A Computational Model of Internal Control Testing Plan Selection

James M. Peters*[1]

Jefferson T. Davis**

January 23, 1998

[1]Correspondence 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, David Kaplan, Ramayya Krishnan, George Krull, Rema Padman, Michael Shields, Stephen Yates, and the participants in seminar series at the Heinz School and Graduate School of Industrial Administration at Carnegie Mellon University and the Katz Graduate School of Business at the University of Pittsburgh for their helpful comments on earlier drafts of this paper.

A Computational Model of Internal Control Testing Plan Selection

Abstract

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. Traditional audit approaches structure the description and documentation of control systems, but do not provide systematic, precise evaluation of accounting information system (AIS) reliability. Although researchers have developed quantitative evaluation methods that provide precise AIS reliability evaluations, practitioners have not adopted them because they become intractable when applied in practice.

This research develops an optimal, mathematical model of an ICE task, internal control testing plan selection, that maintains tractability by solving a part of the overall ICE task and by making simplifying assumptions based on field research. The model selects an optimal control-testing plan given a description of an AIS and the auditor's desired type of assurance in an account balance. The model was validated by comparing its testing plans to both experienced auditors and a professional benchmark. The results indicate that the model'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 the professional benchmark.

A Computational Model of Internal Control Testing Plan Selection

1. 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). Internal control evaluation (ICE) consists of a series of tasks that produce several different outcomes. In performing ICE, auditors:

1. develop and document a preliminary understanding of the client's accounting information system (AIS) and internal controls to determine a level of planned reliance on those controls;

2. select an internal control testing plan that will determine whether the planned reliance is justified; and

3. execute the testing plan, evaluate the results, and assess the level of control risk for the client's AIS (AICPA, 1990).

This process produces a documentation of the client's AIS and internal controls, a documentation of the auditor's testing plan and results, a summary of control weaknesses, and an assessment of the AIS's control risk. The control risk assessment is not an overall assessment of the AIS, but a series of assessments for each account and assertion combination (AICPA, 1990).

One approach to improving ICE in auditing research has been to develop stochastic models of AISs and use them for ICE (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's components. (C.f., Bodnar, 1975; Cushing, 1974; Knechel, 1983, 1985; Lea, Adams, and Boykin, 1992). Practitioners, however, have not used these models because:

1. they require a large number of parameter estimations (e.g., probability assessments);
2. they are considered computationally intractable in spite of their use of simplifying assumptions; and
3. they have not been submitted to field testing to confirm their external validity and robustness (Felix and Niles, 1988; Knechel, 1983; Waller, 1993).

Developing tractable, externally valid quantitative models for ICE reliability assessment is extremely difficult because ICE reliability assessment is an NP-complete[1] problem in which computational complexity increases exponentially with the number of parameters (Bailey, Gerlach, MacAfee, and Whinston, 1981). Auditors also have failed to use these models because they are unaccustomed to making probability assessments (AICPA, 1990; Grant Thornton, 1991, 1996; Waller, 1993).

Practitioners have attempted to improve ICE judgment precision by using structured methods to describe and document internal control systems (e.g., Elliot, 1983; Grant Thornton, 1991, 1996; Mock and Willingham, 1983). The evaluation of internal control systems is still left to auditor judgment (AICPA, 1990; Chen and Lee, 1992; Felix and Niles, 1988; Gadh, Krishnan, and Peters, 1993, Grant Thornton, 1996). Price Waterhouse has recently implemented a control documentation and evaluation system, called COMET, that also selects a control-testing plan and calculates control risks. COMET is a proprietary product that has not been made available for external review and evaluation. In addition, it does not employ a stochastic model that guarantees optimal testing plan selection (Nado, Chams, Delisio, and Hamscher, 1996).

The research reported here develops and tests a tractable stochastic internal control evaluation model. The model was developed and tested by:

1. modeling a single ICE task - testing plan selection - rather than overall AIS reliability assessment;

2. using extensive field research and literature reviews to select a set of simplifying assumptions that are analogous to practitioners' approach to complexity reduction;

3. grouping all possible errors into three mutually exclusive and exhaustive categories, thus reducing the number of parameters in the model; and

4. comparing the model's decisions to those of experienced auditors in realistic cases.

The model 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. The model also produces a list of assertions that cannot be supported for specific accounts, given the AIS's structure and its controls.

To validate the model, 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 model and decision aid could be validated in this study. The results, however, indicate that the model'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 model developed in this study and the third section discusses validation of that model. The final section discusses the model's limitations and directions for future research.

2. *Modeling AIS Reliability Evaluation*

2.1. OVERVIEW OF AIS RELIABILITY EVALUATION

Auditors evaluate the reliability of an AIS to determine how much reliance they should place on the AIS. The ICE process occurs in three steps. First, auditors document the existing AIS and develop a general understanding of its components and the environment in which it functions. Second, based on this preliminary understanding, the auditor decides where to potentially place reliance on the AIS, assuming it is functioning properly. This determination is made for each assertion and account. To determine whether the AIS is functioning properly, the auditor selects a set of controls to test. Third, the auditor evaluates the test results and assesses the level of reliance to place on the AIS (AICPA, 1990). This evaluation process has four primary outcomes: (1) documentation of the client's AIS and internal controls; (2) a control testing plan; (3) a list of control weaknesses to report to management; and (4) an assessment of the risk of error[2] in the financial statements due to the unreliability of the control system given test results (AICPA, 1990).

Most of the prior quantitative ICE research has addressed the fourth outcome listed above, reliability assessment (Bodnar, 1975; Cooley and Cooley, 1982; Cushing, 1974, Hamlen, 1980; Knechel, 1983, 1985; Lea, Adams, and Boykin, 1992). All these studies use probabilistic models to develop an overall assessment of an AIS's reliability[3]. They calculate an overall reliability assessment by combining assessments of reliability for accounts and error types.

Prior quantitative models are very similar. They all specify the components of an AIS; assign one or two probabilistic parameters to those components by types of errors; make some assumptions about the independence, or lack there of, among those components; and then use the laws of probability to combine the parameters for each component into an AIS reliability assessment (Bodnar, 1975; Cooley and Cooley, 1982; Cushing, 1974; Hamlen, 1980; Knechel, 1983, 1985; Lea, et al., 1992). The main differences among these models have been the assumptions on which they are based (Knechel, 1983, 1985).

The prior research, however, does not independently model control testing plan selection and AIS reliability assessment, which is based on testing results evaluation. Auditors separate testing plan selection from AIS reliability assessment to reduce the cost associated with AIS reliability assessment. They test only those controls that are key to providing the required reliability for accounts and assertions where reliance is reasonably possible. This reduces the number of parameters auditors have to estimate to assess the reliability of the AIS. The next section presents the basic model used by prior studies and shows how it can be modified to select a testing plan rather than assess the reliability of an AIS.

