University of Denver
Digital Commons @ DU

Electronic Theses and Dissertations | Graduate Studies

1-1-2017

Meta-Analyses of the Relationship Between Depression and Nine Dimensions of Perfectionism

Gabriel Lynn Hottinger
University of Denver

Follow this and additional works at: https://digitalcommons.du.edu/etd

Part of the Applied Statistics Commons, and the Quantitative Psychology Commons

Recommended Citation

Hottinger, Gabriel Lynn, "Meta-Analyses of the Relationship Between Depression and Nine Dimensions of Perfectionism" (2017). *Electronic Theses and Dissertations*. 1368.
https://digitalcommons.du.edu/etd/1368

This Dissertation is brought to you for free and open access by the Graduate Studies at Digital Commons @ DU. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu,dig-commons@du.edu.

Meta-Analyses of the Relationship Between Depression and Nine Dimensions of Perfectionism

Abstract

Perfectionism has been shown to be related to depression, but perfectionism is multidimensional. Some dimensions are related to positive psychological characteristics and outcomes and other dimensions are related to negative psychological characteristics and outcomes. This study reports results of nine meta-analyses performed to investigate the association between each of nine subscales of perfectionism and depression to determine which dimensions of perfectionism are most strongly associated with depression. The two subscales that were used from the Hewitt and Flett (1991b) Multidimensional Perfectionism scale were Self-Oriented Perfectionism (SOP) and Socially-Prescribed Perfectionism (SPP). The five subscales that were used from the Frost et al. (1991) Multidimensional Perfectionism Scale were Personal Standards (PS), Doubts about Actions (DA), Concern over Mistakes (CM), Parental Expectations (PE), and Parental Criticism (PC). The two subscales that were used from the Slaney et al. (2001) Almost Perfect Scale-Revised were High Personal Standards (HS) and Discrepancy (Dis). The SPP, DA, CM, PE, PC, and DIS subscales are negative dimensions of perfectionism that form the higher-order factor Perfectionistic Concerns (PC). The SOP, PS, and HS subscales are more positive dimensions of perfectionism that form the higher-order factor Perfectionistic Strivings (PS). Knowing the strength of association between depression and various dimensions of perfectionism is important because only negative perfectionism is supposed to be strongly related to depression. Two commercial databases were searched for published studies, and conference proceedings from professional research organizations, gray literature websites, and ProQuest Dissertations and Theses were searched for non-published studies. The total sample consisted of 52 studies, and the search for studies was thorough but not exhaustive. Random-effects models were used for the meta-analyses. Correlations between perfectionism subscales and depression measures that were collected from the studies in the sample were corrected for attenuation due to measurement error.

As anticipated, the six negative dimensions/subscales of Perfectionistic Concerns were shown to be more strongly and directly correlated with depression than the three positive dimensions of Perfectionistic Strivings. Evidence of publication bias was examined using forest plots, funnel plots, statistical tests for asymmetry of funnel plots, and cumulative meta-analyses. Five out of the nine meta-analyses showed evidence of publication bias through the cumulative meta-analyses or the trim and fill procedure. However, none of the meta-analyses showed significant funnel plot asymmetry. In aggregate, results suggest some evidence of publication bias.

Document Type

Dissertation

Degree Name

Ph.D.

Department

Quantitative Research Methods

First Advisor

Antonio Olmos, Ph.D.

Second Advisor

Kathy Green

Third Advisor
Frederique Chevillot

Keywords
Depression, Meta-analysis, Perfectionism

Subject Categories
Applied Statistics | Quantitative Psychology

Publication Statement
Copyright is held by the author. User is responsible for all copyright compliance.

This dissertation is available at Digital Commons @ DU: https://digitalcommons.du.edu/etd/1368

META-ANALYSES OF THE RELATIONSHIP BETWEEN DEPRESSION AND NINE DIMENSIONS OF PERFECTIONISM

A Dissertation

Presented to

the Faculty of the Morgridge College of Education

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Gabriel Lynn Hottinger

November 2017

Advisor: Dr. Antonio Olmos

©Copyright by Gabriel Lynn Hottinger 2017

All Rights Reserved

Author: Gabriel Lynn Hottinger
Title: META-ANALYSES OF THE RELATIONSHIP BETWEEN DEPRESSION AND NINE DIMENSIONS OF PERFECTIONISM
Advisor: Dr. Antonio Olmos
Degree Date: November 2017

Abstract

Perfectionism has been shown to be related to depression, but perfectionism is multidimensional. Some dimensions are related to positive psychological characteristics and outcomes and other dimensions are related to negative psychological characteristics and outcomes. This study reports results of nine meta-analyses performed to investigate the association between each of nine subscales of perfectionism and depression to determine which dimensions of perfectionism are most strongly associated with depression. The two subscales that were used from the Hewitt and Flett (1991b) Multidimensional Perfectionism scale were Self-Oriented Perfectionism (SOP) and Socially-Prescribed Perfectionism (SPP). The five subscales that were used from the Frost et al. (1991) Multidimensional Perfectionism Scale were Personal Standards (PS), Doubts about Actions (DA), Concern over Mistakes (CM), Parental Expectations (PE), and Parental Criticism (PC). The two subscales that were used from the Slaney et al. (2001) Almost Perfect Scale-Revised were High Personal Standards (HS) and Discrepancy (Dis). The SPP, DA, CM, PE, PC, and DIS subscales are negative dimensions of perfectionism that form the higher-order factor Perfectionistic Concerns (PC). The SOP, PS, and HS subscales are more positive dimensions of perfectionism that form the higher-order factor Perfectionistic Strivings (PS). Knowing the strength of association between depression and various dimensions of perfectionism is important because only negative perfectionism is supposed to be strongly related to depression.

Two commercial databases were searched for published studies, and conference proceedings from professional research organizations, gray literature websites, and ProQuest Dissertations and Theses were searched for non-published studies. The total sample consisted of 52 studies, and the search for studies was thorough but not exhaustive. Random-effects models were used for the meta-analyses. Correlations between perfectionism subscales and depression measures that were collected from the studies in the sample were corrected for attenuation due to measurement error.

As anticipated, the six negative dimensions/subscales of Perfectionistic Concerns were shown to be more strongly and directly correlated with depression than the three positive dimensions of Perfectionistic Strivings. Evidence of publication bias was examined using forest plots, funnel plots, statistical tests for asymmetry of funnel plots, and cumulative meta-analyses. Five out of the nine meta-analyses showed evidence of publication bias through the cumulative meta-analyses or the trim and fill procedure. However, none of the meta-analyses showed significant funnel plot asymmetry. In aggregate, results suggest some evidence of publication bias.

Acknowledgments

I would like to thank Dr. Kathy Green, Dr. Antonio Olmos, Dr. Janice Thurn, Dr. Robert Chinisci, Dr. Priyalatha Govindasamy, Dr. Sara Markey, Dr. Frédérique Chevillot, Dr. Trisha Raque-Bogdan, and my Dad, Kenneth William Hottinger, Jr.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction and Literature Review

Depression is a significant health problem worldwide. According to the World Health Organization (WHO, April 2016), depression is the number one cause of disability throughout the world, and an estimated 350 million people of all ages suffer from depression. In its most severe form, depression can lead to suicide, and globally over 800,000 people commit suicide every year (WHO, April 2016). It has been estimated that about 15% of people who struggle with severe depression will eventually commit suicide (Rittberg, 2016; Wryobeck, Haines, Wynkoop, & Swanson, 2013). Of the approximately 30,000 people who commit suicide in the U.S. each year, half of those suicides are linked to episodes of depression (Rittberg, 2016). Depression often co-occurs with generalized anxiety disorder, and this co-occurrence is called *comorbidity* (Goldberg, 2016), and depression is frequently comorbid with other psychological disorders (Rittberg, 2016). Major depressive disorder is not only comorbid with other psychiatric disorders but also with physical health problems (Greenberg, Fournier, Sisitsky, Pike, & Kessler, 2015). Women have a higher rate of depression with a lifetime prevalence rate of depression in women of 20% to 25%, whereas in men the lifetime prevalence rate is 9% to 12% (Ritschel, Gillespie, Arnarson, & Craighead, 2013), and women develop major depressive disorder twice as often as men (Rittberg, 2016). The more episodes of depression an individual has, the more likely it is that that individual will have additional episodes of depression (Rittberg, 2016). In what appear to be the most recent projections for the global burden of disease, by 2030 major depression will be the second largest global burden of disease worldwide, and it will be the first largest burden of disease in high income countries (Mathers & Loncar, 2006).

There is a 50% rule about the diagnosis and treatment of depression in the U.S. that suggests that only 50% of people with depression, who go to their primary caregiver for help, are diagnosed as having depression, and only 50% of those diagnosed are treated, and only 50% of those treated are treated adequately (Rittberg, 2016, p. 82). Greenberg et al. (2015) used propensity score matching and health insurance data to estimate the cost in the United States of people who have been diagnosed with major depressive disorder and who are being treated for major depressive disorder (MDD) and comorbid conditions. According to Greenberg et al., the "incremental burden of MDD," which means the costs directly associated with treating MDD, were $66.2 million in 2005 and $80.3 million in 2010 in the United States. However, the "incremental economic burden of individuals with MDD," that is, the difference between the cost of healthy adults and the cost of adults with MDD (including the cost of their comorbid physical and psychiatric disorders) in the United States was estimated to be $173.2 billion in 2005 and $210.5 billion in 2010 (Greenberg et al., 2005). However, Greenberg et al.'s study only looked at adults with major depression who had been diagnosed and/or who were receiving treatment, and this study did not include people with Medicare coverage, so the cost of major depressive disorder would be greater if the analysis had included people who are not diagnosed and therefore not getting treatment, and if it had included people who had Medicare coverage due to their depression being so severe that it was a legal disability.

One construct found to be correlated with depression is perfectionism. Perfectionism is a transdiagnostic factor that is correlated with many psychological and health disorders such as panic disorder, social anxiety disorder, obsessive compulsive disorder (OCD), generalized anxiety disorder, specific phobias, posttraumatic stress disorder (PTSD), body dysmorphic disorder, chronic fatigue syndrome, obsessive compulsive personality disorder, eating disorders, major depressive disorder, bipolar disorder, and suicidal thoughts and behaviors (Egan, Wade, Shafran, & Antony, 2014; Frost, Glossner, & Maxner, 2010; Kempke, Van Houdenhove, Claes, & Luyten, 2016). Perfectionism is not only correlated with these disorders but can also be part of the cause of such disorders, and it can maintain and impede the psychotherapeutic treatment of these disorders (Blatt & Zuroff, 2002; Egan et al., 2014) Perfectionism has been known to lead to suicide when perfectionists fail to meet their exacting standards (Blatt, 1995; Egan et al., 2014; Flett, Hewitt, & Heisel, 2014). Creating transdiagnostic treatment protocols for comorbid psychological disorders that focus on reducing perfectionism could lead to more efficient and effective types of psychotherapy (Egan, Wade, & Shafran, 2012). However, perfectionism is a multidimensional construct and while some dimensions of perfectionism have been found to be associated with negative psychological characteristics and outcomes, other dimensions have been found to be associated with positive psychological characteristics and outcomes (Lo & Abbott, 2013; Stoeber & Otto, 2006). Because there are both positive and negative aspects or dimensions of perfectionism, and because not all aspects or dimensions of perfectionism have been shown to be associated with depression, it was important to determine which dimensions of perfectionism are most strongly associated with depression. Determining which dimensions of perfectionism are most strongly associated with depression could help inform the development of interventions to prevent depression and treatments to alleviate depression, especially since certain dimensions of perfectionism impede the effective treatment of depression.

Statement of the Problem

Individual empirical studies of perfectionism have found that some dimensions of perfectionism are primarily maladaptive and that there are strong direct associations between these dimensions of perfectionism and depression (Dunkley, Zuroff & Blankstein, 2006). Individual empirical studies have also found that some dimensions of perfectionism are less maladaptive, or in some ways adaptive, and beneficial and that these dimensions are either weakly associated with depression or are not associated with depression (Dunkley, Zuroff et al., 2006; Stoeber & Otto, 2006). The purpose of this study was to use nine separate meta-analyses to estimate the correlations between adaptive and maladaptive dimensions of perfectionism and depression.

Originally perfectionism was theorized to be a unidimensional characteristic that was only maladaptive and was associated with mostly negative psychological characteristics and outcomes (Blankstein & Dunkley, 2002; Burns, 1980; Flett & Hewitt, 2002). However, later research conceptualized perfectionism as multidimensional with some dimensions of perfectionism correlated with negative psychological characteristics and outcomes, and other dimensions of perfectionism correlated with positive psychological characteristics and outcomes (Blankstein & Dunkley, 2002; Flett & Hewitt, 2002). The latter multidimensional conceptions of perfectionism posited that perfectionism has both interpersonal and intrapersonal aspects (Flett & Hewitt, 2002).

However, the results of studies linking perfectionism and depression have been mixed, with some studies showing evidence that most or all aspects of perfectionism are maladaptive and associated with negative psychological characteristics and outcomes, and with other studies showing evidence that only some dimensions of perfectionism as strongly associated with negative psychological characteristics and outcomes, with other dimensions of perfectionism actually associated with positive psychological characteristics and outcomes (Lo & Abbott, 2013: Stoeber & Otto, 2006). There have been numerous individual empirical studies that have examined the associations between different dimensions of perfectionism and depression, and Smith, Sherry, Rnic, Saklofske, Enns, and Gralnick (2016) appear to have done the first meta-analysis on the relationship between perfectionism and depression, but it appears that they only used one database in their literature search (PsycINFO). They did not assess publication bias, which might be substantial since they used only published studies in their meta-analysis. Their meta-analysis used 10 studies with 11 samples. Also, they used the scales for the conceptualization of perfectionism that consists of Self-Critical perfectionism versus Personal Standards perfectionism, which is explained below. Purpose of the Study

The purpose of this study was to identify the strength of the association between depression and both the positive or more adaptive dimensions of perfectionism and the negative or maladaptive dimensions of perfectionism. It is beneficial to investigate the strength of the association between each dimension of perfectionism and depression, in order to determine which dimensions of perfectionism are most strongly associated with depression and to determine if all dimensions of perfectionism are maladaptive, or if some dimensions of perfectionism are maladaptive while other dimensions of perfectionism are neutral, or adaptive and beneficial. Knowing which dimensions of perfectionism are most strongly associated with negative psychological characteristics and outcomes, such as depression, could inform the creation of interventions that target the most maladaptive dimensions of perfectionism in order to improve the treatment of depression in people who also have dimensions of perfectionism that either exacerbate their depression or impede treatment of their depression. A series of meta-analyses summarizing the associations between dimensions of perfectionism and depression gives a better overall estimate of the association between each dimension of perfectionism and depression since estimates are based on all the available and relevant studies and are less influenced by the sampling error of each individual study, so these estimates are more precise (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Research Questions

The research questions for this study were:

1) Does the pattern of correlations for the association of depression with Perfectionistic Strivings (PS) and Perfectionistic Concerns (PC) dimensions of perfectionism differ enough to give evidence that these two types of perfectionism are distinct constructs?
   a) Are all the dimensions of Perfectionistic Concerns (PC) positively and significantly correlated with depression?
   b) Are all the dimensions of Perfectionistic Strivings either not significantly correlated with depression or inversely correlated with depression?

2) Are the two possibly opposite types of perfectionism differentially related to depression?
   a) How strong is the association between the negative (maladaptive) dimensions of perfectionism that comprise Perfectionistic Concerns (PC) and severity of depression?
      i) Which one of the Perfectionistic Concerns (PC) dimensions of perfectionism is most strongly associated with depression?
      ii) Are the associations between the Perfectionistic Concerns (PC) dimensions of perfectionism and depression stronger for women than for men?
      iii) As the research on perfectionism and depression indicates, are the Discrepancy subscale from the APS-R and the Socially Prescribed Perfectionism (SPP) subscale from the HMPS the two dimensions of Perfectionistic Concerns (PC) that are most strongly and positively associated with depression?
   b) How strong is the association between the positive (adaptive) dimensions of perfectionism that comprise Perfectionistic Strivings (PS) and severity of depression?
      i) Are any of the Perfectionistic Strivings (PS) dimensions of perfectionism significantly positively correlated with depression?
      ii) Are any of the Perfectionistic Strivings (PS) dimensions of perfectionism significantly *negatively* correlated with depression?

c) Are the negative dimensions of perfectionism that comprise Perfectionistic Concerns (PC) perfectionism more strongly associated with severity of depression than the positive or neutral dimensions of perfectionism that comprise Perfectionistic Strivings (PS) perfectionism?

All the research questions listed above, except the moderator analysis examining the possible difference between males and females, were answered by doing nine separate meta-analyses in which a summary or mean correlation was calculated for the relationship between each of the nine perfectionism subscales that were the focus of this study and depression.

