

A Meta-Analysis Of Empirical Materiality Studies

David E. Vance, Rutgers University School of Business Camden, USA

ABSTRACT

This paper builds on the work of Messier, Martinov-Bennie and Eilifsen (2005) “A Review and Integration of Empirical Research on Materiality: Two Decades Later.” That paper is an excellent qualitative summary of empirical materiality studies. However, it makes no attempt at quantitative integration.

This paper uses meta-analysis to quantitatively integrate the findings of 42 studies reporting 14,033 materiality decisions by 6,762 individuals. Materiality is broadly defined as an amount that would make a user change his or her mind about a company. Since this definition is subjective, it is no surprise that individual materiality decisions are widely dispersed. What might be surprising is that the mean threshold of materiality over 14,033 decisions was 8.52% percent of net income. The factors which contribute to variability include the interest group making the materiality decision, the accounting issue in play, study method, and the amount of detail provided to the decider.

The contributions of this study are to synthesize prior research quantitatively, to describe the application of meta-analytical techniques to accounting research, and raise questions for further research which might reduce the uncertainty in applying the concept of materiality in the future.

Keywords: Materiality; Meta-analysis; Empirical; Subjective; Opinion

INTRODUCTION

This paper builds on the work of Messier, Martinov-Bennie and Eilifsen (2005) “A Review and Integration of Empirical Research on Materiality: Two Decades Later.” That paper is an excellent qualitative summary of empirical materiality studies. However, it makes no attempt at quantitative integration. This paper uses meta-analysis to quantitatively integrate the findings of 42 studies reporting 14,033 materiality decisions by 6,762 individuals. The common denominator in these studies is that the threshold of materiality measured as a percentage of net income.

There is a large body of literature on the concept of materiality, but little literature on how to operationalize materiality so that it can be applied in a predictable manner. The fairness of financial statements is qualified by the phrase “in all material respects” (AICPA, 1989). Materiality is an amount sufficient to change an investor or creditor’s opinion about a company (FASB 1980 §132). The difficulty is that materiality is defined subjectively

The question of what is material becomes more complex when one considers variables that impact subjective judgments. Different interest groups have different opinions as to what is material. Managers, controllers and board members have mean threshold of materiality of 7.00% of net income, as whereas auditors have a mean materiality threshold of 9.18%. The threshold of materiality also varies by accounting issue from a low of 4.32% of net income for bribes to more three and a half that much for leases. Other variables which impact measured materiality include study method, whether materiality relates to booking or disclosing an item and when the materiality decision was made.

Meta-analysis is a study of studies (Gieles, 1999; Glass, 1976). It is not a single technique, but a collection of techniques to integrate study results quantitatively rather than descriptively. Meta-analysis can be used to identify central tendency, variability, moderating variables and it can sometimes be used predicatively.

Studies included in this meta-analysis report thresholds of materiality ranging from 0.44% to 54.64% with an overall mean of 8.52%. About 9,355 materiality decisions (66.7%) reported materiality thresholds above the much discussed “rule of thumb” of 5%.

The contributions of this study are to (i) quantitatively integrate a number of studies so the broader landscape of materiality comes into sharper relief, (ii) explore the application of meta-analysis to accounting research, and (iii) lay the foundation for research questions as to why the threshold of materiality varies so widely.

The remainder of the paper is organized as follows: (i) the literature review briefly discusses the literature on materiality as well as the literature of meta-analysis, (ii) data and descriptive statistics describes the sources of data for the meta-analysis and provides its overall characteristics, (iii) empirical analysis describes how meta-analysis techniques were applied, (iv) analysis of findings raises questions for future research and (v) the conclusion. Appendix A to this study is a more detailed discussion of estimating confidence intervals using meta-analysis. Appendix B is a more detailed discussion of how inference is applied in meta-analysis.

LITERATURE REVIEW

Several exhaustive studies discuss the policy, issues and philosophy of materiality (Messier, Martiniov-Bennie and Eilfsen 2005, Iskandar and Islin 1999). The common denominator of the studies considered here is that each contains information that can be used to measure materiality as a percent of net income. As used in this study, the percent of net income means the absolute value of an item divided by the absolute value of net income. Studies that did not contain such information were excluded. For example, Estes and Reames (1988) measured the materiality of obsolete inventory in dollars; however, the published facts did not include net income so materiality as a percent of net income could not be computed. Gleason and Mills (2002) was excluded because it estimated net income based on the average return on assets for sampled companies. Other studies such as Cho, Hagerman, Nabar and Patterson (2003) were excluded because they attempted to measure materiality in terms of stock price movements which raises the question of whether it is possible to control for all the variables acting on the market.

There are two advantages to using the magnitude of an item as percentage of net income as a measure of materiality. The first is that many studies cite this measure as the prime factor used by auditors and others to assess whether an item is material (Chewning, Wheeler and Chan 1998; Carpenter, Dirsmith and Gupta 1994; Carpenter and Dirsmith 1992; Icerman and Hillison 1991; Chewning, Pany and Wheeler 1989; Jennings, Kneer and Reckers 1987; Morris, Nichols and Pattillo 1984; Krogstad, Ettenson and Shanteau 1984; Messier 1983; Messier 1981; Moriarity and Barron 1979; Moriarity and Barron 1976; Pattillo 1976; Dyer 1975; Frishkoff 1970; and Woolsey 1954A, 1954B). The second advantage is that this measure becomes more sensitive as a company’s net income approaches zero. This sensitivity means it is more likely to flag items which change the direction of a company’s net income from positive to negative or negative to positive.

The literature of meta-analysis has primarily grown out of the medical and social sciences which have a rich tradition of pioneering new statistical methods. Meta-analysis is an analysis of analyses for the purpose of integrating and interpreting findings. It connotes a rigorous alternative to narrative discussions to make sense out of rapidly expanding research literature. A key assumption of meta-analysis is that each study provides a differing estimate of the relationships between variables in the underlying studies. By summarizing findings across studies one can gain a more accurate representation than can be provided by individual studies (Lyons 1997). The logic is that increased sample size results in statistics that more closely approximate the parameters of the underlying population and by using multiple studies methodological bias that may have crept into any one study is attenuated.

A key question in meta-analysis is whether to include studies of poor quality. In medicine, poor quality can mean the failure to include a control group. But more generally, poor quality means the failure of studies to report statistical measures such as standard deviation, and the results of t and F tests. Quality is not determined by whether

a study is reported in a leading journal. In fact, meta-analysis literature warns against excluding unpublished research. The reason is that research reporting negative relationships is less likely to be published. This is called the “file drawer” effect. Systematically excluding unpublished studies or studies from obscure journals would introduce bias to the analysis (Turner, 2008). The issue of how to deal with studies of varying quality is not fully resolved and represents one of the major challenges for meta-analysis (Whitlock, 2005). Few of the studies in this meta-analysis included measures of variability such as standard deviation, or the results of t and F tests. To overcome these difficulties, first principals were used to develop alternative meta-analytical techniques to test statistical significance. See Appendix A Confidence Intervals.

Hunter, Schmidt and Jackson (1982) and Hunter Schmidt (1990) have developed an expansive theory of meta-analysis including means for correcting sampling error, measurement error and range restriction. Sampling error is error due to sample size. Statistics from small samples tend to depart more widely from population parameters than do large samples. Measurement error is the difference between the reported value of data and the actual value. For example land measured with a yardstick might have more measurement error than land measured with a laser range finder. Range departure is due to the random deviation in the relationship between two variables because the data are truncated in some way. One problem with these last two error corrections is that the data to perform them is not always given in the underlying studies (Lyons, 1997).

While Hunter Schmidt and Jackson (1982) is elegant, this approach has its critics. A significant criticism is that methods which correct for sampling, measurement and range restriction errors also mask differences caused by moderating variables or introduce spurious statistical artifacts. This makes interpretation of meta-analyses difficult (Russell, 1995). Hunter Schmidt (1990) has also been criticized because it leads to substantially different results than do other accepted meta-analysis protocols such as Hedges and Olkin (1985), Rosenthal (1991) and Rosenthal and Rubin (1978) according to research published by Blair (1995). The lesson is to apply meta-analysis carefully and not to stretch it beyond what it can do.

