

DO AUDITORS THINK AS FREQUENTISTS?

Natalia V. Kotchetova
Ph. D. Candidate

William F. Messier, Jr.
Deloitte and Touche LLP Professor

School of Accountancy
J. Mack Robinson College of Business
Georgia State University
Atlanta, GA

*Preliminary draft: do not quote without authors' permission.
Comments welcome.*

December 2001

The second author was supported by a research grant from the Robinson College of Business. We thank Professor Aasmund Eilifsen of Norwegian School of Economics and Business Administration for his help in administering our experiment. Thanks also go to the participants of the workshop at Georgia State University for their comments. We acknowledge assistance provided by J. Allen Jinnette at the early stage of this project.

DO AUDITORS THINK AS FREQUENTISTS?

ABSTRACT

We investigate whether auditors are more likely to make probabilistic judgments in accordance with Bayes rule and less likely to neglect base rates when they make their assessments in a frequency format versus a probability format. We test 3 hypotheses based on research in cognitive psychology by Gigerenzer and his colleagues (e.g., Gigerenzer et al. 1988; Gigerenzer 1991). Four factors are manipulated (base rate, response format, case context, and order) in a 2⁴ mixed factorial design. Sixty-eight auditors participated in the experiment.

The results indicate that (1) auditors have difficulty recognizing the difference between high and low base rates of fraud that are based on realistic data, (2) auditors' assessments are closer to a Bayesian response and have lower base rate neglect when evaluating an auditing case versus a non-auditing case, and most importantly, (3) auditors' responses using a frequency format are closer to the base rate and Bayesian response than those auditors using a probability format.

DO AUDITORS THINK AS FREQUENTISTS?

INTRODUCTION

Research in auditing emphasizes the importance of auditors' ability to make probabilistic inferences about financial statement assertions. A major stream of work in this area examined auditors' susceptibility to heuristics and biases (e.g., Joyce and Biddle 1981a,b; Kida 1984; Frederick and Libby 1986). The most extensively examined heuristics in auditing are anchoring and adjustment, availability, representativeness, and confirmatory strategies. In reviewing this literature, Smith and Kida (1991) report that auditors are susceptible, but do not always fall prey, to heuristics and biases.

In this study, we investigate whether auditors are more likely to make probabilistic judgments in accordance with Bayes rule, and are less likely to neglect base rates when they make their assessments in a frequency format versus a probability format. We draw on research in auditing by Joyce and Biddle (1981b) and in cognitive psychology by Gigerenzer and his colleagues (e.g., Gigerenzer et al. 1988, 1991; Gigerenzer, Hoffrage, and Kleinbölting 1991; Gigerenzer 1991; Gigerenzer and Hoffrage 1995; Gigerenzer and Goldstein 1996) to test three hypotheses. We manipulate four factors (base rate, response format, case context, and order) in a 2^4 mixed factorial design. Sixty-eight auditors participated in the experiment. Our results show that (1) auditors have some difficulty recognizing the difference between high and low base rates of fraud that are based on realistic data, (2) auditors' assessments are closer to a Bayesian response and have lower base rate neglect when solving an auditing case versus a non-auditing case, and most importantly, and (3) auditors' responses using a frequency format are closer to the base rate and Bayesian response than those auditors using a probability format.

The paper is organized as follows. The next section develops our research hypotheses. This is followed by a section that describes the methodology while the fourth section summarizes the results. The last section discusses the results and presents our conclusions.

HYPOTHESES DEVELOPMENT

Single Case Probability vs. Relative Frequency

The *representativeness* heuristic is used when the individual assesses the probability that an item belongs to a population based on the extent to which the item is similar in its essential properties to that population. Events that are more similar to the population (representative) tend to be judged to have a greater probability of occurrence than less similar events, independent of their actual frequency (Libby 1981; Smith and Kida 1991). Tversky and Kahneman (1972) used this heuristic to explain the following biases: (1) neglect of base rates; (2) disregard for the impact of sample size; and (3) insensitivity to the predictive ability and source reliability of data. In a series of experiments, they observed that participants tended to ignore or significantly underweight population base rates, sample size, and data reliability, and tended to make probabilistic judgments on the basis of the similarity between the evidence about an item and possible outcomes.

For most of their experiments, Kahneman and Tversky used “naive” participants, i.e. participants without special knowledge about the task and/or extensive training in probability theory. Auditing studies investigated whether experienced professionals fell prey to the representativeness heuristic and, in particular, neglect of base rates. Unlike “naive” participants, auditors reacted systematically to base rates; however, they did not fully appreciate the

diagnosticity of evidence and made insufficient adjustments (e.g., Joyce and Biddle, 1981b; Swieringa et al. 1976).¹

For example, Joyce and Biddle (JB) (1981b) tested for auditors' base rate sensitivity in a series of experiments using a fraud prediction task (experiments 1, 2a, 2b, and 2c). In JB's experiments, subjects were presented with a diagnostic task to assess the probability of management fraud given individuating data about the client manager's personality profile and a description of a procedure used to test the existence of management involvement in fraudulent activities. Although the auditors were not familiar with the psychological decision aid presented to them in the experiment, the task of assessing client's integrity was known to them. JB manipulated two factors in a between-subjects factorial design: (1) the base rate of fraud, and (2) the diagnosticity of the decision aid. The results suggested that base rates were not ignored by the majority of participants, although the magnitude of adjustment was insufficient. Consistent with normative Bayesian principles, participants in the base rate condition regressed their judgments towards the base rate (correct direction), but the magnitude of the revision was not adequate. Holt (1987) replicated JB's (1981b) experiments using auditors and accounting students as participants. Both subject groups largely ignored base rate information. Holt (1987) concluded that the framing of the case materials, i.e. whether information in the case appears particularly relevant to the participants' statistics training, drove Bayesian integration of evidence in her experiments.

Base rate neglect bias may also be mitigated by the response format. Gigerenzer and his colleagues argue that if participants are asked to estimate the probability of a single event, the question does not connect to probability theory in their minds, whereas the question about

¹ This result was especially observed in designs where base rates were manipulated within subjects.

frequency of such an event does (Gigerenzer et al. 1988, 1991; Gigerenzer, Hoffrage, and Kleinbölting 1991; Gigerenzer 1991; Gigerenzer and Hoffrage 1995; Gigerenzer and Goldstein 1996). This happens because (a) an estimation of confidence in the likelihood of a single-event and the judgment of frequency are cognitively different processes (Gigerenzer, Hoffrage, and Kleinbölting 1991), and because (b) Bayesian computations involve fewer steps and thus are simpler when information is encoded in a frequency format rather than in a standard probability format (Gigerenzer and Hoffrage 1995).

