Electronic Fraud Detection in the U.S. Medicaid Healthcare Program: Lessons Learned from other Industries

Peter Travaille
University of California, San Diego
ptravaille@sdsc.edu
Dallas Thornton
University of California, San Diego
dallas@sdsc.edu

Roland M. Müller
Berlin School of Economics and Law
roland.mueller@hwr-berlin.de
Jos van Hillegersberg
University of Twente
j.vanhillegersberg@utwente.nl

ABSTRACT

It is estimated that between $600 and $850 billion annually is lost to fraud, waste, and abuse in the US healthcare system, with $125 to $175 billion of this due to fraudulent activity (Kelley 2009). Medicaid, a state-run, federally-matched government program which accounts for roughly one-quarter of all healthcare expenses in the US, has been particularly susceptible targets for fraud in recent years. With escalating overall healthcare costs, payers, especially government-run programs, must seek savings throughout the system to maintain reasonable quality of care standards. As such, the need for effective fraud detection and prevention is critical. Electronic fraud detection systems are widely used in the insurance, telecommunications, and financial sectors. What lessons can be learned from these efforts and applied to improve fraud detection in the Medicaid health care program? In this paper, we conduct a systematic literature study to analyze the applicability of existing electronic fraud detection techniques in similar industries to the US Medicaid program.

Keywords

Fraud Detection, Data Mining, Health Care, Medicaid

INTRODUCTION

Healthcare fraud in the United States is a severe problem that costs the government billions of dollars each year. Roughly one-third of all US healthcare costs are attributable to fraud, waste, and abuse (Kelley 2009). Third-party payers for healthcare services (insurance companies and government-run programs) must deal with fraudulent practitioners, organized criminal schemes, and honest providers who make unintended mistakes while billing for their legitimate services. The US Medicaid system is particularly susceptible fraud and abuse, as it is harder to exclude problematic providers and is managed separately and with limited coordination across the states. Each state has sovereignty over its program and maintains its own eligibility and benefits criterion. This makes nation-wide fraud detection and prevention initiatives more complicated.

Despite the complexity and structural challenges, strides can clearly be made through technology to improve fraud and abuse detection and prevention across the Medicaid program. What lessons can be learned from the electronic fraud detection techniques utilized in similar industries? In the insurance, telecommunications, and financial sectors—particularly the credit card industry—fraud detection is vital to sustainability and competitiveness. In general, insurance fraud and abuse are hard to discover because of asymmetric information between the insurer, beneficiary, and provider (Derrig 2002). In the Medicaid program fraud detection is the responsibility of the state and federal government, who each share the cost of the program. Thus, it is their responsibility to reduce the opportunities to commit fraud, improve the detection mechanisms inherent to the system, and impose criminal penalties that serve as a deterrent to fraudulent billing from providers and criminal enterprises.

The structure and the claims process of Medicaid are outlined in section 2. Section 3 presents a systematic literature study of the existing electronic fraud detection techniques applied in related industries. Section 4 discusses fraud schemes discovered in the past. In section 5, we discuss lessons learned from related industries and the advantages, disadvantages, and constraints of the analyzed fraud detection techniques as applied to the Medicaid program.

MEDICAID

Medicaid and Medicare are two government programs that provide medical and health-related services to specific groups of people in the United States. Although the two programs are very different, they are both funded and overseen, in part or whole, by the Centers for Medicare and Medicaid Services (CMS), a division of the U.S. Department of Health and Human Services. Medicare is a federal program which has consistent rules across the fifty states and covers almost everyone 65 years of age or older. Medicaid is a state administered program in which each state provides a unique health care program for individuals and families with low incomes and resources (CMS 2011). Each state sets its own guidelines regarding eligibility and services. With passage of the Affordable Care Act of 2010 (ACA), Medicaid eligibility extended from people with income up to 100% of the federal poverty line (FPL) to people up to 133% of the FPL, adding tens of millions of newly eligible individual. Medicaid also covers special categories of people, such as pregnant women and children (CMS 2011). In 2009, 63 million people were enrolled in Medicaid and the overall costs were $381 billion (Truffer et al. 2010). Due to the federally-expanded enrollment eligibility criteria enacted by ACA, Medicaid's expected costs rise to $587 billion by 2015, with 68.5 million enrolled (US Congressional Budget Office 2010).

The United States faces a current and serious fraud problem concerning the social health care (Sparrow 2000). The current standard detection and control systems are not designed to meet the threats of various criminal fraud types (Hyman 2001). Electronic "edits" and "audits" are built into highly automated claims processing systems which have all been designed with honest providers in mind. They are designed to catch errors and efficiently reimburse honest providers – verifying eligibility, making sure procedure codes match diagnosis, and checking that the price charged is in within bounds – not explore patterns that could flag fraudulent or abusive behaviors (Sparrow 2000). This gives providers and other people with fraudulent intentions the opportunity to easily get away with fraudulent behavior by submitting claims that simply look like they were for appropriate services (Sparrow 2000).