Review of the Literature on Perfectionism

Perfectionism has been seen as being both a unidimensional (Shafran, Cooper, & Fairburn, 2002) and multidimensional construct (Frost et al., 1990). Shafran et al.'s (2002) construct of clinical perfectionism is unidimensional and is defined as "*the overdependence of self-evaluation on the determined pursuit of personally demanding, self-imposed, standards in at least one highly salient domain, despite adverse consequences*" (p. 778, italics original). Most of the unidimensional conceptions of perfectionism view perfectionism as a primarily negative personality characteristic (e.g., Burns, 1980; Shafran et al., 2002). In creating their Multidimensional Perfectionism Scale, Hewitt and Flett (1991b) generated a multidimensional measure of perfectionism by adding interpersonal aspects to the construct (Hewitt, Flett, Besser, Sherry, & McGee, 2003). Shafran et al.'s concept of clinical perfectionism was only intrapersonal. Shafran et al.'s construct of clinical perfectionism specifies that excessively high standards are only a problem in one domain of the clinically perfectionistic person's functioning, but the multidimensional conception of perfectionism, in which excessively high standards are set for a variety of life domains, would logically cause the perfectionistic person more problems and would be more extreme and therefore more detrimental (Hewitt et al, 2003). Shafran et al. indicated that the more areas in one's life in which one has problems with unhealthy perfectionism, the more detrimental that perfectionism is. Shafran et al. argue that a unidimensional perfectionism construct is more appropriate than a multidimensional approach.

Three multidimensional measures of perfectionism. According to other researchers, perfectionism is a multidimensional construct (Frost, Marten, Lahart, & Rosenblate, 1990; Hewitt & Flett, 1991b; Slaney, Rice, Mobley, Trippi, & Ashby, 2001). According to Sirois and Molnar (2016) the three most frequently used measures of perfectionism are The Multidimensional Perfectionism Scale (FMPS) developed by Frost et al. (1990), the Multidimensional Perfectionism Scale (HMPS) developed by Hewitt and Fleet (1991b), and the Almost Perfect Scale-Revised (APS-R), which was developed by Slaney et al. (2001). The Frost et al. (1990) FMPS, the Hewitt and Flett (1991b) HMPS, and the APS-R (Slaney et al., 2001) are all measures of trait perfectionism (Enns & Cox, 2002; Flett, Hewitt, Blankstein, & Gray, 1998).

The Frost et al. (1990) Multidimensional Perfectionism Scale (FMPS) has six subscales that represent six different dimensions of perfectionism: Concern over Mistakes (CM), Personal Standards (PS), Parental Expectations (PE), Parental Criticism (PC), Doubts about Actions (DA), and Organization (O). The Concern over Mistakes (CM) subscale "reflects negative reactions to mistakes, a tendency to interpret mistakes as equivalent to failure, and a tendency to believe that one will lose the respect of others following failure" (Frost, Heinberg, Holt, Mattia, & Neubauer, 1993, p. 121). The Personal Standards (PS) subscale "reflects the setting of very high standards and the importance placed on these high standards for self-evaluation" (Frost et al., 1993, p. 121). The Parental Expectations (PE) subscale reflects the "tendency to believe that one's parents set very high goals" (Frost et al., 1993, p. 121). The Parental Criticism (PC) subscale reflects "the perception that one's parents are (or were) overly critical" (Frost et al., 1993, p. 121). The Doubts about Actions (DA) subscale reflects the "tendency to feel that projects are not completed to satisfaction" (Frost et al., 1990, p. 453). Finally, the Organization (O) subscale reflects "emphasis on the importance of and preference for order and Organization" (Frost et al., 1990, p. 453). The items for the Frost et al. (1990) FMPS consist of several items taken from the Burns Perfectionism Scale (Burns, 1980) and from the perfectionism subscale of the Eating Disorders Inventory (EDI; Garner, Olmstead, & Polivy, 1983, as cited in Frost et al., 1990) and from Rachman and Hodgson's (1983, as cited in Frost et al., 1990) scale measuring obsessionality, along with several newly generated items (Frost et al., 1990). A total score is also reported for the FMPS, but it does not include the Organization subscale (Frost et al., 1993).

Of the same name as the Frost et al. (1990) scale is the Multidimensional Perfectionism Scale (MPS) developed by Hewitt and Flett (1991b), which has three subscales representing three different types of perfectionism: Self-Oriented Perfectionism (SOP), Other-Oriented Perfectionism (OOP), and Socially Prescribed Perfectionism (SPP). Socially prescribed perfectionism consists of "people's belief or perception that significant others have unrealistic standards for them, evaluate them stringently, and exert pressure on them to be perfect" (Hewitt & Flett, 1991b, p. 457). Self-oriented perfectionism involves "setting exacting standards for oneself and stringently evaluating and censuring one's own behavior…[and] striving to attain perfection in one's endeavors as well as striving to avoid failures" (Hewitt & Flett, 1991b, p. 457). A person who has other-oriented perfectionism has "unrealistic standards for significant others, places importance on other people being perfect, and stringently evaluates others' performance" (Hewitt & Flett, 1991b, p. 457). Self-oriented perfectionism has been found to be associated with both positive and negative psychological or personality characteristics; however, socially prescribed perfectionism has been shown to be associated with only negative psychological or personality characteristics and not positive characteristics (Hill, McIntire, & Bacharach, 1997). In summarizing the results of several studies, Blankstein and Dunkley (2002) said that socially-prescribed perfectionism was shown to have the strongest relationships with maladaptive characteristics. Even though some studies have shown that positive perfectionism is mainly associated with positive psychological characteristics and outcomes when negative aspects of perfectionism have been statistically controlled for (Blankstein & Dunkley, 2002; Stoeber & Otto, 2006), other studies have shown that self-oriented perfectionism, which is considered to be an aspect of positive perfectionism, can be directly related to depression (Flett, Hewitt, Blankstein, & Gray, 1998; Hewitt & Flett, 1991a; Hewitt & Flett, 1993; Hewitt, Flett, & Ediger, 1996; Slaney et al., 2002; Stoeber & Otto, 2006). No total score is calculated for the HMPS (Frost et al., 1993).

A third perfectionism scale that measures more than one dimension of perfectionism is the Almost Perfect Scale-Revised (APS-R), which was developed by Slaney, Rice, Mobley, Trippi, and Ashby (2001), and this measure has three subscales:

Discrepancy, High Standards, and Order. The creators of the Almost Perfect Scale and its revised edition, the APS-R, thought that the previous two multidimensional perfectionism scales were based on negative conceptions of perfectionism, so they wanted to allow perfectionism to have positive aspects (Enns & Cox, 2002). According to Slaney, Rice, and Ashby (2002) "the possession of high standards for one's performance has proven to be the dimension of perfectionism about which there is near unanimity in dictionary definitions, the literature, scale development, and interview studies" (p. 69). The High Standards subscale measures whether someone has high personal standards and Order subscale measures a person's preference for orderliness (Slaney et al., 2001). Discrepancy is "defined as the perceived discrepancy or difference between the standards one has for oneself and one's actual performance" (Slaney et al., 2001, p. 133) and it is also defined as "the perception that one consistently fails to meet the high standards one has set for oneself" (Slaney et al., 2002, p. 69). Discrepancy is the central and defining aspect of negative perfectionism (Slaney et al., 2002). High Standards and Order are the central and defining aspects of positive perfectionism (Slaney et al., 2001), but the High Standards subscale is more essential to the concept of perfectionism than Order (Slaney et al., 2002). Discrepancy and High Standards are conceptualized as independent of each other and are considered to be more essential to the construct of perfectionism than is Order (Slaney et al., 2002). According to Slaney et al. (2002), maladaptive perfectionists are people who score high on both High Standards and Discrepancy, and adaptive perfectionists are people who score high on High Standards but not on Discrepancy. The creators of the APS-R thought that the Discrepancy scale could distinguish between adaptive and maladaptive perfectionism (Flett & Hewitt, 2002). The APS-R was made freely available to anyone wanting to do research on perfectionism (Rice, Richardson, & Tueller, 2014).

Two types or higher-order factors of perfectionism. Early research on perfectionism by people such as David Burns (1980) saw perfectionism as being a unidimensional construct that was primarily pathological, or negative, and used only unidimensional measures of perfectionism (Stoeber & Otto, 2006). Hamachek (1978) was an exception to the early unidimensional view of perfectionism and the early research on perfectionism, because Hamachek identified two types of perfectionism: normal and neurotic (as cited in Stoeber & Otto, 2006). In the early 1990s when the Frost et al. (1990) and the Hewitt and Flett (1991b) Multidimensional Perfectionism Scales were developed, perfectionism began to be conceptualized as a multidimensional construct that might have positive attributes (Stoeber & Otto, 2006).

Since the time that perfectionism began to be conceptualized as multidimensional, there have been several studies that reported factor analyses of the subscales of the multidimensional measures of perfectionism to determine which subscales were measuring the same latent factors of perfectionism (Cox, Enns, & Clara, 2002; Frost et al., 1993; Blankstein & Dunkley, 2002; Stoeber & Otto, 2006) There have also been a number of studies that used factor analysis to determine which subscales of the multidimensional perfectionism scales clustered together (Frost et al., 1993). Frost et al. used data combined from the three facets of perfectionism measured by the three subscales of the HMPS and the six facets of perfectionism measured by the six subscales from the FMPS and conducted a single factor analysis on those data. Two higher-order factors emerged from the analysis, and Frost et al. referred to these two factors as *positive* *strivings* and *maladaptive evaluation concerns*. The positive strivings dimension of perfectionism found by Frost et al. consisted of the following subscales from the two different MPS measures: Personal Standards, Organization, Self-Oriented Perfectionism, and Other-Oriented Perfectionism; and Frost et al. found this dimension to be correlated with positive psychological characteristics. The maladaptive evaluation concerns dimension of perfectionism found by Frost et al. consisted of the following subscales from the two different MPS measures: Concern over Mistakes, Doubts about Actions, Socially Prescribed Perfectionism, Parental Expectations, and Parental Criticism; and they found this factor to be associated with negative psychological characteristics. Many other studies used the same combination of subscales that Frost et al. used to explore the two kinds of perfectionism, and some studies found positive strivings perfectionism to be associated with only positive psychological characteristics as Frost et al. did, but other studies found positive strivings perfectionism to be associated with both positive and negative psychological characteristics (Stoeber & Otto, 2006). Also, other studies on perfectionism used different combinations of perfectionism subscales to form two types of perfectionism (Stoeber & Otto, 2006). The two higher-order factors found by Frost et al. are considered by Stoeber and Otto (2006) to be two basic forms of perfectionism, which they refer to as *perfectionistic strivings* and *perfectionistic concerns*, and a single person can have either one or both forms of perfectionism. The same person can have facets of both types of perfectionism—a person can have perfectionistic characteristics that are part of the positive latent factor of perfectionism and at the same time have perfectionistic characteristics that are part of the negative latent factor of perfectionism (Stoeber & Otto, 2006). Stoeber and Otto's review of research on the two basic forms of perfectionism used different combinations of subscales than those Frost et al. used. In Stoeber and Otto's terminology, *healthy perfectionists* have high levels of perfectionistic strivings and low levels of perfectionistic concerns, and *unhealthy perfectionists* have high levels of both perfectionistic strivings and perfectionistic concerns.

In a related conception of perfectionism, Dunkley, Blankstein, Masheb, and Grilo (2006) refer to two different types of perfectionism: *Personal Standards* (PS) perfectionism and *Evaluative Concerns* (EC) perfectionism. EC perfectionism consists of the Concern over Mistakes and the Doubts about Actions subscales from the FMPS and the Socially-Prescribed Perfectionism subscale of the HMPS (Dunkley, Blankstein et al., 2006). PS perfectionism consists of the Self-Oriented Perfectionism subscale from the HMPS and the Personal Standards subscale from the FMPS (Dunkley, Blankstein et al., 2006). According to Dunkley, Blankstein et al., EC perfectionism is maladaptive, but PS perfectionism is not necessarily maladaptive. According to Dunkley, Blankstein et al. "PS perfectionism involves the setting of high standards and goals for oneself" and "EC perfectionism involves overly critical evaluations of one's own behavior, and inability to derive satisfaction from successful performance, and chronic concerns about others' criticism and expectations" (p. 65). According to Dunkley, Blankstein et al. much research has shown evidence that "self-critical evaluative tendencies are the critical component of perfectionism" and it has also shown that EC perfectionism is strongly related to self-criticism (p. 70). Dunkley, Blankstein et al. also said that a substantial amount of research has shown a relationship between EC perfectionism and depression, but PS perfectionism has been shown to have a weak or nonsignificant relationship to depression. According to Dunkley, Blankstein et al., Hamachek's (1978, as cited in

Dunkley, Blankstein et al., 2006) early distinction between normal and neurotic perfectionism was basically the same concept as Dunkley, Blankstein et al.'s PS Perfectionism and EC perfectionism, respectively. Basically, the same two types of perfectionism have been described and defined similarly by other researchers, but have been referred to by different names (Dunkley, Blankstein et al., 2006).

Many factor analyses of the perfectionism measures have found "a two-dimensional, higher order factor structure for the construct" and one of the two factors has been named differently by different authors but "has been suggested to capture the more adaptive and positive facets of perfectionism related to perfectionistic striving and having high personal standards" and "This 'positive' dimension has been shown to be related to positive affect and unrelated to depression" (Lo & Abbott, 2013, p. 98). The other factor, which has also been named differently by different authors, "represents the negative and pathological facets of perfectionism related to critical self-evaluation of one's performance and feelings of discrepancy between one's performance and one's expectations" and this negative factor "has been found to be inversely associated with self-esteem and positively associated with depression and negative affect" (Lo & Abbott, 2013, p. 99). The Self-Oriented Perfectionism and Other-Oriented Perfectionism subscales from the HMPS and the Personal Standards and Organization subscales from the FMPS have been found to load on the positive perfectionism factor, and the Concern over Mistakes, Doubts about Actions, Parental Expectations, and Parental Criticism subscales from the Frost et al. (1990) MPS and the Socially-Prescribed Perfectionism subscale from the Hewitt and Flett (1991b) MPS have been found to load on the negative perfectionism factor (Lo & Abbott, 2013).

Positive perfectionism can also be associated with some negative psychological characteristics and outcomes if the perfectionistic person has both positive and negative aspects of perfectionism, but some studies have shown that when the negative aspects of perfectionism are controlled for statistically, positive perfectionism is mainly associated with positive psychological characteristics and outcomes (Blankstein & Dunkley, 2002; Stoeber & Otto, 2006). Some researchers have found that people with positive perfectionism have stronger positive associations with positive psychological characteristics and outcomes compared to not only people with negative perfectionism but also people who are not perfectionists (Slaney et al., 2002; Stoeber & Otto, 2006).

In the literature on perfectionism, the two different types of perfectionism have been referred to by many different names. The more adaptive type of perfectionism has been referred to as *adaptive perfectionism*, *healthy perfectionism*, *personal standards perfectionism, perfectionistic strivings, positive perfectionism,* and *normal perfectionism,* (Blankstein & Dunkley, 2002; Enns & Cox, 2002; Flett & Hewitt, 2002) The maladaptive type of perfectionism has been referred to as *self-critical perfectionism*, *pathological perfectionism, evaluative concerns perfectionism*, *neurotic perfectionism*, *maladaptive perfectionism*, *clinical perfectionism*, *negative perfectionism,* and *unhealthy perfectionism* (Blankstein & Dunkley, 2002; Enns & Cox, 2002; Flett & Hewitt, 2002).

Positive perfectionism has been shown to be associated with both positive and negative psychological characteristics and outcomes, or with only positive characteristics if dimensions of negative perfectionism have been statistically controlled for, and negative perfectionism has been shown to be associated with only negative psychological characteristics and outcomes (Dunkley et al., 2016: Dunkley, Blankstein et al., 2006;

Frost et al., 1993; Stoeber & Otto, 2006). Factor analyses have shown that positive perfectionism is associated with the following positive psychological characteristics and outcomes: conscientiousness, a sense of well-being, high achievement, high self-esteem, positive affect, and high personal standards (Blankstein & Dunkley, 2002; Dunkley, Zuroff et al., 2006; Enns & Cox, 2002: Flett & Hewitt, 2002; Frost et al., 1993; Stoeber & Otto, 2006). Negative perfectionism or negative dimensions of perfectionism have been shown to be associated with the following negative psychological characteristics and outcomes: self-criticism, maladjustment, avoidant coping, shame, procrastination, depression, anxiety, negative affect, low self-esteem, fear of making mistakes, fear of failure, need for approval, inflexibility, external locus of control, suicide, and eating disorders (Blankstein & Dunkley, 2002; Dunkley, Zuroff et al., 2006; Enns & Cox, 2002: Flett & Hewitt, 2002; Frost et al., 1993; Stoeber & Otto, 2006).

Current conceptions of perfectionism. Even though there is extensive research on perfectionism, and most perfectionism researchers agree that perfectionism is multidimensional, there is no consensus on which combination of perfectionism scales should be used to measure perfectionism, or which dimensions of perfectionism best define the construct, and the different ways that perfectionism is measured affect the empirical results of perfectionism research (Sirois & Molnar, 2016). Most perfectionism researchers agree that the construct of perfectionism is "bidimensional" (Burgess & DiBartolo, 2016, p. 177). These three multidimensional measures of perfectionism that were the focus of this study are the three most popular and most "influential multidimensional models of perfectionism" (Dunkley, Solomon-Krakus, & Moroz, 2016; Molnar & Sirois, 2016, p. 287; Sirois & Molnar, 2016). These three measures are the

Frost et al. (1990) Multidimensional Perfectionism Scale (FMPS), the Hewitt and Flett (1991b) Multidimensional Perfectionism Scale (HMPS) and the Almost Perfect Scale-Revised (APS-R, Slaney et al., 2001). Research using these three most popular multidimensional measures of perfectionism, which were the focus of this study, has repeatedly found that there are two higher-order factors that underlie these three measures (Sirois & Molnar, 2016).

