# Statistical and Non-Statistical Samples: Some Empirical Evidence

Dr. James G. Swearingen, Accounting, Weber State College

Dr. Don R. Hansen, Accounting, Oklahoma State University

## Abstract

*There has been a great deal of discussion of the costs and benefits available from statistical sampling but little discussion as to why auditors continue to use nonstatistical sampling. This study presents a simple theoretical model that offers two possible reasons for the continued use of nonstatistical sampling in spite of the strong advocacy that statistical sampling has received over the past several decades.*

## Introduction

The ultimate (audit) risk that the auditor faces at the completion of an audit is whether the financial statements contain aggregate material error. The professional literature describes audit risk as a function of inherent risk, control risk, and detection risk [SAS, AU 312]. Inherent risk and control risk may only be assessed by the auditor but not controlled in the short run. Detection risk relates to the procedures the auditor uses to examine the financial statements and can be changed at the auditor's discretion. Since detection risk is the only component of audit risk that the auditor may control, it is used as a surrogate for audit risk in the following discussion.

To reduce detection risk to an acceptable low level, the auditor must obtain, among other things, sufficient evidential matter [SAS, AU 326]. Auditors gather evidential matter through a wide variety of techniques including sampling, analytical procedures, observation, walk throughs, etc. Of the various opportunities to gather evidential matter, sampling is perhaps the most important.

If sampling is done, it may either be statistical or nonstatistical sampling [SAS, AU 350]. Of the two sampling approaches, statistical sampling is advocated by many as the more desirable method.[1] Advocates of statistical sampling point to several significant benefits: (1) statistical sampling has the ability to quantify sampling risk; (2) it produces unbiased and representative samples; and, (3) should litigation occur, its scientific approach is more defensible (e.g., experts in statistical sampling can be used to justify the firm's conclusions).

In spite of the purported advantages of statistical sampling and its strong support from academia, statistical sampling has not achieved the widespread support that many have anticipated. In fact, two studies, one in Britain [McRae, 1982], and one in the United States [Akresh, 1980], reveal only moderate use of statistical sampling by auditing practitioners. In the UK study, it was found that only one-third of the large public accounting firms and only ten percent of medium-sized firms use statistical sampling. The U.S. study was equally gloomy in its report on the usage of statistical sampling.

The U.S. study was carried out by an AICPA task force with the explicit goal of assessing the use of sampling procedures. The survey targeted the 15 largest CPA firms and 185 other CPA firms of varying size. Responses were received from 13 of the 15 largest CPA firms and 99 of the 185 smaller firms. The task force found that only one-third of the respondents use statistical sampling; however, 11 of the 13 large firms were included in this user group. Apparently the larger U.S. firms are more receptive to using statistical sampling than are the large U.K. firms; nonetheless, the extent of usage among U.S. firms is not encouraging. The survey also revealed that the firms using statistical sampling did so on a very limited basis. Only seven of the firms in the user group used statistical sampling in more than 5% of their engagements with only three of

these firms using it more than 20% of the time. Thus, a large majority of those in the user group used statistical sampling in less than 20% of their engagements.

The limited use of statistical sampling after several decades of strong advocacy and the continued widespread use of nonstatistical sampling poses a problem that merits some investigation. In the past, advocates of statistical sampling have explained the problem away by arguing that the use of nonstatistical sampling will diminish as auditors are properly trained and as professional inertia is overcome. This response, however, has less appeal than it did a decade ago. Today the auditing profession is acutely aware of and knowledgeable in the application of statistical sampling. Many CPA firms are and have been using statistical sampling in their audits. They employ experts in the area and dedicate resources to train new and senior staff in the use of statistical sampling. Yet, at the same time, it appears that even these CPA firms are not abandoning nonstatistical sampling. It continues to be a major, important tool in carrying out the audit function.

