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 than 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 Eilfesn 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 sample 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 Roeckuck 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 ad 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.

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

<table>
<thead>
<tr>
<th>Study Type</th>
<th>Cohorts</th>
<th>Participants</th>
<th>n</th>
<th>Mean</th>
<th>Q1</th>
<th>Median</th>
<th>Q4</th>
<th>90% Confidence Lower Bound</th>
<th>90% Confidence Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiments</td>
<td>221</td>
<td>3,762</td>
<td>10,721</td>
<td>7.52%</td>
<td>4.00%</td>
<td>6.74%</td>
<td>9.85%</td>
<td>1.50%</td>
<td>16.46%</td>
</tr>
<tr>
<td>Historical data</td>
<td>83</td>
<td>3,000</td>
<td>3,312</td>
<td>11.76%</td>
<td>3.62%</td>
<td>10.58%</td>
<td>17.06%</td>
<td>0.81%</td>
<td>26.37%</td>
</tr>
<tr>
<td>All Studies</td>
<td>304</td>
<td>6,762</td>
<td>14,033</td>
<td>8.52%</td>
<td>4.00%</td>
<td>6.78%</td>
<td>11.00%</td>
<td>1.47%</td>
<td>21.66%</td>
</tr>
</tbody>
</table>
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 QT is given by equation (1).

\[ QT = \sum \left[ w_i \times (T_i - T)^2 \right] \] (1)

Where \( w_i \) is the sample size, \( n_i \), 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, \( w_i \) is the number of materiality decisions in cohort, 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).

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

Where SM is the sub set mean; \( n_i \) is the number of materiality decisions made by cohort, \( M_i \) is the materiality threshold for cohort, 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

<table>
<thead>
<tr>
<th>Cohorts</th>
<th>n</th>
<th>Mean</th>
<th>Q1</th>
<th>Median</th>
<th>Q4</th>
<th>90% Confidence Lower Bound</th>
<th>90% Confidence Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>All study data</td>
<td>304</td>
<td>14,033</td>
<td>8.52%</td>
<td>4.00%</td>
<td>6.78%</td>
<td>11.00%</td>
<td>1.47%</td>
</tr>
<tr>
<td>Largest cohort</td>
<td>303</td>
<td>13,493</td>
<td>8.70%</td>
<td>4.69%</td>
<td>6.81%</td>
<td>12.55%</td>
<td>1.47%</td>
</tr>
<tr>
<td>Three largest cohorts</td>
<td>301</td>
<td>12,692</td>
<td>8.38%</td>
<td>4.83%</td>
<td>6.81%</td>
<td>11.00%</td>
<td>1.33%</td>
</tr>
<tr>
<td>Cohorts with the highest and lowest mean materiality</td>
<td>302</td>
<td>13,978</td>
<td>8.51%</td>
<td>4.00%</td>
<td>6.78%</td>
<td>11.00%</td>
<td>1.47%</td>
</tr>
<tr>
<td>Three cohorts with the highest mean materiality and the three cohorts with the lowest mean materiality</td>
<td>298</td>
<td>13,763</td>
<td>8.54%</td>
<td>4.10%</td>
<td>6.78%</td>
<td>11.00%</td>
<td>1.55%</td>
</tr>
<tr>
<td>500 of the highest reported materiality observations and 500 of the lowest reported materiality observations</td>
<td>NA</td>
<td>13,033</td>
<td>8.16%</td>
<td>4.21%</td>
<td>6.78%</td>
<td>10.50%</td>
<td>2.00%</td>
</tr>
</tbody>
</table>

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.

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 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.
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.

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

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.
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.

<table>
<thead>
<tr>
<th>Issue</th>
<th>Statistically Significantly Different Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asset Impairment or Write-off</td>
<td>Bond Premium Write-off <em><em>&lt;br&gt; Litigation</em>&lt;br&gt; Leases</em>*</td>
</tr>
<tr>
<td>Bribes</td>
<td>Bond Premium Write-off**&lt;br&gt; Casualty Loss**&lt;br&gt; Litigation**&lt;br&gt; Leases**</td>
</tr>
<tr>
<td>Contingent Liability</td>
<td>Bond Premium Write-off*&lt;br&gt; Casualty Loss*&lt;br&gt; Litigation*&lt;br&gt; Leases*</td>
</tr>
</tbody>
</table>

* 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 straightforward the issue, or the more straightforward 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.

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

### 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.
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.

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

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

15. Compustat Database. 2006. Data provided by Standard & Poor's Compustat, a Division of The McGraw-Hill Companies, Inc.


© 2011 The Clute Institute


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 \( n_l \) and \( n_u \) 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).

\[
\text{Lower limit confidence interval (LL)} = n_l = 5\% \times 14,033 = 702
\]  

(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).

\[
\text{Upper limit confidence interval (UL)} = n_u = (100\% - 5\%) \times 14,033 = 13,331
\]  

(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%.