Meta Analysis of Gender-Based Risk Taking

Louangrath, P.I. ★

About the Author

★Louangrath, P.I. is an Assistant Professor in Business Administration at Bangkok University, Bangkok, Thailand. For correspondence, email: Lecturepedia@gmail.com

ABSTRACT

This research is a meta-analysis of gender difference in risk behavior. The research examines 15 individual publications and 5 meta-analysis studies dealing with general difference in risk behavior. This paper attempts to answer the questions of "whether gender difference in risk taking behavior exists? If so, whether male and female entrepreneurs would behave differently as business owners? In prior studies, the answers to these questions had been inconclusive. The objective of this research is to reconcile conflicting results of the current literature on gender difference in risk behavior. This study has found that there is no statistical significance in gender difference in risk taking behavior: $Z(obs) < Z(0.95)$ in both individual papers and meta-analyses. This lack of gender difference is non-random under adjacent tests: $p = 0$. Statistically significant Cohen's d ($\mu_d = 0.38 \pm 0.33$) in prior studies may be reconciled by confounding variables, such as cultural beliefs and stereotypes; meta-analysis of Similarity Index also shows that there is gender similarity in risk behavior: ($\mu_{si} = 0.87 \pm 0.06$). This finding is confirmed by Weilbull statistics with $\beta < 1$ signifying diminishing gender difference in risk behavior; the survival rate for the similarity is $S(t) = 1$ and a score of instantaneous failure of $H(t) = 0$. Assertions about gender difference in risk behavior becomes demonstrative evidence of Type I error. The finding of this paper has significant implications because it dispels the perception of gender difference in risk behavior. The dissolution of such perception would assure equal treatment among male and female entrepreneurs; thus, enhancing social equity in business dealings.

Keywords: Adjacent test, Cohen's *d*, discrete probability, effect size, entrepreneurship, extreme value theory (EVT), gender difference in risk behavior, H distribution, meta-analysis, randomness test, risk behavior, similarity index, Weibull distribution.

CITATION:

Louangrath, P.I. (2015). "Meta Analysis of Gender-Based Risk Taking." *Inter. J. Res. Methodol. Soc. Sci.*, Vol., 1, No. 1: pp. 58-89. (Jan. – Mar. 2015).

1.0 INTRODUCTION

Risk is defined as the uncertainty that affects the outcome of the objective (ISO 31000: 2009, Guideline 73). How a person reacts to or behaves towards risk is known as risk behavior

(Kahneman and Tversky, 1979). The research question presented by this paper is: "whether there is gender difference in risk behavior?"

There are interests in research on gender differences of risk behavior because an affirmative finding of the difference would lead to certain policy agenda and practice. In the context of entrepreneurship, the affirmation of gender difference in risk behavior would justify gender oriented agenda leading to discrimination. Such discrimination may result in gender-based social inequality.

There are three streams of literature on this research issue, namely (i) attempt to prove that gender difference exists, (ii) disprove that the difference does not exist, and (iii) attempting to unify or explain the difference in a more balanced and neutral stance. This paper follows third stream of the literature by using quantitative analysis under meta-analysis to reconcile the differences in prior publications on the issue of gender difference in risk behavior.

The current literature on the issue of gender difference in risk behavior is voluminous and the findings are far from unanimous. Despite prior meta-analysis of the issue, the question of "whether there is a genuine gender difference in risk behavior?" remains unsettled. This research is an attempt in meta-analysis studies to unify these contradictory findings. Unlike prior meta-analysis studies that amalgamated many studies into one set of meta-data, this paper selected 15 individual papers and 5 meta-analysis papers for review. The objective is to find a consensus in their findings. The intended contributions of this paper are its scientific and practical values. First, the methodology for meta-analysis adopted in the paper goes beyond the traditional notion of effect size. This paper uses the effect size as the unit of observations and subjects these observations unto further statistical analysis. Second, the findings of this paper has a practical implication to entrepreneurship study in that it has objectively verified that there is no gender difference in risk behavior; to argue otherwise is to engage in gender discrimination.

This paper is organized into five sections. Section 1 introduces the subject matter, the data set and sample size. Section 2 reviews the current literature relating gender difference on risk behavior. Section 3 outlines the methodology and various statistical tests employed in this paper. Section 4 presents the findings and discussion. Section 5 concludes that gender difference found in prior publications may be explained by confounding factors, such as culture and stereotype or miscalculating of data.

2.0 LITERATURE REVIEW

2.1 Gender Difference Theory

The theoretical foundation behind gender difference in risk behavior may be categorized into two streams: nature and nurture. The nature line of literature suggests that gender difference in risk behavior is innate or biological (Ichino and Moretti, 2009). One nature literature was unequivocal in claiming that gender difference in risk behavior is biological; the Social Issues Research Centre from the UK wrote that "[m]any factors underpin these differences, including neurochemical structures and hormonal processes shaped by evolution …" (SIRC, 2004, p.3). Other publications had been less provocative.

Pawlowski *et al.* (2008) reviewed the nature literature under "Sexual Selection Theory" wrote that:

> "In species like humans where risk-taking may itself become a form of display, this sex difference may be exaggerated and risk-taking may characterize many aspects of behavior. Many studies have noted that young human males are more prone than females to take risks in relation to conflict (Campbell, 1999; Daly and Wilson, 1988; Wilson and Daly, 1993,) and sexual behavior (Clift, Wilkins, and Davidson, 1993; Poppen 1995), as Sex differences in risk-taking well as in such situations as car driving (Chen, Baker, Braver, and Li, 2000; Flisher, Ziervogel, Charlton, Leger, and Roberston, 1993; Harre, Field, and Kirkwood, 1996), accident risk (Fetchenhauer and Rohde, 2002), drug-taking (Tyler and Lichtenstein, 1997), gambling and financial decisions

> (Bruce and Johnson, 1994, Powell and Ansic, 1997) and outdoor activities (Howland, Hingson, Mangione, and Bell, 1996, Wilson, Daly, Gordon, and Pratt, 1996). Indeed, psychological studies have found that females find risky situations more stressful than males do (Kerr and Vlaminkx, 1997). In this context, risktaking by males may be a form of mating display (Hawkes, 1990, 1991)."

The rationale for this one aspect of nature argument is that: "[e]volutionary theory predicts that, in polygamously mating species, young males will be more willing to take risks in an effort to breed successfully than young females." (Pawlowski *et al.* p. 29).

In another study, it was found that men are more risk affine than women (Thom, 2003). Thom wrote that:

> "There is clear evidence of gender differences in responses to risk taking, health, and help seeking for health problems. Male roles and social identities may themselves be 'risk factors' for higher rates of morbidity and mortality among men compared to women. This deserves more attention in preventive approaches and (specifically) in developing responses to substance use and related problems." (Thom, 2003, p. 29).

Thom's data came from non-experimental empirical evidence of heath risk exposure by men and women in the UK. Using time series data on health risk exposure over a period of 5 to 10 years, Thom's conclusion may be collaborated with our recalculation of some of the data sets in that study. For instance, the differences between male and female in health risk exposure by cigarette smoking under d-bar analysis is $T_{d(obs)} = 3.67$ compared to the 0.95 confidence interval for the 7 years period of $T(6, 0.95) = 1.94$. Similar finding was confirmed by sexually transmitted disease (Hepatitise C) comparison in the same study. Using sexually transmitted disease as a measure of risk exposure, we recalculated Thom's data set under d-bar analysis and found that $T_{d(obs)} = 4.24$ compared to the null hypothesis of $T(8, 0.95) = 1.86$. Although Thom's conclusion is correct that "[t]here is clear evidence of gender differences in responses to risk taking," the study is not conclusive whether the difference is due to nature or nurture. Smoking and sexual behavior are not innate behavior; they are socially constructed. The literature offering alternative explanation include:

> "[S]ocial learning theory (Akers 1977, 1998; Bandura 1977), social identity theory (Abrams & Hogg 1990), primary socialization theory (Oetting & Donnermeyer 1998; Oetting et al. 1998) and social network theory (Granovetter 1973; Wasserman & Faust 1994), as well as social bonding theory (Hirschi 1969), a general theory of deviance (Gottfredson & Hirschi 1990), the theory of reasoned action (Fishbein & Ajzen 1975), the theory of planned behavior (Ajzen 1985, 1988), the triadic theory of influence (Flay & Petraitis 1994), peer Peers and adolescent smoking 39 © 2003 Society for the Study of Addiction to Alcohol and Other Drugs Addiction, 98 (Suppl 1), 37–55 cluster theory (Oetting & Beauvais 1986) and social development theory (Hawkins & Weis 1985; Catalano & Hawkins 1996)." (cited in Kobus, 2003, p. 38-39).

In contrast, the nurture line of literature suggests that gender difference comes from the environment. Social pressures and stereotypes on gender role are responsible for the difference. The nurture literature offers a different perspective to the issue of gender difference in risk behavior. Booth and Nolen found "that observed gender differences in behaviour under uncertainty found in previous studies might reflect social learning rather than inherent gender traits." (Booth and Nolen, 2009, p. 17). Other nurture theorists look at culture as the environmental factor that shape's a person's attitude towards risk (Gneezy, Leonard and List, 2008).

Whether nature or nurture theory is accepted, gender difference issue is of interest to researchers. On a practical level, Mincer and Polachek (1974) showed that differences in labor market experience between men and women can account for a substantial share of the gender gap in earnings. In the context of entrepreneurship, the acceptance or rejection of gender difference in risk behavior has a wider implication in training future entrepreneurs and the availability of business opportunities for these entrepreneurs.

2.2 Past Gender Difference Research and Findings

There have been gender studies in many aspects. These studies included the impact of gender on intelligence testing (Born, Bleichrodt and van der Flier 1987), leadership style, evaluation and effectiveness (Eagly and Johnson 1990; Eagly and Karau 1991; Eagly, Makhijani and Klonsky 1992; Eagly, Karau and Makhijani 1995),conformity and social influence (Cooper 1979; Eagly and Carli 1981, Johnson and Eagly 1989), cognitive ability including mathematical, verbal and spatial ability (Hyde 1981; Hyde, Fennema and Lamon 1990; Hyde and Linn 1988; Linn and Peterson 1985), personality development (Cohn 1991, Feingold 1994), self-disclosure (Dindia and Allen 1992), aggressive behavior (Eagly and Steffen 1986, Hyde 1984) and social behavior (Eagly and Wood 1991, Wood 1987).

The findings of these prior studies had been inconclusive. For instance, in economics, researchers use Voluntary Contribution Mechanism (VCM), i.e. retirement plan, as a mechanism to measure gender difference in risk behavior. A willingness to participate in VCM is seen as a willingness to take risk. The findings are inconclusive. Brown-Kruse and Hummels (1993) find that men contribute more toward the public good than women. This effect is also found in Sell and Wilson (1991) and Sell (1997). However, Nowell and Tinkler (1994), and Seguino, Stevens and Lutz (1996) find that women contribute more toward the public good than men. Yet in other studies by Sell, Griffith and Wilson (1993), Cadsby and Maynes (1998), and Solow and Kirkwood (2002) find no significant differences.