In the sense of a positive theory, there may be sound, persuasive reasons why auditors have not abandoned nonstatistical sampling. In this study we develop a conceptual framework which allows us to produce a testable hypothesis concerning the continued use of nonstatistical sampling. We also provide some empirical evidence that is supportive of the hypothesis.

The remainder of the paper is divided into five sections. The first section provides a conceptual framework that can be used to evaluate competing sampling methods and also provides a hypothesis that can be subjected to empirical analysis. The second section describes the empirical analysis and defines the operational measures that will be used. The third section describes how the sample data were gathered to support the analysis. In the fourth section, the data are analyzed and related to the theoretical framework. Finally, our findings and conclusions are summarized in the last section.

## Choice Among Sampling Methods: A Theory and a Proposed Test

In an audit engagement, an auditor is faced with two conflicting objectives: (1) maximizing profits, and (2) minimizing the audit risk. The objectives conflict because an auditor must expend resources (and thus lower profits) in order to reduce the audit risk. We assume that audit risk can be lowered by gathering evidential matter pertaining to the financial statements. However, gathering evidential matter is a costly activity: the more evidence gathered, the lower the risk but the greater the cost and the lower the profit. In every engagement, auditors must evaluate the tradeoff between profits and audit risk. In making this tradeoff evaluation, it is assumed that auditors are expected utility maximizers and choose the profit-risk combination that is optimal.

An auditor's utility function is assumed to be defined as U(P,A), where P = profits and A = audit risk. It is assumed that utility increases with profit and decreases with audit risk. Thus, all rational auditors will prefer a profit-risk combination, (P\*,A\*) to any other profit-risk combination, (P,A), provided P\* $\geq$ P and A\* $\leq$ A with strict inequality holding for at least one of the two conditions. However, if P\* $\leq$ P and A\* $\leq$ A or if P\* $>$ P and A\* $\geq$ A, then each auditor's preference for (P\*, A\*) vis a vis (P, A) depends on the auditor's specific utility function. For some auditors, the profit-risk tradeoffs will result in choosing (P\*, A\*) over (P, A); for other auditors, the tradeoffs, will result in a preference for (P, A) over (P\*, A\*).

These basic principles of rational behavior explain how choices are made among different possibilities for gathering evidential matter, and in particular, for choosing among competing sampling methods. To illustrate, consider choosing between two sampling methods, N and S, for a given engagement. Let (P(N), A(N)), be the profit-risk combination if N is chosen and let (P(S), A(S)), be the profit-risk combination if S is chosen. Ceteris paribus, any difference between P(N) and P(S) is attributable to differences in the cost of using the two methods and any difference in A(N) and A(S) is attributable to the impact each method has on audit risk. Thus, sampling approach N is superior to sampling approach S if P(N) $\geq$ P(S) and A(N) $\leq$ A(S) with strict inequality holding for either the profit comparison or the audit risk comparison. Furthermore, this decision rule holds regardless of the specific form of the utility function. If, however, sampling method N only dominates on either profit or risk (but not both), then either N or S may be chosen, depending on the preference structure of the individual auditor.

If N represents nonstatistical sampling and S represents statistical sampling, then the above analysis suggests the following hypothesis concerning nonstatistical sampling:

*HYPOTHESIS.* The continued wide-spread use of nonstatistical sampling can be explained if one of the following two conditions are satisfied:

(1)     it is less costly to use than the competing statistical methods.
(2)     it lowers audit risk more than the competing methods.

## Proposed Empirical Analysis

Although the decision rule suggests the above hypothesis, it does not establish its validity. The validity of the hypothesis depends on the performance of nonstatistical sampling on the dimensions of cost and audit risk vis a vis the performance of statistical sampling methods on the same two dimensions. Empirical evidence needs to be gathered so that the hypothesis can be tested. The procedures for testing the hypothesized superiority of nonstatistical sampling can be described by the following three steps: (1) definition of operational measures for cost and audit risk; (2) collection of actual values for the operational measures (actual values need to be observed for two types of audits: audits using statistical sampling and audits using nonstatistical sampling); and (3) comparison and evaluation of the different sampling methods based on the observed values of the operational measures.