Gigerenzer, Hoffrage, and Kleinbölting (1991) propose a theory of probabilistic mental models to point out the distinction between confidence (degree of belief) about a single event and a frequency judgment. The theory holds that in order to solve a particular diagnostic problem either a local mental model (LMM) or a probabilistic mental model (PMM) is constructed. A LMM is used if the problem at hand can be solved by retrieving relevant knowledge from memory, or by executing an elementary logical operation, such as the method of exclusion, in order to compensate for missing knowledge. A LMM includes two alternative objects or events a and b , and a target (criterion) variable t ; no probabilities are assigned to the target variable. Once the task is solved, the knowledge produced by LMM is evaluated as certain. For example, the decision rule for LMM may be as follows: if $a > b$ on t , choose b (or a) with 100% confidence.

If no memory cues can be activated for the diagnostic problem at hand, a PMM is constructed. A PMM consists of a *reference class* of objects or events R that includes alternatives a and b , a target variable t , and a list of probability cues that are generated from the reference class with respect to the target variable. First, probability cues are hierarchically ordered according to their cue validities. Second, the list is scanned for a probability cue that

can be activated. Third, once a cue is found that can be activated for the problem under consideration, the judgment is made accordingly and the subject's confidence in this particular judgment is determined by the cue validity of the activated cue. For example, if events a and b belong to the reference class R , and C_i is a probability cue for the target variable in a reference class R (a and b differ on C_i), then a decision rule may be as follows: choose a if $p(a|aC_i b; R) > p(b|aC_i b; R)$, i.e. if probability of a given that a and b differ on cue C_i is greater than probability of b given that a and b differ on cue C_i , assuming that both a and b belong to the same reference class R . In this case, if a is chosen, the confidence that a is correct is given by the cue validity $p(a|aC_i b; R)$.

Gigerenzer, Hoffrage, and Kleinbölting (1991) argue that the structure of a PMM for the probability task described above is different from the structure of a PMM for the frequency task on the following parameters: target variable, reference class, and probability cues. In a frequency task, the reference class is represented by a series of similar questions in similar testing situations or experiences (versus a general type of objects or events in the probability task), the target variable is given by a number of correct responses (versus estimation of cue validity on the activated cue, i.e. probability of a chosen event), and probability cues are base rates known from experience or offered by the problem information (versus a characteristic on which members of the reference class differ). Since evidence suggest that human memory encodes frequencies of events as an automatic cognitive process requiring little attention or conscious effort (Hasher and Zacks 1979; Hintzman, Nozawa, and Irscher 1982; Zacks, Hasher, and Sanft 1982), the subjects who had repeated experience with a reference class, a target variable, and cues in their environment should be better able to perform the frequency task than the probability task. This effect should occur because the subjects should be better able to construct a PMM for the

frequency task than for the probability task due to the differences in target variables and reference classes between the two tasks.

In addition to differences in the structure of the PMM, construction of cue validities for a frequency task PMM involves fewer steps and is computationally simpler than construction of such validities for a probability task (Gigerenzer 1996; Gigerenzer and Hoffrage 1995). First, the frequency format is simpler because it involves calculations with natural numbers, whereas the probability format involves fractions. Second, as documented in a series of experiments by Gigerenzer and Hoffrage (1995), a frequency format enables subjects to invoke computational shortcuts that produce results that are very close to a Bayesian algorithm. For example, in order to calculate $p(\text{disease}|\text{symptom})$, it is often sufficient to divide the number of cases that have both the symptom and the disease by the total number of symptom cases (pp.686-687). In this case a more parsimonious menu of information (number of cases with both symptom and disease and number of false positives) is used, relative to the informational requirement of a standard Bayesian probability format that would entail an additional piece of information: the base rate of symptoms in a population.²

The implications of Gigerenzer et al.'s (1991) and Gigerenzer and Hoffrage's (1995) arguments for experimental research bear upon the link between the response formats of case materials, and evaluating discrepancies of subject's answers from Bayesian calculations as heuristics and biases. We test Gigerenzer and Hoffrage's (1995) hypotheses of whether base rate use is dependent on the properties of the task-specific PMM by manipulating the *response format* of the case materials between a probability format and a frequency format. We propose the following main hypotheses:

² Other shortcut algorithms that produce an outcome very close to Bayesian include pictorial "beam analysis"

H1a: Auditors make fraud likelihood assessments that are closer to the given base rate when they receive case information and make required judgments in *frequency* format rather than in *probability* format.

H1b: Auditors make fraud likelihood assessments that are closer to the Bayesian response when they receive case information and make required judgments in *frequency* format rather than in *probability* format.

Base-Rate Neglect and Case Context

If the long-run frequencies of events are encoded in human memory through experience and/or training (Hasher and Zacks 1979; Hintzman, Nozawa, and Irmscher 1982; Zacks, Hasher, and Sanft 1982), participants should reason more normatively (as Bayesians) when the case context is in the domain of their basic training and/or experience. Therefore, consistent with JB (1981b) and Swieringa et al. (1976), and as a validity check, we expect auditors to be less susceptible to the base rate neglect heuristic when the context of the case is an auditing, rather than a non-auditing problem:

H2a: Auditors make likelihood assessments that are closer to the given base rate when they are faced with an *auditing* problem rather than with a *non-auditing* problem.

H2b: Auditors make likelihood assessments that are closer to the Bayesian response when they are faced with an *auditing* problem rather than with a *non-auditing* problem.

The under-weighting of base-rate information in JB's (1981b) and Holt's (1987) experiments could be also due to the arbitrary manipulation of base rates of fraud in their experiments. If base rates differed from participants' real world experience, they could have raised doubt in participants' minds about the realism of the task.³ Additionally, JB's results in

(Gigerenzer and Hoffrage 1995, 689-690), rare event shortcut (p. 690), big hit-rate shortcut, and comparison shortcut (p. 691).

