A Prototype Decision Aid for Internal Control Testing Plan Selection

James M. Peters*¹

Jefferson T. Davis**

January 5, 1999

¹Correspondence concerning this paper should be directed to James M. Peters; The Heinz School, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213-3890, (412) 268-8473, e-mail - jpeters@cmu.edu.

*Decision Systems Research Institute; H. John Heinz III School of Public Policy and Management; Carnegie Mellon University

**Clarkson University

This research is supported in part by a grant from Grant Thornton, LLP. The authors would like to thank Jon Caulkins, Peter Gillet, David Kaplan, Ramayya Krishnan, George Krull, Rema Padman, Michael Shields, Stephen Yates, the participants in seminar series at the Heinz School and Graduate School of Industrial Administration at Carnegie Mellon University, the Katz Graduate School of Business at the University of Pittsburgh, and the University of Kansas Audit Symposium, as well as two anonymous reviewers for their helpful comments on earlier drafts of this paper.
A Decision Aid for Internal Control Testing Plan Selection

SUMMARY

Since 1977, the importance of internal control evaluation (ICE) has increased due to passage of the Foreign Corrupt Practices Act, the Federal Deposit Insurance Corporation Improvement Act, and other regulatory initiatives. Although professional standards require auditors to assess internal control risk at the assertion level and empirical evidence shows differences in error rates by assertion, empirical evidence also shows that auditors do not vary their control risk assessments across assertions. The inconsistency between standards and experienced error rates, and practice could be caused by the complexity of making assertion-level control risk assessments, the lack of professional guidance for making these assessments, or both.

The research reported here takes a preliminary step to providing auditors with decision support for making assertion-level control risk assessments by developing a prototype decision aid that helps auditors select an optimal control testing plan designed to achieve target assertion-level control risk assessments. The aid supports the control testing plan selection decision by modeling accounting information system components based on their impacts on financial statements assertions, providing an evidence combination algorithm based on reliability theory, and identifying all optimal control testing plans. We validated the aid by comparing its testing plans to both experienced auditors and a professional benchmark. The results indicate that the aid’s testing plans test sufficient controls to provide auditors with their desired assurance but do so by testing fewer controls than either experienced auditors or a professional benchmark.

Key Words: Information systems reliability, internal control evaluation, financial statement assertions

Data Availability: The data used for the validation study is available from the second author. A copy of the decision aid software is available from the first author.
A Decision Aid for Internal Control Testing Plan Selection

INTRODUCTION

Passage of the 1977 Foreign Corrupt Practices Act was the first step in a series of regulatory actions by Congress and various federal agencies to increase the quality and quantity of reporting on internal controls for all types of organizations (Cushing, Graham, Palmrose, Roussey and Solomon 1995; Stachowski 1994; Waggoner 1991). Professional pronouncements require that internal control evaluations (ICEs) be made for each financial statement assertion for each account for which the auditor plans to rely on controls in their audit plan (AICPA 1990).

Assessing control risk at the assertion level facilitates the auditor’s tradeoff decision between relying on the accuracy of the information system (i.e., reliance on controls) and relying on substantive tests of the account balances. If controls and compliance tests are characterized in terms of their effects on financial statement assertions, then their relationship with substantive tests, whose effects are also stated in terms of financial statement assertions, is made clearer. Being precise in making this tradeoff is important because accounts differ in their inherent and control risk by assertion (Waller 1993), and different substantive testing procedures test different assertions.

For example, because of management’s incentives, the completeness assertion tends to be violated more for accounts payable than accounts receivable. Confirmations typically are used to substantively test both accounts, but confirmations can test for existence and valuation but not completeness. They can not test for completeness because the confirmation sample is drawn from the detail supporting the recorded balance, not from the populations of potential accounts receivable, which would be practically impossible to do. This difference in management incentives
between the accounts and the capabilities of confirmations probably explains why substantive
testing of accounts payable is usually augmented with other tests for completeness while
substantive testing of accounts receivable is not to the same degree.

Waller’s (1993) results support these conjectures. He found that the rate of detected errors,
even after controlling for detection risk, varies with assertions and accounts. He also found,
however, that auditors do not appear to vary their control risk assessments to reflect these
differences. Instead, they appear to use a “most important” assertion heuristic. They identify the
“most important” assertion for an account, assess its level of risk, and then assign that risk level to
the rest of the assertions for that account. This lack of sensitivity by auditors for differences in
assertion-level risk could lead to inefficient or ineffective control testing plans.

There are several possible reasons for auditors’ lack of sensitivity to differences in
assertion-level error rates while assessing control risk. Traditionally, control risk assessments have
been based on a concept of control objectives or specific error types and not assertions (Elliot
1983; Grant Thornton 1996; Mock and Willingham 1983; Nado, Chams, Delisio, and Hamscher
1996). No standards have been developed for the definitions of these control objectives or error
types and their definitions vary from firm to firm. In addition, no standards or professional
guidance exists for mapping these control objectives or error types to assertions (e.g., Grant
Thornton 1996). Therefore, very little, if any, guidance exists to help auditors assess controls in
terms of their impact on financial statement assertions.

This lack of guidance is confounded by the complexity of the relationships between the
components of an accounting information system (AIS) and assertions as well as the variation in
assertions tested by various compliance tests mentioned above. The processes within the AIS that
transform information may differ in the assertions they can violate. While most processes can violate three assertions (completeness, existence, and valuation) because they can lose transactions, inject erroneous transactions, or alter the value of a transaction, they may differ in the level of risk associated with each type of assertion violation.

Controls may be able to eliminate more than one assertion violation from the information in the AIS and the assertion violations that controls can eliminate may vary from control to control. For example, a control that accounts for all checks by looking for missing check numbers could help eliminate completeness violations by determining if a check had gotten lost, but would be less effective at eliminating existence and valuation violations because it could not determine if a check were valid or written for the correct amount.

The research reported here develops and tests a decision aid that is based on a reliability theory model of an AIS and that selects control testing plans using assertion-level reliability assessments. The aid selects an internal control testing plan given a description of an AIS, a set of assertions for each account in the AIS where an auditor plans to rely on internal controls, and a level of planned reliance for each assertion-account combination. The plan is a list of controls that the auditor needs to test to achieve target reliance at the lowest possible cost assuming the controls are functioning as designed. The aid also produces a list of assertions that cannot be supported at target reliability levels for specific accounts even if all controls are functioning properly and allows the auditor to test the potential reliability implied by a testing plan.

To validate the aid, its testing plans for three cases were compared to those of experienced auditors and a professional benchmark. Due to data limitations, not all features of the aid and its underlying mathematical model could be validated in this study. The results, however, provide
preliminary evidence that the aid’s plans are more complete (prevent or detect more error types) and more efficient (less costly) than those of experienced auditors or the professional benchmark.

The remainder of this paper is organized in four sections. The first section presents a review of the extant ICE modeling research and discusses how this study extends that research. The second section presents an overview of the aid developed in this study, and the third section discusses validation of the aid. The final section discusses the aid’s limitations and directions for future research.

