Probe Samples and the Minimum Sum Method for Medicare Fraud Investigations

Iliana Ignatova

Department of Statistics, University of South Carolina, Columbia, SC 29208, USA

iliana_ignatova@yahoo.com

Don Edwards

Department of Statistics, University of South Carolina, Columbia, SC 29208, USA

edwards@stat.sc.edu

October 16, 2006

Abstract

Random sampling of paid medicare claims has been legally acceptable for investigating suspicious billing practices by health care providers since 1986. A population of payments made to a given provider during a given time frame is isolated and a probability sample selected for investigation. A lower confidence bound for the total amount overpaid to the provider is then used as a recoupment demand. Edwards *et al.* (Health Services and Outcomes Research Methodology, 2005) show that methods based on the Central Limit Theorem can fail badly and propose an alternative method, called the minimum sum method, for fixed sample sizes. In this paper the sampling is performed in two stages. In case of little abuse in the first stage the investigation is stopped; otherwise a second sample is examined. Based on this strategy a lower confidence bound for the total number of universe payments in error and a corresponding lower bound for the total overpayment amount are defined. Criteria for choosing the sampling parameters are considered. Relative efficiencies are studied.

Keywords: medicare fraud, two-stage sampling, hypergeometric distribution, lower confidence bounds

1 Background

1.1 About Medicare Fraud

Medicare defines fraud as ”an intentional representation that an individual knows to be false or does not believe to be true and makes, knowing that the representation could result in some unauthorized benefit to himself/herself or some other person”. The violator may be a physician or other practitioner, a hospital or other institutional provider, a clinical laboratory or other supplier, an employee of any provider, a billing service, a beneficiary, a Medicare carrier employee, or any person in a position to file a claim for Medicare benefits. Fraud schemes range from those perpetrated by individuals acting alone to broad-based activities by institutions or groups of individuals, sometimes employing sophisticated telemarketing and other promotional techniques to lure consumers into serving as the unwitting tools in the schemes. In Medicare, the most common forms of fraud include billing for services not furnished; misrepresenting the diagnosis to justify payment; soliciting, offering, or receiving a kickback; unbundling or ”exploding” charges; falsifying certificates of medical necessity, plans of treatment, and medical records to justify payment; and billing for a service not furnished as billed (”upcoding”).

Since the early 1990s, health care fraud i.e., the deliberate submittal of false claims to private health insurance plans and/or tax-funded public health insurance programs such as Medicare and Medicaid, has been viewed as a serious and stillgrowing nationwide crime phenomenon, linked directly to the nation's ever-growing annual health care outlay, which in calendar-year 2003 alone amounted to $1.7 trillion (the Office of the Actuary, Centers for Medicare & Medicaid Services [1]).

It is an undisputed reality that some of the more than 4 billion health insurance benefit transactions processed in the United States every year are fraudulent. Although they constitute only a small fraction, those fraudulent claims carry a very high price tag. Each year, for example, the Office of Inspector General(OIG) of the U.S. Department of Health and Human Services (HHS) conducts a formal audit of the Medicare programs fee-for-service claim payment system. On February 21, 2002, the HHS-OIG [1] reported its finding that of the $191.8 billion such claims paid in 2001, 6.3 percent, amounting to $12.1 billion, should not have been paid due to erroneous billing or payment, inadequate provider documentation of services to back up the claims and/or outright fraud. The National Health Care Anti - Fraud Association (NHCAA) estimates that of the nation's annual health care outlay, at least 3 percent or $51 billion in calendar year 2003 is lost to outright fraud. Other estimates by government and law enforcement agencies place the loss as high as 10 percent of our annual expenditure or $ 170 billion each year [1].

1.2 Review: The Minimum Sum Method for Simple Random Samples

The standard approach to investigating suspicious billing practices by health care providers is based on statistical conclusions drawn from random sampling of paid claims. For each payment, the overpayment is defined to be the amount paid minus the amount that should have been paid, given the evidence collected by the investigator. A payment is said to be in error if the overpayment is greater than 0, and if the overpayment is the entire payment, we say "payment completely in error". The current sampling guidelines [6] prescribe a lower 90 percent confidence bound for the total universe overpayment τ as the recoupment demand from the provider under investigation. In this simple random sample setting, the standard approximate lower confidence bound L_{CLT} based on the Central Limit Theorem (CLT) approach as adapted from any standard sampling text, e.g. Cochran [4] or Scheaffer *et al.* [12] is often used.