³ In JB's (1981b) experiments 2a and 2b participants were given descriptions of the management of a company and were told that it was drawn from a collection of 10 management descriptions assembled by an accounting firm as a part of a study to determine whether management descriptions are helpful in detecting management fraud. The base rates in experiment 2a were .30 in the condition A and .70 in the condition B; the population size of the

experiment 1 could have been partially caused by low reliance on the decision aid in the conditions where participants' are cognizant of its notably low performance (Ashton 1990).⁴ The under-weighting of the base rates was especially prevalent when the base rate was low (.01) and the diagnosticity of the decision aid was high (.20).

Following JB's (1981b) and Holt's (1987) experiments, but using realistic fraud base rates based on the *1998 KPMG Fraud Survey* (1998), we propose the following hypotheses that, if supported, will serve as a second internal validity check for our experiment against extant research:

H3a: Auditors make fraud likelihood assessments that are closer to the given base rate when base rates are *high* (.25) rather than when base rates are *low* (.083).

H3b: Auditors make fraud likelihood assessments that are closer to the Bayesian response when base rates are *high* (.25) rather than when base rates are *low* (.083).

METHODOLOGY

Experimental Design and Case Materials

The experimental design is a 2x2x2x2 mixed factorial design (Figure 1). Three factors are manipulated between-subjects and one factor within-subjects. The between-subject variables are: (1) the base rate of fraud (*high* at 0.25 or 3 out of 12 vs. *low* at 0.083 or 1 out of 12); (2) response format (*probability* vs. *frequency*); and (3) order (auditing case presented first vs. non-auditing case presented first). The within-subject factor is case context, manipulated at the two

management descriptions was 10. The base rates in experiment 2b were .10 and .50, in conditions A and B, respectively; with the population increased to 100 descriptions. Holt (1987) used base rates as in JB's (1981b) experiment 2b.

⁴ JB's participants in the experiment 1 were given a psychological decision aid, i.e. a database of personality profiles of key managers that had been designed by clinical psychologists to help the auditors in evaluating the client's management propensity to commit fraud (Joyce and Biddle 1981b, 329).

levels: *auditing* (assessment of the likelihood that client management is fraudulent) vs. *non-auditing* (assessment of the likelihood that a patient has breast cancer).⁵

Insert Figure 1 here

We select .25 and .083 as levels of the base rate (BR), following upper and lower rates of financial statement fraud of the magnitude between \$5,000 and \$999,0000 reported in *1998 KPMG Fraud Survey* (p. 9). We choose these levels versus the more extreme values used by JB (1981b) in order to ensure case realism.

The second between-subject factor, response format (RM), was manipulated by using appropriate language in the body of the management fraud case and in the outcome question requesting assessment of the likelihood of fraud for a client given information provided in the case (see Appendix). In the probability condition, participants received base rates and false positive rates stated in percentages, and answered the following outcome question: “What is your assessment of the probability that the management described above was engaged in fraudulent activity? _____%”. In the frequency condition, participants received the same base rates and the same false positive rates, stated as frequencies, and answered the following question about the likelihood of fraud: “In your opinion, of the 10 randomly selected clients fitting the above management description, how many out of 10 were engaged in fraudulent activity?”⁶ Similar wording was used for the non-auditing case.

The order of cases (ORDER) was varied at two levels: (a) auditing case presented first, followed by non-auditing case; and (b) non-auditing case presented first, followed by auditing case.

⁵ Base rates were manipulated in auditing cases; they were held constant (1%) across all cells in non-auditing cases.

⁶ False positives rates did not vary across conditions, i.e. they were stated at 5% (management fraud case) and 9.6% (breast cancer case) in all cells.

Finally, context of the case (CASE) was manipulated at two levels: auditing and non-auditing. The auditing case was largely based on JB's (1981b) management fraud case, with appropriate modifications for our manipulations. For the non-auditing case, we use a breast cancer diagnosis problem developed by Eddy (1982) to examine physicians' reasoning about single-event probabilities (see Appendix).

Participants and Procedure

The participants in the experiment were students in the graduate program in accounting and auditing at the Institute of Accounting, Auditing and Law, Norwegian School of Economics and Business Administration in Bergen, Norway. To enter this program, the participants must have completed a four-year business administration degree or passed the examinations required for registered auditors (the first level of certification in Norway). Completion of the program, passing a rigorous examination, and three years of practical training allows a candidate to be eligible for state authorization (the highest level of certification). Table 1, Panel A summarizes demographic information about the participants. The majority of participants were either employed as auditors when the experiment was administered (54) or had prior auditing experience (8). Among 54 participants currently working as auditors, 22 (40.7%) were staff or associates and 28 (51.9%) were seniors. Thirteen participants (18.8%) reported that they had encountered a fraudulent client in their audits. The mean auditing experience was 2.4 years. In summary, participants in the experiment are representative of the targeted population, i.e. staff and senior auditors.

Insert Table 1 here

The experiment was administered by the director of the program during a class session. Seventy questionnaires were completed. Two participants' were deleted because one response

was not interpretable and another participant one did not respond to the fraud case (“have no idea”). Thus, 68 responses are used in our analyses.

Since our experiment included the assessment of single-event probabilities/frequencies, participants were asked to report their knowledge of statistics. The mean training in statistics was 49.87 contact hours; the average self-rated proficiency in statistics was 3.03 on a scale from 1 (very low) to 7 (very high). Only 8 subjects (11.6%) reported familiarity with Bayes’ theorem; subsequently, only one participant (1.4%) applied it in solving the experimental cases. Not surprisingly, participants did not appear very confident in their probability/frequency assessments: the mean confidence rating was 2.39 for the fraud task and 2.54 for the medical task, on a scale from 1 (very low) to 5 (very high).

Because we conducted our experiment in Norway and experimental materials were written in English, we asked participants to self-rate their knowledge of the English language.⁷ They appeared to be proficient in English: the mean rating was 4.17 on a scale from 1 (very low) to 7 (very high). Participants were also asked questions about the realism and understandability of case materials. Overall, most participants found the case materials easy to understand (55 or 79.7%) and realistic (54 or 78.3%).

Dependent Variables

We used three measures as dependent variables in order to test our hypotheses. The first measure was the participant’s assessment of the likelihood of management fraud and the likelihood that the patient has cancer: i.e., RAWF and RAWC, respectively.⁸ The second measure was the absolute value of the deviation of the participant’s assessments from respective

⁷ It should be noted that the participants in this program received a major portion of their graduate auditing training in English. They also use an English auditing text and read numerous accounting and auditing articles in English.