USE OF MODELING IN AIS RELIABILITY EVALUATION

Overview of AIS Reliability Evaluation

Auditors perform ICE to determine how much reliance they should place on the AIS as opposed to relying on substantive testing to provide targeted levels of assurance in account balances. A key component of ICE is deciding which controls in the AIS to test. To develop testing plans, auditors first develop a preliminary understanding of the AIS and its control environment to determine if any level of reliance on controls is justified. If it is, their next step is to analyze the design of the control structure in the AIS and determine which controls, if functioning as designed, would provide the level of assurance concerning account balances that the auditor seeks to achieve (i.e., target reliability) (AICPA 1990). Professional pronouncements and firm policies require that these decisions be made by assertion within account (AICPA 1990; Grant Thornton 1996).

Prior Probabilistic Models

One approach to assisting auditors in making AIS reliability assessments is based on reliability theory and other probabilistic modeling methods (e.g., Cooley and Cooley 1982;
Cushing, 1974; Haskins and Nanni, 1987; Knechel 1983). This research has attempted to develop precise, probabilistic models that assess the control risk for an AIS, or portion of an AIS, given the reliabilities of the AIS’ components (c.f., Bodnar 1975; Cushing 1974; Knechel 1983, 1985; Lea, Adams, and Boykin 1992).

For a variety of reasons, these models have not been widely accepted in practice (Felix and Niles 1988; Waller 1993). One common reason cited for auditors’ lack of use of these models is the computational intractability of the problem (Felix and Niles 1988; Bailey, Gerlach, MacAfee, and Whinston 1981). The ICE task theoretically is NP Complete and, therefore, is computationally intractable (Bailey, et. al 1981). In practice, however, several algorithmic solution methods have been developed that render the problem tractable while closely approximating optimal solutions (Nado, et. al 1996; Kaplan, Krishnan, Padman, and Peters 1997).

These models rarely have included the cost of testing controls or considered financial statement assertions, nor have they been used to develop testing plan recommendations. A few studies have focused on AIS design and have considered the cost of operating, but not testing, controls (e.g., Hamlen, 1980). This cost, however, is of secondary importance to auditors. Only Ahituve, Halpern, and Will (1985), Walls and Turbin (1992), and Nado, et al. (1996) have included control testing costs in their models. Nado et. al’s (1996) COMET system assumes that all controls are equally costly to test and minimizes the number of controls selected in a testing plan. The only model to develop reliability assessments at the assertion level was Lea, et. al (1992).

Two prior quantitative models did select testing plans. Ahituve, et al. (1985) developed an optimal algorithmic model that implemented a sequential test planning procedure. Their model,
which assumed that the auditor’s goal was to determine whether at least one control was failing, developed a sequential testing plan designed to minimize the cost of finding that one control. Their model ranked ordered the controls by testing cost and then moved through the controls from lowest cost to highest, testing controls until a failing one was detected. Their model did not represent processes, only controls; did not allow for multiple error types, multiple control coverage, or multiple controls at the same point in the AIS; and did not distinguish among assertions.

PricewaterhouseCoopers has recently implemented an internal control documentation and evaluation system, called COMET, that also selects a control-testing plan and calculates control risks (Nado, et al 1996). COMET is a proprietary product that has not been made available for external review and evaluation. COMET uses an extension of the AIS modeling language developed in the TICOM project to represent processes and controls within an AIS as directed graph structures (Bailey, Duke, Gerlach, Ko, Meservy, and Whinston 1985). These structures can then be used to simulate how errors might be produced and distributed through the AIS as well as how controls might eliminate these errors. COMET also provides guidance in helping auditors assess how effective controls might be in eliminating errors using a set of heuristic rules that are applied to key attributes of the controls. COMET represents a control’s potential effectiveness as a probability; can then propagate these probabilities to determine their impact on account-level reliability; and can compare them to an auditor-supplied target reliability.

COMET models an AIS at a very high level of detail but does not determine a potential error’s impact on financial statement assertions. A typical COMET description of an AIS would include hundreds of activities, potential errors, and controls. However, this description does not
include which financial statement assertions might be affected by a given error and, therefore, provides limited guidance to auditors in developing tradeoff decisions between compliance and substantive testing approaches.

**Prior Deterministic Models**

Two prior research projects have developed deterministic, heuristic-based ICE models (Kelly 1985; Meservy, Bailey, and Johnson 1986). Both of these models were developed by observing how auditors evaluated internal controls and impounding their processes and heuristic rules into a computer simulation of those cognitive processes. While Kelly’s model was based on one auditor and one case and was not validated on a holdout sample (Kelly 1985), Meservy’s model used a variety of data collection techniques and auditors and was validated on cases not used in its development (Meservy 1985). Meservy’s model identifies both control weaknesses as well as controls that should be tested. The heuristic rules in both these models, however, are idiosyncratic to the purchasing cycle and neither model maps control risks onto financial statement assertions.

**DECISION AID PRESENTATION**

**Overview of Decision Aid**

This research developed a decision aid that identifies all optimal control testing plans given an auditor-supplied description of an AIS and assertion-level target reliabilities by account. An optimal testing plan is one that achieves the auditor’s target reliability at the lowest testing cost. The aid is based on a probabilistic model of an AIS that calculates the reliability a control testing plan could achieve at the account and assertion level. A plan’s achieved reliability is the probability that there will be no material violations of an assertion in an account balance due to
failures in the AIS assuming that all controls in the plan are functioning as designed. A control is considered “functioning as designed” if the measured probability that the control will eliminate an error generated by a given process equals the design probability. The aid’s underlying model is very similar to prior probabilistic models reported in the literature. The aid also includes an algorithm for finding all optimum testing plans and spreadsheets for determining whether a plan is feasible and for doing sensitivity analysis.

Aid Development

The aid was developed with funding from a major international CPA firm that wanted to add additional decision support to their existing internal control documentation software. The development process included reviews of the professional literature, reviews of the firm’s audit manuals and procedures, and interviews with field auditors. The interviews were conducted with six auditors (one senior, three managers, and two partners). Detailed analysis of the transcripts from the interviews led to the selection of control plan testing as the internal control evaluation task the field auditors felt was most in need of additional decision support (Kaplan, et. al 1997).

Under the firm’s current audit approach, auditors were required to identify the accounts and assertions targeted by their testing plans, but were provided with no guidance on how to determine the assertions a process might violate or the assertions a control might support. They also needed to develop control risk assessments but were not provided with guidance on how to combine risks associated with AIS components into account and assertion level risk assessments. For these reasons, we elected to develop an aid that would combine process and control level reliability assessments into account and assertion level reliability assessments and to determine all possible optimal testing plans for a given case.
Next a prototype aid was developed with the cooperation of two managers who were not involved in the first set of interviews. They worked through in detail how they had selected the testing plans for a recent engagement. Detailed analysis of the transcripts from these sessions led to an initial deterministic model that was used as a first approximation of the current aid (Kaplan, et. al 1997). Finally, the deterministic model was converted to a probabilistic one and was validated as discussed below. These two development cases were both audits of banks. To demonstrate the aid’s flexibility, these cases were not used in the validation study. The validation cases involved wholesalers and manufacturers.