This paper focuses on a particular issue: gender difference in risk behavior and its relevance to entrepreneurship. The findings on gender difference in risk behavior are also inconclusive. In entrepreneurship research, we often face the issue of gender difference in risk behavior. Some publications show that there is significant difference in risk behavior according to gender (Schubert *et al.*, 1999; Power and Asic, 1997; and Eckel and Grossman, 2002). Other researchers published the opposite findings (Moore and Eckel, 2003; Atkinson *et. al.*, 2003; Johnson and Powell, 1994; Master and Meier, 1988; and Gysler *et al.*, 2002). This paper is an attempt to reconcile these conflicting findings via meta-analysis. Meta-analysis is an analysis of research results from a body of publications on a specific topic through the use of statistical tools for the purpose of consolidating a consensus in prior findings (Greenland S, O'Rourke, 2008).

Guideline 73 of IS0 31000 (2009) defines risk as "the effect of uncertainty on objectives." The research question presented in this paper is "whether there is gender difference in risk behavior? If so, do male and female entrepreneurs behave differently as business owners?" The answers to these two questions would contribute to the current literature in the field of entrepreneurship study.

Common traits of entrepreneurs are leadership, management ability, and team-building to be essential qualities of an entrepreneur (Drucker, 1985). These qualities do not stray from the character content of successful business managers in the corporate world. According to Schumpeter, an entrepreneur is someone who is willing and able to turn an idea into an innovation and shifts the production possibility frontier to a newer height. (Schumpeter, 1934 & 1976). Drucker and Knight see an entrepreneur as a risk taker. (Knight, 2005). Shane pointed out that even with the willingness to take risk in order to succeed an entrepreneur must have proximity to opportunity (Shane, 2000). These two key elements: risk taking and proximity to opportunity provide the motivation for this paper on meta-analysis of gender difference on risk behavior and its application to entrepreneurship. The logic is that if there is a gender difference in risk behavior, such difference would also

influence the degree of exposure to opportunity and, hence, the level of success among male and female entrepreneurs. If that is the case, business opportunity may become gender dependent thereby sex discrimination would be condoned.

3.0 DATA AND METHODOLOGY

3.1 Data selection

One challenge in meta-analysis study is the selection of publications to constitute group members in the study. In this paper, both individual publications and meta-analysis publications are used. It is noted that studies with small sample may not be a good representative of the issue at hand (LeLorier *et al.*, 1997). However, it could not be denied that studies with small sample size may prove significant issues (Lewis and Sauro, 2006). Exclusion of such studies is a clear case of selection bias. For instance, methods to detect publication related biases and assess their potential impact have been well documented for meta-analyses that use extracted aggregated study results (such as treatment effect estimates) (Sterne *et al.*, 2001, Rothstein *et al.*, 2005, and Egger *et al.*, 1997).

It has been written that "[a] known threat to the validity of meta-analysis is publication bias, which occurs when studies with statistically significant or clinically favorable results are more likely to be published than studies with non-significant or unfavorable results." (Ikhlaaq *et al.*, 2012). For this reason, a mixed group of studies with various sample sizes are selected for this paper. This approach assures that the study will not be dominated by large studies (Helenstein, 2002). In addition to size, a second criteria used for inclusion selection is research design. If the elements of the meta-analysis are poorly designed, no matter how good the meta-analysis may be, the result would still be poor (Slavin, 1986). The main data sets used in this paper is tabulated in the table below.

Table 1. Cohen's *d* and Similarity Index used as dataset in meta-analysis

Study Number*	Cohen's *d*	Sample Size	SI Index	Sample Size
1	0.26	38,000	0.98	12,000.00
2	0.46	200	0.91	200.00
3	0.38	300	0.87	300.00
4	0.55	300	0.84	300.00
5	1.45	150	0.88	22,000.00
6	0.48	22,000	0.80	300.00
7	1.13	300	0.91	200.00
8	0.22	200	0.93	120.00
9	0.49	100	0.96	13,000.00
10	0.33	1,200	0.86	2,000.00
11	0.74	700	0.86	200.00
12	0.29	13,000	0.93	100.00
13	0.37	200	0.96	6,000.00
14	0.36	2,000		
15	0.85	150		
16	0.65	200		
17	0.17	100		
18	0.31	400		
19	0.44	50		
20	0.16	6,000		

*A list of studies is provided in the Appendix: Table A3.

The Cohen's *d* looks for the difference between men and women in risk behavior. This method is called the effect size measurement. The SI Index looks for the similarity between two

variables. In order to have these two data sets treated in the same fashion, the SI is subtracted from 1 in order to defined it in terms of male-female difference: $1-SI$. The SI index is extracted from the 20 publications (Nelson, 2012). There are 13 out of 20 publications reported both Cohen's *d* and SI index. Seven publications reported only Cohen's *d*.

3.2 Number of Sample Studies as Sample Size of Meta Analysis

"Most meta-analyses include data from one or more small studies that, individually, do not have power to detect an intervention effect." (Turner and Higgins, 2013). Studies with small sample size tend to over-report the effect of the intervention (Sterne *et al.*, 200). Nevertheless, these small sized studies found their way into publication and passed peer reviews due to their findings of statistical significance (Nygard *et al.*, (1995). However, whether these statistical findings are actual "statistical significance" or a result of experimental design defect remains a contending issue. (Kjaergard *et al.*, 2001). Thus, in meta-analysis study, the challenge of selecting the publications for inclusion in the study becomes problematic. It has been suggested that 90% of prior publication could be discard as unacceptable and accepting 10% for inclusion would suffice. (Stanley *et al.*, 2010). In a less draconian approach to studies selection, it has been suggested that bias resulted from small sample size could be addressed through "statistical methods of adjustment." (Moreno *et al.*, 2009; and Rucker *et al.*, 2011). Sample-studies size used in this meta-analysis is obtained through distribution test under (1) and the issue of "power" is verified through statistical methods.

It has been shown that 80% of published studies drew conclusion from insufficient sample size (Westland, 2010, 476-487). In all research, sample size is an important consideration for acceptable scientific method and ethics in research (Altman, 1994). It has been suggested that there should be at least ten sample counts per indicator variable if the proposed model is multiple regression (Nunnally, (1967, 355). According to Nunnally, in single regression, minimum sample size may be as little as ten counts. This minimum sample size requirement is consistent with the Anderson-Darling test for data distribution where the minimum sample size is $n > 5$ or at least 6 counts for nonparametric case. (Anderson and Darling, 1954, 765-769). Other writers proposed 30 counts in order for the study to avail itself to the full benefits of the Central Limit Theorem (CLT). (Agresti and Min, 2012, 2; and Louangrath, 2013 & 2014).

Minimum sample size in meta-analysis is called sample study size. The current literature is vague on the issue minimum sample (study) size for meta-analysis. The literature focuses on the effect size. Hedges wrote about "sample effect size," but did not elaborate on the method for determining minimum sample size to comprise the meta-analysis. (Hedges and Pigott, p. 205). However, in Hedges it was assumed that the effect size (T_i) is normally distributed: $T_i \cong N(\theta_i, \mathrm{v}_i) \quad i = 1,...,k$. Based on that assumption, this paper adopts n-*omega* method of estimating sample size by using the Cohen's d and SI index values as the basis for the calculation. The n-*omega* method is given by:

$$n_\omega = \sqrt{\frac{\omega}{2}} \qquad (1)$$

where $\omega = \sqrt{(\mathrm{n}_2 - \mathrm{n}_1)/2}$; $n_2 = \sqrt{(Z^2\sigma^2)/E^2}$; and $n_1 = \sqrt{Z\sigma/E}$. (Louangrath, 2014). Sample-studies size is often overlooked in meta-analysis. Most authors focused on statistical power via the effect size and the sample size of the individual studies. However, the question of "how many studies should be included in the meta-analysis?" is overlooked. This design defect belittles the creditability of the meta-analysis studies. This paper overcomes this gap in the literature by including the treatment of minimum sample studies size through equation (1).

3.3 Meta analysis approach

The units of analysis of this meta-analysis study are the effect size and Similarity Index (SI). The tools used in this paper follow conventional statistical testing. Effect size is an indication of statistical power.

Statistical power of statistical tests used in meta-analysis include: effect size across studies (Hedges and Vevea, 1998), test for effect size heterogeneity (Hedges, 1982a), test for contrast in effect size (Hedges, 1982b, Rosenthal and Ruben, 1982), and test for group differences among various studies (Hedges, 1982b). The power test adopted by this study is effect size across studies suggested by Hedges and Vevea. In addition, a second indicator called Similarity Index is used as a back-up test to verify the gender difference in risk behavior. Unlike prior meta-analysis studies, this paper employs Extreme Value Theory (EVT) as an analytical tool to verify data distribution.

3.3.1 Two Approaches in Meta-Analysis

Meta-analysis studies are generally divided into two types: fixed effect, and random effect analyses. No matter which approach is taken, the objective of meta-analysis is to determine the combined effect size or true effect size in order to reconcile the results of various studies. Effect size is the measurement of the strength of the phenomenon (Kelly, 2012). It is considered good practice in research to disclose the effect size of the studies so that the substantive interpretation of the results could be made (Wilkinson, 1999; Nakagawa, 2007; and Ellis, 2010). This paper select 5 meta-analysis and 15 individual studies as the elements for the meta-analysis for gender difference in risk behavior. Two indicators are analyzed: the effect size as measured by Cohen d (Cohen, 1988) and Similarity Index. The Cohen's d is given by:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S} \tag{2}$$

where $$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \tag{3}$$

Detailed methodologies on how to calculate Cohen's d appear in Appendix A3. Some authors remove -2 from the denominator (McGrath, 1988, and Hartung, 2008). Note that equation (2) measures the effect size in units of standard deviation.

In meta-analysis, the common measure used is called the effect size. The effect size is the quantitative measurement of the strength of the phenomenon (Kelly, 2012). There are many methods for measuring the effect size. They are group into five approaches: (i) correlation analysis (Lejnik and Algina, 2003; Steiger, 2004), (ii) difference analysis (McGrath and Meyer, 2006), (iii) categorical data (Deeks, 1998; Greenland, 2004; and Zhang, 1998), (iv) common language effect size (Grissom, 1994), and (v) effect size of ordinal data.

It is incorrect to equate effect size to percentage points. It is a common practice to use the effect size as a final point of the analysis. For instance, Cohen stated that the Pearson correlation coefficient may be used as the measure of effect size and the effect size is classified as: small (0.20), medium (0.50) and large (0.80) (Cohen, 1992 cited in Thalheimer, 2002: 3). This is an incomplete analysis because it falls short of answering the question of whether the effect size is statistically significant. For instance, if $r = 0.50$, it still not meaningful if it is statistically not significant under the T-test. This misreading of the effect size is also repeated in the interpretation of Cohen's d. It is a misunderstanding to interpret as a percentage of the strength of the relationship between two groups. For instance, in the study of the differences in height between men and women, McGraw and Wong wrote:

> "[I]n any random pairing of young adult males and females, the probability of the male being taller than the female is .92, or in simpler terms yet, in 92 out of 100 blind dates

among young adults, the male will be taller than the female." (McGraw and Wong, p. 381).

