MOX-Report No. 06/2017

On the Use of the Concentration Function in Medical Fraud Assessment

Ekin, T.; Ieva, F.; Ruggeri, F.; Soyer, R.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

On the Use of the Concentration Function in Medical Fraud Assessment

Tahir Ekin*
Department of Computer Information Systems and Quantitative Methods
Texas State University-San Marcos, USA
and
Francesca Ieva
Dipartimento di Matematica
Politecnico di Milano, Milano, Italy
and
Fabrizio Ruggeri
Consiglio Nazionale delle Ricerche
Istituto di Matematica Applicata e Tecnologie Informatiche, Milano, Italy
and
Refik Soyer
Department of Decision Sciences
School of Business, The George Washington University, Washington D.C., USA

February 9, 2017

Abstract

We propose a simple, but effective, tool to detect possible anomalies in the services prescribed by a health care provider (HP) compared to his/her colleagues in the same field and environment. Our method is based on the concentration function which is an extension of the Lorenz curve widely used in describing uneven distribution of wealth in a population. The proposed tool provides a graphical illustration of a possible anomalous behavior of the HPs and it can be used as a pre-screening device for further investigations of potential medical fraud.

Keywords: Fraud detection; Lorenz Curve; Health care providers; Unsupervised data mining

*Tahir Ekin is Brandon Dee Roberts Professor and Assistant Professor of Quantitative Methods, Texas State University, San Marcos, TX, USA (e-mail: t_e18@txstate.edu); Francesca Ieva is Researcher, Politecnico di Milano, Italy (e-mail: francesca.ieva@polimi.it); Fabrizio Ruggeri is Research Director, CNR IMATI, Milano, Italy (e-mail: fabrizio@mi.imati.cnr.it); and Refik Soyer is Professor of Decision Sciences and of Statistics, Mitch Blaser Distinguished Scholar in Business Analytics, The George Washington University, Washington, DC, USA (e-mail: soyer@gwu.edu)

1 Introduction

Health care expenditures have increased significantly over recent decades in developed countries. In addition to the aging population and increasing diversity in health services, fraud as well as abuse and waste dramatically contribute to the increase in health care costs. For example, in the United States, the National Health Care Anti-Fraud Association (`www.nhcaa.org`) estimated conservatively that at least 3%, or more than 60 billion dollars, of annual health care expenditures was due to fraud, waste and abuse in 2010. Total health care related spending in the United States was almost 3 trillion dollars corresponding to $9,523$ dollars per person in 2013 (`www.cms.gov`). Therefore, efforts for assessment and reduction of expenses due to medical fraud are crucial in the health care industry.

The systematic use of statistical approaches in medical fraud assessment in the United States gained momentum with the Health Care Fraud and Abuse Control Program in 1996. Sampling and overpayment estimation methods help medical auditors to use sample data and make extrapolations for the population. Woodard (2015) demonstrates the use of sampling by US governmental medical insurance programs. The formal governmental guidelines (`www.cms.gov`) recommend the use of the lower limit of a one sided 90 percent confidence interval of the total overpayments as the recovery amount from the provider under investigation. This protects the provider and results in fair recovery with 95 percent confidence. However, application of the Central Limit Theorem to compute the lower bound is based on the assumption that the overpayment population either follows the Normal distribution or that the sample size of overpayments is reasonably large. It is well known that medical claims data mostly exhibit skewness and non-normal behavior, requiring large sample sizes for a valid application. As alternatives, Edwards et al. (2005) obtain the lower bounds using a non-parametric inferential method whereas Gilliland and Edwards (2011) construct randomized lower bounds. Ignatova and Edwards (2012) propose a sequential sampling framework that aims to make inference on the proportion of claims with overpayments. Ekin et al. (2015) present a zero-one inflated mixture model for estimation of overpayments. Musal and Ekin (2016) provide a recent overview of such references before proposing a Bayesian mixture model.

These studies are mostly interested in the percentages of claims to be reviewed to make

a reliable assessment on the possible fraudulent behavior of a health care provider (HP). In contrast, our paper studies how billing by an HP, split among different prescribed services, differs from the average behavior in a population of HPs.

There are relevant data mining studies that provide sophisticated methods to detect possible fraud. Supervised approaches, including but not limited to neural networks, decision trees and Bayesian networks, are proposed; see Li et al. (2008) for a survey. These methods require labeled data, which correspond to already audited claims. The results of these methods are dependent on a particular data set; therefore they cannot adapt to the dynamic nature of fraud patterns. As a potential remedy, unsupervised methods can be used to extract information about the relationships within medical data. Particularly, these may serve as pre-screening tools to identify a set of potentially fraudulent claims before domain experts are brought during the investigation phase. Onderwater (2010) provides an overview of such outlier detection methods for fraud assessment. The anomaly detection framework for Australia Medicare spatio-temporal data (Ng et al. (2010)), and the use of Benford's Law Distributions to detect anomalies in claim reimbursements (Lu and Boritz (2005)) are some examples. In order to group medical claims data, clustering algorithms can also be used. Musal (2010) provides an illustration of clustering of geographical regions as input to his regression model. The Bayesian co-clustering model of Ekin et al. (2013) investigates the dyadic patterns among providers and beneficiaries. Overall, these methods can decrease personnel costs as fewer transactions are reviewed (Laleh and Azgomi (2009)).