**Aid’s Mathematical Model**

**Model Overview**

This section presents the formulation of the aid’s underlying mathematical model and the additional computational processes used by the aid to provide the model with the information it needs to solve for an optimal testing plan. The presentation begins by specifying the probabilities associated with the model’s components and then shows how they are combined into the final optimal model. As with prior probabilistic models, the mathematical model underlying the aid makes certain independence assumptions to avoid having to specify joint probability distributions. These assumptions are discussed when the formulas that depend on them are presented. This section concludes with a description of the additional computation performed by the aid and an example of how it functions. Table 1 presents a summary of the notation used in this section.
The model represents an AIS as a series of information transformation processes that are linked with information flows that terminate in an account balance. Because processes transform information, either by capturing it at the organization’s boundary, or by augmenting, classifying, summarizing, or posting information once it has been captured by the AIS, they can inject errors into the information flowing out of the process. The model defines an error as a material violation of a financial statement assertion. By defining error types as assertion violations the model limits the number of error types, bases error definitions on a mutually exclusive and exhaustive classification scheme (AICPA 1990), and provides a direct link between compliance and substantive testing.

The model defines controls as procedures that can only eliminate and not generate errors. The firm who supported this research defined controls this way because controls do not transform data and errors are generated when data are transformed and not when data are merely reviewed or checked (e.g., Grant Thornton 1991, 1996). Although some controls may correct information and thereby potentially inject errors, by defining a control’s strength as the probability that it will eliminate errors, any chance it would inject an error would be considered in setting its strength parameter. Therefore, the model presented here does not include an error generation parameter for controls.

Controls are described by three parameters: the process where they can eliminate errors, the error types they can eliminate, and their strength. A single measure of a control’s strength is developed by collapsing a control’s detection and correction phases into one step. Most prior research has decomposed a control’s error elimination activity into two steps, detection and correction. While this decomposition probably is an accurate description of how controls work,
the distinction between detection and correction is not very meaningful to auditors. Their concern is whether the control works or not.

A control’s strength measures how effective the control can be if it is operating as designed, not how effective it is in a given AIS. The purpose of a control-testing plan is to determine whether a control is achieving its designed strength for purposes of assessing the reliability of the AIS not to determine the control’s effectiveness directly.

The model also extends prior research by considering the cost of testing a control in developing a control-testing plan. This allows the model to determine the optimal testing plan. An optimal testing plan is one that achieves the auditor’s target reliability at the lowest testing cost. The model represents the auditor’s target reliability as the probability that an assertion and account combination is free of material errors.

Process Risk
As defined in this research, a process’ risk is the probability that a process $j$ will generate an error of type $i$ or $\Pr(\text{E}_{ij})$, where $\text{E}$ is the presence of an error of type $i$ in information flowing from process $j$. Since error generation is a dichotomous event (i.e., either an error is generated or not), $\Pr(\text{E}_{ij}) = 1 - \Pr(\neg\text{E}_{ij})$.

Control Strength
A control’s strength is represented as the conditional probability a control $k$ will eliminate an error or type $i$ if one is generated by process $j$ or $\Pr(\text{N}_{ijk} \mid \text{E}_{ij})$, where $\text{N}$ is the elimination of an error from an information flow. Since this research assumes that controls cannot inject errors into information flows, controls will not eliminate errors that do not exist or $\Pr(\text{N}_{ijk} \mid \neg\text{E}_{ij}) = 0$ and $\Pr
\( (\neg N_{ijk} \mid \neg E_{ij}) = 1 \). In addition, controls are dichotomous. They will either eliminate an error or not. Therefore, \( \Pr(N_{ijk} \mid E_{ij}) = 1 - \Pr(\neg N_{ijk} \mid E_{ij}) \).

**Risk of Error in an Information Flow**

Based on the above definitions of a process’ risk and a control’s strength, the probability that an error of type \( i \) will exist in an information flow coming from process \( j \) where only control \( k \) is present is \( \Pr(F_{ij}) = \Pr(\neg N_{ijk} \mid E_{ij}) \cdot \Pr(E_{ij}) \). Since controls cannot inject errors into an information flow, the only case where errors can get into information flows is when a process generates them and the control associated with that process fails to eliminate them. Since the existence of an error is a dichotomous event, \( \Pr(F_{ij}) = 1 - \Pr(\neg F_{ij}) \).

If more than one control can eliminate errors for a process, then the probability that no errors will exist in an information flow coming from a process (i.e., the reliability of the information flowing from the process) is \( \Pr(\neg F_{ij}) = 1 - \prod \Pr(\neg N_{ijk} \mid E_{ij}) \cdot \Pr(E_{ij}) \) \( \forall k \) in \( S_{ij} \) where \( S_{ij} \) is the set of all controls\(^3\) that can eliminate errors of type \( i \) from information flowing from process \( j \).

This specification assumes that strengths of the controls associated with a specific process and a specific error type are independent of each other. This assumption would most likely be valid except in the case where two or more of the controls that eliminate the same error type from the information flows coming from the same process are performed by the same person. That is, two, essentially redundant, controls were performed by the same person. If the strengths of two or more of these controls were correlated, then the probability of an error of the type eliminated by the controls would increase.
In the extreme case that two or more of these controls were perfectly correlated, then the probability of error would raise to the same level it would have if only one of the controls existed. We do not feel that these conditions would exist frequently enough in practice to justify performing a sensitivity analysis on the model based on violations of this assumption.

**Multiple Process and Control Model**

An AIS contains multiple processes as well as multiple controls. The processes are generally linked in chains where the information flowing out of one flows into another and terminates in an account balance. The probability that there will be no errors of type \( i \) in an account balance \( l \) is \( \Pr(A_{i\mid l}) = \prod \Pr(\neg F_{ij}) \forall j \text{ in } P_{i\mid l} \) where \( P_{i\mid l} \) is the set of all processes that can inject error type \( i \) into account balance \( l \) (i.e., the account’s *dependency set* for error type \( I \)).

This specification assumes that the probability of there being an error of type \( i \) in an information flow coming from one process in \( P_{i\mid l} \) is independent of the probabilities that there will be errors of type \( i \) in the information flows coming from the remaining processes in \( P_{i\mid l} \). This assumption could be violated if either the risks of the processes in \( P_{i\mid l} \) were correlated or the strengths of the controls associated with the processes in \( P_{i\mid l} \) were correlated. We believe that the former would be fairly rare, but the latter would occur if the same control could eliminate an error of type \( i \) in more than one process in \( P_{i\mid l} \), a situation that is likely to occur in practice. Since our model is designed to select a control testing plan and a control can only be selected once for testing, the model’s selection process is based on the “worse case” scenario that a control’s strength at different processes is perfectly correlated.


**Optimization Model**

Since professional pronouncements require error risk to be assessed by account and error type, the probability that there are no material assertion violations (i.e., errors of that type) in an account balance is the auditor’s ultimate concern (AICPA, 1990). Auditors who want to rely on the AIS as a basis for certifying the account balances need to assess the \( \Pr(A_{il}) \) for each assertion violation (i.e., error type \( i \)) and account \( l \) on which reliance on the AIS is planned to determine the degree of reliance that is justified.