Claims Process

When a provider participates in Medicaid, the provider agrees to the reimbursement rates set by the state and submits claims for payment directly to the state's Medicaid agency. If the provider is not participating in Medicaid, the provider sends the patient the bill which he or she has to pay before requesting reimbursement for partial payment from Medicaid. In both scenarios, the state Medicaid agency processes the claim and sends an explanation of benefits (EOB) to the beneficiary. An EOB is an automatically generated overview of the provided services and corresponding codes and costs.

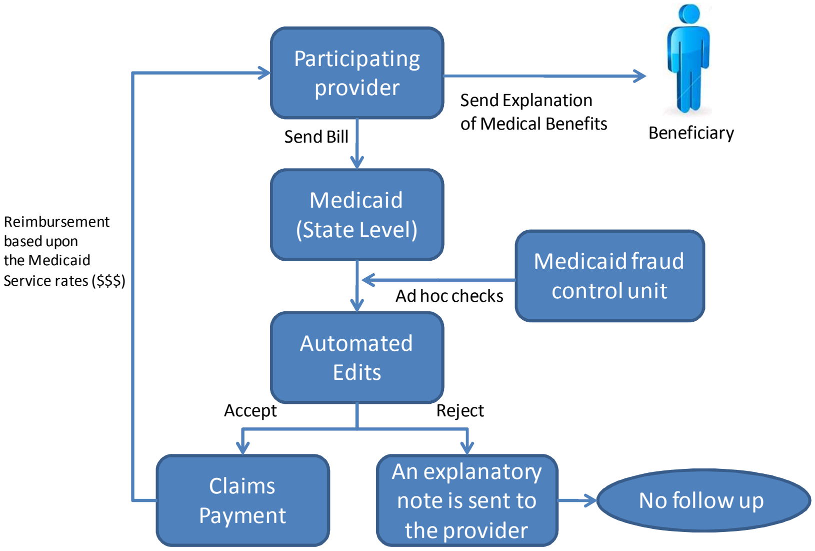
Participating provider
Send Explanation of Medical Benefits
Beneficiary
Send Bill
Medicaid (State Level)
Reimbursement based upon the Medicaid Service rates ($$$)
Medicaid fraud control unit
Ad hoc checks
Automated Edits
Accept
Reject
Claims Payment
An explanatory note is sent to the provider
No follow up

Figure 1: Provider Claim Submission to Medicaid

Every state is responsible for organizing, governing, and operating their Medicaid program. The states process claims with the support of software which is differs state to state. The software performs several prepayment checks and edits to verify if the claim is legitimate. Sparrow (2000) provides some examples of the automated audits:

- Have the mandatory fields been filled in?
- Do the procedure codes match the diagnosis?
- Is the pricing in range with the set boundaries for the service or procedure?
- Has the claim been submitted and paid already (duplicate claims)?

The edits and audits are designed to verify the information with honest providers in mind. However, the system simply lacks effective fraud detection mechanisms (Sparrow 2000). The systems do not verify that the service was provided as claimed, if the diagnosis is correct, or if the patient is aware of the claimed services, as they simply do not possess appropriate, verifiable information. In addition, when a claim is rejected, there is no follow-up investigation as to why an invalid claim was submitted (Sparrow 2000). Instead of vetting these claims, the system sends an explanation to the provider with the reason why the claim was rejected. Thus, instead of flagging what could be fraudulent activity, the system teaches potential fraudsters about the system's billing rules and edits.

EOBs, while well-intentioned, in their current form provide minimal protection against fraud (Sparrow 2000). The beneficiary has little to no financial incentive to pay attention to them. Recipients do not understand the complex computer-generated forms and billing codes contained on them. And, fraudulent providers have even provided incentives to beneficiaries not to read them, including paying $5 per unopened envelope given back to the provider. In addition, many fraud schemes deliberately target vulnerable populations that would be unable to open or understand the EOB or are given kick-backs from the provider not to complain (Kelley 2009).

METHODOLOGY

The foundation of this research is a systematic literature review. In order to systematically review appropriate scientific journals, the following databases have been reviewed, with an initial focus on the top 25 information systems journals (Schwartz and Russo, 2004): Web of Science, Scopus, PiCarta, and Google Scholar. Figure 2 shows the systematic literature review process, with a top down search driven by the keywords, as well as the bottom up search approach using forward and backward citation analysis. Using this methodology, relevant disciplines such as finance, telecommunications, health care, and computer intrusion detection (see table 4) have been included in the review. Exclusion criteria were articles older than 15 years and papers which focus on algorithmic data mining without an emphasis on or application to fraud detection.