2.2. PRIOR QUANTITATIVE MODELS

The components of an AIS that are relevant to ICE are its error types, processes, controls, information flows, and accounts. **Errors** can be classified into **types** that can be generated and prevented. From a quantitative modeling standpoint, the key attributes of error types are how many of them are assumed to be possible in the AIS and whether they are mutually exclusive.

Processes capture or transform information in some way and, therefore, can inject errors into **information flows** (i.e., the information flowing from the process). For quantitative modeling, the important attributes of a process are the probability it will generate an error (i.e., its **risk**), the types of errors it can generate, and its location in the AIS (Grant Thornton, 1991, 1996; Knechel, 1983, 1985).

In specifying a process' risk, researchers have had to specify three independence characteristics of processes and their relationships with other AIS components. These are:

1. whether the generation of one error type by a process is independent of the generation of other error types (**process to error independence**);

2. whether the generation of an error type by a process is independent of the error generating probabilities of other processes in the AIS (**process to process independence)**; and

3. whether the error generating probability of a process is independent of the existence and error prevention abilities of controls in the AIS (**process to control independence)**.

Controls eliminate errors generated by processes from information flows either by detecting and correcting those errors or by preventing[4] processes from generating errors. From a quantitative modeling standpoint, a control's critical attributes are the probability it will eliminate an error given one is generated by a process (i.e., its **strength**), what error types it can eliminate, the processes from which it can eliminate those error types, and the probability that the control will generate an error (i.e., the control's **risk**).

In specifying a control's strength and risk, researchers have had to specify three independence characteristics of controls and their relationships to other AIS components. These are:

1. whether the elimination or generation of one error type by a control is independent of the elimination or generation of other error types (**control to error independence**);

2. whether the elimination or generation of an error type by a control is independent of the error elimination or generation probabilities of other controls in the AIS (**control to control independence**); and

4. whether the error elimination or generation probability of a control is independent of the error generation probabilities of processes (**control to process independence**).

Accounts are the terminal point of the AIS, where the information flows stop. The auditor's primary concern is that the information reaching an account is free of specific types of errors at a desired level of certainty.

The number of probabilities that are required to fully describe an AIS increase multiplicatively with the number of error types generated by processes and/or eliminated by controls and with the existence of dependencies among processes and controls. Therefore, most prior research has assumed independence among AIS components and limited a processes' error-generating ability to either one error type or multiple independent error types (Knechel, 1983, 1985). Table 1 summarizes the assumptions made by prior research.

Insert Table 1 about here

Since most of the prior research has focused on assessing AIS' reliability, cost of testing controls has rarely been considered (Knechel, 1983). A few studies have focused on AIS design and have considered the cost of operating 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. The COMET system assumes that all controls are equally costly to test and minimizes the number of controls selected in a testing plan (Nado, et al., 1996).

Two prior quantitative models did select testing plans. Ahituv, 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 rank 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, and did not allow for multiple error types, multiple control coverage, or multiple controls at the same point in the AIS. COMET (Nado, et al., 1996) also selects a testing plan, but the details of how it does so are not available in the literature.

2.3. MODELING ASSUMPTIONS AND EXTENSIONS OF THIS RESEARCH

This research develops a quantitative model of ICE that achieves tractability by limiting the decision task, selecting simplifying assumptions similar to the ones auditors currently use, and specifying controls, processes, and error types at a more general level of abstraction than previously used. Finally the model includes cost of controls.

2.3.1 Limiting the Decision Task

First, the model is limited to the part of the first and all of the second of ICE discussed in Section 1. above: identifying missing controls and selecting a testing plan. This limitation was made because these tasks naturally precede reliability assessment. Auditors use these two steps to focus their testing efforts on subsets of processes and controls that have audit significance and ignore processes and controls that are not critical to their audit objectives. This reduces the number of parameters that auditors must assess to implement a quantitative ICE model. In addition, the auditors in the field study portion of this research indicated that testing plan selection was problematic and that they had little professional guidance on how to perform it.

2.3.2 Simplifying Assumptions

Second, the model is based on a series of simplifying assumptions about processes, controls, error types, and the interaction of these AIS components. Table 1 summarizes the similarities and differences in assumptions made by prior research and the model. These assumptions were developed using field research and professional literature review. This method for selecting simplifying assumptions allows the model to achieve tractability in a manner similar to how practitioners frame the task. Many of these assumptions are the same as prior research in quantitative modeling. The assumptions used in this model and made in prior research are:

1. A process' risk for one error type is independent of its risk for other error types.
2. A process' risk is independent of other processes' risk.
3. A process' risk is independent of the existence or characteristics of any controls in the system.
4. A control's strength for one error type is independent of its strength for other error types.
5. A control's strength is independent of the strength of any other controls in the system.
6. A control's strength is independent of risk of processes in the AIS.

The model made independence assumptions in addition to those used in prior research to better capture the way auditors view the ICE task. The additional assumptions were based on the field research and professional literature review. These are:

1. Errors are defined as assertion violations.[5]
2. Controls cannot inject errors into information flows.

3. A control's strength is measured by a single probability.

4. The probabilities that different errors may occur in different accounts are not combined into a system-wide reliability assessment.

By **defining of error types as assertion violations** the model limits the number of error types, ensures error types are independent of each other, and provides a direct link between errors injected within the AIS and errors in account balances. Since the professional literature presents standard definitions of the five assertions supporting an account balance, this method of defining error types limits the number of error types the model includes based on accepted practice. Since the auditor's ultimate goal is to assess the probability of assertion violations in account balances, defining error types as assertion violations eliminates the need for auditors to convert internal control error types to assertion violations. This direct link has been missing in much of the prior quantitative AIS reliability research (Knechel, 1993). These prior studies defined error types in terms of how processes can fail (e.g., posting errors), not by how failures can affect an account balance (e.g., valuation or completeness errors).

Most prior research has assumed that control procedures can generate errors as well as eliminate them. Auditors, however, usually define **controls as procedures that can only eliminate and not generate errors** (e.g., Grant Thornton 1991,1996). Therefore, the model presented here does not include an error generation parameter for controls.

The model develops a **single measure of a control's strength** 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. Therefore, the model used in this research characterizes controls with a single parameter, the probability that a control will eliminate an error that has been generated by a process from an information flow.

Finally, **combining probability of errors in different accounts** into a system-wide reliability assessment would defeat the objective of achieving tractability in a manner similar to how practitioners do. The professional literature requires auditors to make separate assessments of reliability on an account and assertion basis (AICPA, 1990).

2.3.3 Defining Processes, Controls, and Error Types

Third, the model also achieves tractability by specifying processes, controls, and error types at a level of abstraction that matches current audit practice. Processes are viewed as collections of actions that are executed together and can lead to errors in information.

2.3.4 Cost of Controls

Finally, 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 desired level of reliance at the lowest testing cost.

3. Model Presentation

This section presents the internal control testing plan selection model developed in this research. The presentation has two parts that describe the methods used to develop the model and the model's components.

