Audit Sampling Methods And Juror Negligence Awards: An Expectation Gap?

David L. Gilbertson, (E-mail: david.gilbertson@wwu.edu), Western Washington University
Terri L. Herron, (E-mail: terri.herron@business.umt.edu), University of Montana

Abstract

Generally accepted auditing standards permit auditors to apply both statistical and nonstatistical sampling techniques in obtaining sufficient, competent evidential matter. However, several recent studies have shown that statistical sampling procedures have nearly disappeared from practice. Despite this trend and the heightened anxiety about professional liability, no studies to date have directly tested the potential implications of sampling method on jurors verdicts, damage awards, or sample size expectations. In this study we investigate the effect of sampling method on jurors' judgments in auditor negligence trials. Overall, in a case where auditors were alleged to have used an insufficient sample size, the sampling method did not affect the likelihood of a "guilty of negligence" verdict. However, as predicted, damage awards were significantly higher when nonstatistical sampling was used compared with statistical sampling. Exploratory analysis revealed that subjects voting "guilty of negligence" would require the auditors to examine over 17% of the population (compared with the 1% examined) in order to change their verdict to "not guilty of negligence." These findings have troubling implications for auditors employing sampling techniques.

1. Introduction

*G*enerally accepted auditing standards (GAAS) permit auditors to apply both statistical and nonstatistical, or judgmental, sampling techniques in obtaining sufficient, competent evidential matter (AICPA 2000, AU§350.04). Research has generally shows that nonstatistical sampling is problematic (e.g., Burgstahler et al 2000; Hall et al. 2000; Nelson 1995; Ponomon and Wendell 1995; Kachelmeier and Messier 1990), and some auditors believe that the ability to statistically measure the risk makes statistical sampling more defensible (Colbert 1991). Ponomon and Wendell (1995) published one of the first auditing studies comparing the performance of judgmental and random sampling methods, finding that statistical methods of determining bounds were superior to auditors' judgments of bounds. They go on to state that "firm management needs to be cautious when employing judgmental procedures for selecting and making subjective inferences about the underlying population under examination" (Ponomon and Wendell 1995, 33). While their study and several others highlight the potential danger in using judgmental sampling in practice, the superiority of statistical audit sampling has yet to be tested in the courts.

Despite the apparent superiority of statistical over nonstatistical sampling, studies consistently show that statistical sampling has nearly disappeared in practice. Sullivan (1992) of the Public Oversight Board reported the he could not recall a single use of statistical sampling in peer reviews of 742 engagements. In their study of 64 audits from three large accounting firms, Elder and Allen (1998) found that auditors often used statistical sample *selection* methods, but none used statistics to *evaluate* sample results. More recently, a survey of audit practitioners from CPA firms, public companies and government offices showed that 85 percent of recent sampling applications were nonstatistical (Hall, et al. 2000).

_____

Readers with comments or questions are encouraged to contact the authors via email.

In recent decades, the volume of litigation alleging auditor failure has swollen to such proportions that national attention has been focused on this issue. In 1992, the AICPA reported that more than 3,000 suits were pending against U.S. accounting firms, seeking damages in excess of $13 billion (AICPA 1992). Palmrose (1991) identified 800 audit failure allegations against 16 of the largest U.S. audit firms from 1960 to 1990. Although most of these cases were resolved without a trial, the preponderance of those going to trial were tried before a jury. Thus, exploration of jury decision-making is highly relevant to our understanding of auditor litigation.

An auditor's failure to comply with GAAS is often conclusive evidence of negligence, though determining if such a failure occurred is highly subjective. In most cases, "any significant error or mistake in judgment will create at least a presumption of negligence that the professional will have to rebut" (Arens and Loebbecke 2000, 116). However, the manner by which two acceptable sampling approaches may affect juror negligence judgments has not been empirically tested.

Given the potential pitfalls of nonstatistical sampling, the widespread use of nonstatistical sampling, and the proliferation of auditor negligence trials, the question of juror judgments involving sampling evidence is important. The purpose of this study is to investigate the effect of sampling method on jurors' judgments in auditor negligence trials. Jurors empanelled in civil trials and junior business students independently determined verdicts and damages in cases involving either statistical or nonstatistical audit sampling methods that were alleged to be insufficient. Findings indicate that the sampling method did not significantly affect the likelihood of an unfavorable verdict, though damages awarded were significantly higher when nonstatistical sampling was used. Further, exploratory analysis of jurors' sample size expectations reveals that such expectations are unreasonably high.

The next section discusses the use of statistical evidence in the courts, followed by sections on hypotheses, method, results, and discussion of results, respectively. The final section includes limitations of the study and concluding remarks.