### Operational Measures

An operational measure of cost is obtained by assuming that the cost of sampling is an increasing function of sample size. Given this assumption, the cost of sampling methods can be compared by comparing the sample sizes produced by each method. This measure is particularly attractive since the sample sizes used in audits are readily available in the working papers. Moreover, the reasonableness of the assumption is difficult to dispute -- costs should increase as sample size increases since gathering and analyzing larger samples requires more time and resources.

Operational measures for the impact on audit risk are obtained by assuming that detection risk is a decreasing function of error-detection ability (where errors are defined as the difference between the book value and the audited value).[2] Failure to detect errors -- particularly material errors -- increases the overall audit risk of an engagement. The above assumption implies that the more errors detected, the less likely a material error will go undiscovered. By decreasing the risk of nondetection of a material error, the overall audit risk is decreased. Error-detection ability will be assessed by defining three different measures. These measures are error rate, error amounts in dollars, and sampling fraction (the percentage of the total population sampled, where population is measured in dollars).

By comparing error rates, we can assess whether one sampling method detects a greater proportion of errors than another sampling method. Although the number of errors detected is an important measure of error-detection ability, the dollar amount of the errors detected is also an important measure; accordingly, the error amounts of competing methods will also be compared. Finally, the proportion of the population sampled will be compared. Comparison of this last measure is based on the assumption that error-detection ability increases as more of the population is sampled.

### Gathering of the Sample Data

In order to compare sampling methods, the error-detection measures and sample sizes were assessed using data from audits of accounts receivable populations. Accounts receivable were selected for study for two reasons. First, the balance in accounts receivable is typically a material portion of the balance sheet, which provides the study with interest; and second, the method of obtaining evidence on accounts receivable, the confirmation process, is one that utilizes outside verification of the account balances. It was felt that the quality of the audit evidence utilized would be superior because of the independent verification. Only verification through positive confirmations was used since it appears that positive confirmations are more reliable in the detection of errors (see [Warren, 1975]).

The sample data were obtained from the working papers of one regional and five "Big 8" CPA firms in Seattle, Washington and one local and three "Big 8" CPA firms in Tulsa, Oklahoma and usually consisted of three years per client. In some cases, only one year's data were available while in others, as many as five years of data were collected. The data made available by the CPA firms were selected by them on a haphazard basis; however, there is no reason to believe that the clients and data selected are not representative of the general population. There is no known selection bias.

The data obtained from the Seattle firms were all the result of selecting the confirmation sample using random or dollar-unit selection while the data from the Tulsa

firms were all the result of using a nonstatistical selection method.[3] Table 1 contains a summary of the number of audits obtained by client type for each sampling method. Manufacturing, service, and merchandising clients were included in each sampling category. Moreover, there is no reason to expect significant differences in the accounts receivable transactions and internal control systems for Seattle and Tulsa clients. Accordingly, it is assumed that differences in sample sizes and error-detection measures are attributable only to differences in sampling methods.

The data obtained from the statistical samples were obtained using two different statistical selection methods (random and dollar- unit) while the data obtained from the nonstatistical samples were obtained using haphazard sampling. The original sample consisted of 26 audits for random samples, 43 audits for dollar-unit samples and 71 audits for nonstatistical samples (as shown in Table 1). However, in order to make meaningful comparisons of dollar error amounts, the samples from the three sampling methods were roughly matched on size as measured by total accounts receivable. For an audit to be retained for a particular sampling method, its total accounts receivable had to be within approximately one million dollars of an accounts receivable in both of the remaining sampling methods.

Appendix A lists the accounts receivable by audit (after matching) for each sampling method. For nonstatistical sampling, receivables range from $323,350 to $11,147,681, with a mean of $3,073,783. For random sampling, receivables range of $1,309,185 to $10,095,382 with a mean of $5,514,135. Finally, for dollar-unit sampling, receivables range of $299,773 to $11,213,780 with a mean of $2,692,954. As can be seen, the distributions of accounts receivable are reasonably similar, and especially so with the distributions associated with the dollar-unit and nonstatistical sampling methods.

