

# The Effects of Audit Quality and Consequence Severity on Juror Evaluations of Auditor Responsibility for Plaintiff Losses

*Kathryn Kadous*  
*University of Washington*

**ABSTRACT:** This study investigates whether providing higher quality audits increases auditors' chances of avoiding legal liability. Negligence rules hold auditors responsible for plaintiff losses only when the quality of the audit provided fails to meet standards of care. The results of my experiment suggest that the *ex post* observed consequences of audit failure can affect the standards of care to which jurors hold auditors. Specifically, participants serving in the role of jurors assessed higher standards of care for auditors when the consequences of audit failure were more severe. Furthermore, when the consequences of audit failure were more severe, participants' evaluations of the auditor did *not* depend on the quality of the audit provided—auditors who provided higher quality audits were evaluated just as negatively as those who provided lower quality audits. In contrast, when audit failure led to only moderately negative consequences, auditors who provided higher quality audits received more favorable evaluations. These results suggest that providing higher quality audits will not necessarily protect auditors from legal liability when the consequences of audit failure are severe.

**Key Words:** *Auditor liability, Litigation, Audit quality, Standard of care.*

**Data Availability:** *Contact author.*

## I. INTRODUCTION

**A**udit failure occurs when an auditor issues an unqualified opinion on financial statements that are subsequently found to have been materially misstated. The legal standard for negligence holds auditors responsible for performing an audit of a

---

This paper is taken from my dissertation, which was completed at the University of Illinois. I thank my dissertation committee, Ira Solomon (Chair), Gerald Clore, Jon Davis, and Dan Stone for their support and guidance. Thanks also to Jackie Hammersley, Jim Jiambalvo, Jane Kennedy, Lisa Koonce, Anne Magro, Linda Myers, Zoe-Vonna Palmrose, and Lisa Sedor for helpful comments, and to Deloitte & Touche for financial support.

*Submitted May 1998*  
*Accepted February 2000*

minimum quality level, and performance of such an audit should relieve the auditor of liability for subsequent audit failure (Causey and Causey 1991, 139).<sup>1</sup> The purpose of this paper is to investigate whether the level of audit quality provided (an *ex ante* decision) affects auditors' chances of being held liable for losses associated with subsequent audit failure.

The minimum audit quality level required to avoid legal liability for audit failure is vague, and thus requires interpretation. Jurors in audit negligence cases must determine what that level is—that is, they must assess standards of care—in order to evaluate whether the audit work performed was sufficient to avoid liability. Although standards of care are vague, the law contemplates that they are set *ex ante*. However, in this paper I argue that juror standards of care depend on the *ex post* observed consequences of audit failure. Specifically, I argue that observing severe adverse consequences of audit failure (such as client firm failure, employee job loss, and widespread losses suffered by parties external to the litigation) causes jurors to assess higher standards of care than does observing less severe consequences.

When standards of care are set *ex ante*, auditors can improve their chances of avoiding liability for audit failure by conducting higher quality audits. However, if standards of care depend on actual outcomes or consequences, then performing higher quality audits may not protect auditors from legal liability when the consequences of audit failure are severe. That is, because jurors set higher standards of care when consequences are severe, they are more likely to find that auditors have not met those standards, even when audits are of relatively high quality. However, when consequences are less severe, jurors are more likely to differentiate among levels of audit quality, and higher levels of audit quality should result in more positive evaluations of auditors.

I investigated these issues by conducting an experiment in which participants serving in the role of jurors evaluated an auditor involved in an audit failure and answered questions about their expectations of auditors. I constructed standard-of-care measures from the expectation items. I manipulated audit quality by varying whether the auditor performed a diagnostic audit procedure that is recommended, but not required, by generally accepted auditing standards. Consequence severity was manipulated by varying the losses suffered by parties external to the litigation. Consequence severity, though irrelevant to the merits of the legal case against the auditor, was expected to influence juror standards of care.

Consistent with expectations, when the consequences of audit failure were severe, participants assessed higher standards of care for two important aspects of an auditor's work: the type of audit tests performed and the level of professional skepticism auditors maintain while performing audits. Further, audit quality and consequence severity jointly influenced participant evaluations of the auditor. Specifically, audit quality influenced evaluations when the adverse consequences of audit failure were moderate, but participants evaluated auditors as if they did not consider audit quality when the consequences were severe.

This study contributes to the accounting literature by demonstrating that standards of care for auditors are moving targets in that they depend on the severity of the consequences of audit failure. Prior research assumes that standards of care are set *ex ante*, and that auditors can increase their chances of meeting such standards, and of avoiding liability, by increasing audit quality (e.g., Schwartz 1997, 390; Radhakrishnan 1999, 228). The current results suggest otherwise. Because standards of care depend on the *ex post* observed consequences of audit failure, auditors may be unable to predict such standards *ex ante*. As a

---

<sup>1</sup> Audit quality is defined as the level of audit assurance provided given the procedures performed and judgments made. This is consistent with DeAngelo's (1981) definition of audit quality as the probability that a financial statement misstatement is both detected and reported.

result, auditors may be unable to bridge the expectation gap by changing their *ex ante* audit decisions, as has been suggested by prior research (e.g., Kinney and Nelson 1996, 293).

In the next section, I develop hypotheses. Section III presents the experimental design and method, and Section IV describes the results of the tests of hypotheses. The final section contains a summary and conclusions.