The CLT method works well for a wide range of sample sizes and underlying populations. However Edwards *et al.* [6] show that there exist situations when the CLT method does not perform well. They give examples from motorized wheelchair claim investigations showing that the confidence level associated with L_{CLT} can be substantially smaller than 90%. Also, there is a nonzero and sometimes substantial probability that L_{CLT} exceeds the total universe payment, which renders it indefensible as a recoupment demand.

To overcome those problems, Edwards *et al.* (2005) [6] proposed a different lower bound called the minimum sum lower bound. For a simple random sample of size n, the number of sampled payments found to be completely in error Y follows a hypergeometric distribution with parameters N, N_E and n, i.e. $Y \sim \mathbf{Hyp}(N, N_E, n)$, where N_E is the number of universe payments completely in error. Thus the probability that the random variable Y assumes any given integer value i, where $0 \leq i \leq n$, and $n < N_E$ and N, is given by

$$P(Y = i) = \frac{\binom{N_E}{i}\binom{N-N_E}{n-i}}{\binom{N}{n}},$$

where $\binom{a}{b} = 0$ if $b > a$ and $\binom{a}{b} = \frac{a!}{(b!)(a-b)!}$ otherwise. The minimum sum lower confidence bound begins with a conservative lower confidence bound for N_E based on the observed error count y^*, obtained by inverting a hypothesis test for H_0: $N_E \leq N_E^0$. The bound is given by

$$L_{NE} = \min\{N_E^0 \,:\, P(Y \geq y^* \,|\, N, N_E^0, n) > \alpha\}.$$

This lower confidence bound for N_E leads immediately to the minimum sum lower confidence bound for the total overpayment τ_Y :

$$\begin{aligned} L_{MS} = &\{\textit{sum of all sample overpayments}\} + \\ &\{\textit{sum of the smallest } L_{NE} - y^* \textit{ nonsampled population payments}\}. \end{aligned}$$

The bound L_{MS} always assumes a sensible value, in the sense that it is always at least as large as the total of all sample overpayments and it cannot be larger than the total of all universe payments. Edwards *et al.* (2005) [6] studied the performance of L_{CLT}, L_{N_E} and L_{MS} on four real payment populations. Their new method provides more than the nominal coverage probability for lower confidence bounds regardless of sample size, and is surprisingly efficient (in terms of recovering overpayment) relative to the CLT bounds in settings where overpayments are essentially all-or-nothing and where the payment population is relatively homogeneous and well separated from zero.

1.3 Motivation for a Two-Stage Procedure

Because auditing is expensive and time consuming, the auditor may first examine a small pilot or probe simple random sample of size n_1. If the probe sample shows little abuse, then it will be best to stop the sampling procedure and the entire investigation. Based on the probe sample, if the number of payments completely in error Y_1 is large enough, say $Y_1 > m_1$, a second simple random sample of size n_2 is selected from the remaining payments. In this paper we discuss lower confidence bounds for the total universe overpayment using the information from both samples.

Section 2 provides the basic theory underlying the two-stage procedure and definitions of the minimum sum lower bounds associated with that procedure. In section 3 we discuss different approaches for proper selection of the overall sample size when using the two-stage procedure. Section 4 provides simulation results and gives general observations on choices for the parameters m_1 and n_1.