The matching criterion had the effect of eliminating eight audits from the random sampling method, four from the dollar-unit sampling method, and ten from the nonstatistical sampling method. Table 2 gives a summary of the types of sampling used by the firms and the number of audits obtained using each method (after applying the matching criterion).

As can be seen from the summary data above and the detail contained in Appendix A, the client size, as measured by total accounts receivable, is relatively small. This may limit the generalizability of the study

and thus, care must be taken as the evaluation of the data in the following sections are reviewed. Seeking to infer the observed outcomes to accounts receivable populations of all sizes may be risky. Further research needs to be done on medium- and large-size populations.

**Data Analysis**

In this section, the data are examined to determine if differences exist depending upon sampling method. Sample size, error rate, error amount, and sampling fraction are examined to determine if these variables differ for samples drawn using random, dollar-unit, and nonstatistical sampling. In particular, the focus is on the differences that exist between nonstatistical sampling and the two statistical sampling methods. Based on the observed evidence, a statement is then made concerning the validity of the hypothesis stated earlier in the paper.

*Sample Size*

To provide information on the sample sizes utilized by the CPA firms, audits were placed into one of three groups, depending on the sampling method used: random selection, dollar-unit selection and nonstatistical selection. Each group was then rank ordered by sample size and broken into quartiles. The range, mean and median sample size were identified for each quartile. The results are contained in Tables 3, 4, and 5.

The above tables reveal that nonstatistical sampling strongly dominates random sampling on the dimension of cost (as measured by sample size). Nonstatistical sample sizes are substantially smaller than random samples for all four quartiles as measured by both the mean and median. The sample sizes for random samples varied from 45 to 601 and the sample sizes for nonstatistical varied from 9 to 110. Note that the minimum value for each random-sampling quartile is larger than the corresponding maximum value for nonstatistical sampling.

Although not shown in the tables, the overall average sample size for random sampling is 186.4 and the overall average for nonstatistical sampling is 30.8. The hypothesis that there is no difference in mean sample sizes between random sampling and nonstatistical sampling can be rejected at a significance level of less than .01.

Comparison of nonstatistical and dollar-unit sampling,

however, results in a significantly different conclusion. Tables 4 and 5 reveal little difference between the sample sizes of the two methods. The comparison does reveal a systematic but small difference in favor of nonstatistical sampling as measured by the mean and median of each quartile (with the exception of the mean of the fourth quartile). The range of the dollar-unit samples is from 7 to 88, while the nonstatistical range is from 9 to 110.

The average sample size for nonstatistical sampling is 30.8 and the overall average for dollar-unit sampling is 31.8. The null hypothesis of no difference between means cannot be rejected at a significance level of .10. However, rejection can be achieved at a level of .14. Based on the evidence of sample size comparisons, the conclusion that there is no difference in the cost of the two methods seems to be warranted.

*Error Rates*

The most striking and possibly the most important differences noted were the error rates. The error rate was obtained by dividing the total number of errors discovered by the total number of confirmations sent. For the random-sampling audits, 38 errors were discovered from 3,355 confirmation for an error rate of 1.13 percent; for the audits using dollar-unit sampling, 53 errors were discovered from 1,242 confirmations for an error rate of 4.27 percent; and for the nonstatistical-sampling audits, 113 errors from 1,883 confirmations or an error rate of 6.00 percent. The nonstatistical samples had an error rate 5.3 times larger than the random sampling method and 1.4 times larger than the dollar-unit method.

