Review of Sampling and Extrapolation Methodologies, Early and Periodic Screening, Diagnosis and Treatment Claims Audits

Prepared for the California Department of Mental Health, Medi-Cal, Epidemiology, and Forecasting Unit

October 2006

By
Randall MacIntosh, Ph.D., Principal Investigator
Senior Faculty Associate &
Professor of Sociology

Institute for Social Research
California State University, Sacramento

Table of Contents

Page

Introduction

This report is prepared at the request of the California Department of Mental Health, Medi-Cal, Epidemiology, and Forecasting unit on the basis of recent legislation. The DMH requested that the review cover:

- The statistical validity of the proposed sampling methodology that addresses the issues of stratified random selection, confidence intervals and related parameters, and sample size;
- the statistical validity of the extrapolation methodology;
- the adequacy of the statistical software programs used for the above processes;
- the consequences to the State and Service Providers due to sampling error; and,
- suggestions for improving the sampling and extrapolation methodologies.

On the basis of that scope of work, interviews were conducted with DMH staff and representatives of providers. Ten technical issues were identified in those conversations. We have sought to explain and evaluate these issues to promote a constructive dialogue between providers and the agency without an overwhelming morass of technical details.
As this is a technical report, however, we have laid out some procedural alternatives for policymakers to consider. We have sought to make it explicit in several places that we wish to avoid making any policy recommendations.

General conclusion

Our general conclusion is that most aspects of the sampling as implemented are at random and scientifically defensible. The precision of the estimates generated by the process, however, appears to be poor. We offer several recommendations for improvement.

Primary recommendations

Our primary recommendation is for policymakers to consider establishing an acceptable precision range. Sample size would then be determined by the level of precision desired for recoupment amounts. Samples could be drawn in an iterative process until a desired precision is achieved. Under the present system, the sample size is fixed for each audit, regardless of the precision it affords.

We also recommend the use of stratification on the basis of dollar amounts. Similarly, if there are several sites that deliver service to different types of clients, it may be desirable to stratify on the basis of location, as well. Other stratification variables may also be tested to reduce sampling error

We also recommend that sample characteristics be compared to population values as part of an on-going monitoring of the sample selection process. Policymakers may consider reporting these values to other stakeholders to promote dialogue and transparency.

It may be more appropriate to use a variable sample calculator, such as the one found in RAT-STATS. An alternative is to use the formulae found in statistical auditing textbooks. (References are provided on page 14.) These computations are straightforward and can be carried out in a spreadsheet.

Issues for Review

Issue 1: ***Does sampling work?***

Most modern science relies on the use of random samples as a means of estimating unknown values in a population. It is possible, for example, to tell what's going on in your body from a few milliliters of blood. It is not necessary to analyze the entire 5 liters of blood that we carry within us. That's because the small sample is representative of the entire supply from which it was randomly selected.

There are many mathematical proofs that demonstrate that the mean (or average) from a properly drawn random sample will be a close estimate of the population average. While the sample mean will be close, it will not be precisely the population value. The discrepancy between the population average and the sample average is because of "sampling error."

Issue 2: ***What is sampling error and what can be done to reduce it?***

Sampling error is the difference that arises when using a sample to estimate an unknown population characteristic. It is, for example, the difference between the population mean and the sample mean. Several factors determine the amount of error present in a sample. One factor is the amount of variation (or differences in magnitudes) in the population. A second factor is the size of the sample.

Sample size can be adjusted to reduce sampling error and obtain a desired level of precision. There is, of course, a trade-off in the cost of the study when sample size is increased. And the increase in precision obtained from a larger sample is not a straight line. In other words, increasing the sample size will only reduce the error up to a certain point. After that point, the increased cost of additional samples greatly exceeds the amount of additional information they would provide. This is shown below in Figure 1.