The first and most basic step in applying meta-analysis is to find or create comparability among reported findings. If the explanatory variable is continuous, Hunter Schmidt and its progeny often convert study findings into a dimensionless parameter, r , which is the correlation coefficient for two groups of variables (Whitlock, 2005). If the explanatory variable is dichotomous, data in the underlying studies are converted to a Cohen d (DeCoster, 2004). Cohen’s d is the difference of means of study variables standardized by dividing them by the pooled standard deviation (Whitlock, 2005). The advantage to using r or Cohen’s d is that the relationship between the measured and explanatory variable in each study is normalized with respect to the variability of the underlying data in each study. A limitation to these techniques is that they rely on the underlying studies to report the t-statistic, F-statistic, standard deviation or z value. Since these measures of variability are not available for most of the studies included in this analysis, a return to more basic meta-analytical techniques is required.

DATA AND DESCRIPTIVE STATISTICS

Data Sources

The data in this study were gleaned from 42 empirical studies of 6,762 individuals who made 14,033 materiality decisions in a wide variety of situations. See Table 1 Studies Included in Meta-Analysis. The subject studies were all conducted in the United States. Empirical studies of materiality conducted in Australia (Martinov and Roebuck 1998), Denmark (Hojskov 1998), Finland, (Adams, Weetman and Gray 1993), Canada (Robinson and Fertuck 1985), the United Kingdom (Firth 1979) and a study examining the reconciliation of international accounting standards to US GAAP by non-US companies (Street, Nichols and Gray 2000) were excluded to avoid the risk that materiality judgments are culturally sensitive.

Table 1 Studies Included in Meta Analysis

C is the number of cohorts in each study. A cohort is the most granular data reported in a study. Each cohort reports the mean materiality threshold for a specific interest group evaluating a particular accounting issue. P is the number of unique individuals participating in each study. The number of materiality decisions in each study is n. The study method is E for questionnaire, case study or some other kind of experiment and H for analysis of historical, documents such as work papers or audited financial statements. Frishkoff-Stringer is a synthesis of Frishkoff (1970) and Stringer (1970) two articles based on one data set.

Study	C	P	n	Participant Description	Study Method	Issues
Baskin, 1972	1	69	69	Investors	E	Various Issues
Bates, 1982	3	67	67	Auditors	E	Litigation
Bernardi, 1996	9	152	152	Auditors	E	Inventory
Boatsman, 1974	2	33	990	Analysts & Auditors	E	Various Issues
Burgstahler, 2000	13	61	183	Auditors	E	Various Issues
Chewning, 1989	7	284	284	Auditors	H	Accounting Changes, Foreign Currency & Absenses
Chewning, 1998	7	139	139	Investors	H	Debt equity swap
Costigan, 1995	1	351	351	Managers	H	Deferred tax liability
Cumming, 1973	20	328	748	Auditors & Managers	E & H	Various Issues
DeZoort, 2003A	1	55	55	Board members	E	Accounts receivable
DeZoort, 2003B	4	262	262	Board members	E	Inventory write-off
DeZoort, 2006	8	160	160	Auditors	E	Accounts receivable Bond premium write off, gains and casualty losse
Dyer, 1975	6	85	245	Auditors	E	Litigation
Fesler, 1989	1	126	126	Controllers	A	
Frishkoff-Stringer 1970	6	180	180	Auditors	A	Consistency of tax treatment
Hatfield, 2007	1	155	155	Auditors	E	Accounts receivable write off Investment in subsidiary write off
Hofstedt, 1977	1	19	19	Professors and students Judges, attorneys & auditors	E	Litigation, gain on sale, inventory and bribes
Jennings, 1987A	12	267	1068	Auditors, bankers, credit managers & analysts	E	Litigation, inventory, gain, loss of product line & bribes
Jennings, 1987B	20	201	1005	Judges	E	Inventory
Jennings, 1991	1	77	77	Judges	E	Stock based compensation & capital leases
Libby, 2005	4	61	61	Auditors	E	Retiree health care costs
Liu, 2002	6	437	437	Managers	H	Various Issues
Messier, 1983	2	29	29	Auditors	E	Depreciation charge
Moriarity, 1976	2	15	15	Auditors	E	General engagement materiality
Moriarity, 1979	20	5	20	Auditors	E	Interest capitalization
Morris, 1984	10	221	221	Auditors	A	Interest capitalization
Morris, 1988	8	334	334	Auditors	A	Reserves & revenue cutoff
Nelson, 2004	8	232	1840	Auditors	E	Reserves, revenue cutoffs contingencies
Nelson, 2005	6	234	681	Auditors	E	Securities valuation & contingencies
Newton, 1977	8	15	15	Auditors Auditors, analysts, bankers, controllers & professors	E	Various Issues
Pattillo, 1975	6	684	684	Professors and students	E	EPS changes
Rose, 1970	4	100	121	Professors and students	E	Loan and inventory write offs and expense accrual
Schneider, 1990	9	170	1494	Managers	E	Inventory write-off, revenue cut-off
Shafer, 2002	2	70	70	Managers	E	Security valuation
Tuttle, 2002	3	72	216	Professors and students	E	

Table 1 cont.

Waters, 1997	2	130	130	Auditors	H	Consistency exception
Wheeler, 1993	27	284	284	Auditors	H	Consistency exception
Woolsey, 1954A	21	69	158	Auditors, analysts, investment bankers, controllers	E	Contingent liabilities, marketable securities & leases
Woolsey, 1954B	21	117	351	Auditors, analysts, investment bankers, controllers, professors & others	E	Bond premiums, casualty loss, gains & casualty losses
Woolsey, 1973	5	176	176	Auditors, analysts, controllers, professors & students	E	Cost of goods sold
Wright, 1983	4	50	100	Auditors	E	Revenue issues
Wright, 1997	2	186	261	Auditors	H	Various Issues
	30	6,76	14,03			
	4	2	3			

Basis of Comparison – A Common Denominator

One of the difficulties in any meta-analysis is that different studies measure and report data in different ways (Whitlock, 2005). To create a common measure of materiality, some of the data in the subject studies had to be adjusted and for other studies reasonable assumptions and estimates had to be made to standardize the data. For example, Rose, Becker and Sorter (1970) measured materiality in terms of sensitivity to changes in earnings per share (EPS). Since EPS is simply net income divided by the number of shares outstanding, and the presumed outstanding shares did not change over the course of the experiment, sensitivity to changes in EPS are directly comparable to sensitivity in changes in net income based on the inclusion or exclusion of items deemed material. Analyses of data in Messier (1983), Hofstedt and Hughes (1977) and Rose, Becker and Sorter (1970) relied on interpretation of graphs. Several studies provided materiality in the form of a percentage of pretax income (DeZoort, Hermanson and Houston 2003; Nelson, Smith and Palmrose 2004; Wright and Wright 1997; Hofstedt and Hughes 1977). The estimated tax rate used to convert pre-tax materiality to after tax materiality was the average corporate tax rate of 3,993 companies on Compustat for the year 2004. Waters (1997) provided information in the form of a regression equation which was used to estimate materiality. Other studies provided materiality information in the form of frequency within materiality brackets that required interpolation (Chewning, Wheeler and Chan 1998; Chewning, Pany and Wheeler 1989; Fesler and Hagler 1989; Morris, Nichols and Pattillo 1984). One study reported the materiality thresholds of six interest groups (Pattillo 1976) but only provided the aggregate number individuals in the study, so results were allocated evenly across the interest groups as recommended by DeCoster (2004). Stringer (1970) reported a more detailed analysis of the data underlying the Frishkoff (1970) study so Stringer’s results were used where possible. The balance of the studies either explicitly provided materiality as a percentage of net income or provided information to compute it.

