Paper 10761-2016

Medicare Fraud Analytics Using Cluster Analysis: How PROC FASTCLUS Can Refine the Identification of Peer Comparison Groups

Paulo Macedo, Ph.D., Integrity Management Services Inc.; Sewit Araia, M.P.H., Integrity Management Services Inc.; Babak Zafari, Ph.D., George Washington University

ABSTRACT

The existence and magnitude of fraud in health insurance programs, although limited to a small fraction of health care providers, requires the utilization of fraud prevention and detection procedures. Data Mining methods are used to uncover odd billing patterns in large databases of health care claims history where efficient fraud discovery involves the preliminary step of deploying automated outlier detection techniques in order to classify identified outliers as potential fraud, before an in-depth investigation. An essential component of the outlier detection procedure is the identification of proper peer comparison groups to classify providers as within-the-norm or outliers.

This study refines the concept of peer comparison group within the category of provider and considers the possibility of the existence of distinct billing patterns associated with medical or surgical procedure codes identifiable by the Berenson-Eggers Type of System (BETOS). The BETOS system, as defined by the Center for Medicare and Medicaid Services (CMS), "covers all HCPCS codes (Health Care Procedure Coding System); assigns a HCPCS code to only one BETOS code; consists of readily understood clinical categories; (and) consists of categories that permit objective assignment."

This study focuses on the specialty "General Practice" and involves two steps: first, the identification of clusters of similar BETOS-based billing patterns; and second, the assessment of the effectiveness of these peer comparison groups in identifying outliers. The working dataset is a sample of the summary of 2013 data of physicians active in health care government programs made publicly available by the CMS through its website. The analysis uses PROC FASTCLUS, the SAS® Cubic Clustering Criterion approach to find the optimal number of clusters in the data and PROC ROBUSTREG to implement a Multivariate Adaptive Threshold Outlier Detection method.

INTRODUCTION

Although limited to a small fraction of health care providers, the existence and magnitude of fraud in Health Insurance programs requires the utilization of fraud prevention and detection procedures. The Fall 2015 Semiannual Report of the Office of Inspector General (OIG of the Department of Health and Human Services) to Congress states that its mission is to work in the identification of "improper payments, which sometimes may be caused by fraud, but may also indicate questionable billing patterns that could lead to waste or signal other problems, such as low quality of care." The document covers work completed in the second half of fiscal year 2015 (April to September) and highlights "expected recoveries of over nearly $3.35 billion, consisting of nearly $1.13 billion in audit receivables, and over $2.22 billion in investigative receivables."

Data Mining methods are now standard tools to uncover questionable billing patterns in large databases of health claims history. Efficient fraud discovery can involve the preliminary step of deploying automated outlier detection techniques to classify identified outliers as potential fraud, before an in-depth investigation. An essential component of the outlier detection procedure is the identification of proper peer comparison groups to classify providers as within-the-norm or outliers.

However, the determination of peer comparison groups needed to identify out-of-the-norm billing behavior may not be straightforward. Even though physicians are typically clustered in their specialties (general practitioner, cardiologist, etc.) as natural peer comparison groups, the boundaries between some medical specialties are not well defined and a number of physicians may be qualified to practice in more than one medical specialty. For example, several general practitioners (GPs) render services related to the diagnosis and management of heart failure, which is also one of the domains of cardiologists.[1] Assigning physicians whose specialty is GP to a single peer comparison group will then overlook the fact that GPs rendering services related to heart failure are not strictly comparable to GPs not doing so.

This study refines the concept of peer comparison group within the category of provider and considers the possibility of the existence of distinct billing patterns associated with medical or surgical procedure codes identifiable by the Berenson-Eggers Type of System (BETOS). "The BETOS system covers all HCPCS codes (Health Care

[1] Differences between general practitioners and cardiologists in diagnosis and management of heart failure: a survey in every-day practice, F. H. Rutten, D. E. Grobbee and A. W. Hoes, The European Journal of Heart Failure 5 (2003) 337–344, http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0ahUKEwjRhuf6g4rLAhUJVD4KHSZfB-oQFgglMAE&url=http%3A%2F%2Fonlinelibrary.wiley.com%2Fdoi%2F10.1016%2FS1388-9842(03)00050-3%2Fpdf&usg=AFQjCNEF7zynyY-SKaa_0m8y2iT6kBa2Qg

Procedure Coding System); assigns a HCPCS code to only one BETOS code; consists of readily understood clinical categories; consists of categories that permit objective assignment…" (Center for Medicare and Medicaid Services, CMS).[2]

The paper focuses on the specialty General Practitioner and involves two steps: first, the identification of clusters of similar BETOS-based billing patterns; and second, the assessment of the effectiveness of these peer comparison groups in identifying outliers. The working dataset is a sample of the summary of 2013 data of physicians active in health care government programs, which was made publicly available by the CMS through its website. The website explains that "CMS has prepared a public data set, the Medicare Provider Utilization and Payment Data: Physician and Other Supplier Public Use File (Physician and Other Supplier PUF), with information on services and procedures provided to Medicare beneficiaries by physicians and other healthcare professionals. The Physician and Other Supplier PUF contains information on utilization, payment (allowed amount and Medicare payment), and submitted charges organized by National Provider Identifier (NPI), Healthcare Common Procedure Coding System (HCPCS) code, and place of service."

