Division of Research
Graduate School of
Business Administration
The University of Michigan

August 1977

SEQUENTIAL SAMPLING METHODS FOR TESTING A COMPLIANCE OBJECTIVE

Working Paper No. 150

by

R.W. Andrews and J.T. Godfrey
The University of Michigan

I. The Problem

In order for an auditor to execute the second standard of field work properly[1] there must be an evaluation of a client's internal control procedures. For this paper we assume that the internal control procedures to be relied upon have already been determined and what remains is to test for a client's compliance to these controls. For the purpose of this paper, we further assume there is only one attribute upon which the auditor wishes to rely. Thus we are conducting a test of compliance on that one attribute.

The problem we are addressing has a special nature in that the internal control procedure we wish to test for is carried out at more than one site. Because the sites are geographically separated, a reasonable question that arises is whether or not an auditor must test all sites in order to come to a conclusion about compliance with the internal control procedure. From our experiences we conclude that this is a common question and to resolve it some auditors still inspect procedures at all the sites, while others subjectively choose a few (maybe two or three) at which to evaluate the internal control procedures. In the latter case, they usually change the sites they inspect from year to year. For example, we can paraphrase a typical statement that we have heard from various auditors as follows, "If the same internal

---

[1] American Institute of Certified Public Accountants, <u>Statement on Auditing Standards No. 1</u> (New York: AICPA, 1973), p. 5.

control procedures are used at different sites, it seems that if we have inspected a couple sites and found them complying with their procedures we should be able to make a statistical statement about all sites as data is obtained sequentially from site to site."

In order to be able to make a statistical statement about all sites, we must develop a model which accounts for the relationship between the error rates at different sites. That is, we expect different sites to have different error rates and our model must reflect this expectation. We do this in Section III below as well as develop a method of stating a conclusion about the upper precision limit achieved. The use of the phrase "upper precision limit" is analogous to the conventional usage, except that we will use it to refer to multiple sites rather than one.

We are only considering the evaluation of internal control, i.e., compliance testing, and are not considering the substantive test. In order to crystallize this point, the reader may consider the proposed procedures as being intended for an interim period in which the substantive test will be accomplished at year-end, and we want an interim evaluation of the upper precision limit for the error rate associated with the attribute on which we want to rely. A situation in which there is only one attribute on which to rely is not unrealistic. If the internal control procedures are not stringent the single attribute might simply be the difference between the audit and book dollar value.

Since we are at interim, however, we will not be considering the substantive error but only whether or not an error exists. In Section II we introduce a specific example problem. In Section III we formulate and apply a method of solution for the problem; in Section IV we test the validity of our solution method through a simulation study. Section V contains a summary and suggestions for further research.

II. Example Problem

Consider a situation in which an auditor is responsible for performing a compliance test of the internal control procedures of a manufacturing company which has six different sites. Cash disbursements are being audited and the auditor has decided that the only attribute on which to rely is the agreement or disagreement of the voucher register with the listing of cash disbursements purported to be correct. It is an interim period and our immediate goal is to make a statement about internal controls, since substantive tests will be executed at year end.

The information contained in Table 1 is available to us. In Table 1, N_j is the number of cash disbursements to date this year at the given site j, and p_j^* is a subjective evaluation of the expected error rate at the site. For example, p_j^* might simply be the error rate which was observed

Table 1

Information Available for Evaluation of Compliance with Internal Control Procedure

Site	N_j	p_j^*
A	1565	.04
B	2313	.03
C	2216	.03
D	1700	.03
E	322	.02
F	312	.02

N_j = number of cash disbursements to date this year at site j

p_j^* = subjective evaluation of error rate at site j

the last time site j was audited. However, given additional information subsequent to the last audit, p_j^* might be the result of subjectively changing the error rate last observed. For example, at Site B assume that the actual error rate observed last year was only 1 percent. However, the volume of inventory at Site B has almost doubled with no additional manpower resources added, therefore, the auditor expects the error rate to increase to approximately 3 percent as given in the table. It follows then that the information content of p_j^* is the result of a history of experience which an auditor will accumulate over a number of periods of auditing the same client. This experience factor is present in most auditing situations and we include it as an important input to the solution approach we present in Section III.

