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

Paul Marks
Indiana University
Bloomington, Indiana
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. (2016 Financial Report). 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

1 INTRODUCTION

The definition of verb waste includes "to spend or use carelessly" (webster). When money spent on healthcare goes to FWA, it is money being spent carelessly. We, FWA is all of our issue, are not doing enough to ensure the money is used for the goods or services provided. FWA 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". (training) In combination these cost the United States healthcare system 80 billion dollars (hfma) annually. Advances in big data technology can help reduce these losses. Big data can offer 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. Specific models targeting specific schemes would identify FWA. Big data can help lower the cost of health-care in the United States by identifying FWA claims and stopping payment before it occurs.

2 FWA IN HEALTHCARE

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. Reducing costs by eliminating as much FWA as possible is one solution. In order to fully utilize advances in technology, the sources of information must be brought together. Sources include claims (current and historic), clinical, provider, geospatial, and other sources of information. The problem is the deluge of information and how to process it fast enough. Payments are generally made in so many days depending on the insurer and their agreement with providers. 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. (2015 Stats)

2.1 Ideas for Big Data

So how can big data be used to approach this issue? The theme could be divide and conquer. Leveraging big data tools such as Hadoop they 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." (ac els cdn)

2.2 Big Data Techniques for FWA

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. (Approch) 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. 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 patters are discovered and verified.

2.3 The Future

Currently there is still a certain amount of honor built into healthcare. If a claim is submitted by a valid entity, using the correct process, and everything is in order then it is most likely paid. 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. It could know what one individual medical history looks like compared to another similar to how facial recognition is able to match like images. This would not be possible without the ability to process massive amounts of data quickly.

3 CONCLUSIONS

While there may be disagreement on aspects of healthcare in America, everyone should agree that eliminating Fraud, Waste, and Abuse within the system is good for everyone. 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. Because of 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. While there will inevitably be privacy concerns, new sources of information must be brought into the fight against FWA. Historically payers of healthcare claims, insurers, have not had the ability to require actual evidence that a service has taken place. By leveraging advances in big data and combining data stores such as electronic health records into the payment process a difference can be made in the amount of money spent on healthcare in America.

REFERENCES

[Figure 1 about here.]

		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