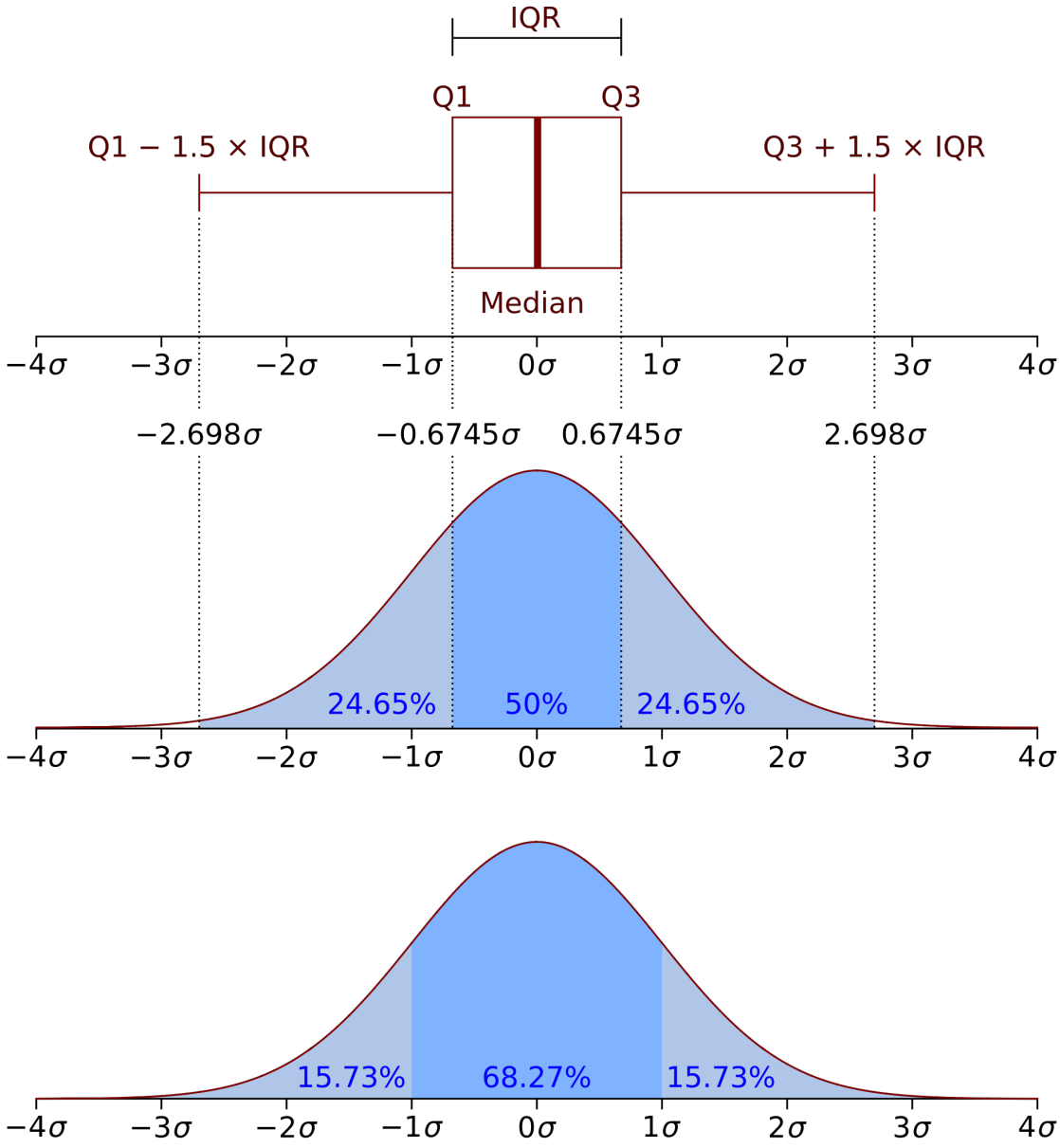
$$Z_i = \frac{|X_i - \mu|}{\sigma}$$

For each significance level, the Z-test has a single critical value (for example, 1.96 for 5% two tailed), which we will use as our outlier criteria. Usually in statistics people are in search of the 2,5%, 1% or 0.5% of data in the tails of the distribution. In general we made use of z-score of 2.33, respectively 1 out of 100, 1% in a two-tailed test, 0.5% in a one-tailed test. Outliers thus may be found using both by parametric models as well as by non-parametric models.

5.2.5 Predictors form suspicion for Provider Fraud Detection

An important aspect is how to report an anomaly and how to relate it to indication of fraud. Chandola defines scoring techniques as procedure to assign an anomaly score to each instance in the test data depending on the degree to which that instance is

Figure 5-2: Box-plot and probability density function of a Normal N(0,σ^2) Population, derived of [57]



considered an anomaly [9]. Our scoring approach is a little different. Once an outlier criteria has been exceeded by a provider we raise flag for the provider in that period. A flag is the identification of an anomaly detected by the data mining algorithm, better known as the outlier. Scoring is the formula of the individual predictors for fraud detection based on the individual results of each of the outliers in provider analysis. The proposed scoring is a way of stacking suspicion. A single provider analytic will mark some honest providers. The assumption of using scoring to find the most interesting provides to investigate for fraud is that it will find those providers that are frequent outliers, in a timely matter, as well as on multiple predictors. The methodology of using outlier criteria, or flags instead of an outlierness score as used by Chandola has advantages as well as some drawbacks. On one hand, using anomaly scoring, entities with only a few high scoring analyses will be compensated by lower scores on all others. Our methodology might leave out providers that score high on all predictors, however just below the thresholds.