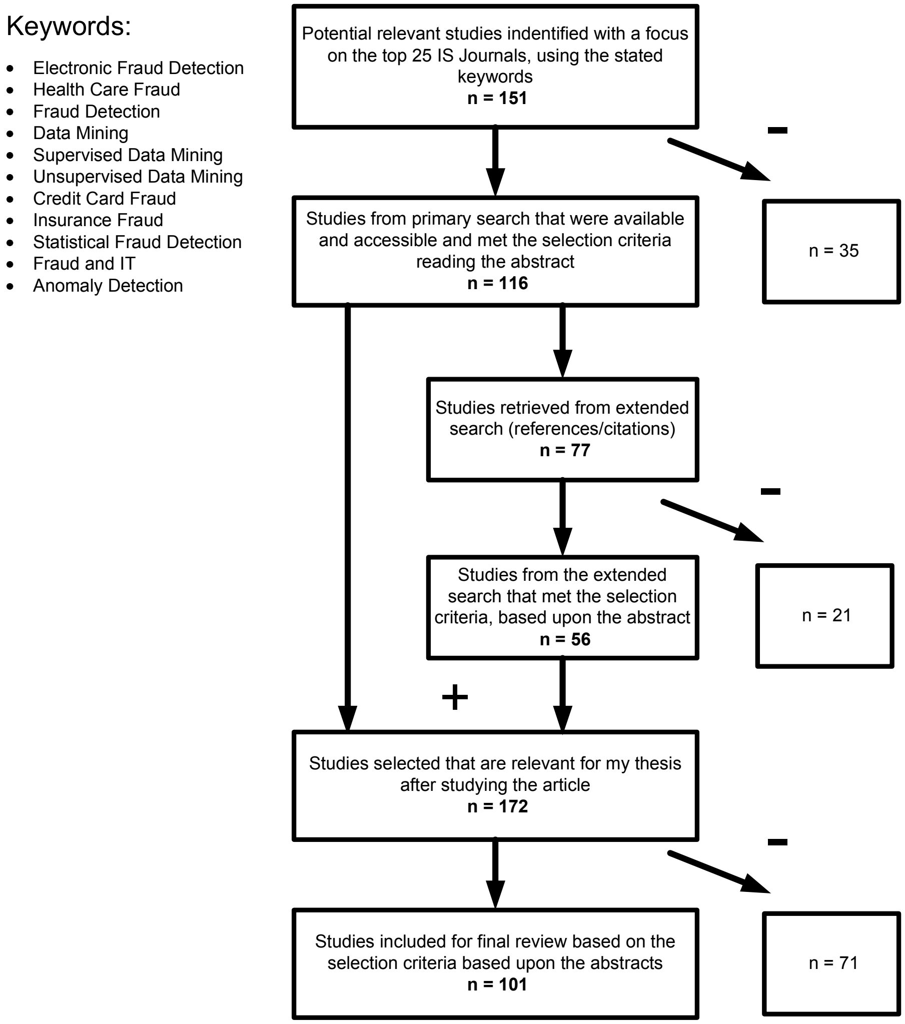
Keywords:
• Electronic Fraud Detection
• Health Care Fraud
• Fraud Detection
• Data Mining
• Supervised Data Mining
• Unsupervised Data Mining
• Credit Card Fraud
• Insurance Fraud
• Statistical Fraud Detection
• Fraud and IT
• Anomaly Detection
Potential relevant studies indentified with a focus on the top 25 IS Journals, using the stated keywords
n = 151
-
n = 35
Studies from primary search that were available and accessible and met the selection criteria reading the abstract
n = 116
Studies retrieved from extended search (references/citations)
n = 77
-
n = 21
Studies from the extended search that met the selection criteria, based upon the abstract
n = 56
+
Studies selected that are relevant for my thesis after studying the article
n = 172
-
n = 71
Studies included for final review based on the selection criteria based upon the abstracts
n = 101

Figure 2: Systematic literature review

HEALTHCARE FRAUD TYPES

Definition of Fraud and Abuse

The terms fraud, waste, and abuse, as used in literature, encompass a wide spectrum of conduct, ranging from intentional misrepresentation of services provided to inadequate documentation of provided care (Hyman 2001). Waste, unnecessary services provided and billed for by a provider, have been explicitly excluded from this research, as they are more difficult to prove, are many times associated with simple inefficiencies and incompatibilities in the healthcare system, and can call into question the subjective medical opinion of providers which can hard to substantiate. While waste is a large problem, it is less tractable and comparable to other industries than fraud and abuse. In our review, we will adopt the CMS definitions for fraud and abuse (CMS 2011):

> *Fraud:* Purposely billing for services that were never given or billing for a service that has a higher reimbursement than the service produced
>
> *Abuse:* Payment for items or services billed by mistake by providers, but should not be paid for by Medicaid

Fraud Strategies

Sparrow (2000) describes two polar extremes in a fraud strategy spectrum: the “hit-and-run” and the “steal a little, all the time” schemes. The hit-and-run is a short term strategy to bill for and acquire large amounts of money quickly and disappear before anyone realizes what happens. At the opposite extreme lies the criminal who steals a little all the time. Legitimate health care providers who provide genuine services use their bulk of legitimate claims to hide the incremental stealing.