2. Statistical Evidence In Courts

2.1. Case Law

Statistical evidence and probabilistic calculations are commonly used in U.S. court rooms to create evidence, establish evidence reliability,, or measure a financial impact of an act (Smith et al. 1996; Fienberg 1996). Typically such evidence is introduced at trial to implicate or exonerate a defendant in a criminal case, though statistical evidence is increasingly being introduced in civil cases such as discrimination, antitrust, and environmental litigation (Fienberg 1989). There is a distinct difference between a statistical method being related to *evidence introduced at trial* (see examples above) and a statistical method being used in an event that is the *subject of the trial* (the focus of this study). The following discussion of statistical evidence introduced at civil trials is provided to give the reader a sense of the courts' acceptance of statistical methods and jurors' interpretation of such methods. For a more complete review and discussion of cases, see Fienberg (1989) and Degroot et al. (1986).

In an often-cited case involving statistical estimation to determine damages, a Sears Roebuck store filed for a $27,000 tax refund for overpaid taxes caused by an error in determining where the taxable boundaries lay in the town (*Sears, Roebuck & Co. v City of Inglewood*, LA Sup. Ct. 1955). A statistician randomly sampled 33 of the 826 sales days resulting in a point estimate for the refund of $28,250, with a 95 percent confidence interval for the amount from approximately $24,000 to $32,500. The court rejected these sampling results and forced the store to conduct a complete audit of the 826 days, resulting in a figure of $26,750.22 to which Sears was entitled. Clearly the courts rejected a statistical estimation that was quite accurate.

In *Putnam Resources v Lloyds of London,*[1] Sammartino, Inc. was a third party custodian that held Putnam's gold. Sammartino management was convicted of theft and the company went bankrupt. Putnam sought payment from its insurer, Lloyds of London, who refused to pay based on grounds that the policy was not in effect

---

[1]   Unpublished decision, U.S.D.C. RI 1991. Olinsky et al. (1996) provide a more detailed discussion of this case.

when the theft occurred. The key issue in the case was determining *when* the gold was stolen. Lloyd's forensic accountants took a sample of invoice records to accomplish this task. However, they did so by haphazard selection. Statistical experts contested the conclusions, suspecting that a bias may have existed in such a method. Lloyds ultimately prevailed. Though it is impossible to determine why the jury found for Lloyds, the bulk of the their case rested on the statistical sampling results of the forensic audit.

With the advancement in medical and technology fields, courts began to exhibit concern in the early 1980s that statistical evidence based on a *single* test may be over-implicating or misunderstood by the jury. For example, one ruling expressed "that statistics on the frequency with which certain blood type combinations occur in a population will be understood by the jury to be a quantification of the likelihood that the defendant, who shares that combination of blood characteristics, is guilty" (*State of Minnesota v Kim* 1987, 548). Today, statistical evidence is "still occasionally viewed with suspicion because of skepticism about whether it really works" (Fienberg 1989, 153). This skepticism highlights the importance of auditors' reliance on sampling procedures to test financial statement assertions if such procedures were to be challenged in court.

2.2. Empirical Research

Much of the research on statistics' use in court has focused on the behaviors jurors exhibit when presented with various forms of statistical evidence. Decision making research, in general, indicates that people often misunderstand or misapply probabilistic information (Tversky and Kahneman 1971; Kahneman and Tversky 1973; Nisbett and Ross 1980). Tversky and Kahneman (1971) found that people will make stronger inferences from a small sample than is warranted. This is termed "the law of small numbers." Nisbett and Ross (1980) found that people overvalue single cases of anecdotal evidence compared to statistical evidence. In a groundbreaking study, Wells (1992) found that jurors are reluctant to find for the plaintiff based solely on probabilistic evidence. Specifically, Wells found a difference "between saying that there is an 80 percent chance that something is true and saying something is true based on evidence that is 80 percent reliable," (Wells 1992, 746), finding jurors more likely to convict in the latter case. The Wells effect has been found to be a robust phenomenon (Niedermeier et al. 1999) and illustrates the difficulty jurors can have in interpreting statistical evidence.

Note that the above empirical studies do not compare statistical to nonstatistical evidence presented *at trials*, and the empirical studies related to evidence at trials did not compare *statistical and nonstatistical* evidence. With this lack of empirical evidence on the subject, we turn to a theory regarding jurors' decision-making processes, the story model, in developing the following hypotheses.

3. Development Of Hypotheses And Other Questions

There is great variation in jurors' use of probabilistic evidence (Smith et al. 1996), and audit sampling results are probabilistic in nature. The following discussion develops the hypotheses regarding how jurors will react to judgmental versus statistical sampling procedures that failed to uncover a fraud.

3.1. The Story Model

One perspective of juror decision-making is offered by Nancy Pennington and Reid Hastie (Pennington and Hastie 1993; Hastie1999). Their theory is called the "story model" because it posits that story construction is the central cognitive process in jury decision-making. The notable claim of this theory is that the story constructed by the juror *determines* the juror's decision. In this view, jurors reach their decision in three stages of processing: (1) evidence evaluation through story construction, (2) representation of the alternative verdicts as categories with a list of attributes, and (3) matching the story with the verdict attributes.