One of the most current and most empirically substantiated conceptualization of perfectionism that also consists of two higher-order factors is Self-Critical perfectionism versus Personal Standards perfectionism (Dunkley et al., 2016). Self-criticism is so much a part of perfectionism that some perfectionism researchers started adding a measure of self-criticism, such as the Self-Criticism subscale of the Depressive Experience Questionnaire (DEQ, Blatt, D'Afflitti, & Quinlan, 1976), to the Perfectionistic Concerns (PC) factor, described in the next paragraph, to create the Self-Critical Perfectionism factor (Molnar, Sirois, & Methot-Jones, 2016). The Self-Critical perfectionism higher-order factor is measured with the Socially Prescribed Perfectionism subscale from the Hewitt and Flett (1991b) MPS, the Concern over Mistakes and the Doubts about Actions subscales of the Frost et al. (1990) MPS, and the Discrepancy subscale of the Slaney et al. (2001) APS-R (Dunkley et al., 2016). The Personal Standards perfectionism higher-order factor is measured with Personal Standards subscale of the Frost et al. (1990) MPS, the High Standards subscale of the Slaney et al. (2001) APS-R, and the Self-Oriented Perfectionism subscale of the Hewitt and Flett (1991b) MPS (Dunkley et al., 2016). This conceptualization of perfectionism was not used in this study because recent studies investigating this topic often use one composite score for all the subscales that constitute

Self-Critical perfectionism and use another composite score for all the subscales that constitute Personal Standards perfectionism (Békés et al., 2015), so it would not be feasible to get the information necessary to calculate the correlation for each individual subscale's association with depression separately because that information would probably not be reported in the journal articles about Self-Critical perfectionism (Dunkley, Berg, & Zuroff, 2012; Dunkley, Mandel, & Ma, 2014; Sherry, Richards, Sherry, & Stewart, 2014; Sherry, Gautreau, Mushquash, Sherry, & Allen, 2014).

Another current conceptualization of perfectionism that also has a lot of empirical support in the literature on perfectionism consists of two higher-order latent factors that are frequently referred to as Perfectionistic Strivings (PS) and Perfectionistic Concerns (PC) (Sirois & Molnar, 2016). The Perfectionistic Strivings (PS) higher-order factor consists of the Personal Standards subscale from the Frost et al. (1990) MPS, the High Standards subscale from the APS-R, and the Self-Oriented Perfectionism subscale from the Hewitt and Flett (1991b) MPS (Sirois & Molnar, 2016). The Perfectionistic Concerns (PC) higher-order factor consists of Parental Expectations, Doubts about Actions, Concern over Mistakes, and Parental Criticism subscales from the Frost et al. (1990) MPS, the Discrepancy subscale from the APS-R, and the Socially Prescribed Perfectionism subscale from the Hewitt and Flett (1991b) MPS (Sirois & Molnar, 2016). Perfectionistic Concerns (PC) is viewed as maladaptive or unhealthy and is correlated with negative psychological characteristics and outcomes (Molnar et al., 2016; Sirois, 2016). Perfectionistic Strivings (PS) is viewed as more adaptive or healthier and is correlated with both positive and negative psychological characteristics and outcomes (Molnar et al., 2016; Sirois, 2016). This conceptualization of perfectionism as having the two higher-order factors of Perfectionistic Strivings (PS) and Perfectionistic Concerns (PC) is the model of perfectionism that was used in this study. Hewitt and Flett (2002) asserted that it would be important to examine the different dimensions of perfectionism from the multidimensional view of perfectionism. Table 1 below gives the characteristics of the included subscales from the three multidimensional measures of perfectionism.

This study adds additional knowledge to the literature on perfectionism above what the meta-analysis on the dimensions of perfectionism by Smith et al. (2016) contributed because unlike the Smith et al.'s meta-analysis, this study used a more thorough literature search because it searched for relevant studies in more than just one database, it included two unpublished studies in the meta-analyses, it assessed for publication bias, it included subscales from the APS-R (Slaney et al., 2001), and it investigated the Perfectionistic Strivings (PS) and Perfectionistic Concerns (PC) dimensions of perfectionism rather than the Self-Critical perfectionism and Personal Standards dimensions of perfectionism that Smith et al. used.

The two most maladaptive dimensions of perfectionism. Of the nine dimensions of perfectionism being investigated in this study, the Discrepancy subscale from the APS-R (Slaney et al., 2001) and the Socially Prescribed Perfectionism (SPP) subscale from the Hewitt and Flett (1991b) HMPS were expected to be the two most maladaptive dimensions of perfectionism.

Discrepancy subscale. According to Enns and Cox (2002) and Burns (1980), black-and-white thinking is a component of maladaptive perfectionism. The concept of Discrepancy involves black-or-white or all-or-nothing thinking because Discrepancy is the difference between a maladaptive perfectionist's impossible-to-reach standard of perfection and his or her actual performance, which falls below the standard of perfection (Slaney et al., 2001; Tangney, 2002). Maladaptive perfectionists view their performance in an all-or-nothing way where either their performance is perfect or else it is a failure (Tangney, 2002). According to Slaney et al. (2002), the Discrepancy concept was posited "to potentially capture the essential defining negative dimension" (p. 69) of negative perfectionism and could be the "defining negative aspect of perfectionism" (p. 80). Enns and Cox (2002) thought that Discrepancy could be very useful in distinguishing between the positive or adaptive type of perfectionism and the negative or maladaptive perfectionism type of perfectionism. According to Slaney et al. (2002), "The research on the APS-R clearly indicates that the discrepancy construct is consistently and substantively related to negative psychological states; conversely, it is negatively related to positive states and measures of achievement" (p. 82). Thus, in this study, it was expected that the Discrepancy subscale from the APS-R would be one of the two subscales that are most highly correlated with severity of depression.

***Socially Prescribed Perfectionism (SPP) subscale*.** As was stated before, Socially Prescribed Perfectionism has been shown to be associated with only negative psychological or personality characteristics and not positive characteristics (Hill et al., 1997). As was also stated before, much research has shown socially-prescribed perfectionism to have very strong relationships with maladaptive characteristics (Blankstein & Dunkley, 2002). According to Tangney (2002) Socially Prescribed Perfectionism is associated with vulnerability to feeling shame. Blatt (1995) thought that feeling of shame might be part of what causes unhealthy perfectionists to become depressed. Thus, in this study, it was expected that Socially Prescribed Perfectionism

(SPP) would be the second of two dimensions of perfectionism that are most strongly correlated with depression.

Table 1 below gives the characteristics of the included perfectionism subscales. In Table 1, the validity coefficients for FMPS subscales (Frost et al., 1990) and HMPS subscales were correlations with The Burns Perfectionism Scale (Burns, 1980, as cited in Flett & Hewitt, 2015), and for APS-R validity coefficients for High Standards was correlation with HMPS Self-Oriented, and for Discrepancy were correlations with CM, DA, PC from FMPS and SPP from HMPS (Flett & Hewitt, 2015).

Table 1

Characteristics of Included Perfectionism Subscales

Name of Multidimensional Perfectionism Measure	**Name of Subscale or Dimension**	**Number of Items**	**Cronbach's Alpha Coefficient**	**Convergent/ Concurrent Validity**
The Almost Perfect Scale-Revised (APS-R), which was developed by Slaney, Rice, Mobley, Trippi, and Ashby (2001)	Discrepancy	12	.92	.47 or greater
	High Standards (HS)	7	.85	.68
The Multidimensional Perfectionism Scale (HMPS) by Hewitt and Flett (1991b)	Socially Prescribed Perfectionism (SPP)	15	.88	.69
	Self-Oriented Perfectionism (SOP)	15	.81	.62
The Multidimensional Perfectionism Scale (FMPS) by Frost, Marten, Lahart and Rosenblate (1990)	Concern over Mistakes (CM)	9	.88	.86
	Parental Expectations (PE)	5	.84	.43
	Parental Criticism (PC)	4	.84	.42
	Doubts about Actions (DA)	4	.77	.47
	Personal Standards (PS)	7	.83	.52
Short Almost Perfect Scale (SAP) by Rice, Richardson & Tueller (2014)	Discrepancy	4	.84-.85	.66
	Standards	4	.87	.62

Review of the Literature on Depression

Depression is a psychiatric disorder that can involve intense feelings of sadness and trouble regulating emotions (Nahas, 2016). Depression is a very heterogeneous disorder with a wide variety of symptoms, and different people with depression can have very different experiences from each other (Goldberg, 2016). A diagnosis of major depressive disorder requires that the individual reports an impaired ability to function in daily life and that he or she has had the symptoms of depression for at least two weeks (Rittberg, 2016). Other symptoms of depression include increased or decreased eating, insomnia or sleeping too much, weight loss without dieting or weight gain, trouble concentrating, agitation or lethargy, persistent sad or depressed mood or anhedonia, which is “the inability to feel pleasure; the loss of interest in formerly pleasurable pursuits,” (“Anhedonia,” 2015) feelings of being worthless, inappropriate guilt, and inability to make decisions (McInnis, Riba, & Greden, 2014; Rittberg, 2016).

Accounts of people suffering from depression go all the way back to the Bible (Ingram, 2012). Until the late 1800s or the early 1900s what is now called major depression was referred to as *melancholia* (Ritschel et al., 2013: Wakefield & Demazeux, 2016). The first written definition of depression is attributed to Hippocrates in the fifth century B.C. E. (Wakefield & Demazeux, 2016). During the time of Hippocrates, people believed in the theory of four humors or bodily fluids that caused disease if they were out of balance in the body (Ritschel et al., 2013; Wakefield & Demazeux, 2016). Black bile was one of the four bodily humors, and melancholia was thought to be caused by too much black bile (Ingram, 2012; Wakefield & Demazeus, 2016).

In modern times, prior to the DSM-5, depression was called unipolar depression, but in the DSM-5 the term *major depressive disorder* (MDD) is used (Rittberg, 2016). In the U.S. the *Diagnostic and Statistical Manual of Mental Disorders*-Fifth Edition (DSM-5; American Psychiatric Association, 2013) criteria for depression are the standards for diagnosing depression, but throughout the world, the World Health Organization's *International Classification for Diseases and Related Disorders* (ICD -10; WHO, 1993) is used (Richards, 2011; Rittberg, 2016).

Depression is thought to develop from a combination of biological, psychological, and social factors (WHO, April 2016). There are at least three types of theories about the etiology and maintenance of major depression: cognitive, behavioral, and biological (Ritschel et al., 2013). Behavioral theories of the etiology and maintenance of depression were developed in the 1960s and 1970s and were based on the concepts of decreased positive reinforcement, increased negative reinforcement, decreased motivation, avoidance, loss of enjoyment for previously enjoyable activities, loss of sources of self-esteem, increased anxiety, and narrowing of one's "behavior repertoire" (Ritschel et al., 2013, p. 293). The behavioral theories posit that depression might develop because people begin to receive less positive reinforcement, possibly because a source of self-esteem is lost, and this makes such individuals start to withdrawal socially and become less motived to engage in behaviors that would give them positive reinforcement (Ritschel et al., 2013). This turns into a vicious cycle because the less they engage in social behavior, the less positive reinforcement they receive until they become totally withdrawn (Ritschel et al., 2013). Increased anxiety can make them avoid engaging in social and other type of behaviors that would lead to positive reinforcement, and this avoidance can lead to a narrowing of a person's behavioral repertoire, causing even more withdrawal and avoidance (Ritschel et al., 2013). Depressed people may have a lack of motivation because they lack energy to do a task, or they believe that completing a task will not be rewarding, or they have the cognitive distortion that they are not capable of completing the task (Ritschel et al., 2013). According to Ritschel et al., depressed people may have a behavioral deficit, such as a deficit in social skills, that makes it hard for them to receive positive reinforcement from their environment. Much research has shown an association between social skills deficits and depression (Hames, Hagen, & Joiner, 2013). Also, depressed people may avoid doing necessary tasks, and the avoidance brings relief in the short term, serving as negative reinforcement, but later there are often long term negative consequences for the avoidance (Ritschel et al., 2013).

The two primary cognitive theories about the development of depression are based on the work of Aaron Beck and Martin Seligman (Ritschel et al., 2013). For Beck's cognitive theory of depression, depressed individuals have negative thoughts about themselves, the future, and the world, and this is called the *cognitive triad* (Ritschel et al., 2013). Beck's cognitive theory of depression specified that the thinking of depressed people has three components that reinforce each other and thereby cause the development and maintenance of depression: automatic negative statements about the self, errors or distortions in thinking, and negative core beliefs, which are called negative *schemas* (Ritschel et al., 2013). Seligman and his colleagues thought that depressed people develop *learned helplessness* because of the pessimistic way in which they explain uncontrollable negative events that happen in their lives (Ritschel et al., 2013). The way people explain the events in their lives is called *explanatory style*, and Seligman thought that a pessimistic explanatory style caused *learned helplessness*, which then is associated with the development and maintenance of depression (Mineka, Pury, & Luten, 1995; Ritschel et al., 2013). Beck's and Seligman's theories about depression, which originated in the late 1960s and early 1970s, were the impetus for an immense amount of research on the cognitive aspects of depression.

Several other cognitive theories of depression were proposed after those of Beck and Seligman (Ritschel et al., 2013). The hopelessness theory of depression by Abramson, Metalsky, and Alloy (1989) updated and revised Seligman's learned helplessness theory. This theory posited that depressed people feel hopeless because they believe that an event that they greatly desire will not happen and that an event that they really do not want will happen and that these events are completely out of their control (Ritschel et al., 2013). The response styles theory (RST) by Nolen-Hoeksema (1987, as cited in Ritschel et al., 2013), described how depressed people spend a lot of time ruminating rather than actively problem-solving. The attention-mediated hopelessness theory by MacCoon, Abramson, Mezulis, Hanking, and Alloy (2006, as cited in Ritschel et al., 2013), revised the hopelessness theory of depression and describes how people with depression focus on the discrepancy between how things in their life are and how they themselves are versus how they desire those things to be. In summary and briefly, the cognitive theories of depression are basically about negative thinking and focusing on the negative (Ritschel et al., 2013)

The etiology of depression is not yet known, but there is considerable evidence that a combination of environmental factors and genetic factors plays a role in the origin of major depressive disorder (Rittberg, 2016). The biological theories of depression propose that depression occurs because of disruptions in the central nervous system, endocrine system, and immune system (Ritschel et al., 2013). Studies of twins and studies of adopted children have shown that there is a genetic or hereditary component to depression that confers a vulnerability to depression (Ritschel et al., 2013). Having a parent or sibling who has depression increases the likelihood that a person will also have depression (Nahas, 2016). A vulnerability to depression can also be caused by environmental factors that exist in childhood such as child abuse or neglect, early trauma, or severe stress (Ritschel et al., 2013). A vulnerability to depression can also be caused by having low levels of neurotransmitters such as norepinephrine and serotonin (Ritschel et al., 2013). Basically, neurotransmitters help brain cells called neurons communicate with each other (Bauer, 2006). Also, stress can change the structure of the brain, the concentration of neurotransmitters, and the way the brain functions, thereby making a person more vulnerable to developing depression (Nahas, 2016). People with a vulnerability to depression show over-reactivity of the hypothalamic-pituitary-adrenal (HPA) axis, and this over-reactivity may be caused by negative events early in a person's life (Ritschel et al., 2013; Rittberg, 2016). The HPA axis is "a collection of neural and endocrine structures that function collectively to facilitate the adaptive response to stress" (Ritschel et al., 2013, p. 308).

Depression is measured in adults using self-report scales, clinical interviews, and clinical rating scales (Ritschel et al., 2013). According to Ritschel et al., researchers frequently use the Hamilton Rating Scale for Depression (HAM-D; Hamilton, 1960, as cited in Ritschel et al., 2013) to measure the severity of depression in adults. The HAM-D is a 17-item clinical rating scale that is intended to measure the severity of depression or changes in severity over time, but it is not intended to diagnose depression (Ritschel et al., 2013). Another clinical rating scale used for adults is the Quick Inventory of Depression Symptomatology (QIDS; Rush et al., 2003, as cited in Ritschel et al., 2013) which has16 items and covers nine domains of depression from the previous version of the *Diagnostic and Statistical Manual* (DSM-IV-TR; American Psychiatric Association, 2000, as cited in Ritschel et al., 2013).

There are at least two frequently used self-report measures of depression (Ritschel et al., 2013). According to Ritschel et al. (2013), the most frequently used self-report measure of depression for adults is the Beck Depression Inventory-II (BDI-II; A. T. Beck, Steer, & Brown, 1996, as cited in Ritschel et al., 2013). The BDI-II has 21 items and measures the severity of cognitive, affective, and somatic symptoms of depression. Another self-report measure of depression is the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977, as cited in Ritschel et al., 2013). The CES-D has 20 items and was designed to be a screen for depression in the general population and was not designed to measure depression severity.