This is not an accurate statement. An effect size of 0.92 does not mean 92%, it means a standard score of 0.92 which corresponds to 82.10% probability in the Z-Table. This misreading of the effect size is also seen in many meta-analysis studies. For instance, in Nelson, the effect size (Cohen's *d*) is read as a raw number or according to the scale suggested by Cohen (small (0.20), medium (0.50) and large (0.80); no significance test was conducted. For that reason, this paper uses the same data set employed by prior publications and takes a step further by re-analyzing the 20 readings of the Cohen's *d* with the test of significance.

The second indicator used in this meta-analysis study is the Similarity Index. The Similarity Index is given by:

$$SI = 1 - \frac{1}{2}\left(\sum_i \left| \frac{f_i}{F} - \frac{m_i}{M} \right| \right) \quad (4)$$

SI measures the "sameness" among the subjects. In order to maintain consistency of logic and language of the Cohen's *d*, which measures the difference among groups, 1 – SI is also used to connote the difference among groups. This second indicator is used as a back-up test to verify the conclusion reached in Cohen's *d* calculation. This two-tier approach is intended to prevent type I & II errors.

3.3.2 Fix Effect Method

There are two approaches to meta-analysis study, namely fixed effect and random effect model. Fixed effect method assumes that the true effect size in the studies is shared by all studies included in the meta-analysis group and the combined effect sized may be estimated through ordinary least square (OLS) method. Weight may be assigned to each study in the meta-analysis. There tends to be sample size bias in this approach. For example, larger sample size would have greater weight and *vice versa*. The error used is the error *within* studies. There is a bias in favor of large sample where the error tends to diminish as the sample size becomes larger. In contrast, smaller sample size would produce larger error (Ugrinowitsch *et al.*, 2004). Due to the selection bias in favor of studies with large sample size, prior publications using fixed effect model were not selected for this paper.

3.3.3 Random Effect Method

In random effect approach, it is assumed that the true effect size may vary from one study to another. The rationale for this assumption is that not all studies were conducted under the same condition, or with the same population or sample size. Therefore, the true effect size may vary from one study to another. The true effect size is the mean effect of the distribution of the effect size included in the study. Since the distribution of the effect size is used, no publication regardless of sample size is excluded. The determination of the combined effect in random effect model consists of two levels of calculation. First, the true effect of each study is calculated. For example, if there are 20 studies in the meta-analysis, there will be 20 separate estimates of true effect values in the first calculation. This is called the true effect *within* each study. Second, the mean of the 20 individual true effects are estimated to constitute the combined true effect. This second calculation is the true effect *between studies* (Senn, 2007).

This paper employs the random effect approach. Since random effect method employs distribution analysis in combined effect determination, and the individual studies in the meta-analytic group consists of studies with large and small samples, extreme value analysis (EVA) approach is used. The specific EVA tool used in this paper is Weibull distribution analysis where $\xi < 0$ (Weibull, 1951) after verifying the tail index under the Pickland and Hill methods.

3.3.4 Combined Effect Size in Random Effect Meta-Analysis

Morris and DeShon wrote that: "[e]xtracting effect sizes from primary research reports is often the most challenging step in conducting a meta-analysis." (Morris and DeShon, 2002). The reason for this difficulty comes from the fact that research designs are not the same or uniform. The individual publications selected for meta-analysis must have comparable designs. Deciding "what is comparable design?" is not an easy task. Further challenges in combine effect determination are not adequate to determine the total effect size because the data came from many studies. The total effect size is obtained through:

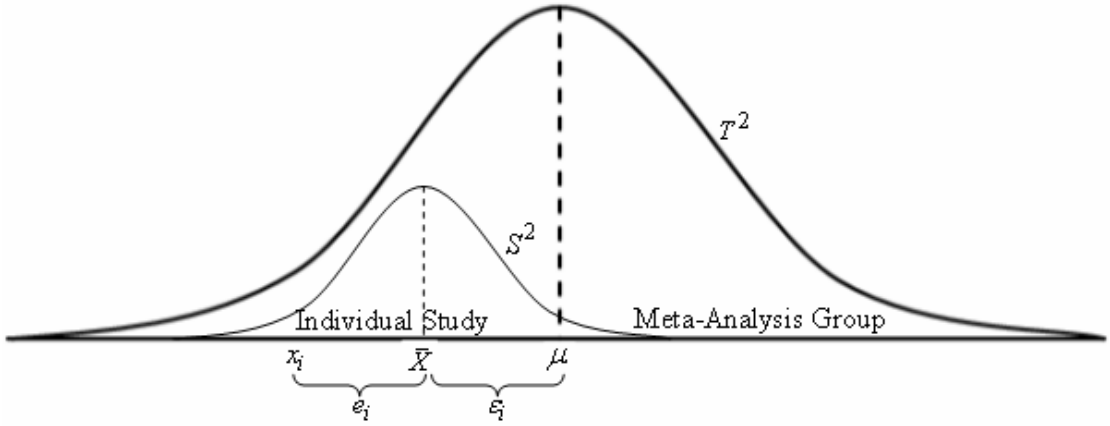
$$T_i = \theta_i + e_i \qquad (5)$$

where T_i = observed effect; θ_i = true effect; and e_i = forecast error. Since the "true effect" θ_i could not be observed directly, it could only be estimated.
Therefore, equation (5) may be written as:

$$T_i = \mu + \varepsilon_i + e_i \qquad (6)$$

Note that there are two error terms in (6): ε_i = error *between* studies, and e_i = sampling error *within* studies.

Fig. 1. Defining true effect for random effect method in meta-analysis



In Figure 1, the true observed effect is $T_1 = \theta_1 + e_1$ and the true effect is $\theta_1 = \mu + \varepsilon_1$. The meta-analysis study size represents the group. The true effect with an individual study θ_1 is distributed about the group mean μ. Note that the Greek letter used here does not represent inferential statistics. This approach is the general method in the literature (Borenstein *et al.*, 2007). However, in this study the combined effect measurement in (5) and (6) are augmented by inferential statistics in order to provide better reading of the measurement. This strengthening procedure consists of two steps: (i) determine inferential statistics for the meta-analysis group; and (ii) determine the true effect for each study.

3.4 Determine Inferential Statistics for the Meta-Analysis Group

Let individual studies in the meta-analysis represented by $X_i : (x_1, x_2, ...x_n)$ with mean μ and standard deviation $S = \sqrt{S^2}$ and $S^2 = T^2$, the estimated mean for the meta-analysis group may be obtained through the sample distribution equation:

$$t = \frac{\mu - \hat{\mu}}{S / \sqrt{n}} \tag{7}$$

The estimated true effect for the meta-analysis group is be obtained by solving for $\hat{\mu}$ (*mu* hat); thus:

$$\hat{\mu} = \mu - t\left(S / \sqrt{n}\right) \tag{8}$$

where μ = group mean; t = critical sample student t-score at a specified percentage confidence and degree of freedom; S = standard deviation of group; and n = group size. The *within* sample study and *between* studies errors (e_i and ε_i) may be rewritten as: $e_i = T_1 - \theta_1$ and $\varepsilon_i = \theta_1 - \hat{\mu}$. Equation (5) may now be written as:

$$T_i = \hat{\mu} + \varepsilon_i + e_i \tag{9}$$

This modification (8) allows us to use the estimated population mean of the group as the reference point for estimating the true effect.

In Figure 1, the variance for the group is represented by T^2, since we modified (6) by substituting μ with $\hat{\mu}$, we will substitute T^2 with the estimated group variance via the Unit Normal Distribution equation:

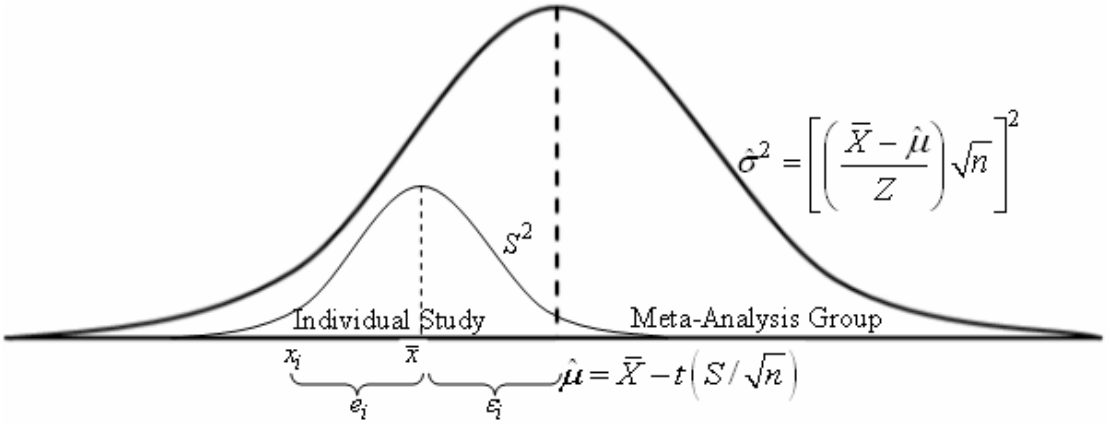
$$Z = \frac{\bar{X} - \hat{\mu}}{\hat{\sigma} / \sqrt{n}} \tag{10}$$

By solving for $\hat{\sigma}$:

$$\hat{\sigma} = \left(\frac{\bar{X} - \hat{\mu}}{Z} \right) \sqrt{n} \tag{11}$$

where Z = standard Z-score for a specified percentage confidence interval; $\bar{X}$ = observed group mean in meta-analysis; $\hat{\mu}$ = estimated mean obtain in (8) and n = meta-analysis group size. The two modifications are represented in Figure 2.

Fig. 2. Contrasting individual study and meta analysis



The rationale for the modification by introducing inferential statistics of the meta-analysis group as reference points (8) and (11) is to prevent selection bias from the meta-analysis group size (Ards *et al.*, 1998, Cortes *et al.*, 2008 & 2014). There are many research publications on gender difference for risk behavior, the group size in the present study may not have covered substantial number of publications; thus, the true mean and variance of the "meta-analysis" could not be readily obtained. (Rosenthal, 1979). However, the second best approximation is to assume that the current meta-analysis group size is a sample whose inferential statistics could represent the true group.