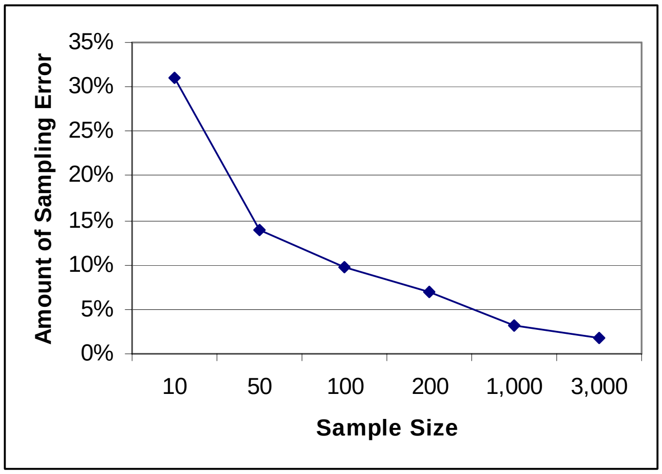


Fig. 1. Relationship between sample size and sampling error

There is a given amount of variation in the population and this cannot be manipulated. There are, however, sampling strategies that can be applied to address this problem. If there is a large amount of variation in the population, it is termed "heterogeneous." Breaking the population into more homogeneous sub-groups is called "stratifying." Sub-samples are then drawn from each stratum (group). This strategy typically results in a substantial reduction in sampling error. If there is a large variation in the dollar value of claims, for example, it might be advisable to stratify on this basis and draw proportionate sub-samples from a pool of small claims and a pool of large claims. Similarly, if there are several sites that deliver service to different types of clients, it may be desirable to stratify on the basis of location.

This idea has already been implemented, in part, by stratifying by service type and restricting most audits to one type. This eliminates the heterogeneity with regard to service function. As outlined above, other stratification variables also may be tested to reduce sampling error.

Issue 3: ***How much sampling error is present under the current process?***

An analysis of a recently completed audit of a legal entity (labeled for purposes of this report LE00A) provides an illustration of sampling error. As shown in Table 1, the mean value of the universe (population) of all 5,839 claims filed within the mental health services category is $272.86. That is to be contrasted with the mean of $296.75 derived from a sample of 190 claims. The sample mean overstates the actual population mean by approximately $24. As will be shown later, this sampling error has implications for extrapolation.

Claims	*Mean*
Claims universe (population)	$272.86
Total sample of 190 claims	$296.75
Disallowed claims only	$323.80

Table 1. Averages (means) from the LE00A audit.

Two points should be noted. First, if the process was repeated and another audit sample drawn, it would be expected to be *as likely to under-estimate* the population mean by a similar amount. Additional study would be required to determine of this over-estimate is simply an unlucky occurrence for the legal entity[1] because of random error, or a systematic defect. The analysis of one audit sample is insufficient to draw any conclusion with regard to the validity of the process.

Table 1 also shows the average amount for the eight disallowed claims ($323.80) is approximately $50 greater than the population average. Additional study would also be necessary to reach any conclusion as to whether auditors are giving greater scrutiny to larger claims or whether larger claims are more complex and therefore more error-prone.

As the three means shown above are simple to obtain, DMH may wish to consider regular reporting of this information on an audit-by-audit basis to stakeholders to facilitate on-going monitoring of the process.

Table 2 shows the values for the eight disallowed claims. There is substantial variation in these amounts. As will be demonstrated later, this wide variation has an important impact on extrapolation precision

Disallowed claims			
1.	$818.12	5.	$81.36
2.	$649.80	6.	$54.72
3.	$576.30	7.	$41.04
4.	$335.16	8.	$33.90

Table 2. Values of the eight disallowed claims

[1] A “legal entity” is a county or a corporation that is the subject of the audit.

The population and sample means for upcoming audits are shown in Table 3.

Audit	*Sample Size*	*LE Claims Population Average*	*LE Claims Sample Average*	*Difference*	*Coefficient of Variation*[1]
1	194	$ 92.35	$ 89.27	-$3.08	74%
2	191	$118.94	$118.94	$0.00	0%
3	192	$203.93	$192.30	-$11.63	73%
4	18	$149.76	$149.76	$0.00	61%
5	194	$111.50	$108.92	-$2.58	52%
6	195	$156.93	$145.60	-$11.33	72%
7	191	$141.92	$141.84	-$0.08	54%
8	195	$ 81.14	$ 88.49	$7.35	75%
9	190	$179.71	$192.39	$12.68	75%
10	195	$135.33	$143.22	$7.89	61%
11	195	$ 83.61	$ 88.60	$4.99	110%
12	195	$131.32	$125.66	-$5.66	87%
13	187	$132.79	$141.26	$8.47	69%
14	195	$164.33	$172.58	$8.25	70%
15	193	$197.76	$196.70	-$1.06	48%
16	195	$147.92	$165.13	$17.21	58%
17	195	$154.28	$149.12	-$5.16	57%
18	192	$156.78	$161.11	$4.33	82%
19	194	$115.07	$106.87	-$8.20	57%
20	195	$292.65	$298.40	$5.75	69%
22	195	$185.00	$164.92	-$20.08	101%
23	194	$130.60	$140.75	$10.15	77%
24	194	$ 97.22	$ 98.83	$1.61	64%
25	194	$152.01	$146.74	-$5.27	75%
26	61	$ 58.39	$ 61.37	$2.98	103%
27	188	$339.52	$363.38	$23.86	64%
Overall Avereage		$150.41	$152.01	$1.59	69%

[1] Coefficient of variation is the standard deviation divided by the population mean. Higher percentages means less consistency among the claim values and this greater variation could be expected to introduce more sampling error. (Standard deviation values are not shown in this table.)