Suggested is to use a formula that makes use of valued flag predictors. It sums the flags of the current analysis, and multiplies each of the metrics with its defined importance factor. In the formula, S is the providers score for potential of fraud detection, f stands for the flag if a provider is detected as an outlier in each of the individual metrics, m is the importance given to a flag and n is the number of metrics used in the predictive model.

$$\mathrm{S} = \sum_{i=1}^{n} \frac{f_i \cdot m_i}{n}$$

It is suggested, however not adopted in our approach to include history of flags of previous analyses. If providers are structural scoring flags, this could be emphasized within the formula. This may be interesting if long term data sets are available.

$$\mathrm{S} = \sum_{i=1}^{n} (\frac{(1-c) \cdot f_i \cdot m_i}{n} + \sum_{j=1}^{h} \frac{1}{h} c \cdot f_{r-j} \cdot m_n)$$

s = Provider score for potential fraud detection

n = total number of metrics

f = boolean outlier flag

m = metric importance score

r = analysis data range starting point

h = number of previous analysis taken as history

c = constant for proportion between current analysis and historical analysis

5.2.6 Report and Present to Fraud Investigators

A successful reporting platform implementation in a business intelligence environment requires input from both business as end users and IT professionals. The simple fact what in failures mostly is experiences is that if users consider the reporting layer as the total of the data warehouse, and if don't like the reporting layer, the application in total will be a waste because it will be ignored. In health insurance, only a small group of fraud investigators represents our business end users. In order to have a successful system, the investigators should be involved in the design of the analytic reports, considering the reporting style and visual analyses. Reporting styles or types concern the characteristics of a report and how a report is delivered to the end user. Current reporting types have been matured within the product industry and can be found in most product organizations. A list of current applied reporting forms are:

1. **Dashboards:** Contain high-level, Strategic data with comparisons, performance indicators in order to make instantaneous and informed decisions at a glance
2. **Interactive, multidimensional OLAP reports:** Provides information on different levels through dynamic drill-down, slicing, dicing and filtering. Analyst go through the levels to find the cause of anomalies.
3. **Ad hoc Reports:** Simple reports created by the end users on demand

4. **Static Reports:** Fixed, Subject oriented, reports precisely defined. Generated either on request by an end user or refreshed periodically from a scheduler.

5. **Technical Reports:** Reports generated to fulfill specific needs. Statistics, meta data, performance, quality, etc.

6. **Writeback Reports:** Interactive reports directly linked to the Data Warehouse which allow modification of the data warehouse data.

Fraud investigators need the flexibility on reporting, there is no specific pre-defined way reporting the data in our view. A combination of dashboards and interactive multidimensional Processing is therefore recommended. On the dashboard level we provide high level information on providers, by presenting the providers metric results, alerts on those provides that score deviant from others or on their history. Within comparative analysis, the business intelligence interactive layer enables the investigators to drill down to the root of a claim to learn how deviations might have occurred and collecting the set of claims that needs further investigations. This mantra calls for astute combinations of analytic approaches together with advanced visualization techniques; "Overview first, Filter and zoom, Details on demand" [25]. A list of alerts and their scores might be a starting point for investigators to begin their analysis. A scheduled update on the provider statistics is expected to be the chosen form of time interval. This is due to the computational interval set in phase four. Real-time updates or triggers are apart from the technical infeasibility, less expected to suit the fraud investigation process since they do not have to monitor, but do a periodical review task.

The other challenge in reporting fraud suspicion is presentation. Visualizations are not only important for usability aspects of the application, but maybe even more in gaining insight in how data behaves, in order to create metrics and compare providers. Visualization is much more effective at showing the differences between these data sets than statistics, although the data sets are synthetic. Example is for instance Anscombes Quartet that demonstrates that shapes in data might differ from the

statistical characterizations alone [3]. Interactive visual interfaces will enable the analytic reasoning process by synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data, detect the expected and discover the unexpected, providing timely, defensible, and understandable assessments and communicate assessment effectively for action [25]. The key is to select the right representations that not only enable provider comparison but in such a manner will support the analytic reasoning for fraud detection. According to Schneiderman, representations are highly dependent on the data characteristics (the seven data types) and the tasks associated (the seven data tasks) [47] and should be designed accordingly. This results in that interactive multidimensional overviews and boxplots are requested to be used for presenting outliers, and interactive drill doown table views are used for analysis.

5.2.7 Metric Evaluation

Evaluation of predictor effectiveness is required for resource allocation and time devotion in order to make decisions for analyses and further metric developments. The measuring of success however is a difficult process since fraud is established as fraud until after a litigation. Therefore measuring convictions as justification for resource allocation and input for iterative improvement is not an ideal situation. Measuring investigations initiated by fraud experts, in which not only claims are gone through, but also provider documentation is reviewed, might be more reliable. If fraud investigation initiations are chosen as evaluation statistics, we may use the formula's of precision and recall to calculate the effectiveness of the method. A downside is that fraud investigations might be systematically wrong initiated distorting the effectiveness measurements. Fraud convictions might eventually provide the contrary evidence, however we believe that fraud experts are capable of interpreting these measurements meaningfully. Thresholds, or configuration of the outlier detection algorithms influence the classification of data point as outliers. Restrictive outlier groups may minimize the number of potential fraud, while less restrained classifica-

tion lead to false positives. The trade-off may be measured in terms of precision and recall [1].

$$\text{Precision(t)} = 100 \cdot \frac{|S(t) \cap G|}{|S(t)|}$$

$$\text{Recall(t) or TPR(t)} = 100 \cdot \frac{|S(t) \cap G|}{|G|}$$

The set of providers is denoted as S(t), where t stands for the threshold, or outlier criteria. The providers classified as the true set (ground-truth) are denoted as G. Plotting the True Positive Rate (FPR(t)), or recall, and the False Positive Rate helps in finding the optimal outlier criteria. The false positive rate (FPR(t)) is the percentage of the falsely reported positives out of the ground-truth negatives and is formulated as:

$$\text{FPR(t)} = 100 \cdot \frac{|S(t) - G|}{|D - G|}$$

Chapter 6

Case Study: Medicaid Dental Providers

– Removed, Confidential –

Chapter 7

Experts Evaluation

– Removed, Confidential –

Chapter 8

Discussion

8.1 General observations

The method that was developed used different forms of outlier detection as predictors of health insurance fraud detection. As suggested by Weng [56] we agree that to detect health insurance fraud, outlier detection is the primary tool use, although should be noted that we believe that other techniques such as profiling may have its appliances in some cases as well. Scientists in [8] showed results on the use of fuzzy logic and neural network that suggested potential for more advanced techniques. Techniques as association rule mining tend to work as well [45]. Within the expert evaluation, we found that even with most simple statistical modeling, education already will be required. Therefore, we believe because of the complexity of such techniques, no systems could be developed that would meet the working criteria of our fraud experts. Forgionne reported that potential users should be involved in developing the metrics and analytical tools and we believe this will take a severe part in the road to success [17].