Medicaid	Telecommunications	Credit Card
Hit and run	Subscription fraud	Application fraud
Steal a little all the time	Superimposed fraud	Behavioral fraud

Table 1: Types of fraud throughout the fraud detection industries

Similarities exist in the telecommunications industry with subscription fraud (false identification and no intention to pay) and superimposed fraud (slow and hidden) (Cahill et al. 2002). Parallels in the credit card industry include application fraud and behavioral fraud (Bolton et al. 2002b) (see table 1). A major difference between the aforementioned “hit and run” and “steal a little all the time” schemes is the degree to which they are self-revealing (Sparrow 2000). The “hit and run” parallels in the telecommunications and credit card industry are self-revealing because customers, not insurance companies or governments, are losing money rapidly and both see and pay the bill. The “steal a little all the time” comparators are likely more applicable, as customers may well not notice small changes in their bills. Table 2 highlights and categorizes some known Medicaid fraud schemes (GAO 2000).

Fraud Scheme	Short Explanation	Type
Identity Theft	Stealing identification information from providers or beneficiaries and using that information to submit fraudulent bills to Medicaid.	Fraud
Fictitious Practitioners	Enrolling and submitting bills to Medicaid on behalf of fictitious practitioners	Fraud
Phantom Billing	Submitting claims for services not provided.	Fraud
Duplicate Billing	Submitting similar claims more than once.	Fraud/ Abuse
Bill Padding	Submitting claims for unneeded ancillary services to Medicaid.	Fraud/ Abuse
Upcoding	Billing for a service with a higher reimbursement rate than the service provided.	Fraud/ Abuse
Unbundling	Submitting several claims for various services that should only be billed as one master claim that includes ancillary services.	Fraud/ Abuse

Table 2: Medicaid Fraud Schemes (partially derived from (GAO 2000))

FRAUD DETECTION SYSTEMS IN OTHER INDUSTRIES: RESULTS OF THE LITERATURE REVIEW

Table 4 shows a typology of fraud detection techniques that were discovered in the literature review. In Table 3 this typology is used to classify the papers.

Fraud Detection Type	Method	Explanation
A	Supervised Classification Techniques	Use training sets with prior information on class membership to learn classification patterns
A1	Linear Discrimination	Regression based on a logistic curve
A2	Support Vector Machines	A kernel method which selects small number of critical boundary instances (support vectors) to construct a separating hyperplane (Sudjianto et al. 2010)
A3	Neural Networks	A set of interconnected nodes that imitate the functioning of a brain (Kou et al. 2005)
A4	Decision Tree Learning	Methods for building a decision tree for classification
B	Unsupervised Data Mining Techniques	Do not assume prior class labels of legitimate or fraudulent behavior
B1	Anomaly Detection	Tries to detect outliers that are inconsistent with the remainder of that data set (Grubbs 1969). (Barnett et al. 1994)
B2	Cluster Analysis	Divide objects into groups (clusters), with objects in a group being similar to one another but dissimilar to the objects in other groups (Ngai et al. 2011).
B3	Peer Group Analysis	Clusters of similar observations (peer groups) are identified and clustered, subsequently the individual behavior is compared to the cluster's behavior (Bolton et al. 2001).
C	Statistical Methods	Statistical methods are more model and theory based than Data Mining methods
C1	Visualization	Allowing users to view the complex patterns or relationships uncovered in the data mining process (Turban et al. 2007)
C2	Profiling	Process of modeling the characteristic aspects of the user (Fawcett et al. 1997).
C3	Benford's Law	The distribution of the first-digit number of a lot of natural phenomena like size of companies, telephone lengths, and invoice amounts will have a characteristic non-uniform distribution. (Hill 1995; Nigrini 1999)
D	Rule Based	Model based on the experience of experts (Bolton et al. 2002b)
D1	Online Analytical Processing (OLAP)	Dynamic ad-hoc multidimensional analysis (Codd et al. 1993)
D2	SQL Queries	Queries designed by domain experts

Table 3: Overview Fraud Detection Techniques

The structured literature review about fraud detection systems in several industries resulted in an overview of applied fraud detection techniques (see Table 4).