II-a. The statistical test

Since the audit objective of the present test is to be able to make a statement about the reliability of the internal control procedure we are primarily concerned with two quantities, the reliability (R) and the upper precision limit (u). When we are finished with this compliance test we want to be able to state that with a reliability of .95 we are confident that no site has an error rate greater than u.

A general description of the strategy of our compliance test is as follows: After we have audited two sites, we will be able to make a statistical statement about the upper precision limit for all six sites at a specific reliability level. If this stated upper precision limit and reliability are within the ranges we wish to accept, then we can stop without auditing the remaining four sites. Otherwise we will proceed sequentially to the third site and calculate a new upper precision limit. It might be necessary to audit all six sites, especially if the error rates are higher than expected and above some prescribed level. On the other hand, if the upper precision limit remains higher than expected after the inspection of three or four sites, it may be realistic to conclude that the internal control procedure is not accomplishing its job and therefore should be relied on less, rather than expending the cost and effort to inspect the remaining sites.

II-b. Selection of sequence of site inspections

As will be seen in Section III, the validity of our solution model does not depend on the order in which sites are inspected; it only depends on knowing N_j and assessing p_j^*. Thus, a decision about the order in which sites should be inspected is not part of our model. This allows an auditor to make that decision based upon other important factors not included in our model. For example, an auditor may wish to consider the total dollar value of transactions at each site, which would reflect an auditor's ever-present concern about materiality, or the time span since a site was last inspected and/or the magnitude of the expected error rate. If a site has not been inspected for two or more periods and/or is expected to have a higher than normal error rate, then it should be a prime candidate for inspection.

These other factors have intuitive appeal and may well lead to the development of a model for optimal sequencing of site inspection. Indeed, this is a potential area for expansion of the model we present here. In this paper, however, the order of site inspection has no impact on the statistical validity of our model. For convenience, in our example problem we select the order of inspection to be the order of sites found in Table 1.

II-c. Determination of sample size

Since our overall objective is to state an upper precision limit at a specified reliability for all sites we must be able to estimate the error rate at each of the sites already audited. Therefore, an objective at each site inspected is estimation, and we must require a sample large enough to accomplish this. The sample size determination at each site audited will be based on the following four factors: N_j, the total number of transactions at each site; p_j^*, the subjectively evaluated expected error rate; R, the required reliability, and u, a desired upper precision limit. The setting of a desired upper precision limit should be understood for its primary function, which is the establishment of a reasonable sample size. Once the data has been observed the auditor will have to decide whether or not the upper precision which is achieved is acceptable.

The proposed random sample at each site is without replacement so we will employ the hypergeometric distribution. If R = reliability, N_j = the total number of transactions, p_j^* = the subjectively evaluated expected error rate, and u = desired upper precision limit, then a reasonable sample size will be the smallest value of n_j such that

$$\sum_{i=0}^{p_j^* n_j} \left[\frac{\binom{uN_j}{i} \binom{N_j - uN_j}{n_j - i}}{\binom{N_j}{n_j}} \right] < 1 - R. \qquad (2.1)$$

If N_j is large compared to the sample size we can approximate the hypergeometric distribution with the normal distribution. That is, in a random sample without replacement of size n_j

from a finite population as described above, the random variable Y_j, the number of observed errors at site j, will have an approximate normal distribution with the following mean and variance:

$$\text{Mean} = n_j u$$

$$\text{Variance} = n_j u(1-u) \frac{N_j-n_j}{N_j-1} . \tag{2.2}$$

Our purpose is to find a sample size which would yield an observed error rate less than p_j^* only (1-R) of the time if the population error rate were u. This means that, if the population error rate is u or larger, we would most likely (R of the time) observe an error rate p_j^* or larger. This stipulation can be stated using our normal approximation as follows. Find the smallest n_j such that

$$P[Y_j \le n_j p_j^*] = 1 - R \tag{2.3}$$

or

$$P\left[\frac{Y_j-n_j u}{\sqrt{n_j u(1-u) \frac{N_j-n_j}{N_j-1}}} \le \frac{n_j p_j^*-n_j u}{\sqrt{n_j u(1-u) \frac{N_j-n_j}{N_j-1}}}\right] = 1 - R. \tag{2.4}$$

Therefore,

$$\frac{n_j p_j^*-n_j u}{\sqrt{n_j u(1-u) \frac{N_j-n_j}{N_j-1}}} = - Z_{1-R} \tag{2.5}$$

where Z_α is the value of the standard normal random variable with a probability of α in one tail. Solving (2.5) for n_j we have

$$n_j = \frac{Z_{1-R}^2 u(1-u) N_j}{Z_{1-R}^2 u(1-u) + (N_j - 1)(p_j^* - u)^2} . \tag{2.6}$$

For the case we are considering with u = .08 and R = .95 we find the sample sizes for the six sites which are set forth in Table 2.