A chi-square test was used to test the hypothesis that the sampling methods detected the same proportion of errors. The null hypothesis of no association was rejected with a level of significance less than .01. Pairwise tests between the nonstatistical sampling method and the two statistical methods were also done. For nonstatistical and random sampling methods, the hypothesis of no difference in error proportions can be rejected at less than the .01 level of significance. For nonstatistical and dollar-unit sampling methods, the hypothesis of no difference in error proportions can be rejected at about the 8% level of significance.

Based on the above evidence, it appears that nonstatistical sampling outperforms both statistical methods when it comes to detecting errors. This observed difference between nonstatistical and statistical sampling methods could be attributable to the fact that nonstatistical sampling requires the use of a judgmental selection method. When auditors employ judgmental selection methods, they are free to include in the sample any elements of the population they feel are likely to contain material errors. Nonstatistical sampling does not limit auditors to randomly selected items--there is no need to select samples blindly. Thus, a possible explanation for the superior error-detection capability of nonstatistical sampling is that it allows auditors to use past experience and knowledge. This view is supported to some extent by [Hylas and Ashton, 1982]. In their study, they found that expectations from prior years and discussions with clients were responsible for detecting 18.5% of all errors and 24.4% of large errors.

*Error Amounts*

Although evidence was found that nonstatistical sampling detects proportionately more errors than statistical methods, it is not yet known how the methods compare in terms of the size of the errors detected. While the number of errors detected has a bearing on audit risk, the size of the errors detected is perhaps more critical. Significant information can be gleaned by comparing the distribution of error amounts for each of the three sampling methods. To provide this information, error amounts were rank ordered by magnitude and broken into quartiles. The range, mean, and median were all computed for each quartile. The results are displayed in Tables 6, 7, and 8. (In comparing errors, we used the absolute values of errors. However, it should be noted that virtually all errors were overstatements and, consequently, positive in sign.)

In comparing Tables 6 and 8, we find that nonstatistical sampling produces higher error amounts than random sampling for the first and fourth quartiles (as measured by the mean and median for each quartile); however, random sampling outperforms nonstatistical sampling for the second and third quartiles. Note also that the range for nonstatistical sampling is much greater than that of random sampling ($273,490 versus $3,993). This suggests that nonstatistical sampling is more successful in detecting larger errors. This observation is reinforced by the fact that the fourth quartile is dominated by the nonstatistical sampling method. In fact, if the upper limit of the fourth quartile for nonstatistical sampling is replaced by the next highest observation, the mean error amount for that quartile is $4,591 which is still much higher than that of random

53

sampling.

The overall average error detected by nonstatistical sampling is $3,853 while the average error for random sampling is $1,390. Nonetheless, the hypothesis that the means are equal cannot be rejected at the .10 level of significance (the significance level for rejection is approximately .31.) Furthermore, the hypothesis that the error variance for random sampling is greater than the error variance for nonstatistical sampling can be rejected at less than a .01 level of significance. Evidence that the error variance for nonstatistical sampling is greater than the error variance for random sampling is consistent with the observation that nonstatistical sampling detects larger errors.

In comparing Tables 7 and 8, nonstatistical sampling dominates dollar-unit sampling on all four quartiles, as measured by the mean and median. Note also that the range for nonstatistical sampling is also much greater ($273,491 versus $58,830). Additionally, the upper limits for the nonstatistical quartiles are all greater than the corresponding upper limits for dollar-unit sampling. The range and the analysis of the fourth quartile suggest that nonstatistical sampling is more successful in detecting large errors.

The overall average error amount for nonstatistical sampling is $3,853 as compared to an average of $1,568 for dollar-unit sampling. However, the hypothesis that the two means are equal cannot be rejected at the .10 level of significance (the level of significance for rejection is approximately .40). Again, consistent with the observation that nonstatistical sampling detects larger errors, the evidence supports the conclusion that the variance of nonstatistical sampling is greater than that of dollar-unit sampling (rejecting the null hypothesis expressing the opposite conclusion at less than a .01 level of significance).