EMPIRICAL ANALYSIS

Meta analysis of this data may be able to tell us four things (i) the measure of central tendency, (ii) the degree of variability, (iii) whether there are moderating variables and (iv) whether studies contain predictive information.

Cohorts

Many studies include sub-sets of information. One subset is the interest group of the person making the materiality judgment. Another subset is the type of accounting issue involved in the materiality judgment. Other subsets include the type of study, the time period in which the study was conducted and whether the materiality decisions involved booking an item on the one hand, or disclosure and classification of already booked items on the other. To extract the maximum amount of information from each study, it was disaggregated into cohorts. Each cohort contains the materiality judgments of only one interest group, on one accounting issue, one study method and

so forth. Cohort characteristics are analyzed to determine whether they explain the variability in materiality decisions. The data in the 42 studies disaggregated into 304 cohorts each with the attributes discussed above. The number of materiality decisions in a cohort ranged from 1 to 540 with a mean of 46 materiality decisions per cohort and median of 113.

Study Method

When looking across a number of studies, a substantial question arises as to whether there is a “best” method to estimate materiality. Studies tend to split into two types. One type of study asks participants to make materiality decisions in an experimental setting. Experiments include questionnaires, case studies and games. The other method is to infer materiality by examining historical documents such as audited financial statements and audit work papers.

The advantage of the experimental approach is that it allows the investigator to directly ask a participant for a threshold of materiality under a given set of circumstances. The drawback is that experimental situations lack the nuance and real world pressures that influence actual decisions.

The advantage of examining historical information is that one is able to see how individuals make materiality judgments in the light of real world pressures. The drawback is that inference is required to estimate the threshold of materiality. See Appendix B Inference for a discussion of how inference is applied. Cumming (1973), for example, used both an analysis of historical materiality decisions and experimental evidence in the form of questionnaires to analyze the issue of materiality.

A fair question is whether materiality estimates based on historical records are substantially different from materiality estimates based on experiments. Table 2 Analysis of materiality by study method addresses this question. Measurements were based on historical evidence for 3,000 study participants (44.4% of 6,762) representing 3,312 materiality decisions (23.6% of 14,033). The remaining materiality decisions were based on experiments. The mean materiality for historical studies was 11.76% as compared to the mean materiality for experiments of 7.52%. This difference raises the question of whether participants provide answers they think experimenters want to hear rather than what they would do when faced with client pressures. If future research found this to be true, it would limit the usefulness of experimental data or at least alert researchers to use experimental data with caution.

Table 2 Analysis of Materiality Study by Method

Materiality has been studied using two different methods. The first method is experiments. Experiments include questionnaires, case studies and games. Experiments have the benefit of measuring the threshold of materiality directly. The drawback is that experiments lack real world pressures and nuance. The other method of studying materiality is to examine historical documents such as audited financial statements and work papers. The drawback to using historical records is that inference is required to impute the threshold of materiality. The benefit of using historical records is that it reflects what people actually do when confronted with real world pressures.

A cohort is the most granular level at which study data are reported. Each cohort reports the mean materiality of one interest group evaluating a single accounting issue. Participants is the number of individuals who made materiality decisions. The number of materiality decisions is n. The upper and lower bound of the 90% confidence intervals for experiments and historical data along with their means indicate that there is no statistically significant difference between materiality thresholds determined through experiments or by analyzing historical records. Never the less the difference of 4.24% raises the question of whether experiments accurately reflect what people will tend to do under real world circumstances.

Study Type	Cohorts	Participants	n	Materiality				90% Confidence	
				Mean	Q1	Median	Q4	Lower Bound	Upper Bound
Experiments	221	3,762	10,721	7.52%	4.00%	6.74%	9.85%	1.50%	16.46%
	83	3,000	3,312	11.76%	3.62%	10.58%	17.06%	0.81%	26.37%
Historical data									
All Studies	304	6,762	14,033	8.52%	4.00%	6.78%	11.00%	1.47%	21.66%

Boundaries for the 90% confidence interval for experimental and historical data were developed using techniques described in Appendix A Confidence Intervals. Based on these confidence intervals, we find no statistically significant difference between experimental and historical data.

Homogeneity

Before proceeding with the meta-analysis, DeCoster (2004) recommends testing the data across studies for homogeneity. Less homogeneity means more sampling error (Whitlock, 2005). In effect, a test of homogeneity provides some assurance that the study is comparing apples to apples or at least fruit to fruit and not fruit to fruit flies. The homogeneity statistic Q_T is given by equation (1).

$$Q_T = \sum [w_i * (T_i - T)^2] \tag{1}$$

Where w_i is the sample size, n , in each study, T_i is the measured statistic in each study, in this case the materiality threshold, and T is the weighted average materiality threshold of the population. The homogeneity statistic follows the chi-squared distribution with $k-1$ degrees of freedom where k is the number of studies included in the meta-analysis. This analysis was modified slightly to substitute cohorts for studies. T_i then becomes the mean materiality threshold for cohort $_i$, w_i is the number of materiality decisions in cohort $_i$, and T is the weighted average mean materiality of all cohorts.

DeCoster (2004) indicates it typically takes removal of around 20% of studies to achieve homogeneity. However, the Q_T for the 304 cohorts in this study was 52.42 far less than the chi-squared cutoff for 303 degrees of freedom. Therefore, no cohorts had to be dropped. According to Whitlock (2005) this greater homogeneity should reduce sampling error.

Measures of Central Tendency

The equations for computing the mean of the overall data and sub-sets of data differ somewhat from the usual equation for the mean as shown in equation (2). The principal modification is to weigh the cohort means with sample size (DeCoster, 2004, Lyons, 1997).

$$\text{Sub set Mean (SM)} = \frac{\sum n_i \times M_i}{\sum n_i} \tag{2}$$

Where SM is the sub set mean; n_i is the number of materiality decisions made by cohort $_i$, and M_i is the materiality threshold for cohort $_i$. In meta-analysis literature this is called the weighted average effect size (Whitlock, 2005). In this study, equation (2) is used to compute the weighed average effect size of all studies as well as subsets of data such as study type, interest groups, accounting issue and decisions over time.

Another measure of central tendency is the median. For each set or subset of data considered in this study the median was estimated by ranking data from lowest materiality to highest materiality and finding the materiality of the cohort that contained the $n / 2$ observation.

Data Stability

The art form in meta-analysis comes not from application of statistical techniques, but from knowing how far to push them before conclusions become unreliable. Data stability is an important issue when determining how far to trust meta-analytic results. If removal of one or two studies radically changes an outcome, for example if it changes an effect from significant to insignificant, or from a positive correlation to a negative correlation the data is said to be unstable.

There are several methods for determining data stability. Muellerleile (2006) for example advocates interpretation of graphic results as a means of determining whether data are stable. However, simpler methods are often better and easier to interpret.

In this study, data stability was analyzed by comparing the mean for all studies to the mean when combinations of cohorts are removed. Removing the largest cohort and three largest cohorts caused the mean of the remaining data to vary less than half a percent from the overall mean of 8.52%. Removing the cohorts with the highest and lowest effects and the three cohorts with the highest and lowest effects barely changed the overall mean. Removal of the 500 highest and lowest materiality decisions changed the overall mean by less than less than 0.4%. The results of removing various combinations of cohorts is provided in Table 3 Data stability. Since the mean materiality varies little when the largest, highest and lowest cohorts are removed, the data appear stable. Further, the confidence intervals for each subset of data at the 90% level indicate that there is no statistically significant difference among subsets of data and the total set.

Table 3 Data Stability

When removal of one or two sets of data from a meta-analysis changes the results from significant to insignificant or insignificant to significant or reverses the sign of a correlation, the data are said to be unstable. The implication is that results are not trustworthy. There are several methods of measuring data stability. The simplest and most reliable way is simply to remove data sets and see how this affects the results. The table below provides statistics for all data, and when various cohorts of data are removed. A cohort is the most granular level at which data are reported. A cohort includes the mean materiality for a single interest group on a single accounting issue. The data indicate that removal of various groups of data impact the mean only minimally.