Table 2

Sample Sizes for Each Site

Site	n_j
A	116
B	78
C	77
D	77
E	48
F	47

In the next section we present a model and procedure for evaluating the sample data.

III. Formulation of a Solution Procedure

In this section we will first describe a model that is assumed for the problem described above; then we will derive estimation procedures used to evaluate this model. This methodology will yield a sequential process which will give an achieved upper precision limit after each site inspection beginning with the second site inspected. We require that at least two sites always be inspected so that we will have some additional evidence as to the variability of the error rate between sites.

III-a. Model formulation

Let p be the random variable that is the proportion of errors generated by the accounting procedures being employed at all sites. We assume a Beta distribution on p with unknown parameters (a,b). We write the density of p to be

$$f(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} \qquad \begin{array}{l} 0 \le p \le 1 \\ a,\ b > 0. \end{array} \tag{3.1}$$

This gives us a rich family of densities for possible distributions on p. It should be noted that the actual proportion of errors, p_j, for each of the sites will be a realization of the random variable p.

Consistent with an earlier definition, we let p_j^* $(j = 1, \cdots, k)$ be the subjectively evaluated expected error rate that has been assessed for each site. We assume that p^* is related to p through a function, i.e., $p^* = h(p)$. Since p is a random variable and p^* is a function of p, then p^* will be a random variable. We do not assume a specific form for $h(\cdot)$; however, as part of the model we assume that $h(\cdot)$ is of a form such that

(i) $E[p] \leq E[h(\cdot)]$

(ii) $\text{Var}[h(\cdot)] = \text{Var}[p] = \dfrac{ab}{(a+b)^2(a+b+1)}$.

The reason for these assumptions is that we will use the p_j^*'s in the initial estimate of the mean and variance of p. Assumption (i) implies that using the p_j^*'s to estimate the mean of p will result in a conservative estimate, since we should get a positively biased estimate and this will lead to a positively biased upper precision limit.

Assumption (ii) allows us to use the variance of the p_j^*'s as a basis for estimating the variance of p. However, as we shall see later, the sample data will be allowed to augment the p_j^*'s in making the variance estimate as well as the estimate of the mean.

After we have sequentially inspected ℓ sites, we will have a problem structure as given in (3.2). The p_j^*'s and $\hat{p}_j$'s form the data set which will be used in our estimation procedures.

$$\begin{array}{ccccccc} \underline{\text{Site } 1} & \underline{\text{Site } 2} & & \underline{\text{Site } \ell} & \underline{\text{Site } \ell+1} & & \underline{\text{Site } k} \\ \begin{bmatrix} p_1 \\ p_1^* \\ \hat{p}_1 \end{bmatrix} & \begin{bmatrix} p_2 \\ p_2^* \\ \hat{p}_2 \end{bmatrix} & \cdots & \begin{bmatrix} p_\ell \\ p_\ell^* \\ \hat{p}_\ell \end{bmatrix} & \begin{bmatrix} p_{\ell+1} \\ p_{\ell+1}^* \\ \end{bmatrix} & \cdots & \begin{bmatrix} p_k \\ p_k^* \\ \end{bmatrix} \end{array} \qquad (3.2)$$

The $p_1, \cdots, p_k$ are unknown realizations from the distribution given in (3.1). The $p_1^*, \cdots, p_k^*$ are observed realizations from the distribution $p^* = h(p)$ and $\hat{p}_j = \frac{Y_j}{n_j}$ $(j = 1, \cdots, \ell)$, the sample proportion of errors at site j for the sites inspected.

Given our data set in (3.2), we want to develop estimates of a and b, the parameters in the distribution of p (3.1). The following discussion will lead to acceptable estimators.