2 Theory Underlying the Two-Stage Sampling Procedure

Table 1 shows a summary and notation for this discussion. The mathematical structure we use follows that of Cohen and Sackrowitz (1996). We begin our procedure by taking a probe simple random sample of size n_1 from the finite population of payments of size N. Given that the event $A = \{Y_1 > m_1\}$ occurs we examine a second simple random sample of size n_2. Based on the total number of payments completely in error $Y = Y_1 + Y_2$ from both samples, and given that event A occurs, we find a lower confidence bound $L_{NE,2S}$ for the total number of population payments completely in error N_E such that the conditional probability of coverage is of at least $1 - \alpha$, i.e.

$$P\{L_{NE,2S} \leq N_E | A\} \geq 1 - \alpha.$$

From now we will assume that payments in error are completely so. The case of partial overpayments is discussed in the final section. As Y_1 is the total number of payments in error in the sample of n_1 elements and N is our population size, we have $Y_1 \sim \mathbf{Hyp}(N, N_E, n_1)$. Further, let Y_2 be the number of payments in error in the second sample given that we have taken the first sample. The conditional distribution of $Y_2|Y_1 = y_1$ is $\mathbf{Hyp}(N - n_1, N_E - y_1, n_2)$. Thus, given that the event $A = \{Y_1 > m_1\}$ occurs in the first sample, the conditional density function of the total number of payments in error Y is

$$P(Y{=}y|A) = \frac{P(\{Y{=}y\} \cap A)}{P(A)} = \frac{P(\{Y{=}y\} \cap \{Y_1{>}m_1\})}{P(Y_1{>}m_1)}$$

$$= \frac{\sum_{j>m_1} P(Y{=}y, Y_1{=}j)}{P(Y_1{>}m_1)} = \frac{\sum_{j>m_1} P(Y_1{+}Y_2{=}y, Y_1{=}j)}{P(Y_1{>}m_1)}$$

$$= \frac{\sum_{j>m_1} P(Y_1 + Y_2{=}y|Y_1{=}j) \cdot P(Y_1{=}j)}{P(Y_1{>}m_1)} = \frac{\sum_{j\geq m_1+1} P(Y_2{=}y{-}j|Y_1{=}j) \cdot P(Y_1{=}j)}{\sum_{j>m_1} P(Y_1{=}j)}. \tag{2.1}$$

The final expression for (2.1) does not simplify, but is not difficult to calculate for any given values of the parameters N, N_E, n_1, m_1 and n_2. A program to evaluate (2.1) in the shareware package R (`www.R-project.com`) is available from the first author.

For specified α, $0 < \alpha < 1$, and any hypothesized value of N_E, say N_E^0, we can define a level-α test of the null hypothesis $H_0 : N_E \leq N_E^0$ versus the alternative hypothesis $H_A : N_E > N_E^0$ to reject H_0 for large values of Y. For an observed sample value y^*, the conditional significance level ('P-value') for the test is

$$P = P(Y \geq y^* \mid A).$$

Given N, n_1, m_1, n_2, α and the observed number of payments completely in error y^*, a confidence region for N_E with confidence coefficient at least $1 - \alpha$ is given by the set of all values of N_E^0 which would not be rejected at level α by the hypothesis test. That is, a $(1 - \alpha) \times 100\%$ confidence region is given by

$$C(y^*) = \{N_E^0 \,:\, P(Y \geq y^* \mid A) > \alpha\}.$$

Thus the desired confidence set is given by the lower confidence bound:

$$L_{NE,2S} = \min\{N_E^0 \,:\, P(Y \geq y^* \mid A) > \alpha\}. \tag{2.2}$$

A program in R computing this lower bound is also available. The corresponding minimum sum lower bound $L_{MS,2S}$ for the total overpayment is

$$\begin{aligned} L_{MS,2S} = &\{\textit{sum of all sample overpayments}\} + \\ &\{\textit{sum of the smallest } L_{NE,2S} - y^* \textit{ nonsampled population payments}\}. \end{aligned} \tag{2.3}$$

3 Criteria for Choosing n_T

In this section we focus on the equal payments setting because it is an important special case and one in which the CLT method does not perform well. Our results here are also completely relevant to the case where payments are approximately equal, for example the frequently encountered case where payments to be sampled all have the same procedure code. Let X be the total amount of payment per claim and let C be the typical cost of investigation per claim. The cost of the pilot sample is n_1C; if a second sample is examined, its cost is n_2C and the recoupment amount is $XL_{NE,2S}$. Thus, the monetary gain for the investigation is

$$G = \mathbf{1}\{Y_1 > m_1\}(X\, L_{NE,2S} - n_2\, C) - n_1\, C \tag{3.1}$$

and the corresponding standardized gain is

$$g = \frac{G}{C} = \mathbf{1}\{Y_1 > m_1\}(\frac{X}{C} L_{NE,2S} - n_2) - n_1. \tag{3.2}$$

