1-1-1984

Discussant's response to an analysis of the audit framework focusing on inherent risk and the role of statistical sampling in compliance testing

William R. Kinney

Discussant's Response to An Analysis of the Audit Framework Focusing on Inherent Risk and the Role of Statistical Sampling in Compliance Testing

William R. Kinney, Jr.
University of Michigan

Don Leslie has prepared a paper with many ideas that give rise to a number of implications for both the theory and practice of auditing and that merit careful thought, discussion, and debate. I agree with most of what he has to say but, of course, have some *caveats*, reservations, and qualifications.

The paper has nine sections, and I plan to comment on parts of all nine, concentrating on three central topics: the audit risk formula, assessment of the prior probability of material error, and the role of compliance tests. My closing comments will address some of the political aspects of Leslie's policy recommendations.

The Audit Risk Formula

Leslie is correct that the audit risk formula of SAP 54 and SAS 39 can understate the "final audit risk" (FAR)* relative to the approach suggested in Leslie, Teitlebaum, and Anderson (1979) and the CICA's *Extent of Audit Testing* (EAT). He is also correct that a reasonable interpretation of SAS 47 may lead to a formula that will always understate FAR relative to the EAT approach. Finally, he is correct that the difference between FAR calculated using SAS 39 and FAR using EAT is small when the prior probabiity of material error (PPE) is small.

Leslie implicity assumes that for the zero error population, every item in the population has zero error and any audit test will confirm this fact. I make a slightly less restrictive assumption that *net* error in the population is zero. Under the latter assumption, substantive tests of details (STD) may lead to incorrect rejection, and, of course, analytical review (AR) may lead to incorrect rejection even if every item has zero error. Under the zero net error assumption, the EAT formula understates FAR relative to a slightly more complete "Bayesian Risk Product" (BRP) calculation* because the EAT approach ignores the risk of incorrect rejection for both analytical review procedures and substantive tests of details.

---

* Since Leslie sometimes uses "final audit risk" and "posterior risk" interchangeably and since posterior risk seems to be a misnomer as used here, I will use "final audit risk" (FAR) throughout.

* I refer to this approach as the Bayesian risk product rather than Bayesian since it assumes that decisions are made on the basis of the likelihood information and not the posterior that combines the likelihood information with the prior distribution.

To illustrate, consider Leslie's Figure 3. Leslie's "ladder-tree diagram" is reproduced as the top part of Figure 1.** The bottom branches diagram the state for which there is zero net error in the account. Leslie calculates the final audit risk by dividing the probability of receiving "accept" signals from both AR and STD under the condition of "intolerable" error (ME) that has not been detected by internal control by the sum of the probabilities that such audit tests would be obtained under either ME or zero net error condition.*** He implicitly assumes that the zero (net) error case will *always* lead to acceptance by the auditor.

While the assumed outcome of no incorrect rejection signals is likely given that there is zero *net* error in an account, it is also true that even in the absence of error the analytical review procedures and/or the substantive tests of details will sometimes indicate that material error may exist. That is, the auditor may incur the cost of needlessly extending the audit, revising the audit plan, suggesting incorrect adjustments, or withdrawing or qualifying the opinion. For specificity, I have included these branches with incorrect rejection probabilities equal to .05 and .10 respectively.

With the inclusion of these additional branches to the bottom branch of Leslie's ladder tree, we see that the possibility of completing the audit with no indication of material error from any test can arise by the top most route (with probability .0075) or via the top route of the bottom branch (with probability .6412 (.75 × .95 × .90)). The denominator for the FAR calculation is the sum of the probability of the top route and the top route of the bottom branch. Thus, the FAR for BRP is:

$$\text{FAR} = \frac{\text{PPE} \times \text{DIC} \times \text{AR} \times \text{STD}}{(\text{PPE} \times \text{DIC} \times \text{AR} \times \text{STD}) + (1-\text{PPE}) \times (1-\text{RIRAR}) \times (1-\text{RIRSTD})}$$

$$= \frac{.0075}{.0075 + .75 \times .95 \times .90} = .0116$$

where RIRAR and RIRSTD are the risks of incorrect rejection. The formula is identical to that of EAT with the exception of the RIRAR and RIRSTD factors. Since these factors are always $\leqslant 1.0$, the Bayesian risk product will always be greater than or equal to the FAR from the EAT formula. How much less will depend, in part, upon the level of the RIR factors. Thus, the BRP approach will always lead to planned sample sizes that are equal to or greater than those using EAT.