## II. DEVELOPMENT OF HYPOTHESES

### Consequence Severity and Standards of Care

Auditors have a duty to exercise the usual judgment, care, skill, and diligence employed by other auditors (Causey and Causey 1991, 18). Thus, jurors evaluating auditors in negligence lawsuits are asked to compare the audit performed with a vague standard of care—what other audit professionals would have done in the circumstances. If the audit falls short of this standard of care, the auditor is found negligent. Researchers typically conceptualize the standard of care as vague and unknown *ex ante*, but independent of realizations such as the severity of consequences of audit failure (e.g., Dye 1995, 81). Under these conditions, auditors cannot precisely predict the standard to which they will be held, but they can improve the probability of avoiding liability by applying additional audit effort (Schwartz 1997, 390; Radhakrishnan 1999, 228) or by making more conservative reporting judgments (Kinney and Nelson 1996, 293).

Research in nonlitigation contexts raises questions about the appropriateness of this conceptualization. In particular, this research indicates that standards of care likely increase with consequence severity, a legally irrelevant case feature. For example, Walster (1966) found that participants rated the importance of regularly checking brakes more highly when told that the consequences of an auto accident involving faulty brakes were more severe. Baldwin and Kleinke (1994) found that a drunk driver was considered more reckless when accident damage was severe vs. moderate. Severe consequences apparently increase the importance of performing certain role-related tasks (e.g., checking safety equipment and remaining sober), thereby increasing standards of care.

In an audit negligence setting, I expect that severe consequences cause a similar retrospective increase in juror standards of care for auditors. Regardless of audit quality, plaintiffs' attorneys argue that if the auditor had performed more extensive tests, performed higher quality tests, or maintained a higher level of professional skepticism, the audit would have revealed material misstatements. Moreover, these attorneys identify specific audit procedures that would have diagnosed the misstatement. By contrast, defense attorneys typically argue that the additional procedures would not have been performed by other auditors and thus were not required. I expect that severe consequences will enhance the apparent importance of omitted procedures, rendering less plausible defense arguments that these procedures were not required. Thus, when consequences are severe, jurors will be more likely to view the omitted procedures as necessary and standards of care will be higher.

**H1:** Standards of care will be higher when the consequences of audit failure are severe than when the consequences of audit failure are moderate.

This prediction is important because it implies that standards of care are moving targets. Recent research suggests that the expectation gap might not be as problematic as previously thought, because auditors can balance high *ex post* expectations with *ex ante* conservatism (Kinney and Nelson 1996, 293). The above hypothesis, however, suggests the contrary. Because severe consequences highlight the importance of what was not done, the audit quality that jurors (and presumably financial statement users) expect from auditors will be

higher when severe consequences occur, making the standards difficult to attain. Thus, *ex ante* conservatism may not protect auditors when severe consequences occur.

### **Audit Quality, Consequence Severity, and Evaluations**

Accounting firms and industry observers express doubts that jurors can provide merit-based verdicts in complex civil cases (e.g., Arthur Andersen & Co. et al. 1992, 20; Lochner 1993, 93). However, limited empirical evidence supports the contention that jurors, with the help of attorneys, are able to understand the essence of complex legal and technical issues. For example, attorneys for both plaintiffs and defendants in medical malpractice cases believe that jurors can follow the evidence, which is typically complex and technical, and can understand the legal issues in such cases (Guinther 1988, 100). Judges and jurors themselves are confident that jurors make fair and appropriate decisions (Guinther 1988, 100). This suggests that jurors, with the help of attorneys, can understand enough evidence about audit issues to identify different levels of audit quality. Accordingly, I expect that when consequences are moderate and standards of care are at normal levels, audit quality will affect juror evaluations of auditors. Jurors will be able to identify auditors who do not meet the standards because expert witnesses and attorneys will point out procedures that others would have done. The auditor's work will fall short and the auditor will receive low evaluations. Auditors who perform higher quality audits in this situation will be rewarded with higher evaluations.

When consequences are severe, however, I expect standards of care to increase (H1). Prior research suggests that severe outcomes can dominate other information normally relevant to attributions of blame. For example, severe consequences appear to cause individuals to ignore causality information and respond with high attributions of blame regardless of other information provided (Fincham 1982). A similar effect could occur in audit litigation if severe consequences increase standards of care to unattainable levels. In this case, jurors would give uniformly low evaluations of auditors regardless of the level of audit quality provided. In other words, when consequences are severe, I expect that juror evaluations of auditors will be low, and they will be made as if audit quality were not considered. In sum, I expect consequence severity, a legally irrelevant case feature, to moderate the relation between audit quality and juror evaluations.

**H2:** When consequences of audit failure are moderate, evaluations of auditors will be higher for higher quality audits than for lower quality audits; however, when consequences are severe, evaluations will be low irrespective of the level of audit quality.

The above prediction implies that performing higher quality audits can protect auditors against liability for audit failures when the consequences of audit failure are moderate, but may not provide protection when the consequences of audit failure are severe.

### **Juror vs. Jury Decisions**

In this study, I examine the effects of two important variables on juror predeliberation evaluations of auditors. Although *jury* evaluations (i.e., group decisions) are of ultimate concern, studying individual predeliberation evaluations rather than jury evaluations allows for more efficient tests with little cost in terms of explaining jury evaluations. Prior research using interview data from jurors in real criminal trials finds that the jury's ultimate verdict matches the first-ballot result in at least 90 percent of the trials studied (Kalven and Zeisel

1966; Sandys and Dillehay 1995). This research also demonstrates that the relation between first-ballot votes and jury decisions is not affected by the nature of the deliberation process or by the amount of discussion that precedes the first ballot (Sandys and Dillehay 1995).