During the first stage, jurors construct a narrative by combining a) evidence presented at trial, b) their own knowledge about similar events, and c) general world knowledge about what makes a plausible story (e.g., knowledge about motives). While jurors may construct more than one story, only one will usually be accepted as the "best." The principles that determine whether a juror will accept a given story are called *certainty* principles.

Certainty consists of *coverage* and *coherence* (Pennington and Hastie 1993). Coverage refers to the extent to which the story accounts for the evidence. A story which leaves evidence unaccounted for will inspire less juror confidence, and thus it is less likely to be accepted. To be considered coherent, a story must be *plausible*, i.e., it must "make sense" to a juror according to his knowledge about what typically happens in the world. The story must also be *consistent:* it must contain no internal contradictions. Finally, the story must be *complete*, containing all the essential elements of a story: initiating events, psychological states, actions, and consequences. In civil litigation, juries pay particular attention to the latter story element (completeness), looking for evidence that the defendant made an explicit choice to deviate from ordinary care (Hastie 1999).

The second processing stage consists of the jurors' representation of the alternative verdicts. Most of the knowledge for this stage is received during the judge's instructions on the law. For jurors with no prior legal knowledge, learning these categories can be difficult. The final stage in jury decision-making consists of matching the "best" story with each of the verdict alternatives. During this stage, jurors will again focus on the key actions of the defendant: those which are alleged to have caused the plaintiffs' loss. This stage may also involve application of the judge's procedural instructions, if there are any.

3.2. Hypotheses

In this study, jurors are to determine whether an auditor is guilty of negligence by reason of the sampling procedures not uncovering fraudulent sales invoices. The plaintiff attorneys allege the sample size was too small, thus precluding the auditors from discovering the fraud. In evaluating an audit failure when the auditors used *statistical* sampling, jurors are likely to encounter difficulty at several points in story construction. First, jurors typically do not comprehend statistical or probabilistic information. Thus, the second knowledge component (knowledge about similar events) will be unavailable to them, and they will find it difficult to construct a plausible story. Second, jurors may find it more problematic to assign a motive to an alleged statistical miscalculation in statistical sampling than to the auditors' judgment error in nonstatistical sampling. Sample size is clearly an auditor *choice* in judgmental sampling. But it is difficult to characterize the auditors as having *chosen* to make an error in the statistical sampling scenario. Without the element of psychological state, the story construction will be incomplete. Thus, the story model would predict that jurors will not construct as plausible a "story" when plaintiffs allege negligence regarding statistical sampling procedures as compared to allegations regarding judgmental sampling. This leads to the following hypothesis:

H1:   In a case alleging insufficiency of audit sampling procedures, the likelihood of a verdict "guilty of negligence" will be higher when nonstatistical methods were used than when statistical methods were used.

While the question of guilt is certainly a central one, a further question is one of the "degree" of guilt. This degree of guilt would be reflected in the damages awarded by jurors. Jurors often have difficulty separating the liability assessment from the damage assessment, even when given specific instructions to consider damages based solely on the plaintiff's injury. This "fusion" of the liability and damage assessments is commonly attributed to jurors discounting the award because of uncertainty about the defendant's liability (Wissler et al. 2001). Research has found that jurors who were doubtful about the defendant's liability compromised with other jurors on liability in exchange for a reduced award (Hans and Lofquist 1992), and compensatory awards were lower when the defendant's fault was ambiguous than when the defendant's fault was clear (MacCoun 1996).

In this study, the story model would predict that evidence related to statistical procedures would be less likely to produce a plausible story when compared to evidence related to nonstatistical procedures, introducing more doubt or ambiguity about the defendant's liability when statistical sampling was used. The increased doubt or ambiguity is expected to result in lower damages. This leads to the following hypothesis.

H2:   In a case alleging insufficiency of audit sampling procedures, damages awarded will be higher when nonstatistical methods were used than when statistical methods were used.

3.3. Other Questions

Jurors are likely to have little prior experience with audit methodologies. The "expectation gap" between what an audit actually entails and what financial statement users expect an audit to entail could certainly extend to sampling procedures. Non-auditors' expectations regarding sample sizes have not been studied, nor have any influences the sampling method may exhibit on those expectations been studied. To explore these questions, this study asks jurors who arrive at guilty verdicts to estimate how many items in the population the auditor should have tested. Specifically, they were asked what sample size would have been adequate for them to change their verdict. This leads to the following exploratory questions:

Q1:   What proportion of the population would be considered to be adequate enough to change a guilty verdict to a not guilty verdict?

Q2:   Does the sampling method affect the jurors' notion of an adequate sample size?

4. Method

4.1. Participants

Eligible jurors for the study were drawn from two sources, an actual juror pool and a group of junior-level business students.   The juror sample was obtained through the cooperation of the jury clerk and the presiding judge of the state trial division in a medium-sized western city. The jurors had been summoned and empanelled to serve on civil cases during a period of 15 months. Frequently, civil cases are settled "on the eve" of trial. The jurors participating in the study had already arrived for their trials when they learned their trials had been settled. They were asked to complete the experimental instruments while their checks for one day's jury pay were being prepared. The cases were numbered, with the two scenarios randomly distributed, before being given to the jury clerk, who read the instructions to the jurors and collected the completed instruments. Instruments were distributed to 88 jurors, with 81 usable instruments being returned.