⁸ In order to have a uniform scale measure of the dependent variable for statistical tests, we convert percentages and frequencies to proportions in probability and frequency conditions, respectively.

base rates; DVBRF = |RAWF – Fraud Base Rate| and DVBRC = |RAWC – Cancer Base Rate|. The third measure used the absolute value of the deviation of the participant’s assessments from respective Bayesian responses; DVBAF = |RAWF – Fraud Bayesian Response| and DVBAC = |RAWC – Cancer Bayesian Response|. ⁹ We performed additional tests on the data by scaling the second and third dependent variables.

RESULTS

Table 2 presents descriptive statistics for each of the three measures of the dependent variable.

Insert Table 2 here

Base Rate Neglect

Since base rates were manipulated only in the auditing case materials by design, we test for the effect of the variable BR first. Hypotheses 3a and 3b state that auditors make fraud likelihood assessments that are closer to the base rate (3a) and the Bayesian response (3b) when base rates are high rather than when base rates are low. Two ANOVAs were run using BR and ORDER as between-subject factors. Table 3, Panel A presents the results using DVBRF as the dependent variable while Panel B presents the results using DVBAF. We use the dependent variables in the deviation form to test these hypotheses because they relate to the proximity of likelihood assessments to base rates and to Bayesian responses.

⁹ Bayesian responses are calculated by the following formula: $f(\theta|y) = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}$, where $f(\theta)$ equals

the true probability of fraud and $f(y)$ equals the fraud base rate. For instance, when base rate is high at .25, Fraud Bayesian Response = $(.25)(1) \div [(.25)(1) + (.75)(.05)] = .87$; when base rate is low at .083, Fraud Bayesian Response = $(.083)(1) \div [(.083)(1) + (.917)(.05)] = .65$. The Bayesian response for the breast cancer case is determined using the same formula as for the management fraud case, and .01 as base rate and .096 as false positive rate: Cancer Bayesian Response = $(.01)(.79) \div [(.01)(.79) + (.99)(.096)] = .08$.

Insert Table 3 here

The main effect of BR is not significant when the dependent variable is the deviation from the base rate ($F = .013$, $p\text{-value} = .910$). BR is also not significant when the dependent variable is the deviation from the Bayesian response (Table 3, Panel B): $F = 1.005$, $p\text{-value} = .320$.¹⁰

In an attempt to examine this issue further, we used scaled versions of the dependent variables, i.e. DVBRF÷Fraud Base Rate and DVBAF÷Fraud Bayesian Response (not tabulated). BR remains insignificant at the conventional levels when the scaled deviations from the base rate and from the base rates and from the Bayesian response are used. Thus, we do not find support for H3a and H3b. Auditors appear to have difficulty recognizing high and low base rates of fraud that are based on realistic data. In our remaining tests, we collapse our design across the levels of BR.

Frequency versus Probability Format

Hypotheses 1a and 1b contend that auditors make fraud likelihood assessments that are closer to base rate and Bayesian response when they receive case information and make required judgments in frequency format rather than in probability format. Tables 4 and 5 report the results of testing H1a and H1b. The main effect for RM is marginally significant in the analyses of auditing cases when either absolute deviation from the base rate of fraud (Table 4, Panel A: $F = 3.949$, $p\text{-value} = .051$) or from the Bayesian response (Table 4, Panel B: $F = 3.049$, $p\text{-value} = .086$) are used as a dependent variable. However, ANOVA results for non-auditing cases show a significant main effect of RM. Panels C and D of Table 4 report a significant main effect for RM

¹⁰ The variable ORDER is not significant in both ANOVAs ($F = 1.213$, $p\text{-value} = .275$ when DVBRF is used as a dependent variable; $F = 1.727$, $p\text{-value} = .193$ when DVBAF is used as a dependent variable). The interaction effects BR x ORDER are not significant at the conventional levels.

when either absolute deviation from the base rate of cancer (Table 4, Panel C: $F = 16.304$, p -value = .000) or from the Bayesian response (Table 4, Panel D: $F = 16.461$, p -value = .000) are used as a dependent variable.^{11, 12}

Insert Tables 4 and 5 here

To complement our ANOVA tests of H3a and H3b, we report means and standard deviations for the dependent variables by response mode (Table 6). Consistent with results reported earlier, we observe smaller mean deviations from base rates or from Bayesian response in frequency cells, than in probability cells, with only two exceptions: DVBAF and DVBAF+Fraud Bayesian Response. Thus, we find full support for H3a. H3b is only supported for the non-auditing case.

Insert Table 6 here

Case Context

H2a and H2b state that auditors make likelihood assessments that are closer to the base rate and Bayesian response, respectively, when they are faced with an auditing problem rather than with a non-auditing problem. We test these hypotheses using a repeated measures ANOVA (Table 5). Table 5, Panel A shows a significant main effect for the variable CASE ($F = 27.439$, p -value = .000) using the deviations from base rate as dependent variables. A Wilcoxon signed rank test also indicates that the number of negative ranks for the differenced deviations from the base rates was greater (44) than the number of positive ranks (23) ($z = -4.142$, p -value = .000) (see Table 7). In other words, participants judgments were further from base rates greater

¹¹ When scaled versions of DVBRF, DVBR, DVBAF, and DVBA are used as dependent variables, main effect of RM remains significant in non-auditing cases only ($F = 2.194$, p -value = .143 for DVBRF; $F = 1.786$, p -value = .186 for DVBAF; $F = 16.304$, p -value = .000 for DVBR; and $F = 16.461$, p -value = .000 for DVBA). These results are not tabulated.

¹² Neither main effect of ORDER nor interaction RM x ORDER are significant at conventional levels in tests of H3a and H3b (Table 4).

number of times when they solved a non-auditing case (44) than the number of times when they solved an auditing case (23). These results support H2a.

The statistical tests of H2b are sensitive to the scaling of the dependent variable. The CASE variable was not significant when the deviation from the Bayesian response was used as the dependent variable (Table 5, Panel B). However, when we scale these deviations by respective normative Bayesian responses (DVBRF÷Fraud Bayesian Response and DVBRC÷Cancer Bayesian Response), the main effect for CASE is significant ($F = 126.882$, $p\text{-value} = .000$) (not tabulated). Additionally, Table 6 reports a greater number of negative ranks for the differenced scaled deviations from the Bayesian responses (54) than positive ranks (13) ($z = -5.847$, $p\text{-value} = .000$). Thus, subject to the scaling effect of the dependent variable, we find support for H2b. Auditors' reasoning is closer to a Bayesian response and base-rate neglect is lower when they solve an auditing problem compared to solving a non-auditing problem.