To calculate this probability, however, the auditor must confirm the control strengths and error-generating probabilities for enough of the processes and controls in \( S^i \) and \( P^i \) to justify the planned degree of reliance (i.e., target reliability). The confirmation of these probabilities is accomplished by testing controls. Auditors typically do not test all of the controls and processes in \( S^i \) and \( P^i \), but only test key controls that, if working, can provide the planned level of reliance (AICPA 1990; Nado, et al. 1996; Grant Thornton 1991, 1996).

Since testing is costly, the auditor’s task is to determine a set of controls to test (i.e., testing plan) that minimizes cost while ensuring that, if the tests are successful, the planned reliance will be justified. The formulation of this decision task is:
\[
\min \sum_{k \in S} C_k
\]

Subject To:
\[
C_k = 1 \text{ if } k \in S^t, \text{ 0 otherwise } \forall k \in S
\]
\[
\Pr(A_{il}) \geq T_{il} \quad \forall i \in E^t, \forall l \in A^t
\]

Where:
- \( S \) = the set of all controls
- \( S^t \) = the set of controls in the testing plan
- \( E^t \) = the set of target error types
- \( A^t \) = the set of target accounts
- \( T_{il} \) = target reliability for each error type and account

**Aid Implementation**

The decision aid was implemented using a Microsoft Access™ database linked to three Microsoft Excel™ spreadsheets. To use the aid, auditors need to provide the following data:

1. risks for all processes and error types;
2. strengths for each control for each process and error type (includes specifying control-to-error type {assertions} mappings);
3. flows of information between processes and accounts; and
4. target reliabilities for each account and assertion.

Auditors can use the database component to enter these data; select from a variety of descriptive reports to facilitate verification of data accuracy; or use one of the three spreadsheets to solve for missing controls, solve for an optimal testing plan; or perform sensitivity analyses.

To solve for this optimal testing plan, the contents of \( P^t \) and \( S^t \) need to be determined. The decision aid uses a depth-first search algorithm that determines the processes in \( P^t \) by calculating
all dependency sets for all target accounts using a description of the AIS. The search algorithm starts with each account in the AIS and follows all information flows coming into the account back to their source. The algorithm determines which controls should be included in each $S^a_i$ using the control to error type and control to process mappings from the same AIS description.

**Example Case**

Figure 1 contains a representation of one of the cases used in the validation portion of this research (Case 1) and will be used to illustrate how the decision aid functions. Solid boxes in the Figure represent processes and dashed boxes include the set of controls that can eliminate errors in a process. Each control box contains a listing of the error types each control can detect along with its testing cost estimate. For the example, all process risks were set to 0.10, all control strengths to 0.95, and all target reliabilities to 0.95. The testing plans selected by the aid (A), professional benchmark (B), and consensus auditor (C) are indicated with bold capital letters before each control. In this case, there was only one optimal plan selected.

For illustration, the achieved reliability for the completeness assertion at each process is shown within each process box. The achieved reliabilities for all accounts and assertions are shown in the account boxes. The achieved reliabilities at the account level do not change from those at the process immediately before the account because the case materials indicated that the auditors did not believe there was any risk of these processes generating errors.

The aid first used its depth first search to determine that the Sales, Sales Register, Accounts Receivable Ledger, Extended Sale Invoice, Packing Slip, and Sales Invoice processes were all in...
the dependency set for information flowing out of the Sales processes (i.e., Path 1). It did the same for Accounts Receivable (i.e., Path 2).

Each dependency set is the $P^d$ matrix for each account. The aid only includes processes that can inject the error types specified in the target assertions entered by the user in the $P^d$ matrix. While the aid is building the dependency sets, it is also building the $S^d$ matrix that indicates which control can eliminate which error type in which process. Again, only error types listed as target error types are included. With these matrices in place, the aid solves the optimization equation specified above and reports the controls selected (i.e., a vector of $C_k$ values) and the achieved reliability for each account and target error type.

Because the cases used for the validation study were relatively small and the goal of the research was to demonstrate that the concept would work, the aid uses a relatively slow grid search algorithm to identify all optimal plans. If the auditor wants to determine if there is a feasible solution (i.e., target reliabilities can be achieved by testing all controls in the case) before running the optimization algorithm, (s)he can use the aid to do so. If the user wants to experiment with different testing plans, the aid also has a function that allows the user to test the achieved reliabilities for target error types in accounts based on different testing plans.

AID VALIDATION

Validation Method

The validation study is exploratory and designed to provide a prima facie case that the aid can produce testing plans comparable to experienced auditors. Because the number of free parameters in any case is large, standard statistical hypothesis tests were not feasible. Testing plans can vary depending on the following parameters of a case:
1. a process’ risk for each error type;
2. a control’s strength for each error type;
3. the structure of the AIS to include how controls are associated with processes and how information flows between processes;
4. the level of reliability the auditor wants to achieve (i.e., target reliability);
5. the cost to test a each control; and
6. the decision rule the auditor uses to combine the above parameters and choose a testing plan.

A complete test of the aid would require sufficient observations to vary each of these parameters independently. Such a test is beyond the scope of this exploratory study.

In addition, there is no established theory of internal control testing plan selection against which to compare the aid’s plans. Therefore, the aid’s plans were compared to three bases: a professional benchmark (explained below), a consensus auditor (explained below), and an average of experienced auditors.

**Data Coding**

The validation study used archival data from a prior, unrelated study (Davis 1993, 1996) that was based on the current audit methods of a large, international accounting firm. None of the cases used in the validation study were used to develop the aid. As is typical with archival data studies, the data needed to be coded for use in this validation study.

Auditors currently do not define error types for internal control purposes as assertion violations. The firm whose auditors were used in the study, however, does code processes by the control objectives that need to be covered at each process and controls by the control objectives...
they cover at each process. These control objective codings were included in the experimental materials given to the auditors. The authors, both of whom have public accounting experience, used the control objective codings and the definition of assertions found in the professional literature (AICPA 1990) to independently code all of the controls in the cases with the error types they could eliminate. In all cases, processes were coded as being capable of violating all assertions. Therefore, only controls needed to be coded with the assertion violations they could eliminate. The kappa coefficient (Cohen 1960) of agreement for these independent codings was 0.81.

The expected costs of testing controls were not included in the experimental materials so the controls in the Davis cases had to be assigned testing cost. Testing costs can be expressed in a variety of ways, but it is the relative cost of testing different controls that matters to the aid reported here. Therefore, controls were cost coded by classifying them into three broad categories, high, medium, and low, where high was three times more costly than low and medium was 1.5 times more costly than low. Controls independently classified into one of these three categories by the authors. The kappa coefficient for this independent differential cost codings was 0.89. The analysis also was run with all testing costs set to one to determine the sensitivity of the aid’s selection algorithm to differential cost weightings.

For both error type codings of controls and testing cost codings, the kappa coefficients were calculated based on the authors’ independent codings. All coding differences between the authors were resolved by mutual agreement before the model was run on the cases. No changes were made to these codings during the validation analysis.
The strength, risk, and reliability parameters were not coded. The validation analysis was run using several different settings for each. In choosing the levels for these parameters, the authors pick levels that generate a range of account-level reliabilities they felt closely approximated the range usually found in practice. The reliability of the information that reaches an account balance (i.e., account-level reliability) is determined by the risks and strengths of the processes and controls in the information flows leading to the account as well as the number of processes in those flows. Control strengths were set at either 0.95 or 0.90 while process risks were set at 0.02, 0.10, or 0.20.