The student sample was obtained by requesting volunteers from two junior-level management information systems classes, 41 of whom returned usable instruments. The students were given class time to complete the complete the instrument, were read instructions identical to those read to the jurors, and completed the instruments in the presence of one of the researchers. Junior-level students were judged to be suitable participants because they are of legal age to serve on a jury. This is consistent with other jury research (e.g., Kadous 2000, 2001).

A questionnaire that accompanied each case solicited demographic information including age, years of education, gender, income bracket, marital status, and general viewpoint (conservative, moderate, liberal). Participants also indicated if they had served on previous civil juries or if they had ever been sued or had sued someone. Demographic information is summarized in Table 1. The demographics for the juror pool are comparable to those obtained by Buckless and Peace (1993). While the students differed from the jurors in several predictable ways (e.g., age, income, marital status), analysis of their voting patterns and damage assessments for each case scenario revealed no significant differences between the students and empanelled jurors. For the analyses that follow, the empanelled jurors and student participants were combined. This is also consistent with current research (e.g., Kadous 2000, 2001). All of the demographic variables were regressed on the dependent variables (verdicts and damages). No demographic variables were significantly related to the verdict (guilty or not guilty). Age was mildly significant ($p<.08$) with regard to the damages variable, where older participants were more harsh in their damage assessments. Thus, the statistical tests of damages will control for the age of the participants.

Table 1: Participant Demographics[a]

	Student (n=41)	Juror (n=81)	Overall (n=122)
Age (mean)[b]	24	41	35
Years of Education (mean)	15.1	14.6	14.7
Gender: M	51 %	34%	40 %
F	49 %	66 %	60 %
Family Income[b]:			
<$50 K	71 %	49 %	57 %
$50 - $100 K	7 %	42 %	30 %
> $100 K	0 %	5 %	3 %
Married[b]	51 %	78 %	69%
Previous Jury Service	7 %	13 %	11%
Previous Involvement In Litigation	5 %	10 %	8 %
Views:			
Conservative	35 %	30 %	32 %
Moderate	52 %	62 %	59 %
Liberal	13 %	8 %	9 %

[a]   Percentages may not sum to 100% because not all participants answered all demographic questions.
[b]   Difference between students and jurors is significant at p<.01 level. Further analysis confirmed that though the students and jurors do demographically differ, student participants did not respond to the dependent variables in a significantly different way than juror participants (p=.83 for verdict dependent variable; p=.58 for damages dependent variable). Thus, statistical tests of hypotheses were performed on the combined participant group.

4.2. Case Instruments

Each participant received a packet of case materials, including general information, jury instructions, a description of the litigation and testimony of expert witnesses, a verdict sheet, and a demographic questionnaire. The general information explained the purpose of an audit, stating that "auditors usually use sampling" and "sampling is a generally accepted audit practice." Participants were also told that expert witnesses have "considerable knowledge and experience in a field." They were told they had been selected as a juror in a civil case involving creditors (plaintiffs) and an accounting firm (defendants), and they were instructed to read the facts of the case and decide whether or not the accounting firm was negligent, and if so, provide a dollar amount of damages the firm should pay. Participants were provided a definition of negligence and instructed to base their decision on the "majority of evidence."

All versions of the case began with a description of the litigation, summarized as follows. The plaintiffs were a group of creditors of Intermountain Restaurant Supply Co., a wholesale restaurant supply company that had recently filed bankruptcy following an employee's disclosure that $84 million (10%) of the company's sales were fictitious. The defendant auditors had considerable experience in the industry, and they had audited Intermountain for four years. Intermountain had nearly 12,000 sales transactions, of which 800 were fictitious. In conducting the audit, 100 sales invoices were examined[2] and no discrepancies were found. The auditors thus concluded they could rely on the sample results. The creditors were suing the auditors for negligence, seeking $8 million in damages. The creditors claim the sample was too small, thus the auditors were negligent.

The case scenarios were reviewed by an attorney specializing in civil litigation, the presiding judge of the

---

[2]   This sample size is sufficient to support a high degree of reliance on the client's internal controls, at the 95% confidence level. (see Arens and Loebbecke 2000, 448 – 457).

trial division, and several former auditors, including one individual who does a substantial amount of expert testimony in auditor negligence cases. The entire instrument was then pre-tested using a group of twenty graduate students.