3.1. MODEL DEVELOPMENT

The model was developed using a combination of reviews of the professional literature, reviews of an international audit firm's audit manuals and procedures, and interviews with field auditors. The interviews, conducted with the staff from one large international CPA firm, were performed in two stages. The first set of interviews identified the control testing plan selection task as a critical one and developed a general understanding of how that task was performed and how it related to the overall audit strategy. These interviews were with six auditors (one senior, three managers, and two partners). In the second set of interviews, two managers, who were not involved in the first set of interviews, helped develop and solve detailed cases based on one of their clients. Both of these cases were based on the deposit transaction cycle of large, regional banks. The model was refined until it produced the same testing plan that was used on the client banks.

3.2. MODEL COMPONENTS

This section presents the mathematical formulation of the model. 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. Table 2 presents a summary of the notation used in this section.

Insert Table 2 about here

3.2.1 **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 $Pr(E_{ij})$, where E is the presence of an error in an information flow. Since error generation is a dichotomous event (i.e., either an error is generated or not), $Pr(E_{ij}) = 1 - Pr(\neg E_{ij})$. The assumption that $Pr(E_{ij})$'s are independent of each other for all *i*'s and *j*'s eliminates the need to specify joint probability distributions among processes and for different error types within processes.

3.2.2 **Control's Strength**

A control's strength is represented as the conditional probability a control *k* will eliminate an error generated by a process or $Pr(N_{ijk} \mid E_{ij})$, where N is the elimination of an error from an information flow. This research assumes that controls cannot inject errors into information flows, which implies that controls will not eliminate errors that do not exist or $Pr(N_{ijk} \mid \neg E_{ij}) = 0$ and $Pr(\neg N_{ijk} \mid \neg E_{ij}) = 1$. In addition, controls are dichotomous. They will either eliminate an error or not. Therefore, $Pr(N_{ijk} \mid E_{ij}) = 1 - Pr(\neg N_{ijk} \mid E_{ij})$.

The assumptions that controls are independent of each other and their strength for one error type is independent of their strength for other error types eliminate the need to specify joint probability distributions among controls and error types. In addition, the assumptions that a process' risk is independent of a control's strength and a control's strength is independent of a process' risk eliminate the need to specify joint probability distributions among controls and processes.

3.2.3 **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 will exist in an information flow coming from process *j* where control *k* is present is $Pr(E_{ij} \mid C_k) = Pr(\neg N_{ijk} \mid E_{ij}) * Pr(E_{ij})$. 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 a control fails to eliminate them.

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 $\Pr(\neg E_{ij} \mid \mathbf{S^{ij}}) = 1 - \prod_{k \in S_{ij}} \Pr(\neg N_{ijk} \mid E_{ij}) * \Pr(E_{ij})$ where $\mathbf{S^{ij}}$ is the set of all controls that can eliminate errors of type *i* from information flowing from process *j*. This specification assumes that control strengths are independent of each other, so the probabilities that they will fail can be multiplied together and joint probabilities do not have to be specified.

3.2.4 **Multiple Process and Control Model**

An AIS contains multiple processes as well as multiple controls. The processes generate information flows that terminate in an account balance. The probability that there will be no errors in an account balance *l* is $\Pr(\neg E_{il} \mid \mathbf{S^{il}}, \mathbf{P^{il}}) = \prod_{j \in \mathbf{P}_{il}} \Pr(\neg E_{ij} \mid \mathbf{S^{ij}})$ where $\mathbf{P^{il}}$ 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*). $\mathbf{S^{il}}$ is the set of controls that can eliminate error type *i* from processes in $\mathbf{P^{il}}$. The assumptions about the independence among processes and controls allow the multiplication of probabilities that a process will inject errors into an information flow given a set of controls. This multiplication of error probabilities in the accounting processes that lead to the general ledger makes it possible to calculate the probability that an account balance will contain an error.

3.2.5 **Optimization Model**

Since professional pronouncements require error risk to be assessed by account and error type, the probability that there are no 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 $\Pr(\neg E_{il} \mid \mathbf{S}^{il}, \mathbf{P}^{il})$ for each assertion violation (i.e., error type) and account 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 $\mathbf{S}^{il}$ and $\mathbf{P}^{il}$ to justify the planned degree of reliance. The confirmation of these probabilities is accomplished by testing controls and process combinations. Auditors typically do not test all of the controls and processes in $\mathbf{S}^{il}$ and $\mathbf{P}^{il}$, but only test key controls that, if working, can provide the planned level of reliance (AICPA, 1990; Nado, et al. 1996; Grant Thornton, 1990, 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:

$$Min \sum_{k \in s} Cost_k C_k$$

Subject To:

$$C_k = 1 \text{ if } k \in \mathbf{S^t}, 0 \text{ otherwise} \quad \forall k \in \mathbf{S}$$

$$\prod_{j \in \mathbf{P}_{il}} [1 - \prod_{k \in Sil} \Pr(\neg N_{ijk} \mid E_{ij}) * \Pr(E_{ij})] \geq T_{il} \quad \forall i \in \mathbf{E^t}, \forall l \in \mathbf{A^t}$$

Where:

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

This model was implemented as a decision aid that analyzes a description of an AIS and produces an optimal testing plan. To solve this formulation, however, the contents of $\mathbf{P^{il}}$ and $\mathbf{S^{il}}$ need to be determined. The decision aid uses a depth-first search algorithm that determines the processes in $\mathbf{P^{il}}$ by calculating all dependency sets for all target accounts using a description of an AIS that includes:

1. the risks for all processes and error types;
2. the strength for each control for each process and error type; and
3. the flows of information between processes and accounts.

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 $\mathbf{S^{il}}$ using the control to error type and control to process mappings from the same AIS description.

The decision aid was implemented using a Microsoft Access™ database linked to two Microsoft Excel™ spreadsheets. The database component allows auditors to enter a description of an AIS and set of target error types; produces a variety of descriptive reports to facilitate verification of data accuracy; uses a spreadsheet to solve for an optimal testing plan[6]; and uses a second spreadsheet to allow auditors to perform sensitivity analyses.

4. Model Validation

4.1. VALIDATION METHOD

The purpose of the validation portion of this research was to provide a *prima facie* case that the model could produce testing plans comparable to experienced auditors. Because the number of free parameters in any test case is large, standard statistical hypothesis tests were not feasible. Testing plans can vary depending on the following parameters of the model:

1. a process' risk for each error type;
2. a control's strength for each error type;
3. the processes where a control can provide coverage;
4. the process and information flow structure of the case; and
5. the decision rule used to combine the above parameters and choose a testing plan.

A complete test of the model 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 model's plans. Therefore, three bases were established against which the model’s plans were compared: a professional benchmark (explained below), a consensus auditor, and an average of auditors. Since existing cases were used to validate the model and since auditors do not make probability assessments as a part of current audit practice, the formulation had to be converted to a deterministic form to process test case data. This was accomplished by setting all $Pr(E_{ij})$'s to 0.5 (the probability that a random process would generate an error) and all $Pr(\neg N_{ijk} \mid E_{ij})$'s to either 1 or 0 depending on whether the control could eliminate error type *i* injected by process *j* or not. While this validation strategy does not test each of the model's assumptions individually, it does provide support for the assertion that the model, taken as a whole, can select internal control testing plans that are similar to both experienced auditors and professional benchmarks.