Our strategy for estimating the parameters a and b will be to use our data set in (3.2) to calculate estimates $(\hat{\mu}, \hat{\sigma}^2)$ of the mean and variance of (3.1). Then since we know that

$$\mu = \frac{a}{a+b} \tag{3.3}$$

and

$$\sigma^2 = \frac{ab}{(a+b)^2(a+b+1)} , \tag{3.4}$$

we can equate these to $\hat{\mu}$ and $\hat{\sigma}^2$ and solve for $\hat{a}$ and $\hat{b}$ yielding

$$\hat{a} = \frac{\hat{\mu}^2(1-\hat{\mu}) - \hat{\mu}\hat{\sigma}^2}{\hat{\sigma}^2} \tag{3.5}$$

and

$$\hat{b} = \frac{\hat{\mu}(1-\hat{\mu})^2 - \hat{\sigma}^2(1-\hat{\mu})}{\hat{\sigma}^2} \tag{3.6}$$

In developing estimators of μ and σ^2 from our data set we have attempted to model typical auditing logic within our statistical estimation procedure. That is, in accomplishing our objective of calculating an upper precision limit u^* for all k sites after $\ell < k$ sites have been inspected, we would rather overstate u^* than understate it at our specified reliability level. Another auditing logic characteristic that we desire in our estimation procedure is that, for a given set of sites inspected, the larger the observed error rates, the larger will be u^*. For example, if sites 1, 2 and 3 are inspected with observed error rates $\hat{p}_1 = .03$, $\hat{p}_2 = .04$, $\hat{p}_3 = .03$, respectively, we want u^* to be larger than it would be if the results would have been $\hat{p}_1 = .02$, $\hat{p}_2 = .01$, $\hat{p}_3 = .03$. These objectives are incorporated into our estimation procedures for μ and σ^2.

As an estimator of μ we use

$$\hat{\mu} = \frac{\hat{p}_1 + \hat{p}_2 + \cdots + \hat{p}_\ell + p^*_{\ell+1} + \cdots + p^*_k}{k} \tag{3.7}$$

This estimator uses the latest data available, the $\hat{p}_j$'s ($j = 1, \cdots, \ell$) for the first ℓ sites inspected (in place of the first ℓ p^*_j's). Since it is a blend of $\hat{p}_j$'s and p^*_j's it will tend, on the average, to be positively biased (as mentioned earlier), because $E[p^*] > E[p]$ and $E[\hat{p}] = E[p]$. An upward biased $\hat{\mu}$ may lead to a larger u^*. However, as more sites are inspected, $\hat{\mu}$ will have less bias as an estimator of μ because we are using more $\hat{p}_j$'s in place of the p^*_j's. This is acceptable from an auditing logic point of view as we are using more currently available data (the $\hat{p}_j$'s) and therefore $\hat{\mu}$ should become a better estimate of μ. When we were using more p^*_j's relative to the $\hat{p}_j$'s we preferred to be conservatively biased ($E[p^*] > E[p]$) rather than otherwise.

A variance estimator ($\hat{\sigma}^2$) is not as straightforward to develop as the mean estimator ($\hat{\mu}$). Since Var[p*] = Var[p] = σ^2, one approach is to use

$$s_{p^*}^2 = \frac{\sum_{i=1}^{k} (p_i^* - \bar{p}^*)^2}{k} \quad . \tag{3.8}$$

Then with $\hat{\sigma}_1^2 = s_{p^*}^2$ as an estimator of σ^2, we could use it in (3.5) and (3.6) along with $\hat{\mu}$ to calculate estimators $\hat{a}$ and $\hat{b}$. However, $s_{p^*}^2$ does not use any data from the ℓ site inspections and, therefore, we elect not to use it and to improve on it using the complete data set. The $\hat{p}_j$'s should be incorporated into $\hat{\sigma}^2$ since they represent the most current data regarding the true error rates (p_j) being generated.

This leads to another approach which is to consider Var[$\hat{p}$].

$$\begin{aligned} \mathrm{Var}[\hat{p}] &= E\{\mathrm{Var}[\hat{p}|p]\} + \mathrm{Var}\{E[\hat{p}|p]\} \\ &= E\{\frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}\} + \mathrm{Var}[p]. \end{aligned} \tag{3.9}$$

So, $\quad \mathrm{Var}[p] = \mathrm{Var}[\hat{p}] - E\{\frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}\}$.