The analysis uses PROC FASTCLUS, the SAS® Cubic Clustering Criterion approach to find the optimal number of clusters in the data and PROC ROBUSTREG to implement a Multivariate Adaptive Threshold Outlier Detection method. The text is divided into four sections: **Identification of Relevant Billing Patterns**, which discusses the use of BETOS-based provider level patterns to identify the relevant peer comparison groups among general practitioners; **Cluster Analysis**, which describes the utilization of PROC FASTCLUS to identify both the optimal number of clusters for the data analyzed and the clusters themselves; **Performance of Pattern-based Peer Comparison Groups**, which compares the performance of pattern-based PGCs with that of a specialty-based PGC in terms of the ability to identify a smaller and more reliable set of outliers in the data; and a **Conclusion**.

IDENTIFICATION OF RELEVANT BILLING PATTERNS

In many applications of anomaly detection analysis the focus of interest is on the identification of data which fit the expected norm, making the discovery of outliers actually an ancillary tool to better define the set of within-the-norm data points. This is the case with a number of production engineering processes in which acceptance sampling defines which products meet the desired quality level. In contrast, there are a growing number of applications where the outliers themselves are the cases of primary interest. This is the case with fraud detection whose main goal is to properly identify the out-of-the-norm individuals/entities.

Improving methods used to discover questionable billing among providers rendering services to beneficiaries of Government Health Care programs on a fee-for-service (FFS) basis – a payment system where services are unbundled and paid for separately – is important for two reasons: 1) the daily operation of the programs generate massive amounts of data; 2) most health care providers comply with the norms but some do not. In statistical terminology the population of providers can be described as a mix of a *baseline distribution* of complying providers with a *contaminating distribution* of non-complying providers.

The task of looking for questionable billing in those programs is subject to the constraint of the large and steadily increasing size of the healthcare claims database and the limited amount of resources available for medical review. Handling this constraint often involves the deployment of automated anomaly detection methods that typically rank providers according to their likelihood of being involved in fraud schemes, making it possible to reduce the number of cases requiring further scrutiny. Additional scrutiny is essential because this type of analysis does not confirm that a particular provider is engaging in fraudulent or abusive practices; in fact some providers may be displaying out-of-the-norm values in payment-related and other indicators for legitimate reasons.

The characterization of providers as having questionable billing patterns requires benchmarking them against proper peer comparison groups. Standard practice in analyzing claims of FFS providers is to group them by their specialties and then proceed with the anomaly detection process. The advantage of doing this is that claims databases in general include the provider specialty as part of their stored information making the identification of the peer comparison group straightforward. As mentioned before, however, the boundaries between some specialties are not well defined and a number of physicians may be qualified to practice in more than one medical specialty. In particular, self-declared general practitioners may have subgroups characterized by medical practice that includes varied but significant shares of services that are related to other specialties such as cardiology, nephrology, orthopedics, lab

[2] In 2016 the CMS discontinued the publication of the crosswalk between HCPCS and BETOS codes. However, the identification of billing patterns can still be achieved in a way similar to what was done in this study by using the natural subdivision embedded in the Current Procedural Terminology (CPT) published by the American Medical Association. There are six major groups of codes that can be used to identify billing patterns - Evaluation and Management, Anesthesia, Surgery, Radiology, Pathology and Laboratories, and Medicine. Also the groups of codes Surgery and Medicine can be further subdivided in a small number of specialty-specific subgroups.

services, etc. Also these subgroups may have a very diverse range of utilization rates of Evaluation and Management services.

Healthcare services rendered in non-facility location (provider office, beneficiary home) under the FFS payment system generally use the Healthcare Common Procedural Coding System (HCPCS) to request reimbursement from healthcare government programs. The Level I of the HCPCS is comprised of Current Procedural Terminology (CPT-4), a five-position numeric coding system maintained by the American Medical Association (AMA).[3] Currently there are over 16,000 procedure codes but this large number notwithstanding it is still possible to identify major subsets in this set of codes that have medical similarity. A number of those subsets, roughly within the range 8-12, is enough to characterize the billing patterns of providers by their shares of services associated with them.