The data used to validate the model was taken from Davis (1993, 1996) and not developed specifically to test this model. A detailed description of the cases, subjects, and research methods used to develop the data can be found in Davis (1993, 1996). The follow sections present a brief overview of the validation data and collection methods.

4.1.1 **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.

4.1.2 **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)[7]. 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 model. 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.

4.1.3 **Professional Benchmark Development**

Because the cases came from two sources, two types of professional benchmarks were used. One of the three cases was based on an actual client of the firm that provided Davis (1993) 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. This study's authors, both of whom have public accounting experience, independently developed testing plans for these cases by reviewing the test results in the examples and inferring the tests that were performed. Differences between the two plans were resolved by mutual agreement.

4.1.4 **Data Coding**

Not all the data needed to run the model were included in the cases. Neither the AICPA examples nor the firm's audit documentation standards included determination of the error types that a control could prevent or the error types that a process could generate. In addition, neither the AICPA examples nor the firm's audit documentation indicated the testing costs for controls.

To provide control-to-error-type mappings, the authors independently coded all of the controls in the cases with the error types they could eliminate. This coding was necessary for the model to determine which $Pr(\neg N_{ijk} \mid E_{ij})$'s to set to 0 and which to 1. Coding differences were resolved by mutual agreement. For the purposes of the validation study, a control's testing cost was set to one and all processes were assumed to be as capable as a random process of injecting all three error types given the absence of controls over the processes (i.e., all $Pr(E_{ij})$'s were set to 0.5).

4.1.5 **Analysis Method**

The model's testing plans were compared for similarity to the auditors' plans and the professional benchmarks. Two types of comparisons were made between the model and the auditors' plans. First, a **consensus audit plan** was developed for each case by including all controls selected by at least half the subjects. The model's plan was compared to this consensus plan. Second, the model's plan was compared to each subject's plan and the results were totaled.

Audit plan comparisons were evaluated using two statistics: the Random Improvement Over Chance (RIOC) statistic (Meehl and Rosen, 1955; Loeber and Dishion, 1983) and a Chi-Square statistic. RIOC differs from Chi Square because it controls for the number of controls selected by the target (i.e., base rate of auditors, consensus auditor, or benchmark) and the number controls selected by the model (i.e., selection rate) in determining the maximum number of matches that the model could produce as well as the number of matches that could be produced by chance. The main advantage of using the RIOC is that comparisons can be made between cases where there are systematic differences in base or selection rates. In addition, the RIOC statistic provides a more convenient way to express relative differences in prediction accuracy for different targets. The formula for RIOC is:

$$\text{RIOC} = \frac{\text{Observed Accuracy} - \text{Random Accuracy}}{\text{Maximum Accuracy} - \text{Random Accuracy}} * 100$$

Chi Square statistics also were calculated because they are more traditionally used in audit research and provide a basis of comparison for readers unfamiliar with RIOC. Chi Square statistics were calculated on the number of matches between the model and the targets compared to a random process. The random process was assumed to generate a correct match 50% of the time. A match occurs when both the model and the target select a control for testing or exclude a control from the testing plan.

4.2. VALIDATION RESULTS

A summary of the RIOC percentages and Chi Square statistics for the study is presented in Table 3. The results show that the model predicts the testing plans of both the experienced auditors and the professional benchmarks statistically significantly greater than chance, although some results are only moderately significant. The model is 55% better than chance in predicting the professional benchmarks, 29% better than chance in predicting the consensus auditor plan, and 11% better than chance in predicting individual auditors' testing plans. The Chi-Square statistics parallel the RIOC results. In addition, the model predicts the professional benchmark plan 43 percentage points better than the individual auditors do and 11 percentage points better than the consensus auditor. The former difference is significant at the 0.03 level; the latter is not statistically significant.

Insert Table 3 about here

The results for the individual cases are similar to the overall results, except for one case where the model predicted the auditors' plans and the consensus auditor plans less accurately than chance and the professional benchmark approximately as well as a chance. Due to the fewer number of observations, the results for the individual cases are also less significant than those for the three cases combined.

The three targets can be ordered in terms of their expected quality. The professional benchmark plans were all developed using some form of review process, either by the audit team on the job or the editorial staff of the AICPA. Therefore, the professional benchmark should be the most reliable standard of quality for the cases. The next most reliable standard should be the consensus auditor's plan, which represents agreement between several experienced auditors. The individual experienced auditors' plans would be the third most reliable since it contains individual variation. The model's predictive accuracy also follows this ordering. It does better in predicting the targets that are potentially more reliable.

4.3. QUALITATIVE ANALYSIS OF VALIDATION DATA

The RIOC and Chi-Square statistics measure the similarity between two testing plans independent of any qualitative characteristics of the two plans. To provide greater insight into the various plans included in the validation study, we analyzed the professional benchmark's and model's plans for completeness and lowest cost, and the auditors' plans for completeness, lowest cost, and consensus. A summary of these analyses is included in Table 4. Although the model's plans are guaranteed to be complete and lowest cost because of the integer program used in the model, its results are included in Table 4 for comparison.

Insert Table 4 about here

4.3.1 **Completeness**

For this analysis, a complete plan was defined as one that tested sufficient controls such that all target error types could be eliminated or "covered" by at least one control at each process (i.e., error type-process combination). Completeness was measured in Table 4, Panel A both as the percentage of auditors whose plans contained no missed error type-process combinations (i.e., complete plans) and as the percentage of all error type-process combinations in the case that were missed by a plan.

Summarizing across all cases, only 10.3% (three of 29 subjects) of the auditors developed complete plans. On average, the auditors' plans missed 42.1% of the error type-process combinations in the cases. The consensus auditor's plan averaged 34.5% missed error type-process combinations. The professional benchmark averaged 12.1% missed error type-process combinations. The benchmark missed no combinations in two of the three cases, but tested no controls that covered final posting to general ledger accounts in Case 3.

4.3.2 **Cost**

For the purposes of the validation study, all controls were assumed to cost the same to test, therefore cost was defined as the number of controls a plan tested. Table 4, Panel B presents the cost data as a percentage of controls in the case that were tested by a given plan.

In total, the model tested the fewest controls, followed closely by the individual auditors and the consensus auditor. All three tested roughly 30% of the possible controls while the benchmark tested 50%.

4.3.3 **Consensus**

Inter-auditor consensus is one measure of the quality of auditors' testing plans. Table 4, Panel C measures auditor consensus as the percentage of times any two auditors picked the same testing plan; the percentage of times any auditor picked either the consensus auditor's, benchmark's, or model's plans; and the range of error type-process combinations missed and controls tested among auditors. The total possible pairings for each case were calculated as N(N-1)/2, where N equaled the number of auditors evaluating the case.