	Cohorts	n	-----Materiality-----				90% Confidence	
			Mean	Q1	Median	Q4	Lower Bound	Upper Bound
All study data	304	14,033	8.52%	4.00%	6.78%	11.00%	1.47%	21.66%
Data Removed	Statistics for remaining data							
Largest cohort	303	13,493	8.70%	4.69%	6.81%	12.55%	1.47%	21.66%
Three largest cohorts	301	12,692	8.38%	4.83%	6.81%	11.00%	1.33%	19.69%
Cohorts with the highest and lowest mean materiality	302	13,978	8.51%	4.00%	6.78%	11.00%	1.47%	21.25%
Three cohorts with the highest mean materiality and the three cohorts with the lowest mean materiality	298	13,763	8.54%	4.10%	6.78%	11.00%	1.55%	21.59%
500 of the highest reported materiality observations and 500 of the lowest reported materiality observations.	NA	13,033	8.16%	4.21%	6.78%	10.50%	2.00%	17.50%

Variability

The degree of variability in underlying study data is often of interest to readers of a meta-analysis. Where the standard deviation or variance is provided for each study, a pooled standard deviation can be computed. Where variance or standard deviations are absent, t-statistic, F-statistic or z values can be used to compute r or Cohen’s d. These standardized measures can then be used to estimate variability. For most of the studies included in this meta-analysis, standard measures of variability were not reported.

Confidence intervals are used to determine whether samples are drawn from the same population. When populations are normally distributed and sample size is greater than about 30, the Central Limit Theorem can be invoked and standard deviation is the usual way to compute confidence intervals. Since standard deviation was not included in underlying study data, confidence intervals had to be computed using the alternative method described in Appendix A. One drawback of this alternative method is that it requires more data than the standard deviation. For example at the 90% confidence interval, a minimum of 80 observations are recommended, at the 95% confidence interval, a minimum of 160 observations are recommended.

ANALYSIS OF RESULTS

Materiality is defined subjectively. As a result, any two people or any two groups of people might set the threshold of materiality at a different level and that level may change based on circumstances. As a consequence, the studies underlying this meta-analysis measured materiality thresholds ranging from 0.44% to 54.64%. This high level of variability raises a number of issues.

- (i) Is there a difference between materiality thresholds in an experimental setting and what people actually do?
- (ii) Do different interest groups have different materiality thresholds?
- (iii) Is the threshold of materiality different based on accounting issue?
- (iv) Are individuals more sensitive to items which, if booked will change net income, or are they more sensitive to proper disclosure and classification of items even though disclosure or classification does not change net income?
- (v) Do materiality thresholds change over time?

Opinions vs. Actions

Experimental methods of estimating the threshold of materiality, such as questionnaires, case studies or games, are really opinion studies. They lack consequence. Studies based on historical documents such as financial statements or audit work papers reflect actions taken as the result of real world pressures and in the face of real world consequences. Table 2 Analysis of study method shows that there are no statistically significant differences between materiality thresholds of experiments, at 7.52%, and historical actions, at 11.76%. Never the less, most people would probably conclude there is a substantial difference between the two thresholds on a subjective basis. This raises several questions for further research such as whether there is any way to make experiments more realistic, how experimenters might adjust findings so that they are more predictive of what people will actually do, and whether actual materiality decisions from financial statements and audit work papers can be measured on a more continuous basis.

Materiality by Interest Group

While auditors are hired by the firms they audit, their real audience is financial statement users. When auditors make assurances that statements fairly represent the financial condition of a company, they undertake a duty to users. Failure to use due care in discharge of that duty can give rise to negligence (Prosser, 1971). This potential liability makes it important for auditors to meet the materiality expectations of the user community.

Table 4 Materiality by interest group analyzes the materiality threshold of auditors and other interest groups. Mean materiality ranges from a low of 7.00% of net income for managers, controllers and board members to a high of 10.76% for bankers and credit managers. Auditors have a mean materiality of 9.18%. Whereas three groups (i) managers, controllers and board members, (ii) attorneys and judges and (iii) investment bankers and analysts have lower mean materiality thresholds of 7.00%, 7.66% and 7.77% respectively.

Table 4 Materiality by Interest Group

A cohort is the most granular level at which an underlying study reported data. One cohort reports the mean materiality for one interest group on one accounting issue. Participants is the number of individuals who made materiality decisions. The number of materiality decisions is n. The upper and lower bound are the boundaries for the 90% confidence interval for each interest group. While there is no statistically significant difference among interest groups, materiality is defined subjectively and one might expect an observer to say there is an important difference between mean auditor materiality at 9.18% and the interest groups with a lower mean materiality.

Interest Group	Cohorts	Participants	n	Materiality				90% Confidence	
				Mean	Q1	Median	Q4	Lower Bound	Upper Bound
Auditors	191	3,764	7,528	9.18%	4.80%	7.07%	11.00%	2.24%	26.37%
Managers, Controllers & Board members	50	1,882	3,599	7.00%	3.32%	6.73%	8.01%	1.05%	17.38%
Investment Bankers, Analysts & Others	27	303	1,004	7.77%	4.00%	4.90%	12.00%	4.00%	20.00%
Bankers and Credit Managers	11	215	619	10.76%	5.30%	11.00%	17.00%	4.00%	20.00%
Attorneys & Judges	9	223	661	7.66%	4.00%	8.60%	11.00%	2.00%	13.00%
Accounting Professors & Students	16	375	622	9.16%	6.60%	10.00%	10.00%	5.00%	16.46%
Overall	304	6,762	14,033	8.52%	4.00%	6.78%	11.00%	1.47%	21.66%

While there is no statistically significant difference among interest groups, statistical significance might not be the point. The materiality threshold for auditors of 9.18% was based on 7,528 materiality decisions by 3,764 individuals. With samples that large, it is hard to ignore the fact that a substantial number of auditors set materiality thresholds higher than many users think reasonable. A question for future research might be whether this disparity has resulted in litigation against accounting firms and whether that litigation has been successful. Another question for further research might be to ask users whether they are satisfied with auditors' mean materiality of 9.18%. If the answer is yes, it might indicate users think a materiality threshold of 9.18% of net income is fair, despite evidence that they prefer a lower threshold.

While the risk of breaching an implied duty to users is one question, another question for further research is why the variability of materiality thresholds within groups is so high? One indication of intra-group variability is the upper and lower bound for the 90% confidence interval. Spreads between these boundaries range from 11.00% for investment bankers, analysts and others to 24.13% for auditors.

Accounting Issues

Table 5 Panel A is an analysis of materiality by accounting issue. The data show that there is a statistically significant difference in the materiality threshold between (i) bribes and (ii) asset impairment or write-off and some other accounting issues at the 95% confidence level. There is also a significant difference between contingent liabilities and some other accounting issues at the 90% confidence level. See Table 5 Panel B Accounting Issue Statistical Significance.

Table 5 Panel A Materiality by Accounting Issue

A cohort is the most granular level at which an underlying study reported data. One cohort reports the mean materiality for one interest group on one accounting issue. Participants is the number of individuals who made materiality decisions. The number of materiality decisions is n. The upper and lower bound are the boundaries for the 90% and 95% confidence interval for each accounting issue respectively. Caution should be use when interpreting data based on a fewer than six studies or when one study contributes more than 20% of the data points. As the number of studies decreases below six and the number of data points from any one study rises above about 20%, the data begin to lose their meta-analytic character. Specifically, peculiarities in the methodology for any one study become much less attenuated. This does not mean the data are invalid, just that corroborating data should be sought.