Two clinical interviews that can assess DSM-IV Axis I disorders or psychopathology including depression are the Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Versions (SCID-CV; First, Spitzer, Gibbon, & Williams, 1997, as cited in Ritschel et al., 2013) and the Longitudinal Interval Follow-Up Evaluation (LIFE; Keller, Lavori, Friedman, Nielsen, Endicott, McDonald-Scott, & Andreasen, 1987, as cited in Ritschel et al., 2013). These interviews can assess depression, but they also assess other types of psychopathology (Ritschel et al., 2013). The SCID-CV is a diagnostic tool, and the LIFE is for measuring frequency and duration of psychopathology in longitudinal research. When PsycINFO was searched for studies using any of the three multidimensional perfectionism measures during the time period from 2007 to 2017, the top three depression scales used in the resulting studies were the BDI-I, the BDI-II, and the CES-D, and these three depression scales were the most commonly used measures of depression in the sample of 52 studies for nine meta-analyses conducted in this study. Table 2 provides an overview of all 11 of the measures of depression used in the nine meta-analyses conducted in this study.

Table 2

Characteristics of Included Depression Scales

Name	**No. of items**	**Type of Scale**	**Estimated Reliability**	**Convergent/ Concurrent Validity**	**Author(s) and Year**
Beck Depression Inventory (BDI-first edition)	21	Self-report	.92-.93[a]	.67	Beck, Ward, Mendelson, Mock, & Erbaugh (1961)
Beck depression Inventory-II (BDI-2nd edition)	21	self-report	.92-.93[b]	.71[b]	Beck, Steer, & Brown (1996)
Center for Epidemiologic Studies Depression Scale (CES-D)	20	self-report	.85-.90	.83	Radloff (1977)
CED-S Short Form	10	Self-report	.75-.82	.74	Cole, Rabin, Smith, & Kaufman (2004)
Depression, Anxiety and Stress Scale (DASS21-D) Depression Subscale	7	Self-report	.91[c]		Lovibond & Lovibond (1995) as cited in Argus & Thompson (2008)
Hamilton Depression Inventory (HDI)	38[d]	Self-report	.89[d]	.93 correlation with BDI-I	Reynolds & Kobak (1995)
KDS	6	Self-report	.79	.72 correlation with SCL-90	Kandel & Davies (1982)

Profile of Mood States POMS-D (Depression subscale) Short Form	8	Self-report	KR 20 values of .84 to .95[e]	.80 correlation with BDI	McNair, Lorr, & Droppleman (1971) and Malouff, Schutte & Ramerth, (1985)
Symptom Checklist SCL-90-R Depression dimension	13[f]	Self-report	.90[f]	.75 with Wiggins Depression scores[f] and .68 with Tryon Cluster Depression Scores of the MMPI[f]	Derogatis (1983) and Derogatis, Rickels & Rock (1976)
Hospital Anxiety and Depression Scale (HADS)—Depression subscale	7	Self-report	.90[g]	.79[g]	Zigmond & Snaith (1983)
The Mood and Anxiety Symptom Questionnaire (MASQ) Short Form—General Distress Depression subscale	12	Self-report	.92	.67 correlation with BDI	Watson & Clark (1991) as cited in Watson, Weber, Assenheimer, Clark, Strauss & McCormick (1995)

[a]Farmer (2001); [b]Arbisi (2001); [c]Lovibond & Lovibond (1995, as cited in Argus & Thompson, 2008); [d]Fernandez (1998); [e]Eichman (1978); [f]Payne (1985); [g]Martin (2003)

Chapter 2: Method

This chapter includes a brief description of the purpose of a meta-analysis, and a more extensive description of the process used to conduct a meta-analysis. The process used to conduct the meta-analyses in this study is described below in terms of steps taken.

Definition of Meta-Analysis

The term meta-analysis was coined by Gene Glass (Cooper & Hedges, 2009), and he defined it as “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings” (Glass, 1976, p. 3). Cooper and Hedges (2009) define meta-analysis as “the quantitative procedures that a research synthesist may use to statistically combine the results of studies” (p. 6). Konstantopoulos (2013) best captured the whole concept with the following definition: “Meta-analysis refers to the statistical methods that are used to combine quantitative evidence from different primary research studies that test comparable hypotheses for the purpose of summarizing evidence and drawing general conclusions” (p. 232).

Meta-Analysis Procedure

Meta-analysis is a multi-step procedure, so there are several decisions to make in conducting a meta-analysis. The steps include (1) searching the literature, (2) coding studies, (3) choosing an appropriate statistical model, (4) combining effect sizes, (5) testing for and explaining heterogeneity of effect sizes, (5) conducting moderator or subgroup analyses, (6) assessing evidence of publication bias (Borenstein et al., 2009;

Card, 2012). Following are the steps associated with the procedures of conducting a meta-analysis. In each step, the rationale for the step is followed by the procedure for implementing the analysis.

Institutional Review Board

This study was determined by the Office of Research and Sponsored Programs at the University of Denver to not require review by the Institutional Review Board (IRB). A copy of the IRB determination letter is can be found in Appendix C.

Literature Search Process

Various search terms and search strings were used for different databases: *perfectionism, multidimensional perfectionism, depression, multidimensional perfectionism scale, almost perfect scale-revised*. Whenever possible in searching the different databases, the results were limited to studies published in English, and studies published between 2007 and June 2017, and studies that used participants of 17 years of age or older (college age or older). The database and website searches were conducted until June 19, 2017.

Searches for published studies. Only two electronic databases were used to search for published studies. The first and main electronic database searched was the American Psychological Association's PsycINFO database because this was the most relevant database available for the topic of the nine meta-analyses. First, PsycINFO was searched for "depress* AND perfect*" without specifying what fields the database should search. When using the EBSCOhost platform to search PsycINFO, if the field to be searched is not specified, all fields are searched (American Psychological Association, 2012). The asterisk is the truncation sign, and the truncation sign returns variations of the roots “depress” and “perfect,” such as *perfectionist* and *depressive* as well as *depression* and *perfectionism* (EBSCO Help retrieved March 18, 2017). Without adding any limits to the results, the search for “depress* AND perfect*” in all fields yielded 1,661 results. The results of the search were then limited to studies published between 2007 to 2017, and that search yielded 1,044 results. The PsycINFO search was then further narrowed to articles in which the participants were of age 18 years or older, and that yielded 714 results. Then the search was further narrowed to studies written in English, and that yielded 682 results. Several different search strings were tried in PsycINFO in order to find the optimal search string that returned the largest number of relevant results. When PsycINFO was searched using the following search string:

> depression AND TM (“multidimensional perfectionism scale” OR “almost perfect scale-revise)

and the results were limited to studies published from 2007 to 2017 and to studies written in English and to studies with participants age of 18 years or older, there were 267 results. The TM specification stands for tests and measures and specifies the exact tests and measures used in the study, and the TM specification of “multidimensional perfectionism scale” retrieved article records with both the Hewitt and Flett (1991b) MPS and the Frost et al. (1990) MPS. However, it was found that not all articles that used either the HMPS or the FMPS were found using the TM specification because the article records did not always contain all the tests and measures used in the studies, so using the TM specification was not retrieving all of the relevant results.

The next search string that was used in PsycINFO was the following:

TX (depression AND perfectionism) AND TX (dimension OR "multidimensional perfectionism scale" OR "almost perfect scale-revised")

with results limited to studies published between 2007 and 2017, and to studies written in English, and to studies with participants age 18 and older, and this search yielded 178 results. However, it was found that having the search term *dimension* in the search string was retrieving too many non-relevant results and not retrieving all of the relevant results because this search string was retrieving results that did not use any of the multidimensional perfectionism measures and was retrieving results that just had the word *dimension* in the PsycINFO article record. The search string that was found to return the largest number of relevant results and the fewest non-relevant results when searching PsycINFO database was the following:

(SU (depression) OR KW (depression)) AND (SU (perfectionism) OR KW (perfectionism)) AND (TX ("multidimensional perfectionism scale" OR "almost perfect scale-revised"))

where TX indicated searching the whole text of the PsycINFO article record (title, abstract, keywords and subject terms), and KW indicated keyword, and SU indicated subject term, and putting search phrases in quotation marks found those phrases with the relevant words appearing in that order and adjacent to each other. The subject term (SU) field in APA databases uses controlled vocabulary from APA's *Thesaurus of Psychological Index Terms* (Retrieved June 15, 2017 from http://www.apa.org/pubs/databases/training/thesaurus.aspx). The results of the above search string were further limited to studies published between 2007 and June of 2017, to studies published in English, and to studies that used participants who were of 18 years of age or older, and this PsycINFO search yielded 129 results. Each study from this last set of results was examined to determine, first, if the study did in fact use one of the three relevant multidimensional perfectionism measures and, second, if the study reported a correlation between a measure of depression and at least one of the nine subscales of interest from the three relevant multidimensional perfectionism scales. Studies that only reported correlations between depression and combined scores from the multidimensional perfectionism subscales (from composite scores), and not for any of the nine relevant subscales individually, were not included.

PsycINFO was also searched extensively for relevant studies on the topic of the relationship between depression and perfectionism during the summer of 2016 and studies found then were included in the sample of studies for this dissertation.

As Card (2012) recommended doing during the literature search process, an Excel spreadsheet was used to create a database of information about the studies found during the literature search. A record for each separate, potentially relevant study found during the literature search was entered as one row in the Excel spreadsheet/database. Each study was given a unique identification number so that the studies could be referenced and cross-referenced by that identification number. The record for each study contained the study's bibliographic/citation information, including the year the article was published, the last names of all the authors, the title of each study, and the abstract for each study when the abstract could be cut and pasted. The date that each study was found and the search terms used to find each study were also incorporated into the study's record. Since not all databases allowed the results to be limited to those published in

English, the language that each study was written in was also recorded if it was other than English. In the process of determining which studies should definitely be included or not be included in one or more of the nine meta-analyses, when it was determined that a study would not be included, the reason for exclusion was entered into that study's record in the Excel spreadsheet/database. When a study used one or more of the relevant perfectionism measures translated into a language other than English, that information was included in the study's record as a reason for exclusion from the meta-analyses.

The other database searched for published studies was ScienceDirect. When ScienceDirect was searched, an advanced search was done with the terms *multidimension perfectionism* AND *depression,* searching for both terms in the title, abstract and keywords of each article record, with the results limited to just psychology and social science journals, and this search yielded 21 results. ScienceDirect does not allow the specification of the age group of the study participants nor does it allow filtering for journal articles published only in English. The information about these 21 resulting studies were added to the Excel spreadsheet database of possibly relevant articles in addition to the many possibly relevant studies found from the various PsycINFO searches, not just the final PsycINFO search, and the published studies located when writing the review of the literature on perfectionism for this study.

At the end of the literature search, any study that reported a correlation that was relevant to one or more of the nine meta-analyses was included even if the relationship between perfectionism and depression was not the focus of the study.

Searches for grey literature and unpublished studies. The primary databases and websites that were searched for grey literature or unpublished studies were ProQuest Dissertations & Theses Global and the American Psychological Association's gray literature database, PsycEXTRA. According to the PsycEXTRA fact sheet (APA, n.d.), PsycEXTRA is APA's best resource for grey literature, and contains the most recent conference presentations and papers, and it uses index terms from the *Thesaurus of Psychological Index Terms*, and the content of PsycEXTRA does not overlap with the content of PsycINFO database. When PsycEXTRA was searched for the terms *perfectionism* (in abstract) and *multidimensional perfectionism* (in abstract), there were 15 results. The PsycEXTRA database advertised that it included the full text of more than 70% of study records it contains (APA, n.d.), but the full text for the relevant studies found in PsycEXTRA usually only included the abstract and sometimes a short summary of the results for each study and did not included the relevant correlations, so the studies that might have had relevant correlations could have been followed up by emailing the researchers with a request for the relevant correlations, but time constraints did not permit this.

Next, an advanced search was done in ProQuest Dissertations and Theses using the following search string:

ab(*multidimensional perfectionism*) AND ab(*depression*)

where *ab* indicated searching in the abstract. Not having the term *multidimensional perfectionism* in quotation marks gave five more results than when quotation marks were used, so quotation marks were not used, and then the number of results was 16, and 14 out of those 16 results were written in English. The full text for one of the dissertations that was written in English could not be retrieved because that dissertation was embargoed. Only two dissertations were found that gave relevant correlations and that were not later published as journal articles, and these two dissertations were the only relevant unpublished studies that were able to be retrieved and used in the set of nine meta-analyses.

Other databases and websites where gray literature or unpublished studies were searched for were National Institute of Health (NIH, http://www.nimh.nih.gov/index), the American Educational Research Association (AERA, http://www.aera.net/) ,the American Evaluation Association (AEA http://www.eval.org/search11/search.asp), OpenGrey (http://www.opengrey.eu/), the National Institute for Mental Health (NIMH, www.nimh.nih.gov), and Grey Literature Report (http://www.greylit.org).

Searching the NIH website for *perfectionism* and *depression* found no relevant results. The American Educational Research Association Annual Meeting Online Portal for years 2017, 2016, 2015, 2014, and 2012 was searched using the terms *perfectionism* AND *depression* because the portals for those years allow searching for topics, and there was a total of 5 results for all of those years. For the year 2013, 2011 and 2010 the AERA website only allowed searching in the title of the session and paper submissions for the terms p*erfectionism* and *depression*, and there were no results. For the American Evaluation Association session titles of conference programs were searched using the terms *perfectionism* and *depression* for the years 2014, 2015 and 2016 because for those years the AEA conference programs only allowed searching in the titles of the sessions, but there were no results. For the years 2007 through 2013, the website allowed searching for a keyword in the titles and abstracts of the sessions of the annual conferences, and when *perfectionism* and *depression* were used as the keywords, there were no results. OpenGrey (http://www.opengrey.eu/) was searched for *perfectionism* and *depression,* and there were four results for the time period 2007 to 2017, and they were all PhD “Thesis,” and the full text for these could not be retrieved. Grey Literature Report (http://www.greylit.org) was produced by the New York Academy of Medicine between 1999 and 2016 and was discontinued in January 2017, but previous documents were still accessible, so it was searched with the terms *perfectionism* and *depression,* and there were zero results. The website for the National Institute for Mental Health was searched using the terms *perfectionism* and *depression*, but there were no results.

In general, while searching electronic databases and websites during summer of 2016 and June of 2017, the terms *multidimensional perfectionism* AND *depression* where searched for in the abstract, title, and keywords of each article record in the databases.

After the literature search begins for a meta-analysis and relevant studies are found, backward searching is conducted. After relevant studies are found, each study is read completely from beginning to end, and additional relevant studies are found from those cited or mentioned (Card, 2012). Backwards searching was performed while conducting the literature review on perfectionism and while preparing for a poster presentation for a 2016 APA poster on the topic of the relationship between perfectionism and depression. It should be noted that backward searching can have the problem of only finding the relevant studies that obtained the results desired by the researchers who conducted them, such as statistically significant results or results that confirmed the researchers’ hypotheses; therefore, the studies found using backwards searching might be a biased sample (Card, 2012), so it is unlikely to be critical that extensive backward searching was not conducted.

Literature search results. From all the searches for both published and unpublished studies, a total of 259 unique and possibly relevant search results were identified in the various database and Internet searches, and a database was created in Microsoft Excel that contained the bibliographic information for those 259 possibly relevant search results. Out of the 259 results, six were excluded because the articles were not written in English. Out of remaining possibly relevant studies, full text for the results could not be retrieved for four results that looked relevant, and one of these studies for which the full text could not be retrieved was an embargoed dissertation. The remaining studies were first checked to see if they used at least one of the three multidimensional perfectionism scales that were the focus of this study, and 14 studies were excluded because they did not use one of those three scales. The remaining studies were checked to see if they used a measure of depression, and 15 studies were excluded because they used no measure of depression. The remaining studies were checked to see if the scales had been translated and administered in a language other than English, and 34 articles were excluded because they had used versions of the scales that had been translated into languages other than English. Because all three of the multidimensional perfectionism scales that were the focus of this study were created in English, the three scales might not be measuring the same constructs if they are translated into languages other than English. Seventeen studies were excluded because they used participants younger than 17 years of age. Treatment-by-control group designs and group-based designs were not included unless they reported correlations for the whole sample, so 11 studies were excluded because they were treatment-by-control group designs that did not report correlations for the whole sample (treatment and control participants together) or did not report correlations at all. Out of the remaining records, 16 were excluded because they were conference posters or presentations that were either not relevant or that there was not time to email the author to ask for the relevant correlations. Full text of the remaining search results was screened to determine if they reported relevant correlations, and 44 studies were excluded because they did not report relevant correlations, and 26 studies were excluded because they only reported correlations for composite scores. Five studies were excluded because they had samples of participants who were either extreme cases or were not from a population to which it was desired to generalize the results of these meta-analyses (e.g., post-partum women in whom the relationship between perfectionism and depression might be expected to differ from the general public). Eight studies were excluded because the researchers modified the multidimensional perfectionism scale to such an extent it could not be determined if the modified versions were measuring the same constructs as the three original multidimension perfectionism scales that were the focus of this study. One study was excluded because it appeared to use the same sample of participants as a previous study by the same authors. One study was excluded because it reported correlations separately for the CES-D subscales and not for the whole scale. Three of the dissertations that had relevant correlations and were not excluded for other reasons were excluded because it was found that they were later published as journal articles, so just the corresponding three journal articles were included. Two relevant studies were misplaced in the search process and not found until the data had been analyzed. The original 259 possible studies were narrowed down to a total of 52 studies to be used in one or more of the nine meta-analyses. The references for the 52 included studies, the one study excluded because it used a duplicate sample, and the two misplaced relevant studies can be found in Appendix B. The demographic characteristics of the participants in the 52 included studies can be found in Appendix D.