The evidence indicates that the average error amount detected by nonstatistical sampling is no different than the amount detected by either of the two statistical sampling methods. However, the evidence also suggests that the nonstatistical sampling method is more successful in detecting larger errors. Detection of larger errors has obvious favorable implications relating to the level of audit risk. Once again, the performance of non-statistical sampling vis a vis the other sampling methods can be explained by the ability of auditors to use subjective input. Knowing where and how errors have occurred in past engagements could guide the auditor in

selecting a sample that displays the performance characteristics described in this section.

*Sampling Fraction*

In principle, audit risk can be minimized by taking a complete census of the population being examined. Clearly, the more of the population examined, the less likely a material error will be missed. Thus, comparing sampling fractions should also provide some insight regarding the audit risk incurred by each sampling method. Sampling fraction is defined as the book value of the accounts audited divided by the total of the accounts in the population. The percentage of audits falling into four sampling fraction classes for each of the three sampling methods is given in Table 9.

According to Table 9, nonstatistical sampling dominates the other two methods for the upper two classes. Over 50% of the audits using nonstatistical sampling sample at least 50% of the population whereas only 39% of the audits using dollar-unit sampling sample at least 50% of the population and none of the audits using random sampling achieved a sampling fraction of at least 50%. However, the hypothesis that there is no significant difference between the proportion of audits that sample at least 50% of the population cannot be rejected at the 10% level of significance when comparing nonstatistical sampling and dollar-unit sampling.

Perhaps a more meaningful analysis is a direct comparison of the sampling proportions for the three sampling methods. The sampling proportion for each method is computed by dividing the total book value of all accounts confirmed in the sample by the total value of the accounts receivable for all audits in the sample. For nonstatistical sampling this computation indicates that an average of 43.3% of the population is sampled; for dollar-unit sampling an average of 36.8% of the population is sampled; and, finally, for random sampling an average of 4.3% of the population is sampled. A chi-square analysis results in the rejection of the null hypothesis of no association at less than a .01 level of significance. Pair-wise tests between nonstatistical sampling and the two statistical methods were also conducted. In both cases the hypothesis of no difference in proportions sampled could be rejected at a level of significance of less than .01.

Both nonstatistical sampling and dollar-unit sampling have the merit of sampling a high proportion of the dollar value of the accounts receivable population.

However, nonstatistical sampling on average, selects a higher proportion of the dollar value of the population than does dollar-unit sampling. An obvious explanation for the performance of nonstatistical sampling is that auditors deliberately choose high-dollar accounts when selecting a sample.

*Implications of the Data Analysis*

The purpose of the data analysis was to assess the performance of nonstatistical sampling vis a vis the two competing sampling methods. By assessing the performance of the three methods, evidence is provided concerning the validity of the hypothesis that purported to explain the continued wide-spread use of non-statistical sampling. The reasons offered by the hypothesis for the continued use of nonstatistical sampling related to its cost and to its ability to reduce audit risk. Operational measures were defined, computed, and analyzed for a sample of audits using each of the three sampling methods. Table 10 summarizes the comparison of the operational measures for the three sampling methods.

From Table 10, it can be concluded that nonstatistical sampling dominates the two statistical sampling methods. It is less costly to use than random sampling and no more costly to use than dollar-unit sampling. Furthermore, it meets or beats the audit risk measures of both statistical methods: it detects proportionately more errors and samples a higher proportion of the population than either statistical method and detects an average error amount at least equal to the amount detected by either of the two methods. Thus, the evidence is generally supportive of the hypothesized dominance of nonstatistical sampling.

Although the evidence is supportive of nonstatistical sampling, the performance of dollar-unit sampling is reasonably close to that of the nonstatistical approach. If the reason for the better performance of nonstatistical sampling is explained by the use of subjective input, then this may suggest that dollar-unit sampling can be significantly improved by building into the method the ability to use subjective input. If this can be successfully accomplished, then the best of both worlds would be available to the auditor.