To determine a rough estimate of the account-level reliabilities these parameter settings would generate in a prototypical case estimated reliabilities were calculated assuming four processes in an information flow and one control per process. The formula for calculating estimated reliabilities was \[1 - (((1 - \text{process risk}) \times \text{control strength})^4]\]. These estimates were used only to gauge the potential account-level reliabilities spanned by the validation data.

Table 2 shows the estimated reliabilities calculated for each of the six combinations of risk and strength as well as the minimum and maximum reliabilities that could be achieved in each of the cases. Maximum reliability assumes all controls are tested and minimum reliability assumes no controls are tested. The reliabilities for each account and error type in a case were then averaged to get an average achievable reliability for each case.

Insert Table 2 about here

The selected ranges capture the upper end of the reliability range fairly well but do not map onto the actual reliabilities for lower reliability scenarios because of the existence in the cases of
redundant controls (i.e., more than one control at a process that can eliminate the same error type). Since auditors would prefer to achieve a reasonable target level or reliability without testing all the controls in the AIS, we believe these parameter settings are adequate to test the aid. In addition, when running the aid through these scenarios, we found a few high reliability scenarios where a target level of reliability could be achieved without testing any controls while there were several low reliability scenarios where the target level of reliability could not be achieved even by testing all the controls. This observation further indicates that our reliability ranges captured most of the scenarios that auditors would have considered in auditing the test cases.

**Benchmark Development**

Davis used cases from two sources, so two types of professional benchmarks were developed. One of the cases was based on an actual client of the firm that provided Davis with subjects. The professional benchmark for that case was the testing plan used on the audit. The other two cases were based on examples from the AICPA’s Audit Guide (AICPA 1990). The Audit Guide examples did not include an explicit control-testing plan. The authors independently developed testing plans for these cases by reviewing the test results in the examples and inferring the tests that were performed. The procedure was performed using a simple decision rule: if the evaluation summary given in the case referred to the results of testing a control, then that control was assumed to have been tested. There were no differences between the two plans independently developed by the authors.

A consensus plan also was developed by using controls that were chosen for each case by at least half the auditors who solved that case. While this method is a crude approximation of a true consensus plan, it does highlight those controls on which the auditors most relied.
Subjects

Although Davis (1993, 1996) studied both experienced and inexperienced auditors, only the data from Davis’ experienced auditor group were used in the validation analysis. In addition, only those auditors who indicated they planned to rely on controls were used because planned reliance was a prerequisite for developing a testing plan.

The auditors were from a single international public accounting firm and participated in Davis’ experiment as part of a training session on internal control documentation and evaluation. The auditors, on average, had 57 months of experience, performed 47 audits, supervised 22 audits, documented the internal control structure in 24 audits, and made preliminary control risk assessments on 32 audits. These auditors also had 18 hours of training with the internal control documentation procedures used in Davis’ study.

Task

Subjects were asked to review a documentation package for one of three cases, indicate whether they planned to rely on controls, and, if so, select a control testing plan (i.e., specify on which controls they planned to rely). Each case presented documentation for the sales stream within the firm’s revenue cycle only. The case documentation was prepared using the firm’s approach, on which the auditors had been trained. The case documentation was not used to develop the aid. These cases included one medium sized retailer and two medium-sized manufacturers. The documentation included some background data on each company, a diagram of the revenue cycle showing all processes, and their connections, and a list of controls and what processes they affected.
The subjects were told to base their testing plans on an audit strategy that would provide support for three assertions: completeness, existence, and valuation. Their plans were limited in this manner because the examples on which the cases were based had limited the audit strategy to these three assertions.

**Analysis Method**

The aid’s testing plans were compared for similarity to the auditors’ plans, the consensus plan, and the professional benchmarks. The comparison was based on three attributes of the plans: the account-level reliability achieved by the plans, the efficiency of the plan, and the correlation between the reliabilities of eliminating the individual error types (defined as assertion violations) within account. We chose to focus on these parameters instead of comparing plans based on the specific controls included in the plans because different combinations of controls could achieve the same level of reliability. Because the aid identifies all possible optimal testing plans and, in several cases, there were multiple equivalent optimal plans, the aid’s statistics were averaged across these equivalent plans.

The account-level reliability a plan can achieve is based on the design strength of the control system and the estimated risk of the processes before compliance testing. It shows the reliability of the AIS assuming that all controls are functioning properly and process risks do not exceed estimated risks. The purpose of the testing plan is to validate this initial reliability assessment by confirming that the controls are functioning as designed and processes are not exceeding their estimated risk. The aid produces separate reliabilities for each account and error-type combination. For simplicity, these were averaged for each case to perform the comparison.
A plan’s efficiency was measured as the ratio of the plan’s cost to the average account-level reliability achieved by the plan. A ratio of cost to reliability was used to reflect the inherent tradeoff between cost and effectiveness of testing plans. In addition, the correlations between the relative reliabilities and efficiencies of plans across accounts and error-types was included to show the similarity of plans in how they balanced these factors among the error-types and accounts in the case.

Validation Results

A summary of the average achieved reliabilities and efficiencies of professional benchmark, consensus auditor, auditors, and aid are presented in Table 3. These statistics were generated by evaluating the benchmark’s, consensus, auditors’, aid’s plans under each of the strength and risk assumptions shown on the left-hand side of the table. Since the aid selects a testing plan that will achieve a target reliability level and the cases did not include a target reliability level, the aid was run at three different target reliability levels. The term “infeasible” indicates combinations of strength and risk where no plan could achieve the target reliability for all accounts and assertions in the case.

Insert Table 3 about here

These data indicate that the professional benchmark achieved a higher level of reliability than either the consensus plan or the auditors. For those combinations of parameter settings where the estimated reliability closely approximates the aids target reliability (i.e., high reliability frontier), the aid’s reliabilities are very similar to the professional benchmark. What these results imply is that the professional benchmark plans are designed to achieve close to the maximum
possible reliability in each case. When the aid is asked to develop a plan near this high reliability frontier, it chooses plans that are very similar to benchmark’s. The aid, however, consistently achieves these reliability levels more efficiently.

Table 4 gives a clearer picture of these comparisons. It shows the differences in reliabilities and efficiencies as a percent of the maximum possible difference for each statistic. For example, the differences in reliabilities for each comparison were divided by the difference between the maximum and minimum possible reliabilities, as described above, for each case and scenario and then converted to a percentage. Standardizing the differences in this way allowed them to be averaged across cases. The differences also were calculated such that negative numbers are considered worse than the professional benchmark and positive numbers better. A positive difference in reliability means that the plan being compared to the benchmark achieved a higher reliability while a positive difference in an efficiency statistic indicates a lower cost than the benchmark.