The studies of Ortega and those of Yang et al. showed positive results for identifying suspicion using supervised data mining modeling techniques [60] [41]. The underlying assumption is that labeled data is available, where we found that in many of the smaller health insurance domains fraud data sets are small or absent. Sec-

ond of all, as sparrow reported [49], *"You should assume the opponents are playing the counter intelligence game and studying carefully what you do, where the normal values of openness, transparency and predictability are in fact losing strategies, and where artfulness and creativity count most."*, which basically means fraud data sets have an expiration date for training purposes. Furthermore the analyses should, apart from known fraud schemes, also look further to new kinds of more sophisticated fraud.

The more similar experiments, working with behavioral heuristics as indicators for fraud, done by researchers as Major and Riedinger [29] and Ng et al. [32] and Tang et al. [51], were found partly to be true in this study. In our 14 experiments, many of the analyses identified suspicious behavior of providers, which led to questionable claims. However, the true effectiveness of the implemented strategy was difficult to measure as it is influenced by many factors such as time, size of fraudulent activity per domain and selection complexity of provider categories.

In the case study that was performed, a set of over 500 providers active in the dental domain, of which 370 were analyzed and compared. From those, 106 providers received flags, of which 35 providers received two or more. The top 50% (17) of these providers have been analyzed by drilling down to transaction level to look for relationships, patters, causes and explanations of the received flags for each of the associated providers. Five providers found were likely to be mis-classifications by reasons of size or logical explanations for their behavior, described in the case study. For nine cases we expressed great doubts by the billing procedures of the provider and recommend further investigation. In three of the occurrences, we found observational fraud, which was clearly odd, that everybody without pre-knowledge would question these providers. If we may take the recommended further analyses cases as true positives and the presumably mis-classifications as false positives, an precision of 71% can be found for this work. In addition, we found ground by fraud experts and other researchers and for the exploration of such technology for longer term and on larger scales [29]. Another important finding related to the validity of the method, was that not all of the insurance domains could be subject to outlier detection. It

requires a wide breath of codes, a subject domain where enough providers are active with assumed similar behavior. At last, the right selection criteria should be chosen for building provider groups in order to do meaningful analysis.

8.2 Limitations

As with most fraud detection problems on which unsupervised data mining techniques are applied, measuring effectiveness to validate the application or methodology remains difficult as conviction cost time. Another problem is that feedback loop of convictions or settlements are not clearly defined. There is no straightforward feedback reporting on such cases to designed systems. Although many researchers, and some of fraud investigators express their believes in the technology, were thus long term research is required to proof the technique's effectiveness above current processes and technology used, for which partly implementation of the technology is a requirement of course.

The proposed method has as limitation that it may only work in those areas were the range and spread of procedures, prices, diagnose, in general the feature set is rich enough to compare providers. Solely the financial metrics as such presented in our experiment will not have enough bases by themselves to initiate investigation, nor provide much advantage over the current decision support systems in place.

8.3 Lessons Learned

In five statements, we outline our most important learned lessons we gained in our research:

1. Although fraud can never be completely eradicated, it can be better managed in terms of detection of fraud as well as on resource allocation for audits.

2. The health insurance domain is too different and complex to take one effective approach of fighting fraud. Rather, fraud detection should consist of a mixture of techniques so each technique can focus on different types of fraud. Moreover, fraud detection is part of a bigger program of fighting fraud, abuse and waste, that should be reflected to practices that could be done in a preventive way, such as the policy limitations.

3. Translating provider suspicion to provider review in the form of an accusation or on site audit requires more proof than solely outlying behavior; It mostly requests for observational fraud or severe suspicion. Fraud investigations are costly and will only be performed once investigators have confidence to recover great amounts that would pay off such investigations.

4. In order to let the method become successful, beside the technical feasibility, fraud investigators involvement is required in development, user education for understanding and performing the analysis by fraud investigators and political alignment that would support in progression of these technologies to be adopted.

5. The proposed model proofed to work within our case study to identify fraud, although mis-classifications were found and the suspicions not always could be translated to objective fraud. The method had some limitations on generalization to some of the fraud insurance sub-domains and it was difficult to measure the real long-term effectiveness.

8.4 Future Work

In the months after the experiments we started on design and deployment of a full operational prototype for the Centers of Medicaid and Medicare. The prospect is to develop the technology for multiple states, and transfer it to different health insurance areas such as drugs prescribing, in and out patient hospitalization, gaining insight in a the method on a larger scale. The involvement of fraud investigators to develop new metrics for each of the areas is a definite requirement. One of the current outstanding

issues is the optimal selection criteria for provider group in order to get meaningful analysis, related to the exploration of automatic clustering on categorical attributes. Also we would suggest to see how the methodology will behave when to look for patterns across multiple dimensions. Research on finding the optimal criteria for outlier detection in the context of designed metrics would also be an interesting area that would be in line with this research. Al last we see the opportunities in the announced ICD-10-CM as a richer data set, which will request for research on feature qualification and metric development potential.

Chapter 9

Conclusion

This thesis, documents our experience of applying outlier detection techniques in order to solve the problem of effective fraud detection process in the health insurance industry. Our first question, which unsupervised techniques are applicable for identifying provider behavior could be answered. We found outlier detection the best candidate for fraud detection and developed a method based on this outlier technique to compare providers on behavioral characteristics. Box-plots on procedure and tooth codes within the case study seemed to provide the best results. Our second question, that searched for the identification of metrics that could be used in the analysis, was answered by the study of cases and validation of fraud experts. Although many metrics could be derived from fraud cases, we found that is was important to include fraud investigators already early in the process of metric and system design in order estimate validity of the metrics on fore hand. Our third question, that looked how the analytic detection technology should be adopted within the full program of fraud initiatives and processes was answered using the expert evaluation. We found that the technology is promising, however it constitutes within a larger whole of fraud fighting initiatives and has limitations concerning generalization to all of the health insurance domains. Together we were able to answer the main research question:

What is an effective method for predictive health insurance fraud detection that identifies suspicious provider claiming behavior using unsupervised data mining tech-

niques?

by conclude the main contributions of the thesis:

1. A method for the health insurance domain has been developed that may be used as a guideline for the iterative development of predictors for detection of fraud in many of the specialty areas of health care in order to identify fraud. The propositionalisation methodology has been shown on a small scale in dentistry domain to effectively mark outliers that may be followed by fraud investigations, which supports to quest of finding fraudulent activity. Within our top 17 result, we found to have an 71% accuracy of actual grounded suspicion that would lead to initiation of investigations.