**Summary and Conclusions**

There has been a great deal of discussion of the costs and benefits available from statistical sampling but little

discussion as to why auditors continue to use non-statistical sampling. This study has presented a simple theoretical model that offers two possible reasons for the continued use of nonstatistical sampling in spite of the strong advocacy that statistical sampling has received over the past several decades. If nonstatistical sampling is less costly (or, at least, no more costly) and if it is perceived as reducing audit risk more than the competing statistical methods, then rational auditors will choose to use nonstatistical sampling.

To test the hypothesized dominance of nonstatistical sampling, operational measures for cost and audit risk were defined and values of these measures were observed and compared for a sample of actual audits. The empirical evidence indicates that nonstatistical sampling may indeed be no more costly and that it may be more successful in reducing audit risk than either of the two competing statistical methods. It appears that the ability to use subjective input with nonstatistical sampling may explain its superior performance.

The conclusions of the study are interesting and provocative--but at this stage they are also somewhat tentative. The evidence provided supporting the hypothesis of the paper is based only on a study of relatively small accounts receivable populations and on a sample of audits that was selected on a nonrandom basis. Although there is no reason to believe that the audits studied are not representative of what generally exists (with the possible limitation of population size), the possibility of some unknown selection bias does exist. Nonetheless, the evidence is intriguing and hopefully will stimulate additional empirical research that will shed further light on the validity of the hypothesized dominance of nonstatistical sampling.

The results of the study also suggests that additional research into the use of subjective input with dollar-unit sampling may be worthwhile. Currently little seems to be known about what auditors consider in selecting a judgmental sample. Identifying the factors that are used to select a judgmental sample may help considerably in building a statistical method that incorporates subjective input. A statistical method with subjective input (e.g., some form of Bayesian dollar-unit sampling) may produce a sampling method that dominates the nonstatistical approach.

Table 1
Client Type Summary

| Client Type | Random | Dollar-Unit | Nonstatistical | Total |
|---|---|---|---|---|
| | | Number of Audits | | |
| Manufacturing | 5 | 28 | 23 | 56 |
| Service | 11 | 3 | 3 | 17 |
| Merchandising | 10 | 12 | 45 | 67 |
| Totals | 26 | 43 | 71 | 140 |

Table 2
Sample Selection Methods

Nonstatistical Samples:
    Haphazard Sample                       61

Statistical Samples:
    Random Sample             18
    Dollar-Unit Sample       39      57

    Total                          118

Table 3
Random Sample Sizes

| Group | Sample Size Range | Mean | Median |
|---|---|---|---|
| 1 (4 audits) | 45 - 48 | 46 | 46 |
| 2 (4 audits) | 50 - 88 | 65 | 60 |
| 3 (5 audits) | 146 - 217 | 172 | 160 |
| 4 (5 audits) | 345 - 601 | 410 | 352 |

Table 4
Dollar-Unit Sample Sizes

| Group | Sample Size Range | Mean | Median |
|---|---|---|---|
| 1 ( 9 audits) | 7 - 18 | 15 | 16 |
| 2 (10 audits) | 20 - 26 | 22 | 23 |
| 3 (10 audits) | 29 - 38 | 32 | 32 |
| 4 (10 audits) | 40 - 88 | 57 | 52 |

Table 5
Non-Statistical Sample Sizes

| Group | Sample Size Range | Mean | Median |
|---|---|---|---|
| 1 (15 audits) | 9 - 17 | 13 | 13 |
| 2 (15 audits) | 18 - 24 | 20 | 19 |
| 3 (15 audits) | 24 - 39 | 30 | 28 |
| 4 (16 audits) | 40 - 110 | 58 | 54 |

Table 6
Random Error Amounts

| Group | Error Amount Range | Mean | Median |
|---|---|---|---|
| 1 ( 9 errors) | $ 2 - 47 | $ 21 | 20 |
| 2 ( 9 errors) | 59 - 622 | 336 | 266 |
| 3 (10 errors) | 674 - 2,657 | 1,992 | 2,273 |
| 4 (10 errors) | 2,657 - 3,995 | 2,967 | 2,700 |