This study used the Berenson-Eggers Type of Service (BETOS) to identify relevant sets of codes grouped by their medical similarity. Every HCPCS was assigned to one of a set of 106 BETOS codes and a BETOS/HCPCS crosswalk was publicly available by CMS until 2015 when the agency discontinued its web-publication.[4] For example, BETOS codes having initial characters “P2” had the word “cardiovascular” in their descriptions making it straightforward to identify cardio related codes. Still the same identification of cardio-related codes could be obtained using the intrinsic organization of the Current Procedure Terminology (CPT) codebook (published by the American Medical Association, AMA) to make the proper selection of HCPCS ranges of values – for example the 2014 edition of the CPT codebook had the codes for surgery of the cardiovascular system located within the range 33010-37799 and the codes for cardiovascular medicine within the range 92920-93799. Either way, it is possible to identify about ten or twelve major sets of codes medically similar and to add a catchall group to include codes not part of any of the well-defined sets previously identified. The utilization level of these major groups of codes by providers is reflected in their shares of the total services rendered and characterizes their billing patterns.

CLUSTER ANALYSIS

The working data set for the study is a sample of summary data on FFS services rendered by general practitioners to Medicare beneficiaries. The population from which the sample was drawn is the 2013 Physician and Other Supplier (PUF) public data set, which “contains information on utilization, payment (allowed amount and Medicare payment), and submitted charges organized by National Provider Identifier (NPI), Healthcare Common Procedure Coding System (HCPCS) code, and place of service. This PUF is based on information from CMS’s National Claims History Standard Analytic Files”. The original sample included 8,975 providers and the analysis was done on a subset of 5,249 providers whose reimbursement amounts were at least $36,000 in 2013.

The study used shares of ten major sets of procedures codes – *Evaluation and Management (E and M), Cardiology (Cardio), Laboratory Services (Lab), Imaging, Pharmaceutical Drugs (Pharma Drugs), Gastroenterology (Gastro), Orthopedics (Ortho), Oncology (Onco), Nephrology (Nephro), Dermatology (Derm)* to characterize the providers’ billing patterns - plus a catchall group (labeled *Other*) of codes not included in any of those ten. All providers’ information related to those ten major sets were input variables to the cluster analysis, which used PROC FASTCLUS to determine the optimal number of clusters and to identify their membership.

The determination of the optimal number of clusters used the Cubic Clustering Criterion (CCC) method, which is one of the statistics generated by the SAS® procedure FASTCLUS. The CCC method was developed by Sarle (1983) “as a comparative measure of the deviation of the clusters from the distribution expected if data points were obtained from a uniform distribution”, and it is discussed at length in the SAS® Technical Report A-108. In the analysis of the sample drawn from the PUF public data a *list of candidates for the optimal number of clusters* was obtained by: 1) running the procedure for a sequence of predefined numbers of clusters k = 2, 3… 12; and 2) finding *the local peaks of the CCC statistic* in the plot of CCC against the number of clusters. The upper bound 12 was good enough for this study but, as a discretionary choice, it can be set to any other value depending on the overall number of items being analyzed. What follows shows the code to implement the exploratory cluster analysis using PROC FASTCLUS for different predefined numbers of clusters:

[3] Level II codes are maintained by the Center for Medicare and Medicaid Services (CMS) and used primarily to identify products, supplies, and equipment not included in the CPT codes. They have five characters - four digits followed by the letter 'F' in the last field. Level III codes are maintained by the American Medical Association and used to classify emerging technologies, services, and procedures. They also have five characters - four initial digits followed by the letter 'T' in the last field.

[4] Before the discontinuation of the publication of the BETOS/HCPCS crosswalk in 2015 the CMS described in its website the BETOS system as “developed primarily for analyzing the growth in Medicare expenditures. The coding system covers all HCPCS codes; assigns a HCPCS code to only one BETOS code; consists of readily understood clinical categories (as opposed to statistical or financial categories); consists of categories that permit objective assignment; is stable over time; and is relatively immune to minor changes in technology or practice patterns”.

Code for PROC FASTCLUS exploratory cluster analysis

```
/************************************************************************************************************************/
The input file to the clustering procedure includes the provider identifier, summary figures for six performance
indicators supplied by the Physician and Other Supplier public data set and the shares (per provider) of the ten
major sets of procedure codes that have clinical similarity; the output files, through the range of choices for the
number of clusters, add to those provider-specific variables the cluster membership at each type of analysis
(k = 2, 3,,,, 12). In addition the code generates a dataset displaying the CCC statistic for the diverse values of k.
/************************************************************************************************************************/
%macro cluster(input_file);

%do i = 2 %to 12;
proc fastclus data=&input_file out=cluster_&i
  outstat = cluster_stats_&i
  maxclusters=&i maxiter=100 noprint;
  var Share_E_and_M Share_Cardio Share_Lab Share_Img  Share_Gastro
  Share_Pharma_drugs Share_Ortho Share_Nephro Share_Onco  Share_Derm;
run;

Data cluster_stats_&i;
set cluster_stats_&i;
format max_num_clusters comma8.;
  if _TYPE_= 'CCC';
  keep max_num_clusters _TYPE_ OVER_ALL;
  max_num_clusters = &i;
run;

%if &i = 2 %then %do;
Data Cubic_clustering_crterion;
set cluster_stats_&i;
run;
%end;

%if &i > 2 %then %do;
Data Cubic_clustering_crterion;
set Cubic_clustering_crterion cluster_stats_&i;
run;
%end;
Proc SQL; Drop table cluster_stats_&i; quit;
%end;

%mend cluster;

%cluster(input_file);
/************************************************************************************************************************/
```