Title Paper	Industry	Objective paper	Type of Fraud	Technique	Results/Findings/Problems
Statistical Methods for fighting Financial Crimes (Sudjianto et al. 2010)	Financial	To provide a survey of statistical techniques and data mining.	Money Laundering, Retail banking fraud	A B1 B2 C2	To provide an overview of financial fraud.
Fraud detection in the telecommunications: History and Lessons learned (Becker et al. 2010)	Telecom	To discuss major fraud schemes and fraud detection techniques used to address them.	Subscription and Superimposed fraud (both telecom)	C1 C2 D	Use simple understandable models, heavy use of visualization, involvement of humans.

Title Paper	Industry	Objective paper	Type of Fraud	Technique	Results/Findings/Problems
Holistic Approach to Fraud Management in Health Insurance (Furlan et al. 2008)	Health care	Holistic overview of fraud management: fraud detection is just one step in the process of fraud management	Health care fraud	A B C	Fraud management is just as important as fraud detection. A case study supports their prepositions
A Comprehensive Survey of Data Mining-based Fraud Detection Research (Phua et al. 2005)	General	To define existing challenges in the fraud detection domain for diff types of large data sets.	Insurance fraud, Credit card fraud, Telecommunications	A B2 B3 C2	Overview of Supervised, semi-supervised and unsupervised techniques.
Novel Techniques for Fraud Detection in mobile telecommunications Networks (Moreau et al. 1997)	Telecom	To explore the detection of fraudulent behavior based upon a combination of absolute and differential behavior.	Telecom fraud	A C2 D	Obtaining a significant amount of fraudulent data and labeling it as such is a significant effort and often a problem.
EFD: A Hybrid Knowledge/Statistical based System for the Detection of Fraud (Major et al. 1992)	Health care insurance	Electronic fraud detection.	Medical insurance fraud	C D	With the applied set of heuristics true positive rates are approximately 50%.
A Taxonomy of Frauds and Fraud Detection Techniques (Laleh et al. 2009)	General	Provide a taxonomy of (new) frauds and fraud detection techniques.	Internal, customs, insurance, credit card, computer, telecommunication	High level overview of A & B	The result is an overview of several types of fraud and different fraud detection techniques on a high level. High-level overview of (un)supervised and semi-supervised techniques.
Survey of Fraud Detection Techniques (Kou et al. 2005)	Credit card Computer Intrusion Telecom	To provide a comprehensive review of different fraud detection techniques.	Credit card Computer intrusion Telecommunications	A3 B1 C1 D	Neural network is an important tool however difficult to implement due to a lack of data. Profiling to detect fraud from call pattern is effective.
Adaptive Fraud detection (Fawcett et al. 1997)	Telecom	To describe a design of user profiling methods.	Superimposed fraud (Telecom)	C2 D	Fraud detection systems must be adaptive and people must determine (trial-and-error) how to profile and which rules are effective.
Unsupervised Profiling Methods for fraud detection (Bolton et al. 2002a)	Credit Card	To apply unsupervised techniques because labeled data is not always available.	Credit card transaction fraud	B3 C2	Both analysis and visualization have the ability to detect anomalies and detect changes in spending trends.
Statistical Fraud Detection: A Review (Bolton et al. 2002b)	Financial Computer intrusion	To describe the statistical tools available in the	Credit Card Money laundering	A B1 C1	The speed of detection is important so the time of detection should be

Title Paper	Industry	Objective paper	Type of Fraud	Technique	Results/Findings/Problems
	Telecom	different areas.	Computer intrusion Telecom	C2 D	measured. How effective a statistical tool is depends on the type of problem.
Neural Fraud Detection in Credit Card Operations (Dorronsoro et al. 1997)	Credit card	To present an applied on-line fraud detection system (Minerva).	Credit card transactions fraud	A3	Positive result; it detects 40% of all the fraudulent transactions and, can be used as a basis for other models.
Establishing Fraud Detection Patterns Based on Signatures (Ferreira et al. 2006)	Telecom	To detect deviate behaviors within a useful time span	Superimposed Fraud (Telecom)	B1 C2	The anomaly detection with the signature as a basis supports the detection of telecom fraud.
Data mining for credit card fraud: a comparative study (Bhattacharyya et al. 2011)	Credit card	To evaluation Random Forests and Support Vector Machines.	Credit card fraud	A2 A4	Random forests based methods are able to obtain good (the best of the 3) overall performances.
The application of data mining techniques in financial fraud detection (Ngai et al. 2011)	Financial	To review data mining techniques in order to discover financial fraud.	Financial and insurance fraud	A B1 B2 C1	A review of 49 articles to categorize financial fraud and an overview of applicable data mining techniques.

Table 4: Overview relevant fraud detection papers

Some papers discuss fraud detection in the health care industry. Major and Riedinger (1992) addressed this topic 19 years ago, and, more recently, Furlan and Bajec (2008) touched this topic from a holistic point of view, highlighting the importance of fraud detection as well as the broader scope of fraud management.