	Cohorts	n	----- Materiality -----				90% Confidence		95% Confidence	
			Mean	Q1	Median	Q4	Lower bound	Upper bound	Lower bound	Upper bound
Accounting Changes & General	107	4,295	8.91%	4.00%	6.20%	11.88%	1.50%	26.37%	0.81%	27.37%
Asset Impairment or write off	59	3,902	7.82%	6.43%	6.81%	8.86%	3.97%	12.00%	3.69%	13.00%
Bond premium write off	9	202	13.52%	5.90%	15.32%	19.69%	5.90%	22.15%	5.90%	28.31%
Bribes	7	468	4.32%	3.00%	3.00%	6.00%	2.00%	9.00%	2.00%	9.00%
Casualty loss	9	193	11.64%	5.83%	12.77%	13.54%	5.83%	21.08%	5.83%	21.08%
Contingent liability (i)	11	91	5.90%	3.91%	4.86%	7.54%	2.32%	9.38%	(iv)	(iv)
Expense	50	2,178	7.87%	1.47%	4.73%	12.11%	0.97%	22.70%	0.73%	25.00%
Litigation (ii)	11	661	12.55%	7.00%	10.00%	21.66%	4.00%	22.33%	4.00%	40.56%
Leases (iii)	8	81	15.20%	10.15%	14.46%	18.92%	5.40%	27.08%	(iv)	(iv)
Revenue or gain treatment	33	1,962	8.42%	3.62%	8.57%	9.00%	1.75%	16.77%	1.75%	17.00%
All Accounting Issues	304	14,033	8.52%	4.00%	6.78%	11.00%	1.47%	21.66%	1.08%	26.37%

Notes:

(i) Contingent liability data was based on 3 studies of which one study contributed 41% of the data.

(ii) Litigation data was based on 4 studies of which one study contributed 79% of the data.

(iii) Lease data was based on 4 studies of which one study contributed 40% of the data.

(iv) Insufficient data to generate 95% confidence interval bounds.

Table 5 Panel B Statistical Significance by Accounting Issue

There is a statistically significant difference between six pairs of accounting issues at the 95% confidence level and a statistically significant difference between five pairs of accounting issues at the 90% confidence level.

Issue	Statistically Significantly Different Issues
Asset Impairment or Write-off	Bond Premium Write-off ** Litigation* Leases**
Bribes	Bond Premium Write-off** Casualty Loss** Litigation** Leases**
Contingent Liability	Bond Premium Write-off* Casualty Loss* Litigation* Leases*

* Statistically significant at the 0.1 level

** Statistically significant at the 0.05 level

Bribes are an illegal act which makes them different in kind from other materiality issues. The question is what drives the other differences?

One possibility is that the more straight forward the issue, or the more straight forward the impact on net income, the lower the threshold of materiality. Supporting that possibility is the fact that the threshold for expenses and revenue or gain treatment are less than the materiality threshold for more complex accounting issues like casualty loss, litigation, bond premium write off and leases.

Areas for future research might be to confirm or refute this possibility and to determine whether better explanation and disclosure of complex accounting issues, or more education would reduce variability.

Materiality of Items Booked versus Disclosed

The concept of materiality is applied both to booking and item on the one hand, and how to classify or disclose an item on the other hand. For example, an audit may find \$10,000 of unrecorded phone bills. The question is whether this is material enough to book. If booked, it will directly impact net income. On the other hand, there are some items which have already impacted net income. The question is either how to classify these items or whether to disclose them. A bribe run through accounts payable has already impacted net income. The question is whether it should be disclosed. Retiree health benefits may have already been paid, the question is whether these costs should be disclosed (Morris, 1984).

Classification and treatment of an item often turns on whether it is material. A \$500 gain on the sale of a truck may not warrant special disclosure whereas a \$50 million gain on the sale of an unused factory might be classified as an extraordinary item. The issue with these items is not their impact on net income, but how they are classified and disclosed.

Table 6 Materiality in Booking or Disclosure analyzes whether people tend to set the threshold of materiality lower if it will impact net income than if it is simply a disclosure item.

The mean threshold of materiality for booking an item that will impact net income is 9.05% and the threshold for disclosure of an item which has already impacted net income is 9.02%. There is virtually no difference between them. The more interesting finding concerns studies that asked about materiality in a theoretical sense and in which it was unclear whether an item would impact net income or simply be a disclosure or classification item. The threshold of materiality for these uncategorized decisions was only 6.23%. About 87% of uncategorized decisions were based on experiments which indicates that people tend to be more conservative in setting theoretical

materiality thresholds than when confronted with specific facts. This is consistent with the finding that materiality thresholds based on historical evidence are higher than materiality thresholds developed through experiments.

Table 6 Materiality in Booking or Disclosure

A cohort is the most granular level at which an underlying study reported data. A cohort reports the mean materiality for one interest group on one accounting issue. Participants is the number of individuals who made materiality decisions. The number of materiality decisions is n. The upper and lower bound are the boundaries for the 90% confidence interval. While the differences between groups is not statistically significant, it is striking that the mean materiality threshold for general theoretical questions is much lower than for items which involve more concrete fact patterns involving whether to book or to disclose or classify an item.

	Cohorts	n	-----Materiality -----				90% Confidence	
			Mean	Q1	Median	Q4	Lower bound	Upper bound
Materiality involving booking and entry and affecting net income.	134	7,625	9.05%	5.40%	6.84%	11.50%	1.33%	26.37%
Materiality involving disclosure or classification. No net income effect.	110	3,806	9.02%	3.91%	8.57%	14.00%	1.05%	20.00%
General materiality questions such as planning materiality or decisions that could not be categorized as book or disclose/classify decisions.	60	2,602	6.23%	4.00%	4.90%	6.85%	4.00%	11.88%
All Materiality Decisions	304	14,033	8.52%	4.00%	6.78%	11.00%	1.47%	21.66%

Materiality Over Time

The studies included in this analysis were produced over the period 1954 to 2007. A fair question would be whether the threshold of materiality has gotten higher as the result of a growing push for profits or whether the threshold of materiality has gotten lower as lessons from business failures have accumulated.

Table 7 Materiality over time analyzes materiality by decade. No studies were found from the 1960s. Overall, there seems to be a trend toward setting more conservative materiality thresholds. In fact there is a statistically significant difference at the 90% confidence level between 2000s studies, which had the lowest mean materiality at 6.57%, compared to the 1950s and 1980s which had mean materialities of 15.12% and 10.76% respectively.

On the other hand, it would inappropriate to conclude there is a secular decline in materiality thresholds. The 1970s which accounted for about 25.1% of all materiality decisions had the second lowest mean materiality over the six decades for which data was available. It could be that the subjective nature of materiality means that it is responsive to the ebb and flow of political, social and economic circumstances. Whether there is a correlation between these factors and mean materiality offers another question for future research.

This meta-analysis did not analyze materiality thresholds pre and post the collapse of Enron, a seminal event in corporate governance. Nor did it attempt to assess the impact of Sarbanes Oxley on materiality threshold. There were just not enough studies which used post Enron, post Sarbanes-Oxley data to perform such a meta-analysis. A question for future research might be whether Enron and Sarbanes-Oxley will accelerate the trend toward more conservative accounting practices in the form of lower materiality thresholds or whether materiality thresholds will begin to creep upward over time.

Table 7 Materiality Over Time

A cohort is the most granular level at which an underlying study reported data. One cohort reports the mean materiality for one interest group on one accounting issue. Participants is the number of individuals who made materiality decisions. The number of materiality decisions is n. The upper and lower bound are the boundaries for the 90% confidence interval for materiality decisions. There appears to be a downward trend in the mean materiality. However, data from 1970s studies could indicate that subjective views of materiality are cyclical in nature. There is a statistically significant difference between studies published since 2000 as compared to studies published in the 1950s and the 1980s. However, the statistical difference between 2000s and 1980s studies is slight. There were no studies from the 1960s.