Table 1 displays the results of the exploratory cluster analysis, which is used here as a tool to identify the range of possibilities for the optimal number of clusters. As mentioned before the best choices are the ones in which the CCC statistic reaches local maxima. In spite of an initial declining trend as the number of clusters get higher, the statistic achieves peaks for k = 3, 5 and 7. The initial downward trend is reversed as an uptrend starts at k = 9 but the granularity of a higher partition of data offsets the benefit of having a more refined identification of peer comparison groups, so the ensuing analysis discusses billing patterns for 3, 5, and 7 clusters.

Table 1. Plot of CCC Statistic and Number of Clusters

CCC Statistic	Number of Clusters
34.9100	2
50.7392	3
28.3880	4
37.6287	5
33.5250	6
34.7285	7
31.4875	8
34.0180	9
51.8645	10
52.7835	11
53.1288	12

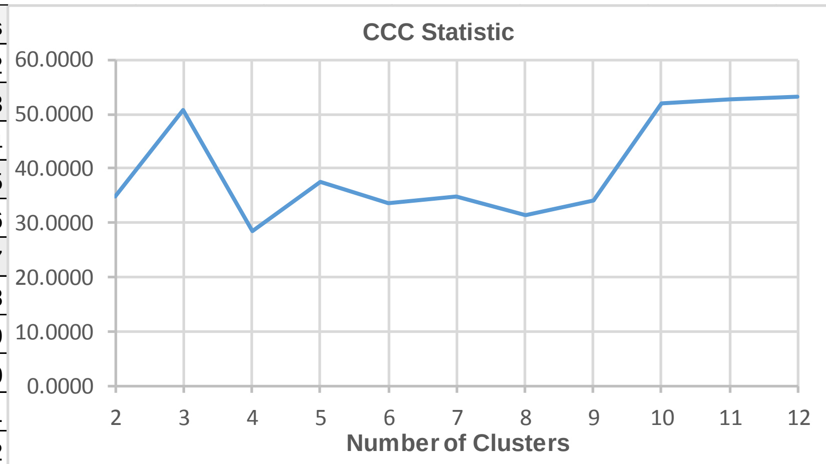

Table 2 shows the averages shares of major sets of clinically related procedure codes per cluster for different type of analysis - “No Clusters” (cell background has no fill), “3 Clusters (cells colored orange), “5 Clusters” (cells colored blue) and “7 Clusters (cells colored purple). The effect of going from “No Clusters” to other types of analysis with increasing number of clusters (3, 5 or 7) can be benchmarked by the average figures for the share of Evaluation and Management (E & M) codes. E & M codes are related to services that are associated with “1) the reason for the patient encounter or Chief Complaint; 2) an exam along with the medical history aids to determine the correct diagnosis and to devise a treatment plan; and 3) medical decision making related to the complexity of establishing a diagnosis and/or selecting a management option”.

Most general practitioners are expected to have a significant share of their services related to E & M codes and figures in Table 2 confirm that expectation. However, it is remarkable that the average share of E & M codes in the “No Clusters” type of analysis, 62.32%, is substantially distinct from the figures computed in the “3 Clusters” analysis - 47.92%, 23.13% and 94.34% (for clusters 1, 2 and 3). Also the cluster identified as overwhelmingly E & M in the type of analysis “3 Clusters” remains approximately consistent in its share and membership number in analyses “5 Clusters” and “7 Clusters” (shares = 94.34%, 96.59% and 96.41%; number of members = 2,241, 2,012 and 2,031).

The type of analysis “5 Clusters” includes in its cluster 1 a set of 47 providers having an average share of Nephrology codes of approximately 33% and identifies in its cluster 2 a set of providers having an average share of Lab codes of almost 50%. Taking the analysis to the “7 Clusters” level does not change the share figures and membership of the Nephrology-oriented set of providers (membership = 47 in both cases) and also keeps approximately constant the share figures of the Lab-leaning group of providers although the membership of cluster 2 drops from 668 to 571. It is also noteworthy that the “7 Clusters” analysis identifies a Cardiology-oriented group of providers as its cluster 3 has 92 members whose average share of Cardiology codes is approximately 41%.

The next Section compares the performance of pattern-based peer comparison groups (PCGs) with that of specialty-based PGC in terms of the ability of identifying a more reliable set of outliers in the data.