Our paper presents the use of the concentration function as a pre-screening tool to aid in medical fraud assessment. It does not suffer from the issues of supervised methods; in particular, it does not require labeled data and can easily adapt to changes if it is run with a different data set. This simple unsupervised tool lets the auditor analyze the billing patterns of a particular doctor, which can reveal potential unusual behaviors. Compared to existing data mining approaches, our approach provides a simple tool, both in implementation and understanding, which could be used to detect an anomalous behavior and might hinder potential fraud.

The idea behind the proposed approach is quite simple. We assume that a group of health care providers (HPs) with similar characteristics (age, specialty, years in the area,

etc.) are providing similar services to patient populations that are similar in terms of distribution of age, income, gender, etc. We are aware that this assumption might not fully reflect reality, but our goal is to provide a tool that can detect unusual behavior by an HP in terms of deviating from a population of HPs who are expected to display similar prescribed service patterns. Further analyses will be needed to prove if such an anomaly is due to fraud or due to heterogeneity in population of patients. It is possible that an HP might prescribe particular services, and charge consequently, more frequently than other HPs with no fraudulent intent. For example, a provider may prescribe a significantly larger number of prostate exams in one year due to a larger number of elderly patients in need in his/her area. Assessment of potential causes of such different behavior is the next step to be undertaken through careful review of the claims by the HP. In Section 2 we will introduce the concentration function, an extension of the Lorenz curve, which will be used in Section 3 to analyze real data. Final remarks will be presented in Section 4.

2 Concentration function and Lorenz curve

The proposed tool is based on the concentration function, which is a generalization of the Lorenz curve (see, e.g., Marshall and Olkin, 1979, p. 5) and is well known in the statistical literature. The Lorenz curve is a graphical tool used to describe the discrepancy between a discrete probability measure $\mathbf{\Pi}$ and a discrete uniform measure $\mathbf{\Pi}_0$. Its typical application is about the comparison of the actual income distribution in a population ($\mathbf{\Pi}$) with an income that is evenly distributed across the population ($\mathbf{\Pi}_0$). The Lorenz curve is obtained by plotting the cumulative wealth of the poorest individuals in the population. In particular, we consider a population of n individuals with wealth (income) $x_i, i = 1, \ldots, n$, assuming no ties for simplicity. We order their incomes in ascending order and obtain the ordered wealths $x_{(1)}, \ldots, x_{(n)}$, from the poorest to the richest individual. We define $S_0 = 0$ and $S_k = \sum_{i=1}^{k} x_{(i)}$. Therefore, S_n is the total income of the population and S_k/S_n is the fraction of wealth owned by the k poorest individuals. We plot the curve connecting the points $(k/n, S_k/S_n), k = 0, \ldots, n$. For a given k, the plot displays the fraction S_k/S_n of the total income owned by the $k/n \cdot 100\%$ of the poorest part of the population. We obtain a

convex, increasing function connecting the points $(0,0)$ and $(1,1)$. It is worth mentioning that evenly distributed wealth implies a straight line since $S_k = k/n \cdot S_n$ for all k. When the discrepancy in wealth distribution gets larger, then S_k/S_n gets smaller and further away from k/n. Therefore, the Lorenz curve deviates farther away from the straight line as the disparity in income distribution increases. Uneven distribution of wealth can be detected not only through visual inspection of the plot but also through some summary indices. In particular, we consider Gini's area of concentration (Gini, 1914) which measures the area between the Lorenz curve and the straight line, i.e., $(n+1)/(2n) - (1/n)\sum_{1\leq k\leq n} S_k/S_n$, and an index proposed by Pietra (1915), which measures the maximum distance between the curve and the straight line, i.e., $\sup_{1\leq k\leq n-1}(k/n - S_k/S_n)$. Larger values of the indices denote larger disparities. Gini's and Pietra's indices are defined here when comparing the income distribution with respect to a discrete uniform one. Later, we will present the extension to the case of any (not necessarily uniform) discrete distribution.

We can compare the Lorenz curves and related indices for two or more populations. The population with the lowest Lorenz curve, if that exists, is the one with the largest disparity in income distribution. Sometimes, Lorenz curves might intersect and this requires the use of Gini's area of concentration and Pietra's index for comparison of the income distributions. The first case is well represented by two populations, A and B, of three individuals each, where the wealth is distributed according to the following, already ordered, percentages of $(10\%, 30\%, 60\%)$ in A and $(20\%, 30\%, 50\%)$ in B.