2. A demonstration of the method in the form of a case study was shown. The study resulted in a deeper understanding of dental fraud and successful analyses that led us to questionable claim sets of potential fraudulent providers. Analyses that cross validated procedure and tooth codes between providers using box-plot outlier techniques were found to be promising, though noted that statistical knowledge will be required in order to be used by most fraud experts.

3. We learned from the fraud subject matter experts that predictors for fraud detection should be seen within a program of multiple fraud detection methodologies. Because of size, complexity and changing nature of fraud, there is no single effective method to detect fraud and therefore multiple approaches are required. Predictors are not likely to succeed as fraud classification technology, though it might explore an important role as decision supportive technology for resource allocation of fraud investigations and audits.

Appendix A

Tables

– Removed, Confidential –

Appendix B

Figures

Figure B-1: Medicaid Tooth Numbering System-Permanent Teeth.

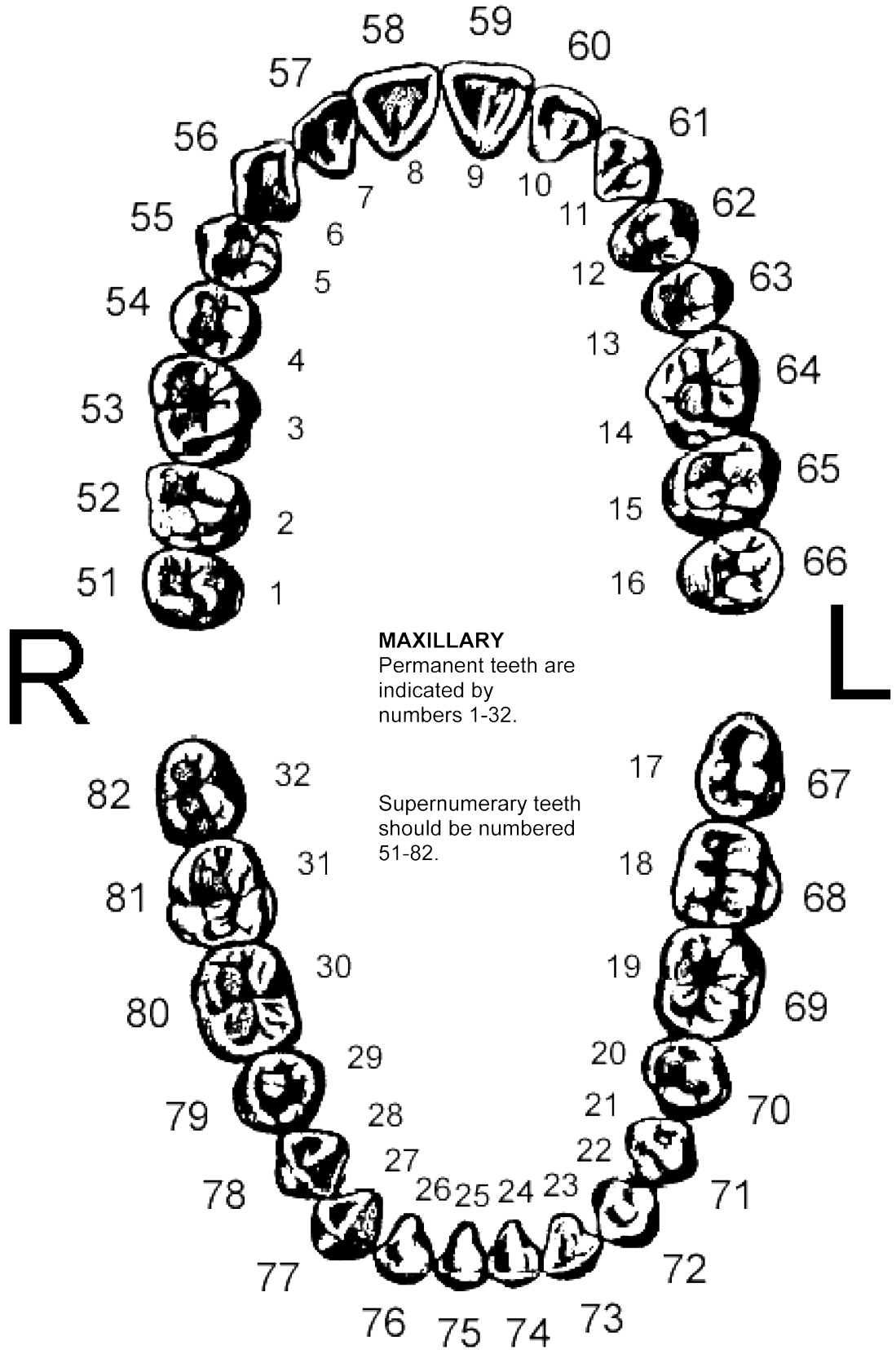
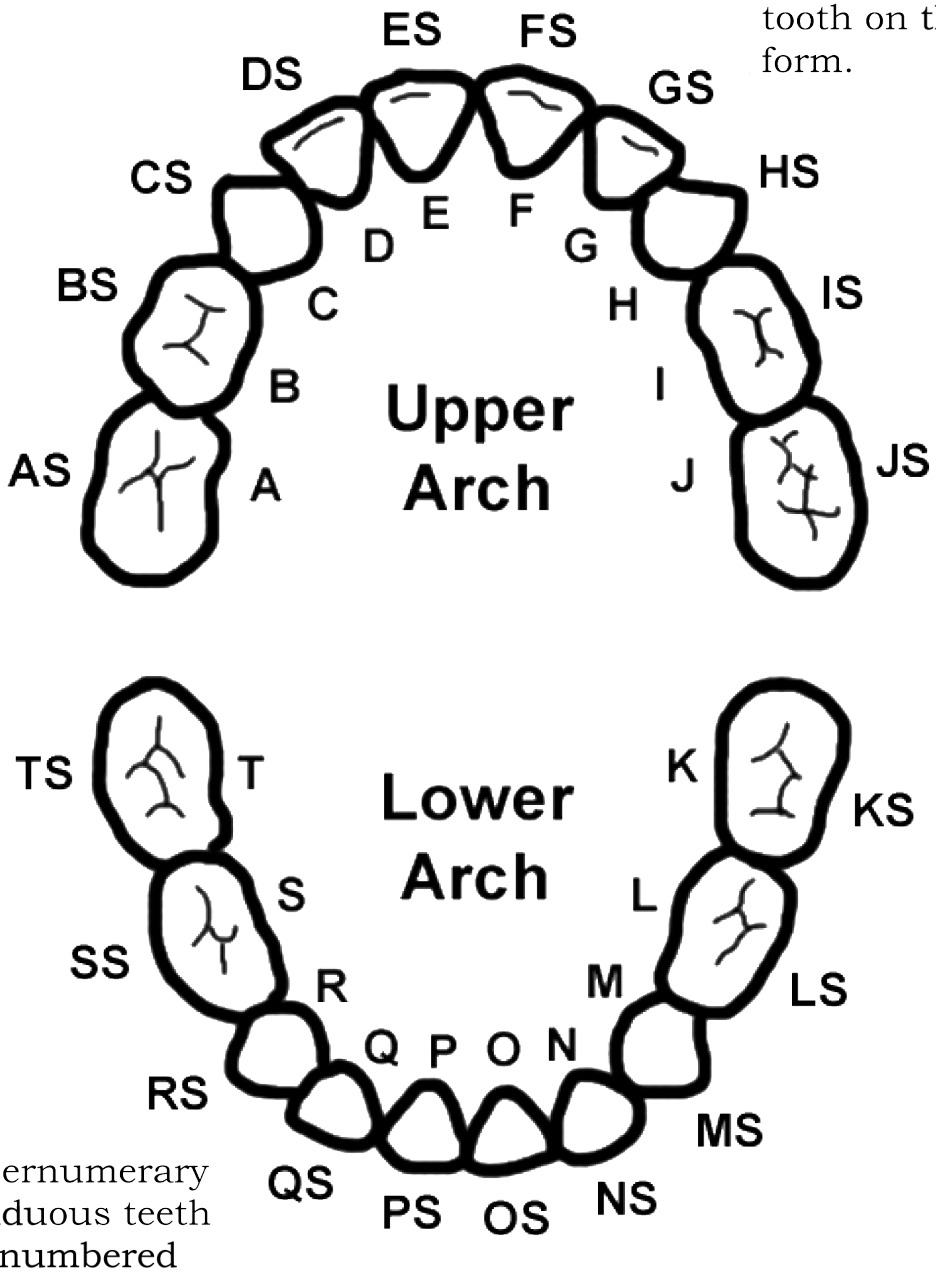
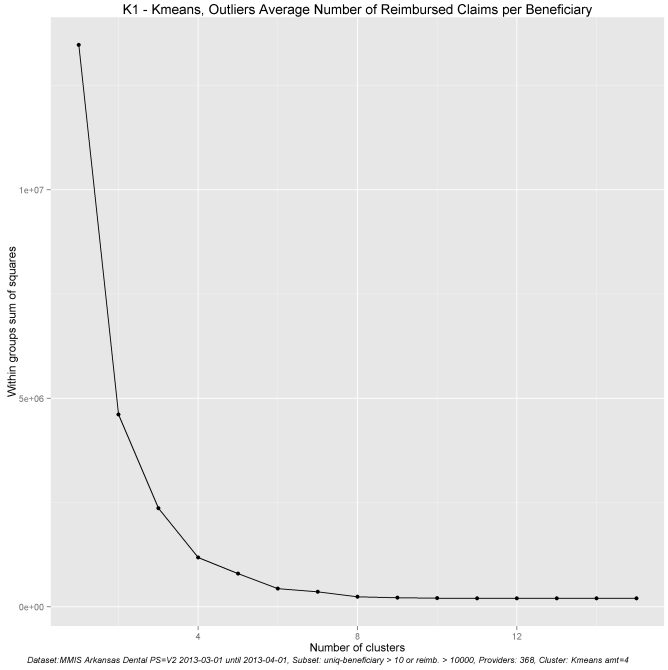