The population proportion of fraudulent claims is denoted by $\pi = \frac{N_E}{N}$, with N_E being the population number of payments completely in error. We begin with methods for choosing an appropriate total sample size n_T making the following observations:

(1) If $\pi = 1$ then with probability one $L_{NE,2S}\big|_{\pi=1} = L_{NE}\big|_{\pi=1} = L = \min\{N_E^0 : P(Y = n_T) > \alpha\}$, with $Y \sim \mathbf{Hyp}(N, N_E^0, n_T)$.

(2) If in addition the population size $N \longrightarrow \infty$ and n_T is held constant then $\frac{L}{N} \longrightarrow \alpha^{\frac{1}{n_T}}$ where α is the confidence level.

To see that (1) holds note that when $\pi = 1, Y_1{=}n_1{>}m_1$ with probability 1 and thus $P(Y_1{>}m_1) = 1$. Also, $Y_2{=}n_2$, the observed total number of payments completely in error is n_T and the set $\{Y_1{+}Y_2{\geq}y^*\}$ in the expression 2.2 for $L_{NE,2S}$ (with $Y = Y_1 + Y_2$) becomes $\{Y_1 + Y_2 = n_T\}$ and $\{Y_1 > m_1\} \supset \{Y_1 + Y_2 = n_T\}$. Therefore

$$\begin{aligned} P(Y{=}n_T|Y_1{>}m_1) &= \frac{P(\{Y{=}n_T\} \cap \{Y_1{>}m_1\})}{P(Y_1{>}m_1)} \\ &= P(Y{=}n_T). \end{aligned} \tag{3.3}$$

The last implies that $L_{NE,2S}\big|_{\pi=1} = \min\{N_E^0 \,:\, P(Y = n_T \,|\, A) > \alpha\} = \min\{N_E^0 \,:\, P(Y = n_T) > \alpha\} = L_{NE}\big|_{\pi=1}$, which for simplicity we denote by L.

To see (2) we just use the fact that $Y \sim \mathbf{Hyp}(N, N_E, n_T)$. Thus

$$\begin{aligned} P(Y{=}n_T) &= \frac{\binom{N_E}{n_T}\binom{N-N_E}{n_T-n_T}}{\binom{N}{n_T}} \\ &= \frac{N_E}{N}\frac{N_E-1}{N-1}\cdots\frac{N_E-n_T+1}{N-n_T+1} \\ &\approx \left(\frac{N_E}{N}\right)^{n_T}, \end{aligned} \tag{3.4}$$

as $N \longrightarrow \infty$. Therefore,

$$\begin{aligned} L &= \min\{N_E^0 \,:\, P(Y = n_T) > \alpha\} \\ &\approx \min\{N_E^0 \,:\, \left(\frac{N_E^0}{N}\right)^{n_T} > \alpha\} \\ &\approx \min\{N_E^0 \,:\, N_E^0 > N\alpha^{\frac{1}{n_T}}\}, \end{aligned} \tag{3.5}$$

which yields $\frac{L}{N} \longrightarrow \alpha^{\frac{1}{n_T}}$.

The observations from (1) and (2) suggest a useful common-sense way to choose the overall sample size n_T. The idea is to choose n_T to recoup a specified percentage of the amount overpaid to a fraudulent provider ($\pi \approx 1$). With $\alpha = 0.1$ a fixed governmental standard [2], if we wish to obtain a 90% recoupment, solving $\frac{N\alpha^{\frac{1}{n_T}}}{N} = 0.9$ we should choose a total sample of size $n_T = 22$. Similarly, a total sample size of $n_T = 45$ should give a 95% return. These calculations apply for highly abusive providers in the nearly-equal-payments case using either the minimum sum method or the two-stage minimum sum method.