The Lorenz curves for the two populations are presented in Figure 1, where the lowest dashed curve corresponds to the wealth distribution in A, the middle dotted one to the distribution in B and the highest (straight) solid curve corresponds to the case in which wealth is evenly distributed among the three individuals. In this case it is evident that both populations have an unequal distribution of wealth, and it is more uneven in population A.

An example of intersecting Lorenz curves is obtained when the population A and B have percentages of $(15\%, 40\%, 45\%)$ and $(20\%, 25\%, 55\%)$, respectively. The corresponding curves are presented in Figure 2. In this case there is no ordering of Lorenz curves. It is true that the poorest individual in A has less income than the corresponding one in B (15%

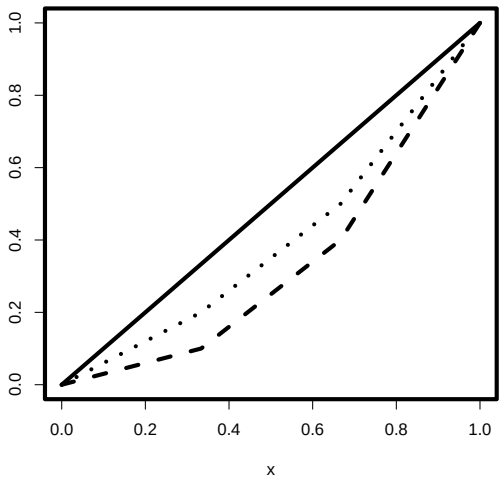

Figure 1: Lorenz curve for wealth distribution in population A (dashed) and B (dotted) w.r.t. uniform distribution.

vs 20%) but the reverse holds when considering the two poorest individuals (55% vs 45%). Gini's area is 0.1 for A and 0.117 for B, whereas Pietra's index is 0.183 for A and 0.217 for B. The indices show that the income distribution is, in general, more uneven in B, where the largest values are obtained.

The Lorenz curve can be extended to compare any pair of probability measures on the same measurable space, as described in Cifarelli and Regazzini (1987). The same authors showed that Pietra's index is equal to the total variation norm distance between the two probability measures. We do not refer to their elegant, but mathematically sophisticated, definition but we rather prefer to illustrate its use with a very simple example related to HPs' prescribed services.

Suppose that HPs in a homogeneous region are prescribing only 3 tests (blood, urine and ECG) for their patients. We are looking at the percentage of the billing for each test with respect to (w.r.t.) the total. We are interested in discovering if an HP has a different pattern w.r.t. the group and, therefore, if further investigation of the individual is worthwhile in order to detect possible fraud. We suppose the billings in the HPs group for blood test, urine test and ECG account for 20%, 40% and 40% of the total amount,

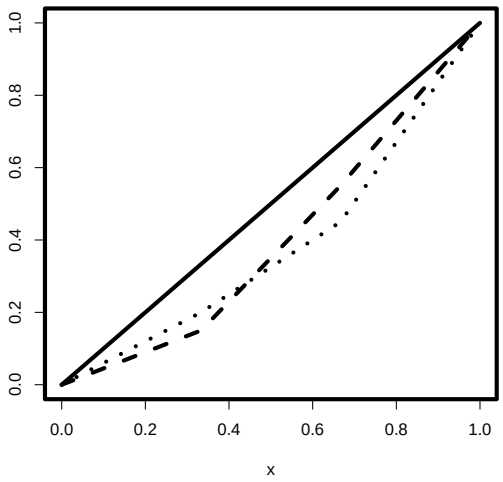

Figure 2: Lorenz curve for wealth distribution in population A (dashed) and B (dotted) w.r.t. uniform distribution.

respectively.

We consider two HPs and we expect that they behave in a similar way w.r.t. the group. The billing of the first HP (called A) is split into 20% for blood tests, 70% for urine tests and 10% for ECG, whereas the percentages for the second HP (called B) are 30%, 50% and 20%, respectively. In probabilistic terms we are interested in comparing two probability measures: a reference one, $\mathbf{\Pi}_0$, related to the whole group of HPs with probabilities $(0.2, 0.4, 0.4)$ (for blood, urine and ECG, respectively) w.r.t. $\mathbf{\Pi}$ for the selected HP, given by $(0.2, 0.7, 0.1)$ for A and $(0.3, 0.5, 0.2)$ for B.

Therefore, we consider two probability measures, $\mathbf{\Pi}$ and $\mathbf{\Pi}_0$, assigning probabilities $\underline{p} = (p_1, \ldots, p_n)$ and $\underline{q} = (q_1, \ldots, q_n)$, respectively, to the same outcomes $(x_1, \ldots, x_n)$ of a statistical experiment (here the billing for different services, i.e., blood, urine and ECG). We suppose $\underline{q}$ represents the distribution for the group and we are interested in *measuring* how far $\underline{p}$ is from it. Earlier, the Lorenz curve was constructed summing the income of the individuals x_i starting from the poorest. The concentration function is constructed summing the probabilities of the outcomes x_i which are more unlikely under $\mathbf{\Pi}$ than under $\mathbf{\Pi}_0$ (i.e., the values where $\mathbf{\Pi}$ is *less concentrated* than $\mathbf{\Pi}_0$).

