Using Big Data to minimize Fraud, Waste, and Abuse (FWA) in United States Healthcare

Paul Marks
Indiana University
Online Student
Shepherdsville, Kentucky 40165
pcmarks@iu.edu

ABSTRACT

The cost of healthcare includes the loss of billions of dollars due to Fraud, Waste, and Abuse (FWA). Many of the schemes to commit FWA are very intricate and require the analysis of many data sources simultaneously. The question answered here is "How can we use big data analysis to help minimize these costs and thus optimize the money spent on healthcare?"

KEYWORDS

i523, hid327, fraud, waste, abuse, healthcare, health insurance

1 INTRODUCTION

FWA is an issue that affects everyone in the U.S. since healthcare services are leveraged by everyone at some point and the costs for those services include the money lost to FWA. The three components of FWA are varying degrees of culpability. The Centers for Medicare and Medicaid Services (CMS) in part defines fraud as "knowingly and willfully executing, or attempting to execute, a scheme or artifice to defraud any health care benefit program", Waste as "overusing services, or other practices that, directly or indirectly, result in unnecessary costs", and Abuse as "involves payment for items or services when there is not legal entitlement to that payment and the provider has not knowingly and/or intentionally misrepresented facts".[9] While the percentage of cost attributable to FWA can vary from insurer to insurer, Medicare estimates that 11 percent of its payments for Original Medicare are improper primarily due to FWA.[8] In combination these cost the United States healthcare system 80 billion dollars[6] annually.

Advances in big data technology can help reduce these losses. Big data offers the ability to look at data in real time to determine if a claim is legitimate or not. Historically, due to the amount of data involved, this type of analysis would have to happen after the claims have been paid with specific models targeting specific schemes to identify FWA. Big data can help lower the cost of health-care in the United States by identifying FWA claims and stopping payments before they occur.

2 HEALTHCARE FRAUD, WASTE, AND ABUSE ENVIRONMENT

It is easy to understand the problem FWA poses. Healthcare funds are of limited quantity. Insurance helps to spread the cost among groups of people, but does not provide limitless funds. As costs increase, so do premiums or direct payments for health-care. In order for as many people to be able to have access to healthcare costs have to be managed. There are many ideas for helping to provide affordable healthcare, but there is much discussion and disagreement on exactly how to do that. Reducing costs by eliminating as much FWA as possible is one solution that everyone, except for those participating in and profiting from FWA schemes, can agree on.

Data to fight FWA is not just the information gathered by a doctor or other provider while working with a patient. In order to fully utilize advances in technology, multiple sources of information must be brought together. Sources include claims (current and historic), clinical, provider, geospatial, and other sources of information. This allows for data analytics to take a deeper look into not only a single participant, but others who may be related to that participant. "If Provider A is involved in improper billing, it is not uncommon for other providers with which they associate to also be engaged in bad behavior. Thus, many payers will work to analyze connected providers. Information on corporate ownership, billing and management companies, social media interactions of physicians and staff can reveal whether other physicians, pharmacies, radiology centers, home infusion agencies, etc. are engaged in a broader pattern of referral and collusion."[13]

The problem for big data to solve is the size of all this data and how to process it fast enough. Using CMS as an example, being a government entity much of their data is available publicly, it is easy to get an idea of the amount of data. Medicare processed 1.2 billion claims in 2014, covering 53.8 million beneficiaries, with 6,142 hospitals, and 1,173,802 non-institutional providers[7]. In addtion payments must be made within a specific timeframe depending on the insurer and their agreement with providers. This time includes all the normal steps to verify and process a claim so the time available to examine the data for FWA is very limited.

It must be noted that when working with this type of data, Protected Health Information (PHI) and Personally Identifiable Information (PII), that there are many regulations about the ability to access and secure it which must be followed. While this makes it more difficult to get access to the data it can be overcome by working cooperatively with the various data owners.

