LESLIE KISH, UNIVERSITY OF MICHIGAN
JOEL W. AGER, JR., WAYNE STATE UNIVERSITY
DAVID L. WILLIAMS, WAYNE STATE UNIVERSITY
A. DAVID BAUMHART, MAY, GOWING, MOSER AND SIMPSON

Introduction

This case exemplifies an intimate interplay between some technical issues of statistical sampling and fair judicial process through mediation. The plaintiffs comprised a group of mental health clinics who contested the reductions of prepayments by a large insurance carrier, the defendant, for services performed by the clinics. The reductions were based on estimates of total overpayment ($\hat{Y}$) found by audit teams of the carrier in ten percent samples of the bills on contracts from each clinic. The methods of sampling were valid, reasonable, and common and they yielded enough precision for the mean of all clinics. However, it was the precision of $\hat{Y}$ for each clinic and its proper estimation that became the critical and contested issue. Mediation by a judge, a psychiatrist and a statistician was agreed on. The mediators proposed a compromise based on statistical principles: proper (and higher) estimates of sampling variability, but a reduced-value for statistical confidence. The mediation was accepted, the payments made and the dispute kept out of the courts.

We first give a brief history of the origin of the suit brought by the mental health clinics against the carrier. We then go through the procedures used in the carrier's audit including the sampling and the analyses leading to the overpayment projections. Following this description the statistical issues raised by the plaintiff (i.e., the clinics) will be outlined particularly with regard to the accuracy of the overpayment estimates. Then we discuss the mediation of the dispute and indicate the statistical grounds on which the compromise was based. The general legal and statistical issues raised by the litigation are then discussed. We conclude by offering some recommendations as to how such audits might more effectively be designed and analyzed in future cases.

The Audit

Origin of the Audit

The relationship between the carrier and the clinics was based on a contractual agreement emanating from collective bargaining efforts in the automotive industry for mental health treatment and facilities. The agreement did not provide any guidelines which would be directly applicable to audits.

Certain types of counseling would be a permitted benefit at outpatient psychotherapy clinics and certain types not permitted. For instance, family counseling, marriage counseling and independent psychological testing would not be permitted benefits. Treatment for other, general mental health illnesses were permitted and, specifically, alcohol and substance abuse were treatable as a benefit.

Previous audits had been conducted by the carrier of the clinics. These audits were "educational" or instructional in nature. The audit at issues was a "financial" audit and resulted in a claim by the carrier that the clinics had been overpaid for services rendered pursuant to benefit utilization by subscribers during years 1979, 1980 and 1981.

The manner and scope in which the audit was conducted created the controversy: the clinics believed that (a) a breach of patient confidentiality had occurred as a result of the auditors review of patient records, (b) the auditors were not qualified (peer review would have been preferred) and (c) the results were inaccurate based on the carriers failure to apply appropriate, effective error standards.

The suit was started by one clinic for injunctive relief and on breach of contract theory. For that clinic the injunction was sustained. Several clinics joined this suit and legal proceedings progressed including substantial discovery of evidence from both sides to the litigation. Ultimately, the clinics requested mediation of the case by special panel and the court and carrier accepted the proposal.

How Audit was Conducted by Carrier

To provide an adequate background for our discussion, we turn now to a more detailed description of the steps which the carrier followed in conducting its audit. First, the situation provides four possible units for sampling and analysis. Each of these units, which have a hierarchical relationship, was used at one or more points during the audit process. The broadest unit is the contract which refers to those individuals covered together, typically the insured and his or her immediate family. The second unit is the case, the collection of treatments to the individual patient for a specific illness.

The next unit, the claim, is the billing unit or the group of treatments submitted to the carrier for payment at one time. Finally, the lowest unit in the hierarchy is the procedure which is the smallest unit in which the service could be measured such as a single therapy session or the administration of a psychological test.

The sampling of services provided by a clinic was done using contract as the unit. The procedure followed was to develop, using the carrier's computer records of claims paid, a list of all contracts active during the calendar year audit period. This list of active contracts served as the population, from which a 10% random sample was selected. For small clinics with fewer than 200 active contracts a larger sampling proportion was used to yield a sample size of a least 20 contracts for each clinic for each year audited.

The use of contract as the unit of sampling is sensible as it provides for the most efficient use of the auditor's time given that the clinic's records are organized by contract. Sampling by claim, for example, would require skipping around in patient charts. Since most sampled claims would occur in the middle of a course of treatment, the auditor would have to review the patient's treatment history to assess the appropriateness of the procedures being checked.

Next, for each of the contracts sampled, audit personnel conducted a complete examination of all records and medical charts maintained in the clinic's files. Thus for any contract in the sample, all cases, claims, and procedures were reviewed. The auditor disallowed any previously billed procedure which was, in his or her view, medically unjustified for the patient not within the carrier's coverage or for which documentation was inadequate to justify payment.

Following the audit, estimates were made of the total overpayments for the audit period. In contrast to the use of contracts as the unit in sampling, these projections were made using procedure as the unit of analysis. Specifically, for each category of procedure, a ratio was computed as the procedures of that type disallowed to the total of these procedures reviewed. This ratio was then applied to the total amount billed during the period for this category of procedure. Summing these amounts across all procedural categories produced the projected total overpayment to that clinic (see figure).