Insert Table 7 here

DISCUSSION AND CONCLUSIONS

Conclusions

In this study, our primary investigation was how the response format of base rate information impacts the proximity of auditors' fraud likelihood assessments to base rates and to Bayesian computations. Using auditing staff and seniors as participants in an experiment, we found that the argument by Gigerenzer and colleagues (Gigerenzer et al. 1988, 1991; Gigerenzer, Hoffrage, and Kleinbölting 1991, Gigerenzer 1991, Gigerenzer and Hoffrage 1995, Gigerenzer and Goldstein 1996) about the frequentist nature of human statistical intuition has merit. In the context of significant management level fraud likelihood assessments, participants' likelihood

assessments were closer to base rates when they were supplied with base rates and were required to respond in frequency format rather than in probability format. Our results are weaker and tend to be sensitive to scaling effects when we focus on the proximity of auditors' judgments to Bayesian calculations. We suspect that these findings are due to the low familiarity of our participants with Bayes' Theorem and low self-rated knowledge of statistics.

Also as expected, our participants performed better when presented with an auditing problem than when faced with a medical problem. However, when solving an auditing problem they appeared to be insensitive to realistic base rates of management fraud.

Contributions and Limitations

Aside from investigating the effect of response format (frequency vs. probability) on audit judgment, our study contributes to knowledge about the application of frequency information by auditors. The extant research suggests that frequency information can be acquired either via individual experiences or from individual statistics. The method of information acquisition affects the way in which it is (a) stored in memory, (b) recalled (and accuracy of recall), and (c) is applied in audit judgment (Nelson 1994). Our study focuses on the latter, the application and the articulation of summary-form frequency information by the auditors in the task of performing a management fraud probability assessment. We recognize that use of summary form frequency information is a limitation of our study because prior research found that summary-form frequency information is learned less accurately than frequency information learned experientially. Butt (1988) and Hintzman et al. (1982) compared the accuracy of error-cause frequency knowledge learned through instance-by-instance presentations, summary data, and a mixture of the two. They found that frequency knowledge is

learned most accurately through instance-by-instance presentations, and least accurately when encountered in purely summary form.

Another limitation is that we do not distinguish between frequency of causes and frequency of effects of management fraud; this lack of distinction is by design because our research question concentrates mainly on the effect of internal problem representation, driven by response format, on probability judgment and less so on the nature of the problem.

Finally, we do not examine sub-population frequency perceptions and do not differentiate between the use of sub-population and population frequency knowledge as was done in the literature on recall and evaluation of error explanations (Tuttle 1996). The base-rates in this study apply to the general population of clients; no special sub-population of frauds is assumed.

The aforementioned limitations should provide fruitful ground for future research.

References

- Ashton, R. H. 1990. Pressure and performance in accounting decision settings: paradoxical effects of incentives, feedback, and justification. *Journal of Accounting Research* 28 (Supplement): 148-180.
- Ajzen, I. 1977. Intuitive theories of events and the effects of rate base information on prediction. *Journal of Personality and Social Psychology* 35: 303-314.
- Butt, J.L. 1988. Frequency judgments in auditing-related task. *Journal of Accounting Research* 26: 315-330.
- Eddy, D.M. 1982. Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgments Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Frederick, D.M., and R. Libby. 1986. Expertise and auditors' judgments of conjunctive events. *Journal of Accounting Research* 24(2): 1986.
- Gigerenzer, G., W. Hell, and H. Blank. 1988. Presentation and content: the use of base rates as a continuous variable. *Journal of Experimental Psychology* 14 (3): 513-525.
- _____, 1991. How to make cognitive illusions disappear. *European Review of Social Psychology* 2: 83-111.
- _____, U. Hoffrage, and H. Kleinbölting. 1991. Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review* 98(4): 506-528.
- _____, and U. Hoffrage. 1995. How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review* 102 (4): 684-704.
- _____, and D. G. Goldstein. 1996. Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review* 103(4): 650-669.
- Hackenbrack, K. 1992. Implications of seemingly irrelevant evidence in audit judgment. *Journal of Accounting Research* 30: 126-136.
- _____. 1993. The effect of experience with different sized clients on auditor evaluations of fraudulent financial reporting indicators. *Auditing: A Journal of Practice and Theory* 12: 99-110.
- Hasher, L., and R. T. Zacks. 1979. Frequency and effortful processes in memory. *Journal of Experimental Psychology: General* 108: 356-388.
- Hintzman, D., G. Nozawa, and M. Irmscher. 1982. Frequency as a non-propositional attribute of

- memory. *Journal of Verbal Learning and Verbal Behavior* 21: 127-141.
- Hodges, J.L., and E.L. Lehman. 1970. *Basic Concepts of Probability and Statistics* San Francisco, CA: Holden-Day, Inc.
- Holt, D. 1987. Auditors base rates revisited. *Accounting, Organizations, and Society* 12(6): 571-578.
- Joyce, E., and G. Biddle. 1981a. Anchoring and adjustment in probabilistic inference in auditing. *Journal of Accounting Research* 19: 120-145.
- Joyce, E., and G. Biddle. 1981b. Are auditors' judgments sufficiently regressive? *Journal of Accounting Research* 19: 323-349.
- Kahneman, D., and A. Tversky. 1972. Subjective probability: a judgment of representativeness. *Cognitive Psychology* 3: 430-454.
- Keppel, G. 1982. *Design and Analysis: A Researcher's Handbook* Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Kida, T. 1984. The effect of causality and specificity on data use. *Journal of Accounting Research* 22(1): 145-152.
- KPMG. *KPMG Fraud Survey*. 1998. KPMG Peat Marwick LLP.
- Locksley, A., and C. Stangor. 1984. Why versus how often: causal reasoning and the incidence of judgmental bias. *Journal of Experimental Social Psychology* 20: 470-483.
- Nelson, M. 1994. The learning and application of frequency knowledge in audit judgment. *Journal of Accounting Literature* 13: 185-211.
- Nisbett, R., and E. Borgida. 1975. Attribution and the psychology of prediction. *Journal of Personality and Social Psychology* 32: 932-943.
- _____, and L. Ross. 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Smith, J. F., and T. Kida. 1991. Heuristics and biases: expertise and task realism in auditing. *Psychological Bulletin* 109(3): 472-489.
- Swieringa, R., Gibbins, M., Larsson, L., and J.L. Sweeney. 1976. Experiments in the heuristics of human information processing. *Journal of Accounting Research* 4(Supplement): 159-187.
- Tuttle, B. M. 1996. Using base rate frequency perceptions to diagnose financial statement error

causes. *Auditing: A Journal of Practice and Theory* 15(1): 104-121.