2.1 Big Data Techniques for FWA

So how can big data be used to approach this issue? Leveraging big data tools such as Hadoop, analysts could divide the different sources of information into data lakes, looking at each source separately, and then combining the results. Table 1 on page 6 shows sources of information and what level of FWA they are generally related to. The highest level combines sets of data. "Level 7 combines all previous data views and concerns all fraud that is part of criminal networks which involve many different beneficiaries and/or providers. This much larger data view, spanning billions of claims in the case of Medicaid, is the most rich, delivering the ability to perform complex network analysis that could detect intricate conspiracies. However, performance of analysis here will be much lower than in previous levels."[14]

While there are simple cases of fraud which follow a typical known pattern, this is only a portion of the problem. Fraud schemes change and can involve many different entities which may not seem to be related on the surface. The more data which can be combined and analyzed, the more fraud that can be found. "Much of the FWA that plague health care payers is the result of organized, sophisticated and collusive activities among providers and between providers and patients. Social network analysis can help identify relationships, links and hidden patterns of information sharing and interactions within potentially fraudulent clusters, including:

- Patient relationships with known perpetrators of health care fraud;
- Links between recipients, businesses, assets and relatives and associates;
- Links between licensed and non-licensed and sanctioned providers; and
- Inappropriate relationships between patients, providers, employees, suppliers and partners"[5]

In order to keep up with organized fraud activities, there must be a dedicated practice of data analytics which is ever evolving.

Traditionally programs have been written to look for specific sets of circumstances. Leveraging existing knowledge about the data and using it to look for specific patterns is known as supervised in big data terms. "There are several supervised fraud detection methods such as: Bayesian Networks, Neural Networks (NNs), Decision Trees, and Fuzzy Logic. NNs and decision trees are the most popular fraud detection methods because of their high tolerance of noisy data and huge data set handling."[3] There are also unsupervised methods in which data is fed into the system without preexisting notions of what to look for[3]. Unsupervised methods sort through data and find relationships and groupings of related information, find clusters of what could be considered normal, and determine where the outliers are.

Because unsupervisord methods only identify outliers, applying unsupervised methods to healthcare data will require that the outliers will then have to be verified as FWA or acceptable patterns. "Patrick McIntyre, SVP of Health Care Analytics at Anthem, one of the country's biggest payers, credits machine learning and big data with their ability to "identify potentially fraudulent or wasteful claims on a daily basis." The algorithms are run at the same time as claims are batch processed, so questionable claims are immediately identified, flagged and sent to the clinical coding experts for review."[4] This greatly increases the ability to fight FWA by having the machine pinpoint where to look in all the data available to the reviewer. Suddenly the task of finding fraud is not as daunting. By leveraging both of these techniques FWA can be discovered at an accelerated pace. The number of models the system knows will grow over time as more data is fed into it and more patters are discovered and verified.

2.2 Current Solutions

Many companies currently offer solutions for detecting FWA in healthcare payment systems. They include the ability to identify FWA claims during the payment cycle so that payment is not made to suspect claims. Truven Health[1], Healthcare Fraud Shield[12], and SAS[10], just to name a few, all have systems they offer based on big data. The specifics of the systems they offer are proprietary in nature so many of the descriptions are generic. Truven Health for instance claims "Our approach to FWA analytics is to infuse our solution with healthcare intelligence. As one example, Truven Health pioneered a 'medical model' for the healthcare fraud detection process. A key component of this model is our Medical Episode Grouper (MEG). MEG is an advanced healthcare analytic methodology that groups all the services for a clinically-relevant episode of illness, including inpatient, outpatient, and drug. The MEG episodes are constructed using rigorously defined clinical classifications based on current medical literature. Episodes analysis can expose patterns of clinical and billing abuse that are otherwise difficult to detect such as wasteful or unnecessary services. Using MEG, the analyst can understand the entire range and cost of services provided to a patient during a single episode of illness, which then is aggregated to profile a provider's entire practice."[1] SAS materials include "Detect improper payments before the money goes out the door, and get potential savings of hundreds of millions of dollars. Take an enterprise approach to detecting and preventing fraud, waste and abuse with a hybrid analytics solution"[10] and "By making analytical models and rules engines part of the process, you can spot more payment integrity breaches than ever before. Process all data (not just a sample) through rules and analytical models. Use customized models to detect previously unknown schemes. And spot linked entities and crime rings, which can help stem larger losses."[11]. The U.S. Federal government is also investing heavily into FWA analytics. "The CMS has awarded defense contractor Northrop Grumman Corp. a $ 91 million contract to develop and implement a second generation of an advanced analytics system, called the Fraud Prevention System, to help identify high-risk claims in Medicaid and Medicare."[2]