LESSONS LEARNED FOR MEDICAID

The foundations of fraud detection across the various industries studied are underpinned by electronic fraud detection mechanisms which flag suspicious transactions for further review. These sophisticated systems must evaluate mass amounts of information and match patterns both simple and complex. Systems must be paired with humans knowledgeable of appropriate and inappropriate practices to interpret what the data means and to judge if a transaction should be flagged as fraudulent (Hand 2010). While Medicaid possesses its own structural complexities, a great deal of progress can be made with the help of electronic and human data-driven fraud detection techniques.

Stakeholder feedback, or the lack thereof, makes automated electronic mechanisms even more important in government healthcare fraud control. Ideally, stakeholders should be incentivized, willing, and able to offer information that would indicate fraudulent behaviors. In the credit card and telecommunications industries, customers immediately report fraud, as it is in their personal financial best interests to do so. With health insurance, even if a beneficiary notices a mistake on an EOB, they are inclined to think that someone else is paying, so why worry about it (Sparrow 2000). Thus, little feedback is provided from beneficiaries on the legitimacy of claims to state Medicaid agencies.

The credit card and telecommunications industries possess real time data, resolve reported cases of fraud quickly, and, as such, are able to maintain high-quality databases of labeled data which can be used for supervised learning. Medicaid data is dispersed and unlabeled, and there are no signals that this will change in the near future. Multiple stakeholders at the federal and local level, misaligned incentives, and fragmented responsibility hamper the process of labeling and sharing data. Thus, supervised learning techniques are severely restricted.

Clearly improvement is needed in the feedback loop of prosecutions and post-payment adjustment to label the source claims data with high-certainty adjudications that could be leveraged for supervised learning. This should be a joint effort of the federal government, states, and the commercial health insurance industry to improve the data supply and enable the co-development and sharing of fraud models that could apply across the health care industry.

It should be noted that the insurance industry as a whole has much tighter controls around the providers of services, be they healthcare practitioners, auto body shops, or home construction contractors. Providers are modeled and compared, and providers with costs above an acceptable range are excluded from participation and reimbursement under the insurance policy. In contrast, all providers are welcome to participate in Medicaid programs and can only be excluded based on fraudulent activities.

Applicability of Fraud Detection Techniques in Medicaid

Supervised classification models are particularly appropriate for use in health care fraud, as they can be trained and adjusted to detect sophisticated and evolving fraud schemes. In the credit card industry, supervised classification techniques like neural networks, support vector machines, and random forests form the basis for sophisticated and effective fraud detection. The drawback to these techniques is that new fraud schemes are not immediately detectable due to the lag of discovering and labeling new fraud in training data. In the telecommunications industry, unsupervised techniques such as profiling and anomaly detection are applied to complement supervised learning. In the telecommunications industry, extensive, high-quality data is available that is used to construct accurate profiles. Computer security and intrusion detection utilize supervised techniques to discover and detect known patterns and anomaly detection to detect new, unique intrusions. Unfortunately, with health care's more diverse set of outcomes and patterns, applying unsupervised techniques suffers from a high false-alarm rate, because outliers not necessarily implying fraudulent or abusive behaviors but rather the diversity of patterns of care and practitioner prerogatives.

All of these industries have an important advantage over Medicaid: they all possess accurate, real-time, and largely labeled data. Furthermore, these industries are supported by stakeholders who report unusual events and behavior because it affects them directly. These commercial industries and their customers do not want to lose profit, and, therefore, they are willing to allocate the necessary resources to ensure fraud is removed from the system. These industries and companies realize that fraud detection is a vital aspect of doing and staying in business. Medicaid's prioritization of timely payments over accurate, fraud-free payments put the program at a disadvantage from the start. Additionally, the number of stakeholders involved and the fragmented responsibilities further complicate fraud control. That said, with today's technologies and the cooperation of those with knowledge of ground-truths, much progress can be made in fighting Medicaid fraud using both supervised and unsupervised techniques guided by subject-matter and data experts.

Modern modeling, scoring, and business intelligence tools can be used to apply some of these techniques. For example, practical anomaly detection and peer group analysis can be performed and automated when combining claims history with geographically and socioeconomically adjusted provider models. Using dashboards and visualization tools, problematic providers quickly stand out and raise flags for targeting. Business intelligence tools can serve as an important monitoring instrument for payment trends by various dimensions that could be signaling fraud. For example, if, for a specific geographic area, the Medicaid payment profile across provider types suddenly diverges from historical norms and recent national trends, a localized criminal enterprise may be at play. Clearly developing these models with appropriate environmental variables is a challenge, but today's business intelligence, modeling, and scoring tools make their real-world application practical and achievable.