	Cohorts	Participants	n	----- Materiality -----				90% Confidence	
				Mean	Q1	Median	Q4	Lower Bound	Upper Bound
1950s studies	42	186	509	15.12%	11.54%	15.54%	18.92%	6.77%	23.62%
1960s studies	None								
1970s studies	75	1,880	3,528	7.42%	4.00%	5.30%	7.50%	2.75%	17.06%
1980s studies	67	1,579	3,234	10.76%	4.86%	11.00%	14.00%	1.55%	21.66%
1990s studies	64	1,318	2,642	9.12%	4.64%	6.84%	10.71%	1.08%	26.37%
2000s studies	56	1,799	4,120	6.57%	4.06%	6.55%	8.60%	1.67%	10.72%
All Studies	304	6,762	14,033	8.52%	4.00%	6.78%	11.00%	1.47%	21.66%

Overall mean, and median materiality thresholds in this meta-analysis are 8.52% and 6.78% respectively. Both are far above oft discussed rules of thumb for materiality planning of 3% to 5% (Turner 2000, Levitt 1998). This raises the question of whether auditors’ materiality decisions meet user expectations as to the fairness of financial statements. Other questions for future research are whether the current flexibility in setting materiality thresholds serves the public interest and how markets would react if materiality thresholds were disclosed on a company by company, and audit by audit basis. While this study outlines the broad landscape of materiality decisions, there is a vast and undiscovered country of materiality knowledge awaiting the intrepid researcher.

CONCLUSION

Meta-analysis as a robust collection of tools for synthesizing the results of studies in a quantitative manner. Meta-analysis can provide measures of central tendency such as mean and median, and variability including confidence intervals. It can also identify moderating variables and trends.

Not every meta-analysis technique can be used on every set of studies. Techniques should not be applied in a mechanical manner, but should be applied with an understanding of the limits of each technique. For example, results improve when at least six studies are used and no more than 20% of data points come from a single study. If fewer than six studies are used or more than 20% of data are from a single study, it is difficult for meta-analysis to attenuate the methodological peculiarities of any one study. Confidence intervals can be estimated by returning to first principals, and recalling that at a 90% confidence each tail of a two-tailed distribution should contain about 5% of the data points in any subset of data. A limitation on this technique is that it requires much more data than the standard deviation.

There is a high degree of variability in materiality judgments ranging from a low of 0.44% net income to a high of 54.65% of net income. While this meta-analysis was unable to account for all of the variability, some moderating variables emerged. Studies based on experiments such as questionnaires, case studies and games tended to produce lower materiality thresholds than studies based on work papers or audited financial statements. Materiality varies among interest groups. Managers, controllers and board members have the lowest mean materiality and bankers and creditors have the highest. There are statistically significant differences among some accounting issues. Bribes have the lowest mean materiality and leases have the highest. There is virtually no

difference in materiality threshold between items which directly impact net income items which have already impacted net income so the materiality issues involve disclosure or classification. There also appears to be a trend over the last six decades toward tightening materiality standards. Whether this trend continues, accelerates or reverses is left for a future study.

AUTHOR INFORMATION

David E. Vance is an MBA, CPA, attorney and former Naval Officer. He has been corporate controller, Chief Financial Officer and trial attorney. Prior to joining Rutgers University he was in corporate restructuring. He is currently an Assistant Professor at Rutgers University School of Business Camden where he teaches in the MBA program. Recent journal articles include “Capital for Asset Based Lenders,” “Evidence Companies are Not Analyzing Goodwill for Impairment” and “Return on Goodwill.” He has published four books the most recent of which are *Ratios for Analysis Control and Profit Planning*, Global Professional Publishing and *Corporate Restructuring*, Springer. E-mail: dave.vance1@verizon.net

REFERENCES

1. Adams, Carol A., Pauline Weetman and Sidney J. Gray. 1993. “Reconciling national with international accounting standards,” *European Accounting Review*, Dec., Vol. 2 Issue 3, p471-494, 24p
2. American Institute of Certified Public Accountants. 1989. *Reports on Audited Financial Statements*. AICPA: New York. SAS58, AU Section 508, p.814.
3. Baskin, Elba F. 1972. “The Communicative Effectiveness of Consistency Exceptions,” *The Accounting Review*. Jan. pp.38-51.
4. Bates, H. L., R.W. Ingram, and P. M. J. Reckers. 1982. “Auditor-Client Affiliation: The Impact on Materiality.” *Journal of Accountancy*. April, pp.60-63.
5. Bernardi, R. A., and K. V. Pincus. 1996. “The relationship between materiality thresholds and judgments of fraud risk,” *Managerial Finance* 22(9):1-15.
6. Blair, Johnson T., Brian Mullen and Eduardo Salas. 1995 “Comparison of Three Major Meta-Analytic Approaches,” *Journal of Applied Psychology* Vol.80(1) Feb. p94-106.
7. Boatsman, J. R. and J. C. Robertson. 1974. “Policy-Capturing on Selected Materiality Judgments.” *The Accounting Review*. April, pp.342-352.
8. Burgstahler, David, Steven M. Glover and James Jiambalvo. 2000. “Error Projection and Uncertainty in the Evaluation of Aggregate Error,” *Auditing: A Journal of Practice and Theory*. Vol.19 No.1 Spring. pp.79-99.
9. Carpenter, B. W., M. W. Dirsmith, and P. P. Gupta. 1994. “Materiality Judgments and Audit Firm Culture: Social-behavioral and Political Perspectives,” *Accounting Organizations and Society* Vol. 19 Issues 4 & 5 pp.355-380.
10. Carpenter, B. W. and M. W. Dirsmith. 1992. “Early Debt Extinguishment Transactions and Auditor Materiality Judgments: A Bounded Rationality Perspective.” *Accounting Organizations and Society*. Vol. 17 Issue 8 pp.709-739.
11. Chewning, Eugene, G., Wheeler, Stephen W. and Chan, Kam C. 1998. “Evidence on Auditor and Investor Materiality Thresholds Resulting from Equity for Debt Swaps,” *Auditing: A Journal of Practice and Theory* Vol. 17. No.1 Spring, pp.39-53.
12. Chewning, Gene, Kurt Pany and Stephen Wheeler. 1989. “Auditor Reporting Decisions Involving Accounting Principle Changes: Some Evidence on Materiality Thresholds.” *Journal of Accounting Research*. Vol. 27 No. 1Spring pp.78-96.
13. Cho, Seong-Yeon, Robert Hagerman, Sandeep Nabar, and Evelyn Patterson, Evelyn, R. 2003. “Measuring Stockholder Materiality,” *Accounting Horizons* Supplement pp.63-76.
14. Costigan, M. L. and D. T. Simon. 1995. “Auditor materiality judgment and consistency modifications: Further evidence from SFAS No. 96.” *Advances in Accounting* Vol. 13 pp.207-222.
15. Compustat Database. 2006. Data provided by Standard & Poor's Compustat, a Division of The McGraw-Hill Companies, Inc.