Tversky, A., and D. Kahneman. 1982. Evidential impact of base rates. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgments Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Zacks, R. T., L. Hasher, and H. Sanft. 1982. Automatic encoding of event frequency: Further findings. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8: 106-116.

Figure 1
Experimental Design

		BASE RATE	
		High	Low
Response Mode	Probability	Order 1: AUD/NAUD Order 2: NAUD/AUD	Order 1: AUD/NAUD Order 2: NAUD/AUD
	Frequency	Order 1: AUD/NAUD Order 2: NAUD/AUD	Order 1: AUD/NAUD Order 2: NAUD/AUD

AUD = auditing case
NAUD = non-auditing case

Table 1
Information about Participants
(n = 69)

Panel A

Demographic Question	Minimum	Maximum	Mean	Standard Deviation
Experience as an auditor, in years	0.0	9.0	2.40	1.58
Age, in years	22	37	27.29	2.82
Training in statistics, in contact hours	0	150	49.87	37.24
Self-rated knowledge of statistics, on a Likert's scale from 1 (very low) to 7 (very high)	1.00	5.50	3.03	1.17
Self-rated knowledge of English, on a Likert's scale from 1 (very low) to 7 (very high)	1.00	7.00	4.17	1.17
Confidence in likelihood assessment made during the fraud task, on a Likert's scale from 1 (very low) to 5 (very high)	1	5	2.39	.99
Confidence in likelihood assessment made during the medical task, on a Likert's scale from 1 (very low) to 5 (very high)	1	5	2.54	1.16

Panel B

Demographic Question	Categories	Number of Participants	Percent
Gender	Male	39	56.5%
	Female	30	43.5%
Currently working as an auditor	Yes	54	78.3%
	No	15	21.7%
If currently working as an auditor, level in the firm	Staff/Associate	22	31.9% (40.7%) ¹
	Senior	28	40.6% (51.9%) ¹
	Response Missing	4	5.8% (7.41%) ¹
If currently not working as an auditor, past auditing experience	Yes	8	11.6% (53.3%) ²
	No	7	10.14% (46.7%) ²
Encountered fraudulent client	Yes	13	18.8%
	No	46	66.7%
	Response Missing	10	14.5%
Familiar with Bayes' Theorem	Yes	8	11.6%
	No	61	88.4%

Applied Bayes' Theorem in solving the experimental case	Yes	1	1.4%
	No	66	88.4%
	Response Missing	3	4.3%
Easily understood fraud case materials	Yes	55	79.7%
	No	10	14.5%
	Response Missing	4	5.8%
Realism of fraud case materials	Yes	54	78.3%
	No	9	14.3%
	Response Missing	6	8.7%

¹ The first number represents % of the sample. The number reported in parentheses represents percent of those participants who answered "yes" to the question "Do you currently work as an auditor?", i.e. 54 participants.

² The first number represents % of the sample. The number reported in parentheses represents percent of those participants who answered "no" to the question "Do you currently work as an auditor?", i.e. 15 participants.

Table 2
Descriptive Statistics for the Dependent Variable

Variable	N	Minimum	Maximum	Mean	Standard Deviation
RAWF	68	.0000	.9500	.3421	.3078
RAWC	68	.0000	1.0000	.4935	.3925
DVBRF	68	.0000	.8667	.2171	.2798
DVBRC	68	.0000	.9900	.4841	.3917
DVBAF	68	.0014	.8696	.4670	.2096
DVBAC	68	.0032	.9232	.4481	.3557
DVBRF÷Fraud Base Rate	68	.0000	10.4046	1.8041	2.8965
DVBRC÷Cancer Base Rate	68	.0000	99.0000	48.4106	39.1749
DVBAF÷Fraud Bayesian Response	68	.0022	1.0000	.6307	.2733
DVBAC÷Cancer Bayesian Response	68	.0417	12.0208	5.8341	4.6319

Notes:

RAWF is participant's assessment of the likelihood of fraud in a fraud task

RAWC is participant's assessment of the likelihood of breast cancer in a medical task

DVBRF= |RAWF – Fraud Base Rate|

DVBRC = |RAWC – Cancer Base Rate|

DVBAF = |RAWF – Fraud Bayesian Response|. Fraud Bayesian Response is defined as the following:

$$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}, \text{ where } f(\theta) \text{ equals the true probability of fraud and } f(y) \text{ equals the fraud base rate}$$

rate

DVBAC = |RAWC – Cancer Bayesian Response|. Cancer Bayesian Response is defined as the following:

$$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}, \text{ where } f(\theta) \text{ equals the true probability of breast cancer and } f(y) \text{ equals the cancer base rate}$$

cancer base rate

DVBRF÷.25 for high base rate conditions, and DVBRF÷.083 for low base rate conditions

DVBRC÷.01

DVBAF÷.87 for high base rate conditions, and DVBAF÷.65 for low base rate conditions. Also, see footnote 8 in the text

DVBAC÷.08. Also, see footnote 8 in the text.

Table 3
ANOVA Results for Base Rate Tests (Auditing Cases Only)
(n = 68)

Panel A: DVBRF = | Participants' Raw Response- Fraud Base Rate |

Source	SS	df	F	p -value
Intercept	3.242	1	40.341	.000
BR	1.039E-03	1	.013	.910
ORDER	9.748E-02	1	1.213	.275
BR * ORDER	5.819E-04	1	.007	.932
Error	5.144	64		

Panel B: DVBAF = | Participants' Raw Response- Fraud Bayesian Rate |

Source	SS	df	F	p -value
Intercept	14.700	1	335.500	.000
BR	4.403E-02	1	1.005	.320
ORDER	7.569E-02	1	1.727	.193
BR * ORDER	1.862E-02	1	.425	.517
Error	2.804	64		

Computed using alpha = .05

Notes:

DVBRF= |RAWF – Fraud Base Rate|

DVBAF = |RAWF – Fraud Bayesian Response|. Fraud Bayesian Response is defined as the following:

$$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}, \text{ where } f(\theta) \text{ equals the true probability of fraud and } f(y) \text{ equals the fraud base}$$

rate.