3.5 Determine the True Effect for Each Study

The modifications in (8) and (11) become useful in our variance decomposition for purposes of determining total variance (τ) and weight assignment (w). The total variance *within* and *between* studies is obtained by series of steps beginning with Q. The Q statistic is given by the sum of weighted sum squared deviation of the individual studies from the combined mean ($\bar{T}$). If the source of variance comes only from the error of *within* studies, then $Q = df_{group}$. However, in many instances, the variance also come from *between* studies. The variance *between* studies is obtained by:

$$\tau^2 = \frac{Q - df}{C} \tag{12}$$

where the terms are defined as:

$$C = \sum w_i - \left(\frac{w_i^2}{\sum wi}\right) \tag{13}$$

$$Q = \sum_{i=1}^{k} w_i \left(T_i - \bar{T}\right)^2 \tag{14}$$

3.6 Assigning Weights to Studies

In meta-analysis, we examine many prior research publications. These publications have different experimental designs and sample sizes. In order to give fair treatment of all publications selected for the meta-analysis, we need to give weight to each publication. The weight of random effect model is given by:

$$w_i = \frac{1}{V_i} \tag{15}$$

where V_i = within study variance (see Figure 2) or $S^2 + \tau^2$. The weighted mean is given by:

$$\bar{T}_w = \frac{\sum w_i T_i}{\sum w_i} \tag{16}$$

The variance combined effect is calculated by taking the reciprocal of the sum of the weight:

$$V = \frac{1}{\sum w_i} \tag{17}$$

By taking the square root of the variance of the combined effect, we obtain the standard error:

$$SE(\bar{T}_*) = \sqrt{V_*} \tag{18}$$

By setting the percent confidence level to 95% with the corresponding Z-critical value of 1.65, the upper and lower confidence intervals are given by: $-1.65(SE(\bar{T}_*)) < \bar{T}_* < +1.65(SE(\bar{T}_*))$ where $Lower < \bar{T}_* < Upper$. Thus, the test statistic to verify if the combined effect is statistically significant is given by:

$$Z = \frac{\bar{T}_*}{SE(\bar{T}_*)} \tag{19}$$

The decision rule used in this paper is $H_0 : Z(obs) < 1.65$ and $H_A : Z(obs) > 1.65$.

3.7 Determination and Treatment of Extreme Values

Prior meta-analysis publications did not consider extreme values even though both fixed effect and random effect models use OLS as their building block. OLS is based on mean difference square. If there is an extreme value in the set, large variance will result. The reading of the result becomes unreliable (Ugrinowitsch *et al.*, 2004, p. 2144-48). It is a common knowledge that extreme values cause bias in the final estimate. Thus, the present in a group may cause large variance and inaccurate reading of the effect size. This paper addresses this inadequacy in the literature concerning meta-analysis methodology by employing Extreme Value Theory as part of the analytical tools.

Extreme values may cause bias and inaccuracy in the study especially when OLS is employed as the building block of the analytical tool. This paper used two indicators from various publications: Cohen's *d* and the Similarity Index. The Cohen's *d* in the select studies ranges from the low of 0.16 to the high of 1.13. Similarly, the Similarity Index ranges from 0.80 to 0.98. In the present study, both the Cohen's *d* and Similarity Index had been tested for extreme values through standard Z-score using maxima and minima as testing values. If the standard score proves that some of the values in the array are significantly large or small, the array is treated as a suspect for extreme value distribution analysis. The standard score is given by:

$$Z = \frac{X_i - \bar{X}}{S} \tag{20}$$

To discover extremely large number the subscript i is substituted with maxima in the array and *vise versa* for the extremely small value. After extreme values are discovered, we proceed to verifying the type of extreme value distribution.

Extreme values may be analyzed under the generalized extreme value (GEV) distribution proposed by Fisher-Tippett-Gnedenko:

$$H(x;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{21}$$

where μ = location; σ = scale; and ξ = shape. If $\xi > 0$, H becomes a cumulative distribution function (CDF); if $\xi < 0$, it is valid for $x < \mu + \sigma/(-\xi)$; and if $\xi = 0$, H is undefined (Bensalah, 2000). However, if $\xi \to 0$, then (21) is reduced to:

$$H(x;\mu,\sigma,0) = \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)\right\} \tag{22}$$

The parameter ξ is the tail index of the distribution. This index may be used to classify the type of extreme value distribution. If $\xi = 0$, the H distribution is Gumbel distribution, also known as Type I where $x \in \Re$ and $\xi = 0$. The Gumbel distribution is given by:

$$H(x;\mu,\sigma,0) = \exp\left\{-\exp\left(\frac{x-\mu}{\sigma}\right)\right\} \tag{23}$$

If $\xi > 0$, the H distribution is a Fréchet distribution or Type II. The Fréchet distribution is given by:

$$H(x;\mu,\sigma,\xi) = \begin{cases} 0 & \text{for } x < \mu \\ \exp\left\{\left(\frac{x-\mu}{\sigma}\right)\right\}^{-\alpha} & \text{for } x > \mu \end{cases} \tag{24}$$

If $\xi < 0$, the H distribution is Weibull distribution or Type III. The Weibull distribution is given by:

$$H(x;\mu,\sigma,\xi) = \begin{cases} \exp\left\{-\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right\}^{-\alpha} & \text{for } x < \mu \\ 1 & \text{for } x \geq \mu \end{cases} \tag{25}$$

The next step was to classify the type of extreme value distribution of the series through the use of the tail index. The tail index may be estimated. There are two methods for the tail index estimation: the Pickland method (Pickland, 1975), and the Hill method (Wagner and Marsh, 2000). Firstly, the Pickland method is given by:

$$\hat{\xi}_{k,\mathrm{m}} = \frac{1}{m}\sum_{i=1}^{k}\left(\ln X_{n-i+1} - \ln X_{n-m}\right) \quad (26)$$

where m = number of observations whose tail is to be observed and k = sample size. Secondly, the Hill method is given by:

$$\hat{\xi}_{k,T} = \frac{1}{k}\sum_{i=1}^{k}\left(\ln R_{i,T} - \ln R_{k,T}\right) \quad (27)$$

where $R = \sigma Z$; recall that σ is the estimated population standard deviation and Z is the standard score of the series. Both methods follows the same conditions in providing the decision rule for classifying the type of extreme value distribution: $Frechet = \xi > 0$, $Weibull = \xi < 0$ and $Gumbel = \xi = 0$. Both methods of estimating the tail index were used.

4.0 FINDINDS AND DICUSSION

This section is comprised of three parts: sub-section 4.1 and 4.2 points out some spurious conclusion in the current literature, and 4.3 presents the general findings of this meta-analysis study.

In this meta-analysis study, it was found that there are at least two publications showing spurious findings. Fletschner involves a Type I error where the evidence points to non-significant findings, the author concludes that there is a significant gender difference in risk behavior. A second case comes from a meta-analysis study in which the author review 28 published articles which 20 Cohen's *d* and 13 SI index were reported (Nelson, 2012). The author reported that evidence is inclusive on the issue of gender difference in risk behavior. A re-analysis in this paper shows that there is conclusive empirical evidence to reach a conclusion. Nelson is a case of Type II error.

4.1 Spurious Findings Warranting Type I Error

In this meta-analysis study, it has been found that there are instances the findings contradict the evidence. It could be summarized that there are three lines of faulty findings in prior study results on gender difference on risk behavior: (i) no gender difference, (ii) gender difference, and (iii) wrong conclusion. The first two types may be unified and explained by confounding factors, such as cultural beliefs or stereotypes. Conclusions reached in these (i) and (ii) lines of research may conclusively be classified as Type I inferential error. Equally interesting is the third line of research which reached wrong conclusions. This third group of literature may be classified as Type II inferential error.

An illustrative case comes from a field study by three American researchers who conducted a study in Vietnam involving 500 rural couples (Fletschners et al., 2000). The study concludes that: "… we find that women are more risk averse than men, compare to men, women are less likely to choose to compete, irrespective of how they are likely to succeed." (*id.*, p. 1459). However, a re-examination of the data leads to an opposite conclusion. Table 1 (Willingness to Take Risks) in that study is reproduced below for illustration.

	Men	Women
Optimistic about the future	81.7%	75.5%
Worry about health problems	16.4%	19.3%
Worry about low yield because of weather	63.5%	74.0%
Worry about pests	24.6%	29.8%
Worry about livestock disease	28.3%	29.7%

Worry about low output prices	15.0%	12.7%

Using the paired means difference test, it is verified that there is no statistical difference between men and women. The paired means difference test is given by:

$$T_d = \frac{\bar{d}}{S_d / \sqrt{n}} \tag{28}$$

where $\bar{d} = -1.92$; $S = 5.82$; $n = 6$. The result of the T-test with degrees of difference of $df = n - 1 = 6 - 1 = 5$ at 95% confidence interval is -0.81. Compared to the standard t-score $T(df, 0.95) = 2.02$, the observed t-score falls far short of being significant. A F-test could also verify whether the two groups (men and women) are statistically different:

$$F = \frac{S_1^2}{S_2^2} \tag{29}$$

Since there are six evaluating factors, the degree of freedom is 5; the observed value of the F-test is 1.01 compared to the theoretical value of 5.05. Again, the result of the test shows that there is no statistically significant difference between the two groups. To conclude that there is a significant difference between men and women in the "willingness to take risk" would be a clear evidence of Type I error, i.e. insisting that the alternative hypothesis is correct when empirical test result shows otherwise. This error in calculation is confirmed by the Similarity Index (4).

The result of SI analysis shows that six items in Fletschners' Table 1 has SI value of 0.97, 0.99, 0.95, 0.98, 0.99 and 0.99 respectively; men and women are not different in their "willingness to take risk." These calculations (28), (29) and (4) did not appear in Fletschners' analysis.

4.2 Spurious Findings Warranting Type II Error

In Nelson, 20 Cohen's *d* and 13 SI were presented from 28 published articles (Nelson, 2012). While Nelson maintained that "[i]n regards to issues of risk, it is argued that exaggerated and stereotyped beliefs in the existence of sex-based differences may lead to suboptimal results in economic efficiency and equity." (*Id.*, p. 29). She was unable to conclude whether there is a definite answer from the review of 28 articles. She wrote that the issue of gender difference in risk behavior "cannot be empirically proven or disapproven." (*id.*). Nelson is a clear case of Type II error. Type II error occurs when the evidence is presented in a clear direction, but the researcher is does not know how to conclude. The analysis of Cohen *d* and SI index involves two steps.

Firstly, the Cohen's *d* array consisting of 20 counts should be checked for evidence of statistical significance through use of standard score. From 20 Cohen's *d* cited in the study, there are 2 instances of statistical significance. The question needed to be answer is: "do these two instances of significant finding represent significant probability?" This question may be answer by the Z-binary test:

$$Z_{bin} = \frac{\frac{X}{n} - p}{\sqrt{\frac{pq}{n}}} \tag{30}$$

The reference Z-critical at 0.95 confidence interval is 1.65; the observed value is $Z = 0.47$ which represents a probability of 68.10%. Thus, there is no statistical significant among the 20 items where the significant finding consists of 2 counts. This same test was used to analyze the SI

index where there is only one instance of statistical significance among 13 counts of SI values. The observed Z-score for the SI index is $Z = -0.60$ or 25.78% probability which is far less than the threshold value of 95%.