Experimental studies allow researchers to develop and fit models mapping the initial pattern of juror predeliberation verdicts into jury verdicts. These studies typically find that the predeliberation verdict pattern is the only significant explanatory variable for jury decisions, and that juries behave as though they were assigned a decision rule requiring them to decide with a two-thirds majority or otherwise hang, regardless of the assigned decision rule or jury size (see Davis [1980] for a summary). Thus, research based on both retrospective interviews of jurors in real trials and on experimental trials implies that juror initial verdict choices map into jury verdict choices in a straightforward way. Therefore, juror predeliberation evaluations are a reasonable proxy for jury decisions. Further, focus on individual juror decisions may be appropriate since jury size and decision rule vary across jurisdictions (Kassin and Wrightsman 1988).

### III. EXPERIMENTAL DESIGN AND METHOD

#### Participants

I tested the hypotheses in an experiment in which I manipulated audit quality and consequence severity between subjects in a  $2 \times 2$  full-factorial design. I recruited jury-eligible participants through a mailing to people who had recently been called to jury duty (23 participants), a mailing to addresses randomly selected from the telephone directory (55), and personal contact in an introductory economics course and in campus study areas (29).<sup>2</sup> Participants completed the study in groups of one to 20. I was present to give instructions and answer questions at all sessions. All participants were paid \$10 for their time, and one participant was randomly selected to win a \$500 prize.

#### Materials and Experimental Procedure

Participants listened to an audio tape summarizing a negligence lawsuit that a client company's creditor had filed against an auditor. The suit alleged that the company's inventory account was materially misstated. A written copy (transcript) of the case was also provided. The case included the plaintiff's complaint, the respondent's answer, attorneys' opening and closing statements, witness testimony providing background information, expert witness testimony for each party to the litigation, and the judge's instructions to the jury.

An extensive body of juror decision-making research demonstrates that case format influences verdicts, confidence in verdicts, and memory for evidence (see Pennington and Hastie [1993] for a summary). My experimental materials enhanced external validity by incorporating important structural features of trials and by maintaining the natural order of testimony. For example, attorneys summarized information important to their cases in their opening and closing statements, expert witnesses gave opposing views regarding the quality of the audit work based on the same fact pattern, and a judge's instructions were provided. An attorney who practices primarily in audit negligence cases reviewed the pilot materials for realism, and I incorporated his suggestions into the final materials.

---

<sup>2</sup> This study was part of a larger one with 330 participants. Two hundred nineteen participants were assigned to experimental conditions not reported in this paper. Data from four participants assigned to reported conditions were eliminated from the analysis. One of these four participants gave contradictory data (finding the auditor not responsible for the plaintiff's loss but making a monetary award to the plaintiff), while the other three (two attorneys and one CPA) reported professions that would prevent them from serving as jury members in an audit negligence case.

The main dependent measures are standard of care measures and evaluations of the auditor. Evaluations include ratings of the extent to which the auditor was “guilty” and “innocent” of negligence and verdicts in favor of or against the auditor.<sup>3,4</sup> Participants completed evaluations of the auditor and assessments of the auditor’s monetary liability to the plaintiff, if applicable, immediately after hearing the case. They next performed a filler task and then answered a post-test questionnaire, which included a manipulation check for consequence severity and questions about their participation in the study. Finally, participants responded to standard of care and demographic items.

### **Consequence Severity**

The consequences of audit failure were either moderate or severe. The two consequence levels arose from the same underlying events and auditor decisions. In both conditions, audit-client inventory was materially overstated due to fraud.<sup>5</sup> I manipulated consequence severity by varying the scope of the reported consequences. In the severe-consequences condition, late discovery of the overstatement resulted in large investor and creditor losses, failure of the client company, and loss of jobs for the company’s employees. In the moderate-consequences condition, the same late discovery led to large creditor losses, but the client company was acquired by another firm and continued to operate. The larger scope of the more severe loss is irrelevant to the plaintiff’s lawsuit because the losses were experienced by parties external to the suit (i.e., investors and employees). The experimental materials for the severe-consequences condition explained this to participants. The dollar amount of the plaintiff’s loss was held constant across consequence descriptions to ensure that any observed effects are attributable to consequence severity, as defined above, rather than to the magnitude of the loss.

### **Audit Quality**

I also manipulated audit quality at two levels. In the higher-audit-quality condition, the auditor consulted a specialist regarding inventory measurement. In the lower-audit-quality condition, the auditor did not consult a specialist. Consultation with a specialist is a stronger test of inventory quantities than estimation without a specialist’s help, especially for the type of inventory described in the case (which was gravel stored in piles). Generally accepted auditing standards specifically mention, but do not require, consultation for this type of inventory (AICPA 1994). There were no other differences in audit procedures between the two levels of audit quality.

Buchman (1985, 269) notes that hindsight makes it easy to identify changes to an audit that would have facilitated discovery of misstatements. This is true even for audits of relatively high quality. Accordingly, experimental materials for both levels of audit quality

---

<sup>3</sup> The instrument asked participants to rate how guilty and innocent the auditor was of negligence. Although the terms “guilty” and “innocent” would not be used in a civil case, I used these terms in the study to capture both positive and negative aspects of evaluations. Research in social psychology implies that positive (innocence) and negative (guilt) evaluations may be processed separately (Larsen and Ketelaar 1991, 133), and thus may be based on different types of evidence. For example, individuals might search for evidence of intent in assessing guilt, but be more interested in evidence of lack of causality in assessing innocence. In this study, guilt and innocence ratings were slightly super-additive when consequences were moderate and slightly sub-additive when consequences were severe, on average (see Table 3). Such noncomplementarity supports the contention that guilt and innocence ratings are based on different evidence.