4.3. Experimental Conditions

The study employed a between subject, two-level single factorial design. Jurors were randomly assigned to one of the two experimental conditions that differed in the type of sampling method used (statistical or nonstatistical). Both versions of the case contained a brief statement that the "plaintiffs claim that this sample is clearly too small, since it did not detect any of the fraudulent sales." Such a statement was necessary to establish the basic foundation for the negligence allegation. The cases included one of the following paragraphs to introduce the sampling approach taken by the auditor and the plaintiffs' allegations:

Statistical Condition: An expert witness for the auditors testified that sampling is a widely accepted audit technique, and that the audit firm had used statistical methods to determine the sample size. Based on their statistical analysis, the auditors were 95% confident that there were no serious errors in the sales invoices. Put another way, the auditors would have detected a fraud such as this in 19 out of 20 cases. However, an expert witness for the plaintiffs testified that the auditors had made an error in their statistical calculations. According to the expert, if the appropriate statistical method had been chosen, a larger sample size would have been used.

Nonstatistical Condition: An expert witness for the auditors testified that sampling is a widely accepted audit technique, that determining sample size is a matter of professional judgment, and that he agreed with their decision to sample 100 items. An expert witness for the plaintiff testified that, while sampling is a widely accepted audit technique, and determining sample size is a matter of professional judgment, he believed the auditors' decision to sample only 100 items was clearly a mistake. In his judgment the sample should have been larger.

After reading the case materials, jurors were asked if they believe the defendant auditors are "guilty or not guilty of negligence." The participants who selected "guilty" were then asked to indicate the damages the audit firm should pay (from zero to $8 million). This provided two dependent variable measures corresponding to H1 and H2, respectively: a dichotomous measure of guilt and a continuous measure of damages. Further, if a guilty verdict was selected, the participant was asked to indicate "how many invoices would need to be sampled to change your verdict to 'not guilty'."

5. Results

5.1. Juror vs. Jury Decisions

In testing the hypotheses, individual responses were treated as surrogates for jury decisions. Prior research has indicated that the jury's ultimate verdict corresponds to the first-ballot results in at least 90 percent of the trials studied (Kalven and Keisel 1966; Sandys and Dillehay 1995). Thus, though jurors act in concert to produce a *jury* evaluation (i.e., group decision), using the individual participant assessments is an appropriate and more efficient approach in analyzing verdicts (see Kadous 2000, 2001). Likewise, civil damage awards are the result of a group process with each juror bringing his or her own damage estimates into the deliberation process. Those that consider the defendant as "not guilty" and thus not responsible for the act will be deliberating the damage amounts alongside those who view the defendant as "guilty" and thus financially responsible. . Therefore, combining and averaging all participants' damage awards is appropriate. The jury damage award can be viewed as an expected loss value of the defendant audit firm

5.2. Hypothesis Tests

Overall, 76 participants (62 %) recommended guilty verdicts. H1 is tested using a chi-square test for equal proportions of guilty verdicts, the results of which are presented in Table 2. The auditors were found "guilty" of negligence *more often* when judgmental sampling was used. However, the difference in treatment proportions compared to the population proportions was not statistically significant (p<.37). Thus, hypothesis one was not supported.

Recall that in each case instrument, if the subjects determined the auditors to be negligent, they were asked to also determine the extent of damages (ranging from zero to $8 million) to be assessed against the auditors. This provides a measure of "how negligent" the participants perceived the auditors to be. H2 was tested using a t-test for differences in mean damages (see Table 3, Panel A) and using a regression with a control variable for age (see Table 3, Panel B). Mean damages assessed against auditors are $2,043,933 when statistical sampling is used, and $2,929,032 when nonstatistical sampling is used. This difference is significant (p<.05, one-tailed) and in the predicted direction. Hypothesis two is supported.

5.3. Exploratory Analysis

Participants arriving at "guilty" verdicts were asked, "how many invoices would need to be sampled to change your verdict to 'not guilty'?" Five of the 76 participants with "guilty" verdicts did not answer this question, resulting in a sample of 71 for this portion of the analysis. Recall that the auditors sampled 100 of the 12,000 sales invoices in their testing (0.8%). This would be an unusually large sample in practice (Elder and Allen 1998). Our first exploratory question addressed what the participants would consider to be an adequate sample size. The mean number of invoices that would need to be sampled to change the verdict was 2,070 (17.25% of the population), with the range of the responses being quite wide (see Table 4, Panel A). Fifty-one percent would require 1000 or fewer invoices to be tested to change their verdict, 27% would require 1,200 to be examined, and 17% would require more than 1,200 but less than 12,000 to be examined. Four participants (6%) would require the *entire population* to be examined in order to change their verdicts. There was nothing distinct about these four participants, either demographically or treatment-wise. Interestingly, the most frequent response was 1,200 invoices, a sampling fraction of 10%. It appears that these participants were using a heuristic measure in determining what an adequate sample size would be, yet 10% is a very high sampling fraction rarely seen in practice.

Our second exploratory question addressed whether the sample size required to change the verdict was different depending on the type of sampling method used. First, the demographic variables were examined for possible influences. Second, the number of invoices was regressed on sampling method, and several demographic variables. Results are presented in Table 4, Panel B. The demographic variables that affected participants' response to the question were whether they had served on a jury (resulting in lower sample size expectations, p<.07), whether they had been directly involved in a lawsuit (resulting in higher sample size expectations, p<.07), and age (older jurors expecting larger sample sizes, p<.10). The sampling method did not significantly affect jurors' sample size expectations (p<.51).