Secondly, the Cohen's *d* may be compared with the obverse of SI index. Since the SI index measures the similarity, the obverse of SI or 1 – SI is the measure for the difference. Using the T-test for two population means, the Cohen's *d* and the reciprocal of SI index could verify whether they are statistically different. The rationale of this test is to verify whether the two indicators convey the same information. T-test for two population means is given by:

$$T = \frac{\left(\bar{X}_d - \bar{X}_{si-1}\right) - \left(\mu_d - \mu_{si-1}\right)}{S\sqrt{\frac{1}{n_d} + \frac{1}{n_{si-1}}}} \tag{31}$$

where … $$SS_d = \sum_{i=1}^{n_1}\left(X_{di} - \bar{X}_d\right) = SS_1 \tag{31.1}$$

$$SS_{si-1} = \sum_{i=1}^{n_2}\left(X_{i(si-1)} - \bar{X}_2\right) = SS_2 \tag{31.2}$$

$$S^2 = \frac{\left(SS_1 + SS_2\right)}{\left(n_1 - n_2 - 2\right)} \quad \text{and} \quad S = \sqrt{S^2} \tag{31.3}$$

The Cohen's *d* is used as X_{1i} and the differenced interpolated from SI index: $(1 - SI)$ is used as X_{2i}. The observed T-score under (31) is 1.24 while the standard T-score for the two arrays at 0.95 confidence interval is 1.70. The result shows that the difference under the two indicators is not statistically significant. This finding contradicts Nelson who looked at the same data and concluded that the result is inconclusive. The conclusion reached in Nelson is a Type II error.

4.3 Findings from the Meta-Analysis of the Present Study

4.3.1 Minimum Sample Size

In meta-analysis, the literature does not cover minimum sample size as in individual studies. However, in meta-analysis, each publication selected comprises the element of the sample to be used for the analysis; therefore, the issue of minimum sample size should not be ignored. The fact that some publications would be left out makes the issue of minimum sample size (minimum number of publications needed to be selected) equally important.

Meta-analysis literature focuses on power per study and power per meta-analysis instead of sample size of the individual publication or sample-studies size of the meta-analysis. Turner et al., for instance, suggests that the calculation for power per study and per meta-analysis should be as follows:

$$POWER_{study} = \Phi\left(\sqrt{\frac{n_i\left(\tilde{p}_{j_0} - \theta_R\tilde{p}_{j_0}\right)}{\tilde{p}_{j_0}\left(1 - \tilde{p}_{j_0}\right) + \theta_R\tilde{p}_{j_0}\left(1 - \theta\tilde{p}_{j_0}\right)}}\right) - C_{\alpha/2} \tag{32}$$

… where $\tilde{p}_{j_0}$ is a fixed baseline event (higher median value); θ_R is relative risk; Φ is the cumulative standard normal distribution function and $C_{\alpha/2} = \Phi^{-1}\left(1 - \alpha/2\right)$. The decision rule is that acceptable power per study must score $\geq 50\%$.

$$POWER_{MA} = 1 - \Phi\left(C_{\alpha/2} - \frac{\delta}{\sqrt{V_j}}\right) + \Phi\left(-C_{\alpha/2} - \frac{\delta}{\sqrt{V_j}}\right) \quad (33)$$

… where Φ is the cumulative standard normal distribution function and $C_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$ (Turner, 2013 citing Hedges and Piggot, 2001: 203-17).

However, in both (32) and (33) the assumption of "normal distribution" in individual study and meta-analysis study may not be reasonable if the data are actually distributed non-normally. It would be more acceptable if the argument had been modified so that Φ is the cumulative probability distribution of whatever type of distribution so verified by Hill's tail index according to the value of ξ ((27) *supra*).

In this study, sample-study size, not power, is used. Following the n-omega method (1), the minimum publication counts required for a meta-analysis may be determined by:

$$n_{\omega} = \sqrt{\frac{\omega}{2}}$$

where $\omega = (n_3 / 0.01) - (n_3 / 0.99)$, $n_3 = \sqrt{(n_2 - n_1)/2}$, $n_2 = Z^2\sigma^2 / E^2$, and $n_1 = Z\sigma / E$. The values for Cohen's *d*, SI Index and 1 – SI were used as the initial sample size. The estimated minimum sample sizes are 13.78, 13.78 and 17.71, respectively; the mean is 15.09 ± 2.27. The sample size used in this research is $n = 20$ publications.

4.3.2 Extreme Value Analysis

As part of preliminary data test, each data set was verified if there were any extreme values. Extreme value is defined as values whose probabilities lie outside of the specified confidence interval. In the paper, a confidence interval of 95% is used. Extreme value test shows that both Cohen's d and SI index series show extreme index. Once extreme values were verified, Extreme Value Analysis (EVA) was used with the application of Pickland and Hill methods for tail index estimation.

Table 2: Estimation of the tail index (ξ) to classify Extreme Value Distribution

Data Series	Pickland Method	Hill Method	Distribution Type
Cohen's *d*	-0.9700	-0.3900	*Weibull* = $\xi < 0$
1 – SI	-0.0016	-0.5200	*Weibull* = $\xi < 0$

After it was concluded that both series were Weibull distributed, Webull's shape analysis is used via β. The standard rule for interpretation of β is that: if $\beta > 1$ there is an increasing rate of failure with respect to time; if $\beta < 1$, there is an decreasing rate of failure with respect to time; and if $\beta = 1$ the failure rate of the process is stabilized with neither increase nor decrease with respect to time. Success in Cohen's *d* means gender difference in risk behavior and failure means no gender difference. In the present case, the Cohen's *d* is greater than 1.00; it means that the process manifests an increasing failure which means that there is a diminishing gender difference in risk behavior. This conclusion is confirmed by the fact that the immediate failure rate for Cohen's *d* is $H(t) > 1.00$ and zero system reliability. In contrast, the Similarity Index shows a decreasing failure with respect to time. The beta for SI series is less than zero. For SI series, success is defined as similarity in gender risk behavior (or no gender difference). Failure means there is gender difference. In the present case, the value $\beta < 1$ means that the gender difference, if any, decreases

with time without failure. Both series were analyzed under Weibull distribution model. The Weibull statistics are summarized in the table below.

Table 3: Weibull statistics for Cohen's *d* and obverse of SI

	β	η	*CDF*	*PDF*	$H(t)$	$S(t)$	*R*
Cohen's *d*	2.11	0.55	0.63	1.40	2.81	0.06	0.37
1 – SI*	1.75	0.12	0.63	5.56	14.10	0.00	0.37
SI	19.46	0.92	0.63	7.76	20.10	0.00	0.37

*SI measure the similarity of sameness and 1 – SI measures the difference.

The findings from Cohen's *d* and SI series reconcile conflicting findings of prior studies in individual publications and meta-analysis works on the issue of gender difference in risk behavior. Objectively, when analyzing the effect size as a system, stripping away the cultural nuances and stereotypes, gender difference in risk behavior becomes a nullity. The Cohen's *d* has an observed mean of 0.50 and estimated population mean of $\mu = 0.38$. These numbers tell us that there is no significant tendency for gender difference. Out of 20 studies, there were only two instances where the probability of *d* exceeds 0.95. Were these two incidences significant? This question may be answered by:

$$Z_{bin} = \frac{\frac{X}{n} - p}{\sqrt{\frac{pq}{n}}}$$

where Z = observed critical value to be compared with $Z(0.95) = 1.65$; X = number of incidents to be tested for significance; p = probability of success defined under Laplace Rule of Succession as $p = (s+1)/(n+2)$; $q = 1 - p$; and n = sample size. The calculation follows:

$$Z_{bin} = \frac{\frac{X}{n} - p}{\sqrt{\frac{pq}{n}}} = \frac{\frac{2}{20} - 0.14}{\sqrt{\frac{0.14(0.96)}{20}}} = \frac{0.10 - 0.14}{\sqrt{\frac{0.1344}{20}}} = \frac{-0.04}{\sqrt{0.00672}}$$

$$Z_{bin} = \frac{-0.04}{0.0819} = 0.4884 \quad \text{or} \quad \cong 68.8\%$$

The confidence interval is 95%, the observed probability is about 50%. The prior meta-analysis should have alerted the research community that among 20 studies of Cohen's *d*, the evidence to prove that there is gender difference in risk behavior has no statistical significance.

Does the probability of 68.8% convey useful information? By employing the Boltzmann *H* theorem, we can verify whether the information conveyed was valuable (Borda, 2011; Han, 2002). Under information entropy theory, the Boltzmann *H* may be determined by:

$$H(X) = E[-\ln P(X))] \tag{34}$$

The average information value is: $H(X) = -\ln(0.54) = -(-0.69) = 0.69$. A guess work is said to be 50/50. A research result should convey more information than blind guessing. In this case, the result of prior publications on gender difference in risk behavior contributed 0.37 points which is low compare to a heighten standard of 0.95 or in a more relaxed standard of 0.80 under 80/20 rule. The current state of literature on the subject is far from Pareto efficient (Bar, 2012). Yet, this

apparent inefficiency would most likely not be improved because the values of the effect size in these studies are not statistically significant.

Hedges and Pigott wrote that: "… a statistically significant finding in a previous review is not necessarily an indication of adequate power in a later meta-analysis conducted for a difference purpose." (Hedges and Pigott, p. 205). This is indeed applicable in the present case. The apparent significant findings of prior studies had been used for re-analysis; it is found that even had the Cohen's *d* in each study had been significant, as a group this significance disappears. With such a disappearance, a generalization of prior studies could be made: they were false positive readings (Type I error).

What is the implication of continued affirmation of gender difference in risk behavior? For existing entrepreneurs and in the training of future entrepreneurs, this insistence is perpetuating false pretense of a moot issue. If left uncorrected, it will perpetuate gender discrimination among male and female business owners. In light of a complete lack of empirical evidence supporting the claim of gender difference in risk behavior, any attempt to sustain such an argument is an attempt to torn asunder the fragile achievement we made in gender equality. Such discrimination, standing behind the façade of academic research has no place in modern society. Success in business must be built upon a solid foundation of meritocracy. The use of cultural beliefs and stereotypes as the means to maintain inequality mars social progress.

An examination of 20 studies shows that the claim for gender difference in risk behavior through the use of effect size (Cohen's *d*) is not random occurrence. This was confirmed by the adjacent test for randomness:

$$L_{n<25} = \frac{\sum_{i+1}^{n-1} (X_{i+1} - X_i)^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \qquad (35)$$

The standard reference is $1.30 < L(0.95) < 2.70$ for Cohen's *d* and $1.21 < L(0.95) < 2.79$ for SI. The null hypothesis assumes that the data is random. In the present case, the observed value is $L(obs) = 11.80$ for Cohen's *d* and $L(obs) = 30.28$; both are non-random under the Adjacent Test (34). For reference table of (34), see Appendix A4.

The result of the test shows that the gender difference is non-random and it is not statistically significant. This result implies that the non-random gender difference suggests that the claimed gender difference resulted from structural design in society or organization. It is not a random process; this is "nurture." Second, the gender difference resulted from this social structure does not succeed in producing statistically significant difference in gender-based risk behavior. However, it is beyond the scope of this paper to conclude whether the structural design or discrimination causes injuries. In order to achieve and maintain social equity, discrimination of all kinds should not be practiced.