For each $i, i = 1, \ldots, n$, we compute the (likelihood) ratios $r_i = p_i/q_i$ and order the x_i's according to ascending values of r_i. We therefore order the outcomes from the ones where $\mathbf{\Pi}$ assigns much less probability than $\mathbf{\Pi}_0$ towards the ones where $\mathbf{\Pi}$ assigns much more probability than $\mathbf{\Pi}_0$. The ordered values are denoted as $x_{(1)}, \ldots, x_{(n)}$, and the corresponding probabilities are $q_{(1)}, \ldots, q_{(n)}$ and $p_{(1)}, \ldots, p_{(n)}$. Similar to the Lorenz curve, we plot the curve connecting the points $(Q_k, P_k), k = 0, \ldots, n$, where $Q_0 = P_0 = 0$, $Q_k = \sum_{i=1}^{k} q_{(i)}$ and $P_k = \sum_{i=1}^{k} p_{(i)}$. As before, we obtain a convex, increasing function connecting the points $(0, 0)$ and $(1, 1)$ and we call it the *concentration function of* $\mathbf{\Pi}$ *w.r.t.* $\mathbf{\Pi}_0$.

Let us focus on a particular value k and consider Q_k and P_k to better understand their meaning. When considering all possible percentages $Q_k 100\%$ of billings for services prescribed by the whole group of HPs, there are some services which have been charged less by the HP under scrutiny. In our case, the 60% of billings by the entire group could be represented either by blood and urine tests or by blood test and ECG. Looking at HP A, the first pair corresponds to 90% of his/her billing for the prescribed services and the second to just 30%. We are interested in the second pair since it provides the smallest possible percentage, $P_k 100\%$, of billings for A among all those with a given percentage $Q_k 100\%$ for the group of HPs. Such choice corresponds to the largest possible dissimilarity in behavior, and this is what we are after. In probabilistic terms, we are interested in the set of outcomes which assign less probability under $\mathbf{\Pi}$ (here 0.3) among all those with a given probability under $\mathbf{\Pi}_0$ (here 0.6). Such outcomes are given by $x_{(1)}, \ldots, x_{(k)}$, whose probabilities sum up to P_k under $\mathbf{\Pi}$ and Q_k under $\mathbf{\Pi}_0$.

Like the Lorenz curve, the concentration function can be used to analyze the distance between the probability measures $\mathbf{\Pi}$ and $\mathbf{\Pi}_0$. If the distance between the concentration function and the straight line is short, then the distribution of the services prescribed by an individual HP is similar to that of the group of HPs. No warning should be issued in this case since the HP's behavior is very similar to the group's. Otherwise, future investigation may be suggested in search of possible causes, potentially including fraud.

In the previous example about two HPs, it can be shown that Q_1 is obtained in both cases for the ECG billing, accounting for 40% of the group total, whereas the value of P_1

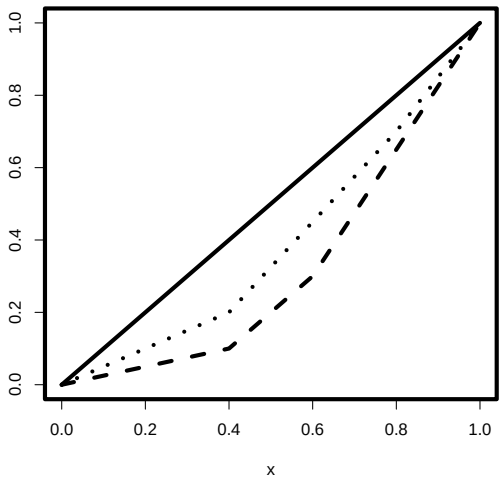

Figure 3: Concentration function for HP A (dashed) and HP B (dotted) w.r.t. population.

is given by 0.1 for A and 0.2 for B. Whereas B is not too far from the group, A should be further investigated because of the very small amount charged for ECG and the larger amount for the urine tests that are prescribed.

The concentration curves for the two HPs are presented in Figure 3, where the lowest dashed curve corresponds to A, the middle dotted one to B and the highest (straight) solid curve corresponds to the case in which the HP charges tests in the same percentages as the group.

The different behavior of HP A is confirmed by looking at Gini's area of concentration and Pietra's index. Their previous definitions can be extended to the current situation. Pietra's index becomes $\sup_{1\leq k\leq n-1}(Q_k - P_k)$ whereas Gini's area of concentration becomes $1/2 - 1/2\sum_{1\leq k\leq n}(P_i + P_{i-1})(Q_i - Q_{i-1})$. The latter is 0.18 for A and 0.11 for B, whereas the former is 0.3 for A and 0.2 for B.