Figure B-2: Medicaid Tooth Numbering System-Deciduous Teeth

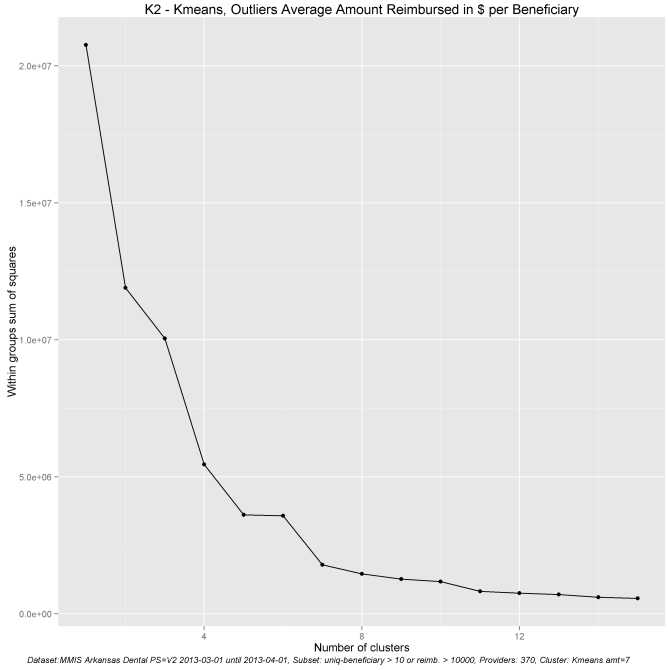
When requesting treatment on a deciduous tooth, indicate the tooth on the ADA form.



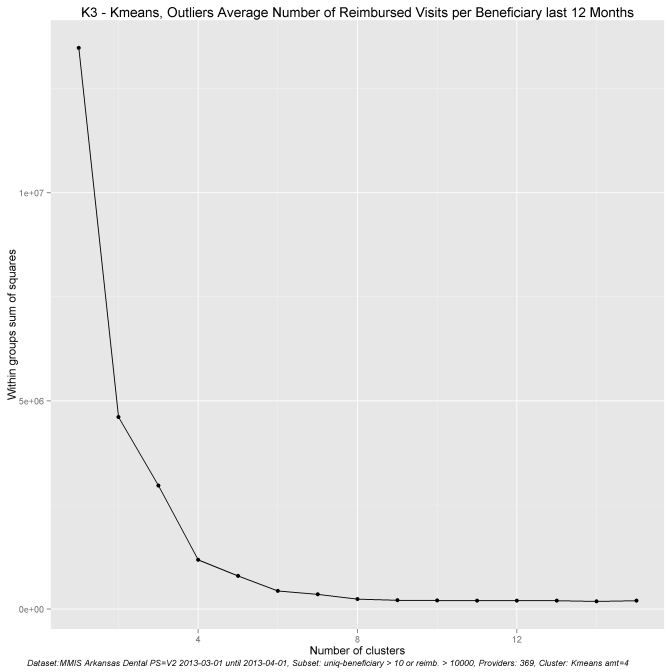
Supernumerary deciduous teeth are numbered AS, BS, CS, etc.