16. Cumming, John. 1973. "An Empirical Evaluation of Possible Explanations for the Differing Treatment of Apparently Similar Unusual Events," *Journal of Accounting Research*, Supplement, Vol. 11 Issue 3, pp.60-95.
17. DeCoster, Jamie. 2004. "Meta-Analysis Notes," Department of Psychology, University of Alabama, Tuscaloosa. downloaded 12/31/2006 www.stat-help.com/notes.html
18. DeZoort, Todd, Paul Harrison and Mark Taylor. 2006. "Accountability and Auditor' Materiality Judgments: The Effects of Differential Pressure Strength on Conservatism, Variability and Effort," *Accounting, Organizations and Society*. Vol.31. Iss. 4-5. July-August. pp.373-390.
19. DeZoort, Todd F., Dana R. Hermanson and Richard W. Houston. 2003B. "Audit Committee Member Support for Proposed Audit Adjustments: A Source Credibility Perspective," *Auditing: A Journal of Theory and Practice*. Vol.22 No. 2 Sept. pp.189-205.
20. DeZoort A, Todd F., Dana R. Hermanson and Richard W. Houston. 2003A. "Audit Committee Support For Auditors: The Effects of Materiality Justification and Accounting Precision," *Journal of Accounting and Public Policy*. Vol.22 Iss.2 March-April. pp.175-199.
21. Dyer, Jack L. 1975. "Toward the Development of Objective Materiality Norms." *The Arthur Andersen Chronicle*. Oct. pp38-49.
22. Estes, R. and DD. Reames. 1988. "Effects of Personal Characteristics on Materiality Decisions: A Multivariate Analysis." *Accounting and Business Research*. Autumn, pp.291-296.
23. Fesler, R. D. and J. L. Hagler. 1989. "Litigation disclosures under SFAS No. 5: A study of actual cases." *Accounting Horizons* 3(1):10-21.
24. Financial Accounting Standards Board (FASB) 1980. *Statement of Financial Accounting Concepts No. 2, Qualitative Characteristics of Accounting Information*. sections 58-97 and 123-132.
25. Firth, M. 1979. "Consensus Views and Judgment Models in Materiality Decisions," *Accounting, Organizations and Society*. pp.283-295.
26. Frishkoff, P., 1970. "An Empirical Investigation of the Concept of Materiality in Accounting," *Empirical Research in Accounting*, pp.116-129.
27. Gieles, Frans. 1999 "An explanation of the statistics used in the meta-analysis" *IPCE Newsletter* E7. Dec.
28. Glass, G. V. 1976. "Primary, secondary and meta-analysis of research," *Educational Researcher*, 5: 3-8.
29. Gleason, Cristi A. and Lillian F. Mills. 2002. "Materiality and Contingent Tax Liability Reporting," *The Accounting Review*, Vol. 77, No. 2, April. pp.317-342.
30. Hatfield, Richard C., Scott B. Jackson and Scott D. Vandervelde. 2006. "The Effects of Auditor Rotation and Client Pressure on Proposed Audit Adjustments," Feb. unpublished working paper. Corresponding Author S. B. Jackson, University of South Carolina, Columbia, SC. scott.jackson@moore.sc.edu.
31. Hedges, L. V., and I. Olkin. 1985. *Statistical Methods for Meta-analysis*. Orlando, FL, Academic Press
32. Hofstedt, T. R. and G. D. Hughes. 1977. "An Experimental Study of the Judgment Element in Disclosure Decisions," *The Accounting Review* Vol. 52 No.2 April, pp.379-395.
33. Hojskov, Lief. 1998. "The expectation gap between users' and auditors' materiality judgments in Denmark," Paper presented at the Second Asian Pacific Interdisciplinary Research in Accounting Conference, in association with *Accounting, Auditing and Accountability Journal*. Osaka, Japan. August 4-6. pp.1-23.
34. Hunter, J.E. and Schmidt, F.L. (1990). *Methods of meta-analysis: correcting error and bias in research findings*. Newbury Park (CA): SAGE Publications.
35. Hunter, John E., Frank L. Schmidt, and Gregg B. Jackson. 1982. *Meta-analysis: Cumulating Research Findings Across Studies*, New Delhi, Sage Publication India Pvt. Ltd.
36. Icerman, R. C. and W. A. Hillison. 1991. "Disposition of Audit-detected errors: Some Evidence on Evaluative Materiality." *Auditing a Journal of Practice and Theory*, Spring, pp.22-34.
37. Iskandar, Takiah M. and Errol R. Islin. "A Review of Materiality Research," *Accounting Forum* Vol.23 No. 3 (Sept. 1999) pp.209-239.
38. Jennings, Marianne, M., Phillip M. J. Reckers, and Daniel C. Kneer. 1991. "The Auditor's Dilemma: The incongruous Judicial Notions of the Auditing Profession and Actual Auditor Practice," *American Law Business Journal*, 1991. pp99-125
39. Jennings, Marianne M., Dan C. Kneer and Philip M. J. Reckers. 1991. "Selected Auditor Communications and Perceptions of Legal Liability," *Contemporary Accounting Research*. Vol.7 No.2 pp.449-465.

40. Jennings, M., D. C. Kneer, and P. M. J. Reckers. 1987. "A Re-examination of the Court of Materiality: Views of Auditors, Users and Officers of the Court." *Studies A and B Auditing: A Journal of Practice and Theory*. Spring, pp.104-115.
41. Krogstad, J., R. Ettenson, and J. Shanteau. 1984. "Context and Experience in Auditors' Materiality Judgments." *Auditing: A Journal of Practice and Theory*. Fall, pp.611-618.
42. Levitt, Arthur. 1998 "The Numbers Game," Speech by SEC Chairman Arthur Levitt to the NYU Center for Law and Business, New York, NY September 28, 1998.
43. Libby, Robert; Mark W. Nelson, and James E. Hunton. 2005 "Recognition v. Disclosure and Auditor Misstatement Correction: The cases of Stock Compensation and Leases" *Social Sciences Research Network Working paper series* March 23, updated July 31.
44. Liu, C. S., and H. F. Mittelstaedt. 2002 "Materiality judgments and disclosure of retiree health care costs under SFAS No. 81," *Review of Accounting Studies*. Vol.7 pp.405-434.
45. Lyons, Larry C. 1997. "Meta-Analysis: Methods of Accumulating Results Across Research Domains," Manassas, Virginia. <http://www.lyonsmorris.com/MetaA/index.htm>
46. Martinov, N. and P. Roebuck. 1998 "The assessment and integration of materiality and inherent risk: An analysis of major firms' audit practices." *International Journal of Auditing* 2(2) 103-126.
47. Messier, William F., Nonna Martinov-Bennie, and Aasmund Eilifsn. 2005. "A Review and Integration of Empirical Research on Materiality: Two Decades Later," *Auditing, A Journal of Theory and Practice*. Vol. 24 Iss. 1 Nov. pp.153-187.
48. Messier, W. F., Jr. 1983. "The Effect of Firm Type and Experience on Materiality/Disclosure Judgments." *Journal of Accounting Research*. Autumn, pp.320-341.
49. Messier, William F. 1981. "An Analysis of Expert Judgment in the Materiality/Disclosure Decision," *Proceedings – Southeast Section, American Institute for Decision Sciences*. pp.117-119.
50. Moriarity, S. and Barron, F. H. 1979. "A Judgment Based Definition of Materiality," *Journal of Accounting Research* Vol.17 Supplement pp114-136.
51. Moriarity, S. and F. Barron. 1976. "Modeling the Materiality Judgments of Audit Partners." *Journal of Accounting Research*. Autumn, Vol. 14 Issue 2, p320-341
52. Morris, M. H. and W. D. Nichols. 1988. "Consistency Exceptions: Materiality Judgments and Audit Firm Structure," *The Accounting Review*. April, pp.237-254.
53. Morris, M. H., Nichols, W. D. and Pattillo, J. W. 1984. "Capitalization of Interest, Materiality Judgment Divergence and Users' Information Needs," *Journal of Business Finance and Accounting*. Winter, Vol. 11 Issue 4, pp547-555.
54. Muellerleile, Paige and Brian Mullen. 2006. "Sufficiency and Stability of Evidence for Public Health Interventions Using Cumulative Meta-Analysis," *American Journal of Public Health*, Vol.96 Iss.3, Mar. pp515-522
55. Nelson, Mark W., Steven D. Smith and Zoe-Vonna Palmrose. 2005. "The Effect of Quantitative Materiality Approach on Auditor's Adjustment Decisions," *The Accounting Review*. Vol.80 No.3. pp897-920.
56. Nelson, Mark W., Steven D. Smith and Zoe-Vonna Palmrose. 2004. "Quantitative Materiality Perspectives and Auditors' Disposition of Detected Misstatements," *Social Sciences Research Network Electronic Paper Collection*. Jan.29, 2004.
57. Newton, Lauren K. 1977. "The Risk Factor in Materiality Decisions," *The Accounting Review*. Vol. LII No.1 Jan. pp.97-108.
58. Pattillo, J. W. 1976. *The Concept of Materiality in Financial Reporting*. New York: Financial Executives Research Foundation.
59. Pattillo, James W. 1975. "Materiality: The (Formerly) Elusive Standard." *The Financial Executive*. August pp.20-27.
60. Pattillo, J. W. and J. D. Siebel. 1973. "The Concept of Materiality," *Financial Executive*. October, pp.28-30.
61. Prosser, William L. 1971 *The Law of Torts*. West Publishing Company. St. Paul, Minn. pp.622-627.
62. Robinson, C., and L. Fertuck. 1985. "Materiality: An empirical study of actual auditor decisions." *Canadian Certified General Accountants' Research Foundation*. Research Monograph No. 12
63. Romano, Albert. 1977. *Applied Statistics for Science and Industry*. Allen Bacon, Inc. Boston, p14.