4.3.3 Meta-Analysis under Random Effect Approach

Most meta-analyses focus on the effect size and terminate the analysis at the effect size analysis. However, this paper uses the effect size: Cohen's *d* and SI index as the observed value for general statistical test. These Cohen's *d* and SI index were treated as "observed values" in order to verify their statistical significance. The combined effect from all selected studies was tested for their statistical significance. It was found that both Cohen's *d* and SI index series were not statistically significant. The findings of the present study may be summarized by the following two tables in the Appendix: Table A1 (Cohen's *d*) and Table A2 (1- SI).

The result of the test statistics for the Cohen's *d* of the 20 publications that appeared in the Nelson article and elsewhere shows that gender difference in risk behavior has no statistical significance. With a Z-score of 0.31, it means that the probability is less than 61.80% while the standard of review requires 95% confidence interval. This finding is determined by the verification of results from published articles. The conclusion reached here contradicts those in prior publications. The methodology adopted in this paper follows conventional statistical method in testing for statistical significance using 0.95 confidence interval.

Similar finding was made from the calculation of the difference among male and female subjects by using the SI Index as the basis. Table A2 (Appendix) uses $(1-SI)$ as the basis for the testing. The result shows that the Z-score is below what is required by the 0.95 confidence interval. At 0.95 confidence interval, the standard Z-score is 1.65, the observed value for the Z-score from the $(1-SI)$ series is 0.06 or about 52.40%. The $(1-SI)$ series confirms that there is no statistical significance in the gender difference in risk behavior. From the selected publications, both individual papers and meta-analysis studies, it has been found that the current literature is littered with defective analysis.

5.0 CONCLUSION

There are many publications on the issue of gender difference in risk behavior. A meta-analysis is used to summarize and reconcile these different findings. This paper answers the question of "whether there is gender difference in risk behavior? If so, whether such difference contributes to different behavior in male and female entrepreneurs as business owners?" The result of the meta-analysis shows that there is no significant difference between male and female in risk behavior. Many claims in prior publications were made as the result of spurious findings or failure to do proper statistical analysis. The current literature in this area of gender research is littered with Type I and II errors. Having found no statistical significance as the answer to the first research question, the second research question becomes moot by default. This paper contributes to the literature by helping to reconcile and unify contradictory findings on this research topic. The practical implications of this paper are two folds. First, having found that there is no gender difference in risk behavior, the issue should become moot. Second, since there is no significant gender difference among male and female entrepreneurs, the beliefs and stereotypes towards male and female should cease.

REFERENCE

Altman D.G. (1994). The scandal of poor medical research. *B.M.J.* 308: 283. doi: 10.1136/bmj.308.6924.283.

Anderson, T.W. and Darling, D.A. (1954). "A Test of Goodness-of-Fit." *Journal of the American Statistical Association* **49**: 765–769.

Ards, Sheila; Chung, Chanjin; Myers, Samuel L. (1998). "The effects of sample selection bias on racial differences in child abuse reporting." *Child Abuse & Neglect* **22** (2): 103–115.

Barr, Nicholas (2012). "3.2.2 The relevance of efficiency to different theories of society." *Economics of the Welfare State* (5th ed.). Oxford University Press. p. 46. ISBN 978-0-19-929781-8.

Bensalah, Younes (2000). Steps in Applying Extreme Value Theory to Finance: Review. Working Paper 2000-20/Document de travail 2000-20. Bank of Canada; pp. 2-12.

Borda, Monica (2011). *Fundamentals in Information Theory and Coding*. Springer. p. 11. ISBN 978-3-642-20346-6.

Born, Marise; Nico Bleichrodt, and Henk van der Flier (1987). "Cross-Cultural Comparison of Sex-Related Differences on Intelligence Tests: A Meta-Analysis," *Journal of Cross-Cultural Psychology* 18, pp.283-314.

Charness G., and Gneezy, U. (2012). "Strong Evidence for Gender Differences in Risk Taking." *Journal of Economic Behavior & Organization*. Volume 83, Issue 1, June 2012, pp. 50–58.

Cliff, Norman (1993). "Dominance statistics: Ordinal analyses to answer ordinal questions." *Psychological Bulletin* **114** (3): 494.

Cortes, Corinna; Mohri, Mehryar; Riley, Michael; Rostamizadeh, Afshin (2008). "Sample Selection Bias Correction Theory." *Algorithmic Learning Theory* **5254**: 38–53.

Cortes, Corinna; Mohri, Mehryar (2014). "Domain adaptation and sample bias correction theory and algorithm for regression." Theoretical Computer Science 519: 103–126.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (second ed.). Lawrence Erlbaum Associates. p. 67.

Cohen, J. (1992). "A power primer." *Psychological Bulletin* **112** (1): 155–159.

Cohn, Lawrence D. (1991). "Sex Differences in the Course of Personality Development: A Meta-Analysis," *Psychological Bulletin* 109, pp. 252-266.

Cooper, H. M. (1979). "Statistically Combining Independent Studies: A Meta-Analysis of Sex Differences in Conformity Research" *Journal of Personality and Social Psychology* 37:1, pp. 131-146.

Deeks J (1998). "When can odds ratios mislead? : Odds ratios should be used only in case-control studies and logistic regression analyses." *BMJ* **317** (7166): 1155–6.

Drucker, Peter F. (1985). *Innovation and Entrepreneurship*. attributes the coining and defining of "entrepreneur" to Jean-Baptiste Say in his *A Treatise on Political Economy*; (1834).

Dindia, Kathryn and Michael Allen (1992). "Sex Differences in Self-Disclosure: A Meta-Analysis," *Psychological Bulletin* 112, pp. 106-124.

Eagly, Alice and Linda Carli (1981). "Sex of Researchers and Sex-Typed Communications as Determinants of Sex Differences in Influenceability: A Meta-Analysis of Social Influence Studies," *Psychological Bulletin* 90, pp. 1-20.

Eagly, Alice and M. Crowley (1986). "Gender and Helping Behavior: A Meta-Analytic Review of the Social Psychological Literature," *Psychological Bulletin* 100, pp. 283-308.

Eagly, Alice and Blair T. Johnson (1990). "Gender and Leadership Styles: A Meta-Analysis," *Psychological Bulletin* 108, pp. 233-256.

Eagly, Alice and S. J. Karau (1991). "Gender and the Emergence of Leaders: A Meta-Analysis," *Journal of Personality and Social Psychology* 60, pp. 685-710.

Eagly, Alice and V. J. Steffen V. J. (1986). "Gender and Aggressive Behavior: A Meta-Analytic Review of the Social Psychology Literature," *Psychological Bulletin* 100, pp. 309-330.

Eagly, Alice and Wendy Wood (1991). "Explaining Sex Differences in Social Behavior: A Meta-Analytic Perspective," *Personality and Social Psychololgy Bulletin* 17, pp. 306-315.

Eagly, Alice; S. J. Karau and M. G. Makhijani (1995). "Gender and the Effectiveness of Leaders: A Meta-Analysis," *Psychological Bulletin* 117, pp. 125-145.

Eagly, Alice; M. G. Makhijani and B. G. Klonsky (1992). "Gender and the Evaluation of Leaders: A Meta-Analysis," *Psychological Bulletin* 111, pp. 3-22.

Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test.BMJ1997;315:629-34.

Ellis, Paul D. (2010). *The Essential Guide to Effect Sizes: An Introduction to Statistical Power, Meta-Analysis and the Interpretation of Research Results*. United Kingdom: Cambridge University Press.

Feingold, Alan. (1994). "Gender Differences in Personality: A Meta-Analysis," *Psychological Bulletin* 116, pp. 429-456.

Fletschner, D., Anderson, C. L., and Cullen, A. (2000). "Are Women as Likely to Take Risk and Compete? Behavioral Findings from Central Vietnam." Journal of Development Studies, **46**: 1459-1479.

Greenland, S. (2004). "Model-based Estimation of Relative Risks and Other Epidemiologic Measures in Studies of Common Outcomes and in Case-Control Studies.' *American Journal of Epidemiology* **160** (4): 301–305.

Grissom R.J. (1994). "Statistical analysis of ordinal categorical status after therapies." *Journal of Consulting and Clinical Psychology* **62** (2): 281–284.

Han, Te Sun & Kobayashi, Kingo (2002). Mathematics of Information and Coding. American Mathematical Society. pp. 19–20. ISBN 978-0-8218-4256-0.

Hartung, J., Knapp, G. and Sinha, B.K. (2008). *Statistical Meta-Analysis with Application*. Hoboken, New Jersey: Wiley. p. 14.

Helfenstein U. (2002). "Data and models determine treatment proposals--an illustration from meta-analysis." *Postgrad Med. J.* **78** (917): 131–4.

Hyde J. S. (1981). "How Large are Cognitive Gender Differences? A Meta-Analysis using ω^2 and d." *American Psychologist* 36, pp. 892-901.

Hyde J. S. (1984). "How Large are Gender Differences in Aggression? A Developmental Meta-Analysis," *Developmental Psychology* 20, pp. 722-736.

Hyde J. S. and Marcia C. Linn. (1988). "Gender Differences in Verbal Ability: A Meta-Analysis," *Psychological Bulletin* 104, pp. 53-69.

Hyde J.S., E. Fennema and S. J. Lamon (1990). "Gender Differences in Mathematics Performance: A Meta-Analysis," *Psychological Bulletin* 107, pp. 139-155.

Ichino, A. and Moretti, E. (2009). "Biological Gender Differences, Absenteeism, and the Earnings Gap." *American Economic Journal: Applied Economics*; 1:1, 183–218.

Ikhlaaq A., Sutton, A.J., and Riley, R.D. (2012). "Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey." *B.M.J.* 2012; 344: d7762.

Johnson, Blair T. and Eagly, Alice. (1989). "Effects of Involvement on Persuasion: A Meta-Analysis," *Psychological Bulletin* 106, pp. 290-314.

Johnson, J. E. V. and P. L. Powell. (1994). "Decision Making, Risk and Gender: Are Managers Different?" *British Journal of Management* 5, pp. 123-138.

Kahneman, Daniel; Tversky, Amos (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* **47** (2): 263.

Kelley, Ken; Preacher, Kristopher J. (2012). "On Effect Size." *Psychological Methods* **17** (2): 137–152.

Kjaergard L.L., Villumsen J., Gluud C. (2001). Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Annals of Internal Medicine* 135: 982–9. doi: 10.7326/0003-4819-135-11-200112040-00010.

Knight, Frank Hyneman (2005). *Risk, Uncertainty and Profit*. Cosimo, Inc. ISBN 978-1-59605-242-0.

Kobus, K. (2003). "Peers and adolescent smoking." *Addiction*, 98 (Suppl 1), 37–55.

LeLorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F (1997). "Discrepancies between Meta-Analyses and Subsequent Large Randomized, Controlled Trials." *New England Journal of Medicine* **337** (8): 536–542.

Lewis, R.J. and Sauro, J. (2006). "When 100% Really Isn't 100%: Improving the Accuracy of Small-Sample Estimates of Completion Rates." *JUS*, Issue 1, vol. 1, May 2006, pp. 136-150.