Once a synthetic index exceeds a fixed threshold then a warning should be issued and the individual HP could be subject to more detailed investigation to detect the causes of the different behavior than the group of HPs.

3 Application of the concentration function

In this study, we utilize the public data set, *Provider Utilization and Payment Data Physician and Other Supplier Public Use File*, that was prepared by *The Centers for Medicare & Medicaid Services* (`www.cms.gov`). It includes information related to payment, number of services and number of beneficiaries for each provider and prescribed service. We rearrange the data and we consider here a small dataset, with 30 medical doctors (MDs) in Diagnostic Radiology in Vermont and the percentages of their billings within a set of 61 prescribed services. In the Appendix, Table 3 contains the service descriptions. We chose MDs in the same specialty and in an area with a small number of people (Vermont is the second least populous state in the United States) to make the assumption about uniformity of behavior among HPs more reasonable. The prescribed services include X-rays, Computed Tomography, Magnetic Resonance Imaging for different parts of the human body. We consider just two MDs (named MD1 and MD2 respectively) out of 30 since we would like to show how the concentration function works in the simplest case. The extension to all the MDs is straightforward.

Table 1 presents the percentages of each MD's billing for each i^{th} prescribed service (p_i^1 and p_i^2 respectively, $i = 1, 2, \ldots, 61$) and the average percentages q_i among the population. Whereas, Table 2 lists the likelihood ratios for each MD ($r_i^1 = p_i^1/q_i$ and $r_i^2 = p_i^2/q_i$, respectively).

First of all, the likelihood ratios r_i^1 and r_i^2 provide information on how the charges of each MD differ from the average charges of the population. A value close to 1 denotes a similar behavior in terms of percentage of billing for such service, whereas a smaller (larger) one shows that the MD is charging less (more) than the average. Note that we are not considering the billing in dollars, but are normalizing the figures considering just the percentage w.r.t. the total billing.

There are some services never charged by the MDs but our interest is about those which are overcharged w.r.t. the population: those that might be due to fraud, or, at least, abuse and waste. In particular, we set a threshold (5 in our case) on the likelihood ratio whose exceedance should trigger further investigations. When the likelihood ratio is more than 5, then the percentage of charges for the related service prescribed by an MD is at least five

Table 1: Percentages of prescriptions for MD1 (p_i^1), MD2 (p_i^2) and population (q_i)

Type								
p_i^1	0.0089	**0.0319**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.0144	0.0000	0.0047	0.0123	**0.0208**	0.0000	0.0187	0.0000
	0.0119	0.0098	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.0000	**0.0344**	**0.0514**	0.0000	0.0000	0.0144	0.0000	0.0000
	0.0242	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.0238	0.0000	0.0000	0.0000	0.0637	0.0246	0.0000	0.0331
	0.1865	0.0412	0.0251	0.0467	0.0416	0.0000	0.0000	0.0514
	0.0000	0.0000	0.0000	0.1155	0.0888			
p_i^2	0.0030	0.0000	0.0000	0.0000	0.0070	0.0076	0.0042	0.0000
	0.0030	0.0061	0.0065	0.0056	0.0070	0.0035	0.0060	0.0056
	0.0052	0.0039	0.0018	**0.0222**	0.0000	0.0033	0.0018	0.0073
	0.0030	0.0138	0.0000	0.0020	0.0000	0.0083	0.0000	0.0202
	0.0088	0.0026	0.0066	0.0000	0.0000	0.0169	0.0178	0.0000
	0.0140	0.0182	0.0030	0.0205	0.0028	0.0018	0.0378	0.0201
	0.0127	0.0249	0.0130	0.0253	0.0188	0.0030	0.0081	0.0308
	0.0159	0.0886	0.0888	0.1082	0.2333			
q_i	0.0028	**0.0034**	0.0032	0.0034	0.0035	0.0033	0.0033	0.0037
	0.0039	0.0035	0.0035	0.0039	**0.0042**	0.0046	0.0044	0.0039
	0.0046	0.0048	0.0045	**0.0039**	0.0043	0.0047	0.0058	0.0060
	0.0061	**0.0065**	**0.0088**	0.0068	0.0077	0.0076	0.0075	0.0073
	0.0087	0.0088	0.0096	0.0122	0.0111	0.0104	0.0098	0.0126
	0.0132	0.0127	0.0134	0.0132	0.0151	0.0163	0.0140	0.0172
	0.0193	0.0178	0.0190	0.0200	0.0220	0.0216	0.0284	0.0293
	0.0379	0.0736	0.0855	0.1046	0.1676			