<sup>4</sup> I elicited ratings because they allow more sensitive statistical tests than does the binary verdict variable.

<sup>5</sup> Inventory fraud is typical of important litigation cases against auditors. Management fraud is involved in about half of all litigation cases against auditors (Palmrose 1987, 98) and the most common type of management fraud is overvalued assets (Loebbecke et al. 1989, 11).

included expert witness testimony for the plaintiff claiming that other auditors in the defendant's position might have performed different or additional procedures. The materials mentioned performing tests at more inventory locations (since not all locations were tested) and giving the client less lead time regarding which locations would be tested. Consulting a specialist was also mentioned in the lower-audit-quality condition. In both audit-quality conditions, the expert witness for the defense argued that the defendant complied with auditing standards and performed the procedures that other auditors would have performed in the circumstances.

### Standards of Care

In accordance with recommended psychometric practice, I used multiple measures to develop and assess standard of care dependent measures (Churchill 1979). Participants were expected to develop standards for the *type* and *extent* of work that an auditor must perform before making a judgment about the fairness of presentation of the financial statements, and for the degree of professional *skepticism* that auditors must maintain during an audit. Accordingly, I developed test items to measure the three standards of care (type, extent, and skepticism), as well as the extent to which auditors should serve as insurers against investor and creditor losses.<sup>6</sup> Most items were newly constructed, but five were adapted from Anderson et al. (1993). Next, two individuals with auditing experience reviewed each item to ensure that it addressed the construct of interest. I included the remaining items in a pilot test. Because reliability measures were within acceptable ranges for the pilot test (Cronbach's Alpha = 0.60), I included all items in the final instrument.

I used confirmatory factor analysis on data from the main study to construct standard of care dependent measures. The analysis confirmed that there were four important factors, and that the factors coincided with the three standards of care and the insurer role expectancy.<sup>7</sup> I computed the three standard of care dependent measures (type, extent, and skepticism) as equally weighted sums of responses to items with loadings of at least 0.5 on the relevant factor. Table 1 lists the items comprising each dependent measure.

## IV. EMPIRICAL RESULTS

### Manipulation Check for Severity of Consequences

The post-test questionnaire asked participants to choose the category that best described the consequences in the case they had read. Of five alternatives, three described neutral consequences, and two described losses of increasing severity accompanying a financial statement misstatement. Participants in the severe-consequences condition identified more

---

<sup>6</sup> Beliefs about the extent to which auditors should serve as insurers against investor and creditor losses should not change with consequence severity. However, I measured these beliefs because they can affect the processing of audit-quality information. If jurors believe that auditors are always financially responsible for client losses, their verdicts will not be influenced by audit quality—as long as a loss occurs, such jurors will find auditors responsible for reimbursing plaintiffs. Participants with this extreme view would have an insurer role expectancy score of 40. The maximum observed score was 34, and the mean (standard deviation) of observed expectancies was 14.00 (8.20), indicating that participants did not hold this extreme view. Consistent with expectations, the dependent measure constructed for the insurer role expectancy (following the same method described for the standards of care) was not significantly affected by the experimental manipulations.

<sup>7</sup> I performed the factor analysis on data from the larger sample of 330 participants. A large sample size is required to ensure that factor results are reliable because of the large number of test items (Nunnally and Bernstein 1994, 534).

**TABLE 1**  
**Standard of Care Survey Items<sup>a</sup>**

*Panel A: Type-of-Work Items*

- \*\* T1. Auditors should always create and perform special tests designed to discover fraud, even if the company's owners and managers seem to be honest.
- \*\* T2. In the performance of an audit, it is the auditor's responsibility to actively search for instances of fraud in financial reporting, no matter how small.
- \*\*\*T3. Auditors should always perform a complete review of the client's accounting system and of the controls over the system, even if they can verify the numbers on the financial statements without doing so.
- \* T4. Auditors must examine all aspects of a company's financial health in order to properly do their jobs.
- \*\* T5. Auditors should gather all of the evidence that they can possibly get before drawing any conclusions about the financial statements.

*Panel B: Extent-of-Work Items*

- E1. Auditors can tell a lot about a company's financial records by inspecting only a few documents if those few documents are selected carefully.
- E2. Auditors cannot examine every client transaction. They must rely on samples and tests of relationships in conducting an audit.
- E3. When a client stores inventory in several different places, auditors cannot and do not need to observe inventory at every site.
- E4. Auditors must make their decisions after inspecting only a small sample of the documents that are available because this is more efficient than looking at every document.
- E5. By planning carefully and using sophisticated techniques for choosing which documents to examine, auditors are able to reduce the number of tests that they do without sacrificing quality of their work.
- \* E6. If auditors always required only the strongest type of evidence, audits would be too expensive.
- E7. Auditors cannot always use the strongest tests; they need to balance the strength of the tests with the cost and convenience of the tests.
- E8. Auditors cannot be expected to discover small misstatements in financial statements, because the level of detail of their tests has to be limited.

*Panel C: Professional-Skepticism Items*

- \*\* S1. Auditors should be completely independent of their clients, so they should not accommodate their clients' wishes in designing their tests.
- \*\* S2. Auditors should be completely objective and unaffected by their clients' wishes.
- \*\*\*S3. Auditors work for the companies whose financial statements they audit, so they have to allow their clients some latitude in what they report.
- S4. Even though auditors are hired by and paid by their clients, their responsibility is to the public; they should try to put the public's interests before the client's in designing their audits.

\*. \*\* \*\*\* indicate a higher standard of care for the severe- vs. moderate-consequences condition (one-tailed test) at the 0.10, 0.05 and 0.01 level, respectively.