Table 3. Sample and population means for planned audits

Table 3 shows the 27 sample averages slightly overstate the overall population average by about 1 percent. This is well within the expected range of sampling error. Notice, however, that the over- and under-estimate is substantially greater for some individual audit subjects. This is because the amount of variation within the claims is greater. The “coefficient of variation” column gives a measure of how much variation there is among claim values for each legal entity. As these values vary widely among the LEs, the number of sampled units would also vary to provide similar levels of precision. Resources could be better managed by reallocating portions of samples from entities with a smaller coefficient of variation to those with more variability.

Issue 4: ***Is the sample size sufficient for extrapolation?***

A review of reports issued by Office of Audit Services within the Inspector General's Office of the U.S. Department of Health and Human Services (DHHS) shows the proposed California DMH sample size of 200 to 250 is at the upper end of the range used by the federal agency in its audits. Some of the federal audit reports show the DHHS extrapolation procedure differs from that of CA DMH. The DHHS has used the lower limit of a 90 percent confidence interval as the repayment amount. The current DMH method does not take sampling error into account and therefore does not use any confidence interval when establishing repayment amounts through extrapolation.

Audit	***Sample Size***
PCH Health Systems outpatient physical and occupational therapy	114
Hospital of the University of Pennsylvania air ambulance services	100
Regent Care Center skilled services	50
Independent Diagnostic Testing Facilities	230
New Jersey school-based health services	150

Table 4. Sample sizes used in U.S. DHHS audits.

The level of desired precision and the equity of the extrapolation methodology are policy matters and beyond the scope of this review. Consequently, we make no recommendations in these areas.

Issue 5: ***How much precision is there in the extrapolation?***

One method of evaluating the extrapolation precision is to use the audit sample to estimate the total value of claims in the mental health services category for LE00A (again for confidentiality this is a hypothetical number). This is useful because we know what the true value is and it can be compared to the extrapolated value. Panel A of Table 5 shows the extrapolated value and the true value, which differ by $139,524. The extrapolation overestimates the true value because of the sampling error discussed above under Issue 3 and is reflected in the discrepancy between the sample and population means shown in Table 1.

A.	*Value*
Extrapolated estimate of total claim value	$1,732,728
True total claim value	$1,593,204
Difference	$139,524

B. *Extrapolated ranges*	*Confidence Level*
	80%
Lower limit	$1,641,779
Upper limit	$1,823,677
	90%
Lower limit	$1,615,831
Upper limit	$1,849,625
	95%
Lower limit	$1,593,227
Upper limit	$1,872,230

Table 5. Extrapolation of total claims value based on the audit sample.

The amount of sampling error can be taken into account by constructing a confidence interval around the extrapolated value. It gives a range of plausible values based on the amount of variation in the sample. Panel B of Table 5 shows the confidence intervals for the extrapolated amount obtained by adding and subtracting a "margin of error." The margin of error is computed based on the sampling error. Several confidence intervals are available depending on the level of confidence desired. The wider the interval, the greater the confidence we have that the true value will fall within it.

Notice that only the lower limit of the 95 percent confidence interval ($1,593,227) approaches the true value of $1,593,204. The upper limit substantially overstates the true value and note that the other more narrow confidence intervals do not include the true value.

The same procedure, with the same level of confidence, can be applied to the disallowed values. Doing so produces a wide range between the upper and lower limits. This is because there is a wide disparity in the disallowed values (shown previously in Table 2).

When the value of the eight claims disallowed in the audit ($2,590.40) is extrapolated to the universe of 5,839 mental health services claims, the recoupment amount is estimated to be $79,607[2]. Taking sampling error into account and computing a margin of error reveals the imprecision of the estimate, as shown in Table 6. A 95 percent confidence interval is obtained for the extrapolation of disallowed values. The range for the true value is between a lower limit of $6,067 and an upper limit of $153,147. Recall that is was the lower limit that approached the true value for the estimate of all claims shown in Table 5. But the lower limit in the table below is 1/13th of the extrapolated disallowed value. The difference between these two values is not inconsequential.

The level of precision can be summarized. The summaries are part of the standard output of RAT-STATS, a program specifically designed to analyze audit data written by the DHHS Office of Audit Services and distributed free.

