

Developing Transparent Health Care Reimbursement Auditing Procedures

Yonah Wilamowsky, Seton Hall University, USA

Aliza Rotenstein, Yeshiva University, USA

Sheldon Epstein, Seton Hall University, USA

ABSTRACT

The continued computerization of health care records has enabled easier sampling and analysis of large sets of medical records, making it easier than ever for Medicare, Medicaid and other private insurers to use statistical audits to determine and demand return of alleged overpayments to health care providers. However, there are sometimes statistical difficulties with the audits, and there is frequently not sufficient transparency in the procedures or their application to reproduce the results in order to determine whether they have been carried out correctly.

This paper addresses concerns in sampling and analysis of data records by looking at the case of a specific audit of a medical practice carried out by a private insurer. If done properly, statistical audits can be a very useful tool, but often the methodologies are vague and the implementation is either wrong or not explained fully enough to reproduce and analyze.

Keywords: Auditing; Sampling; Healthcare; Statistics

INTRODUCTION

Fraud in reimbursement requests from Medicare, Medicaid and private insurers by health care providers has been - and continues to be - a tremendous problem. In recent years, new legislation in various states has led to increased demand for return of alleged overpayments by these insurers. A controversy that is currently roiling the industry is the question of whether these audits have become overly aggressive due to political considerations, or whether they are statistically sound methods of recapturing legitimate overcharges including those that are blatantly fraudulent.¹ Questions are raised as to whether some of the claimed overcharges are in fact legitimate and, for some of those that are not, whether they are the result of fraud or just simple clerical error. To address some of these concerns, specifically in the sampling and analysis of data, this paper examines the case of a specific audit of a medical practice carried out by a private insurer.

Some health care providers claim that forms that need to be filled out have become so complex that miscoding of procedures can easily occur. With the strong encouragement of the Federal Administration, health care records are becoming computerized to a greater extent. It was hoped that such computerization would lead to more accurate diagnoses, enable a reduction in duplicate testing and procedures, and generally make for better health care at a reduced cost. However, there is some evidence that as an unintended consequence of computerization of health care records, charges to insurers have in fact increased.² Some claim that this is a result of it being easier to check boxes for procedures and examinations that have either not occurred at all or to exaggerate what has taken place. One can cut and paste from one patient to another or for a certain patient from one visit to another.³ On the other hand, others say that computerization of records has led to tests or procedures that might have otherwise been missed or forgotten, and therefore charges have legitimately increased.

¹ Under pressure, New York moves to soften tough Medicaid audits. *New York Times*, March 19, 2012, Page A1.

² Medicare bills rise as records turn electronic. *New York Times*, September 21, 2012.

³ A shortcut to wasted time. *New York Times*, November 22, 2012, Op-Ed.

A second issue involving computerization is separate and apart from the question of what types of erroneous charges should be included or categorized as fraud. Computerization has enabled easier sampling and analysis of large sets of medical records. However, the statistical procedures and their implementation by auditors have sometimes been criticized by health care providers and their advocates. See, for example, “Auditing the Auditors in Medicare and Medicaid”⁴ for a brief discussion of some statistical difficulties with audits. Additionally, there is frequently not sufficient transparency in the procedures or their application to reproduce the results in order to determine whether they have been carried out correctly. Recently, there has been increased pressure on state governments to make the procedures clearer and more transparent. For example, Connecticut passed a law in 2005 to regulate these audits which, until then, had been entirely unregulated. In 2010, the state approved legislation that gives providers the right to appeal the result of audits. It also forced the Department of Social Services (DSS)—the governmental entity that administers Medicaid in Connecticut and audits the providers—to issue regulations “to ensure the fairness of the audit process, including, but not limited to, the sampling methodologies associated with the process.”⁵ Other states have similar regulations on the books.

Methodologies for auditing health care providers typically involve sampling of charges and extrapolation of errors (overcharges or undercharges) to determine total errors in charges. Most commonly, either a single random sample is taken from the population of all charges, or a stratified sample is taken from each of several different kinds of services. The average error in the sample is found and multiplied by the total number in the population. (In the case of stratified sampling, this is done for each stratum and summed.) The health care provider is billed for the total of the overcharges found in this way. Many states have permitted the recovering of payments to insurers based, in general, on this type of methodology. If done correctly, it can be an accurate way of determining total overcharges and is significantly cheaper than having to examine all bills. The process of examining charges in bills to verify whether they are correct is time consuming and expensive. A description of a typical process can be found in a presentation by the Connecticut Department of Social Services.⁶