Though the above method for choosing n_T to provide a specified percentage return has great appeal, it does not depend on the sampling cost -to-payment ratio $\frac{C}{X}$ in the gain function g defined by (3.2). In particular, for large $\frac{C}{X}$ values the cost of the investigation could exceed any possible return. Because of this we propose a second method for choosing the parameter n_T. This method is constructed to provide maximal gain g when the proportion of unjustified payments $\pi = 1$. By (1) and (2) we have $L \approx N\alpha^{\frac{1}{n_T}}$. Therefore, the standardized gain for the investigation

$$g = \frac{G}{C} = \mathbf{1}\{Y_1 > m_1\}(\frac{X}{C}L_{NE,2S} - n_2) - n_1$$

at $\pi = 1$ becomes

$$\begin{aligned} g(1) &= \frac{X}{C}L - n_T \\ &\approx \frac{X}{C}N\alpha^{\frac{1}{n_T}} - n_T \end{aligned} \tag{3.6}$$

Maximizing $g(1)$ over n_T and choosing $\alpha = 0.1$ as before, we obtain an optimal n_T

depending on the population size N and the ratio $\frac{C}{X}$. This maximization can be done numerically yielding the following results:

$N = 300$:

C/X	1	2	3	4	5	6	7	8	9	$\geq$10
n_T	26	18	14	12	11	10	9	9	8	$\leq$8

$N = 1000$:

C/X	1	2	3	4	5	6	7	8	9	$\geq$10
n_T	47	33	27	23	21	19	17	16	15	$\leq$14

It is interesting to observe that relatively modest total sample sizes, on the order of the value 30 often used in practice, are again recommended using the minimum sum or two-stage minimum sum methods for highly abusive providers when payment amounts are nearly equal.

The actual value of C and hence $\frac{C}{X}$ is likely larger than one might initially guess, because it is not simply the money spent to investigate a claim, but more importantly the investigator's time spent. Most investigations involve truly abusive providers and hence consistently large recoupment amounts. So, there is at the least high monetary cost (lost recoupment opportunity) when an investigator spends time investigating a non-abusive provider.

To support the claim that $\frac{C}{X}$ is not small, we note that a current typical strategy is to choose $n_T = 30$, $n_1 = 10$ and $m_1 = 3$. For $N = 300$ this leads to the following heuristic argument that $\frac{C}{X} \approx 3$. We have shown that $L_{NE,2S}\big|_{\pi=1} = L \approx N\alpha^{\frac{1}{n_T}}$. If we approximate the expected value of $L_{NE,2S}$ at other π by the linear function $\pi N\alpha^{\frac{1}{n_T}}$ the expected gain function would be $g \approx \frac{X}{C}\pi N\alpha^{\frac{1}{n_T}} - n_T$. Considering the π needed to "break even" on average, solving the equation $g = 0$ with respect to $\frac{C}{X}$, we obtain $\frac{C}{X} = \pi\frac{N}{n_T}\alpha^{\frac{1}{n_T}} \approx 3$, if we interpret the typical strategy to imply that important π values to detect are close to $\frac{m_1}{n_1} = 0.3$ or above. Generalizations of this argument suggest that reasonable values for $\frac{C}{X}$ are between 1 and 9. If $\frac{C}{X}$ is much larger than this we should not sample at all, since the gain g defined by (3.2) then becomes negative for most, if not all, widely encountered values of π, N and n_T. The simulation studies in the next section support this discussion.

4 Performance Studies and Choice of n_1 and m_1

When planning a particular sample we will know the population size N, and $\alpha = 0.1$ is given as well. So we can choose the appropriate total sample size n_T by a method from section 3. We can thus explore, via Monte Carlo simulations, the effects of choices of n_1 and m_1 on $L_{NE,2S}$ versus a range of $\frac{C}{X}$ values and π values. In this section we report some simulation studies of this sort and make general observations on choices of n_1 and m_1 suggested by these.

We performed a simulation study for 16 different cases, based on the following settings: $N = 300, 1000$; $n_T = 30, 100$; $n_1 = 5, 10$; $C/X = 2, 4$. For each case we consider values of $\pi \in [0, 1]$ in increments of 0.1, and consider all choices of m_1. Each simulation is of size 5000.