The results confirm the observations from Table 3. On the high reliability frontier, the aid’s plans achieve virtually the same level of reliability as the benchmark while the consensus plan and auditors’ plans are less reliable. The aid achieved reliability levels similar to the benchmark more efficiently. The consensus plans and auditors’ plans also are less costly than benchmark in all cases, but give up substantial reliability in achieving the lower costs. The aid’s plans are generally about as efficient as the consensus and auditor plans, but achieve a much higher level of reliability.
The correlations in Table 5 show how closely the pattern of reliability and efficiency statistics across accounts and error-types are related between the benchmark, consensus auditor, auditors, and aid. The results are similar to Tables 3 and 4. On the high reliability frontier, the aid shows moderately strong correlation with the professional benchmark while the consensus and auditor plans are much less similar.

**Validation Summary**

The achieved reliability levels of the professional benchmark plans imply that these plans are designed to achieve reliabilities fairly close to the maximum possible for the AIS being tested. When the aid’s target reliability is set to a similar level, it produces plans that are very similar to the professional benchmark except that they are less costly. The aid does a better job of approximating the benchmark than either the consensus audit plan or the average of the experienced auditors in Davis’ study. We believe these data provide sufficient support for the aid to justify further testing.

An additional strength of the aid is that it forces the user to specify control to error type mappings, probability assessments, and target reliabilities, steps that are very important in selecting an optimal testing plan and are not systematically documented in current audit practice. However, the professional literature gives very little guidance on what the control to error type mappings should be, leaving room for future research into whether or not defining errors as assertion violations would improve consensus in making these mappings.
SUMMARY AND CONCLUSIONS

Contributions

This research has led to the development of a prototype decision aid for the internal control testing plan selection task of ICE. This task is an important part of ICE that has not been extensively studied. The aid can provide auditors with a proposed solution to a specific key control selection task as well as a method to test for feasible solutions and perform sensitivity analyses on their own testing plans. The aid’s outputs should help auditors ensure both the effectiveness and efficiency of their control testing plans.

The aid was developed using field research and literature reviews to help ensure external and ecological validity. The aid used a reliability theory-based and optimal plan selection algorithm to ensure its accuracy, completeness, and internal consistency. The decision aid also includes algorithms to produce the data matrices required by the optimal solution algorithm. The aid has been implemented in a database management system that helps establish the validity of the audit approach underlying it by allowing auditors to set up cases and providing them with a variety of descriptive and diagnostic reports.

The aid was partially validated by comparing its testing plans to both experienced auditors and professional benchmarks. The results indicated that the aid compares favorably with both measures. The results, however, are exploratory in nature and may not generalize to all audit clients because:

1. subjects in the validation study came from only one accounting firm;

2. the data in the test cases were incomplete and required additional judgmental coding by the authors;

3. not all aspects of the aid were tested due to unavailability of data in the test cases;
4. only three error-types (completeness, existence, valuation) were used in the test cases; and

5. only three test cases were used for validation.

**Future Research**

More detailed testing of the aid is needed to gain clearer picture of how it qualitatively compares with the auditors’ and professional benchmark’s plans and to determine whether auditors using the aid will develop better testing plans. In addition, the aid implements an audit approach that requires coding of process to the financial statement assertions they can violate and controls to the financial statement assertions they can eliminate. Finally, the aid requires auditors to make probability assessments, a task with which they are probably not familiar.

Several follow-on research studies are needed to further validate the aid and the audit approach it embodies. First, studies are needed to determine whether using assertion violations as a definition of error types will improve consensus among auditors and the efficiency and effectiveness of their testing plans. The auditors in the Davis study used to validate the aid exhibited very low levels of consensus in the controls they included in their testing plans. We speculate that this lack of consensus might be due to ambiguities in the definitions of control objectives, lack of guidance on how to select a testing plan, or both. We are running a study to determine whether auditors can be trained to use assertion violations as definitions of error types in making internal control decisions and whether doing so increases consensus over using control objectives. The CPA firm conducting the data collection portion of the study has not returned the data set to us yet.
Second, a better test of the qualitative nature of the aid’s testing plan should be run by having a panel of experts do a blind review of the testing plan it selects in several cases. To properly test the aid, the cases should be developed to include all the information the aid uses to select a testing plan. Due to the extensive amount of expert practitioner time needed to develop these cases and evaluate the aid’s results, we have not initiated this study yet and are waiting to review the results of the study mentioned above.

Finally, more studies are needed to determine how effective auditors can be in assessing the probabilities needed by the aid. There is a rich set of research studies on eliciting probabilities for human subjects that show mixed results, but there are no studies that model the specific context of this decision task. With training and the use of standardized heuristic rules like those used in the COMET system, auditors may be able to reliably assess the necessary probabilities.
Footnotes

1. To improve readability, we use the term "eliminate" to mean either prevent or detect and correct for the balance of this paper. Our model does not distinguish between preventative and detective controls. It focuses instead on a control's ability to ensure that errors are not present in an information flow, regardless of whether the control prevents them from entering the flow or detects them in a flow and corrects them.

2. NP stands for Non-deterministic Polynomial-time. A problem is NP-complete if it admits only exponential-time solutions. NP-complete problems can be solved in a reasonable amount of computing time only when the problem is small. Therefore, NP-complete problems are considered computationally intractable.

3. More accurately, $S^{ij}, P^i$, and $S^i$ are the set of events where each event is the existence of a control or process in an AIS. We used simplified terminology to improve readability.

4. In Davis’ study, subjects also were asked to provide a preliminary control risk assessment, but those data were not used in this study.
Bibliography


Grant Thornton. 1996. *Accounting and Auditing Manual*, Grant Thornton LLP.


Accounts Receivable
Achieved:
C = 0.980149
E = 0.980149
V = 0.994254

Sales
Achieved:
C = 0.975249
E = 0.975249
V = 0.989238

sales register
Achieved:
C = 0.975249

Sales Register
Achieved:
C = 0.975249

A/R Ledger
Achieved:
C = 0.980149

Extended Sale Invoice
Achieved:
C = 0.985075

Packing Slip
Achieved:
C = 0.990025

Sales Invoice
Achieved:
C = 0.995

Parameter Settings:
Process risks = 0.1
Control strengths = 0.95
Target reliabilities = 0.95

Key:
Solid boxes are processes
Dashed boxes contain controls associated with processes
Bold letters after controls are the assertions they cover. Numbers are costs.