Louangrath, P.T.I. (2014). "Sample Size Determination in Nonfinite Population." *International Conference on Discrete Mathematics & Applied Sciences.* 21st - 23rd May 2014. University of the Thai Chamber of Commerce, Bangkok, Thailand. ICDMAS 2014. Book of Abstract, Contributed Talk 2.
doi: 10.13140/2.1.4785.7286.

Louangrath, P.T.I. and Rewtrakunpaiboon W. (2013). "Determination of minimum Sample Size for Film-Induced Tourism Research." *Silapkorn 70th Anniversary International Conference 2013. Towards the Next Decade of Hospitality & Creative economics: Looking Forward to 2020.* 1st – 3rd December 2013. Bangkok, Thailand. Conference Proceedings; pp. 127-139. doi: 10.13140/2.1.2090.5605

McGrath, R.E. and Meyer, G.J. (2006). "When Effect Sizes Disagree: The Case of r and d." *Psychological Methods* **11** (4): 386–401.

McGraw K.O., Wong S.P. (1992). "A common language effect size statistic." *Psychological Bulletin* **111** (2): 361–365.

Mincer, Jacob, and Solomon Polachek (1974). "Family Investment in Human Capital: Earnings of Women." *Journal of Political Economy*, 82(2): S76–108.

Moreno SG, Sutton AJ, Ades AE, Stanley TD, Abrams KR, et al.. (2009) Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC Medical Research Methodology 9(2).

Morris, S.B. and DeShon, R.P. (2002). "Combining Effect Size Estimates in Meta-Analysis With Repeated Measures and Independent-Groups Designs." *Psychological Methods*. Vol. 7, No. 1, 105-125.

Nakagawa, Shinichi; Cuthill, Innes C (2007). "Effect size, confidence interval and statistical significance: a practical guide for biologists." *Biological Reviews Cambridge Philosophical Society* **82** (4): 591–605.

Nunnally, J. C. (1967). "Psychometric Theory." *McGraw-Hill, New York*: 355.

Nygard O., Vollset S.E., Refsum H., Stensvold I., Tverdal A., et al. (1995). Total plasma homocysteine and cardiovascular risk profile. *J.A.M.A.* 274: 1526–33. doi: 10.1001/jama.1995.03530190040032.

Olejnik, S. and Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs *Psychological Methods*. 8:(4)434-447.

Pawlowski, B., Atwal, R., and Dunbar, R.I.M. (2008). "Sex Differences in Everyday Risk-Taking Behavior in Humans." *Evolutionary Psychology*, **6** (1): 29-42.

Picklands, J. (1975) "Statistical Inference using extreme order statistics." *Annals of Statistics*, 3,119-131.

Robert E. McGrath and Gregory J. Meyer (2006). "When Effect Sizes Disagree: The Case of r and d." *Psychological Methods* **11** (4): 386–401.

Rosenthal R (1979). "The 'File Drawer Problem' and the Tolerance for Null Results." *Psychological Bulletin* **86** (3): 638–641.

Rothstein HR, Sutton A.J., Borenstein M (eds) (2005). *Publication bias in meta-analysis*. Wiley.

Rucker G, Carpenter JR, Schwarzer G (2011) Detecting and adjusting for small-study effects in meta-analysis. Biometrical Journal 53: 351–68. doi: 10.1002/bimj.201000151.

Schumpeter, Joseph Alois (1976). *Capitalism, Socialism and Democracy*. Routledge. ISBN 978-0-415-10762-4.

Schumpeter, Joseph Alois (1934). *The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle*. Transaction Publishers. ISBN 978-0-87855-698-4.

Senn S. (2007). "Trying to be precise about vagueness." *Stat. Med.* **26**: 1417–30.

Slavin RE (1986). "Best-Evidence Synthesis: An Alternative to Meta-Analytic and Traditional Reviews." *Educational Researcher* **15** (9): 5–9.

Shane, S. (2000). "The Promise of entrepreneurship as a Field of research." *Academy of Management Review* **25** (1): 217–226.

Stanley T.D., Jarrell S.B., Doucouliagos H. (2010). Could it be better to discard 90% of the data? A statistical paradox. *The American Statistician* 64: 70–7. doi: 10.1198/tast.2009.08205.

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods* 9:(2) 164-182.

Stephens, M. A. (1974). "EDF Statistics for Goodness of Fit and Some Comparisons." *Journal of the American Statistical Association* **69**: 730–737.

Sterne J.A, Egger M, Smith G.D. (2001). Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. BMJ2001; 323:101-5.

Sterne J.A.C., Gavaghan D., Egger M. (2000). Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology* 53: 1119–29. doi: 10.1016/s0895-4356(00)00242-0.

Turner R.M., Bird S.M., Higgins J.P.T. (2013). The Impact of Study Size on Meta-analyses: Examination of Underpowered Studies in Cochrane Reviews. *PLoS ONE* 8(3): e59202. doi:10.1371/journal.pone.0059202.

Ugrinowitsch, C., G.W. Fellingham, and M.D. Ricard. (2004). "Limitations of Ordinary Least Squares Models in Analyzing Repeated Measures Data." *Med. Sci. Sports. Exerc.*, Vol. 36, No. 12, pp. 2144–2148.

Wagner, N., and Marsh, T. (2000). "On Adaptive Tail Index Estimation for Financial Return Models." Research Program in Finance Working Paper RPF-295; pp. 5-8.

Weibull, W. (1951), "A statistical distribution function of wide applicability." *J. Appl. Mech.-Trans. ASME* **18** (3): 293–297.

Westland, J. Christopher (2010). "Lower bounds on sample size in structural equation modeling." *Electron. Comm. Res. Appl.* **9** (6): 476–487.

Wilkinson, Leland (1999). APA Task Force on Statistical Inference. "Statistical methods in psychology journals: Guidelines and explanations." *American Psychologist* **54** (8): 594–604.

Yamane, Taro (1967). *Statistics: An Introductory Analysis*, 2nd Ed., New York: Harper and Row; p. 886.

Zhang, J.; Yu, K. (1998). "What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes." *J.A.M.A.: the Journal of the American Medical Association* **280** (19): 1690–1691.

APPENDIX

Table A1: Test of Significance for Cohen's *d*

Item	d	n_d	w_d	T_i	$\bar{T}$	$w_i(T_i-\bar{T})^2$
1	0.26	8,000.00	0.03	0.91	0.50	0.0043
2	0.46	200.00	0.05	1.11	0.50	0.0169
3	0.38	300.00	0.04	1.03	0.50	0.0105
4	0.55	300.00	0.05	1.20	0.50	0.0266
5	1.45	150.00	0.14	2.10	0.50	0.3671
6	0.48	22,000.00	0.05	1.13	0.50	0.0188
7	1.13	300.00	0.11	1.78	0.50	0.1830
8	0.22	200.00	0.02	0.87	0.50	0.0030
9	0.49	100.00	0.05	1.14	0.50	0.0198
10	0.33	1,200.00	0.03	0.98	0.50	0.0075
11	0.74	700.00	0.07	1.39	0.50	0.0579
12	0.29	13,000.00	0.03	0.94	0.50	0.0055
13	0.37	200.00	0.04	1.02	0.50	0.0098
14	0.36	2,000.00	0.04	1.01	0.50	0.0092
15	0.85	150.00	0.08	1.50	0.50	0.0839
16	0.65	200.00	0.06	1.30	0.50	0.0410
17	0.17	100.00	0.02	0.82	0.50	0.0017
18	0.31	400.00	0.03	0.96	0.50	0.0064
19	0.44	50.00	0.04	1.09	0.50	0.0151
20	0.16	6,000.00	0.02	0.81	0.50	0.0015
$Q = \sum w_i - \left(T_i - \bar{T}\right)^2$		***0.89***				
$C = \sum w_i - \left(w_i^2 / \sum w_i\right)$		***19.93***				
$\tau^2 = (Q - df) / C = 0.91$		***0.91***				
$V = S^2 + \tau^2$		***1.47***				
$SE(\bar{T}_*) = \sqrt{V_*}$		***1.21***				
$Z = \bar{T}_* / SE(\bar{T}_*)$		***0.31***				

Table A2: Test of Significance for the Gender Difference through *1 - SI*

Item	$1-SI$	n_{si}	w_{si}	T_i	$\bar{T}$	$w_i(T_i-\bar{T})^2$
1	0.02	12,000.00	0.02	0.12	0.10	0.0000
2	0.09	200.00	0.07	0.19	0.10	0.0006
3	0.13	300.00	0.10	0.23	0.10	0.0018
4	0.16	300.00	0.12	0.26	0.10	0.0033
5	0.12	22,000.00	0.09	0.22	0.10	0.0014
6	0.20	300.00	0.15	0.30	0.10	0.0063
7	0.09	200.00	0.07	0.19	0.10	0.0006
8	0.07	120.00	0.05	0.17	0.10	0.0003
9	0.04	13,000.00	0.03	0.14	0.10	0.0001
10	0.14	2,000.00	0.11	0.24	0.10	0.0022
11	0.14	200.00	0.11	0.24	0.10	0.0022
12	0.07	100.00	0.05	0.17	0.10	0.0003
13	0.04	6,000.00	0.03	0.14	0.10	0.0001

$Q = \sum w_i - \left(T_i - \bar{T}\right)^2$	***0.02***
$C = \sum w_i - \left(w_i^2 / \sum w_i\right)$	***12.90***
$\tau^2 = (Q - df) / C = 0.91$	***1.47***
$V = S^2 + \tau^2$	***1.70***
$SE(\bar{T}_*) = \sqrt{V_*}$	***1.30***
$Z = \bar{T}_* / SE(\bar{T}_*)$	***0.06***

Table A3: List of Prior Publications

No.	Author(s)	Cohen'd	SI	n
1	Areno *et al*. (2010)	NSS	-	400
2	Barber and Odean (2001)	-0.09 – 0.26	-	38,000
3	Barsky, Juster, *et al*. (2008)	-	0.98	12,000
4	Beckmann and Menkhoff (2008)	NSS – 0.46	0.67 – 0.91	200
5	Bernasek and Schwiff (2001)	NSS	0.87	300
6	Booth and Nolen (2012)	NSS – 0.38	0.84	300
7	Borhans, Golsteyn, *et al*. (1999)	0.32 – 0.55	-	300
8	Byrnes, Niller, *et al*. (1999)	-1.23 – 1.45 Mean 0.13	-	-
9	Dohmen, Falk, *et al*. (2011)	NSS – 0.48	0.80 – 0.88	500 – 22,000
10	Eckel and Grossman (2008)	0.55 – 1.13	0.60 – 0.80	300
11	Eriksson and Simpson (2010)	0.19 – 0.22	0.89 – 0.91	200
12	Fehr-Duda, De Gennaro, *et al*. (2000)	-0.25 – NSS – 0.49	-	100
13	Finucane, Slovic, *et al*. (2000)	0.11 – 0.33	0.86 – 0.93	1,200
14	Harris, Jenkins, *et al*. (2006)	-0.34 – NSS – 0.74	-	700
15	Hartog, Ferreri-Carbonell, *et al*. (2002)	0.22 – 0.29	0.85 – 0.96	1,500 – 13,000
16	Holt and Laury (2002)	NSS – 0.37	0.83 – 0.86	200
17	Kahan, Brahman, *et al*. (2007)	0.15 – 0.36	-	2,000
18	Linquist and Sav-Soderbergh (2011)	NSS	-	600
19	Meier-Pesti and Penz (2008)	NSS – 0.85	-	150
20	Olsen and Cox (2001)	NSS – 0.65	0.60 – 0.86	200
21	Powell and Ansic (1997)	0.06 – 0.17	0.90 – 0.93	100
22	Rivers, Arvai, *et al*. (2010)	0.25 – 0.31	-	400
23	Ronay and Kim (2006)	NSS – 0.44	-	50
24	Hartog, Ferreri-Carbonell, *et al*. (2002)	0.08 – 0.16	0.95 – 0.96	6,000
25				
26				
27				
28				
29				
30				

APPENDIX: A3
Effect Size under Cohen's *d*

There are various approaches to calculating Cohen's *d* depending on the circumstances and the availability of data. These methods have been provided by Thalheimer and Cook (2002: 1-9). These various approaches may be summarized below as t-test and F-test approaches to Cohen's *d* effect size calculation.