<sup>a</sup> Participants indicated the extent to which they agreed with each item on a 0 (completely disagree) to 10 (completely agree) scale. Each panel contains a subset of items chosen from a larger set using confirmatory factor analysis. Cronbach's Alpha is 0.67, 0.78, and 0.60 for the type-of-work, extent-of-work, and professional-skepticism subsets, respectively.



severe outcomes as consistent with the case they read (collapsing across the neutral-consequence choices,  $X^2_2 = 59.04$ ,  $p < 0.001$ ), indicating that the consequence-severity manipulation was effective.<sup>8</sup>

### Tests of Hypothesis 1: Standards of Care

To analyze the first hypothesis, I use the dependent measures constructed using confirmatory factor analysis as described above. Hypothesis 1 predicts that standards of care are higher when the consequences of audit failure are severe. Recall that there are three dimensions to the standard of care: type of work, extent of work, and level of professional skepticism. Table 1 shows that the type-of-work standard assesses the extent to which auditors should actively search for fraud and evaluate the company's financial health in addition to performing more "normal" audit duties. The extent-of-work standard focuses on the appropriateness of testing on a sample basis. Professional skepticism refers to the importance of independence from the client and the client's demands, and duty to the public over duty to the client.

Table 2 provides means and standard deviations for each standard of care for the two levels of consequence severity. Both the type-of-work and professional-skepticism standards are higher for severe vs. moderate consequences ( $t_{104} = 3.27$ ,  $p < 0.001$  for type;  $t_{104}$

\* The neutral consequences were: (1) that the participants did not learn whether the financial statements were misstated; (2) that the financial statements were not misstated; and (3) that the financial statements were misstated, but with no financial consequences for the plaintiff. Overall, 80 percent of the participants identified the consequence description that most closely matched the case they had read. When data from the participants who did not correctly identify their condition are omitted from the analysis, inferences about the hypotheses are identical to the reported results except as otherwise noted.

**TABLE 2**  
**Standard of Care Assessments by Consequence Severity<sup>a</sup>**

	<i>Moderate Consequences<sup>b</sup></i>	<i>Severe Consequences<sup>b</sup></i>
Type of Work	32.47 (9.42) n = 52	37.84*** (7.39) n = 54
Extent of Work	33.20 (14.49) n = 52	32.72 (14.66) n = 53
Professional Skepticism	19.25 (7.39) n = 53	22.93*** (6.27) n = 53

\*\*\* indicates a higher standard of care (one-tailed test) for the severe vs. moderate consequences condition at the 0.01 level.

<sup>a</sup> Table values are the mean (standard deviation) of standard of care measures constructed as equally weighted linear combinations of participants' responses to items in each panel of Table 1. All items are scored such that higher scores reflect higher standards of care.

<sup>b</sup> In the moderate-consequences condition, the creditor suffered large financial losses when the fraud was discovered. In the severe-consequences condition, the creditor suffered the same losses, but, in addition, the company failed, employees lost their jobs, and stockholders incurred large losses. The differences in conditions are legally irrelevant because the additional losses in the severe-consequences condition were borne by parties external to the lawsuit described in the experimental materials.

= 2.76,  $p = 0.003$  for skepticism).<sup>9</sup> However, the extent-of-work standard did not differ between the two consequences conditions ( $t_{103} = -0.17$ ,  $p = 0.568$ ).<sup>10</sup> Consistent with H1, severe outcomes highlight the importance of auditor independence and searching for fraud. Participants assessed higher standards for these aspects of auditor work when consequences were severe.<sup>11</sup>

Evidence that the extent-of-work standard did not vary with consequence severity is surprising. In all experimental conditions, the auditor observed inventory at only a sample of locations, and this decision allowed the fraud to remain undetected. I expected that severe consequences would highlight the importance of exhaustive testing, causing participants to have less favorable views toward sampling; however, this did not occur.

### **Additional Analysis: Standards of Care**

Results of tests of H1 indicate that the type-of-work and professional-skepticism standards of care increased with consequence severity. This phenomenon may be attributable to severe consequences motivating participants to lay blame. In other words, omitted procedures may have become more salient to participants when the consequences of audit failure were severe *because* they wanted to assign blame and they could do so if they increased standards of care sufficiently.<sup>12</sup>

To test for this “blame-laying” mechanism, the post-test questionnaire asked participants how important it was to “figure out who was to blame for the events described in the case.” Participants in the severe-consequences condition reported higher motivation to lay blame than did participants in the moderate-consequences condition (means [standard deviations] were 7.65 [2.07] for severe consequences and 6.80 [2.02] for moderate consequences,  $t_{105} = 2.14$ ,  $p = 0.017$ ). This suggests that the increase in standards of care may be even more troublesome than is implied by the H1 result. If severe consequences motivate jurors to lay blame, standards of care may increase beyond the level of care provided, even if that level is extremely high.

### **Tests of Hypothesis 2: Evaluations**

Hypothesis 2 predicts an ordinal interaction between audit quality and consequence severity. Specifically, I expect juror evaluations to be more positive for higher quality audits than for lower quality audits when consequences are moderate. However, I expect juror evaluations of auditors to be low and insensitive to audit quality when consequences are severe. I use three types of evaluations (guilt ratings, innocence ratings, and verdicts) to

<sup>9</sup> Several participants failed to answer one or more questions. Each analysis includes the maximum number of usable responses.

<sup>10</sup> The theory predicts that severity of consequences will increase standards of care, but provides no expectation that audit quality or the interaction of audit quality and consequence severity will affect the standards. ANOVA models for the standards of care confirm that the main effect of consequence severity is the only significant effect for the type-of-work and professional-skepticism standards. No effects are significant for the extent-of-work standard.