2.3 Future uses of Big Data Analytics

Currently there is still a certain amount of honor built into healthcare. "The system's inherent structure of trust enables both simple billings errors and illicit actors to hide in the shadows of the murky deep as overpayments quietly siphon money away from legitimate care."[13] If a claim is submitted by a valid entity, using the correct process, and everything is in order then it is most likely paid. For many claims this is done without any specific proof of the services being provided. With more and more healthcare information being digitized this may not be the case in the future. X-rays, lab tests, clinical notes, etc. are all being stored digitally. Computers are now able to interpret images and unstructured text very accurately. By linking this data to claims data the clinical information could be required as part of claims payment. An x-ray of broken bone, notes which support a diagnosis, Magnetic Resonance Imaging files, could all be interpreted automatically. Not only would the data be used to compare to the claims information, but to other images/notes on file to ensure that the same files were not being submitted with multiple claims. The system could know what one individual medical history looks like compared to another similar to how facial recognition is able to match like images. Requiring and being able to validate more information before services are paid for would help the reduce the ability of perpetrators of FWA to be able to get reimbursed for services they should not. This level of verification would not be possible without the ability to process massive amounts of data quickly.

Historically the payers of most healthcare claims, insurers, have not had the ability to examine actual evidence that a service has taken place on a broad scale. (It is done manually on a specific case or audit basis.) Through the use of advances in big data and combining current and new data stores such as electronic health records into the payment process a difference can be made in the amount of money lost to FWA in healthcare. "By combining identity and entity resolution, rules-based claim and clinical review, complex linking analysis and predictive analytics into a seamless workflow, we will come closer to migrating an integrated pre-pay fraud solution to a real risk control environment with the potential to eliminate billions of dollars in improper payments due to FWA. This is not just a health care imperative, but a national economic imperative that must be addressed immediately. The analytics exist. It is time for those analytics to be implemented and the hard choices that enable that implementation to be made to insure that we remain at the forefront of quality care for all Americans."[5]

3 CONCLUSIONS

While there may be disagreement on many aspects of healthcare in America, everyone should agree that eliminating Fraud, Waste, and Abuse within the system is the right thing to do. FWA costs billions of dollars annually. Just a 1 percent reduction in the estimated 80 billion dollars annually would result in 800 million dollars in savings. With this amount of money at stake significant investments should continue to be made in leveraging advanced big data technologies into solving this problem. Due to the continued rise in the amount of data collected traditional programming cannot keep up with the pace. Advanced techniques must be leveraged which can learn in an unsupervised manner. The future of the best methods for fighting FWA in healthcare will be a combination of this analysis and teams specializing in the rules and regulations of healthcare in the United States. The unsupervised methods will work through massive amounts of structured and unstructured data breaking it down into cases and schemes which are most like FWA. These will be reviewed, confirmed or denied as accurate, and fed back into overall FWA platform. As this cycle continues over and over the ability to fight FWA in United States Healthcare will get better. While Big Data may never eliminate FWA in Healthcare it can help to minimize it and save the country billions of dollars a year.

ACKNOWLEDGMENTS

The author would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper. It has helped to expand my knowledge in how modern data analytics can help to save FWA which has plagued the healthcare system.