CONCLUSION

Given the fact that the Medicaid is the payer of last resort and receives little feedback from the actual beneficiary of paid healthcare services, the dependence on electronic fraud detection is significantly greater than in similar studied industries. As learned from the credit card industry, telecommunications, and computer security, fraud detection using supervised classification can be extremely effective. However, the base requirement for this approach (labeled data) is currently not available across the Medicaid program. The clear benefits of supervised learning techniques should be weighed against the costs of streamlining data acquisition and closing the feedback loop from adjudicated claims to labeled claims data. Given the high rate of fraud estimates across Medicaid and the program's overall expenditures, it is unfathomable that these IT and business process problems could not be overcome for orders of magnitude less investment than the dollars lost to fraudulent behavior in the program.

In section 3, the analysis showed that supervised techniques are necessary for an effective fraud detection system. Furthermore, the extensive application of classification techniques in various domains proves the effectiveness and utility to contribute to fraud detection. However, no one technique, supervised or unsupervised, is applicable to discover all fraud strategies and schemes. A fraud detection system consisting of multiple techniques, with a flexible, modular approach capable of adapting to the continuous changes in the fraud detection field, must be employed to effectively combat fraud and abuse.

Over time and with increasing levels of sophistication in the fraud control systems, empirical testing must be performed to evaluate their efficacy. Evaluation criteria should include the detection rate, effort, interpretability, and return on investment. The corresponding costs of developing a fraud detection system should be offset and weighed against the resulting benefits of the fraud detection system. What are the strengths and weaknesses of the system, and how can its performance be enhanced? Applying data collection and fraud detection techniques in practice through a state-centric pilot program would help determine the effectiveness of various approaches.

A major limitation of this study is its theoretical approach; a systematic literature study has been conducted to provide an overview of the current worrisome situation in Medicaid and what electronic fraud detection techniques exist in related fraud detection domains. The published fraud frameworks are limited since fraudsters would benefit from being able to easily access the information and would undoubtedly attempt to use that sensitive information to enhance their fraud techniques. While not an ideal investigation, the literature review does provide a proper first impression and overview of the existing fraud schemes and detection techniques as currently applied across similar industries.

Future research should be undertaken to evaluate the current methodologies and tools employed by states and CMS to detect and prevent fraud as well as to assess the potential impact of the methodologies discussed in this paper. In addition, while not a technology problem, an in-depth assessment should evaluate the effects of Medicaid policy changes such as increasing Medicaid provider enrollment standards, delaying payment to allow for more claim review time, or providing incentives to report fraudulent activity found on EOBs.

The high number of stakeholders, 50 states with unique legislation and eligibility rules, and the sheer magnitude of the program complicate Medicaid fraud control efforts. As Sparrow (2000) explains, fraud should be properly measured to create a realistic impression of the current situation and an estimation of the amount of fraud and abuse in the system. The Thompson Reuters estimation (Kelley 2009) of $600 to $850 billion lost to fraud, waste, and abuse annually is only an estimate. Without significant, periodic audits of randomly sampled claims across the Medicaid system, it is impossible to accurately estimate the level of fraud, waste, and abuse in the system and its change over time. Although fraud will never be completely eradicated, it can be better managed with systematic improvements in data collection, applied detection and prevention tools, better incentive structures, and enforcement actions.

REFERENCES

1. Barnett, V., and Lewis, T. *Outliers in statistical data*, (3rd ed.) John Wiley & Sons, Chichester, 1994.
2. Becker, R.A., Volinsky, C., and Wilks, A.R. "Fraud Detection in Telecommunications: A Historical Perspective and Lessons Learned," *Technometrics* (52) 2010, pp 20–33.
3. Bhattacharyya, S., Jha, S., Tharakunnel, K., and Westland, J.C. "Data mining for credit card fraud: A comparative study," *Decision Support Systems* (50:3) 2011, pp 602-613.
4. Bolton, R., and Hand, D. "Unsupervised Profiling Methods for Fraud Detection," *Statistical Science* (17:3) 2002a, pp 235-255.
5. Bolton, R.J., and Hand, D.J. "Peer Group Analysis – Local Anomaly Detection in Longitudinal Data," Technical Report, Department of Mathematics, Imperial College, London, 2001.
6. Bolton, R.J., and Hand, D.J. "Statistical fraud detection: A review," *Statistical Science* (17:3) 2002b, pp 235-255.
7. Cahill, M., Chen, F., Lambert, D., Pinheiro, J.C., and Sun, D.X. "Detecting Fraud in the Real World," in: *Handbook of Massive Datasets,* J. Abello, P. Pardalos and M. Resende (eds.), Kluwer Press, New York, 2002.
8. CMS "Centers for Medicaid and Medicare Services," Retrieved 1/20/2011, from www.cms.gov, 2011.
9. Codd, E.F., Codd, S.B., and Salley, C.T. "Providing OLAP (Online Analytical Processing) to User-Analysts: An IT Mandate," Codd and Date, Inc., 1993.
10. Derrig, R.A. "Insurance fraud," *Journal of Risk and Insurance* (69:3) 2002, pp 271-287.
11. Dorronsoro, J.R., Ginel, F., Sanches, C., and Cruz, C.S. "Neural Fraud Detection in Credit Card Operations," *IEEE Transaction on Neural Networks* (8:4) 1997, pp 827 - 834.
12. Fawcett, T., and Provost, F. "Adaptive fraud detection," *Data mining and knowledge discovery* (1:3) 1997, pp 291-316.
13. Ferreira, P., Alves, R., Belo, O., and Cortesao, L. "Establishing Fraud Detection Patterns Based on Signatures," in: *Proceedings of the 7th Industrial Conference on Data Mining, ICDM 2006*, Springer, Leipzig, 2006, pp. 526-538.
14. Furlan, S., and Bajec, M. "Holistic Approach to Fraud Management in Health Insurance," *Journal of Information and Organizational Sciences* (32:2) 2008, pp 99- 114.
15. GAO "United States General Accounting Office. Health Care Fraud: Schemes to defraud Medicare, Medicaid, and private health care insurers," in: *T-OSI-00-15*, 2000.
16. Grubbs, F.E. "Procedures for detecting outlying observations in samples," *Technometrics* (11:1) 1969, pp 1-21.