However, there are numerous pitfalls that must be avoided in the implementation of such a study. The number of items to be sampled in the Connecticut presentation is anywhere from 33 for a small population (up to 50 items) to a maximum of 96 for the largest population. For simplicity, auditors generally take samples of size 100, which covers any size population. In the example illustrated in the presentation, a sample of 100 items was taken from a population of 20,000. The presentation assumes that the auditor found 10 of the 100 cases considered to be incorrect with an overpayment on these 10 items of \$300. This gives an average error for the entire sample of \$3 (i.e. $\$300/100 = \3). The \$3 is then multiplied by 20,000 - the total size of the population - to produce an extrapolated audit adjustment of \$60,000. It is clear that small errors in the sampling procedure are magnified many times over in the final result. A single miscoding by the auditor can have a significant effect on the outcome. For instance, if only one of the 10 was not actually an error, then the sample average would be about \$2.70 and the adjustment should only be \$54,000 (a \$6,000 difference). If it were only a clerical error, rather than fraud, then it should not be included. This is sometimes a judgment call, and a single case can, again, make a great difference. In addition, if the sample was not taken completely randomly, then the result would have little to no relationship with reality. Further, if a confidence level in the result is not taken into account - for example, by finding a confidence interval for the mean rather than just a point estimate - then the result can be off by a significant amount. Therefore, the standard deviation of overcharges in the sample is as important as the mean in calculating the extrapolation. In cases of stratified sampling, where sample sizes are often smaller, these problems can lead to even greater errors in the result. The remainder of this paper addresses some of these concerns in sampling and analysis of data records by looking at the case of a specific audit of a medical practice carried out by a private insurer.

In July 2010, a major insurance company (insurer) informed a health practitioner (doctor) that an audit of his billing was undertaken by the insurer. It explained that it audits doctors whose billing practices were “abusive.” It determines that a particular doctor’s billing is abusive by checking which doctors’ bills are outliers with respect to the universe; that is, those whose codes are generally inconsistent with those of other doctors. Under normal circumstances, the auditors may only review billings dating back two years. However, if an abusive pattern is

⁴ Rotenstein, Epstein, & Wilamowsky, (2012). Auditing the auditors in Medicare and Medicaid. *Journal of Business Case Studies*, 8(1).

⁵ Acevedo, J. M. (2010, May 24). Service providers seek their day in court: State legislature passes law that allows challenges to Medicaid audit findings. *Connecticut Law Review*, 36(21).

⁶ https://www.nampj.org/members/2011presentations/1%20Monday%20Morning/3%20McCormick_Audit%20Prac.pdf

determined, then the audit may include billings dating back further in time. In this case, the insurance company determined that the doctor's billings were abusive and conducted an audit of the last six years of billings. A sample of 100 records was divided into three categories: new visits, consultations and established visits. The sample of 100 records included:

- 8 new visits out of a total of 818 in the universe of new visits
- 11 consultations out of 607 in the universe of consultations
- 81 established visits in the universe of 5,167 established visits

Out of the 100 in the sample, the insurer claimed that 26 were miscoded, which further indicates abusive billing practices. The insurer claimed that the eight new visits averaged an overcharge of \$28.87; consultations, \$41.74; and established visits, \$2.18. The insurer was therefore demanding a return of $818 \times \$28.87 + 607 \times \$41.74 + 5,167 \times \$2.18 = \$60,215.9$. The sample of 100 records was supplied to the doctor and the doctor was then given 45 days to dispute the findings.

The doctor asked for reconsideration and in March 2011, the insurer sent a new document with revised figures. They said that the new audit was conducted by an independent coding specialist not involved in the first review. The results of this audit were significantly different as indicated in Table 1.

Table 1: Audit Results

	New Visits	Consultations	Established Visits
Sample size	8	11	81
Universe size	818	607	5,167
Average overcharge	\$23.25	\$39.48	(\$0.47)
Extrapolation	\$19,018.5	\$23,964.36	(\$2,428.49)

The insurer now demanded \$40,554.37, or about \$20,000 less than the previous demand. The average mischarge for established visits was now negative; that is, there was an average undercharge for this stratum. Of the 26 claimed miscodings showing abusive practices, 21 were overcharges; i.e., all of the new visits and consultations as well as two of the established visits. In addition, five of the established visits were undercharges. The doctor next consulted with statisticians to help with an appeal of this extrapolation as performed. What follows is a brief analysis of the procedure. If it is representative of how these extrapolation audits are done, it has important ramifications for the industry.

As an initial comment, the fact that a second independent audit could reduce the claimed overcharges by \$20,000, or fully 33% of the total, demonstrates that there is a problem with the objectivity of the audit. Approximately 33% of the upcodings found in the first audit were changed or eliminated. At least one upcoding became a downcoding. One of the bases for showing "abusive" practices was the fact that 26% of the 100 in the sample were miscoded. Of these, five were in fact downcodings. Should these be included in showing abusive practices? Moreover, there is no stated definition of what or how many is considered "abusive." This seems to be left up to the insurer.

The most crucial issue, of course, is the sampling procedure and the analysis leading to the extrapolation. A general problem in analyzing the audit procedures of insurers is that details of the sampling procedure are typically not made available to the health care provider. The sample is not reproducible. The only way to check for randomness is by observation of the sample and by seeing if it satisfies statistical tests for randomness. In the current case, simple observation showed that the 100 records chosen in the sample included only 31 patients out of many hundreds in the population. A number of the patients had numerous visits included in the sample, some having more than ten. A simple random sample of 100 taken from the population of over 6,500 visits could not possibly have led to only 31 patients being chosen, with many of them having all their visits in the sample. When confronted with this fact, the statistician for the insurer admitted that the sample was taken in the following manner. Rather than taking a random sample of 100 from the universe of visits, a sample was chosen of the patients. Some, most, or all visit records were then used for each of the patients chosen until the total of 100 visit records was reached. This was done for the ease of sampling. A more complex analysis using cluster sampling methods might have been appropriate;

however, as shown previously, this was not done. In addition, even had cluster sampling methods been used, larger samples would have been necessary. A goodness of fit analysis of the number of records taken from each of the six years again showed that the data were not randomly chosen from the six years.