For each level of confidence, RAT-STATS computes a margin of error (labeled "Precision Amount") and computes the size of the margin of error relative to the estimate (labeled "Precision Percent"). The margin of error for the 95 percent confidence interval for the disallowed extrapolation is $73,540 or an amount that is 92.4% of the extrapolated estimate of $79,607 (shown in Table 6). Put more simply, the margin of error is almost equal to the recoupment estimate. (The complete RAT-STATS output is contained in the Appendix.)

	Value
Extrapolated disallowed estimate	$79,607
Margin of error	$73,540
	Confidence Level
	95%
Lower limit	$6,067
Upper limit	$153,147

Table 6. Extrapolated disallowed amount based on audit sample average.

Issue 6: ***How much larger should the sample be if more precision is desired?***

Keep in mind that sample stratification would reduce the sampling error and provide some additional precision at little additional cost. Additional research could reveal which stratification variables would be the most useful. If claims are stratified on value, Roberts (1978) recommends that the top stratum be sampled on a 100 percent basis. For LE00A (again, number

[2] This value is for exposition purposes only. Because the value of the eight disallowed claims is 4.6 percent of the total sample value, no extrapolation is being applied in this audit. (Another $230 in disallowed claims would have triggered extrapolation.) The values discussed above are to demonstrate how the process would work if the disallowed amount exceeded the 5 percent threshold.

is hypothetical for reasons of confidentiality), there are 138 claims larger than $600. The virtue of examining all large claims is that the sample error for this stratum is reduced to zero.[3]

Beyond stratification, increasing the sample size would also deliver estimates that are more precise. Each audit will be different because the variation in the disallowed claims is expected to differ. With that caveat in mind, presented below is a simulation of how various sample sizes would affect the extrapolation precision for the LE00A audit. This analysis is based on two assumptions. First, it is assumed that larger samples would uncover the same proportion of disallowed claims. Second, it is also assumed that the variation of those additional disallowed claims would approximate the variation shown for the disallowed claims reported in Table 2.

Sample Size	***Extrapolation Margin of Error[1]***	***Precision Percentage[2]***	***Expected Disallowed Claims***
190	$73,540	92%	8
200	$71,620	90%	8
225	$67,370	85%	9
250	$63,770	80%	11
275	$60,670	76%	12
300	$57,950	73%	13
325	$55,560	70%	14
350	$53,410	67%	15
375	$51,480	65%	16
400	$49,740	62%	17
425	$48,140	60%	18
450	$46,670	59%	19
475	$45,320	57%	20
500	$44,070	55%	21
525	$42,910	54%	22
550	$41,830	53%	23
575	$40,810	51%	24
600	$39,850	50%	25

[1] Margin of error is the "Precision Amount" in RAT-STATS output. It is the standard error of the estimate multiplied by a t-value of 1.97. This value, when added to and subtracted from the extrapolated estimate, provides a 95 percent confidence interval.
[2] The precision percentage represents the size of the margin of error relative to the estimated recoupment amount. Smaller values in this column represent more precision.

Table 7. Extrapolation precision estimates for various sample sizes

As shown in Table 7, doubling the sample size to 400 would be expected to reduce the margin of error for the extrapolation by about one-third (92% versus 62%). A sample of 600 claims would be expected to reduce the margin of error to approximately 50 percent of the extrapolated

[3] Details on determining stratum boundaries as recommended by Gunning, Horgan and Yancey (2004) are found in the Appendix.

recoupment amount. The cost of increasing the sample size could be offset by conducting fewer audits. Samples could be drawn in an iterative process until a desired precision is achieved.

Issue 7: ***Does it matter if more than one claim is selected from a client's chart?***

All of the precision computations presented and discussed in previous sections of this report are based on the assumption of simple random sampling (SRS). Under SRS, each selected claim is assumed unrelated to the others in the sample. If, however, this is not the case, the computations for variance (and precision) estimates become much more complicated. If the claims selected in the sample are somehow related to each other, then the variance is understated and the precision is overstated.

If some of the sample claims come from one client's record, then they are said to be "clustered." Additional research could describe the exact nature of how this clustering affects the claim contents. But it is likely, for example, that multiple claims from one client's chart are processed by the same clerical persons and documented by the same professionals. If one claim in this file is problematic, then it is likely others could be as well. In statistical terms, the claims' quality within a client's file are said to be "correlated." It is this correlation among sampled claims that leads to the understatement of variance and overstatement of precision.