17. Hand, D.J. "Fraud Detection in Telecommunications and Banking: Discussion of Beacker, Volinsky, and Wilks (2010) and Sudjianto et al. (2010)," *Technometrics* (52:1) 2010, pp 34 – 38.
18. Hill, T.P. "A Statistical derivation of the significant-digit law," *Statistical Science* (10:4) 1995, pp 354 - 363.
19. Hyman, D.A. "Health Care Fraud and Abuse: Market Change, Social Norms, and the Trust" Reposed in the Workmen"," *The Journal of Legal Studies* (30:2) 2001, pp 531-567.
20. Kelley, R.R. "Where can $700 Billion in Waste be cut annually from the US Healthcare System?," *Ann Arbor, MI: Thomson Reuters* (24) 2009.
21. Kou, Y., Lu, C.T., Sirwongwattana, S., and Huang, Y.P. "Survey of fraud detection techniques," Networking, Sensing and Control, 2004 IEEE International Conference on, 2005, pp. 749-754.
22. Laleh, N., and Azomi, M.A. "A Taxonomy of Frauds and Fraud Detection Techniques," in: *Proceedings of the third International Conference, ICISTM 2009,* S.K. Prasad, S. Routray, R. Khurana and S. Sahni (eds.), Springer, Ghaziabad, India, 2009, pp. 256-267.
23. Major, J.A., and Riedinger, D.R. "EFD: A Hybrid Knowledge/Statistical-Based System for the Detection of Fraud," *International Journal of Intelligent System* (7) 1992, pp 687 – 703.
24. Moreau, Y., Preneel, B., P., B., Shawe-Taylor, J., Stoermann, C., and Cooke, C. "Novel Techniques for fraud detection in mobile telecommunications networks," in: *ACTS Mobile Summit*, Grenada Spain, 1997.
25. Ngai, E.W.T., Hu, Y., Wong, Y.H., Chen, Y., and Sun, X. "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems* (50:3) 2011, pp 559-569.
26. Nigrini, M.J. "I've Got Your Number," *Journal of Accountancy* (187:5) 1999, pp 79 - 83.
27. Phua, C., Lee, V., Smith, K., and Gayler, R. "A comprehensive survey of data mining-based fraud detection research," *Artificial Intelligence Review*) 2005, pp 1-14.
28. Sparrow, M.K. *License To Steal: How Fraud bleeds america's health care system* Westview Press, Boulder, 2000.
29. Sudjianto, A., Nair, S., Yuan, M., Zhang, A., Kern, D., and Cela-Díaz, F. "Statistical Methods for Fighting Financial Crimes," *Technometrics* (52) 2010, pp 5–19.
30. Truffer, C.J., Klemm, J.D., Wolfe, C.J., and Rennie, K.E. "2010 Actuarial Report on the Financial Outlook for Medicaid," Office of the Actuary, Centers for Medicare & Medicaid Services, Department of Health & Human Services Washington, 2010.
31. Turban, E., Aronson, J.E., Liang, T.P., and Sharda, T. *Decision Support and Business Intelligence Systems*, (9th ed.) Pearson Education, Upper Saddle River, 2007.
32. US Congressional Budget Office "Spending and Enrollment Detail for CBO's March 2010 Baseline: Medicaid," retrieved 2/22/2011 from http://www.cbo.gov/budget/factsheets/2010b/medicaidBaseline.pdf, 2010.