We first comment in detail on the average gain operating characteristics under the cases $N = 300$, $n_T = 30$, $n_1 = 5, 10$ and $C/X = 2$. We also include the gain function for the fixed sample procedure. The solid line in Figure 1 represents the average gain for the fixed sample procedure, which could be considered a special case of the two-stage procedure with $m_1 = -1$. This gain is nearly linear in π and the greatest loss $-n_T$ is at $\pi = 0$. We note also that as provided in Section 3, when π approaches 1 the gain approaches $\frac{X}{C} N \alpha^{\frac{1}{n_T}} - n_T$. Figure 1 also provides the average gain for the two stage procedure at several different values for the threshold m_1. At $m_1 = 0$ it can be seen that the two-stage method gives smaller loss than the fixed sample procedure when π is small; for moderate and large π the two curves coincide. Looking at the two-stage procedure when $m_1 = 1$, we see that for small π the loss is still smaller, and the curve is slightly more convex than the previous two for moderate values of π, but still very close to both of them. Similar changes in the gain function are observed for the next three values of m_1 $(3, 4, 5)$. A desirable choice of m_1 would feature small loss ($g \approx -n_1$) for small π and average gain close to that of the fixed sample procedure for the rest of the cases. For this reasons we would recommend

$m_1 = 1$ or $m_1 = 2$ for the setting in Figure 1 ($N = 300$, $n_T = 30$, $n_1 = 10$, $\frac{C}{X} = 2$).

Figure 2 represents a similar setting as in Figure 1 with the only difference being that the pilot sample is of size $n_1 = 5$. The fixed sample gain curve is identical to that of Figure 1. The shape/behavior of the other curves as m_1 grows from 0 to 4 is similar to the pattern observed in the previous plot; however for small π, the loss from the two-stage is smaller than the one with $n_1 = 10$. Choosing $m_1 = 1$, we notice that as π increases the two-stage procedure has comparable gain to the fixed sample procedure and for highly abusive providers ($\pi > 0.8$), the two methods essentially perform equally well.

Similar results to those observed in Figures 1 and 2 follow when $N = 1000$, $C/X = 4$ and $n_T = 100$. The simulation study implications are thus that we should choose a small pilot sample size, say $n_1 = 5$, when using the two-stage minimum sum method. The threshold m_1 should also be small. The performance of any particular plan can always be investigated via Monte Carlo; a program in R to do this is also available from the first author.

5 Discussion

We have proposed a two-stage minimum sum method and have considered optimal choices for the design parameters. Relative efficiencies with the fixed sample procedure suggest that for low denial rates π the two-stage procedure gives smaller loss than the fixed sample method. For moderate π the two procedures perform equally well, as long as the threshold m_1 is chosen wisely. In the case of highly abusive providers the two methods perform equally well.

It may seem that the new method can be applied only to the "all or nothing" payments situation. Following Edwards *et al.* (2005) [6] an adjustment to the partial overpayment case for the two-stage method may be achieved by a generalized approach in which one may consider/define a sample payment to be "seriously in error" if the overpayment/payment ratio is at least as large as some pre-specified proportion q, $0 < q < 1$. For example, we may consider a payment to be seriously in error if it is at least 80% ($q = 0.8$) in error. In any case, Y_1, Y_2 and $Y = Y_1 + Y_2$ then count the number of payments found to be seriously in error and the bound $L_{NE,2S}$ given by (2.2) is a lower confidence bound for the number of universe payments seriously in error. The corresponding two-stage minimum sum lower bound would be generalized to a two-stage "q-minimum" sum bound, defined by equation (2.3) with the second bracketed term multiplied by q. This will again give a valid lower confidence bound for the total population overpayment, though it may sacrifice some recoupment. This q-minimum sum method would be particularly appropriate for providers who consistently "upcode", i.e. perform one service but bill for a much more expensive service.

We have seen that the loss of the investigation process when the proportion of overpayments π is small is reduced when we use a two-stage sample with small pilot sample size. Thus a natural extension to the proposed method would be to consider a sequential sampling plan, or a sequential first-stage followed by a "jump" to a full sample. Under such a scenario, one would examine the claims one by one until either a specified number of justified claims are observed or all n_T sample claims are examined. Lower confidence bounds for total overpayment under sequential and/or partially sequential sampling plans are currently under study by the authors.