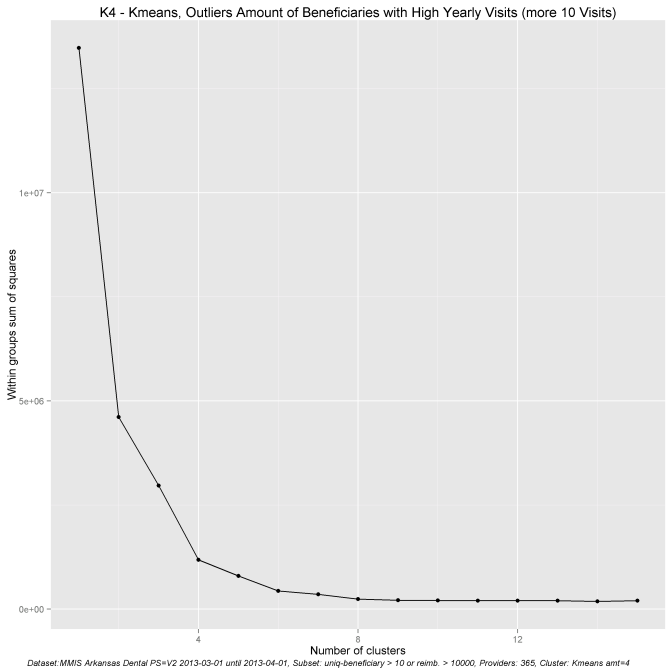
(a) K1 elbow



(b) K2 elbow

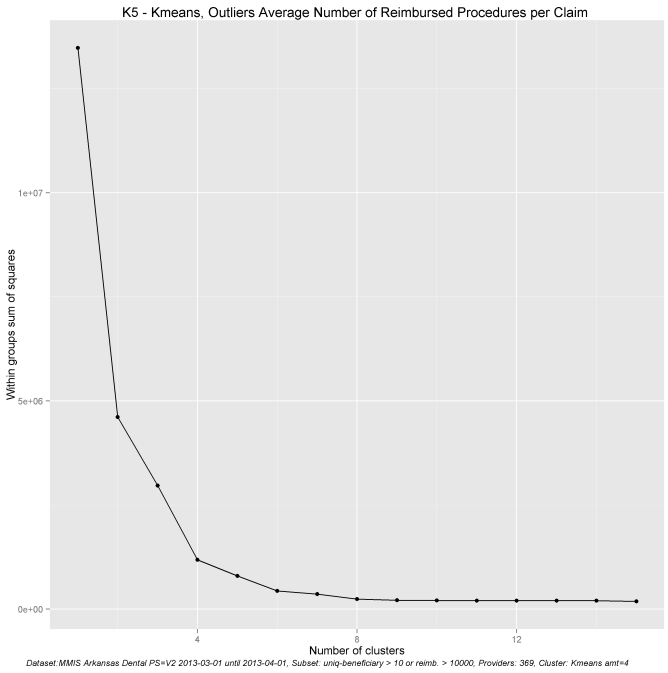


(c) K3 elbow

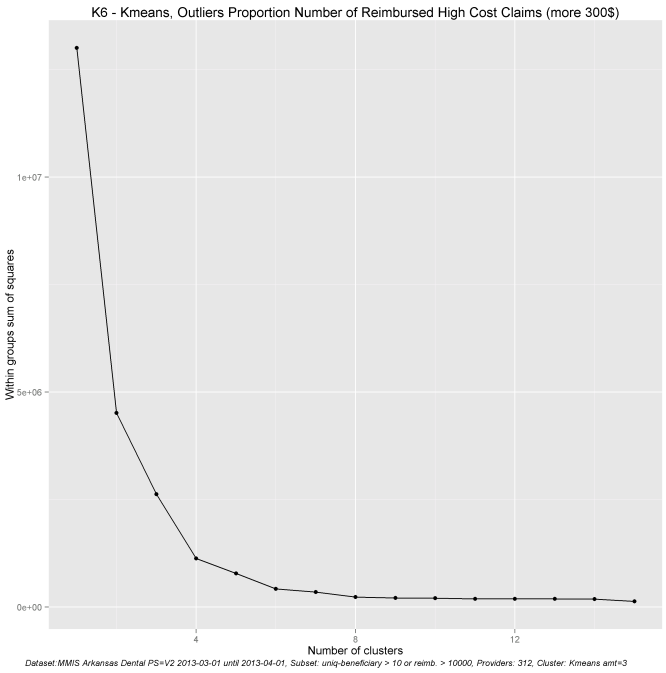


(d) K4 elbow

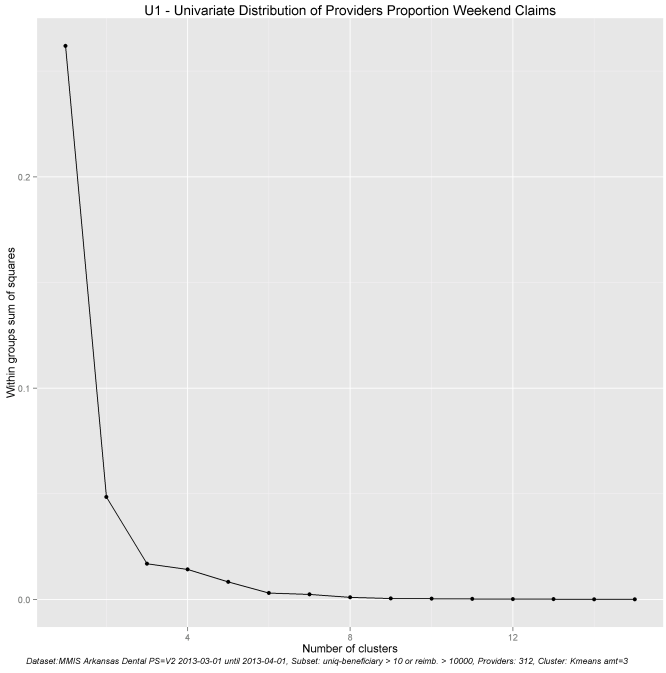
Figure B-3: Elbow Figures K1-4



(a) K5 elbow



(b) K6 elbow



(c) U1 elbow

Figure B-4: Elbow Figures K5, K6, U1

Appendix C

Interview Questions

1. What is the general view you have on the exploration of data mining technology for the use of fraud prediction? (good, bad, promising, waste of resources, too complex, not feasible etc.)