BR – base rate (low vs. high)

ORDER – order in which participants completed experimental cases (auditing case first vs. non-auditing case first)

Table 4
ANOVA Results for Tests of Response Mode (Auditing Cases Only)
(n=68)

Panel A: DVBRF = | Participants' Raw Response- Base Rate |

Source	SS	df	F	p-value
Intercept	3.292	1	43.661	.000
RM	.298	1	3.949	.051
ORDER	.120	1	1.592	.212
RM * ORDER	3.005E-02	1	.398	.530
Error	4.828	64		

Panel B: DVBAF = | Participants' Raw Response- Bayesian Response |

Source	SS	df	F	p-value
Intercept	14.635	1	343.088	.000
RM	.130	1	3.049	.086
ORDER	8.341E-02	1	1.955	.167
RM * ORDER	6.387E-03	1	.150	.700
Error	2.730	64		

Computed using alpha = .05

Notes:

DVBRF= |RAWF – Fraud Base Rate|

DVBAF = |RAWF – Fraud Bayesian Response|. Fraud Bayesian Response is defined as the following:

$$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}, \text{ where } f(\theta) \text{ equals the true probability of fraud and } f(y) \text{ equals the fraud base}$$

rate

RM – response mode (probability vs. frequency)

ORDER – order in which participants completed experimental cases (auditing case first vs. non-auditing case first)

Table 4 (continued)
ANOVA Results for Tests of Response Mode (Non-Auditing Cases Only)
(n=68)

Panel C: DVBRC = | Participants' Raw Response- Base Rate |

Source	SS	df	F	p-value
Intercept	16.244	1	137.800	.000
RM	1.922	1	16.304	.000
ORDER	.417	1	3.538	.065
RM * ORDER	.314	1	2.662	.108
Error	7.545	64		

Panel D: DVBAC = | Participants' Raw Response- Bayesian Response |

Source	SS	df	F	p-value
Intercept	13.880	1	141.776	.000
RM	1.612	1	16.461	.000
ORDER	.284	1	2.899	.093
RM * ORDER	.244	1	2.497	.119
Error	6.266	64		

Computed using alpha = .05

Notes:

DVBRC = |RAWC – Cancer Base Rate|

DVBAC = |RAWC – Cancer Bayesian Response|. Cancer Bayesian Response is defined as the following:

$$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}, \text{ where } f(\theta) \text{ equals the true probability of breast cancer and } f(y) \text{ equals the}$$

cancer base rate

RM – response mode (probability vs. frequency)

ORDER – order in which participants completed experimental cases (auditing case first vs. non-auditing case first)

Table 5
Repeated Measures ANOVA Results
(n=68)

Panel A: DVBRF, DVBRC = | Participants' Raw Response- Base Rate |

Tests of Between-Subjects Effects

Source	SS	df	F	p - value
Intercept	17.225	1	160.275	.000
RM	1.891	1	17.597	.000
ORDER	.512	1	4.766	.033
RM * ORDER	4.968E-02	1	.462	.499
Error	6.771	63		

Tests of Within-Subjects Effects

Source	SS	df	F	p - value
CASE	2.374	1	27.439	.000
CASE * RM	.294	1	3.401	.070
CASE * ORDER	2.617E-02	1	.303	.584
CASE * RM * ORDER	.285	1	3.299	.074
Error(CASE)	5.450	63		

Computed using alpha = .05

Notes:

DVBRF= |RAWF – Fraud Base Rate|.

DVBRC = |RAWC – Cancer Base Rate|

RM – response mode (probability vs. frequency)

ORDER – order in which participants completed experimental cases (auditing case first vs. non-auditing case first)

CASE – case context (auditing vs. non-auditing)

Table 5
(continued)

Panel B: DVBAF, DVBAC = |Participants' Raw Response- Bayesian Response|

Tests of Between-Subjects Effects

Source	SS	df	F	p -value
Intercept	27.970	1	388.020	.000
RM	.378	1	5.241	.025
ORDER	2.177E-02	1	.302	.585
RM * ORDER	8.849E-02	1	1.228	.272
Error	4.541	63		

Tests of Within-Subjects Effects

Source	SS	df	F	p -value
CASE	1.398E-03	1	.020	.888
CASE * RM	1.323	1	18.865	.000
CASE * ORDER	.339	1	4.839	.032
CASE * RM * ORDER	.143	1	2.046	.158
Error(CASE)	4.418	63		

Computed using alpha = .05

Notes:

DVBAF = |RAWF – Fraud Bayesian Response|. Fraud Bayesian Response is defined as the following:

$$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}, \text{ where } f(\theta) \text{ equals the true probability of fraud and } f(y) \text{ equals the fraud base}$$

rate

DVBAC = |RAWC – Cancer Bayesian Response|. Cancer Bayesian Response is defined as the following:

$$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}, \text{ where } f(\theta) \text{ equals the true probability of breast cancer and } f(y) \text{ equals the}$$

cancer base rate

RM – response mode (probability vs. frequency)

ORDER – order in which participants completed experimental cases (auditing case first vs. non-auditing case first)

CASE – case context (auditing vs. non-auditing)

Table 6
Descriptive Statistics by Response Mode

Dependent Variable	RM	N	Mean	Standard Deviation
DVBRF	Probability	34	.2798	.3348
	Frequency	34	.1544	.1967
DVBAF	Probability	34	.4247	.2312
	Frequency	34	.5094	.1789
DVBRF÷Fraud Base Rate	Probability	34	2.2841	3.3922
	Frequency	34	1.3241	2.2476
DVBAF ÷ Fraud Bayesian Response	Probability	34	.5880	.3195
	Frequency	34	.6733	.2140
DVBRC	Probability	34	.6597	.3165
	Frequency	34	.3186	.3873
DVBAF	Probability	34	.6091	.2834
	Frequency	34	.2963	.3534
DVBRC÷ Cancer Base Rate	Probability	34	65.9673	31.6525
	Frequency	34	31.8571	38.7299
DVBAF÷ Cancer Bayesian Response	Probability	34	7.9306	3.6896
	Frequency	34	3.8574	4.6019

Notes:

DVBRF= |RAWF – Fraud Base Rate|

DVBRC = |RAWC – Cancer Base Rate|

DVBAF = |RAWF – Fraud Bayesian Response|. Fraud Bayesian Response is defined as the following:

$$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}, \text{ where } f(\theta) \text{ equals the true probability of fraud and } f(y) \text{ equals the fraud base$$

rate

DVBAC = |RAWC – Cancer Bayesian Response|. Cancer Bayesian Response is defined as the following:

$$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}, \text{ where } f(\theta) \text{ equals the true probability of breast cancer and } f(y) \text{ equals the$$

cancer base rate

DVBRF÷.25 for high base rate conditions, and DVBRF÷.083 for low base rate conditions

DVBRC÷.01

DVBAF÷.87 for high base rate conditions, and DVBAF÷.65 for low base rate conditions. Also, see footnote 9 in the text

DVBAC÷.08. Also, see footnote 9 in the text.