<sup>11</sup> The increase in standards of care was not limited to specific issues raised in the case. For example, univariate tests of the effect of consequence severity on each component of the standards of care (Table 1) reveal that in light of severe consequences, participants had higher expectations regarding the extent to which auditors should search for fraud (T1, T2); however, they also expected auditors to perform a more complete review of the accounting system and internal controls (T3) and allow clients less latitude in reporting (S3). The experimental materials did not mention either internal controls or reporting issues.

<sup>12</sup> Fincham (1982) provides evidence consistent with the assertion that severe consequences cause higher motivation to lay blame. Fincham (1982) found that individuals provided with one cause of an intentional act of aggression were more likely to infer a second cause when consequences were severe, particularly if the provided cause was external. External causes, such as being provoked, do not allow unambiguous attribution of blame to the aggressor, while internal causes, such as being in a bad mood, do. Thus, when severe consequences occurred, participants inferred additional causes that allowed them to unambiguously assign blame to the aggressor.

test this hypothesis. Table 3, Panel A provides means and standard deviations for guilt and innocence ratings by consequence severity and audit quality. Planned contrasts for guilt and innocence ratings appear in Panels B and C, respectively.

The first contrast reported in Panel B of Table 3 is a comprehensive test of the predicted pattern described above. The planned contrast shows that guilt ratings were lowest when consequences were moderate and audit quality was higher, and higher in the other three

**TABLE 3**  
**Guilt and Innocence Ratings by Consequence Severity and Audit Quality**

*Panel A: Mean (Standard Deviation of) Ratings*

	<u>Moderate Consequences<sup>a</sup></u>	<u>Severe Consequences<sup>a</sup></u>
<i>Guilt<sup>b</sup></i>		
Lower Audit Quality <sup>c</sup>	5.00 (2.69) n = 26	5.26 (2.85) n = 27
Higher Audit Quality <sup>c</sup>	3.83 (3.21) n = 24	5.35 (3.15) n = 26
<i>Innocence<sup>d</sup></i>		
Lower Audit Quality <sup>c</sup>	5.59 (2.48) n = 27	4.61 (2.82) n = 27
Higher Audit Quality <sup>c</sup>	6.74 (2.62) n = 25	4.46 (2.65) n = 27

*Panel B: Planned Contrasts for Guilt Ratings*

<u>Contrast</u>	<u>t-statistic</u>	<u>df</u>	<u>p-value<sup>e</sup></u>
H2: Higher-audit-quality, moderate-consequences cell lowest	1.97	99	0.026
H2a: Lower vs. higher audit quality for moderate consequences	1.39	99	0.085
H2b: Lower vs. higher audit quality for severe consequences	-0.10	99	0.542

*Panel C: Planned Contrasts for Innocence Ratings*

<u>Contrast</u>	<u>t-statistic</u>	<u>df</u>	<u>p-value<sup>e</sup></u>
H2: Higher-audit-quality, moderate-consequences cell highest	3.06	102	0.001
H2a: Higher vs. lower audit quality for moderate consequences	1.56	102	0.061
H2b: Higher vs. lower audit quality for severe consequences	-0.20	102	0.581

<sup>a</sup> Table 2 describes the moderate- and severe-consequences conditions.

<sup>b</sup> The dependent measure is the participant's response on a 0–10 scale to "How guilty of negligence do you believe [the auditor] is?"

<sup>c</sup> In the higher-audit-quality condition auditors performed a procedure that generally accepted auditing standards specifically recommend, but do not require. In the lower-audit-quality condition, auditors did not perform this procedure.

<sup>d</sup> The dependent measure is the participant's response on a 0–10 scale to "How innocent of negligence do you believe [the auditor] is?"

<sup>e</sup> p-values are one-sided.

conditions ( $t_{99} = 1.97$ ,  $p = 0.026$ ).<sup>13</sup> Additional tests reveal that when consequences were moderate (H2a), guilt ratings were lower for higher quality audits than for lower quality audits ( $t_{99} = 1.39$ ,  $p = 0.085$ ). When consequences were severe (H2b), however, guilt ratings were not influenced by audit quality ( $t_{99} = -0.10$ ,  $p = 0.542$ ).

Similarly, results for innocence ratings support the hypothesis. Panel C of Table 3 demonstrates that innocence ratings were highest when consequences were moderate and audit quality was higher ( $t_{102} = 3.06$ ,  $p = 0.001$ ). Further, when consequences were moderate (H2a), innocence ratings were higher for higher quality audits than for lower quality audits ( $t_{102} = 1.56$ ,  $p = 0.061$ ), but innocence ratings were not influenced by audit quality when consequences were severe ( $t_{102} = -0.20$ ,  $p = 0.581$ ). Thus, participants evaluated auditors who performed higher quality audits more favorably only when the consequences of audit failure were not severe.