Table 2. Average Shares per Cluster for Major Sets of Clinically Related Procedure Codes

Type of Analysis	Cluster	N	E and M	Cardio	Lab	Imaging	Pharma Drugs	Gastro	Ortho	Onco	Nephro	Derm	Other
NO CLUSTERS		5,249	62.32%	4.50%	12.05%	1.50%	2.63%	0.40%	1.49%	0.08%	0.35%	0.69%	14.00%
3 CLUSTERS	1	1,860	47.92%	7.45%	9.98%	1.85%	3.97%	0.95%	1.92%	0.12%	0.72%	1.31%	23.81%
3 CLUSTERS	2	1,148	23.13%	5.53%	37.66%	3.73%	5.39%	0.18%	3.50%	0.15%	0.12%	0.89%	19.72%
3 CLUSTERS	3	2,241	94.34%	1.54%	0.66%	0.06%	0.10%	0.05%	0.09%	0.00%	0.15%	0.08%	2.92%
5 CLUSTERS	1	47	59.39%	0.29%	3.37%	0.00%	0.61%	0.00%	0.11%	0.00%	33.45%	0.00%	2.80%
5 CLUSTERS	2	668	21.12%	4.50%	48.34%	3.21%	3.70%	0.20%	2.40%	0.05%	0.02%	0.46%	15.98%
5 CLUSTERS	3	1,169	60.60%	6.04%	7.52%	0.74%	1.95%	0.91%	0.78%	0.01%	0.06%	0.94%	20.45%
5 CLUSTERS	4	1,353	33.28%	8.40%	15.86%	3.53%	6.58%	0.60%	3.83%	0.27%	0.06%	1.57%	26.02%
5 CLUSTERS	5	2,012	96.59%	1.09%	0.28%	0.04%	0.05%	0.03%	0.05%	0.00%	0.05%	0.05%	1.76%
7 CLUSTERS	1	229	21.26%	5.23%	11.44%	10.45%	13.00%	0.37%	14.38%	0.75%	0.03%	1.70%	21.38%
7 CLUSTERS	2	571	20.27%	4.39%	50.94%	3.21%	3.34%	0.20%	2.31%	0.05%	0.02%	0.42%	14.84%
7 CLUSTERS	3	92	36.11%	41.29%	7.21%	7.33%	1.89%	0.29%	0.18%	0.00%	0.00%	0.18%	5.52%
7 CLUSTERS	4	47	59.39%	0.29%	3.37%	0.00%	0.61%	0.00%	0.11%	0.00%	33.45%	0.00%	2.80%
7 CLUSTERS	5	1,164	59.91%	5.68%	7.35%	0.76%	2.10%	1.08%	0.85%	0.01%	0.04%	1.05%	21.16%
7 CLUSTERS	6	1,115	34.98%	6.49%	19.35%	1.79%	5.51%	0.47%	1.85%	0.17%	0.07%	1.49%	27.83%
7 CLUSTERS	7	2,031	96.41%	1.12%	0.30%	0.05%	0.06%	0.03%	0.06%	0.00%	0.05%	0.05%	1.87%

PERFORMANCE OF PATTERN-BASED PEER COMPARISON GROUPS

The study used information from a sample of 8,975 general practitioners drawn from the CMS[5] Physician and Other Supplier Public Use File (PUF), which has information on the provider specialty, Healthcare Common Procedure Coding System (HCPCS) codes, payment (allowed amount and Medicare payment), and submitted charges all organized by National Provider Identifier (NPI). The "PUF is based on information from CMS's National Claims History Standard Analytic Files. The data in the Physician and Other Supplier PUF covers calendar year 2013 and contains 100% final-action physician/supplier Part B non-institutional line items for the Medicare fee-for-service population."[6]

The working dataset had information on 5,249 providers whose reimbursement in 2013 was at least $36,000. The set of variables available in the dataset and considered for the multivariate outlier detection analysis included Number of services, Amount paid per service, Total payment, Number of distinct procedures and Amount paid per beneficiary/day service. The possibility of reducing the number of variables and still perform a meaningful analysis was validated by a principal component analysis (PCA) implemented with the use of PROC PRINCOMP. The PCA highlighted two indicators – Total payment and Amount paid per service – for their large role in accounting for the variability of data across providers.

The multivariate outlier detection phase of the study considered three methods: 1) the standard Mahalanobis Distance (MD), which uses the available data points to estimate the center (location) and the shape (the covariance matrix) of the multivariate distribution from which the sample being analyzed comes from and proceeds to measure the distance of each data point to the center; data points within large *shape-adjusted* distances from the center are considered outliers; 2) the Robust Distance (RD), which excludes *influential points* in its estimation of the center and shape of the population to avoid the bias brought by those points to the estimation process; 3) the Robust Distance with Adaptive Threshold (RDAT), which builds on the RD approach by classifying data points as outlier candidates using an adaptive threshold dependent on the size of the sample **n** and the number of variables **p**; the RDAT selects the squared distances that clear the adaptive threshold rather than the ones that clear a standard threshold such as the quantile of the chi-square distribution $\chi^2_{p,97.5\%}$ - as both the MD and RD methods do.