Table 4, Panel C shows a low level of consensus among the auditors. Two auditors selected the same testing plan in only 2.2% of the possible pairings (three of 137 times). Only one of the 38 subjects (3.6%) picked the consensus auditor plan and none of the auditors picked either the benchmark's or model's plans. Auditors also exhibited wide ranges in both the number of error type-process combinations missed in their plans and the number of controls they tested.

Table 5 helps put these observations in context by showing the distribution of all possible plans and complete plans by the number of controls included (i.e., cost level) in each plan. The table also shows the distribution of the experienced auditors' plans by the number of controls tested in each case. This table highlights the fact that the model always picks the optimal plan; that the professional benchmark always picks plans that are at least as costly, and usually more costly, than the model's; and that there is considerable variation in the number of controls the experienced auditors select to test.

Insert Table 5 about here

4.3.4 Validation Summary

Table 6 presents a scatter plot of the validation data that helps to visually present the major findings of the validation study. Since the auditors tend to test fewer controls that either the model or the professional benchmark, one possible explanation is that they are not trying to achieve as high a level of reliance on controls as either the model or the benchmark. Therefore, Table 6 contains a Pareto frontier that shows the minimum number of controls that must be tested for each case to achieve various levels of completeness. The conclusions from this analysis are:

1. The model, by design, always selects plans that covers all error types at all processes. It is always on the Pareto frontier at the 100% level. Neither the auditors, the consensus auditor, nor the professional benchmark are consistently on the Pareto frontier.

2. The model tests more controls than the auditors or the consensus auditor, but both the auditors and the consensus auditor fail to cover roughly 40% of the error type by process combinations.

3. The professional benchmark tests more controls than the auditors, the consensus auditor, or the model, and fails to cover roughly 12% of the error type by process combinations.

4. There is very little consensus among auditors as to which controls should be tested in a given situation.

One reason the auditors' and professional benchmark plans may not have provided completeness or consensus is that their control to error type mappings were different than the mappings used by the model. The model has specific control to error type (control to assertions) mappings that can be specified by the user and are used by the model to search for the set of controls that will eliminate as many of the error types as possible with the least number of controls. Since the auditors had very little testing plan consensus even on the number of controls chosen, auditors likely have very little agreement on what the control to error type mappings should be. Interviews with auditors also indicated that auditors have very little agreement concerning control to error type mappings. One of the strengths of the model is that it does force the user to specify control to error type mappings, a step that is very important in determining a testing plan containing controls that will eliminate the most error types. However, professional literature gives very little guidance on what these mappings should be.

The lack of consensus among auditors could also be caused by significant variation in the level of reliance planned by each auditor. Since no measure of level of reliance nor auditors' controls to error type mappings were gathered in Davis' study, these explanations cannot be tested with the existing data.

5. *Summary and Conclusions*

5.1. CONTRIBUTIONS

This research has led to the development of a model of, and 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 previously studied. The model can provide auditors with a proposed solution to a specific key control selection task as well as a method to perform sensitivity analyses on their own testing plans. The model's outputs should help auditors ensure both the effectiveness and efficiency of their key control selections.

The model was developed using field research and literature reviews to help ensure external and ecological validity. The model used mathematical formulations to help ensure its internal consistency. The decision aid also includes algorithms to produce the data matrices required by the optimal solution algorithm. The model has been implemented in a database management system that allows auditors to set up cases easily and provides them with a variety of descriptive and diagnostic reports.

The model was partially validated by comparing its testing plans to both experienced auditors and professional benchmarks. The results indicated that the model 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 model were tested due to unavailability of data in the test cases;

4. only three error-types were used in the test cases; and

5. only three test cases were used for validation.

5.2. FUTURE RESEARCH

The validation portion of the study was designed to provide exploratory evidence that the model's recommendations compared favorably with experienced auditors and professional benchmarks. More detailed testing of the model 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 model will develop better testing plans. In addition, the model is based on a series of assumptions concerning controls and processes that should be individually tested to determine whether they are valid in describing a broad variety of cases. Finally, the probabilistic features of the model were not tested due to the nature of the validation data. Cases should be developed that include probability assessments to test whether the model also will predict auditors' testing plan selection behavior in stochastic environments.

Footnotes

1. NP stands for Nondeterministic 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.

2. Although auditors differentiate between errors that are unintentional and irregularities that are intentional, this paper does not make that distinction. In this paper, "errors" refers to any inaccuracy in accounting information regardless of whether that inaccuracy is intentional or unintentional.

3. Prior quantitative research has been based on two approaches, reliability theory and Markov processes. As Knechel (1983) points out, these two approaches usually yield consistent, and often completely reconcilable, conclusions. Therefore, we do not distinguish between these two approaches in this paper.

4. 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.

5. Assertions are implied representations about account balances that management makes when they produce financial statements. For example, when reporting an inventory balance, management implies that all items in inventory that they own have been included (completeness), no items have been included that they do not exist (existence), and the items have been properly valued (valuation). Violations of each one of these assertions are a different error class.

6. Since the model is NP complete, solution of large problems may be intractable. The sizes of the test cases were kept to a level that Excel could solve within a few minutes by limiting the cases to one transaction cycle. Since the purpose of this research was to demonstrate feasibility of the overall strategy and not to develop a tool that could be efficiently used in the field, this level of speed was adequate for our purposes. For larger, more complex cases either a more efficient solver would have to be used or an heuristic algorithm instead of the solver. Price Waterhouse took the latter approach in building the COMET system (Nado, et al., 1996).

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

Bibliography

Ahituv, N., J. Halpern, and H. Will. 1985. Audit Planning: an Algorithmic Approach. *Contemporary Accounting Research*, 2:1: 95-110.

AICPA. 1990. Consideration of the Internal Control Structure in a Financial Statement Audit, AICPA: New York, NY.

Bailey, A.D., Jr., J.H. Gerlach, R.P. MacAfee, and A.B. Whinston. 1981. An Application of Complexity Theory to the Analysis of Internal Control Systems. *Auditing: A Journal of Practice and Theory*, (Summer): 38-52.

Bodner, G. 1975. Reliability Modeling of Internal Control Systems. *The Accounting Review*, (October): 747-757.

Chen, K. and R.M. Lee. 1992. *Schematic Evaluation of Internal Accounting Control Systems*, Erasmus University Research Institute of Decision and Information Systems: Rotterdam.

Cooley, J.W. and B.J. Cooley. 1982. Internal Accounting Control Systems: A Simulation Program for Assessing Their Reliabilities. *Simulation and Games*, (June): 211-231.

Cushing, B.E. 1974. A Mathematical Approach to the Analysis and Design of Internal Control Systems. *The Accounting Review*, (January): 24-41.

Cushing, B.E., L.E. Graham, Jr., A. Palmrose, R.S. Roussey and I Solomon. 1995. Risk Orientation. in *Auditing Practice, Research and Education: A Productive Collaboration*, eds. T.B. Bell and A.M. Wright, AICPA: New York: 11-54.