Inclusion and Exclusion Criteria. Studies that reported at least one correlation between depression and one of the nine subscales from the relevant multidimensional perfectionism measures could potentially be included in the one or more of the nine separate meta-analyses even if the relationship between perfectionism and depression was not the focus of that study. To be included in any of the meta-analyses for this study, individual empirical studies had to use one or more of the nine subscales from the three multidimensional perfectionism measures that were the focus of this dissertation. Studies that reported correlations that were based on composite scores from a combination of the relevant perfectionism subscales were not included because the focus of this study was to examine the relationship between depression and each dimension of perfectionism separately from the other dimensions of perfectionism. The three multidimensional perfectionism measures that were the focus of this study were originally created in English, so studies in which the relevant perfectionism subscales were translated into other languages and not administered in English were excluded from the nine meta-analyses even if they reported a relevant correlation. Studies were excluded if they used participants who were younger than 17 years old.

The sample of studies used in these meta-analyses were both published and unpublished studies conducted between 2007 and 2017 that reported a correlation between one of the specified dimensional subscales of perfectionism and depression.

Wilson (2009) stated that the selected time frame should not be arbitrary, but should be based on theory. The best theoretical time frame for the meta-analyses would have been from 1990 to present because 1990 was when the first of the three multidimensional measures of perfectionism was created, the Frost et al. (1990) FMPS, but that would have been a 27 year-long time frame, which would not have been feasible for this dissertation because of time constraints. In addition, 1027 results were found when PsycINFO was searched using the following search string: TM multidimensional perfectionism scale" OR TM "almost perfect scale-revised" In this search string, TM searches for tests and measures listed in the study record. Since no other theory-based time frame was found, a ten-year time frame was chosen: 2007 to June of 2017. Finally, only studies written in English were used, and this may have created some bias in the set of meta-analyses that were conducted (Card, 2012).

Also, because a Fisher's z_r transformation of the correlation coefficient was used and because for that transformation the large sample approximations are accurate for samples of at least 20 participants per study, only studies with sample sizes of at least 20 participants were used (Hedges, 2009). Since the purpose of these meta-analyses was not to infer causation, many different types of research designs were appropriate for use in the meta-analyses; therefore, as many different types of research designs as possible were included in determining the strength of association (Cooper, 2009). However, according to Stoeber and Otto (2006), there are two basic types of research designs in the literature on perfectionism: group-based designs and dimensional designs. According to Stoeber and Otto (2006), in group-based designs the participants are separated into groups of adaptive perfectionists, maladaptive perfectionists, and non-perfectionists based on cutoff scores on the Hewitt and Flett (1991b) MPS and the Frost et al. (1990) MPS and/or based on participants' scores on the Discrepancy and Standards subscales of the APS-R by Slaney et al. (2001). Treatment group studies were only included if the relevant correlations were given for the whole sample of participants and not just for the separate groups. These were not group-based designs that divided groups based on cut-off scores on the three multidimensional measures of perfectionism that were the focus of this study. Group-based designs that divided participants into groups based on cut-off scores on the three multidimensional measures of perfectionism were not used.

For the dissertations that were later published as journal articles, those journal articles were used in the meta-analyses rather than the preceding dissertations. Only two dissertations were found that were relevant and that were not later published as journal articles, and those two dissertations were the only unpublished studies that were found that had full text available.

Again, studies included were limited to those that reported a correlation between one of the nine relevant perfectionism subscales and a measure of depression, so studies that did not report relevant correlations were excluded.

The age range for participants in the included studies was college-aged students and older persons, which meant participants who were 17 years old and older. Studies of children with depression and/or perfectionism were excluded because perfectionism may be related to depression in a different way in children than in adults and because there are different scales for measuring perfectionism in adults than in children (Flett et al., 2016). This study investigated the relationship between perfectionism and depression in adults only. Since depression occurs more frequently in women than in men (Rittberg, 2016), both males and females were included. It was originally planned that a moderator or subgroup analysis would be conducted to determine whether the relationship between dimensions of perfectionism and depression is different for females than it is for males (Borenstein et al., 2009; Card, 2012); however, not enough studies were found that used only men and that used only women to conduct such a moderator analysis.

Duplicate studies. Studies were checked to see if they used the same participant data as other studies by some or all of the same authors. To determine if studies used duplicate data the heuristic by Wood (2008) was used. This heuristic assumes that researchers are not trying to be deceptive and are not trying to unethically produce more than one publication from each dataset (Wood, 2008). The heuristic by Wood (2008) asked whether some or all of the authors were the same, whether the measures were the same, whether the participants were recruited in the same way, whether the research questions were the same, and then the last question in the heuristic was “Are matched study effects sufficiently different to exclude the study?” (p. 81), and if the answer is “yes” the studies are not considered to be duplicates. When duplicate studies were found, the study that had the most information was chosen (Vanchu-Orosco, 2012), meaning the study with the largest sample size or the study that reported the most relevant correlations between perfectionism subscales and depression. Of all the published studies that were found to meet all the criteria for inclusion in at least one of the nine meta-analyses, only one pair of studies seemed to be duplicates according to Wood’s (2008) heuristic for identifying duplicate studies. The studies by Akram, Ellis, and Barclay (2015) and by Akram, Ellis, Myachykov, Chapman, and Barclay (2017) both studied the same topic and the sample sizes were almost equal and they used the same scales, and the titles were the same except that the 2015 study had the subtitle “A longitudinal study,” and the samples participants were recruited in the exact same way, and they were published in different journals. Akram et al. (2017) cited Akram et al. (2015) and said that the Akram et al. (2017) study added additional information to the findings of Akram et al. (2015) study, but Akram et al. (2017) did not say it used the same data as Akram et al. (2015). Both Akram et al. (2017) and Akram et al. (2015) provided correlations for the relationship of depression with CM, DA, PE, PC, PS from the Frost et al. (1990) FMPS and SOP and SPP from the Hewitt and Flett (1991b) HMPS, and the pattern of the correlations was slightly different for the two studies. Because of all the similarities between these two studies and because the correlations differed by such a small amount, it was decided that these two studies must have been using at least some of the same participants. Akram et al. (2015) was included instead of Akram et al. (2017) because Akram et al. (2015) gave more information, such as the mean age and age range for all participants combined.

Measures used for the analysis. According to Sirois and Molnar (2016), the three most commonly used measures of perfectionism are the FMPS, the HMPS, and the Almost Perfect Scale-Revised (APS-R). These three measures of perfectionism were the only measures of perfectionism used in this study.

Articles referencing use of both the first and second versions of the Beck Depression Inventory (BDI and BDI-II) were included. Also, articles referencing use of the CES-D (Radloff, 1977) were used as well as other measures of depression that were found in research that also used the three specified multidimensional measures of perfectionism. A total of 11 different measures of depression were included in the nine meta-analyses.

Coding Process

Developing a coding form and coding protocol. Developing the coding form and codebook was an iterative process. It started with a rough outline of the codebook, and then a few studies were coded, and the codebook and coding form were revised. Two people coded studies for these nine meta-analyses. The first person was the primary researcher, and the second person who coded studies was a Ph.D. student in the same Statistics and Research Methods Program as the primary researcher. This second coder had extensive knowledge about and experience with coding because she had done a meta-analysis for her Master's Degree thesis, so she had many good suggestions about how to gradually improve the coding process and the codebook and coding sheet. The codebook and coding form were created with an Excel spreadsheet because the two coders emailed copies of that spreadsheet back and forth to each other. After both coders coded the same first ten studies in order to calculate interrater reliability, the two coders coded five different studies each and exchanged questions and suggestions for improving the coding process via email. Discussing these questions and suggestions via multiple email messages led to the coding process and the Excel codebook and coding sheet being continuously improved throughout the coding process in an iterative manner. A copy of the final version of the codebook is provided in Appendix A.

According to Orwin and Vevea (2009) and Brown et al. (2003), coders need substantive expertise in order to improve accuracy of coding judgments. For this study, substantive expertise was gained by the primary researcher by reading several literature reviews and empirical research articles on perfectionism and depression and the relationship between those two constructs while writing the literature review in the first chapter. Less substantive expertise was needed for these nine meta-analyses because only low inferences codes were used, and low inference codes reduce the need for substantive expertise (Orwin & Vevea, 2009). Low inference codes “require the coder only to locate the needed information in the research report and transfer it to the database” (Cooper, 2009, p.33). Low inference codes also reduce coder error (Orwin & Vevea, 2009), and they improve reliability (Wilson, 2009).

It was initially proposed that a coding scheme would be developed by following Brown et al.’s (2003) example, which involves taking a random sample of studies from the relevant literature and using this sample of studies to determine all the relevant variables that should be included in the coding form. In developing a coding form and coding book or protocol for one of their meta-analyses, Brown et al. started by thoroughly reviewing 50% of the relevant studies in order to determine all the variables that should be included in their coding form and that should be defined in their coding book/protocol. When thoroughly reviewing a sample of the relevant literature for relevant variables to be coded, Brown et al. recommended starting with the following “methodological and substantive features…for the purpose of relating these characteristics to study findings” (p. 207): study source, publication year, type of research design, and characteristics of authors/investigators such as discipline and educational credentials. It was found that the characteristics of authors/investigators was not usually apparent from the studies, except that most of them had Ph.D.s or were graduate students at universities, so information about characteristics of authors/investigators was not coded, but study source, publication year, and type of research design were coded. As studies were coded, the codebook and coding sheet were incrementally improved, and additional variables were added to the codebook, and the coding of variables that were already included was improved.

Variables that were coded during the literature search. The location where the paper was found (which database or other location) and the date that study was found were coded (Card, 2012). Also, when a study was excluded, the identifying information for that study and the reason for its exclusion were coded (Card, 2012). Each study was given an identification number rather than organizing studies by the surnames of the authors. Card (2012) said that giving each study an identification number helps to organize all the papers found in the literature search. The following citation information for each paper found in the literature search was recorded in an Excel spreadsheet: year of publication, author(s), title of the paper, and the source from which the paper came (Card, 2012). These were columns in the Excel spreadsheet in which each row was a separate paper (Card, 2012). Also, the type of papers found in the literature search (such as, empirical, theoretical, conference presentation, dissertation, thesis, or book chapter) were coded because the reference lists from the theoretical papers and literature reviews on the relationship between perfectionism and depression were useful for finding more studies to include as data in the meta-analyses (Card, 2012)

Study characteristics coded. Card (2012) recommends including at least the following four study characteristics: “characteristics of the sample, measurement, design, and source” (pp. 65-68). All of the selected studies were observational or nonexperimental because it is neither ethical nor possible to randomly assign perfectionism or depression to study participants and because there was no attempt to infer causation (Gliner, Morgan, & Leech, 2009), and as was stated earlier, it was best to include as many studies as possible because internal validity was not an issue (Cooper, 2009). Also, as Brown et al. (2003) recommended, the study year, the source of the study, and the type of study were coded. Also, whether the study was published or not and the format in which the study was written was coded so that a moderator or subgroup analysis could have been used to look for evidence of publication bias if there had been enough unpublished studies to do a moderator analysis, but there were not enough studies to do moderator analyses (Borenstein et al., 2009: Card, 2012). The specific measures of depression and perfectionism that were used in the study were also coded.

When a study reported the Cronbach's alpha or internal consistency reliability coefficients estimated from the sample for the different measures used in the study, those reliability coefficients were coded for the different measures of perfectionism and depression. When a study gave these reliability coefficients for the sample, those values were used for the reliability of the scales, and when studies did not report the reliability estimated from the sample, the reliability estimate from the psychometric development of the scales was used.

As the purpose of the present study was to provide an estimate of a correlation, internal and external validity of the source studies was not a focus.

Study participant characteristics coded. Card (2012) also recommend coding characteristics of the sample of participants in each study included in the meta-analysis in order to know to what populations of study participants you can generalize the results of the meta-analysis. Characteristics of the study participants, such as ethnicity/race, gender/sex, status of the participants (such as inpatient and outpatient, whether participants had a clinical diagnosis of depression or not, community members, or college students), country of origin/nationality and age range and mean age were coded in order to know how far the results of the meta-analyses can be generalized.

Coding reliability. Since there were two coders for these meta-analyses, interrater reliability was relevant here (Card, 2012). Two people, including the primary researcher, coded studies for these meta-analyses. At the beginning of the coding process, both of the two coders coded the same ten studies separately, and two reliability coefficients were calculated: one for all continuous variables and one for the variables used to calculate effect size because the variables used to calculate effect size are the most important because measurement error is introduced when those variables are coded inaccurately (Hedges & Olkin, 1985, as cited in Yeaton & Wortman, 1993). Since a reliability coefficient for interrater reliability does not measure exact agreement but rather measures the covariance between coded values (Orwin & Vevea, 2009), the reliability coefficient was estimated conservatively by entering a zero for any value that one of the two coders completely missed coding, so that there was zero covariance between those two coded values. For sample size and correlations (the important information for calculating effect size) there was only one disagreement between the primary researcher and the second coder, and interrater reliability was estimated for coding of these two variables, and it was $r = .999$ because the two coded values for the one disagreement were so close. A correlation coefficient was calculated for the reliability of the coding for all continuous variables for the first ten studies, the reliability coefficient was $r = .987$. Also, after coding only ten studies each, the percentage of agreement was calculated for all the categorical variables and there was 80% agreement between the two coders on the coding of the categorical variables. The two disagreements were for the *Sample Type* code, so the primary researcher revised the codes for Sample Type in the Excel codebook to 0=Not specified at all, and 2=Non-college adults/general, so the two coders had 80% agreement on Sample Type before that code was revised in the codebook and 100% agreement on all other categorical codes. In the final version of the codebook, the Sample Type code was eventually taken out and replaced by *Population* code, where the directions where, “code a few words that describe the population that the study participants represent.”

Combining Effect Sizes

Summary statistics used to estimate effect size. The correlation coefficient *r* can be considered an effect size (Borenstein et al., 2009). The correlation coefficient *r* is standardized, so it is unit free, and it allows comparison between measures that are on different scales (Bobko, 2001). The correlation coefficient *r* captures the strength and direction of the association between two continuous variables (Bobko, 2001).

Calculating the correlation effect size for each study. Because the absolute value of a correlation is limited to the range between 0.0 and 1.0, the sampling distribution of the correlation coefficient *r* is not normal but is skewed (Cohen, Cohen, West, & Aiken, 2003). Therefore, before combining correlation coefficients as effect sizes in a meta-analysis, the correlation coefficients are usually transformed using Fisher’s z_r transformation (Borenstein et al., 2009). Schmidt and Hunter (2015) recommended not using Fisher’s z_r transformation of correlations because they said it causes the mean correlation from a meta-analysis to be upwardly biased, but Card (2012) said that Fisher's z_r transformation is an effect size that is "roughly normally distributed around a population effect size" and therefore is beneficial for use in meta-analyses and in creating funnel plots to look for evidence of publication bias (p. 264). Fisher's z_r transformation for the correlation coefficient was used in these meta-analyses because when combining the correlation coefficient effect sizes from the separate studies, large sample approximations for correlation coefficients are only accurate for samples of at least several hundred participants (Hedges, 2009). For Fisher's z_r transformation of correlation coefficients, the large sample approximations are accurate for samples of at least 20 participants per study (Hedges, 2009). The formula for Fisher's z_r transformation of the correlation coefficient is as follows:

$$z = 0.5 \times ln\left(\frac{1+r}{1-r}\right) \tag{1}$$

"where $ln(x)$ is the natural (base e) logarithm of x" (Shadish & Haddock, 2009, p. 264). Fisher's z_r transformation has a sampling distribution that is approximately normal (Bobko, 2001). An approximation of the variance for Fisher's z_r transformation of a correlation coefficient is as follows:

$$V_z = \frac{1}{n-3} \tag{2}$$

where *n* is the sample size for the study (Borenstein et al., 2009).