2. By means of the case results of this research, which of the metrics and/or outliers found are in your view the most promising results?

3. For which types of fraud would you think the methodology could be effective in predicting fraud, and which schemes are possible unable to detect with this kind of methodology and require other approaches.

4. What would be a general pitfall using the proposed methodology and what would be the pitfalls in designing metrics for provider comparison.

5. Where in the process of fraud detection would you make use of such technology? In which tasks could the technology support your work? In which extend would you make use of such technology/ services?

6. To measure effectiveness:

   What are your thoughts on the amount of new case you could differ from these predictive analysis (single case dental Arkansas vs. applied on whole state) (effectiveness)

In how far might it replace other fraud indication work, in such that it will provide support for your work and time savings (efficiency)

7. To which extend can the method be generalized to the entire health insurance area? Are there certain constrains in applying the methodology to certain providers, provider groups or health care areas?

8. What kind of visualization should be chosen to display the the provider metric results? Which descriptions / trainings you think are needed in order to let experts work with these kind of tools?

9. What kind of challenges do you see when we would like to adapt the general methodology of provider comparison on certain behavioral metrics concerning:

   Technological and General feasibility

   Provider Group formation (providers to cross compare on their behavior)

   Metric Identification (developing metrics for each of the provider groups)

   Resource Availability (money-wise as human-wise)

   Stakeholder/Management

   Process and Expert adaptation (human adoption)

Bibliography

[1] A. Aggarwal. *Outlier Analysis*. Springer, 2013.

[2] N. Aldrich and B. Benson. Medicare/medicaid improper payments exceed $64 billion a year, 2012. Accessed: 2013-04-18.

[3] F. J. Anscombe. Graphs in statistical analysis. *The American Statistician*, 27(1):pp. 17–21, 1973.

[4] K. D. Aral, H. A. Güvenir, İ. Sabuncuoğlu, and A. R. Akar. A prescription fraud detection model. *Comput. Methods Prog. Biomed.*, 106(1):37–46, April 2012.

[5] Medicaid Arkansas. What is a claim adjustment? https://www.medicaid.state.ar.us/InternetSolution/provider/faq/faq.aspx, 2013. Accessed: 2013-06-10.

[6] L. Ben-Gal. *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*, chapter 1, pages 1–12. Kluwer Academic Publishers, 2005.

[7] R. J. Bolton and J. H. David. Statistical fraud detection: A review. *Statistical Science*, 17:2002, 2002.

[8] P. L. Brockett, X. Xia, and R. A. Derrig. Using kohonens selforganizing feature map to uncover automobile bodily injury claims fraud. *The Journal of Risk and Insurance*, pages 245–274, 1998.

[9] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3):15:1–15:58, July 2009.

[10] C. Colin, R. Ecochard, F. Delahaye, G. Landrivon, P. Messy, E. Morgon, and Y. Matillon. Data quality in a drg-based information system. *International Journal for Quality in Health Care*, 6(3):275–280, 1994.

[11] L. Copeland, D. Edberg, A.K. Panorska, and J. Wendel. Applying business intelligence concepts to medicaid claim fraud detection. *Journal of Information Systems Applied Research*.

[12] W. H. Delone and E. R. McLean. The delone and mclean model of information systems success: A ten-year update. *J. Manage. Inf. Syst.*, 19(4):9–30, April 2003.

[13] E. Elmer. *Profiling Machines: Mapping the Personal Information Economy.*, volume 1. MIT Press, 2004.

[14] U.S. Centers for Medicare & Medicaid Services. 2012 actuarial report on the financial outlook for medicaid, 2012.

[15] U.S. Centers for Medicare & Medicaid Services. 2012 annual report ofthe boards of trustees of the federal hospital insurance and federal supplementary medical insurance trust funds, 2012.

[16] U.S. Centers for Medicare & Medicaid Services. 2012 medicare fraud & abuse: Prevention, detection, and reporting, 2012.

[17] G. A. Forgionne, A. Gangopadhyay, and M. Adya. An intelligent data mining system to detect healthcare fraud. *Healthcare Information Systems: Challenges of the New Millennium*, (1):149–168, January 2000.

[18] F. E. Grubbs. Procedures for detecting outlying observations in samples. *Technometrics*, 11:1–21, 1969.

[19] A. S. Hadi. Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54(3):pp. 761–771, 1992.

[20] D. J. Hand. Rejoinder: Classifier technology and the illusion of progress. *Statistical Science*, 21(1):pp. 30–34, 2006.

[21] J. W. Henderson. *Health Economics and Policy.* Mason, OH: South-Western Cengage Learning, 2009.

[22] M. A. Hernández and S. J. Stolfo. Real-world data is dirty: Data cleansing and the merge/purge problem. *Data Min. Knowl. Discov.*, 2(1):9–37, January 1998.

[23] A. R. Hevner. Design science in information systems research. *MIS quaterly*, 28(1):75–105, 3 2004.

[24] V. Iyengar, K. Hermiz, and R. Natarajan. Computer-aided auditing of prescription drug claims. *Health Care Management Science*, pages 1–12, 2013.

[25] D. Keim, G. Andrienko, J. Fekete, C. Grg, J. Kohlhammer, and G. Melanon. Visual analytics: Definition, process, and challenges. In Andreas Kerren, JohnT. Stasko, Jean-Daniel Fekete, and Chris North, editors, *Information Visualization*, volume 4950 of *Lecture Notes in Computer Science*, pages 154–175. Springer Berlin Heidelberg, 2008.

[26] R. Kelley. Where can $700 billion dollar in waste be cut annually from the us healthcare system? http://www.larson.house.gov/images/pdf/700billioninwaste.pdf', 2013. Accessed: 2013-04-18.

[27] N. Laleh and M. A. Azgomi. A taxonomy of frauds and fraud detection techniques. In Sushil K. Prasad, Susmi Routray, Reema Khurana, and Sartaj Sahni, editors, *ICISTM*, volume 31 of *Communications in Computer and Information Science*, pages 256–267. Springer, 2009.