Davis, J.T. 1993. Auditor Experience in Preliminary Control Risk Assessments: Neural Network Models of Auditors' Knowledge Structures, Unpublished Ph.D. Dissertation, The University of Tennessee, Knoxville.

Davis, J.T. 1996. Experience and Auditors' Selection of Relevant Information for Preliminary Control Risk Assessments. *Auditing: A Journal of Practice & Theory*, Spring. 16-37.

Elliot, R.K., 1983. Unique Audit Methods: Peat Marwick International. *Auditing: A Journal of Practice and Theory* (Spring): 1-12.

Felix, W.L., Jr. and M.S. Niles. 1988. Research in Internal Control Evaluation. *Auditing: A Journal of Practice & Theory*, (Spring): 43-60.

Gadh, V.M., R. Krishnan, and J.M. Peters. 1993. Modeling Internal Controls and Their Evaluation. *Auditing: A Journal of Practice and Theory*, (Supplement): 113-129.

Grant Thornton. 1991. *INFOCUS Instruction Manual*, Grant Thornton LLP.

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

Hamlen, S.S. 1980. A Chance-Constrained Mixed Integer Programming Model for Internal Control Design. *The Accounting Review*, (October): 578-593.

Haskins, M.E. and A.J. Nanni, Jr. 1987. Toward Attribute Models of Accounting Control Systems: Qualitative versus Quantitative Approaches. *Journal of Accounting Literature*, 6: 111-130.

Knechel, W.R. 1983. The Use of Quantitative Models in the Review and Evaluation of Internal Control: A Survey and Review. *Journal of Accounting Literature,* 205-219.

Knechel, W.R. 1985. An Analysis of Alternative Error Assumptions in Modeling the Reliability of Accounting Systems. *Journal of Accounting Research*, (Spring): 194-212.

Lea, R.B., S.J. Adams, and R.F. Boykin. 1992. Modeling of the Audit Risk Assessment Process at the Assertion Level Within an Account Balance. *Auditing: A Journal of Practice and Theory,* (Supplement): 152-179.

Loeber, R. and T. Dishion. 1983. Early Predictors of Male Delinquency: A Review. *Psychological Bulletin*, 94:2, 69-99.

Meehl, P.E. and A. Rosen. 1955. Antecedent Probability and the Efficiency of Psychometric Signs, Patterns or Cutting Scores. *Psychological Bulletin*, 52:3, 194-216.

Mock, T.J. and J.J. Willingham. 1983. An Improved Method of Documenting and Evaluating A System of Internal Accounting Controls. *Auditing: A Journal of Practice and Theory*, (Spring): 91-99.

Nado, R., M. Chams, J. Delisio, and W. Hamscher. 1996. COMET: An Application of Model-Based Reasoning to Accounting Systems. in *Proceedings of the Eight Innovative Applications of Artificial Intelligence Conference,* Menlo Park, CA: AAAI Press. 1482-1490.

Stachowski, J. 1994. Evaluation of Internal Controls: The Changing Role of the External Auditor. Massachusetts CPA Review, Spring: 12-14.

Waggoner, J. 1991. Evaluating Control Procedure Effectiveness. *Internal Auditing*, Winter: 38-43.

Waller, W.S. 1993. Auditors' Assessments of Inherent and Control Risk in Field Settings. *The Accounting Review*, 68:4, 783-802.

Walls, J.G. and E. Turban. 1992. Selecting Controls for Computer-based Information Systems Using Heuristics and Logic Programming. *Accounting, Management, and Information Technology*, 2:4: 214-261.

Table 1
Summary of Model Assumptions

Assumption	Prior Research	T
Process to error independence	Assumed independence	Assumes indepe
Process to process independence	Assumed independence	Assumes indepe
Process to control independence	Assumed independence	Assumes indepe
Control to error independence	Assumed independence	Assumes indepe
Control to control independence	Assumed independence	Assumes indepe
Control to process independence	Assumed independence	Assumes indepe
Number of error types a process can generate or Number of error types a control can eliminate	Either assumed only one or an unlimited number of independent error types	Limited to four defined by asser
Control's risk	Controls can create errors in information flows	Controls cannot information flov
Determination of control strength probability	Control strength based on two probabilities: detection and correction of an error	Control strengtl probability: elir
Combining probabilities of different errors occurring in different accounts	Combined	Not combined

Table 2
Summary of Model Notation

Element	Explanation
i	Index for error types
j	Index for processes
k	Index for controls
l	Index for accounts
E	The event that an error is injected into an information flow
N	The event that a control eliminates an error from an information flow
C_k	The presence of control *k*
¬	Not
S	Set of all controls
$\mathbf{S}^{ij}$	Set of all controls that can eliminate errors of type *i* from information flows coming from process *j*.
$\mathbf{S}^{il}$	The set of all controls that can eliminate errors or type *i* from one or more processes in $\mathbf{P}^{il}$
$\mathbf{P}^{il}$	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
$\mathbf{S}^{t}$	Set of all controls in the testing plan
$\mathbf{E}^{t}$	The set of target error types
$\mathbf{A}^{t}$	The set of target accounts
$\Pr(E_{ij})$	Probability that process *j* will generate an error of type *i*, i.e. the process' risk
$\Pr(N_{ijk} \mid E_{ij})$	The probability that control *k* will eliminate errors of type *i* generated by process *j* from an information flow, i.e. the controls strength
$\Pr(E_{ij} \mid C_k)$	Probability that process *j* will inject an error of type *i* into an information flow given control *k* is present
$\Pr(\neg E_{ij} \mid \mathbf{S}^{ij})$	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 $\mathbf{S}^{ij}$ exists.
$\Pr(\neg E_{il} \mid \mathbf{S}^{il}, \mathbf{P}^{il})$	Probability that there will be no errors of type *i* in account balance *l*.

Table 3
Summary of Validation Results

	Case 1		Case 2		Case 3	
	RIOC % **(N[1], p)**	**% right** **(X^2, p)**	**RIOC %** **(N, p)**	**% Right** **(X^2, p)**	**RIOC %** **(N, p)**	**% Right** **(X^2, p)**
Model Predicts Auditors	30.00 (35, 0.04)	65.7 (0.76, ns)	15.13 (270, 0.01)	62.3 (18.15, <0.01)	-20.31 (99, 0.06)	44.4 (1.22, ns)
Model Predicts Consensus Auditor	41.67 (7, ns)	71.4 (1.33, ns)	62.50 (18, 0.02)	77.8 (5.56, 0.02)	-37.50 (11, ns)	36.4 (0.82, ns)
Model Predicts Benchmark	100.00 (7, < 0.01)	100.0 (7.00, <0.01)	100.00 (18, 0.02)	66.7 (2.00, ns)	8.33 (11, ns)	54.5 (0.09, ns)
Auditors Predict Benchmark	30.00 (35, 0.04)	65.7 (0.76, ns)	5.26 (270, ns)	43.7 (4.28, 0.04)	16.52 (99, 0.07)	61.6 (5.34, 0.02)
Consensus Auditor predicts Benchmark	41.67 (7, ns)	71.4 (1.33, ns)	100.00 (18, 0.05)	55.6 (0.22, ns)	21.43 (11, ns)	63.6 (0.82, ns)
Difference in Auditors and Model to Benchmark	70.00 (7, 0.05)	34.3 (0.72, ns)	94.74 (18, 0.03)	23.0 (9.51,<0.01)	- 8.18 (11, ns)	- 7.1 (0.20, ns)
Difference in Consensus Auditor and Model to Benchmark	58.33 (7, ns)	28.6 (0.43, ns)	0.00 (18, ns)	11.1 (2.14, ns)	-13.10 (11, ns)	- 9.1 (0.33, ns)
Industry	Retail/Wholesale		Manufacturing		Manufacturing	
Description	Lumber Yard		Bottled Water		Women's Apparel	