Table 2: Guilty and Not Guilty Verdicts

	N	Not Guilty[a]	Guilty[a]
Statistical Sampling Used	60	25 (41.7%)	35 (58.3%)
Judgmental Sampling Used	62	21 (33.9%)	41 (66.1%)
	122	46 (37.7%)	76 (62.3%)

$\Pi^2 = 0.79$ $p < .37$

[a]   Response to question, "As a juror in this case, do you believe the defendant auditors are guilty or not guilty of negligence?"

Table 3: Damage Awards
Panel A: Means

	N	Mean Damages[a]	Std. Deviation
Statistical Sampling Used	60	$2,043,933	2,955.200
Judgmental Sampling Used	62	$2,929,032	3,088,970
	122		

t = 1.62, p < .05, df = 120

Panel B: H2 Regression Results[c]

Parameter	Estimate	Standard Error	t	p-value
Intercept	1,330,320	806,787	1.65	0.102
Age	44,674	19,699	2.27	0.025
Sampling Method (1=statistical)	-929,568	545,536	-1.70	0.046[b]

Dependent Variable = Damages

Adj. R^2 = .054, Model F-value = 4.29; p<.02, df = 115

[a]   Response to question, "If you answered "guilty" to question one: please indicate the amount of damages (between zero and $8 million) you believe the audit firm should pay." Subjects responding "not guilty" were considered to have assessed zero damage values.

[b]   Test of H2 is one-tailed; results are in the predicted direction.

[c]   Six participants did not provide their age, thus this analysis has n=116.

Table 4: Sample Size Sufficiency
Panel A: Descriptive Statistics -- Number of Invoices to Change Verdict [a]

Descriptive Measure	Number of Invoices
Mean	2,070
Standard Deviation	3,006
Maximum	12,000
Minimum	120
Median	1000

Response Ranges	Number (%) Responding [b]
120 – 1000 invoices	36 (50.7%)
1,200 invoices	19 (26.8%)
2,000 – 9000 invoices	12 (16.9%)
12,000 invoices	4 (5.6%)
TOTAL	71 (100%)

Panel B: Regression Results

Parameter	Estimate	Standard Error	t	p-value
Intercept	322	972	0.33	0.742
Jury (1=yes)[c]	-2,020	1,083	-1.86	0.067
Suit (1=yes)[d]	2,455	1,324	1.85	0.068
Age	42	25	1.69	0.097
Sampling Method (1=statistical)	455	665	0.69	0.500

Dependent Variable = Invoices
Adj. R^2 = .057, Model F-value = 1.99; p-value = 0.106, df = 66
[a]   Response to question: "If you answered 'guilty' to question one: please indicate how many invoices would need to be sampled to change your verdict to 'not guilty'."
[b]   Of the 76 participants who believed the defendant auditors to be "guilty" of negligence, five did not respond to this question. Thus the total n=71.
[c]   Response to question: "Have you ever served on a civil jury?"
[d]   Response to question: "Have you ever sued someone or been sued?"

6. Discussion

The central issue of this study is whether the method of sampling affects jurors' judgments against auditors in negligence trials. As posited by Pennington and Hastie's (1993) story model, the use of statistical methods (which are more difficult to understand) was expected to make the construction of a plausible story associating the defendant with a guilty verdict less likely. The results are in the direction expected with both the percentage of guilty verdicts and damage awards, however, only the damage awards were significantly affected by the sampling method. One might question whether the use of statistical methods makes either verdict, guilty or not guilty, less likely. Recall the story model's element of *completeness* when constructing the stories. Specifically, jurors in a civil negligence trial look for evidence that the defendant made an explicit choice to deviate from ordinary care (Hastie 1999). It is possible that participants, given the general difficulty people have in understanding and interpreting statistics, could imagine themselves making an honest mistake in a statistical calculation but could not see themselves making such a mistake in judgment. Another perspective on the jury decision process may be provided by attribution theories.

An *attribution* is defined as "an inference about why an event occurred or about a person's dispositions or

other psychological states" (Weary et al. 1989, 3). In a jury trial, the juror is an ex post "observer" of a reconstructed event making an attribution about that event. It is the jurors' job to consider the evidence, making a final attribution in the form of a verdict and damages, if applicable. While a number of potential biases or errors in attributions have been noted, the best documented is the "fundamental attribution error" (Ross 1977; Weary et al. 1989). Attributions can be categorized as *external* (physical, social circumstances surrounding the action) or *internal* (actor's ability, motivation, attitude, emotional state). Attributors have a tendency to "overestimate the importance of personal or dispositional factors relative to environmental influences" (Ross 1977, 184). The fundamental attribution error is a very robust finding (Weary et al. 1989).