The MD method uses all the data points to estimate the center and the shape of the distribution and these estimates are subject to the pull of influential points, which are outliers whose presence may mask the existence of other outliers. It is desirable that outlier identification procedures be robust against the possible presence of m*asking effects* – cases in which outlier observations go undetected, and *swamping effects* - cases in which non-outliers are deemed outliers. The RD method improves on the MD approach because it estimates the center and shape of the population distribution by using data points that include *a subset of more representative observations*, which increases its ability to identify outliers undetected by the MD method. In this study the robust distances are computed using PROC ROBUSTREG, which implements the FAST-MCD (minimum covariance determinant) algorithm proposed by Rousseeuw and Van Driessen (1999). The FAST-MCD determines the optimal subset of sample data points of size **h < n** (sample size) to

[5] Center for Medicare and Medicaid Services, Department of Health and Human Services
[6] Source: Center for Medicare and Medicaid Services

estimate the parameters of the center and shape of the population by searching efficiently for the **h** observations that have the minimum covariance determinant among all possible distinct subsets of size **h** drawn from the **n** data points.

The option "LEVERAGE" in the MODEL statement in PROC ROBUSTREG combined with the OUTPUT options "RD" and "MD" generate the RD and MD sets of values and assign them to the chosen variable names. Although the main purpose of PROC ROBUSTREG is to provide stable results in the presence of outliers in a regression analysis context it can be used as an effective tool to implement in the SAS®/STAT environment the MCD algorithm, which is called by the function MCD in the SAS®/IML context. Since the implementation here is not actually about doing a regression analysis, the left hand side of the model can be any stochastic variable such as a random draw from a normal distribution.

The robustness of estimators – their ability to be resilient against the pull of influential but non-representative outlier observations - is often benchmarked by the breakdown point – which measures the smallest fraction of observations that need to be replaced by arbitrary values to carry the estimate beyond all bounds. Whereas Mahalanobis distances are very sensitive to the presence of influential points, robust distances - estimated using the MCD algorithm - are resilient to the level of 25% breakdown point, which means that the estimator does not generate biased estimates of the location and shape of the population as long as the fraction of outliers in the data remains below 25%. As mentioned above the population of providers can be described as a mix of a baseline distribution of complying providers with a contaminating distribution of non-complying providers. The fraction of non-complying providers is very small but non-zero, so it is likely that statistical analyses using the MD approach will underestimate the number of outlier candidates (masking effect) while analyses using the RD method will overestimate their number (swamping effect).

A compromise solution between the two alternatives is the Robust Distance with Adaptive Threshold method proposed by Gervini (2003). The method addresses the problem that robust distance approaches tend to identify more outlier candidates as data sets grow larger even if those observations are plausibly part of the baseline distribution. The basic idea is to relate the threshold to the sample size **n**. This study uses an implementation of the method discussed in Filzmoser, Garrett and Reimann (2005), who derived by simulation an empirical relationship between the threshold, the sample size **n** and the number of variables **p** in the analysis.[7] The source code to implement the approach is available as an appendix to the paper.

Table **3** displays the results of applying the three outlier detection analysis methods to the sample of 5,249 providers drawn from the CMS Physician and Other Supplier Public Use File.

Table 3. Two Types of Analysis, NO CLUSTERS and 7 CLUSTERS, and Three Outlier Detection Methods*