Table 2: Likelihood ratios for MD1 (r_i^1) and MD2 (r_i^2) w.r.t. population

Type								
r_i^1	3.2364	**9.2643**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3.7370	0.0000	1.3289	3.1351	**5.0040**	0.0000	4.2212	0.0000
	2.5907	2.0574	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.0000	**5.2653**	**5.8476**	0.0000	0.0000	1.8881	0.0000	0.0000
	2.7880	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1.8049	0.0000	0.0000	0.0000	4.2120	1.5120	0.0000	1.9252
	9.6515	2.3163	1.3229	2.3307	1.8918	0.0000	0.0000	1.7561
	0.0000	0.0000	0.0000	1.1039	0.5299			
r_i^2	1.0909	0.0000	0.0000	0.0000	1.9886	2.3194	1.2625	0.0000
	0.7785	1.7313	1.8379	1.4274	1.6840	0.7538	1.3544	1.4213
	1.1321	0.8188	0.3959	**5.6633**	0.0000	0.7031	0.3116	1.2133
	0.4953	2.1122	0.0000	0.2950	0.0000	1.0883	0.0000	2.7811
	1.0138	0.2945	0.6899	0.0000	0.0000	1.6307	1.8244	0.0000
	1.0617	1.4308	0.2231	1.5561	0.1851	0.1106	2.6936	1.1691
	0.6573	1.3999	0.6852	1.2627	0.8549	0.1391	0.2856	1.0523
	0.4196	1.2046	1.0382	1.0341	1.3922			

times larger than the average charge for that service.

In Table 2 we highlight the values of likelihood ratios exceeding the threshold, using boldface for r_i^1 (MD1) and r_i^2 (MD2). Those providers and services with high likelihood ratios are also given in bold in Table 1 to highlight their percentages p_i^1 and p_i^2 for the bills by the two MDs and the average billing q_i by the MDs population. This may facilitate a more thorough examination about the anomalous billings.

First of all, for MD1 there are two ratios exceeding 9: they correspond to Computed Tomography of the abdomen and pelvis (9.2643) and X-ray exam of abdomen (9.6515). The former accounts for more than 3% of the charges by MD1 whereas the latter represents almost 19%. Although both account for charges almost ten times more than average, the major concern is about the X-ray exam of abdomen since it accounts for almost one fifth of the total billing by MD1. Further investigation in this case should be about number of prescriptions and diagnoses, to check for possible overcharges (and fraud, therefore) and waste in prescribing the X-ray. The other three prescriptions with ratios larger than 5 are X-ray exam of thigh (5.0040), a different X-ray exam of abdomen (5.2653) and X-ray exam of hip (5.8476). MD2 is prescribing just an exam, the ultrasound exam of the abdomen back wall, charged well above average. Given that there is just one anomalous exam, accounting only for 2% of the charges, then MD2 can be hardly considered at risk of fraudulent behavior, unlike MD1.

The analysis of the likelihood ratios allows us to identify a different behavior of MD1 w.r.t. the MDs population regarding the charges for five prescribed services, out of 61. The concentration function is a graphical tool which allows for an immediate recognition, at a glance, of the anomalous behavior when considering all the prescribed services.

The concentration functions, plotted in Figure 4, denote clearly how MD1 (dashed line) differs significantly from the population since the corresponding curve is quite far from the straight line, unlike MD2 (dotted line). Such a distance is summarized by Pietra's index of 0.5504, compared to 0.2198 for MD2. A similar result is obtained when considering Gini's index: 0.3580 for MD1 and 0.1593 for MD2.

A look at the concentration function in Figure 4 can provide further information on the behavior of the MDs. The flat (dashed) line from 0 to almost 0.5 tells us that MD1 is never

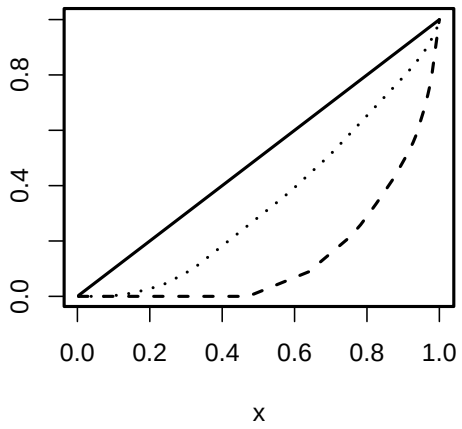

Figure 4: Concentration function for MD1 (dashed) and MD2 (dotted) w.r.t. population.

giving prescriptions accounting for almost 50% of the billings by the average population of MDs. The sharp increase around 1 is interpreted as an excess of charges by MD1 w.r.t. average (mostly due to X-ray exam of abdomen, as discussed earlier). The dotted curve is almost parallel to the straight line: this is due to very few prescribed services which are not charged by MD2, unlike the MDs population, implying the initial lowering of the curve, whereas all the other services are charged quite similarly to the average, except for the ultrasound exam of the abdomen back wall, as discussed earlier.

In Figure 5 we present the histogram of Pietra's index values for the 30 MDs in the group. It can be seen that 20 MDs have their index in the first two bins, denoting a substantial concordance among themselves about percentages of charges for the services they prescribe. There are just a few values (7) exceeding 0.5, including the one corresponding to MD1. If a warning limit is set to 0.5, then all the MDs whose Pietra's index exceeds that value could be subject to further investigation.