In this study, the auditors were alleged to have chosen an insufficient sample size. In the case of statistical sampling, the allegations stated that the auditors "made an error in their statistical calculations" and if the correct statistical methods had been chosen "a larger sample size would have been used." This alleged calculation error is clearly an internal factor. In the case of judgmental sampling, the allegations stated that the "decision to sample only 100 items was clearly a mistake" and "the sample should have been larger." This is also clearly an internal factor. However, in the case of judgmental sampling, instead of a statistical miscalculation, the auditors were alleged to have made a mistake in judgment. Given that auditor judgment is a conscious act, this level of consciousness highlights the factor from an attribution perspective (as well as from a story model perspective). This leads jurors to over-attribute the judgmental error to the auditor vis a vis the statistical error.

Attribution theory also recognizes that people often try to infer the *motivations* of others in making an attribution (Kelley 1973). In the context of this study, if the auditors are perceived as being motivated to cut corners by examining fewer items,[3] that motivation would contribute to the jurors attributing the company's loss to the auditors. Kelley (1973) discusses several dimensions of motivation in attribution. One dimension that applies in this study is the ability to freely choose the behavior. If the actor's behavior is perceived as being largely the result of forces beyond the actor's control, then motivation would be perceived as low and, in turn, the behavior would be less likely to be attributed to the actor. In this study, judgmental sample sizes are entirely under the control of the auditor, while making statistical sample size mistakes might be perceived as being more accidental than intentional. Thus, the judgmental sampling approach provides more motivation for the auditor to use insufficient sample sizes, and damage awards in judgmental sampling are larger.

7. Limitations And Suggestions For Future Research

One limitation of this study is the simplistic design of the cases. This was intentional given that the effect of sampling method on juror judgments has not been studied, and a more simplified case lessened the chance of confounding variables being present. In this study, we chose to strengthen the internal validity, a trade-off that often decreases external validity. Future research in this area might include simplistic and more complex cases, as prior research demonstrates that case format influences verdicts (Pennington and Hastie 1993).

Another limitation of the study is that the case was devised so that the evidence was stacked against the auditor. This was considered necessary to give merit to the allegation that the sample size was insufficient in the statistical sampling case. In that case, the argument for sample size insufficiency due to a miscalculation was the most straightforward. However, other factors that drive statistical sample sizes could have been challenged by the plaintiff attorneys. It might be interesting to see if the results would also be noted in a case where the evidence against the auditor was more ambiguous. Future research might also include defendant expert witnesses rebutting the allegations of insufficient sample size.

8. Summary and Implications for Practice

This study found that allegations of errors of judgment in judgmental sampling applications resulted in jurors awarding 43% higher damages than when statistical miscalculations in statistical sampling applications were alleged. Further, for those who considered the auditors guilty, the sample sizes that would be necessary to convert their verdicts to not guilty was excessively large regardless of the sampling method employed.

---

[3]   Sullivan (1992) states that "the most plausible reason for (auditors using judgmental sampling) "is that nonstatistical sampling plans are less expensive."

Judgmental sampling is the most frequently used method of sampling employed by external, internal, and governmental auditors. External auditors, in particular, use judgmental sampling with the defense that it is allowable by GAAS, in effect extending this to mean that it would be defensible in court. But auditors should remember that jurors are perhaps operating under decision making models that result in a more favorable legal outcome for one GAAS-approved method (statistical) over another GAAS-approved method (judgmental) of sampling. This study used cases where the auditors were alleged to have insufficiently sampled sales invoices, the result of which was failure to discover a fraud. Overall, regardless of the sampling method used, the participants were more likely to find the auditors guilty of negligence than not guilty. Participants also had very high expectations regarding the sample size that they would consider sufficient to exonerate the auditor of the negligence charge. However, perhaps the most important implication that this study brings to the audit practitioner is in regard to damages. Errors in judgmental sampling resulted in nearly $900,000 (43%) more in damages being awarded the plaintiff, compared to damages awarded when the error was in a statistical calculation. This study involved a relatively small monetary amount, where the plaintiffs were seeking only $8 million in damages. In today's litigious environment, extrapolating the difference between statistical and judgmental sampling mistakes could easily be a much larger number. 📖