Type of Analysis	Cluster Number	Cluster Main Patterns	Number of Providers	Multivariate Outlier Detection Method	# of Identified Outliers	# of Outliers per Cluster identified in the NO CLUSTERS Analysis
NO CLUSTERS		E & M*** (62.3%), Other** (14%) and Lab (12%)	5,249	STD MAHAL DIST	105	
NO CLUSTERS		E & M*** (62.3%), Other** (14%) and Lab (12%)	5,249	ROBUST DIST	515	
NO CLUSTERS		E & M*** (62.3%), Other** (14%) and Lab (12%)	5,249	ADAPTIVE THRSL	450	
7 CLUSTERS	1	Other** (21.4%), E & M*** (21.3%), Ortho (14.4%) and Pharma (13%)	229	STD MAHAL DIST	4	20
7 CLUSTERS	1	Other** (21.4%), E & M*** (21.3%), Ortho (14.4%) and Pharma (13%)	229	ROBUST DIST	32	47
7 CLUSTERS	1	Other** (21.4%), E & M*** (21.3%), Ortho (14.4%) and Pharma (13%)	229	ADAPTIVE THRSL	29	45
7 CLUSTERS	2	Lab (50.9%), E & M*** (20.3%) and Other** (14.8%)	571	STD MAHAL DIST	9	19
7 CLUSTERS	2	Lab (50.9%), E & M*** (20.3%) and Other** (14.8%)	571	ROBUST DIST	60	87
7 CLUSTERS	2	Lab (50.9%), E & M*** (20.3%) and Other** (14.8%)	571	ADAPTIVE THRSL	53	78
7 CLUSTERS	3	Cardio (41.2%) and E & M*** (36.1%)	92	STD MAHAL DIST	1	4
7 CLUSTERS	3	Cardio (41.2%) and E & M*** (36.1%)	92	ROBUST DIST	8	17
7 CLUSTERS	3	Cardio (41.2%) and E & M*** (36.1%)	92	ADAPTIVE THRSL	7	14
7 CLUSTERS	4	E & M*** (59.4%) and Nephro (33.5%)	47	STD MAHAL DIST	1	1
7 CLUSTERS	4	E & M*** (59.4%) and Nephro (33.5%)	47	ROBUST DIST	2	11
7 CLUSTERS	4	E & M*** (59.4%) and Nephro (33.5%)	47	ADAPTIVE THRSL	1	10
7 CLUSTERS	5	E & M*** (59.9%) and Other** (21.1%)	1,164	STD MAHAL DIST	31	23
7 CLUSTERS	5	E & M*** (59.9%) and Other** (21.1%)	1,164	ROBUST DIST	119	138
7 CLUSTERS	5	E & M*** (59.9%) and Other** (21.1%)	1,164	ADAPTIVE THRSL	104	115
7 CLUSTERS	6	E & M*** (35%), Other** (27.8%) and Lab (19.4%)	1,115	STD MAHAL DIST	22	25
7 CLUSTERS	6	E & M*** (35%), Other** (27.8%) and Lab (19.4%)	1,115	ROBUST DIST	121	117
7 CLUSTERS	6	E & M*** (35%), Other** (27.8%) and Lab (19.4%)	1,115	ADAPTIVE THRSL	108	106
7 CLUSTERS	7	E & M*** (96.4%)	2,031	STD MAHAL DIST	55	13
7 CLUSTERS	7	E & M*** (96.4%)	2,031	ROBUST DIST	190	98
7 CLUSTERS	7	E & M*** (96.4%)	2,031	ADAPTIVE THRSL	165	82

***Data Source:** Sample of CMS Physician Public Use File **; **Other** stands for services not in any of the main identified patterns; ***E & M** stands for Evaluation and Management

[7] Macedo (2015) discusses these papers on the adaptive threshold method in detail in "Using SAS/STAT to implement a multivariate adaptive outlier detection approach to distinguish outliers from extreme values," Proceedings of the SAS Global Forum 2015

The multivariate analysis includes two variables, payment value and number of services per beneficiary, so some payment values in data points identified as outlier candidates may be close to the lower bound of their range ($36,000) in association with high number of services per beneficiary. Following the standard practice in audit of bearing in mind the notion of return on (audit) investment when selecting cases for further scrutiny – after all audit resources are limited - the figures computed in Table 3 refer to the number of outlier candidates associated with the top 50% in payment values. This kind of analysis can be seen as an automated filter that screens cases that are likely to be the best candidates for additional investigation. The choice of focusing on top values notwithstanding, additional scrutiny in actual audit practice is essential after a first automated filtering because high payment values may be legit.

Table 3 reports results according to billing patterns and two types of analysis, one involving "NO CLUSTERS" and another dividing the data in "7 CLUSTERS". Each cluster has its billing pattern characterized by a distinct set of average shares of sets of codes clinically similar in a broad sense – the top major sets of codes are reported in the column "Cluster Main Patterns".

The column "# of Identified Outliers" displays the number of candidate outliers computed using the three detection methods across all types of analysis. The Mahalanobis distances figures in the column are the lowest as compared to the other two outlier detection techniques for every type of analysis. It is likely that some potential outliers go undetected by the Mahalanobis method. In contrast the robust distance numbers consistently rank as the highest with respect to the other methods and it is possible that they are classifying some non-outliers as outliers - this could be because their 25% breakdown point hedges too much against the expected small fraction of non-complying providers. The robust distances computed with the adaptive threshold approach have more plausible numbers, not as low as the Mahalanobis method and not as high as the standard robust distance (fixed threshold). The difference between the number of outliers detected by the standard robust distance approach and those detected by the robust distance with adaptive threshold grows large as the sample grows - this is illustrated by the figures computed for clusters 6, 5 and 7.

The column "# of Outliers per Cluster identified in the NO CLUSTERS Analysis" reports the allocation of providers previously labeled by their cluster membership (1, 2, .,7) resulting from an outlier detection procedure that uses the "NO CLUSTERS" type of analysis. It is worth comparing the column's figures for two clusters with clearly distinct billing patterns - Cluster 7, which has providers rendering overwhelmingly Evaluation and Management services (96.4% on average) and Cluster 3, which has providers rendering a mix of Cardio (41.2%) and Evaluation and Management (36.1%) services. Following the NO CLUSTERS approach the adaptive threshold method was able to identify 82 members of Cluster 7 as outlier candidates (out of 5,249 overall sample size) whereas the cluster-specific figures displayed in column "# of Identified Outliers" identified 165 outlier candidates (out of 2,031 cluster members). The corresponding numbers for Cluster 3 were 14 (out of 5,249) and 7 (out of 92). These results suggest once more the importance of going beyond statistical analysis at the aggregate level whenever the identification of well-defined sublevels is possible.