Therefore, another possible estimator of Var[p] would be

$$\hat{\sigma}_2^2 = S_{\hat{p}}^2 - \frac{1}{\ell} \sum_{j=1}^{\ell} \left[\frac{\hat{p}_j(1-\hat{p}_j)}{n_j} \quad \frac{N_j - n_j}{N_j - 1} \right] \tag{3.10}$$

where $$S_{\hat{p}}^2 = \frac{\ell \left[\sum_{j=1}^{\ell} \hat{p}_j^2 \right] - \left[\sum_{j=1}^{\ell} \hat{p}_j \right]^2}{\ell^2} \quad .$$

We also object to using $\hat{\sigma}_2^2$ as an estimate of Var[p]. The main reason is that, because of the relatively low values of $\hat{p}_j$ with which we are working, it is not unusual for our sample data to yield a $\hat{\sigma}_2^2 < 0$. This, of course, is not a usable result. Two other objections to $\hat{\sigma}_2^2$ are (1) that it does not make use of any of the prior knowledge embodied in the p_j^*'s and (2) for small ℓ there may be too few $\hat{p}_j$'s available for an adequate estimate.

Besides the short-comings of $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ discussed above, we also desire that our estimator of Var[p] help us to achieve our auditing objectives mentioned earlier. In brief, these objectives were not to understate an achieved upper precision limit (u*) and that our calculation of u* be consistent with the error rates observed at sites inspected.

With all of our objectives in mind, we propose the following estimator of Var[p]

$$\hat{\sigma}_3^2 = \frac{\sum_A (\hat{p}_j - \bar{p}^*)^2 + \sum_B (p_j^* - \bar{p}^*)^2}{k} \tag{3.11}$$

where the set A consists of all sites inspected where the observed $\hat{p}_j > p_j^*$ and $|\hat{p}_j - \bar{p}^*| > |p_j^* - \bar{p}^*|$ and set B consists of all sites not included in A.

An observed $\hat{p}_j$ will enter the calculation only if

(1) $\hat{p}_j$ is larger than its corresponding p_j^*

and (2) the distance that $\hat{p}_j$ is from $\bar{p}^*$ is larger than the distance the corresponding p_j^* is from $\bar{p}^*$.

We elect to keep $\bar{p}^*$ constant in our calculation of $\hat{\sigma}_3^2$ (i.e., not recalculated using the $\hat{p}_j$'s) so that the value of $\hat{\sigma}_3^2$ will always be at least as large as the estimator based only on the p_j^*'s ($\hat{\sigma}_1^2$). Then since we assume Var[p] = Var[p*] and $\hat{\sigma}_1^2$ is an estimate of Var[p*], we are confident that $\hat{\sigma}_3^2$ will be an upward biased (conservative) estimated of Var[p]. This should have a favorable impact on our auditing objectives as well as meet our statistical requirements.

Now that we have determined our estimators $\hat{\mu}$ [in (3.7)] and $\hat{\sigma}_3^2$ [in (3.11)] we can use them to calculate estimates of a and b, the parameters of the Beta distribution which is generating the error rates at the k sites. With $\hat{a}$ and $\hat{b}$ we have an estimate of the distribution of p and, therefore, can make a statement concerning the probability that any site will have an error rate above a certain level. After inspecting ℓ sites, we can state with reliability R that u* is the achieved upper precision limit for all k sites if

$$u^* = \min_u\{\text{Prob}[\max p_j \leq u] \geq R\}. \tag{3.12}$$

This expression states that the probability that the largest p_j will be less than or equal to u* is greater than or equal to the specified reliability. The probability calculation uses the Beta distribution with the estimated values of a and b. In order to evaluate this procedure, we need to use an approximation to an incomplete Beta integral. Assuming independence of the realization of the p_j's we have

$$\text{Prob}[\max p_j \leq u] = \{\text{Prob}[p \leq u]\}^k$$

$$= \{ \int_0^u \frac{\Gamma(\hat{a} + \hat{b})}{\Gamma(\hat{a}) + \Gamma(\hat{b})} p^{\hat{a}-1}(1 - p)^{\hat{b}-1} dp\}^k \quad . \qquad (3.13)$$

Setting this probability equal to R, we are left with the numerical problem of finding the value of u* such that,

$$\int_0^{u^*} \frac{\Gamma(\hat{a} + \hat{b})}{\Gamma(\hat{a}) + \Gamma(\hat{b})} p^{\hat{a}-1}(1 - p)^{\hat{b}-1} dp = R^{\frac{1}{k}} \quad . \qquad (3.14)$$

An algorithm has been given[2] which solves this problem and which we use in our numerical example. The solution, u*, is the achieved upper precision for all sites. We next consider calculating u* for the example problem introduced in Section II.