A, B, C Matching to previously valid extended sale invoice (C, E, V:3)

Compare A/R batch totals (C, E, V:1)

A, B, C Checking to previously valid extended sale invoice (C, E, V:3)

A, B Checking invoice one to one (C, V:3)

A, B Driver obtains signature on sales invoice (E, V:1.5)

B, C Matching to previously valid invoice (C, V:3.0)

A, B, C Checking invoice one to one (C, V:3)

A, B Driver obtains signature on sales invoice (E, V:1.5)

Reviews and compares invoice to packing slip (C, E, V:3)

C Matching to authorized list (V:1.5)

A, B, C Checking invoice one to one (C, V:3)

A, B Driver obtains signature on sales invoice (E, V:1.5)

Figure 1
Test Case 1
<table>
<thead>
<tr>
<th>Element</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>Index for error types</td>
</tr>
<tr>
<td>$j$</td>
<td>Index for processes</td>
</tr>
<tr>
<td>$k$</td>
<td>Index for controls</td>
</tr>
<tr>
<td>$l$</td>
<td>Index for accounts</td>
</tr>
<tr>
<td>$E$</td>
<td>A process injects an error into an information flow</td>
</tr>
<tr>
<td>$N$</td>
<td>A control eliminates an error from an information flow</td>
</tr>
<tr>
<td>$F$</td>
<td>An error exists in an information flow</td>
</tr>
<tr>
<td>$A$</td>
<td>No errors in an account balance</td>
</tr>
<tr>
<td>$\neg$</td>
<td>Not</td>
</tr>
<tr>
<td>$S$</td>
<td>Set of all controls</td>
</tr>
<tr>
<td>$S^i$</td>
<td>Set of all controls that can eliminate errors of type $i$ from information flows coming from process $j$.</td>
</tr>
<tr>
<td>$S^d$</td>
<td>The set of all controls that can eliminate errors or type $i$ from one or more processes in $P^d$.</td>
</tr>
<tr>
<td>$P^d$</td>
<td>The set of all processes that can inject errors of type $i$ into information flows that reach account $l$, i.e., the accounts dependency set</td>
</tr>
<tr>
<td>$S^t$</td>
<td>Set of all controls in the testing plan</td>
</tr>
<tr>
<td>$E^t$</td>
<td>The set of target error types</td>
</tr>
<tr>
<td>$A^t$</td>
<td>The set of target accounts</td>
</tr>
<tr>
<td>$\Pr(E_j)$</td>
<td>Probability that process $j$ will generate an error of type $i$, i.e. the process’ risk</td>
</tr>
<tr>
<td>$\Pr(N_{jk} \mid E_j)$</td>
<td>The probability that control $k$ will eliminate errors of type $i$ generated by process $j$ from an information flow, i.e. the controls strength</td>
</tr>
<tr>
<td>$\Pr(E_j \mid C_k)$</td>
<td>Probability that process $j$ will inject an error of type $i$ into an information flow given control $k$ is present</td>
</tr>
<tr>
<td>$\Pr(\neg E_j \mid S^i)$</td>
<td>The reliability of information in a flow or the probability that no error of type $i$ will exist in the information flowing from process $j$ given the set of controls $S^i$ exists.</td>
</tr>
<tr>
<td>$\Pr(\neg E_j \mid S^i, P^d)$</td>
<td>Probability that there will be no errors of type $i$ in account balance $l$.</td>
</tr>
</tbody>
</table>
Table 2

<table>
<thead>
<tr>
<th>Control Strength</th>
<th>Process Risk</th>
<th>Estimated Reliability</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Average</th>
<th>Max/Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.950</td>
<td>0.020</td>
<td>0.996</td>
<td>0.9981</td>
<td>0.9989</td>
<td>0.9976</td>
<td>0.9982</td>
<td>Maximum/Minimum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.992</td>
<td>0.9959</td>
<td>0.9974</td>
<td>0.9951</td>
<td>0.9961</td>
<td>Maximum/Minimum</td>
</tr>
<tr>
<td>0.900</td>
<td>0.020</td>
<td>0.980</td>
<td>0.9903</td>
<td>0.9943</td>
<td>0.9881</td>
<td>0.9909</td>
<td>Maximum/Minimum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.961</td>
<td>0.9795</td>
<td>0.9871</td>
<td>0.9758</td>
<td>0.9808</td>
<td>Maximum/Minimum</td>
</tr>
<tr>
<td>0.950</td>
<td>0.100</td>
<td>0.961</td>
<td>0.9806</td>
<td>0.9886</td>
<td>0.9763</td>
<td>0.9819</td>
<td>Maximum/Minimum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.959</td>
<td>0.9593</td>
<td>0.9744</td>
<td>0.9519</td>
<td>0.9619</td>
<td>Maximum/Minimum</td>
</tr>
<tr>
<td>0.900</td>
<td>0.200</td>
<td>0.922</td>
<td>0.986</td>
<td>0.9550</td>
<td>0.3195</td>
<td>0.3477</td>
<td>Maximum/Minimum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.368</td>
<td>0.3686</td>
<td>0.3550</td>
<td>0.3195</td>
<td>0.3477</td>
<td>Maximum/Minimum</td>
</tr>
</tbody>
</table>
## Table 3

<table>
<thead>
<tr>
<th>Control Strength</th>
<th>Process Risk</th>
<th>Estimated Reliability</th>
<th>Professional Benchmark</th>
<th>Consensus Auditor</th>
<th>Auditors</th>
<th>Model Target .90</th>
<th>Model Target .95</th>
<th>Model Target .99</th>
<th>Reliability</th>
<th>Unweighted Costs</th>
<th>Weighted Costs</th>
<th>Efficiency</th>
<th>Unweighted Costs</th>
<th>Weighted Costs</th>
<th>Efficiency</th>
<th>Unweighted Costs</th>
<th>Weighted Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.950</td>
<td>0.020</td>
<td>0.996</td>
<td>0.9950</td>
<td>0.9664</td>
<td>0.9589</td>
<td>0.9126</td>
<td>0.9638</td>
<td>0.9963</td>
<td></td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7.7055</td>
<td>3.1017</td>
<td>4.0859</td>
<td>0.3645</td>
<td>2.4223</td>
<td>5.0195</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11.7244</td>
<td>6.7037</td>
<td>7.1766</td>
<td>0.3646</td>
<td>3.6280</td>
<td>7.5269</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.900</td>
<td>0.020</td>
<td>0.992</td>
<td>0.9920</td>
<td>0.9634</td>
<td>0.9560</td>
<td>0.9124</td>
<td>0.9611</td>
<td>0.9930</td>
<td></td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7.7276</td>
<td>3.1114</td>
<td>4.6568</td>
<td>0.3647</td>
<td>2.4292</td>
<td>5.3722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11.7592</td>
<td>6.7245</td>
<td>7.2023</td>
<td>0.3649</td>
<td>3.6069</td>
<td>8.2150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.950</td>
<td>0.100</td>
<td>0.980</td>
<td>0.9750</td>
<td>0.8418</td>
<td>0.8106</td>
<td>0.9815</td>
<td>0.9767</td>
<td>infeasible</td>
<td></td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7.8703</td>
<td>3.5896</td>
<td>4.7283</td>
<td>4.7572</td>
<td>4.7572</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11.9686</td>
<td>7.6807</td>
<td>8.2253</td>
<td>7.6819</td>
<td>7.6819</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.900</td>
<td>0.100</td>
<td>0.961</td>
<td>0.9606</td>
<td>0.8285</td>
<td>0.8008</td>
<td>0.9628</td>
<td>0.9641</td>
<td>infeasible</td>
<td></td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7.9841</td>
<td>3.6443</td>
<td>4.7884</td>
<td>4.8532</td>
<td>5.1929</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12.1482</td>
<td>7.7961</td>
<td>8.3035</td>
<td>7.7804</td>
<td>8.1214</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.950</td>
<td>0.200</td>
<td>0.961</td>
<td>0.9504</td>
<td>0.7068</td>
<td>0.6651</td>
<td>0.9634</td>
<td>0.9647</td>
<td>infeasible</td>
<td></td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8.1027</td>
<td>4.4589</td>
<td>5.9448</td>
<td>4.8501</td>
<td>5.1898</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12.3108</td>
<td>9.4248</td>
<td>10.1281</td>
<td>7.7757</td>
<td>8.1162</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.900</td>
<td>0.200</td>
<td>0.922</td>
<td>0.9222</td>
<td>0.6835</td>
<td>0.6478</td>
<td>0.9293</td>
<td>infeasible¹</td>
<td>infeasible</td>
<td></td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8.3396</td>
<td>4.5892</td>
<td>6.0870</td>
<td>5.3958</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12.6845</td>
<td>9.6939</td>
<td>10.3689</td>
<td>8.4239</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹Infeasible in two of three cases. Data from one feasible case excluded to be comparable with other cells.