CONCLUSION

This paper utilizes a sample from the 2013 Physician and Other Supplier public data set, made available by the Center for Medicare and Medicaid Services, to suggest a refinement of the standard approach in the setting of peer comparison groups for detection of questionable billing pattern. Rather than settle for the customary specialty-focused definition of peer comparison groups, it proposes the use of well-defined subspecialties, identified by grouping providers by their shares of clinically similar procedure codes.

The study first does a Cluster Analysis with PROC FASTCLUS to identify the number and membership of clusters whose members share a common billing pattern structure. Then it proceeds to assess the performance of the pattern-based peer comparison groups by means of three multivariate outlier detection methods – the Mahalanobis distance, the robust distance, and the robust distance with adaptive threshold. The results of the no-clustering analysis are distinct from the cluster-based exercise for all methods. Additionally, the counts of the cluster-labeled providers in the set of outlier candidates identified by the no-clustering analysis are quite distinct from the counts computed by the cluster-specific analyses for every cluster.

In the quest to find questionable billing patterns, the question whether it is worth going from statistical analysis at the aggregate level to analysis at well-defined sublevels was already answered by the industry practice of using specialty-based peer comparison groups rather than lumping all specialties together. This paper suggests that there is room for improvement in going one notch down by identifying similar billing pattern structures within the same specialty.

REFERENCES

Filzmoser P., Garrett R.G. and C. Reimann. 2005. "Multivariate outlier detection in exploration geochemistry", Computers & Geosciences 31 579–587

http://www.statistik.tuwien.ac.at/public/filz/papers/ArticleFGR05.pdf

Gervini D. 2003. "A robust and efficient adaptive reweighted estimator of multivariate location and scatter", Journal of Multivariate Analysis 84 116–144

https://pantherfile.uwm.edu/gervini/www/papers/JMVA2080.pdf

Macedo, P. 2015. "Using SAS/STAT to implement a multivariate adaptive outlier detection approach to distinguish outliers from extreme values." *Proceedings of the SAS Global Forum 2015 Conference*, Dallas, TX: The SAS Institute.

http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjU4_rLh7zLAhWGXB4KHWhJDD8QFggcMAA&url=http%3A%2F%2Fsupport.sas.com%2Fresources%2Fpapers%2Fproceedings15%2F3208-2015.pdf&usg=AFQjCNEQ-xWCHOfuwyYLsR_xX7JvMU7V2w&bvm=bv.116636494,d.dmo

Rousseeuw PJ, Van Driessen K. 1999. "A fast algorithm for the Minimum Covariance Determinant estimator",. Technometrics, 41:212–223

Rutten F.H., Grobbee D.E. and Hoes A.W. 2003. "Differences between general practitioners and cardiologists in diagnosis and management of heart failure: a survey in every-day practice", The European Journal of Heart Failure 5 337–344

http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0ahUKEwjRhuf6g4rLAhUJVD4KHSZfB-oQFggIMAE&url=http%3A%2F%2Fonlinelibrary.wiley.com%2Fdoi%2F10.1016%2FS1388-9842(03)00050-3%2Fpdf&usg=AFQjCNEF7zynyY-SKaa_0m8y2iT6kBa2Qg

Sarle, W. S. 1983. *Cubic Clustering Criterion*, SAS® Technical Report A-108, Cary, NC: SAS Institute Inc.
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiJkPnNl8XLAhUFHh4KHanWCksQFggcMAA&url=https%3A%2F%2Fsupport.sas.com%2Fdocumentation%2Fonlinedoc%2Fv82%2Ftechreport_a108.pdf&usg=AFQjCNGXTtcE82Gg_S9G1T3aAOnDTPldzQ

SAS Institute. "The ROBUSTREG Procedure".
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0ahUKEwjeh_P1g8jLAhWJPD4KHS3dAoYQFggiMAE&url=https%3A%2F%2Fsupport.sas.com%2Fdocumentation%2Fcdl%2Fen%2Fstatug%2F63033%2FHTML%2Fdefault%2Fviewer.htm%23rreg_toc.htm&usg=AFQjCNGwGCSKQ4W2nEX1gQFJZA12ZkklOQ

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Paulo Macedo
Enterprise: Integrity Management Services Inc.
Address: 5911 Kingstowne Village Parkway
City, State ZIP: 22315
Work Phone: (703) 683-9600 Ext. 408
E-mail: pmacedo@integritym.com