Acknowledgement

The authors wish to thank Petko Kostadinov, Jane Naig, Jody Jaco, and Robert D'Zio for advice during the development of these methods.

References

[1] National Health Care Anti-Fraud Association(NHCAA). Health care fraud: A serious and costly reality for all americans. http://www.nhcaa.org/content/Files/HealthCareFraudArticle2005.pdf, 2005.

[2] Program Memorandum Transmittal B-01-01. Use of statistical sampling for overpayment estimation when performing administrative reviews of part b claims. Technical report, United State Department of Health and Human Services, 2001.

[3] G. Casella and R. Burger. *Statistical Inference*. Duxbary, Pacific Grove, CA, 2nd edition, 2002.

[4] W.G. Cochran. *Sampling Techniques*. John Wiley and Sons, New York, 3rd edition, 1977.

[5] Arthur Cohen and H.B. Sackrowitz. Lower confidence bounds using pilot samples with an application to auditing. *Journal of the American Statistical Association*, 91(433):338–342, March 1996.

[6] Don Edwards, Gail Ward-Besser, Jennifer Lasecki, Brenda Parker, Kristin Wieduwilt, Fuming Wu, and Philip Moorhead. The minimum sum method: A distribution-free sampling procedure for medicare fraud investigation. *Health Services & Outcomes Research Methodology*, 4:241 – 263, 2005.

[7] Center for Medicare, U.S. Dept. of Health Medicaid Services, Program Memorandum Carriers Change Request 1363. Office of the Inspector General, and Human Services. Medicare definition of fraud, 2001.

[8] William C. Guenther. The average sample number for truncated double sample attribute plans. *Journal of the American Statistical Association*, 13(4):811–816, Nov 1971.

[9] Norman L. Johnson and Samuel Kotz. *Discrete Distributions.* Noughton Mifflin Company, Boston, 1969.

[10] Norman L. Johnson, Samuel Kotz, and Kemp Adrienne W. *Univariate Discrete Distributions.* John Wiley & Sons Inc., second edition, 1992.

[11] RAT-STATS program Office of the Inspector General(HHS). http://www.oig.hhs.gov/organization/oas/ratstat.html, 2003.

[12] R.E. Scheaffer, W. Mendenhall, and L. Ott. *Elementary Servey Sampling.* Duxbury, 4th edition, 1990.

Table 1: Notation

Notation	Meaning
N	finite population (universe) size
N_E	total number of payments completely in error in the population
n_1	first stage (pilot sample) size
n_2	second stage sample size
n_T	overall sample size; $n_T = n_1 + n_2$
Y_1	number of payments completely in error in the first sample
Y_2	number of payments completely in error in the second sample
Y	total number of sample payments completely in error, $Y = Y_1 + Y_2$
m_1	threshold value for drawing a second sample, $0 \leq m_1 \leq n_1$
A	the event that a second sample is taken, i.e. $\{Y_1 > m_1\}$
y^*	observed payments completely in error from the entire sample of size $n_T = n_1 + n_2$
α	lower bound error rate
$L_{NE,2S}$	the $(1-\alpha) \times 100\%$ lower confidence bound for N_E based on the two-stage procedure
$L_{MS,2S}$	the minimum sum $(1-\alpha) \times 100\%$ lower confidence bound for total overpayment based on the two-stage procedure