Indicates high reliability frontier.
Table 4

<table>
<thead>
<tr>
<th>Control Strength</th>
<th>Process Risk</th>
<th>Estimated Reliability</th>
<th>Consensus Auditor to Benchmark</th>
<th>Auditors to Benchmark</th>
<th>Model to Benchmark .90</th>
<th>Model to Benchmark .95</th>
<th>Model to Benchmark .99</th>
<th>Reliability</th>
<th>Unweighted Costs</th>
<th>Weighted Costs</th>
<th>Efficiency</th>
<th>Unweighted Costs</th>
<th>Weighted Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.950</td>
<td>0.020</td>
<td>0.996</td>
<td>(31.33)</td>
<td>(167.83)</td>
<td>(91.66)</td>
<td>(34.61)</td>
<td>1.52</td>
<td>Reliability</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td>Efficiency</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.900</td>
<td>0.020</td>
<td>0.992</td>
<td>(32.06)</td>
<td>(37.87)</td>
<td>(90.65)</td>
<td>(35.06)</td>
<td>1.13</td>
<td>Reliability</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td>Efficiency</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.950</td>
<td>0.100</td>
<td>0.980</td>
<td>(34.23)</td>
<td>(39.62)</td>
<td>1.75</td>
<td>1.75</td>
<td>infeasible</td>
<td>Reliability</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td>Efficiency</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.900</td>
<td>0.100</td>
<td>0.961</td>
<td>(34.85)</td>
<td>(39.57)</td>
<td>0.66</td>
<td>1.92</td>
<td>infeasible</td>
<td>Reliability</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td>Efficiency</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.950</td>
<td>0.200</td>
<td>0.961</td>
<td>(38.11)</td>
<td>(41.96)</td>
<td>2.08</td>
<td>2.28</td>
<td>infeasible</td>
<td>Reliability</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td>Efficiency</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.900</td>
<td>0.200</td>
<td>0.922</td>
<td>(38.58)</td>
<td>(41.74)</td>
<td>1.17</td>
<td>1.15</td>
<td>infeasible</td>
<td>Reliability</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td>Efficiency</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
</tr>
</tbody>
</table>

Infeasible in two of three cases. Data from one feasible case excluded to be comparable with other cells.  
Indicates high reliability frontier.
## Table 5

### Average Correlations Across Cases for Account and Error Type with Benchmark

<table>
<thead>
<tr>
<th>Control Strength</th>
<th>Process Risk</th>
<th>Estimated Reliability</th>
<th>Consensus Auditor</th>
<th>Auditors</th>
<th>Model Target .90</th>
<th>Model Target .95</th>
<th>Model Target .99</th>
<th>Reliability</th>
<th>Unweighted Costs</th>
<th>Weighted Costs</th>
<th>Efficiency</th>
<th>Unweighted Costs</th>
<th>Weighted Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.950</td>
<td>0.020</td>
<td>0.996</td>
<td>0.2975</td>
<td>0.3634</td>
<td>0.0119</td>
<td>(0.0330)</td>
<td>0.7835</td>
<td>Reliability</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td>Efficiency</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2985</td>
<td>0.3699</td>
<td>0.0035</td>
<td>(0.0312)</td>
<td>0.7814</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2985</td>
<td>0.3727</td>
<td>0.3430</td>
<td>(0.0328)</td>
<td>0.7834</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3305</td>
<td>(0.0310)</td>
<td>0.7834</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.900</td>
<td>0.020</td>
<td>0.992</td>
<td>0.3039</td>
<td>0.3801</td>
<td>0.0302</td>
<td>0.0310</td>
<td>0.6545</td>
<td>Reliability</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td>Efficiency</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3047</td>
<td>0.3641</td>
<td>0.0227</td>
<td>0.6456</td>
<td>0.7906</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3047</td>
<td>0.3602</td>
<td>0.3593</td>
<td>0.0312</td>
<td>0.6544</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3481</td>
<td>0.4547</td>
<td>0.7923</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.950</td>
<td>0.100</td>
<td>0.980</td>
<td>0.2952</td>
<td>0.3608</td>
<td>0.7836</td>
<td>0.7836</td>
<td>infeasible</td>
<td>Reliability</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td>Efficiency</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3005</td>
<td>0.3412</td>
<td>0.7836</td>
<td>0.7836</td>
<td>0.7836</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3005</td>
<td>0.3384</td>
<td>0.7833</td>
<td>0.7833</td>
<td>0.7833</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.900</td>
<td>0.100</td>
<td>0.961</td>
<td>0.3018</td>
<td>0.3816</td>
<td>0.7978</td>
<td>0.7854</td>
<td>infeasible</td>
<td>Reliability</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td>Efficiency</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3057</td>
<td>0.3601</td>
<td>0.8017</td>
<td>0.7346</td>
<td>0.7346</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3057</td>
<td>0.3565</td>
<td>0.7971</td>
<td>0.7850</td>
<td>0.7850</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8011</td>
<td>0.7341</td>
<td>0.7341</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.950</td>
<td>0.200</td>
<td>0.961</td>
<td>0.2921</td>
<td>0.3624</td>
<td>0.7838</td>
<td>0.7609</td>
<td>infeasible</td>
<td>Reliability</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td>Efficiency</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3034</td>
<td>0.3376</td>
<td>0.7838</td>
<td>0.7609</td>
<td>0.7609</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3034</td>
<td>0.3351</td>
<td>0.7831</td>
<td>0.7605</td>
<td>0.7605</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.7831</td>
<td>0.7043</td>
<td>0.7043</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.900</td>
<td>0.200</td>
<td>0.922</td>
<td>0.2991</td>
<td>0.3835</td>
<td>0.7666</td>
<td>0.7184</td>
<td>infeasible(^1)</td>
<td>Reliability</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
<td>Efficiency</td>
<td>Unweighted Costs</td>
<td>Weighted Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3075</td>
<td>0.3549</td>
<td>0.7656</td>
<td>0.7184</td>
<td>0.7184</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3075</td>
<td>0.3516</td>
<td>0.7174</td>
<td>0.7174</td>
<td>0.7174</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)Infeasible in two of three cases. Data from one feasible case excluded to be comparable with other cells.

Indicates high reliability frontier.