Client	***Claims***	***Client***	***Claims***	***Client***	***Claims***
8323	13	6532	3	6740	2
8311	12	8462	3	1297	1
8388	10	2616	3	5066	1
7413	10	3256	3	8164	1
0887	9	3388	3	2767	1
6091	7	9345	3	7122	1
0556	6	5862	3	3024	1
9703	6	0568	2	4614	1
8071	5	5753	2	1870	1
0491	5	0585	2	8006	1
1613	4	4864	2	5724	1
9466	4	3576	2	9901	1
0880	4	5348	2	8378	1
3077	4	6173	2	8614	1
5591	4	6405	2	9661	1
4777	4	6445	2	9663	1
6115	4	6909	2	3402	1
6281	4	4582	2	4223	1
0508	4	1424	2	7435	1
2204	3	1526	2	9263	1

Table 8. Distribution of sampled claims by client.

Table 8 shows that nearly one-third of the audit sample claims came from six clients and only 19 of the 190 sampled claims fit the SRS assumption.

Client	*Disallowed claims*	*Claims in the sample*
0887	2	9
8311	1	12
1613	1	4
5591	1	4
0508	1	4
6405	1	2
1424	1	2

Table 9. Disallowed claims and number of claims sampled by client.

Table 9 shows that all the disallowed claims came from charts where more than one claim was sampled. Two of the eight disallowed claims came from one chart. None of the disallowed claims came from the 19 client records that met the SRS assumption.

Table 10 illustrates the effect clustering has on the sample variance estimate. Under the assumption of SRS, the variance estimate is 12.3. When taking clustering into account, the variance is inflated by about one-fourth[4] and is estimated to be 15.5. (This inflation is called the "design effect.") The practical import of this issue is that the variance computations are the basis for precision estimates. If the variance is understated, the precision will be overstated.

Sample mean	*Standard error assuming SRS*	*Standard error assuming clustering*	*Design Effect*
296.75	12.3	15.5	1.64

Table 10. Effect of clustering on sample variance estimates.

Should policymakers determine that some precision standard is desired, the continued use of clustered data would make subsequent analysis tasks more complex. The computational difficulties associated with clustered data, however, could be avoided by limiting the sample selection to one claim per client record. If there are an insufficient number of clients to achieve this and clustered samples are unavoidable, then complex sample analysis routines could be utilized, such as PROC SURVEYMEANS and SURVEYREG in SAS, to obtain appropriate variance estimates.

The clustering has a smaller effect on the variance estimates of the disallowed claims because there is only one instance where more than one disallowed claim came from the same client's record (as shown in Table 10). This may not be the case, however, in other audits.

[4] The square root of the design effect gives the amount that clustering inflates the variance estimate. In this instance $1.64^{.5}$=1.28 or 28%.

Issue 8: ***Is the Raosoft calculator the appropriate software to determine sample size?***

Preliminarily, a distinction must be made between attribute sampling and variable sampling. Attribute sampling is used when the goal of the study is to simply estimate the proportion of disallowed claims, where each claim is evaluated on a "pass/fail" basis. Variable sampling, on the other hand, is used when the goal is to estimate a dollar value for disallowed claims.

The Raosoft sample size calculator implements standard statistical formulae to determine sample sizes for a given confidence level for attribute sampling.[5] As the goal of the audits is to estimate the value of disallowed claims, it may be more appropriate to use a variable sample calculator. RAT-STATS has such a module that will give sample sizes for various precision levels. One limitation of the RAT-STATS module is that the user cannot specify additional alternative levels of desired precision. The computations for sample sizes based on desired precision levels are presented in textbooks on audit sampling, such as Roberts (1978) or Guy, Carmichael and Whittington (1998).

As with the Raosoft calculator, the RAT-STATS sample calculator requires the input of population sizes and several assumed values: strata means and standard deviations. These values can be computed using the claims database. It will be difficult, however, to calculate sample sizes to arrive at a desired precision based on the mean and standard deviation of disallowed claims. The only way to estimate these values is to use a probe sample to uncover some disallowances.

Issue 9: ***What are the advantages and disadvantages of probe samples?***

Providers expressed great interest in using probe samples. From a technical point of view, a probe sample would allow for overall sample sizes to be determined for a desired level of precision (as discussed above), which is a departure from current practice. The use of probe samples would also allow DMH to more efficiently allocate its resources in that portions of samples from providers with low disallowance rates could be reallocated to providers with higher disallowance rates to obtain more precise recoupment extrapolation.

DMH expressed concern about adding to the project management complexity if probe samples were used. The primary concerns expressed were in the area of audit notification to the providers and timeframe management. As these are policy matters, they are beyond the scope of this report and we make no recommendation.

Issue 10: ***What factors could invalidate the legal entity selection process?***

A two-stage sampling process is currently used by DMH (a detailed description is found in the Appendix). In the first stage, legal entities are selected to be audited. Briefly, entities are stratified by the claim dollar amounts and the number of entities selected for an audit in each stratum is roughly proportional to the total claims values within the service category. Not all

[5] The parameters used by DMH are a 95% confidence interval and an assumed failure (disallowance) proportion of 15 percent. Details are found in the Appendix.

entities are subject to audit selection, but DMH does not want to disclose which legal entities fall into this category.