[1]**N =** number of observations (number of controls or number of auditors times the number of controls)

Table 4
Panel A
Summary of Completeness Data

	Case 1	**Case 2**	**Case 3**
% Auditors with complete plans	60.0 (5)[1]	0.0 (15)	0.0 (9)
Average % missed error type-process combinations by auditors	22.0 (15)[2]	48.90 (18)	54.2 (17)
% Missed error type-process combinations by consensus auditor	20.0 (15)[2]	55.6 (18)	41.2 (17)
% Missed error type-process combinations by benchmark	0.0 (15)[2]	0.0 (18)	41.2 (17)
% Missed error type-process combinations by model	0.0 (15)[2]	0.0 (18)	0.0 (17)

[1]Number of subjects
[2]Number of process times number of target error types

Table 4
Panel B
Summary of Cost Data

	Case 1	Case 2	Case 3
Average % of controls tested by auditors	57.1 (7)[3]	22.2 (18)	31.8 (11)
% Controls tested by consensus auditor	57.1 (7)[3]	22.2 (18)	36.4 (11)
% Controls tested by benchmark	57.1 (7)[3]	66.7 (18)	36.4 (11)
% Controls tested by model	57.1 (7)[3]	18.2 (18)	27.3 (11)

[3]Number of controls

Table 4
Panel C
Summary of Consensus Data

	Case 1	Case 2	Case 3
% Auditors with the same plan	0.0 (10)[4]	3.3 (91)	0.0 (36)
% Auditors picking consensus plan	0.0 (5)[1]	0.0 (14)	11.1 (9)
% Auditors picking benchmark plan	0.0 (5)[1]	0.0 (14)	0.0 (9)
% Auditors picking model's plan	0.0 (5)[1]	0.0 (14)	0.0 (9)
Range of auditor-missed error type-process combinations	0 - 16 (19)[2]	3 - 22 (22)	1 - 16 (17)
Range of controls tested by auditors	1 - 5 (7)[3]	0 - 9 (18)	1 - 6 (11)

[1]Number of subjects
[2]Number of target error types times the number of processes
[3]Number of controls
[4]Number of possible auditor combinations - N(N-1)/2

Table 5
Panel A
Distribution of Testing Plans for Case 1

	Controls in the Plan	Possible Plans	%	Possible Complete Plans	% of Possible Plans	Number of Auditors Selecting Plans	%	Num Con Au Pl
	-	1	0.78	-	-	-	-	
	1	7	5.47	-	-	1	20.00	
	2	21	16.41	-	-	-	-	
	3	35	27.34	-	-	-	-	
M,B,*C*	4	35	27.34	2	1.56	1	20.00	
	5	21	16.41	5	3.91	3	60.00	
	6	7	5.47	4	3.13	-	-	
	7	1	0.78	1	0.78	-	-	
	Totals	128	100.00	12	9.38	5	100.00	

M = Model
C = Consensus Auditor
B = Benchmark
Bold indicates a complete plan, *italic indicates an incomplete plan*

Table 5
Panel B
Distribution of Testing Plans for Case 2

	Controls in the plan	Possible plans	%	Possible Complete Plans	% of Possible Plans	Number of Auditors Selecting Plans	%	Numbe Compl Audit Plan
	-	1	0.00	-	-	1	6.67	
	1	18	0.01	-	-	-	-	
	2	153	0.06	-	-	2	13.33	
	3	816	0.31	-	-	1	6.67	
C	4	3,060	1.17	-	-	1	6.67	
	5	8,568	3.27	-	-	1	6.67	
M	6	18,564	7.08	4	0.00	5	33.33	
	7	31,824	12.14	64	0.02	3	20.00	
	8	43,758	16.69	417	0.16	-	-	
	9	48,620	18.55	1,463	0.56	1	6.67	
	10	43,758	16.69	3,125	1.19	-	-	
	11	31,824	12.14	4,397	1.68	-	-	
B	12	18,564	7.08	4,282	1.63	-	-	
	13	8,568	3.27	2,961	1.13	-	-	
	14	3,060	1.17	1,460	0.56	-	-	
	15	816	0.31	504	0.19	-	-	
	16	153	0.06	116	0.04	-	-	
	17	18	0.01	16	0.01	-	-	
	18	1	0.00	1	0.00	-	-	
Totals		262,144	100.00	18,810	7.18	15	100.00	

Table 5
Panel C
Distribution of Testing Plans for Case 3

	Controls in the plan	Possible plans	%	Possible Complete Plans	% of Possible Plans	Number of Auditors Selecting Plans	%	Numbe Compl Audit Plans
	-	1	0.05	-	-	1	6.67	
	1	11	0.54	-	-	-	-	
	2	55	2.69	-	-	2	13.33	
	3	165	8.06	-	-	1	6.67	
C	4	330	16.11	-	-	1	6.67	
	5	462	22.56	3	0.15	1	6.67	
M	6	462	22.56	15	0.73	5	33.33	
	7	330	16.11	31	1.51	3	20.00	
	8	165	8.06	34	1.66	-	-	
	9	55	2.69	21	1.03	1	6.67	
	10	11	0.54	7	0.34	-	-	
	11	1	0.05	1	0.05	-	-	
Totals		2,048	100.00	112	5.47	15	100.00	