64. Rose, J., W. Beaver, S. Becker and G. Sorter. 1970. "Toward and Empirical Measure of Materiality," *Empirical Research in Accounting Selected Studies*. Supplement of *Journal of Accounting Research* pp.138-48.
65. Rosenthal, R. 1991. *Meta-analytic procedures for social research*. (Rev. ed). Sage Publications. Beverly Hills, CA.
66. Rosenthal, R. and D. Rubin. 1978. "Interpersonal expectancy effects: The first 345 studies." *Behavioral and Brain Sciences*. 3, 377-415.
67. Russell, Craig J. and Stephen W. Gilliland, 1995. "Why meta-analysis doesn't tell us what the data really mean: distinguishing between moderator effects and moderator processes." *Journal of Management*, Winter.
68. Schneider, Arnold and Neil Wilner. 1990. "A Test of Audit Deterrent to Financial Reporting Irregularities Using Randomized Response Technique," *The Accounting Review*, July pp.68-681.
69. Shafer, William E. 2002. "Effects of Materiality, Risk, and Ethical Perceptions on Fraudulent Reporting by Financial Executives," *Journal of Business Ethics*. July. pp.243-262.
70. Street, Donna L., Nancy B. Nichols, and Sidney J. Gray. 2000. "Assessing the Acceptability of International Accounting Standards in the US: An Empirical Study of Materiality of US GAAP Reconciliations by Non-US Companies Complying with IASC Standards," *International Journal of Accounting*. Vol.35 Iss.1 pp27-62.
71. Stringer, Kenneth W. 1970. "Discussion of an Empirical Investigation of the Concept of Materiality in Accounting," *Journal of Accounting Research*, Vol.8, *Empirical Research in Accounting: Selected Studies*. pp.133-137.
72. Turner, Erick H., Annette M. Matthews, Eftihia Linardatos, Robert A. Tell, and Robert Rosenthal. 2008. "Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy," *The New England Journal of Medicine*. Vol.258. No.3 Jan. pp.252-260.
73. Turner, Lynn. 2000. "A QT report card for high quality financial reporting." Remarks at the Hylton Lecture Series on Accountancy: Critical Issues in Accounting Forum, Wake Forest University, Winston-Salem, NC, April 25.
74. Tuttle, Brad, Maribeth Coller and R. David Plumlee. 2002. "The Effect of Misstatements on Decisions of Financial Statement Users: An Experimental Investigation of Auditor Materiality Thresholds," *Auditing: A Journal of Practice & Theory*. Vol.21. No.1. March pp.11-27.
75. Waters II, John M.; Mikel G. Tiller. 1997. "Auditors' Materiality Thresholds: Some Empirical Findings Based on Real Data," *American Business Review*, Jun. Vol. 15 Issue 2, pp115-119.
76. Wheeler, Stephen, Kurt Pany and Eugene Chewning. 1993. "Inter-Firm Differences in Propensities to Modify Audit Opinions for Pre-SAS No. 58 Accounting Principals Change," *Accounting Horizons*. Vol.7 No.3 Sept. pp.46-54.
77. Whitlock, Michael and Dolph Schluster. 2005 "The Analysis of Biological Data, Part 5 Modern Statistical Methods," University of British Columbia <http://jackalope.zoology.ubc.ca/~whitlock/ABD/WhiSchPart5.pdf>
78. Woolsey, S. M. 1973. "Materiality Survey." *Journal of Accountancy*. Sept. pp.91-92.
79. Woolsey, Samuel M. 1954A. "Judging Materiality in Determining Requirements for Full Disclosure," *Journal of Accountancy*. Dec. pp.745-750.
80. Woolsey, Samuel M. 1954B. "Development of Criteria to Guide the Accountant in Judging Materiality." *Journal of Accountancy*. Feb. pp.167-173.
81. Wright, Arnold and Sally Wright. 1997. "Factors Affecting the Decision to Waive Audit Adjustments" *Journal of Accounting, Auditing & Finance*, Winter. Vol. 12 Issue 1, p15-36, 22p
82. Wright, Arnold. 1983. "The Impact of CPA-Firm Size on Auditor Disclosure Preferences," *The Accounting Review*. Vol. LVIII, No.53. July, pp.621-632.

APPENDIX A

Confidence Intervals

The conventional means of determining a confidence interval is to compute the standard deviation of a sample and use it and the appropriate Z-table value to estimate the high and low interval limits. However, in a meta-analysis the pooled standard deviation may not be available and there may no way to compute it from the information in the underlying studies, so alternative means of estimating confidence intervals must be employed.

The more data in a sample, the more likely the characteristics of the sample are to mimic the characteristics of the underlying population (Lyons, 1997) and the Central Limit Theorem says a sample size of 30 or more usually makes a good approximation of a population's distribution (Romano, 1977.) However, a standard deviation is simply a model of how data are distributed. If we cannot build this model, why not look directly at the data itself to find n_h and n_l which correspond to the observation numbers for the high and low limits of a confidence interval? Given these, confidence interval boundaries can be estimated.

Suppose a set of data contains 14,033 observations, which are sorted from low effect value to high effect value, and the desired confidence interval is 90% over a two-tailed distribution. Each tail should contain 5% of the observations. The lower limit of the confidence interval can be computed using equation (A1).

$$\text{Lower limit confidence interval (LL)} = n_l = 5\% * 14,033 = 702 \tag{A1}$$

The effect value at observation $n = 702$ is the lower limit of the 90% confidence interval. Similarly the high limit of the confidence interval can be computed using equation (A2).

$$\text{Upper limit confidence interval (UL)} = n_u = (100\% - 5\%) * 14,033 = 13,331 \tag{A2}$$

The effect value at observation $n = 13,331$ is the high limit of the confidence interval. This method is superior to using standard deviation because it automatically adjusts confidence interval boundaries for skewed data. Confidence intervals for subsets of data make it possible to determine whether there are statistically significant differences among them.

APPENDIX B

Inference

Inference is necessary to analyze and interpret historical materiality decisions in the form of financial statements and audit work papers. If a person classifies an item that is 12% of net income as material, all we really know is that 12% exceeds his or her threshold of materiality. On the other hand if one classifies a transaction of 2% as immaterial, we cannot know that individual's exact threshold of materiality, only that it is greater than 2%. However, suppose ten people are presented with items between 2% and 12%, two of these individuals classify items as immaterial and eight individuals classify transactions as material, one can use inference to estimate the threshold of materiality for the group. Keep two things in mind. First, the specific materiality threshold of any one individual is of little value. What we are seeking is a consensus on a materiality threshold. Second, a materiality threshold is an estimate. That estimate may vary from person to person and circumstance to circumstance so it would be imprudent to demand too much precision of any one materiality estimate.

Inference is aided if we make the reasonable assumption that individuals are more likely to view smaller items as immaterial and larger items as material. If eight out of ten individuals classify transactions between 2% and 12% as material, it would be reasonable to believe that the threshold of materiality is closer to 2% than to 12%. A simple proportion can be used to estimate the threshold of materiality for this group at about 4%.