III-b. Numerical results of example problem

In order to demonstrate our suggested procedure, we will estimate the upper precision limit for all sites after

[2] K.L. Majumder and G.P. Bhattacharjee, "Inverse of the Incomplete Beta Function Ratio," Applied Statistics 22, No. 3 (1973): 411-14.

the first two sites have been inspected ($\ell = 2$). Recall that there are six sites in total ($k = 6$) with population sizes and p_j^*'s given in Table 1 and sample sizes given in Table 2. Table 3 below gives a calculation of u* after the inspection of two sites and for seven possible error outcomes at each site (number of errors $Y_j = 0, 1, \cdots, 6$).

Table 3

Upper Precision Limit for All Six Sites
after First Two Sites Inspected
(Reliability = .95)

Site B ($n_B = 78$)

Site A ($n_A = 116$)

Y_j		0	1	2	3	4	5	6
	$\hat{p}_j$	0	.0128	.0256	.0385	.0513	.0641	.0769
0	0	.0372	.0388	.0406	.0462	.0606	.0792	.1000
1	.0086	.0383	.0400	.0418	.0474	.0616	.0800	.1007
2	.0172	.0394	.0412	.0431	.0486	.0627	.0808	.1011
3	.0259	.0407	.0425	.0444	.0498	.0637	.0818	.1020
4	.0345	.0419	.0437	.0457	.0511	.0649	.0827	.1026
5	.0431	.0462	.0480	.0499	.0550	.0680	.0852	.1046
6	.0517	.0576	.0592	.0608	.0651	.0764	.0921	.1103

Table 3 may be read as follows. If at Site A we observe $Y_A = 2$ errors and at Site B, $Y_B = 3$ errors, we calculate $u^* = .0486$ as our achieved upper precision limit. That is, we believe that with probability .95 none of the six sites has an error rate (p_j) larger than .0486. This limit is calculated, of course, from a specific Beta distribution with parameters $\hat{a}$ and $\hat{b}$ whose values were dependent on our estimators $\hat{\mu}$ and $\hat{\sigma}_3^2$.

Notice that u* increases monotonically as we read down and to the right across Table 3. This reflects one of the auditing objectives stated earlier in Section III-a. Our other primary auditing objective (that u* not be understated) cannot be demonstrated in Table 3, but we will comment on it in the report of our simulation runs in Section IV.

To further illustrate the monotonic property, we calculated the u*'s after three sites had been inspected (Sites A, B, and C), given that two possible occurrences had been observed at each of Sites A and B. That is, we assumed that $Y_A = 2$ and $Y_B = 5$ at Sites A and B and then calculated u* for the possible outcomes of $Y_C = 0, 1, \cdots, 6$. We then repeated the calculations assuming $Y_A = 5$ and $Y_B = 5$. The results are given in Tables 4 and 5.

Table 4

Upper Precision Limit for All Six Sites
after First Three Sites Inspected and
$Y_A = 2$, $Y_B = 5$
(Reliability = .95)

Site C
($n_C = 77$)

Y_C	0	1	2	3	4	5	6
$\hat{p}$	0	.0130	.0260	.0390	.0519	.0649	.0779
u*	.0782	.0791	.0805	.0839	.0933	.1069	.1237

Table 5

Upper Precision Limit for All Six Sites
after First Three Sites Inspected and
$Y_A = 5$, $Y_B = 5$
(Reliability = .95)

Site C
($n_C = 77$)

Y_C	0	1	2	3	4	5	6
$\hat{p}_C$	0	.0130	.0260	.0390	.0519	.0649	.0779
u*	.0821	.0834	.0848	.0882	.0973	.1105	.1269

When only two sites were inspected with $Y_A = 2$ and $Y_B = 5$, u* was .0808 (see Table 3). Then in Table 4, we see that with zero errors observed at site C, u* decreases to .0782. This is not surprising since our new data from Site C is an improvement over the error rates observed at Sites A and B. However, we see that u* increases monotonically as the error rate increases at Site C and for $\hat{p}_C = .0390$, u* = .0839, which is larger than it was at the end of inspection of Site B.