The standard error for Fisher's z_r transformation is as follows (Borenstein et al., 2009):

$$SE_z = \sqrt{V_z} \tag{3}$$

The Fisher's z_r values were used in the analysis to calculate a mean or summary effect size and confidence intervals for the meta-analysis, and then those results were transformed from Fisher's z_r back to the correlation coefficient using the following formula (Borenstein et al., 2009):

$$r = \frac{e^{2z}-1}{e^{2z}+1} \tag{4}$$

Correction of effect sizes for artifacts. According to Card (2012), there is a debate about whether to correct effect sizes for study artifacts such as reliability of the measures used, imperfect validity of the measures used, artificial dichotomization of naturally continuous variables used in computing effect sizes, and range restriction of the measured variables. According to Card, some researchers, including Card himself, argue that the meta-analyst should correct for study artifacts because the interest should be in the association or effect size between the latent constructs that are measured and not the association between the specific scales used to measure the latent constructs. Also, according to Card, some disciplines customarily correct study effect sizes for artifacts and other disciplines do not, but Card says that the decision about whether to correct for study artifacts should be based on the conceptual knowledge that the meta-analyst has about the topic of the meta-analysis and the empirical information found in the sample of studies used in the meta-analysis and not on whether the researcher's disciplinary field traditionally does or does not correct effect sizes for study artifacts. According to Card, if effect sizes for the meta-analysis are corrected for study artifacts, the standard error for each study needs to be adjusted. The strength of a correlation coefficient is attenuated by measurement error and range restriction (Glass & Hopkins, 1996). Rosenthal (1991) argued that meta-analysts should not correct for study artifacts because the interest should be in the results of studies that actually exist and not in the results of hypothetical ideal studies. Rosenthal (1991) also argued against correction for artifacts by making the point that such corrections can yield inaccurate results, for example, a correction for reliability attenuation can yield a correlation greater than 1.0. However, the correlations from the studies used in these nine meta-analysis were corrected for measurement error and so was their corresponding Fisher's z_r variances. When the reliability estimates for the scales were not provided by the individual studies, the reliability estimate from the psychometric development of each scale was used. The correlations for the nine meta-analyses were corrected with the following formula:

$$unatt\_ESr = \frac{ES_r}{\sqrt{r_{XX}}\sqrt{r_{YY}}} \tag{5}$$

and the Fisher's z_r variance was corrected using the following formula:

$$unatt\_Vzr = \frac{V_{Zr}}{(r_{XX})(r_{YY})} \tag{6}$$

These corrections were done in Microsoft Excel prior to importing each dataset into the R statistical software. In the R statistical software, it was specified with syntax that the unattenuated effect sizes (study correlations that had been corrected for attenuation due to measurement error) were used with the "escalc" command, and it was also specified with syntax that the unattenuated Fisher's z_r variances, which had to be calculated in Excel, were used in running the meta-analyses with the "rma" command in Metafor (Viechtbauer, 2010)

Data Analysis

Setting up the data. Raw correlation coefficients along with coding of study descriptors and potential moderator variables were coded directly into an Excel file by the two people who coded the studies.

Software for the statistical analyses. R Statistical Software version 3.4.1 (2017-06-30) — "Single Candle" (R Core Team, 2017), which is a "free software environment for statistical computing and graphics" (https://www.r-project.org/), and RStudio version 1.0.153 open source edition, which is a "an integrated development environment (IDE) for R" (https://www.rstudio.com/products/RStudio/), and version 2.0-0 (2017-06-22) of the R package "metaphor" (Viechtbauer, 2010) were used to run the nine separate meta-analyses to estimate the correlation mean effect size between each of the nine relevant dimensions of perfectionism (or nine perfectionism subscales) and the relevant measures of depression.

Selection of the model for the meta-analyses. When conducting a meta-analysis, a choice must be made about what statistical model to use (Borenstein et al., 2009; Card, 2012). A researcher can choose between a fixed-effects model, a random-effects model, or a mixed-effects model (Card, 2012).

Fixed-effects model. The fixed-effects model assumes that all studies share a common true effect size in the population (Borenstein et al., 2009). Under the fixed-effects model, the effect size for each study is as follows:

$$Y_i = \theta + \varepsilon_i \tag{7}$$

where Y_i is the observed effect size for study i, and θ (the Greek small letter Theta) is the one common true population effect size, and ε_i is the within-study sampling error for study *i* (Borenstein et al., 2009). Or according to Viechtbauer (2010) the fixed-effects model is as follows:

$$y_i = \theta_i + e_i \tag{8}$$

"where y_i denotes the observed effect in the *i-th* study, θ_i the corresponding (unknown) true effect, e_i is the sampling error, and $e_i \sim N(0, v_i)$" (p. 3).

The random-effects model. The random-effects model allows each study to have its own true effect size. Keeping with the notation from Borenstein et al. (2009), the random-effects models is:

$$Y_i = \mu + \zeta_i + e_i \tag{9}$$

where Y_i is the observed effect size for study i, and μ is the mean of all the effect sizes in the population distribution of effect sizes because each study is assumed to estimate a separate effect size, and ζ_i (the Greek small letter Zeta) is the deviation of each study's true unique effect size from the mean effect size for the distribution of effect sizes, and e_i is the deviation of the observed effect size for each study from its true effect size parameter in the population (e_i is the sampling error) (Borenstein et al., 2009). Or according to Viechtbauer (2010) the random-effects model is as follows:

$$\theta_i = \mu + u_i \tag{10}$$

"where $u_i \sim N(0, \tau^2)$. Therefore, the true effects are assumed to be normally distributed with mean μ and variance τ^2" (p. 3). The effect sizes for a random-effects model are hypothetically a random sample from a distribution of effect sizes (Hedges & Vevea, 1998). A random-effects model has two sources of variance: sampling error or within-study variation and between-studies variation (Borenstein et al., 2009). The population parameters for the variance of the sampling error or within-study variation is σ^2 (with sample statistic V_{Y_i}) and for the between-studies variance is τ^2 (with sample statistic T^2) (Borenstein et al., 2009).

Mixed-effects model. Mixed-effects models have both a fixed-effects component and a random-effects component (Raudenbush, 2009). According to Borenstein et al. (2009) a mixed-effects model would be a subgroup analysis where the within-group summary effect (or mean effect size within-group) is sampled from a random distribution of effect sizes, and if the meta-analysis were replicated, the exact same subgroups would not be used, and the summary effect across groups is fixed. Specifying that the within-group summary effect (or mean effect size within-group) is random allows generalization to subgroups not included in the analysis (Borenstein et al., 2009). According to Viechtbauer (2010) the formula for a mixed-effects model is as follows:

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p'} x_{ip'} + u_i \quad (11)$$

According to Hedges (1992) ‘Statistical methods for mixed effects meta-analyses have received less complete treatment in the literature than have fixed and random effects models” (p. 292).

Estimating summary or mean effect size. When the effect sizes for a set of studies in a meta-analysis are combined, each effect size is weighted to take into account study characteristics such as the precision of the effect size estimate (Shadish & Haddock, 2009). The precision of the estimate due to the within-study sample size can be taken into account with inverse-variance weights (Shadish & Haddock, 2009). The inverse-variance weight for a fixed-effects model is as follows:

$$W_i = \frac{1}{V_{Y_i}} \quad (12)$$

where V_{Y_i} is the within-study variance for study i (Borenstein et al., 2009). For a fixed-effects model, the weighted mean effect size is calculated with the following formula:

$$M = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \tag{13}$$

where W_i is the weight for study i, and Y_i is the relevant effect size for study i, and M is the summary effect or mean effect size for the meta-analysis (Borenstein et al., 2009). Then the variance of the summary effect or mean effect size for the meta-analysis is calculated with the following formula:

$$V_M = \frac{1}{\sum_{i=1}^{k} W_i} \tag{14}$$

The estimated standard error for mean effect size or summary effect is calculated by taking the square root of the above variance of the mean effect size:

$$SE_M = \sqrt{V_M} \tag{15}$$

Then the upper and lower limit of the confidence interval around the summary effect or mean effect size for the meta-analysis are calculated with the following formula at $\alpha = .05$ level of significance or 95% level of confidence:

$$LL_M = M - 1.96 \times SE_M \tag{16}$$

and

$$UL_M = M + 1.96 \times SE_M \tag{17}$$

(Borenstein et al., 2009). Then a z statistic to test the null hypothesis that the common population effect size is zero can be calculated with the following formula:

$$z = \frac{M}{SE_M} \tag{18}$$

(Borenstein et al., 2009).

A random-effects model uses the same formulas to combine effect sizes and to estimate the variance and standard error of the estimate of the mean effect size for the meta-analysis except that all the places where the variance appears in the formula and all the formulas based on the variance are marked by an asterisk to denote that they include the between-studies variance as well as the within-study variance (Borenstein et al., 2009). Thus, the formula for the inverse-variance weights for combining study effect sizes in a random-effects metal-analysis is as follows:

$$W_i^* = \frac{1}{V_{Y_i}^*} \tag{19}$$

Borenstein et al. (2009) use the asterisk in the superscript of W_i^* to distinguish random-effects inverse-variance weights from fixed-effects inverse-variance weights while at the same time showing the similarity between the inverse-variance weights for the two models. The formula for the variance of the estimate of the summary effect size for the random-effects meta-analysis is as follows:

$$V_{Y_i}^* = V_{Y_i} + T^2 \tag{20}$$

where V_{Y_i} is the within-study variance, which differs from study to study, and T^2 is the estimate of the between-studies variance τ^2, which is the same value for all studies in the meta-analysis (Borenstein et al., 2009). The formula for the mean or summary effect for the random-effects meta-analysis is as follows:

$$M^* = \frac{\sum_{i=1}^{k} W_i^* Y_i}{\sum_{i=1}^{k} W_i^*} \tag{21}$$

where the summary effect for the random-effects model is the mean of a distribution of effect sizes (Borenstein et al., 2009).

In a meta-analysis using a fixed-effects model, the precision of the estimate of the effect size increases as the total sample size increases (Borenstein et al., 2009). In a meta-analysis using a random-effects model, increasing the precision of the estimated mean effect size depends not only on the sample size of each study included but also the total number of studies included in the meta-analysis (Borenstein et al., 2009).

Heterogeneity of effect sizes. There can be variation in the effect sizes from studies used in a meta-analysis due to within-study variation (sampling error) and also due to between-studies variation (Borenstein et al., 2009). If there is significant between-studies variation, it indicates that there are real differences in the population effect sizes that are estimated using the observed effect sizes from the sample of studies actually used in the meta-analysis (Borenstein et al., 2009). Between-studies variation needs to be explained or accounted for (Shadish & Haddock, 2009). Between-studies variation can be explained using categorical-level study characteristics as moderators (Card, 2012).

Testing for homogeneity of effect sizes. The significance of the between-studies variance in the effect size is tested with the Q statistic (Borenstein et al., 2009), which is sometimes referred to as a "homogeneity test statistic" (Shadish & Haddock, 2009). The conceptual formula for the Q statistic is as follows:

$$Q = \sum_{i=1}^{k} W_i (Y_i - M)^2 \quad (22)$$

where W_i is the inverse variance weight for study *i*, and Y_i is the observed effect size for study *i*, *k* is the number of studies in the meta-analysis, and *M* is the summary effect for the meta-analysis (Borenstein et al., 2009). The Q statistic is a weighted sum of squares (*WSS*) and not a mean, so it is dependent on the number of studies, and it is also on a standardized scale (Borenstein et al., 2009). The formula for the Q statistic can be written in the following way to show that it is standardized (Borenstein et al., 2009):

$$Q = \sum_{i=1}^{k} \left(\frac{Y_i - M}{S_i}\right)^2 \tag{23}$$

The Q statistic tests the null hypothesis that "all studies share a common effect size" (Borenstein et al., 2009, p. 110). Under the null hypothesis, the Q statistic follows a central chi-square distribution with the degrees of freedom being the number of studies minus one or $(k - 1)$ (Borenstein et al., 2009). Also, $k - 1$ is the expected value of the Q statistic under the null hypothesis that there is no true between-study variance and all observed differences between effect sizes are due solely to sampling error (Borenstein et al., 2009). If the Q statistic is significant, then the null hypothesis that all studies share a common effect size, or that there is no significant between-studies variance, is rejected (Borenstein et al., 2009). However, a non-significant Q statistic does not mean there is no between-studies variance in effect sizes because the Q statistic is a significance test and does not indicate the actual amount of between-studies variation independent of sample size (Borenstein et al., 2009). Also, under certain circumstances, the Q statistic can have low power and fail to detect a meaningful amount of between-studies variation, or it can indicate that a non-meaningful amount of between-studies variation is statistically significant (Borenstein et al., 2009). According to Borenstein et al. (2009) the difference between Q, which is the observed weighted sum of squares (*WSS*), and df, which is the expected weighted sum of squares (*WSS*), is the true difference between the study effect sizes:

$$Q - df \tag{24}$$

Estimating the between-studies variance. According to Borenstein et al. (2009) if a random-effects model is deemed appropriate for the meta-analysis, the between-studies variance can be estimated with the following formula:

$$T^2 = \frac{Q - df}{C} \tag{25}$$

where Q is the observed weighted sum of squares, and *df* is the expected weighted sum of squares, and C is calculated using the following formula:

$$C = \sum W_i - \frac{\sum W_i^2}{\sum W_i} \tag{26}$$

This way of calculating the between-studies variance is often used and is called the *method of moments* or the *DerSimonian and Laird method* (Borenstein et al., 2009, p. 115). The formulas for the DerSimonian and Laird method are included because the DerSimonian and Laird method is conceptually easier to understand because it can be calculated by hand, but restricted maximum likelihood (REML) is often preferred (Borenstein et al., 2009), and REML was used to estimate the between-studies variance in the nine meta-analyses in this study because it has been shown to perform better than most other common methods (Viechtbauer, 2005, as cited in Shadish and Haddock, 2009).

Quantifying and describing heterogeneity in effect sizes. According to Shadish and Haddock (2009), I^2 is a descriptive statistic that does not estimate any underlying population value, and it quantifies the “proportion of total variation in the estimate of treatment effects that is due to heterogeneity rather than to chance” (p. 263). They recommended reporting I^2 as a supplement to the value of the Q statistic in part because, unlike the Q statistic, the value of I^2 does not depend on the metric of the effect size used nor on the number of studies used in the meta-analysis (Shadish & Haddock, 2009). Shadish and Haddock cited guidelines by Higgins and Thompson (2002) for interpreting values of I^2, where $I^2 = 25\%$ indicates a small amount of heterogeneity, $I^2 = 50\%$ indicates a medium amount of heterogeneity, and $I^2 = 75\%$ indicates a large amount of heterogeneity. And, according to Borenstein et al. (2009) "I^2 is the ratio of true heterogeneity to total variance in observed effects, a kind of signal to noise ratio" (p. 120). The formula for computing I^2 is as follows:

$$I^2 = \left(\frac{Q-df}{Q}\right) \times 100\% \tag{27}$$

(Borenstein et al., 2009). According to Borenstein et al. (2009) I^2 can be conceptually understood with the following formula:

$$I^2 = \left(\frac{Variance_{bet}}{Variance_{total}}\right) \times 100\% = \left(\frac{\tau^2}{\tau^2+V_Y}\right) \times 100\% \tag{28}$$

Credibility intervals (*CrI*) were also estimated to describe the distribution of effect sizes (Schmidt & Hunter, 2015) because random-effects models were used in all nine meta-analyses in this study. According to Viechtbauer (2010), the 95% credibility interval "estimates where 95% of the true outcomes would fall in the hypothetical population of studies" (p. 17). In Viechtbauer's (2010) Metafor package for R, estimation of the credibility interval assumes that τ^2 is known rather than estimated, but in actuality, τ^2 is estimated. Credibility intervals are important in random-effects meta-analyses because random-effects models assume a distribution of population effect sizes, whereas the width of a credibility interval in a fixed-effect meta-analysis would be zero because a fixed-effects model assumes one true population value for the effect size, and the value of both the upper bound and lower bound of a credibility interval for a fixed-effect model would be equal and would be the estimate of the one true population effect size, and because in a fixed-effect model, $SD_\rho = 0$ (Schmidt & Hunter, 2015). Credibility intervals are different than confidence intervals because credibility intervals are estimated using the standard deviation of the population correlation whereas confidence intervals are estimated using the standard error of the estimate of the population correlation. According to Schmidt and Hunter (2015) "The credibility interval refers to the distribution of parameter values, while the confidence interval refers to *estimates* of a single value—the value of $\bar{\rho}$" (italics original, p. 228). Schmidt and Hunter (2015) said that the 80% credibility interval is frequently used and is calculated by adding and subtracting $1.28 \times SD_\rho$ from the mean correlation (i.e., that is the critical value for an 80% confidence level times the standard deviation of the population correlation), and "an 80% credibility interval would contain the middle 80% of values in the distribution of population true score correlations" (p. 171). This study used 95% credibility intervals because the R package Metafor only reports 95% credibility intervals (Viechtbauer, 2010).

Explaining heterogeneity in effect sizes. When the assumption of the fixed-effects model that all variation in effect sizes is due to subject-level sampling error is rejected for either theoretical or statistical reasons, there are three options for how to proceed with the meta-analysis: (1) the researcher can use a fixed-effects model and then try to explain the excess variability among effect sizes using coded study characteristics as moderator variables, which is also called a subgroup analysis by Borenstein et al. (2009), (2) the researcher can use a random-effects model, or (3) the researcher can use a mixed-effects model that incorporates both random effects and study characteristics as moderator variables to explain the variability in effect sizes that is not due to subject-level sampling error (Lipsey & Wilson, 2001).

When study effect sizes in a meta-analysis have variability that cannot be explained as subject-level sampling error, that variability could either be systematic variation or random (non-systematic) variation or a combination of both systematic and random variation (Lipsey & Wilson, 2001). The systematic variation that is in excess to the subject-level sampling error could be explained using coded study characteristics as moderators and a fixed-effects model for the meta-analysis; the non-systematic or random variation could be explained using a random-effects model for the meta-analysis; and a combination of systematic and random or non-systematic sources of variation in effect sizes could be explained using a mixed-effects model for the meta-analysis (Lipsey & Wilson, 2001).