I also use contrasts to test my hypothesis for verdicts. Results are comparable to those for ratings. Panel A of Table 4 reports the proportion and percentage of verdicts against the auditor by experimental condition, and Panel B reports tests of H2. When consequences were moderate and audit quality was higher, 38 percent of the participants chose verdicts

<sup>13</sup> I assigned contrast weights for the guilt dependent measure as follows: -3 for the moderate-consequences, higher-audit-quality cell, and +1 for the other three cells (as in Buckless and Ravenscroft 1990, 939). Weights for the innocence dependent measure were +3 for the moderate-consequences, higher-audit-quality cell, and -1 for the other three cells. This contrast is a stringent test in that it assumes responses in the two severe-consequences cells are equal to those in the moderate-consequences, lower-audit-quality condition, which is not strictly necessary for support of the hypothesis. If I relax the assumption that the moderate-consequences, lower-audit-quality cell is equivalent to the two severe-consequences cells, then the relevant comparison is that the moderate-consequences, higher-audit-quality cell is different from the two severe-consequences cells (i.e., contrast weights for guilt (innocence) ratings are -2 (+2) for the moderate-consequences, higher-audit-quality cell, 0 for the moderate-consequences, lower-audit-quality cell, and +1 (-1) for the other cells). The data also support this less stringent contrast ( $t_{99} = 2.01$ ,  $p = 0.024$  for guilt,  $t_{102} = 3.44$ ,  $p < 0.001$  for innocence)

**TABLE 4**  
**Verdicts against Auditor by Consequence Severity and Audit Quality**

*Panel A: Proportion (Percentage) of Verdicts against Auditor*

	<u>Moderate Consequences<sup>a</sup></u>	<u>Severe Consequences<sup>a</sup></u>
Lower Audit Quality <sup>b</sup>	16/27 (59.25)	19/27 (70.37)
Higher Audit Quality <sup>b</sup>	10/26 (38.46)	15/27 (55.56)

*Panel B: Planned Contrasts for Verdicts*

<u>Contrast</u>	<u>p-value from Fisher's Exact Test</u>
H2: Higher-audit-quality, moderate-consequences cell lowest	.032
H2a: Lower vs. higher audit quality for moderate consequences	.107 <sup>d</sup>
H2b: Lower vs. higher audit quality for severe consequences	.199 <sup>d</sup>

<sup>a</sup> Table 2 describes consequence-severity conditions.

<sup>b</sup> Table 3 describes audit-quality conditions.

<sup>c</sup> p-values are one-sided.

<sup>d</sup> Omitting the data from the 21 participants who did not correctly identify their consequence condition changes the p-values to 0.043 for H2a and 0.274 for H2b.

against the auditor. This is significantly less than the 62 percent of participants who chose verdicts against the auditor in the other three conditions, aggregated (Fisher's Exact Test,  $p = 0.032$ ).<sup>14</sup> Additional tests reveal that when consequences were moderate (H2a), audit quality marginally influenced verdict choice (38 percent against the auditor for higher audit quality vs. 59 percent for lower audit quality,  $p = 0.107$ ). When consequences were severe, however, audit quality did not affect verdicts (70 percent against the auditor for higher audit quality vs. 56 percent for lower audit quality,  $p = 0.199$ ). When I omit data from participants who did not correctly identify their consequence-severity condition, the differential reliance on audit-quality information across the two levels of consequence severity is more striking. For moderate consequences, 71 percent of the participants chose verdicts against the auditor when audit quality was lower, while only 40 percent found against the auditor when audit quality was higher. This difference is significant ( $p = 0.043$ ). When consequences were severe, 71 percent chose verdicts against the auditor when audit quality was lower, compared with 58 percent when audit quality was higher. As expected, these percentages are not significantly different ( $p = 0.274$ ). Auditors were rewarded for higher-audit quality, but only when consequences were moderate.<sup>15</sup>

## V. SUMMARY AND CONCLUSIONS

In this study I investigate whether standards of care for auditors depend on the severity of the consequences of audit failure. Prior research assumes that standards of care are determined *ex ante*, and that auditors can improve their chances of meeting or exceeding such standards by performing higher quality audits. In this study, participants evaluating an auditor involved in an audit failure with severe adverse consequences set higher standards of care for the type of work that auditors should do and the level of professional skepticism that they should maintain than did participants evaluating an auditor involved in an audit failure with moderate consequences.

If standards of care depend on the *ex post* consequences of audit failure, auditors may be unable to meet the higher standards that arise when consequences are severe, even by performing higher quality audits. This is consistent with my evidence that participants evaluated auditors more positively when the auditor performed a higher quality audit only when the consequences of audit failure were moderate. When the consequences of audit failure were severe, participants evaluated auditors as if they did not consider audit quality.

Additional analysis suggests that the effect of consequence severity on standards of care likely arose because participants were more motivated to lay blame when the consequences of audit failure were severe. If jurors are highly motivated to lay blame, auditors may be unable to improve their evaluations when severe consequences occur, even by providing audits of extremely high quality. Future research could further investigate the

---

<sup>14</sup> This contrast is analogous to that reported in footnote 13 for ratings, and thus may be overly restrictive. If I relax the requirement that the moderate-consequences, lower-audit-quality cell is equivalent to the two severe-consequences cells, then the relevant comparison is that the moderate-consequences, higher-audit-quality cell is different from the two severe-consequences cells ( $34/54 = 63$  percent), and the contrast remains significant ( $p = 0.034$ ).

<sup>15</sup> The juror's task can be seen as having three components: assessing standards of care, assessing audit quality, and comparing the two. Severe consequences could have caused the effects observed in tests of H2 by increasing standards of care (as in H1), reducing perceptions of audit quality, or both. I did not measure perceived audit quality, and so cannot disentangle the two effects. However, the fact that higher standards of care were observed for more severe consequences suggests that higher standards of care are at least partly responsible for the observed effects. Regardless of the mechanism, the results demonstrate that severe consequences, though legally irrelevant, caused participants to evaluate auditors as if they did not consider audit quality.

importance of the blame-laying mechanism by determining whether the results hold in conditions with extremely high audit quality.

Future research could extend this study by incorporating additional factors expected to affect juror evaluations of auditors. Such variables might include the source of professional standards (e.g., government vs. the profession [Buckless and Peace 1993]), the reputation of the auditor, whether the misstatement is caused by auditor error or client fraud, and, if by fraud, the type of fraud (Bonner et al. 1998).