Table 7
Wilcoxon Signed Ranks Test¹

Dependent Variables	No. of Times where Fraud > Cancer	No. of Times where Fraud < Cancer	z (p-value)
DVBRF & DVBRC	23	44	-4.142 (.000)
DVBAF & DVBAC	34	33	-0.206 (.837)
DVBRF÷Fraud Base Rate & DVBRC÷Cancer Base Rate	13	54	-6.447 (.000)
DVBAF÷Fraud Base Rate & DVBAC÷Cancer Base Rate	13	54	-5.847 (.000)

¹ No ties were reported.

Notes:

DVBRF = |RAWF – Fraud Base Rate|

DVBRC = |RAWC – Cancer Base Rate|

DVBAF = |RAWF – Fraud Bayesian Response|. Fraud Bayesian Response is defined as the following:

$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}$, where $f(\theta)$ equals the true probability of fraud and $f(y)$ equals the fraud base rate

DVBAC = |RAWC – Cancer Bayesian Response|. Cancer Bayesian Response is defined as the following:

$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}$, where $f(\theta)$ equals the true probability of breast cancer and $f(y)$ equals the cancer base rate

cancer base rate

DVBRF÷.25 for high base rate conditions, and DVBRF÷.083 for low base rate conditions

DVBRC÷.01

DVBAF÷.87 for high base rate conditions, and DVBAF÷.65 for low base rate conditions. Also, see footnote 8 in the text

DVBAC÷.08. Also, see footnote 8 in the text.

Appendix

Management Fraud Case – Probability Response Format

A public accounting firm is assembling a database of management descriptions as a part of a study to determine whether management descriptions are helpful in detecting management fraud. Each is an accurate description of the management of an actual company.

The accounting firm has documented that of 12 descriptions documented so far, 3 involve a company whose managers were found to have been involved in fraudulent activities of a nature sufficient to affect the opinion on their financial statements for that year. The other 9 descriptions are of companies whose managers were not involved in fraudulent activities that year.

The accounting firm subsequently completed the database of all its clients. The accounting firm has determined that the original 12 clients are representative of the accounting firm's entire client base, i.e. the rate of management fraud in the firm's client base is 25% (3/12). However, the firm estimates that in 5% of cases the descriptions in the database will identify the client management as fraudulent when it is, in fact, honest.

A client description below was selected at random from the database. The database indicates this client management may be involved in fraud.

This company is a new retail client. Its management has undergone extensive changes in the past year. A new president was selected by the board of directors to replace the former president and founder of the firm, who retired. The new president is young, aggressive, and growth oriented. His experience in the retail trade is relatively limited (3 years with another firm) but sufficient for him to have acquired a reputation as a risk taker. He has been given a free hand by the board of directors which wishes to change the rather conservative image of the firm. The new president has made a number of key personnel changes (including controller and treasurer), replacing long-time employees of the firm with hand-picked individuals, most of whom were with him at his previous employer. The new president arranged an incentive-type compensation package from the board that is heavily dependent on reported firm profits. Økokrim has begun an investigation of the new president's former employer, although he has not been implicated in any wrongdoing at the present time.

What is your assessment of the probability that the management described above was engaged in fraudulent activity? _____%

^a Norwegian organization similar to the SEC in the U.S.

Management Fraud Case – Frequency Response Format

A public accounting firm is assembling a database of management descriptions as a part of a study to determine whether management descriptions are helpful in detecting management fraud. Each is an accurate description of the management of an actual company.

The accounting firm has documented that of 12 descriptions documented so far, 3 involve a company whose managers were found to have been involved in fraudulent activities of a nature sufficient to affect the opinion on their financial statements for that year. The other 9 descriptions are of companies whose managers were not involved in fraudulent activities that year.

The accounting firm subsequently completed the database of all its clients. The accounting firm has determined that the original 12 clients are representative of the accounting firm's entire client base. However, the firm estimates that out of each 100 descriptions of honest management, the database will identify 5 as fraudulent.

A client description below was selected at random from the database. The database indicates this client management may be involved in fraud.

This company is a new retail client. Its management has undergone extensive changes in the past year. A new president was selected by the board of directors to replace the former president and founder of the firm, who retired. The new president is young, aggressive, and growth oriented. His experience in the industry is relatively limited (3 years with another firm) but sufficient for him to have acquired a reputation as a risk taker. He has been given a free hand by the board of directors which wishes to change the rather conservative image of the firm. The new president has made a number of key personnel changes (including controller and treasurer), replacing long-time employees of the firm with hand-picked individuals, most of whom were with him at his previous employer. The new president arranged an incentive-type compensation package from the board that is heavily dependent on reported firm profits. Økokrim has begun an investigation of the new president's former employer, although he has not been implicated in any wrongdoing at the present time.

In your opinion, of the 10 randomly selected clients fitting the above management description, how many out of 10 were engaged in fraudulent activity?

_____ of 10

Breast Cancer Case – Probability Response Format

The prevalence of breast cancer among women at age forty who participate in routine screening is 1%. If a woman has breast cancer, the probability that a mammography is positive is 79%. If a woman does not have breast cancer, the probability is 9.6% that she will also get a positive mammography. A woman in this age group had a positive mammography during routine screening.

What is the probability that she actually has breast cancer? _____%

Breast Cancer Case – Frequency Response Format

Ten (10) out of every 1000 women at age forty who participate in routine screening has breast cancer. Eight (8) of every 10 women with breast cancer will get a positive mammography. Ninety-five (95) out of every 990 women without breast cancer will also get positive mammography. You have a representative sample of 100 women in this age group who had a positive mammography during routine screening.

How many of these women out of 100 do you expect to actually have breast cancer?
____ out of 100