For the legal entities subject to selection, a random number is assigned as a means of carrying out selection. Deviations could invalidate the random selection process, such as repeatedly drawing the sample of selected audit subjects and choosing a particular subject sample based on undocumented selection criteria.

References

Gunning, Patricia, Horgan, Jane M., and Yancey, W. 2004. A Simple Method for Setting Stratum Boundaries for Statistical Auditing, Proc. *Of Amer, Acc, Association Southern Region Meeting*, Austin Texas, March

Guy, Dan M., Douglas R Carmichael and O. Ray Whittington. 1998. *Audit Sampling: An Introduction.* Wiley: New York.

Roberts, Donald M. 1978. *Statistical Auditing.* American Institute of Certified Public Accountants: New York.

APPENDIX

RAT-STATS OUTPUT

```
                 DEPARTMENT OF HEALTH & HUMAN SERVICES
                  OIG - OFFICE OF AUDIT SERVICES
Date: 9/4/2006     VARIABLE UNRESTRICTED APPRAISAL          Time: 21:20
                        AUDIT/REVIEW: LE00A (fictitious number for confidentiality)

     DATA FILE USED: C:\My Documents\auditing\ExaminedAuditedLE00A.xls

     SAMPLE         EXAMINED    NONZERO       TOTAL OF          TOTAL OF
      SIZE            VALUE      DIFFS        DIFF VALUES        AUD VALUES
       190        56,382.66         8           2,590.40         53,792.26
           ----------------------- E X A M I N E D ------------------------
           MEAN / UNIVERSE                       296.75              5,839
           STANDARD DEVIATION                    169.73
           STANDARD ERROR                         12.11
           SKEWNESS                                 .59
           KURTOSIS                                3.05
           POINT ESTIMATE                     1,732,728

                                   CONFIDENCE LIMITS
                                  80% CONFIDENCE LEVEL
           LOWER LIMIT                           1,641,779
           UPPER LIMIT                           1,823,677
           PRECISION AMOUNT                         90,949
           PRECISION PERCENT                          5.25%
           T-VALUE USED                     1.286046870294

                                  90% CONFIDENCE LEVEL
           LOWER LIMIT                           1,615,831
           UPPER LIMIT                           1,849,625
           PRECISION AMOUNT                        116,897
           PRECISION PERCENT                          6.75%
           T-VALUE USED                     1.652955801726

                                  95% CONFIDENCE LEVEL
           LOWER LIMIT                           1,593,227
           UPPER LIMIT                           1,872,230
           PRECISION AMOUNT                        139,502
           PRECISION PERCENT                          8.05%
           T-VALUE USED                     1.972595079100

           ----------------------- A U D I T E D --------------------------
           MEAN / UNIVERSE                       283.12              5,839
           STANDARD DEVIATION                    169.01
           STANDARD ERROR                         12.06
           SKEWNESS                                 .45
           KURTOSIS                                2.89
           POINT ESTIMATE                     1,653,121

                                   CONFIDENCE LIMITS
                                  80% CONFIDENCE LEVEL
           LOWER LIMIT                           1,562,560
           UPPER LIMIT                           1,743,682
           PRECISION AMOUNT                         90,561
           PRECISION PERCENT                          5.48%
           T-VALUE USED                     1.286046870294

                                  90% CONFIDENCE LEVEL
           LOWER LIMIT                           1,536,723
           UPPER LIMIT                           1,769,519
           PRECISION AMOUNT                        116,398
           PRECISION PERCENT                          7.04%
           T-VALUE USED                     1.652955801726

                                  95% CONFIDENCE LEVEL
           LOWER LIMIT                           1,514,215
           UPPER LIMIT                           1,792,027
           PRECISION AMOUNT                        138,906
           PRECISION PERCENT                          8.40%
           T-VALUE USED                     1.972595079100
```

---------------------- D I F F E R E N C E ----------------------		
MEAN / UNIVERSE	13.63	5,839
STANDARD DEVIATION	89.48	
STANDARD ERROR	6.38	
SKEWNESS	7.33	
KURTOSIS	57.56	
POINT ESTIMATE	79,607	

	CONFIDENCE LIMITS 80% CONFIDENCE LEVEL
LOWER LIMIT	31,662
UPPER LIMIT	127,552
PRECISION AMOUNT	47,945
PRECISION PERCENT	60.23%
T-VALUE USED	1.286046870294

	90% CONFIDENCE LEVEL
LOWER LIMIT	17,983
UPPER LIMIT	141,231
PRECISION AMOUNT	61,624
PRECISION PERCENT	77.41%
T-VALUE USED	1.652955801726

	95% CONFIDENCE LEVEL
LOWER LIMIT	6,067
UPPER LIMIT	153,147
PRECISION AMOUNT	73,540
PRECISION PERCENT	92.38%
T-VALUE USED	1.972595079100

Defining Stratum Boundaries

(from Gunning, Horgan and Yancey 2004)

The following assumes the top stratum (greater than $600) is sampled on a 100 percent basis.