References

1.   AICPA. 1992. *Annual Report of the Public Oversight Board of the SEC Practice Section*. New York: AICPA.
2.   _____. 2000. *AICPA Professional Standards.* New York: AICPA.
3.   Arens, A. and Loebbecke, J. 2000. *Auditing: An Integrated Approach* (8th ed). Upper Saddle River, NJ: Prentice Hall.
4.   Buckless, F. and R. Peace. 1993. "The Influence of the Source of Professional Standards on Juror Decision Making". *The Accounting Review* (January): 164-175.
5.   Burgstahler, D., Glover, S. and Jiambalvo, J. 2000. "Error Projection and Uncertainty in the Evaluation of Aggregate Error". *Auditing: A Journal of Practice and Theory* (Spring) v19(1): 79 – 99.
6.   Colbert, J. 1991. "Statistical or non-statistical sampling: Which approach is best?" *The Journal of Applied Business Research* 7(2): 117-120.
7.   DeGroot, M, Fienberg, S. and Kadane, J. (Eds.). 1986. *Statistics and the Law.* New York: John Wiley and Sons.
8.   Elder, R. and R. Allen. 1998. "An Empirical Investigation of the Auditor's Decision to Project Errors". *Auditing: A Journal of Practice and Theory* (Fall) v17(2). 71 – 87.
9.   Fienberg, S. (Ed). 1989. *The Evolving Role of Statistical Assessments as Evidence in the Courts*. New York: Springer-Verlag.
10.   Hall, T., Hunton, J., and Pierce, B. 2000. "The use of and selection biases associated with nonstatistical sampling in auditing". *Behavioral Research in Accounting* 12: 231-256.
11.   Hans, V. and Lofquist, W. 1992. "Jurors' judgments of business liability in tort cases: Implications for the litigation explosion debate". *Law and Society Review* 26: 85-115.
12.   Hastie, Reid 1999. "The Role of 'Stories' in Civil Jury Judgments". *University of Michigan Journal of Law Reform*. V32(2). Winter. 227 - 239.
13.   Kachelmeier, S. and Messier, W., Jr. 1990. "An investigation of the influence of a nonstatistical decision aid on auditor sample size decisions". *The Accounting Review* 65(1): 209-226.
14.   Kadous, K. 2000. "The effects of audit quality and consequence severity on juror evaluations of auditor responsibility for plaintiff losses". *The Accounting Review* 74(3): 327-342.
15.   _____ 2001. "Improving Jurors' Evaluations of Auditors in Negligence Cases". *Contemporary Accounting Research.* (Fall) v18(3) 425-44.
16.   Kalven, H., Jr. and Zeisel, H. 1966. *The American Jury.* Boston, MA: Little, Brown.
17.   Kahneman, D. and Tversky, A. 1973. "On the psychology of prediction". *Psychological Review* 80: 237-251.
18.   Kelley, H. 1973. "The processes of causal attribution". *American Psychologist* 28: 107-128.
19.   MacCoun, R. 1996. "Differential treatment of corporate defendants by juries: An examination of the 'deep-pockets' hypothesis". *Law and Society Review* 30: 121-161.
20.   Niedermeier, K., Kerr, N., and Messe, L. 1999. "Jurors use of naked statistical evidence: exploring bases

and implications of the wells effect". *Journal of Personality and Social Psychology* 76(4): 533-542.

21.   Nelson, M. 1995. "Strategies of auditors: Evaluation of sample results". *Auditing: A Journal of Practice & Theory* 14(1): 34-49.

22.   Nisbett, R. and Ross, L. 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall.

23.   Olinsky, A., Mangiameli, P., and Chen, S. 1996. "Statistical support of forensic auditing". *Interfaces* 26(6): 95-104.

24.   Palmrose, Z. 1991. "Trials of legal disputes involving independent auditors: Some empirical evidence". *Journal of Accounting Research* 29 (supp): 149-185.

25.   Pennington, N. and Hastie, R. 1993. "The story model for juror decision making". In R. Hastie (Ed), *Inside the juror: The psychology of juror decision making* (192-221). New York: Cambridge University Press.

26.   Ponomon, L. and Wendell, J. 1995. "Judgmental versus random sampling in auditing: An experimental investigation". *Auditing: A Journal of Practice & Theory* 14(2): 17-34.

27.   Ross, L. 1977. "The intuitive psychologist and his shortcomings: Distortions in the attribution process". In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology (Vol. 10)* (173-220). New York: Academic Press.

28.   Sandys, M. and Dillehay, R.C. 1995. "First-ballot votes, predeliberation dispositions, and final verdicts in jury trials". *Law and Human Behavior* 19 (2): 175-195.

29.   *Sears, Roebuck, & Co. v City of Inglewood* (L.A. Superior Court, 1955)

30.   Smith, R., Penrod, S., Otto, A., and Park, R. 1996. "Jurors' use of probabilistic evidence". *Law and Human Behavior* 20 (1): 49-82.

31.   *State of Minnesota v. Kim*, (398 N.W.2d 544, Minn. 1987)

32.   Sullivan, J. 1992. "Litigation risk broadly considered". In R. Srivastava (ed.) *Auditing Symposium XI: Proceedings of the 1992 Deloitte & Touche/University of Kansas Symposium on Auditing Problems* (49-59). Lawrence, KS: University of Kansas.

33.   Tversky, A. and Kahneman, D. 1971. "The belief in the 'law of small numbers'". *Psychological Bulletin* 75: 27-36.

34.   Weary, G., Stanley, M., and Harvey, J. 1989. *Attribution*. New York: Springer-Verlag.

35.   Wells, G. 1992. "Naked statistical evidence of liability: Is subjective probability enough?" *Journal of Personality and Social Psychology* 62: 739-752.

36.   Wissler, R, Rector, K., and Saks, M. 2001. "The impact of jury instructions on the fusion of liability and compensatory damages". *Law and Human Behavior* 25(2): 125-139.

Notes