Figure

Formula for Total Projected Overpayment

$$\hat{Y} = \sum_{i=1}^{k} \frac{d_i}{r_i} \cdot T_i$$

$\hat{Y}$ = total projected overpayment in dollars

k = number of procedure types

d_i = number of type i procedures disallowed

r_i = number of type i procedures reviewed

T_i = amount billed of type i procedure during audit period

Arguments Raised for the Clinics

The major problem with the analysis just described leading to the estimate of overpayment was that is was based on procedures. This unit of analysis was inappropriate for several reasons. In the first place, it was contracts and not procedures (or even cases) that were originally sampled and therefore should have been used as the unit of analysis. Because there were often a large number of disallowed procedures for a given case and often several individual cases (i.e., members of the same family) nested within a contract, the procedures and cases tended to be highly clustered. Whenever a patient's record was found by the auditor to be deficient for a particular procedure it was very likely that other instances of that procedure (and other procedures) for the patient would be found to be deficient also. The resulting non-independence of the procedures would make it difficult to construct estimates of the sampling error associated with the overpayments for a given clinic and, indeed, no attempt was made by the carrier to calculate such estimates. Of course the use of procedures as the basis for calculating estimated overpayment greatly overstates the effective sample size. This was not contested in the mediation.

When the appropriate unit of analysis, i.e., the contract, is used to calculate the estimated overpayment, $\hat{Y}$, and its standard error for a given clinic, we find that for most of the clinics the standard errors are very large indeed. In the Table we give the results of our calculation of estimated overpayments and their associated standard errors for two clinics. The standard error assumes a normal sampling distribution of the average overpayment and is corrected for finite sampling. The first clinic is typical of most; the sample size is relatively small and the average overpayment is also small. As can be seen the frequency distribution of total amount disallowed per contract is highly skewed with the mode at zero. The second clinic shown was atypical. It was larger with a correspondingly larger audit sample size; the average disallowed amount per contract was also large.

In the case of the smaller clinic the standard error of the estimated overpayment was not only large but the 95% Confidence Interval based on an assumed normal sampling distribution included zero. Zero of course is an inadmissible population value and occurs not only because of the small sample size but also because of the non-normality of the sampling distribution. For the larger clinic we observe that the relative standard error is smaller. Also the sampling distribution was somewhat more symmetrical. Even here, however, it might be argued that the sampling error is still too large to ensure reasonable accuracy of the overpayment estimate. What may be viewed as reasonable accuracy is discussed later.

In addition to the large sampling error discussed above there were, potentially, other sources of error which were not evaluated or taken into account. One of these is the error due to differences between auditors. To some extent auditors may tend to systematically differ among themselves with regard to what procedures they will allow or disallow. Some auditors may be "tougher" than others. To determine the extent of such error would require estimates of inter-rater agreement. Such inter-rater agreement studies might be based on sets of auditors evaluating the same contracts and the extent of agreement assessed by the appropriate intraclass correlation statistics.

The validity of the auditor assessments themselves, though an important issue in the litigation and certainly a potentially large source of both error and bias will not be further discussed here.

Table
Frequency Distribution of Overpayments, Descriptive Statistics and
Confidence Intervals for Two Selected Clinics for Two Years in Dollars

| | Clinic One | | | | Clinic Two | | | |
| | Year 1 | | Year 2 | | Year 1 | | Year 2 | |
	X	freq	X	freq	X	freq	X	freq
	0	20	0	16	0	11	0	10
	40	1	40	2	43	4	22	1
	80	1	60	2	50	1	43	9
	120	1	125	2	83	1	49	1
	200	1	320	1	86	3	86	2
			555	1	125	3	125	9
			600	1	129	1	129	1
			880	1	172	2	211	1
					250	1	216	1
					258	1	293	1
					301	1	829	1
					337	1	1266	1
					440	1	1398	1
					552	1	4325	1
					1158	1		
					2706	1		

	Clinic One Year 1	Clinic One Year 2	Clinic Two Year 1	Clinic Two Year 2
Sample Size	24	31	34	40
Population Size	154	172	569	750
Mean Overpayment per Contract	18.33	90.48	215.09	260.55
Projected Overpayment (Y)	2823	15,563	122,385	195,412
Standard Error of Y	1404	5892	46,965	84,027
1.96 SE Y	2752	11,548	92,051	164,693
1.96 Lower Limit	-72	4015	30,334	30,719
1 SD Lower Limit	1420	9666	75,421	111,385

The Adjudication of the Dispute Through Mediation

The Mediation Procedure

Mediation, as used by the courts in Michigan, assists with the disposition of litigation. Cases are assigned by judges to mediation panels which are constituted of three lawyers. In exceptional instances, special mediation panels may be constituted. In those special mediation panels, experts in the subject matter as well as in the legal issues are generally included. To this mediation which involved health care carrier policies, outpatient psychotherapy clinics and legal issues, there was a suggestion by the parties that in addition to an attorney who was a former judge, that persons highly experienced in statistics and in outpatient psychotherapy clinic administration be included on the panel. This was done.