The determinants of standards of care may be another fruitful area for future research. At present, we know very little about the formation of standards of care. This study used psychometric methods to develop standard of care measures, and shows that standards of care depend on legally irrelevant information. Future work could provide a more detailed investigation of the factors that are important in determining each standard of care.

Finally, future work could investigate determinants of jury award amounts. It is likely that once a juror has determined that an auditor is financially responsible for a plaintiff's loss, the juror will use a second set of factors to determine the amount of the liability. These factors likely overlap with those used to evaluate responsibility, but the two sets may not be coincident. For example, the dollar value of the plaintiff's loss (independent of consequence severity), the auditor's ability to pay, and the availability of additional solvent defendants in the suit would likely affect award amounts, but not evaluations.

## REFERENCES

- American Institute of Certified Public Accountants (AICPA). 1994. *Using the Work of a Specialist*. SAS No. 73. New York, NY: AICPA.
- Anderson, J. C., D. J. Lowe, and P. M. J. Reckers. 1993. Evaluation of auditor decisions: Hindsight bias effects and the expectation gap. *Journal of Economic Psychology* 14 (4): 711-737.
- Arthur Andersen & Co., Coopers & Lybrand, Deloitte & Touche, Ernst & Young, KPMG Peat Marwick, and Price Waterhouse. 1992. *The Liability Crisis in the United States: Impact on the Accounting Profession (A Statement of Position)*. Reprinted in *Journal of Accountancy* (November): 19-23.
- Baldwin, M. R., and C. L. Kleinke. 1994. Effects of severity of accident, intent, and "alcoholism is a disease" excuse on judgments of a drunk driver. *Journal of Applied Social Psychology* 24 (23): 2097-2109.
- Bonner, S. E., Z-V. Palmrose, and S. M. Young. 1998. Fraud type and auditor litigation: An analysis of SEC accounting and auditing enforcement releases. *The Accounting Review* 73(4): 503-532.
- Buchman, T. A. 1985. An effect of hindsight on predicting bankruptcy with accounting information. *Accounting, Organizations and Society* 10 (3): 267-285.
- Buckless, F. A., and S. P. Ravenscroft. 1990. Contrast coding: A refinement of ANOVA in behavioral analysis. *The Accounting Review* 65 (4): 933-945.
- , and R. L. Peace. 1993. The influence of the source of professional standards on juror decision making. *The Accounting Review* 68 (1): 164-175.
- Causey, D. Y., Jr., and S. A. Causey. 1991. *Duties and Liabilities of Public Accountants*. 4th edition. Mississippi State, MS: Accountant's Press.
- Churchill, G. A., Jr. 1979. A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research* 16: 64-73.
- Davis, J. H. 1980. Group decision and procedural justice. In *Progress in Social Psychology*, Vol. 1, edited by M. Fishbein, 157-229. Hillsdale, NJ: Erlbaum.
- DeAngelo, L. 1981. Auditor size and audit quality. *Journal of Accounting and Economics* 3: 183-199.
- Dye, R. A. 1995. Incorporation and the audit market. *Journal of Accounting and Economics* 19: 75-114.

- Fincham, F. 1982. Moral judgment and the development of causal schemes. *European Journal of Social Psychology* 12: 47–61.
- Guinther, J. 1988. *The Jury in America*. New York, NY: Facts on File.
- Kalven, H., Jr., and H. Zeisel. 1966. *The American Jury*. Boston, MA: Little, Brown.
- Kassin, S. M., and L. S. Wrightsman. 1988. *The American Jury on Trial: Psychological Perspectives*. New York, NY: Hemisphere.
- Kinney, W. R., Jr., and M. W. Nelson. 1996. Outcome information and the “expectation gap”: The case of loss contingencies. *Journal of Accounting Research* 34 (2): 281–299.
- Larsen, R. J., and T. Ketelaar. 1991. Personality and susceptibility to positive and negative emotional states. *Journal of Personality and Social Psychology* 61 (1): 132–140.
- Lochner, P. R., Jr. 1993. Accountants’ legal liability: A crisis that must be addressed. *Accounting Horizons* 7 (2): 92–96.
- Loebbecke, J. K., M. M. Eining, and J. J. Willingham. 1989. Auditors’ experience with material irregularities: Frequency, nature, and detectability. *Auditing: A Journal of Practice & Theory* 9 (1): 1–28.
- Nunnally, J. C., and I. H. Bernstein. 1994. *Psychometric Theory*. 3rd edition. New York, NY: McGraw-Hill.
- Palmrose, Z-V. 1987. Litigation and independent auditors: The role of business failures and management fraud. *Auditing: A Journal of Practice & Theory* 6 (2): 90–103.
- Pennington, N., and R. Hastie. 1993. The story model for juror decision making. In *Inside the Juror: The Psychology of Juror Decision Making*, edited by R. Hastie. Cambridge, U.K.: Cambridge University Press.
- Radhakrishnan, S. 1999. Investors’ recovery friction and auditor liability rules. *The Accounting Review* 74 (2): 225–240.
- Sandys, M., and R. C. Dillehay. 1995. First-ballot votes, pre-deliberation dispositions, and final verdicts in jury trials. *Law and Human Behavior* 19 (2): 175–195.
- Schwartz, R. 1997. Legal regimes, audit quality and investment. *The Accounting Review* 72 (3): 385–406.
- Walster, E. 1966. Assignment of responsibility for an accident. *Journal of Personality and Social Psychology* 3 (1): 73–79.