[28] F. Lu and J. E. Boritz. Detecting fraud in health insurance data: Learning to model incomplete benfords law distributions. In Joo Gama, Rui Camacho, PavelB. Brazdil, AlpioMrio Jorge, and Lus Torgo, editors, *Machine Learning: ECML 2005*, volume 3720 of *Lecture Notes in Computer Science*, pages 633–640. Springer Berlin Heidelberg, 2005.

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

[30] M. Matthews. Medicare and medicaid fraud is costing taxpayers billions. http://www.forbes.com/sites/merrillmatthews/2012/05/31/medicare-and-medicaid-fraud-is-costing-taxpayers-billions/2/, 2013. Accessed: 2013-04-18.

[31] R. M. Musal. Two models to investigate medicare fraud within unsupervised databases. *Expert Syst. Appl.*, 37(12):8628–8633, December 2010.

[32] K. S. Ng, Y. Shan, D. W. Murray, A. Sutinen, B. Schwarz, D. Jeacocke, and J. Farrugia. Detecting non-compliant consumers in spatio-temporal health data: A case study from medicare australia. In Wei Fan, Wynne Hsu, Geoffrey I. Webb, Bing Liu 0001, Chengqi Zhang, Dimitrios Gunopulos, and Xindong Wu, editors, *ICDM Workshops*, pages 613–622. IEEE Computer Society, 2010.

[33] R. T. Ng and J. Han. Efficient and effective clustering methods for spatial data mining. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94:)*, pages 144–155, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc.

[34] OECD. Oecd health data october 2012, 2012.

[35] U.S. Department of Health and Human Services. Advancing the health, safety, and well-being of our people: Hhs budget prospective. http://www.hhs.gov/budget/budget-brief-fy2013.pdf, 2011. Accessed: 2013-04-18.

[36] U.S. Office of Inspector General. Medicaid fraud control units - mfcus. http://oig.hhs.gov/fraud/medicaid-fraud-control-units-mfcu/, 2013. Accessed: 2013-04-18.

[37] U.S. Federal Bureau of Investigation. Financial crime report 2010-2011. www.fbi.gov/stats-services/publications/financial-crimes-report-2, 2012. Accessed: 2013-04-18.

[38] U.S. Federal Bureau of Investigation. Fbi news blog, 2013. Accessed: 2013-04-18.

[39] U.S. House of Representatives. 31 usc 3729: False claims, 2013. Accessed: 2013-04-18.

[40] U.S. Government Accountability Office. Medicare fraud prevention: Cms has implemented a predictive analytics system, but needs to define measures to determine its effectiveness, 2012.

[41] P. A. Ortega, C. J. Figueroa, and G. A. Ruz. A medical claim fraud/abuse detection system based on data mining: A case study in chile. In Sven F. Crone, Stefan Lessmann, and Robert Stahlbock, editors, *DMIN*, pages 224–231. CSREA Press, 2006.

[42] C. Phua, V. Lee, K. Smith-Miles, and R. Gayler. A comprehensive survey of data mining-based fraud detection research. 2005.

[43] E. Scheiner. Gao estimates $44 billion in improper medicare payments. http://cnsnews.com/news/article/gao-estimates-44-billion-improper-medicare-payments, 2012. Accessed: 2013-04-18.

[44] S.S. Schiffman, M. L. Reynolds, and F. W. Young. *Introduction to multidimensional scaling*. Academic Press, Orlando [u.a.], 1981.

[45] Y. Shan, D. Jeacocke, D. W. Murray, and A. Sutinen. Mining medical specialist billing patterns for health service management. In John F. Roddick, Jiuyong Li, Peter Christen, and Paul J. Kennedy, editors, *AusDM*, volume 87 of *CRPIT*, pages 105–110. Australian Computer Society, 2008.

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

[47] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

[48] SMPresource.org. Summary of anti-fraud provisions in the affordable care act, 2013. Accessed: 2013-04-18.

[49] M. K. Sparrow. *License to Steal: How Fraud Bleeds America's Health Care System*, volume 2000. Westview press, 2000.

[50] P. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.

[51] M. Tang, B. S. U. Mendis, D. W. Murray, Y. Hu, and A. Sutinen. Unsupervised fraud detection in medicare australia. In *Proceedings of the Ninth Australasian Data Mining Conference - Volume 121*, AusDM '11, pages 103–110, Darlinghurst, Australia, Australia, 2011. Australian Computer Society, Inc.

[52] D. Thornton, R. M. Müller, P. Schoutsen, and J. van Hillegersberg. Prediction healthcare fraud in medicaid: A multidimensional data model and analysis technique for fraud detection. In *CENTERIS 2013, Conference on ENTERprise Informations Systems / HIST 2013 International Conference on Health and Social Care Information Systems and Technologies*. Elsevier, 2013.

[53] P. Travaille, R. M. Müller, D. Thornton, and J. van Hillegersberg. Electronic fraud detection in the u.s. medicaid healthcare program: Lessons learned from other industries. In Vallabh Sambamurthy and Mohan Tanniru, editors, *AMCIS*. Association for Information Systems, 2011.

[54] District of New Jersey U.S. Attorneys Office. South jersey doctor admits making half-a-million dollars in fraud scheme involving home health care for elderly patients. http://www.fbi.gov/newark/press-releases/2013/south-jersey-doctor-admits-making-half-a-million-dollars-in-fraud-scheme-involving-home-health-care-for-elderly-patients, 2013. Accessed: 2013-03-28.

[55] District of Texas U.S. Attorneys Office. Physician pleads guilty to role in health care fraud conspiracy. http://www.fbi.gov/dallas/press-releases/2013/physician-pleads-guilty-to-role-in-health-care-fraud-conspiracy, 2013. Accessed: 2013-03-01.

[56] X. Weng and J. Shen. Detecting outlier samples in multivariate time series dataset. *Knowl.-Based Syst.*, 21(8):807–812, 2008.

[57] Wikipedia. Boxplot and a probability density function (pdf) of a normal n(0,1σ^2) population. https://en.wikipedia.org/wiki/File:Boxplot_vs_PDF.svg, 2013. Accessed: 2013-10-28.

[58] G. Williams, R¿ Baxter, H. He, S. Hawkins, and L. Gu. A comparative study of rnn for outlier detection in data mining. In *in ICDM*, page 709, 2002.

[59] K. Yamanishi, J. Ichi Takeuchi, G. Williams, and P. Milne. On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. *Data Mining and Knowledge Discovery*, 8:275–300, 2004.

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