To elaborate on the comparison, the relative sample sizes for SAS 39, SAS 47, EAT, and BRP (with PPE = .75, DIC = .30, AR = .40, and FAR = .03) would be 1.00, .79, 1.77 and 1.88. Thus the sample size for BRP is about 6% greater than that for EAT. A similar calculation using PPE = .25 would yield 1.0, 0, .19 and .31. Thus, for the low PPE case, SAS 47, EAT and BRP require much less substantive tests of details than does SAS 39.

As shown in the figure, the difference in FAR for PPE = .25 is .0017 or about 17% more than FAR using EAT. For the PPE = .75 case, FAR using

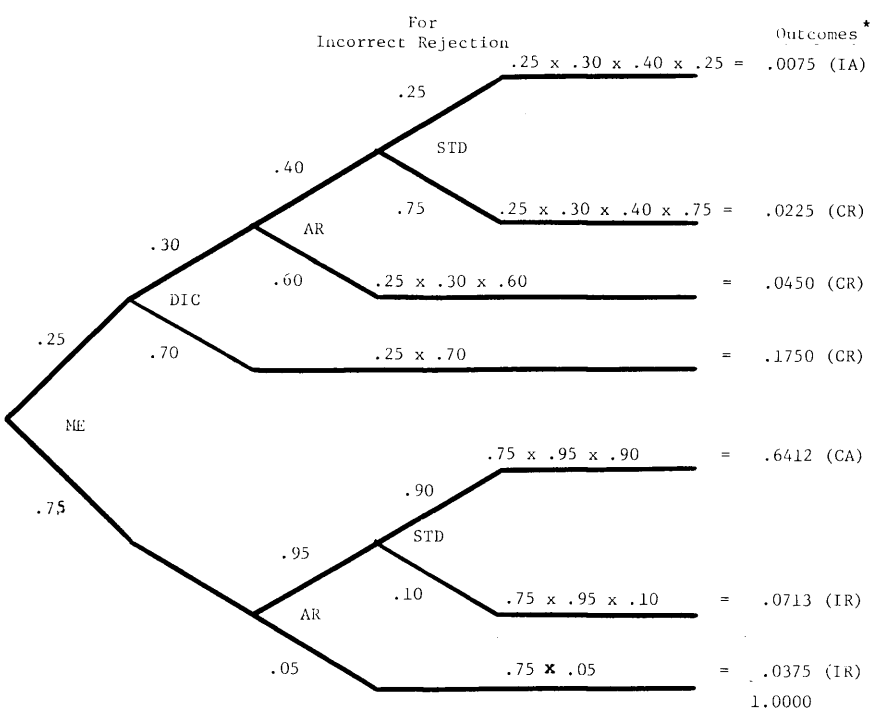
---

** For those familiar with Bayesian decision trees, the Leslie approach differs in that his "ladder tree" places the prior probabilities of states at the start of the tree rather than at the end of each branch.

*** I will use ME to denote "just intolerable" monetary error for the account. Thus it is the single account counterpart of materiality for the financial statements taken as a whole.

Figure 1

Leslie's 'Ladder Tree' Modified



FAR (SAS 39) = 1 x .3 x .4 x .25 = .0300

FAR (SAS 47) = .25 x .3 x .4 x .25 = .0075

FAR (EAT) = $\frac{.0075}{.0075 + .75}$ = .0099

FAR (BRP) = $\frac{.0075}{.0075 + .6412}$ = .0116

* IA denotes "incorrect acceptance", CR denotes "correct rejection", etc.

BRP is .0116 or 15.3% more than EAT. While these differences are not huge, they aren't negligible either. If the EAT formula provides appropriate correction of the SAS 39 formula then the BRP correction to EAT would also seem justified.

As to the direction of the misstatement of risk, Leslie shows that, under his assumptions, the SAS 39 formula is conservative relative to EAT for PPE $\leqslant$ .5 and is not conservative for PPE > .5. He is especially concerned about the use of SAS 39 for PPE > .5 since statement users may not have adequate protection against material error in audited financial statements. Yet it seems

likely that, in the presence of a high expectation of material error, the auditor would probably not use the risk formula in an "hypothesis testing" mode.[1] Rather, the auditor might use an "estimation" approach to adequately estimate the extent of error so that adjustments can be made or perhaps some other strategy pursued in which the client attempts to correct errors. In the latter case, sampling may not even be applicable. Thus, apropriate use of the SAS 39 formula may not lead to excessive risk of incorrect acceptance but may, of course, lead to costly overauditing for the PPE $\leqslant$ .5 cases and would be of concern to an auditor who faces competition.