REFERENCES

[1] Truven Health Analytics. 2017. Program Integrity. Online. (2017). http://truvenhealth.com/your-healthcare-focus/government/program-integrity
[2] Virgil Dickson. 2016. Northrop Grumman wins $ 91 million CMS contract for fraud detection. Online. (04 2016). http://www.modernhealthcare.com/article/20160407/NEWS/160409912
[3] Namrata Ghuse, Pranali Pawar, and Amol Potgantwar. 2017. An Improved Approch For Fraud Detection In Health Insurance Using Data Mining Techniques. *International Journal of Scientific Research in Network Security and Communication* 5, 3 (06 2017), 27–33.
[4] Erin Hitchcock. 2017. The Role of Big Data in Preventing Healthcare Fraud, Waste and Abuse. Online. (09 2017). https://www.datameer.com/company/datameer-blog/role-big-data-preventing-healthcare-fraud-waste-abuse/
[5] Mark Isbitts. 2017. Preventing Health Care Fraud with Big Data and Analytics. Online. (2017). http://www.lexisnexis.com/risk/insights/health-care-fraud-layered-approach.aspx
[6] Vinil Menon and Parikshi Sheth. 2016. Big Data Analytics Can Be a Game Changer for Healthcare Fraud, Waste, and Abuse. Online. (04 2016). https://www.hfma.org/Content.aspx?id=47523
[7] United States Department of Health and Human Services. 2015. 2015 CMS Statistics. Online. (12 2015). https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMS-Statistics-Reference-Booklet/Downloads/2015CMSStatistics.pdf
[8] United States Department of Health and Human Services. 2016. FY 2016 Agency Financial Report. Online. (11 2016). https://www.hhs.gov/sites/default/files/fy-2016-hhs-agency-financial-report.pdf
[9] United States Department of Health and Human Services. 2017. Combating Medicare Parts C and D Fraud, Waste, and Abuse Web-Based Training Course. Online. (01 2017). https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/CombMedCandDFWAdownload.pdf
[10] Inc. SAS Institute. 2017. Health Insurance Get full coverage for your big data challenges. Online. (2017). https://www.sas.com/en_us/industry/health-insurance.html
[11] Inc. SAS Institute. 2017. SAS Fraud Framework for Health Care. Online. (2017). https://www.sas.com/en_us/software/fraud-framework-for-health-care.html
[12] Healthcare Fraud Shield. 2017. FWAShield A Fully Integrated Fraud, Waste and Abuse Software System. Online. (2017). http://www.hcfraudshield.com/fwashield.aspx
[13] Rodger Smith. 2016. Using Big Data in the Hunt for Healthcare Fraud, Waste, and Abuse Payers must leverage all the big data analytics tools at their disposal to hunt down healthcare fraud, waste, and abuse. Online. (04 2016). https://revcycleintelligence.com/news/using-big-data-in-the-hunt-for-healthcare-fraud-waste-and-abuse
[14] Dallas Thornton, Roland M. Mueller, Paulus Schoutsen, and Jos van Hillegersberg. 2013. Predicting Healthcare Fraud in Medicaid: A Multidimensional Data Model and Analysis Techniques for Fraud Detection. *Procedia Technology* 9, Supplement C (2013), 1252 – 1264. https://doi.org/10.1016/j.protcy.2013.12.140 CENTERIS 2013 - Conference on ENTERprise Information Systems / ProjMAN 2013 - International Conference on Project MANagement/ HCIST 2013 - International Conference on Health and Social Care Information Systems and Technologies.

[Table 1 about here.]

4 BIBTEX ISSUES

5 ISSUES

bibtex labels can not have any spaces, _ or in it

no title

you do not use "quotes" properly

you need to *emphasize* and not "quote"

this is cool

Have you written the report in the speciffied format?

Have you included an acknowledgement section?

Have you included the paper in the submission system (In our class it is git)?

Have you specified proper identification in the submission system. THis is typically a form or ASCII text that needs to be filled out (In our case it is a README.md file that includes a homework ID, names of the authors, and e-mails)?

Have you included all images in native and PDF format in the submission system?

Have you added the bibliography file that you managed (In our case jabref to make it simple for you)?

In case you used word have you also provided the jabref?

In case of a class and if you do a multi-author paper, have you added an appendix describing who did what in the paper?

Have you spellchecked the paper?

Are you using and, a, the properly?

Have you made sure you do not plagiarize?

Is the title properly capitalized?

Have you not used phrases such as shown in the Figure below, but instead used as shown in Figure 3 when referring to the 3rd figure?

Have you capitalized Figure 3, Table 1, ... ?

Have you removed any figure that is not referred explicitly in the text (As shown in Figure ..)

Are the figure captions bellow the figures and not on top. (Do not include the titles of the figures in the figure itself but instead use the caption or that information?

When using tables have you put the table caption on top?

Make the figures large enough so we can read the details. If needed make the figure over two columns?

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. If you want you can place all figures at the end of the report?

Are all figures and tables at the end?

In case you copied a figure from another paper you need to ask for copyright permission. IN case of a class paper youmustinclude a reference to the original in the caption.

Do not use the word *I* instead use we even if you are the sole author?

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned.

Do not artificially inflate your paper if you are bellow the page limit and have nothing to say anymore.

Is the paper shorter than 2 pages text, e.g. no images, tables figures

Is the paper more than 6 pages text, e.g. no images, tables figures

Do not use the characters & # % _ put a bakslash berfore them

If you want to say and do not use & but use the word and.

Latex uses double single open quotes and double single closed quotes for quotes. Have you made sure you replaced them?

Pasting and copying from the Web often results in non ascii characters to be used in your text, please remove them and replace accordingly.

Figures shoudl be reasonably sized and ofteh you just need to add width columnwidth

e.g.

```
[width=\columnwidth]
```

List of Tables

Table 1: Types of Fraud and their related Sources[14]

		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	**		**	*	**	***

Usefulness: * Low ** Medium *** High