In the other case when only two sites were inspected with $Y_A = 5$ and $Y_B = 5$, u* was .0852 (see Table 3). Again, observing $\hat{p}_C = 0$ reduces u* to .0821, but then u* increases as more errors are observed, and at $\hat{p}_C = .0390$, u* = .0882, which is greater than it was at the end of inspection of Site B.

Thus, our calculation of u* is responsive to variation in error rates observed between sites inspected. Also, u* is consistent for a given site since it increases monotonically as the observed error rate ($\hat{p}_j$) increases.

In Tables 6 and 7 below, we compare our results for the example problem with the calculation of the upper precision limit by traditional methods. In Table 6 we consider the case of inspecting Sites A and B. Notice in Table 6 that upper precision limits for Sites A and B are calculated independently of each other by the traditional method. Then the decision to inspect more sites would depend on, among other things, a subjective evaluation of the error rates at the remaining sites. Our proposed u* is an upper precision limit for all sites, whether or not inspected. There may be other criteria that will influence an auditor's decision about whether or not to inspect more sites. However, in u* he does have a statistically derived error rate which applies to all sites and which can be used to help make his decision.

In Table 7 consider the case where Sites A and B have been inspected with two different outcomes for each and then consider seven possible outcomes for Site C. In the top part of Table 7 we have error rates of $\hat{p}_A = .0172$ and $\hat{p}_B = .0641$ and seven different error rates for Site C. Again, all of the traditional upper precision limits are calculated independently of each other while each of our proposed upper

Table 6

Upper Precision Limit (U.P.L.) Comparisons
Two Sites Inspected
(Reliability = .95)

	Sites					
	A	B	C	D	E	F
Sample Size	116	78	----	----	----	----
Observed Errors	2	5	----	----	----	----
Error Rate	.0172	.0641	----	----	----	----
Traditional U.P.L.*	.0600	.1400	----	----	----	----
Proposed U.P.L.	.0808	.0808	.0808	.0808	.0808	.0808

	Sites					
	A	B	C	D	E	F
Sample Size	116	78	----	----	----	----
Observed Errors	5	5	----	----	----	----
Error Rate	.0431	.0641	----	----	----	----
Traditional U.P.L.*	.0900	.1400	----	----	----	----
Proposed U.P.L.	.0852	.0852	.0852	.0852	.0852	.0852

*Sampling for Attributes (Supplementary Section), AICPA, 1967, page S-21, Table 2-B. Upper precision limits for n = 80 and 120 are used since the source table does not include n = 78 and 116.

Table 7

Upper Precision Limit Comparisons
Three Sites Inspected
(Reliability = .95)

	Sites								
	A	B	C						
Sample Size	116	78	77	77	77	77	77	77	77
Observed Errors	2	5	0	1	2	3	4	5	6
Error Rate	.0172	.0641	.0000	.0130	.0260	.0390	.0519	.0649	.0779
Traditional U.P.L.*	.0600	.1400	.0400	.0600	.0800	.1000	.1200	.1400	.1500**
Proposed U.P.L. (for all sites)	----	----	.0782	.0791	.0805	.0839	.0933	.1069	.1237

	Sites								
	A	B	C						
Sample Size	116	78	77	77	77	77	77	77	77
Observed Errors	5	5	0	1	2	3	4	5	6
Error Rate	.0431	.0641	.0000	.0130	.0260	.0390	.0519	.0649	.0779
Traditional U.P.L.*	.0900	.1400	.0400	.0600	.0800	.1000	.1200	.1400	.1500**
Proposed U.P.L. (for all sites)	----	----	.0821	.0834	.0848	.0882	.0973	.1105	.1269

*Sampling for Attributes (Supplementary Section), AICPA, 1967, page S-21, Table 2-B. Upper precision limits for n = 80 and 120 are used since source table does not include n = 77, 78 and 116.

** Interpolated since no table value given for 6 errors.

precision limits is dependent on the outcomes at all three sites and pertain to all six sites.

In the next section we use a simulation study to test the validity of our approach.