1. Arrange the stratification variable *X* in ascending order;
2. Take the minimum value as the first term, and the maximum value as the last term of the geometric series with L+1 terms;
3. Calculate the common ratio: $r = (max/min)^{1/L}$;
4. Take the boundaries of each stratum to be the *X* values corresponding to the terms in the geometric progression with this common ratio:

$$\text{Minimum } k_0 = a, ar, ar^2 \ldots\ldots ar^L = \text{maximum } k_L.$$

$$L=3, \quad k_0=11, \ldots\ldots, k_3=600:$$

thus $r = (600/11)^{1/3} = 54.55^{1/3} = 3.89$, and $k_h = 11 * 3.89^h$ $(h=0,1,2,3)$.

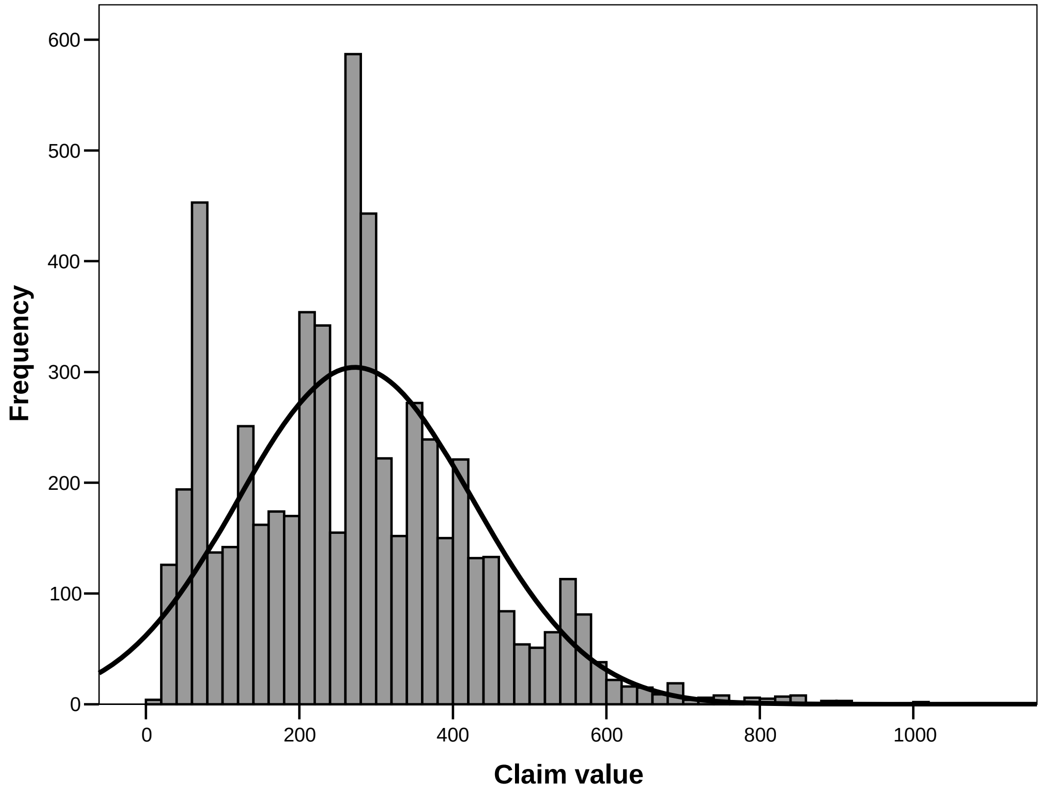
Sample sizes should correspond to the percent of the population within each stratum. (Research has shown up to five strata provide the most cost-effective stratification.)