A reader familiar with Bayesian statistics will note at least two deficiencies in all four formulations of the auditor's problem. First, all approaches assume that ME that is detected by DIC cannot lead to incorrect rejection (i.e., Leslie's original assumption). A more complete formulation would direct the .70 branch from the DIC node to the lower AR node. This change would lead to FAR = .0080 for EAT and FAR = .0094 for BRP. Second, all of the approaches ignore the possibility that the auditor might misassess IR or misassess the reliance to be placed on internal control. For example, the auditor might conclude, on the basis of a study and evaluation of internal control, that DIC is .30 when the correct probability is .50. A complete Bayesian formulation would consider such misevaluations.

Finally, the AICPA models, EAT, and BRP implicitly assume that one of two states must occur—either net error is zero or net error is equal to ME. Outcomes between zero and ME as well as error greater than ME cannot exist. Leslie believes that such simplification may lead to substantially different answers than would a continuous outcome space. I am not convinced that relative sample sizes will be substantially affected by allowing continuous state spaces. I base this view on the lack of meaningful differences for several continuous state follow-up studies of the Kinney (1975) discrete Bayesian model. Also, the imprecision inherent in the auditor's subjective assessments may overshadow any differences due to continuous state modelling.

How Should IR and Internal Control Risks be Assessed?

Another area of considerable controversy in the profession concerns the appropriate way to consider the three client-specific factors related to PPE. These factors are inherent risk (IR), preventive internal controls (PIC), and detective internal controls (DIC). In SAS 47, para. 24, the Auditing Standards Board suggested that the auditor may separately assess IR and "control risk" (PIC and DIC) and presumably multiply them together to get the risk that ME will occur and not be detected by IC. Leslie suggests that IR and PIC be considered together and multiplied by DIC to yield the same risk. A third alternative, also mentioned in SAS 47, para. 24, is to jointly assess the effects of all three factors. That is, the auditor would assess the overall risk that ME would occur in the accounting process and not be prevented or detected by internal controls. The joint method requires no separate assessment of individual factors that have not been separately observed and does not specify a particular combination algorithm (e.g., multiplication).[2]

Using as an analogy, gun control at Kansas City International Airport, we might ask a gun control worker to assess the joint probability that a plane

departing the airport has one or more guns on board. Alternatively, we might ask the worker to estimate the number that would, in the absence of all controls, carry a gun on board and the number that would be caught by the controls. The worker's response to the alternative question is likely to be "How should I know?—I've never observed would-be passengers without the control." Furthermore, the worker is not likely to be able to separately assess PIC and DIC (it seems clear to me that the X-ray machine is detective but has preventive aspects as well). Finally, it is not clear that simple multiplication of separately assessed risks is in order due to lack of independence, because passengers' behavior is conditioned by their expectation of the control components.

I favor the alternative of joint assessment and argued for it at the Auditing Standards Board deliberations on SAS 47. As discussed by the ASB and Don Leslie, behavioral research may lead to a determination of the best method.

Compliance Testing

Leslie's concern about the "5 and 5 will get ya 60" syndrome deserves comment. The example of a .05 tolerable rate (which, in practice, would have been set based on the auditor's judgment about the relationship of the prescribed control to ME) and .05 risk of overreliance due to compliance deviations originated in SAP 54 in 1972 as an example of reasonable levels of the factors. These levels were continued in SAS 39, and an example sample size of 60 was added. Sixty was chosen to show nonstatistical samplers that a compliance sample size that would satisfy the required criteria of "acceptably low" risk that the error rate for the population does not exceed a "reasonable" tolerable rate requires a compliance sample size of considerably greater than, say, 5 or 10—even if no deviations are observed. It is *only* an example and not intended to be a standard itself. Leslie does not cite evidence to support his claim that it has become the standard the world over. Even if it has, however, it has probably led to an increase in compliance sample sizes and, thus, is not necessarily bad.

Furthermore, it is not clear to me that a "smoke to fire" ratio is the only way to assess the risk of overreliance due to the compliance portion of the joint risk of overreliance. There would seem to be various ways to map the design evaluation and rate of deviation (whether based on documents or dollars of book value) to the posterior probability of error.

Is compliance testing *always* necessary for reliance? Leslie raises some challenging questions with respect to required compliance testing for preventive controls. He suggests that a preferred alternative is to look for evidence that such controls have actually detected errors. While such a procedure may be acceptable under some circumstances, in a generally poor control environment employees may document phony "prevention hits" or detected errors that were not really errors. That is, evidence of effectiveness may not indicate application of a control at all. Again, joint assessment of risk may help.

