

Waste in Healthcare Using Big Data to help optimize our Healthcare spend

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

ABSTRACT

The cost of healthcare includes the cost of inefficient services. While this could include topics such as misdiagnosis, less effective treatment plans, and more efficient use of types of services (emergency room vs. immediate care vs. telemedicine) the purpose of this paper is the cost of Fraud, Waste, and Abuse (FWA) within the system. The estimate for what percentage of cost are 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.[7]. The question 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, Medicare, Medicaid, FWA, health insurance

1 INTRODUCTION

The definition of verb waste includes "to spend or use carelessly"[5]. When money spent on healthcare goes to FWA, it is money being spent carelessly. FWA is all of our issue 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 Fraud, Waste and Abuse are varying degrees of culpability of waste. The Centers for Medicare and Medicaid Services (CMS) in part defines fraud as "is 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"[8]. In combination these cost the United States healthcare system 80 billion dollars[4] 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 healthcare in the United States by identifying FWA claims and stopping payment before it occurs.

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."[9]

The problem for big data to solve is the deluge of information 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 million hospitals, and 1,173,802 non-institutional providers[6]. Payments are generally 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.

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. Figure 1 on page 5 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." [10]

Traditionally programs are 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." There are also unsupervised methods in which data is fed into the system without preexisting notions of what to look for [1]. 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.

Applying unsupervised methods to healthcare data will identify patterns that 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." [2] This greatly increases the ability to fight FWA by having the machine pinpoint where to look in all the data available. 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 patterns are discovered and verified.

Specifically fraud can be particularly hard to find. There are simple cases which follow a typical known pattern. However 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

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

2.2 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." [9] 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.

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.

There will inevitably be privacy concerns, but new sources of information must be brought into the fight against FWA. 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 spent on 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." [3]

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 waste which has plagued the healthcare system.

REFERENCES

- [1] Namrata Ghuse, Pranali Pawar, and Amol Potgantwar. 2017. An Improved Approach 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.
- [2] 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/>
- [3] 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>
- [4] 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>
- [5] Merriam-Webster. 2017. Waste. Online. (09 2017). <https://www.merriam-webster.com/dictionary/waste>
- [6] 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>
- [7] 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>
- [8] 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/CombMedCandDFWdownload.pdf>
- [9] 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>
- [10] Dallas Thornton, Roland M. Mueller, Paulus Schoutsen, and Jos van Hilleberg. 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.]

LIST OF TABLES

1	Types of Fraud and their related Sources[10]
---	--

5

Table 1: Types of Fraud and their related Sources[10]

		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