Stratum	*Range*	*Population Percent*
1	$11-$41.70	2.4
2	$41.71-$158.19	22.3
3	$158.20-$600*	75.3
4	$600-$1,100	

Table 11. Stratified sample distribution for LE00A

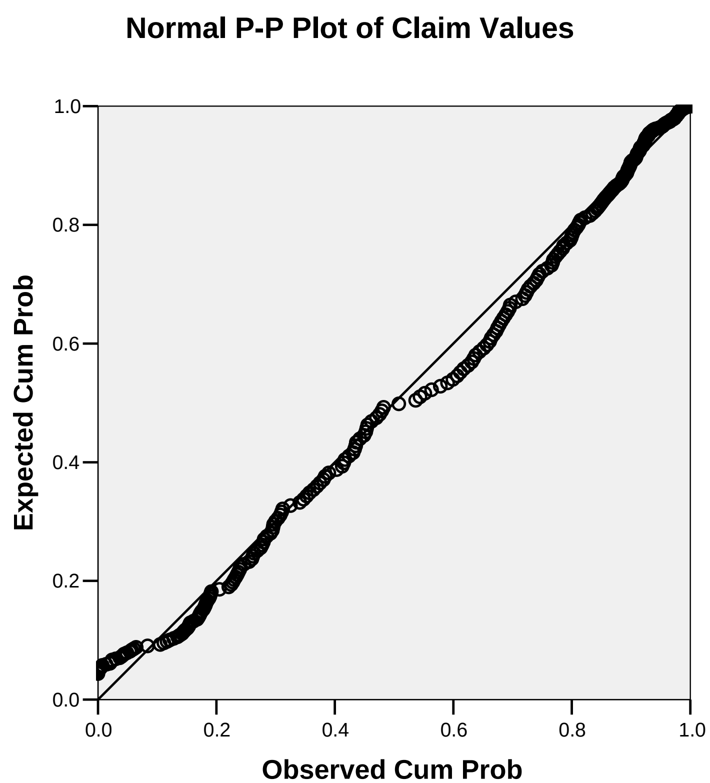
*All claims greater than $600 are examined. This threshold could be adjusted upward for LEs with a large number of high-value claims.

Fig. 2. Histogram of all claim values for LE00A.



Skew: .72 Kurtosis: .98.

Fig. 3. Non-normal distribution of all claim values LE00A



DMH Documents

Sampling Process

1. A Legal Entity is a corporation or a county. A corporation may have one or more providers. Counties usually have multiple providers, but small counties may have only one provider.

2. Legal Entities are assigned into three categories based on their total approved amount during the review fiscal year. The data source is the SD/MC approved claims file.

3. The LEs are assigned two random numbers. The first number determines which LEs will be reviewed during the review period. The second number determines when the selected LEs will be reviewed.

4. About a month prior to the review, several SAS programs are run to obtain the sample.
   - Claims are selected for a specific LE within county and saved in a dataset.
   - During the selection those claims with a disallowed indicator are deleted and excluded from the dataset.
   - Claims in the dataset are assigned a random number, using the SAS software random number generator.
   - The RaoSoft software is used to determine the sample size based on the number of claims for the LE within county and type of service for the review time period.
   - A maximum of 250 claims will be selected, plus a maximum of 20 replacement claims per service category. These claims are selected based on the random number.
   - The replacement claims are used if the LE disallowed a claim prior to notification of the review and the LE has a receipt for the disallowed claim.
   - Usually a LE will have one or more types of service reviewed to a maximum of 250 claims.
   - The sample claims and replacement claims are unduplicated by client and provider to obtain a list of clients and charts that will be used in the audit.

Sample Size Calculation

1) The sample size will be determined by using the Raosoft web site at www.raosoft.com.
2) Where there is more than one county per Legal Entity (LE), each county will have their own sample size calculated.
3) The sample Size will be calculated for each type of service within county within LE.
4) Where there is more than one county per LE any disallowance will be extrapolated to each county using the county's sample of claims.

The parameters for the Raosoft product will be the following:
Error level will be set to 5%
The confidence level will be set to 95%
The distribution level will be set to 85%. This level was selected based on past reviews of charts. It is expected that at least 85% of the claims will not have a disallowance.

Enabling Legislation

SEC. 81. The State Department of Mental Health shall revise its method for auditing entities that provide specialty mental health services under the Early and Periodic Screening, Diagnosis, and Treatment Program, and its method for extrapolating data obtained from those audits, pursuant to this section. Commencing July 1, 2006, and continuing thereafter, the following provisions shall apply:

(a) The department shall select statistically valid stratified samples by service function for each entity to be audited.

(b) The department shall not extrapolate the results of any audit to the full audited service function unless the error rate determined by the audit is five percent or greater. If the error rate is less than five percent, the department shall disallow only the specific claims found to be in error.

(c) The department, in consultation with stakeholders, shall select an independent statistician to review the sampling methodology and extrapolation methodology used by the department. No later than October 1, 2006, the statistician shall prepare a public report on the statistical validity of those methodologies. If the statistician determines either methodology to be invalid, the department shall adopt a new methodology, which shall be used by the department only after its validity is verified by the statistician.