Conclusions

To summarize, I agree with Leslie that the audit risk formula can be improved. For example, it can be modified to be less conservative for low PPE

audit situations. Also, I agree that future Auditing Standards Board consideration should be given to elaboration of the PIC and DIC concepts. Whether official improvement of the formula is possible within the present ASB is, of course, an open question.

It has been my experience at the ASB that the firms that have the most structured audit approaches generally vote in support of structured standards. Similarly, those with the least structured approaches regularly vote against further structure. In drafting standards, there is a desire to make them "read" better by simplifying the wording, and yet complex concepts often require seemingly difficult wording to be correct. The audit risk formula seems to be a victim of such a conflict. The concepts of audit planning and evaluation are complex—a simple formula and few words are not likely to be technically correct. Yet complex formulae and discussion will not receive Board support.

While there is little disagreement that auditors tend to behave as if they are Bayesians, there is much disagreement as to how this should be expressed in professional standards. There is a valid question of how much "theory" should be in professional standards. I do not believe that much more structure is likely to be forthcoming from the ASB. Elaboration of the PIC and DIC concepts is more likely since they are easier to understand.

While the audit risk formula of SAP 54 and SAS 39 may be subject to misuse, it has led to considerable research and experimentation as well as to enlightenment of scholars and practitioners. It indicates relevant audit planning factors and the direction of their effects on auditing procedures. Furthermore, even if a practitioner uses the formula "simplistically," I know of no evidence or even claims that those not using the formula take *larger* samples or conduct more thorough or extensive audits. That is, I have never heard critics of the formula claim that they do not use it because they feel that they should be taking larger samples.

I have heard the opposite claim—that a firm does not use the formula because it leads to sample sizes that are too large. This latter result, however, is often because the auditor has not taken advantage of his or her judgment in designing the sampling applications, rather than the potential problem that Leslie mentions. SAS 39 is quite clear in the statement that nonstatistical sample sizes for a given level of effectiveness cannot be *smaller* than a well designed statistical sample.

I am concerned that those who have been using the SAP 54 and SAS 39 formula may now apply the implicit formula of SAS 47 in a way that increases FAR. It seems to me that many practitioners have already included some allowance for inherent risk in assessing IC under the SAS 39 guidance. That is, they have *implicitly adjusted* the IC factor to include an allowance for inherent risk as well as internal controls. In effect, they use IC as the PPE when planning AR and STD. If they now use the same value for the IC factor and multiply it by an IR less than one, they will double count IR and will unduly restrict STD. The extent of this potential problem can also be addressed by behavioral research, of course.

In conclusion, I believe that Don Leslie has written a thoughtful and very timely paper. Increasing competition and the micro computer are making operational many of the models that we have discussed at this conference for the last ten years, beginning with Felix (1974). Field workers now have the

computing power necessary to do sophisticated audit planning and evidence integration. We scholars can no longer avoid these issues by using the excuse of computational impracticality. Consideration of Leslie's analysis, criticisms, and suggestions can help all of us improve the practice, theory, and, perhaps, the regulation of auditing.

End Notes

1. See Cushing and Loebbecke (1983), for additional discussion.
2. A similar problem with multiplication of risks is noted in Jiambalvo and Waller (1984), p. 87.

References

Canadian Institute of Chartered Accountants, *Extent of Audit Testing: A Research Study* (CICA, 1980).

Cushing, B.E., and Loebbecke, J.K., "Analytical Approaches to Audit Risk: A Survey and Analysis," *Auditing: A Journal of Practice and Theory* (Fall, 1983), pp. 23-41.

Felix, W.L., Jr., "A Decision Theory View of Auditing," *Contemporary Auditing Problems*, Proceedings of the 1974 Arthur Andersen/University of Kansas Symposium on Auditing Problems (1974), pp. 63-71.

Kinney, W.R., Jr., "Decision Theory Aspects of Internal Control System Design/Compliance and Substantive Tests," *Journal of Accounting Research Supplement* (1975), pp. 14-37.

Jiambalvo, J., and Waller, W., "Decomposition and Assessment of Audit Risk," *Auditing: A Journal of Practice and Theory*, Spring, 1984, pp. 80-88.

Leslie, D.A., Teitlebaum, A.D., and Anderson, R.J., *Dollar-Unit Sampling: A Practical Guide for Auditors* (Copp Clark Pitman: 1979).