Description of moderator or subgroup analysis. When using coded study characteristics as moderator variables in order to explain variability in effect sizes, there are two options for moderator analyses: an analog to analysis of variance (ANOVA) that resembles a one-way ANOVA can be used for a few categorical variables reflecting coded study characteristics, or a modified weighted least squares regression can be used for continuous coded and/or dichotomous coded study characteristics (Lipsey & Wilson, 2001). The weighted regression approach can test multiple continuous or dummy coded study characteristics all in one analysis to see if they explain the variability in effect sizes that is not due to subject-level sampling error, and the analog to ANOVA can test one categorical coded study characteristic at a time to determine whether that study characteristic explains variability in effect sizes (Lipsey & Wilson, 2001). When coded study characteristics are used to explain the variability in effect sizes that is not due to subject-level sampling error, that variability is seen as study-level sampling error and is considered to be systematic variability (Lipsey & Wilson, 2001).

The meta-analyst should not test all possible coded study characteristics, searching for which ones are significant, because that would capitalize on chance, and if the researcher tested enough study characteristics, some would be significant moderators of effect size simply due to chance (Lipsey & Wilson, 2001). Also, as stated earlier, if it is desired to explain between-studies variation with continuous variables rather than categorical variables, meta-regression can be used (Shadish & Haddock, 2009). If neither moderator analyses nor meta-regression is capable of explaining enough of the between-studies variation, a random-effects model can be used for the meta-analysis to take into account the uncertainty with which the mean effect size and confidence intervals are estimated when there is unexplained between-studies variation (Shadish & Haddock, 2009).

It was planned that moderator analyses would be conducted by gender to determine if there was a stronger relationship between depression and the set of dimensions of perfectionism for women than for men, but there were only four studies that had samples that were all women and there was only one study that had a sample of all male participants, so a moderator analysis could not be conducted. However, one of the meta-analyses had four studies with only women participants, so that meta-analysis was run both with those four studies with all women participants and without them to see if it would have an effect on the estimates from this particular meta-analysis.

Publication Bias

Description of publication bias. Publication bias occurs because studies with significant results, larger effect sizes, and/or larger sample sizes are more likely to be published than studies that do not have these characteristics (Sutton, 2009). Publication bias can make the estimate of the summary or mean effect size in a meta-analysis have a larger absolute magnitude than the effect size in the population, causing it to be biased in favor of there being an effect of treatment or a significant correlation when in fact there is no effect or substantial correlation in the population (Borenstein et al., 2009). Publication bias is a serious threat to the validity of the conclusions from a meta-analysis (Sutton, 2009).

Preventing publication bias. The best way to prevent publication bias in a meta-analysis is to perform a very comprehensive search of the literature in order to retrieve all studies relevant to the current meta-analysis (Borenstein et al., 2009; Sutton, 2009). The literature search for these meta-analyses was thorough but not exhaustive. These meta-analyses used any of the methods for detecting publication bias explained below that were appropriate based on the number of studies included and the type of statistical model used, as recommended by Borenstein et al. (2009). The different pieces of evidence for publication bias were synthesized (Borenstein et al., 2009).

A sensitivity analysis can be used to detect the presence of publication bias (Sutton, 2009). It is necessary to look for evidence of publication bias when doing a meta-analysis because if a researcher conducting a meta-analysis does not look for evidence of publication bias, the results of that meta-analysis may falsely indicate that a particular treatment or intervention is effective (Borenstein et al., 2009). When publication bias causes the results of a meta-analysis to be inaccurate, this is a “major threat to the validity of meta-analysis” as a statistical methodology (Sutton, 2009, p. 436). According to Borenstein et al., the different methods for looking at publication bias ask different questions, and the different information given by these methods should be synthesized. The six methods for evaluating evidence of publication bias that are described below are all based on the assumption that there is a relationship between effect size and sample size for each study, and effects should be interpreted in light of that assumption (Borenstein et al., 2009; Card, 2012).

Assessing evidence of publication bias. There are several ways to assess possible publication bias (Borenstein et al., 2009; Card, 2012). Many different methods for assessing publication bias were used because the different methods for assessing publication bias give different types of information about the existence and effect of publication bias (Borenstein et al., 2009; Card, 2012).

Forest plots. A way to start visually inspecting the data from a meta-analysis to determine if there is evidence of publication bias is to construct a forest plot (Borenstein et al., 2009). A forest plot has the studies plotted from most precise at the top to least precise at the bottom, and it shows the effect size, confidence interval, and the relative weight for each study with each study on a separate line (Borenstein et al., 2009). The forest plot can be visually inspected to see if there is a relationship between study size and effect size, and the presence of such a relationship may be seen as evidence of publication bias (Borenstein et al., 2009).

Funnel plots. Another way to visually assess publication bias is to create a funnel plot in order to visually examine if there is a relationship between study effect size and study sample size (Borenstein et al, 2009). Small studies that found big effect sizes are more likely to get published than small studies that found only small or medium effect sizes (Borenstein et al., 2009). If only the small studies that found large effect sizes are used in the meta-analysis, this could make the estimated mean or summary effect size for the meta-analysis biased by making it larger than it really is in the population (Borenstein et al., 2009). The funnel plot should have the shape of an upside-down funnel with larger studies close together at the narrow end of the tunnel at the top and moderate sized studies being more spread out in the middle of the upside-down funnel, and small studies being most spread out at the bottom wide end of the funnel (Borenstein et al., 2009). For graphing a funnel plot, a measure of the precision of the studies included in the meta-analysis goes on the y-axis, and the effect sizes of all the studies in the meta-analysis go on the x-axis (Borenstein et al., 2009). The measure of precision on the y-axis can be the standard error, the variance, or the sample size of each study. If the standard error is used on the y-axis as the measure of precision for each study included in the meta-analysis, this spreads out the small studies at the bottom of the funnel plot so that asymmetry can be more easily spotted. If the funnel plot shows asymmetry among the smaller studies, this means that there is a relationship between the sample size of each study and the corresponding effect size, and this may be seen as evidence of publication bias (Borenstein et al., 2009). However, the smaller studies could truly have larger effect sizes for reasons other than publication bias, so an asymmetrical funnel plot does not give definitive evidence of publication bias (Borenstein et al., 2009). If the direction of the effect is positive, more effect sizes for small studies will be in the lower right side of the funnel plot than on the lower left side of the funnel plot, indicating that mostly only small studies that found large positive effect sizes have been included in the meta-analysis, and thus there is a relationship between study effect size and sample size (Borenstein et al., 2009).

Fail-safe N. Funnel plots only give a subjective indication of publication bias because they are only visually inspected for asymmetry (Borenstein et al., 2009). Another way to examine whether there is evidence of publication bias is to calculate the number of studies having some specified value for effect size, a specified value that is either not statistically significant or not practically significant, that would be needed to make the summary effect size for the meta-analysis either not statistically significant or not practically significant, and this number is called the *Fail-Safe N* (Borenstein et al., 2009). The Fail-Safe *N* developed by Orwin (1983 as cited in Borenstein et al., 2009) is more appropriate than the Fail-Safe *N* developed earlier by Rosenthal (1979 as cited in Borenstein et al., 2009) because Orwin's Fail-Safe *N* allows specification of a value other than zero as the null value for the effect size of the necessary number of studies that would make the summary effect estimated by the meta-analysis become not practically significant, in other words, a finding that there is no substantial relationship between the two variables that are the focus of the meta-analysis (Borenstein et al., 2009). Unlike Rosenthal's Fail-Safe *N*, Orwin's Fail-Safe *N* is not based on *p*-values from significance tests of the effect sizes of the studies used in the meta-analysis (Borenstein et al. 2009). If it had been appropriate for these meta-analyses to use Orwin's Fail Safe *N*, the target value for Orwin's Fail-Safe *N* would have been .09 because Cohen's (1988 as cited in Cohen et al., 2003) guidelines for correlation coefficients see a correlation coefficient of .10 to be a small effect (Cohen et al., 2003). The difference between the funnel plot and the Fail-Safe *N* is that the funnel plot is subjective because its pattern is visually inspected for asymmetry, so it gives no quantitative evidence regarding publication bias, but Fail-Safe *N* does give quantitative evidence of publication bias (Borenstein et al., 2009). Because Card (2012) said that Fail-Safe-*N* becomes problematic if there is extensive heterogeneity in the study effect sizes used to calculate the separate meta-analyses, so as Card recommended, Fail-Safe *N* was not used in the meta-analyses because random-effects models were used.

Egger's linear regression**.** To test for significance of funnel plot asymmetry and possible publication bias, Egger's linear regression approach was used (Card, 2012). However, according to Card's (2012) rough guidelines at least 17 studies are needed to have adequate power for Egger's linear regression to detect severe publication bias and find significant asymmetry in a funnel plot. According to Card's (2012) rough guidelines, only three of the nine meta-analyses had enough statistical power to detect severe funnel plot asymmetry and none of the meta-analyses had enough studies to detect moderate funnel plot asymmetry, but Egger's linear regression approach was used in all nine of the meta-analyses in this study. Also, according to Card's (2012) rough guidelines, none of the nine meta-analyses had enough studies to detect even severe funnel plot asymmetry using Kendall's rank correlation approach, so that approach was not used at all.

Trim and fill method. Another way to examine the extent of possible publication bias and to estimate what the summary or mean effect size for the meta-analysis would be if there were no publication bias is the *Trim and Fill* method developed by Duval and Tweedie (2000a, 2000b as cited in Borenstein et al., 2009). With the trim and fill method the most extreme small studies with the biggest effect sizes are removed from the funnel plot, and the summary or mean effect size for the meta-analysis is iteratively re-estimated until the distribution of study effect sizes in the funnel plot is evenly distributed around this adjusted summary or mean effect size for the meta-analysis (Borenstein et al., 2009). However, this trimming process causes the variance for the meta-analysis to be underestimated, so to correct for this, the small studies with extreme effect sizes are added back in, and a mirror reflection of the effect sizes of these extreme studies are imputed into the opposite side of the funnel plot from where the extreme studies were trimmed (Borenstein et al., 2009). These two steps correct the underestimation of the variance and create a visual display of the distribution of studies that would occur if there were no publication bias (Borenstein et al., 2009). The trim and fill method also gives an estimate of what the summary or mean effect size for the meta-analysis would be if there were no publication bias. Then the original summary or mean effect size for the meta-analysis with possible publication bias and the adjusted estimate of what the summary or mean effect of the meta-analysis would be without publication bias are compared to see if these two estimates are substantially different (Borenstein et al., 2009).

With the trim and fill method, the idea is to determine whether the effect of publication bias on the results of a meta-analysis is, in the words of Borenstein et al. (2009, p. 286), “trivial,” “modest” or “substantial.” If the effect of publication bias on the results of a meta-analysis are trivial, this would indicate that the estimated mean or summary effect for the meta-analysis is not significantly different than it would be if all existing studies were included, and meta-analysis of all existing studies would reach the same conclusions as the actual meta-analysis with an unknown number of excluded studies (Borenstein et al., 2009). If the effects of publication bias on the results of the actual meta-analysis is modest, then if that meta-analysis were conducted using all existing relevant studies, the estimated mean or summary effect for the actual meta-analysis and the hypothetical meta-analysis with all existing studies included would only be slightly different, and they would reach the same conclusion (Borenstein et al., 2009). If the effect of publication bias on the results of the actual meta-analysis is substantial, then the estimated mean or summary effect of the actual meta-analysis would be substantially different than if the meta-analysis had included all existing studies, and the conclusion of the meta-analysis would be different (Borenstein et al., 2009).

Cumulative meta-analysis method. A final method for investigating the evidence for or against the presence of publication bias is conducting a cumulative meta-analysis with studies ordered from largest sample size to smallest sample size (Borenstein et al., 2009). In a cumulative meta-analysis, first the meta-analysis is conducted on the study with the largest sample size to obtain an estimate of the summary or mean effect size for a meta-analysis based on just that one study (Borenstein et al., 2009). In the next step, the study with the second largest sample size is added in, and the meta-analysis is conducted on the two studies with the largest sample sizes, and a summary or mean effect size is calculated (Borenstein et al., 2009). Then the study with the third largest sample size is added and this process of adding the study with the next largest sample size and calculating the mean or summary effect size based on the included studies is repeated until all studies are included, and the results are displayed on a cumulative forest plot (Borenstein et al., 2009). Each line of the cumulative forest plot shows what the summary or mean effect size would be for a meta-analysis based on the study listed on that line and all the studies above it, which have larger sample sizes, if only those studies were included in the meta-analysis (Borenstein et al., 2009). The cumulative forest plot can sometimes also show the cumulative percentage of relative weight given for the total of each study and all the studies above it (Borenstein et al., 2009) but this was not the case with the cumulative forest plots in this study. The forest plot for a cumulative meta-analysis allows one to see what the summary or mean effect size of a meta-analysis based on only the large studies would be without having to decide on a cut-off for what constitutes a large study, and one can also see if inclusion of the smaller, less precise studies shifts or biases the mean or summary effect size for the meta-analysis due to the existence of a relationship between study size and effect size (Borenstein et al., 2009). Thus, one can get an estimate of the unbiased summary or mean effect size for the meta-analysis and see if the conclusion of the meta-analysis would be substantially different with a biased set of retrieved studies versus an unbiased set of studies (Borenstein et al., 2009). In this way, cumulative meta-analyses are a transparent method for assessing the presence of publication bias (Borenstein et al., 2009). The R package Metafor by Viechtbauer (2010) was used to run the cumulative meta-analyses for this study.

However, a cumulative meta-analysis may only be effective in determining whether there is evidence of publication bias and in getting an estimate of what the unbiased summary or mean effect size for the meta-analysis would be if a fixed-effects model were used for the meta-analysis because if there is significant heterogeneity in the distribution of effect sizes and random-effects weights are used, the cumulative meta-analysis might not accurately estimate the unbiased summary or mean effect size (Borenstein et al., 2009). This is because random-effects weights give relatively less weight to larger studies and relatively more weight to smaller studies because random effects weights have the addition of a between-studies variance component (Borenstein et al., 2009). In a fixed-effects meta-analysis smaller studies are given less weight in the estimate of the summary effect size, so if the smaller studies have upwardly biased effect sizes because smaller studies with larger effect sizes are more likely to be published than smaller studies with moderate or small effect sizes, the summary effect is protected somewhat from publication bias because most of the weight is given to the larger studies (Borenstein et al., 2009). Table 3 shows a hypothetical example using variance estimates based on Fisher's z_r transformation formula for a correlation and the respective fixed-effects and random-effects weights for a study with a small sample size and for a study with a large sample size:

Table 3

Example of Inverse Variance Weights under Different Conditions

	Study Size	
	Small n=20 $V_z = .0588$	Large n=100 $V_z = .0103$
Fixed	$W_i = 17.007$	$W_i = 97.087$
Random $T^2 = .012$	$W_i^* = 14.124$	$W_i^* = 44.843$
Random $T^2 = .0819$	$W_i^* = 7.107$	$W_i^* = 10.846$

The estimate T^2 of the between-studies variance τ^2 is a constant value for all studies in a particular meta-analysis (Borenstein et al., 2009).

In general, publication bias is more of a problem when random-effects models are used because random-effects models give more weight to less precise studies that have smaller sample sizes than do fixed-effects models (Card, 2012; Sutton, 2009). Thus, if studies with small sample sizes that did not get significant effects are missing from the meta-analysis, the studies with small sample sizes and larger effect sizes that did get included in the meta-analysis will upwardly bias the mean effect size estimate more in a random-effects model than in a fixed-effects model for a meta-analysis on the same set of studies because smaller studies get relatively more weight in a random-effects model than they do in a fixed-effects model.

It was originally intended that evidence of publication bias would be further examined by conducting separate meta-analyses for published versus unpublished studies or conducting a moderator analysis with published studies coded 1 and unpublished studies coded as zero (Card, 2012; Matt & Cook, 2009). In using this approach, it is important that there is a sufficient number of unpublished studies (Card, 2012). If the meta-analyses on published studies were to give a larger mean effect size estimate than the meta-analyses on unpublished studies, that would be evidence that there is a problem with publication bias (Matt & Cook, 2009). However, not enough unpublished studies were found to conduct separate meta-analyses for unpublished studies, so when one of the meta-analyses had an unpublished study in it, that meta-analysis was run both with and without the unpublished study to determine if the unpublished study affected the results of that meta-analysis.

Procedure

Separate datasets were created for the nine meta-analyses using Excel, with one or two datasets for each of the nine meta-analyses. The datasets for APS-R Discrepancy-depression correlations had one unpublished dissertation by Garrison (2014). The datatsets for HMPS SOP-depression and HMPS SPP-depression relationships both had an unpublished dissertation by Leventhal (2007). These two datasets also had correlations from a study by Blankstein and Lumley (2008), in which the results were reported separately for males and females, and the correlation for males was chosen because that was the only study that gave correlations between any of the relevant perfectionism subscales and depression for only males. The dataset for the FMPS Personal Standards-depression correlations had four studies with only women but it had no studies with only men. It was decided not to use treatment/control group designs if they reported correlations separately for the two groups because it would be hard to decide which group to pick without biasing the results, but if a treatment/control group design reported correlations for the whole sample, that type of study was used. A bibliographic database was constructed using Excel, and it had an entry for each of the 259 search results with the title, the database in which the study record was found, sometimes the abstract if it could be cut and pasted into the Excel file, the title of the study, the date of publication or the date the study was completed, and notes about whether the study was included or excluded in the meta-analyses and if excluded the reason the study was excluded. Correlations were interpreted using Cohen's (1988, as cited in Cohen et al., 2003) guideline for the size of correlations, where a value of $r = .10$ is a small effect size, a value of $r = .30$ is a medium effect size, and a value or $r = .50$ is a large effect size.