IV. Simulation Study

The main purpose of this simulation study will be to demonstrate the validity of the procedures described in the last section. In the last section we reported procedures for calculating an upper precision limit u* at a specified reliability R. Using simulation we will show that u* actually has that reliability. Because of the complex nature of the estimation procedure an analytic proof is intractable; therefore stochastic simulation methods will be employed.

IV-a. Simulation design

We have designed a simulation experiment which performs 600 replications of the specific problem discussed in Section III. In order to investigate this specific problem we have fixed: (1) the number of sites at six, (2) the population sizes and the subjectively evaluated error rates as given in Table 1, and (3) the sample sizes as given in Table 2. We could use other values for these quantities; however a favorable conclusion about the reliability for this particular problem will give credence to the statistical validity of the procedure.

Each replication can be considered as a compliance test for a company with six sites. It still holds that there is only one attribute on which we have decided to rely. Since the procedures for arriving at an upper precision limit, as given in the last section, will be applied to each application the reliability levels will be stated as a percentage of the 600 outcomes.

The first random element of each replication is a Beta distribution which generates the actual error rates at each site. We decided to set the mean and variance of the generating distribution at the mean and variance of the six p^*'s. Since the mean and variance of the p^*'s are .028333333 and .000047222 respectively, the resulting parameters of the Beta distribution are $a = 16.49$ and $b = 565.51$. Therefore for this simulation study we are generating the p_j's from a distribution for which $E[p]$ is equal to the sample mean of the p^*'s and the $Var[p]$ is set at the sample variance of the p^*'s. In our model we assumed that $Var[p] = Var[p^*]$; therefore, for simulation purposes it is only reasonable that our initial study should set the generating variance at the sample variance of the p^*'s.

If we wanted the simulation study to conform explicitly with our model assumptions, then we should set the mean of the generating distribution at a value smaller than the mean of the p^*'s. Since the difference between $E[p]$ and $E[p^*]$ was not specified in the model, we set the generating mean equal to the average of the p*'s. A sensitivity analysis, not reported here, which investigates different parameter settings for the generating distribution is underway.

The second random element of each replication is a sample taken without replacement at each site. Based on these samples and using the procedures given in Section III, an upper precision limit is calculated after the 2nd, 3rd, 4th, 5th and 6th sites. Since we know the actual error rates we can determine the percentage of times that our stated upper precision limit exceeded the error rates at all of the sites. These percentages are an estimate of the reliability of our procedure.

IV-b. Simulation results

Throughout this study we used R = .95. Table 8 gives the proportion of times that none of the six actual error rates exceeded our u* for $\ell = 2, 3, 4, 5, 6$, where ℓ is the number of the sites inspected.

Table 8

Simulation Study
Achieved Reliability

Number of Sites Inspected	Proportion of Times None of Actual Error Rates Exceeded u*
2	.94
3	.94
4	.945
5	.962
6	.978

These values indicate that our procedure has a reliability of approximately .95. Notice the increase in reliability as the number of sites inspected increases. This is due to the fact that the estimate of the variance of the generating distribution can only increase as ℓ, the number of sites inspected, increases. However, the importance of these results is that u* will be reported at approximately a .95 reliability. This can be further emphasized by considering the following. For each of the 600 replications a u* was reported after the 2nd site was inspected. Of the 600 u*'s only 36 (.06) failed to exceed the error rates of all six sites. With this data we would accept the hypothesis that the reliability is .95.

V. Conclusions and Further Research

We have modeled the situation in which the same internal control processes are being applied to several different sites. On the basis of our model we have developed a method of stating an upper precision limit (u*) for all the sites at a specified reliability (R). The interpretation of this (u*,R) pair should be that we are R confident that no site has an error rate which exceeds u*. The estimation procedure was developed on the basis of logical auditing properties and was not developed with any particular statistical property in mind. However, we were able to show the statistical validity of our reliability statement by using simulation. The simulation study indicated that when we set R = .95, our procedure gave an upper precision limit that exceeded all error rates approximately .95 of the time.

We have developed a procedure for stating an upper precision limit for a single attribute when there are multiple sites. This same procedure could be used independently on more than one attribute; however, we believe that there should be more research in the multiple attributes case. In fact, very little work has been done for the multiple attributes case when there is only one site.