As far as sample size is concerned, as was mentioned previously, samples of size 100 are often typical and sufficient. However, in this case, the 100 was broken down into three groups - eight for new visits, 11 for consultations and 81 for established visits. While the 100 is a sufficient sample size for a single sample, 8 and 11 are not. In addition, in this case, the bulk of the 100 records is in the established visits category which, in fact, had an undercharge on the part of the doctor. Thus, the entire first \$60,000, and subsequently \$40,000, was being demanded on the basis of 19 items in the sample. These numbers are clearly insufficient to draw statistically significant conclusions using the methodology of the auditor.

Beyond this, we must look at the analysis that was done. Even if the sample size had been sufficient, one cannot draw conclusions simply by looking at the mean overcharge and multiplying by the number in the population. The confidence level that the overcharge was at the mean or greater was only 50%. In order to reclaim an alleged overcharge on the part of a health care provider, there should be a confidence level⁷ of at least 90 or 95%. That would entail finding a confidence interval and using the lower boundary as the average. Of course, the confidence interval is highly dependent on the standard deviation, which was not even considered in this case, and on the sample sizes, which were small. In any event, the use of the lower end of a 95% confidence interval would result in an extrapolation far less than the one given.

Each of these problems is the same or a variation of one that was found in the Medicare study mentioned earlier.⁸ These problems with the statistical analyses in audits would seem to be not just localized to a single case or insurer. In this case, the doctor settled for \$25,000 because the cost of legal fees to continue the appeal and the 9% yearly interest charges placed on the final decision amount deemed as an overcharge were prohibitively high. As mentioned in the *New York Times* article⁹, at one legislative hearing, Gov. Eliot Spitzer's appointee for conducting these audits was accused of "gangster tactics" for demanding that providers settle with the state or risk having to pay more based on findings extrapolated from statistical samples.

In a recent case, Administrative Law Judge John Terepka, in New York Department of Health, Appeal No. 08-3739, dated July 9, 2010, ruled in favor of Rite Aid of New York and refused to allow the Office of the Medicaid Inspector General (OMIG) to extrapolate from a sample of audited pharmacy claims. In that case, the OMIG wanted to extrapolate from a sample to a population of around 135,000 and collect nearly \$300,000. The ALJ ruled that because the OMIG refused to disclose the details of the procedure by which it had selected the allegedly random sample, it could recoup only the amount of money actually found to be invalid in the sample, i.e., \$433.06.

If done properly, statistical audits can be a very useful tool, but often the methodologies are vague and the implementation is either wrong or not explained fully enough to reproduce and analyze.

AUTHOR INFORMATION

Yonah Wilamowsky is a professor and former chairperson of Computing and Decision Sciences at the Stillman School of Business, Seton Hall University. He holds a PhD in operations research from New York University. His research and consulting interests center on applications of statistics and quantitative methods to the law, business processes, biblical studies, and higher education. Among other journals, his work has appeared in *Naval Research Logistics*, *Journal of the Operational Research Society*, *American Journal of Mathematical and Management Sciences*, *Tradition*, *Journal of College Teaching and Learning*, *Property Tax Journal*, *Location Sciences*, and *Computers and Operations Research*. E-mail: wilamoyo@shu.edu (Corresponding author)

⁷ See e.g. footnote 4 which describes the Medicare/Medicaid methodology.

⁸ See footnote 4.

⁹ See footnote 1.

Aliza Rotenstein is an assistant professor of Accounting at the Sy Syms School of Business, Yeshiva University. She holds a PhD in business/accounting from the CUNY Graduate Center and teaches courses in financial and management accounting. Her main research interests include corporate governance, accounting restatements, and earnings management, and she is currently involved in research projects relating to accounting for gift cards and accounting for non-profit organizations. Her research has appeared in *Journal of Accounting Education*, *Journal of Applied Business Research*, *Academy of Business Journal*, and *Journal of Business Case Studies*. E-mail: arotenst@yu.edu

Sheldon Epstein is a professor of Computing and Decision Sciences at the Stillman School of Business, Seton Hall University. He holds a PhD in operations research from New York University and has done extensive business consulting. His research has appeared in: *Computers and Operations Research*, *Naval Research Logistics*, *Journal of the Operational Research Society*, *American Journal of Mathematical and Management Sciences*, *Opsearch*, *Annals of the Society of Logistics Engineers*, *Journal of Property Tax Assessment and Administration*, *Property Tax Journal*, *Location Sciences*, and *Interface*. E-mail: epsteish@shu.edu

NOTES