Table 6
Panel A
Case 1 - Coverage Cost Tradeoff

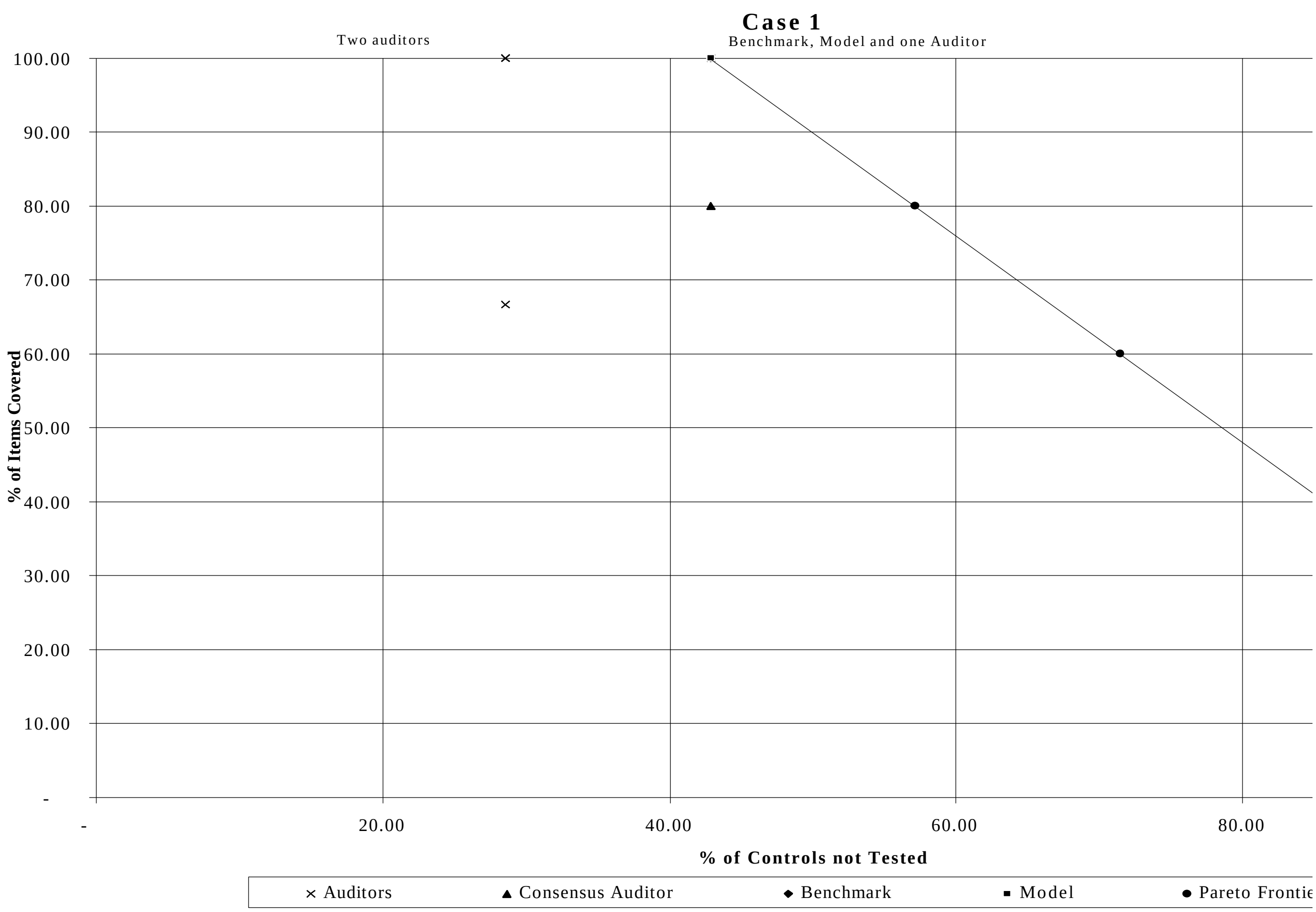

Table 6
Panel B
Case 2 - Coverage Cost Tradeoff

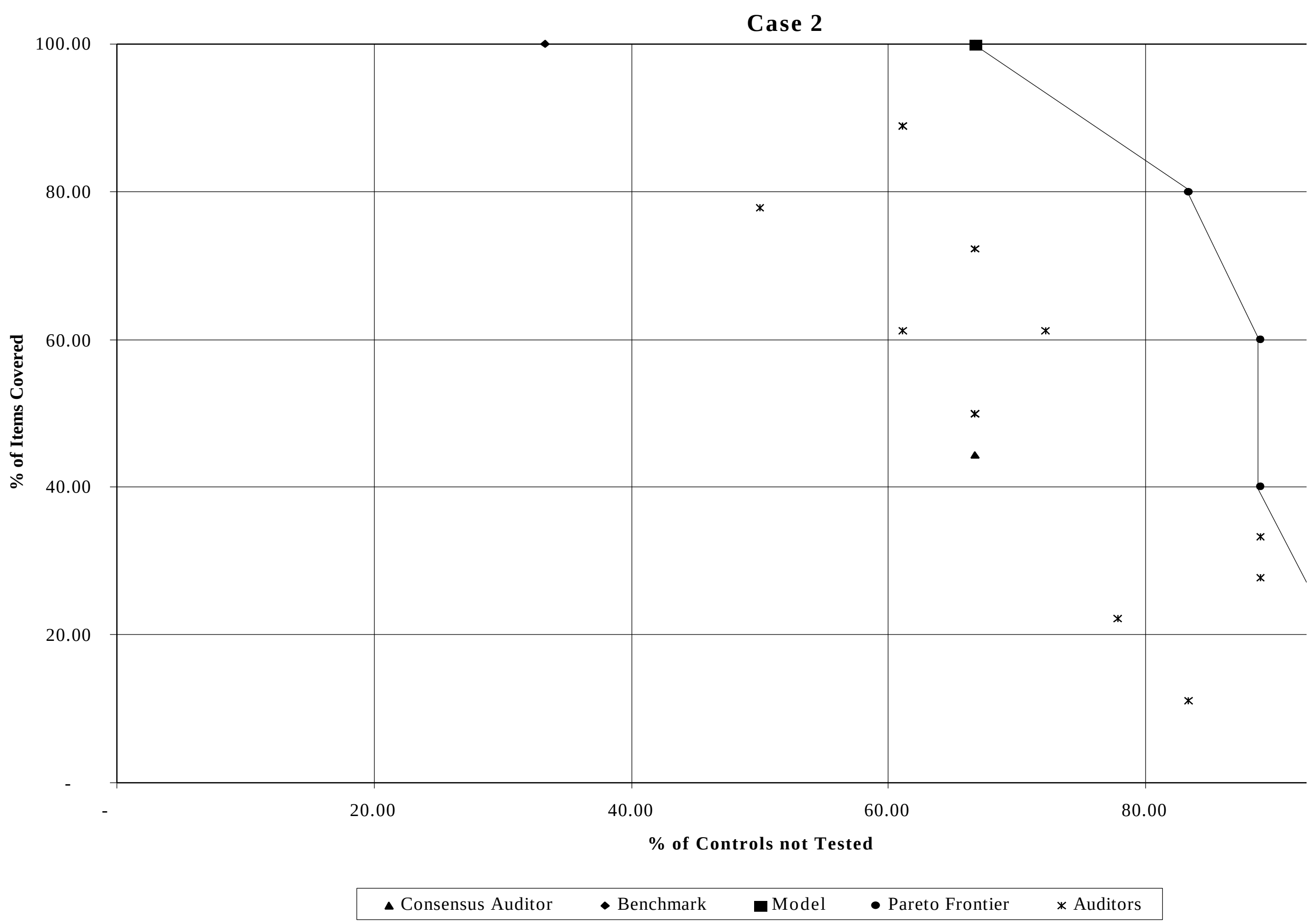

Table 6
Panel C
Case 3 - Coverage Cost Tradeoff