Attorneys for the parties prepared mediation summaries, which are statements of fact and applicable laws as well as the theories of the case. The summaries were submitted to the mediators in advance of the scheduled oral hearing/argument so that the mediators could review them and be familiar with the facts, the applicable law and the theories of the case. Oral presentation was made to the three mediators who heard argument from both sides and who inquired further of counsel and the parties' experts as to facts and applicable law as well as the theories of the case. Finally, a private discussion by the mediators was conducted and an evaluation/recommendation regarding liability, damages and/or proposed results was rendered. The mediation panel reduced its evaluation/recommendation to writing and made this available to the parties. Thereafter, the mediation evaluation was reviewed by the parties and their counsel and could be either accepted or rejected. There are court imposed sanctions or incentives for rejecting or accepting a mediation evaluation/recommendation which may include costs to the prevailing party, assuming a party has rejected, at trial.

The Panel Recommendation and Statistical Rationale

First the three mediators and the two lawyer-statistician teams quickly covered several approaches with only marginal promises. Samples large enough to yield sufficient precision would have led to 100 percent inspection within the clinics. Furthermore, the methods for measuring overpayment, though not perfect, were also judged as reasonable in the light of cost constraints. It was judged uneconomical to increase precision either with complete inspection or with more expensive measurements. Neither expense was judged justifiable.

One resolution might have been a judgment based on the average for all clinics and its standard error. Though those statistics might be sufficiently precise and unbiased for the average over all, the individual clinics would certainly have deemed it "unfair" to penalize better clinics

75

for mistakes of the poorer clinics. One may wonder if the principle of averaging or "shrinking" separate predictions would win in a court trial, but the parties in this case wanted to avoid a costly contest to find that out.

It is also possible that the appropriate but higher estimates of standard errors based on selections of clinics could also become subject to litigation with costly court proceedings, and with possible references to common use of simple random estimates by respected statisticians. In the present case this would involve the assumption of independently sampled procedures. But this time happily the clustered estimates based on contracts were accepted after only a brief disquisition.

The appropriate clustered estimates S_c were of course much larger than the simple random estimates based on procedures would have yielded. The carrier team argued that the 10% sampling used insured adequate precision in the overpayment estimates. But with the appropriate standard errors S_c in hand it became clear that the amount of overpayment could not be estimated with acceptable precision. Hence a different use of statistical inference was proposed and accepted as a reasonable compromise after discussion.

Although the existence of overpayments was established as an average over all clinics we have seen that the amount could not be accurately estimated for any given clinic. Consequently the mediators suggested that for each clinic the value of $\hat{Y} - S_c$ be computed and be designated as a likely overpayment with odds roughly 5 to 1. These values also amounted in most cases to cutting about in half the accepted overpayment.

Though not beyond "all reasonable doubt" for the individual clinics, this Solomon-like decision appeared reasonable on the cumulated evidence for the group of client clinics. It was so accepted by the board of mediation and later by both parties also.

Further Recommendations

We now add some remarks on how such audits of clinics and contracts may be better performed in the future, when such audits seem desirable. Such audits may be indicated when and where the overpayments may be suspected, perhaps with pilot tests, to be large enough to pay for the audits. Also perhaps as precautionary devices to discourage the practice of overcharging. We avoid discussing here the details of better and more costly auditing procedures and concentrate on better sampling procedures for contracts and for clinics.

A. Stratified sampling of contracts seems plausible within clinics, with oversampling of large and of suspect contracts. This would probably mean complete selections for the highest strata, plus one or two strata for other strata with lower rates.
B. Stratified sampling of clinics may also be feasible with higher selection rates for clinics which become suspect on the basis of some screening procedure. Some clinics then may be subjects of complete audits of all contracts, or with high enough selection rates to yield estimates with sufficient precision. To estimate average overpayment some selection in all strata would be needed.
C. It is also possible to search for economic sampling plans with a compromise between two views of fairness: between overall average estimates of overpayment that seems fair to the carrier and individual estimates that seems fairest to the clinic. But high precision estimates for each clinic are too costly to obtain separately for each clinic. Perhaps methods proposed for small domains may provide acceptable compromises; these exist with various names, such as "shrinking" Bayes, empirical Bayes, or Stein-James estimates (Thompson, 1968; Efron and Morris, 1973). They propose statistically designed compromises between precise overall estimates and less precise, specific, local estimates.

Conclusion

In conclusion we are happy to emphasize that a reasonable compromise was found between two conflicting views of fairness. Also that this compromise avoided a costly court trial. And most of all that sound statistical sampling principles provided the basis for the settlement.

References

Efron, B. and Morris, C., 1973. Stein's Estimation and Its Competitors - An Empirical Bayes Approach. Journal of the American Statistical Association, 68, 117-130.

Thompson, J. R., 1968. Some Shrinkage Techniques for Estimating the Mean. Journal of the American Statistical Association, 63, 113-122.