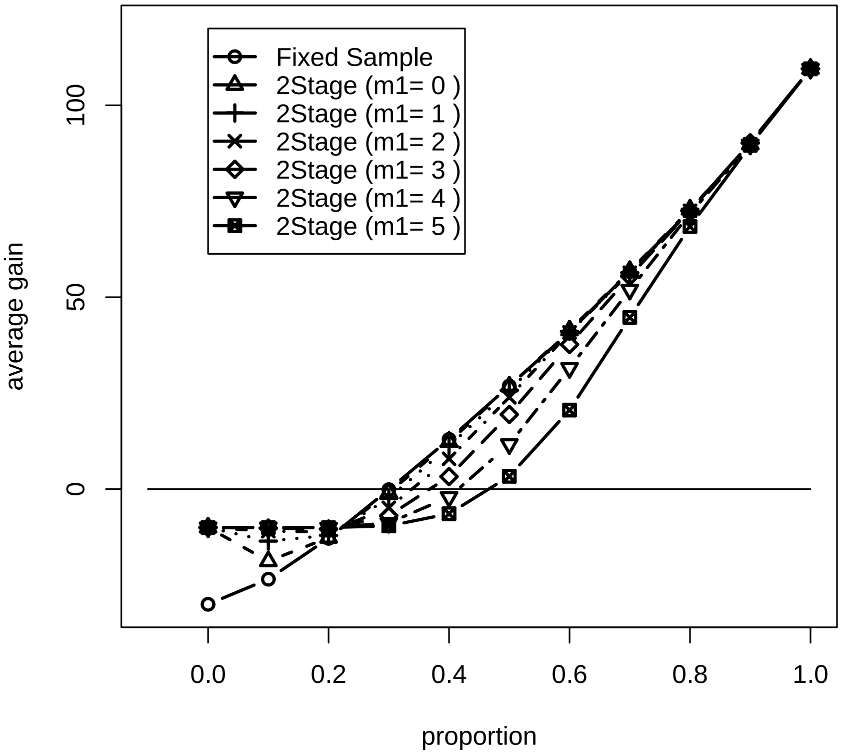

Figure 1: Average gain curves for the fixed sample and the two-stage procedures for the case when $N = 300$, $n_T = 30$, $n1 = 10$, $C/X = 2$ and threshold values of $m_1 = 0, 1, 2, 3, 4, 5$ based on 5000 repetitions with $\pi \in [0, 1]$ in increments of 0.1.

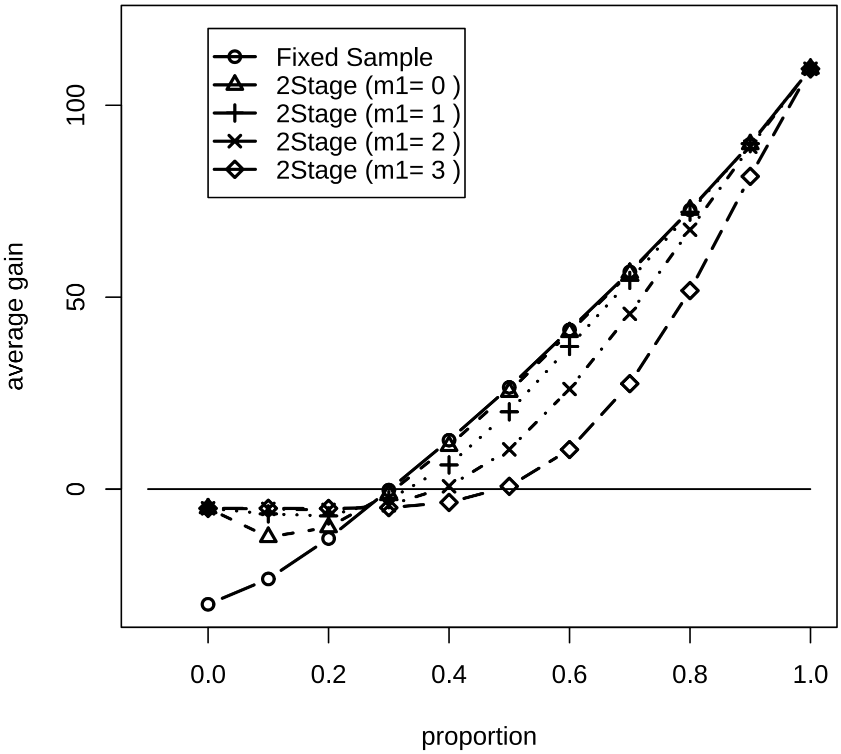

Figure 2: Average gain curves for the fixed sample and the two-stage procedures for the case when $N = 300$, $n_T = 30$, $n1 = 5$, $C/X = 2$ and threshold values of $m_1 = 0, 1, 2, 3$, based on 5000 repetitions with $\pi \in [0, 1]$ in increments of 0.1.