Table 7
Dollar-Unit Error Amounts

| Group | Error Amount Range | Mean | Median |
|---|---|---|---|
| 1 (13 errors) | $ 1 - 28 | $ 9 | 8 |
| 2 (13 errors) | 36 - 201 | 98 | 84 |
| 3 (13 errors) | 219 - 430 | 297 | 286 |
| 4 (14 errors) | 430 - 58,831 | 5,562 | 761 |

Table 8
Nonstatistical Error Amounts

| Group | Error Amount Range | Mean | Median |
|---|---|---|---|
| 1 (28 errors) | $ 1 - 148 | $ 40 | 26 |
| 2 (28 errors) | 153 - 433 | 272 | 251 |
| 3 (28 errors) | 437 - 1,822 | 815 | 700 |
| 4 (28 errors) | 1,887 - 273,491 | 14,284 | 3,087 |

Table 9
Sampling Fraction

| Fraction Class | Percentage of audits in each class | | |
|---|---|---|---|
| | Nonstatistical | Dollar-Unit | Random |
| 0% to 25% | 29% | 28% | 100% |
| 25% to 50% | 20% | 33% | - |
| 50% to 75% | 38% | 28% | - |
| 75% to 100% | 13% | 11% | - |

Table 10
Summary of Operational Measures

| | Nonstatistical | Random | Dollar-Unit |
|---|---|---|---|
| Average sample size | 30.8** | 186.4 | 31.8 |
| Error rate | 6.0%* | 1.1% | 4.3% |
| Average error amount | $3,853 | $1,390 | $1,568 |
| Sampling fraction | 43.3%* | 4.3% | 36.8% |

*Statistically significant differences when compared to each
statistical method.
**Statistically significant difference when compared to
random sampling only.

## Footnotes

1    The focus of this study is on variables sampling; thus any reference to sampling or statistical sampling means variables sampling.

2    Once a sample is taken and the errors observed, the auditor must use this information to decide whether the account is fairly stated or not. Projection of the information contained in the sample to the overall population may be done through the use of formal statistical procedures (e.g., creation of confidence intervals, upper bounds, etc.), through the use of informal, subjective analysis, or a combination of the two. Admittedly, how auditors use error information contained in a sample can affect the audit risk. Unfortunately, exactly how auditors use sample information to assess the fairness of accounts is unknown at this time; however, the data in the working papers we examined suggest that subjective input is a critical component even when formal statistical projections are made. Of the 39 audits using dollar-unit sampling, the information available allowed the computation of the 90% and 95% upper bounds for accounts receivable. In 22 of the 39 audits, the estimated error (the upper bound) in accounts receivable exceeded the tolerable error - yet in none of these 39 audits was a judgement made that accounts receivable was materially misstated. Because of the uncertainty of how auditors process sample information, we have decided to restrict our analysis to the error-detection abilities of the competing sampling methods.

3    While no formal analysis of the number of statistical vs. nonstatistical audits was performed, it was the authors' impression that the number of statistical audits obtained from the Seattle firms represented a very small portion of the population of audits performed.

## Bibliography

1    Akresh, A. D., "Statistical Sampling in Public Accounting," *The CPA Journal* (July, 1980), pp. 20-26.

2    American Institute of Certified Public Accountants, *Statements of Auditing Standards Nos. 1 to 51*. New York: AICPA, 1987.

3    Hylas, R. E. and R. H. Ashton, "Audit Detection of Financial Statement Errors, " *The Accounting Review* (October, 1982), pp. 751-765.

4    McRae, T. W., "A Study of the Application of Statistical Sampling to External Auditing," *The Institute of Chartered Accountants in England and Wales* (1982), pp. 12-21.

5    Warren, C. S., "Confirmation Reliability - The Evidence," *The Journal of Accountancy* (February 1975), pp. 85-89.