As shown in this section, the concentration function, and the likelihood ratios needed to construct it, can be easily computed and can provide insights on the general behavior of an MD and also on the details of his/her prescribed services and related charges.

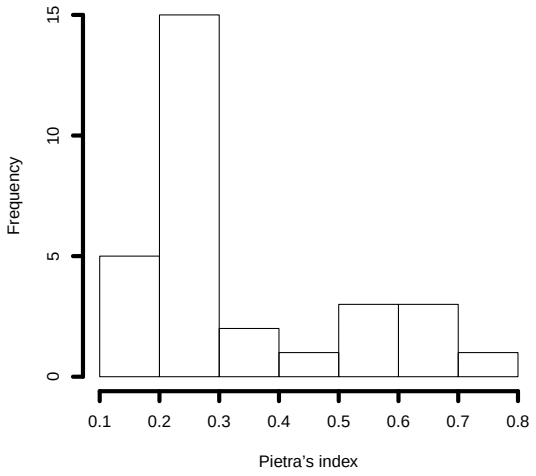

Figure 5: Histogram of Pietra's index for 30 MDs.

4 Discussion

In this paper we present a simple tool which could detect anomalous behavior of health care providers w.r.t. a population of providers believed homogeneous, with a similar pattern in terms of charges for prescribed services. As discussed in the paper, the tool does not provide formal evidence of fraud or abuse but it can be used as a pre-screening device to detect possible anomalies in the pattern of charges, which could be further investigated. It should be noted that heterogeneous behavior in terms of charges for prescribed services can be totally legitimate, and may be due to specialists having sicker patients and performing necessary operations. However, the tool still provides us information about different doctor billing patterns which may also be the result of incentives that lead doctors to overcharge for a procedure.

We are working on the assessment of potential fraud in medical practice by different approaches. One approach that will be presented in a forthcoming paper considers sophisticated methods based on Bayesian co-clustering to link groups of providers and prescribed services. However, it is important to note that one also needs to provide simple tools that can be used by practitioners and the current work is an attempt to meet this need. At the

same time, it is also worth considering critical aspects of data. In fact, data pre-processing takes a fair amount of time before conducting the statistical analysis. Furthermore, another important aspect is medical data security. The data analyst should adhere to proper security, access and privacy controls. Personally identifiable information that can distinguish or trace any identity should be dealt with caution and protected properly.

Those issues should be kept in mind when considering other possible applications. As an example, the proposed tool can be used to analyze deviations within any given category. Therefore we can conduct the analysis for differently defined peer groups of providers. For example, Berenson-Eggers Type of Service (BETOS) categories have been used to categorize the providers and analyze U.S. Medicare costs. These clinical categories are a collection of Health Care Financing Administration Common Procedure Coding System procedure codes and can serve as peer groups. Another potential extension is to construct sub-peer groups with respect to a co-variate, such as patient profiles. Then, the billings of a given provider can be compared to the peer group to extract potential patterns. These insights can be helpful before bringing domain experts into the investigation. Close cooperation between physicians, statisticians and people involved in decision making is essential while interpreting the results.

Appendix

X-ray exam of humerus	Ct abd & pelv 1/ > regns
Ct abdomen w/dye	Us exam pelvic complete
Bone imaging whole body	Transvaginal us non-ob
Us exam of head and neck	Mri jnt of lwr extre w/o dye
X-ray exam of neckspine 2 view	X-ray exam of neckspine 4 view
X-ray exam of ribs/chest	X-ray exam of elbow
X-ray exam of thigh	Pet image w/ct skull-thigh
X-ray exam of thoracic spine	Extremity study
X-ray exam of lower leg	X-ray exam of finger(s)
Ct angiography chest	Us exam abdo back wall lim
Mri brain w/o & w/dye	Diagnosticmammographyuniteral
Mri brain w/o dye	Us exam abdo back wall comp
Ct neck spine w/o dye	X-ray exam of abdomen 2
X-ray exam of hip	Ct thorax w/o dye
X-ray exam of lower spine	X-ray exam of hand
Mri lumbar spine w/o dye	Echo exam of abdomen
X-ray exam of wrist	Us exam abdom complete
Computer dx mammogram add-on	Ht muscle image spect mult
X-ray exam of knee 3	Us exam breast(s)
Extracranial study	Mammogram screening
X-ray exam of ankle	Diagnosticmammographybilateral
Ct abd & pelvis	Extremity study
X-ray exam series abdomen	X-ray exam of knee 1 or 2
Dxa bone density axial	X-ray exam of pelvis
X-ray exam of abdomen	X-ray exam knee 4 or more
X-ray exam of foot	X-ray exam of lower spine
X-ray exam of shoulder	X-ray exam of shoulder
Ct abd & pelv w/contrast	X-ray exam of hip
Ct head/brain w/o dye	Comp screen mammogram add-on
Total knee arthroplasty	Repair of wound or lesion
Inject spine w/cath c/t	

Table 3: Service descriptions

5 Bibliography

Cifarelli, D.M., and Regazzini, E. (1987), "On a general definition of concentration function", *Sankhyā B*, 49, 307-319.