(1) If the treatment and control arrays are known and their respective variances are give, the Cohen's *d* is reduced to the mean difference comparison study, thus:

$$d_t = \frac{\bar{X}_t - \bar{X}_c}{S_{pooled}} \qquad \text{(A1)}$$

… where the term S_{pooled} is the pooled standard deviation of the treatment and control groups; S_{pooled} is given by:

$$S_{pooled} = \sqrt{\frac{(n_t - 1)S_t^2 + (n_t - 1)S_c^2}{n_t + n_2}} \qquad \text{(A2)}$$

(2) If the standard deviation or standard error are not given, the Cohen's *d* may be obtained through:

$$d_t = \sqrt{t\left(\frac{n_t + n_c}{n_t n_c}\right)\left(\frac{n_t + n_c}{n_t + n_c - 2}\right)} \qquad \text{(A3)}$$

(3) If the standard error is given, but the standard deviation is not available, the value for the standard deviation may be obtained indirectly via: $S = SE\sqrt{n}$.

The Cohen's *d* obtained through A1 and A3 described above is known as the *t-test* method. The second approach is known as the *F-test* method because it uses the F-test. In case where the mean of the treatment and control groups are known, the F-test approach for Cohen's *d* is given by:

$$d_F = \frac{\bar{X}_t - \bar{X}_c}{\sqrt{MSE\left(\frac{n_t + n_c - 2}{n_t + n_c}\right)}} \qquad \text{(A4)}$$

where the mean square error (MSE) is defined as: $MSE = SSE / n - 1$ and $SSE = \sum\left(Y_i - \hat{Y}\right)^2$. In case where the MSE is not available, the F-test for Cohen's *d* is obtained through the frequency count, thus:

$$d_F = \sqrt{F\left(\frac{n_t + n_c}{n_t n_c}\right)\left(\frac{n_t + n_c}{n_t + n_c - 2}\right)} \qquad \text{(A5)}$$

See Thalheimer, W., & Cook, S. (2002, August). *How to calculate effect sizes from published research articles: A simplified methodology*. Retrieved June5, 2015 from http://work-learning.com/effect_sizes.htm.

Cf. Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159; Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods, 1*, 331-340; and Rosnow, R. L., Rosenthal, R., & Rubin, D. B. (2000). Contrasts and correlations in effect-size estimation. *Psychological Science, 11*, 446-453. [cited in Thalheimer: p.9].

APPENDIX A4

Significance Level for the Adjacent Test

	Significance Level: α			
Two-sided	0.10		0.02	
One-sided	0.05		0.01	
n	*a*	*b*	*a*	*b*
4	0.78	3.22	0.63	3.37
5	0.82	3.18	0.54	3.46
6	0.89	3.11	0.56	3.44
7	0.94	3.06	0.61	3.39
8	0.98	3.02	0.66	3.34
9	1.02	2.98	0.71	3.29
10	1.06	2.94	0.75	3.25
11	1.10	2.90	0.79	3.21
12	**1.13**	**2.87**	0.83	3.17
15	**1.21**	**2.79**	0.92	3.08
20	1.30	2.70	1.04	2.98
25	1.37	2.63	1.13	2.87

Table 1.0: The critical value of L at various significance levels. Lower bound = a and upper bound = b. Source: Hart, B.I. (1942). "Significance Level for the Mean Square Successive Difference to the Variance." *Annals of Mathematical Statistics*, **13**: 445-7.

APPENDIX 5

Selected works of prior publication used or reviewed for meta-analysis

[1] Barua, A., Davidson, L.F., Rama, D.V. and Thiruvadi, S. (2010). "CFO gender and accruals quality." *Accounting Horizons.* American Accounting Association, 24(1): 25-39.doi: 10.2308/acch.2010.24.1.25.

[2] Croson, R. and Gneezy U. (2009). "Differences in Preferences." *Journal of Economic Literature* 47: 448-74.

[3] Cross, C.P., Lee, T., and Campbell A. (2011). "Sex Differences in Impulsivity: A Meta-Analysis." *Psychological Bulletin*, American Psychological Association, 137(1): 97–130.doi: 10.1037/a0021591.

[4] Duda, H.F., Gennaro, M. and Schubert, R. (2004). "Gender, Financial Risk, ad Probability Weights." W.I.F. - *Institute of Economic Research*. Working Paper 04/31.

[5] Falaschetti, D. (2005). "A Sex-Difference in Risk-Taking and Promotions in Hierarchies: Evidence from Females in Legislatures (2012)." *Journal of Law and Economics* 55(3):477-502. Lead article.

[6] Fletschner, Diana, Leigh Anderson, and Alison Cullen. 2010. "Are Women as Likely to Take Risks and Compete? Behavioral Findings from Central Vietnam," Journal of Development Studies 46 (8): 1459-1479.

[7] Harris, C.R., Jenkins, M. and Glaser D. (2006). "Gender Differences in Risk Assessment: Why do Women Take Fewer Risks than Men?" *Judgment and Decision Making*, Vol. 1, No. 1, July 2006, pp. 48–63.

[8] Hedges, L.V. and Pigott, T.D. (2001). "The Power of Statistical Tests in Meta-Analysis." *Psychological Methods*, Vol.6, No. 3,203-217; doi: 10.1037//1082-989X.6.3.203.

[9] Hirschberger G., Florian, V., Mikulincer, M.; Goldenberg, J.L. and Pyszynski T. (2002). "Gender differences in the willingness to engage in risky behavior: A terror management perspective." *Death Studies*, 26:1177141.

[10] Johnson, J.G., Wilke, A., and Weber E.U. (2004). "Beyond a trait view of risk takig: A domain-specific scale measuring risk perceptions, expected benefits, and perceived risk attitudes in German-speaking populations." *Polish Psychological Bulletin*, 35(3): 153-163.

[11] Julie A. Nelson (2012). "Are Women Really More Risk-Averse than Men?" *Global development and Environmental Institute*; Working Paper No. 12-05.

[12] Moore, S. and Chater, N. (2003). *The influence of affect on risky behavior: From the lab to real world financial behavior*. In: 25th Annual Conference of the Cognitive-Science-Society, Boston, MA, JUL 31-AUG 02, 2003. Published in: PROCEEDINGS OF THE TWENTY-FIFTH ANNUAL CONFERENCE OF THE COGNITIVE SCIENCE SOCIETY, Pts 1 and 2 pp. 822-827.

[13] Ronay R. and Kim, D.Y. (2010).
"Gender differences in explicit and implicit risk attitudes: A socially facilitated phenomenon." British Journal of Psychology, 45(2): 397-419. doi: 10.1348/014466605X66420.

[14] Sapienza, P., Zingales, L. and Maestripieri D. (209). "Gender differences in financial risk aversion and career choices are affected by testosterone." *P.N.A.S.* 106(36): 15268-15273.

[15] Arano, Kathleen, Carl Parker, et al. (2010). "Gender-Based Risk Aversion and Retirement Asset Allocation." *Economic Inquiry* 48(1): 147-155.

[16] Barber, Brad M. and Terrence Odean (2001). "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment." *The Quarterly Journal of Economics* 116(1): 261-292.

[17] Barsky, Robert B., F. Thomas Juster, et al. (1997). "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *The Quarterly Journal of Economics* 112(2): 537-579.

[18] Beckmann, Daniela and Lukas Menkhoff (2008). "Will Women Be Women? Analyzing the Gender Difference among Financial Experts." *Kyklos* 61(3): 364-384.

[19] Bernasek, Alexandra and Stephanie Shwiff (2001). "Gender, Risk, and Retirement." *Journal of Economic Issues* 35(2): 345-356.

[20] Booth, Alison L. and Patrick Nolen (2012). "Gender Differences in Risk Behaviour: Does Nurture Matter?" *The Economic Journal* 122(558).

[21] Borghans, Lex, Bart H. H. Golsteyn, et al. (2009). "Gender Differences in Risk Aversion and Ambiguity Aversion." *Journal of the European Economic Association* 7(2-3): 649-658.

[22] Dohmen, Thomas, Armin Falk, et al. (2011). "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9(3): 522-550.

[23] Eckel, Catherine C. and Philip J. Grossman (2002). "Sex differences and statistical stereotyping in attitudes toward financial risk." *Evolution and Human Behavior* 23: 281-295.

[24] Eckel, Catherine C. and Philip J. Grossman (2008). "Forecasting risk attitudes: An experimental study using actual and forecast gamble choices." *Journal of Economic Behavior & Organization* 68(1): 1-17.

[25] Eriksson, Kimmo and Brent Simpson (2010). "Emotional reactions to losing explain gender differences in entering a risky lottery." *Judgment and Decision Making* 5(3): 159-163.

[26] Finucane, Melissa L., Paul Slovic, et al. (2000). "Gender, race, and perceived risk: The 'white male' effect." *Health, Risk & Society* 2(2): 159-172.

[27] Harris, Christine R., Michael Jenkins, et al. (2006). "Gender Differences in Risk Assessment: Why do Women Take Fewer Risks than Men?" *Judgment and Decision Making* 1(1): 48-63.

[28] Hartog, Joop, Ada Ferrer-i-Carbonell, et al. (2002). "Linking Measured Risk Aversion to Individual Characteristics." *Kyklos* 55(1): 3-26.

[29] Holt, Charles A. and Susan K. Laury (2002). "Risk Aversion and Incentive Effects." *The American Economic Review* 92(5): 1644-1655.

[30] Kahan, Dan M., Donald Braman, et al. (2007). "Culture and Identity-Protective Cognition: Explaining the White-Male Effect in Risk Perception." *Journal of Empirical Legal Studies* 4(3): 465-505.