Edwards, D., Ward-Besser, G., Lasecki, J., Parker, B., Wieduwilt, K., Wu, F., and Moorhead, P. (2003), "The minimum sum method: a distribution-free sampling procedure for medicare fraud investigations". *Health Services and Outcomes Research Methodology*, 4(4), 241-263.

Ekin, T., Ieva, F., Ruggeri, F., and Soyer, R. (2013), "Application of Bayesian Methods in Detection of Healthcare Fraud", *Chemical Engineering Transactions*, 33, 151-156.

Ekin, T., Musal, R.M., and Fulton, L.V. (2015), "Overpayment models for medical audits: multiple scenarios", *Journal of Applied Statistics*, 42(11), 2391-2405.

Gilliland, D., and Edwards, D. (2011), "Using randomized confidence limits to balance risk: an application to Medicare investigations, *The American Statistician*, 65, 149-153.

Gini, C. (1914), "Sulla misura della concentrazione della variabilità dei caratteri", *Atti del Reale Istituto Veneto di S.L.A., A.A. 1913-1914*, 73, parte II, 1203-1248.

Ignatova, I., Deutsch, R.C., and Edwards, D. (2012), "Closed sequential and multistage inference on binary responses with or without replacement", *The American Statistician*, 66, 163-172.

Laleh, N., and Azgomi, M.A. (2009), *A taxonomy of frauds and fraud detection techniques*. In Information Systems, Technology and Management: ICISTM 2009, 256-267, Springer: Berlin.

Li, J., Huang, K-Y., Jin, J., and Shi, J. (2008), "A survey on statistical methods for health care fraud detection", *Health Care Management Science*, 11, 275-287.

Lu, F., and Boritz, J.E. (2005), *Detecting fraud in health insurance data: Learning to model incomplete Benford's law distributions*. In Machine Learning: ECML 2005,

633-640, Springer: Berlin.

Marshall, A.W., and Olkin, I. (1979), *Inequalities: Theory of Majorization and its Applications*, New York, NY: Academic Press.

Musal, R.M. (2010), "Two models to investigate Medicare fraud within unsupervised databases", *Expert Systems with Applications*, 37(12), 8628-8633.

Musal, R.M., and Ekin, T. (in press), "Medical Overpayment Estimation: A Bayesian Approach", *Statistical Modelling*.

Ng, K.S., Shan, Y., Murray, D.W., Sutinen, A., Schwarz, B., Jeacocke, D., and Farrugia, J. (2010), *Detecting non-compliant consumers in spatio-temporal health data: A case study from medicare Australia*. In Data Mining Workshops (ICDMW), 2010 IEEE International Conference on, 613-622, IEEE.

Onderwater, M. (2010), *Detecting unusual user profiles with outlier detection techniques*, Amsterdam: VU University.

Pietra, G. (1915), "Delle relazioni tra gli indici di variabilità", *Atti del Reale Istituto Veneto di S.L.A. A.A. 1914-1915*, 74, parte II, 775-792.

Woodard, B. (2015), "Fighting healthcare fraud with Statistics", *Significance*, 12:3, 22-25.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

05/2017 Menafoglio, A.; Hron, K.; Filzmoser, P.
Logratio approach to distributional modeling

04/2017 Dede', L; Garcke, H.; Lam K.F.
A Hele-Shaw-Cahn-Hilliard model for incompressible two-phase flows with different densities

02/2017 Arena, M.; Calissano, A.; Vantini, S.
Monitoring Rare Categories in Sentiment and Opinion Analysis - Expo Milano 2015 on Twitter Platform.

03/2017 Fumagalli, I.; Parolini, N.; Verani, M.
On a free-surface problem with moving contact line: from variational principles to stable numerical approximations

01/2017 Riccobelli, D.; Ciarletta, P.
Rayleigh-Taylor instability in soft elastic layers

58/2016 Antonietti, P. F.; Bruggi, M. ; Scacchi, S.; Verani, M.
On the Virtual Element Method for Topology Optimization on polygonal meshes: a numerical study

56/2016 Guerciotti, B.; Vergara, C.; Ippolito, S.; Quarteroni, A.; Antona, C.; Scrofani, R.
A computational fluid-structure interaction analysis of coronary Y-grafts

57/2016 Bassi, C.; Abbà, A.; Bonaventura, L.; Valdettaro, L.
Large Eddy Simulation of gravity currents with a high order DG method

55/2016 Antonietti, P. F.; Facciola' C.; Russo A.; Verani M.;
Discontinuous Galerkin approximation of flows in fractured porous media on polytopic grids

54/2016 Vergara, C.; Le Van, D.; Quadrio, M.; Formaggia, L.; Domanin, M.
Large Eddy Simulations of blood dynamics in abdominal aortic aneurysms