Model Selection

A random-effects model was chosen prior to running the meta-analyses because it was desired to generalize the results beyond the specific studies used in each meta-analysis, because the studies were not identical and differed in more than just the research participants, and because they used a total of 11 different measures of depression.

Information about each of these measures of depression can be found in Table 2 (above). Borenstein et al. (2009) said that if there is no true between-studies variability, the fixed-effects and random-effects model give the same results. Borenstein et al. (2009) also said that if there were a default model, it should be the random-effects model rather than the fixed-effects model because it is rarely the case that studies are identical except for the specific participants used and because the random-effects model is more conservative.

Chapter 3: Results

First Meta-Analysis—APS-R HS Subscale and Depression

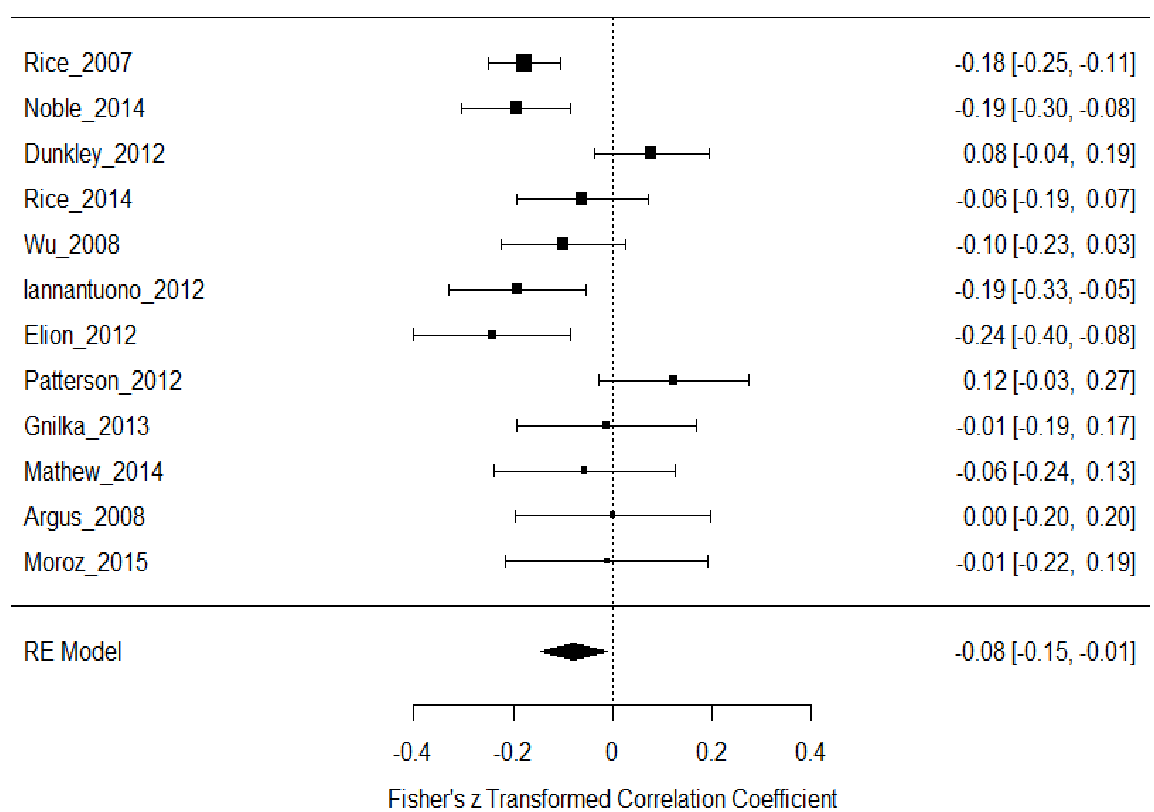
The first of nine meta-analyses estimated the mean correlation between the APS-R High Standards (HS) subscale and depression using a random-effects model and restricted maximum likelihood estimation (REML). It had a total of 12 studies and a total sample size of $N = 3{,}678$. For meta-analyses of the correlations between depression and High Standards and between depression and Discrepancy, the correlations contributed by the Rice et al (2014) study were from developing the Short Almost Perfect Scale (SAP), which is s shortened version of the APS-R. The estimate of the mean effect size for the relationship between HS and depression using a random-effects model was $r = -.08$, 95% CI $[-.14, -.01]$. By Cohen's (1988) guidelines for correlation coefficients, this is smaller than a small effect size. For this meta-analysis, the approximate 95% credibility interval had a lower bound of $r = -.27$ and an upper bound of $r = .12$. The estimate for τ^2, the between-studies variance or total heterogeneity was $T^2 = 0.0088$, and the estimated between-studies standard deviation τ was $T = 0.094$. The result of the test of heterogeneity was $Q(df = 11) = 32.58$, $p < .001$, and $I^2 = 65.67\%$. As was recommended as an option by Borenstein et al. (2009) the critical value for the test of heterogeneity was set at $\alpha = .10$ to give the Q statistic more power to detect heterogeneity, especially since some of the nine meta-analyses had only a small sample of studies. According to Higgins and Thompson's (2002, as cited in Shadish and

Haddock, 2009) guidelines for interpreting the descriptive statistic I^2, the I^2 value of about 66% for this meta-analysis is between a medium and a large amount of heterogeneity.

Figure 1 provides a forest plot of the 12 studies used in the meta-analysis that estimated the mean correlation between the APS-R High Standards subscale and depression with the effect sizes in Fisher's z_r transformed correlation coefficients. Figure 2 provides the same type of plot as Figure 1 except that the effect sizes are raw correlations rather than Fisher's z_r transformed correlation coefficients. Figure 3 provides a funnel plot with the Fisher's z_r transformed correlation coefficients on the x-axis and the standard error on the y-axis, and Figure 4 provides another funnel plot that has Fisher's z_r transformed correlation coefficients on the x-axis and sample size, instead of the standard error, on the y-axis. Figure 5 provides the funnel plot that resulted from doing a trim and fill analysis to obtain an unbiased estimate of the mean correlation between the APS-R High Standards subscale and depression. Figures 6 is a forest plot for a cumulative meta-analysis with all 12 studies that was done with a random-effects model. Figure 7 is a forest plot for a cumulative meta-analysis using the same 12 studies with a fixed-effects model. All figures for this meta-analysis include all 12 studies.

Figure 1

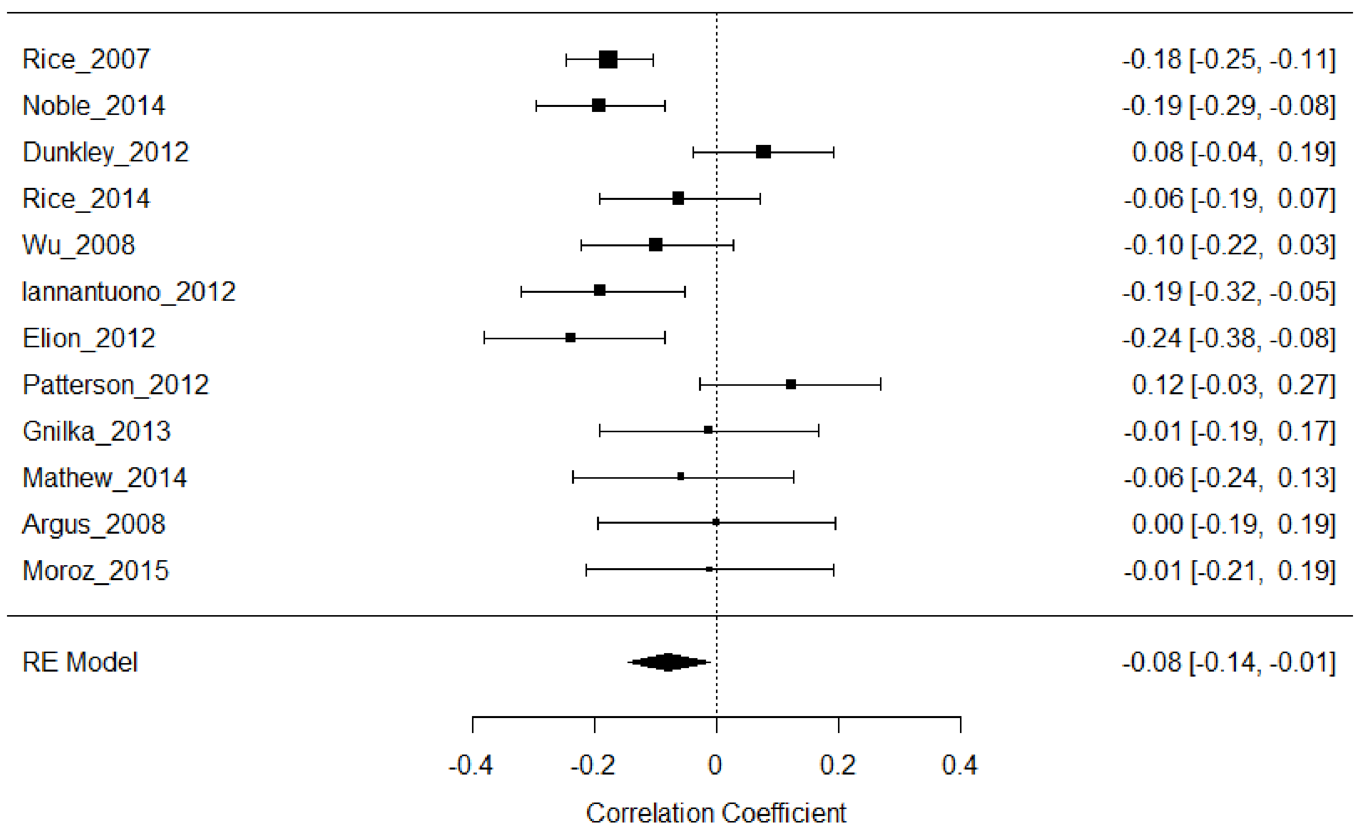
Forest Plot HS_D Fisher's Zr Correlations Corrected for Attenuation Random-Effects Model



The 12 studies in Figure 1 are sorted by sample size with largest sample size at the top and smallest sample size at the bottom as recommended by Borenstein et al. (2009). Ordering the studies from largest sample size on the top to smallest sample size on the bottom allows visual inspection of the relationship between sample size and effect size (Borenstein et al., 2009). The size of the box for each study is proportional to the size of the weight that the meta-analysis gave to each study (and is also proportional to the inverse of the study's variance), with a larger box area indicating greater weight given to a study when combining studies in the meta-analysis and also with a larger box indicating that a study's effect size was estimated with more precision because of having a larger sample size (Borenstein et al., 2009). The values on the right side of the forest plot give (from left to right) the effect size estimate for the study on that line and then also the lower bound and upper bound of the 95% confidence interval in brackets for the study on the same line. Meta-analyses that use random-effects models give a narrower range of weights compared to meta-analyses that use fixed-effects models (Borenstein et al., 2009). The dashed vertical line down the middle of the forest plot represents a correlation of zero, or more generally, an effect size of zero (Card, 2012). The horizontal line extending through the square for each study's effect size represents that study's 95% confidence interval, with shorter lines indicating more precise estimates of that study's effect size (Borenstein et al., 2009; Card, 2012). Borenstein et al. call the bottom row of the forest plot the *summary line*, and the center of the black diamond on the summary line represents the mean effect size in Fisher's z_r transformed correlation coefficients (in Figure 1) for this sample of studies, and the width of the diamond indicates the 95% confidence interval for the mean effect size (Borenstein et al., 2009). The numeric values on the right side of the forest plot's summary line give the exact values for the mean effect size (here in Fisher's z_r transformed correlation coefficients) for this meta-analysis and the upper and lower bound of the 95% confidence interval for that mean effect size (also in Fisher's z_r transformed correlation coefficients). Figure 2 provides similar information but using raw correlations.

Figure 2

Forest Plot HS_D Correlations Corrected for Attenuation Random-Effects Model



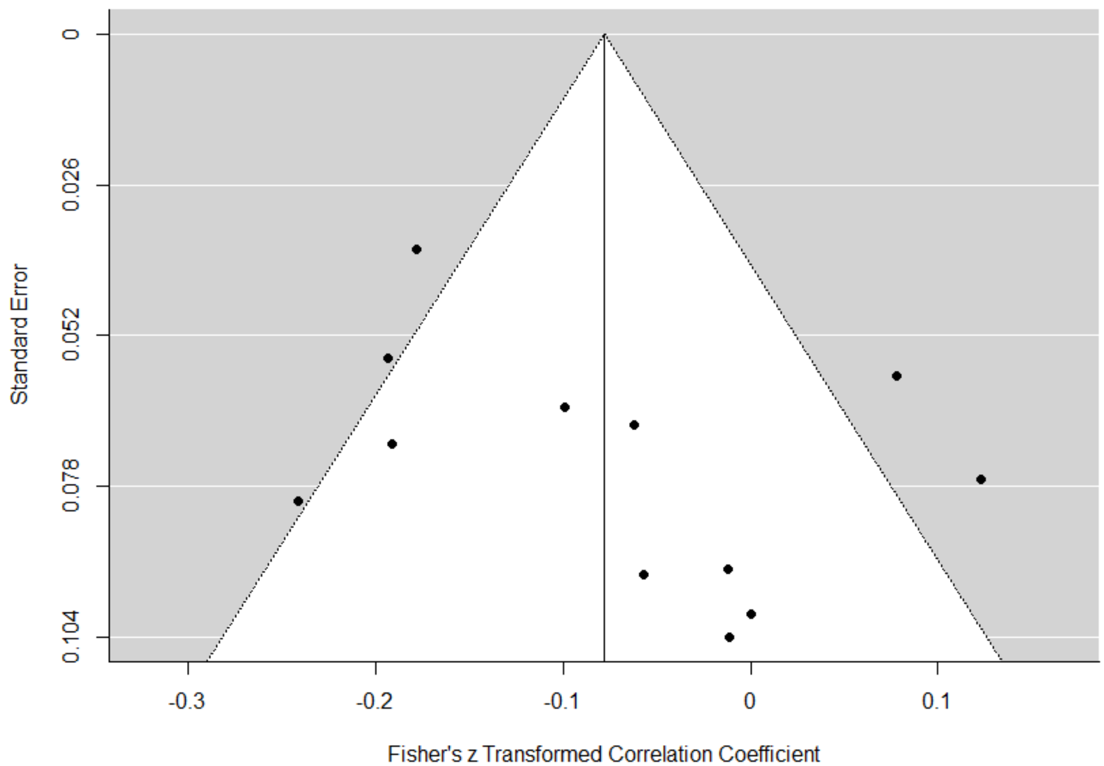
Publication bias. To look at the evidence for publication bias, the effect size can be predicted by the sample size in a regression analysis (Card, 2012). Since the correlation between High Standards and depression was negative, a positive relationship between the corresponding effect size and sample size would indicate possible publication bias for this meta-analysis (Card, 2012)**.** In a linear regression sample size was used to predict the unattenuated correlation effect size. The unattenuated correlation was chosen as the effect size to be predicted because measurement error causes attenuation of correlations (Bobko, 2001) The estimated regression coefficient for sample size in predicting the unattenuated correlation effect size was $b = -.00017$, and it was not statistically significant, $p = .25$, so there was no significant association between sample size and the effect size for the High Standards-depression relationship, and as long as there was adequate power for this significance test, this is evidence against the existence of publication bias (Card, 2012).

According to Sutton (2009) "There is also evidence to suggest that studies with significant outcomes are published more quickly than those with nonsignificant outcomes (Stern & Simes, 1997)" (p. 436) The three multidimensional perfectionism scales that were the topic of this study were published in 1990, 1991 and 2001, so the these measures were around for several years prior to the 2007 to 2017 time frame of the nine meta-analyses in this study, so there has been time for studies with nonsignificant results to be published. Thus, the studies obtained from the time period 2007 to 2017 might contain less publication bias since the three measures of perfectionism have already been used for at least six years prior to the time frame for this study.

Figures 3-7 address potential bias.

Figure 3

Funnel Plot Random-Effects HS_D using Standard Error



The funnel plot in Figure 3 addresses the question of whether bias exists (Borenstein et al., 2009). The direction of the effect is to the left, so a gap on the right lower side would indicate possible publication bias (Borenstein et al., 2009). However, there is a gap in the lower left side of the funnel plot which indicates that small studies with negative correlations of greater absolute value were less likely to be published. According to Sutton (2009), funnel plot asymmetry can be caused by things other than publication bias, so funnel plot asymmetry does not definitively indicate publication bias. According to Borenstein et al., using the standard error, rather than the sample size or variance, on the y-axis of the funnel plot spreads out the studies with smaller sample sizes and thereby makes it easier to look for gaps in the funnel plot where small studies with small or non-significant results should be if all available studies had been retrieved.

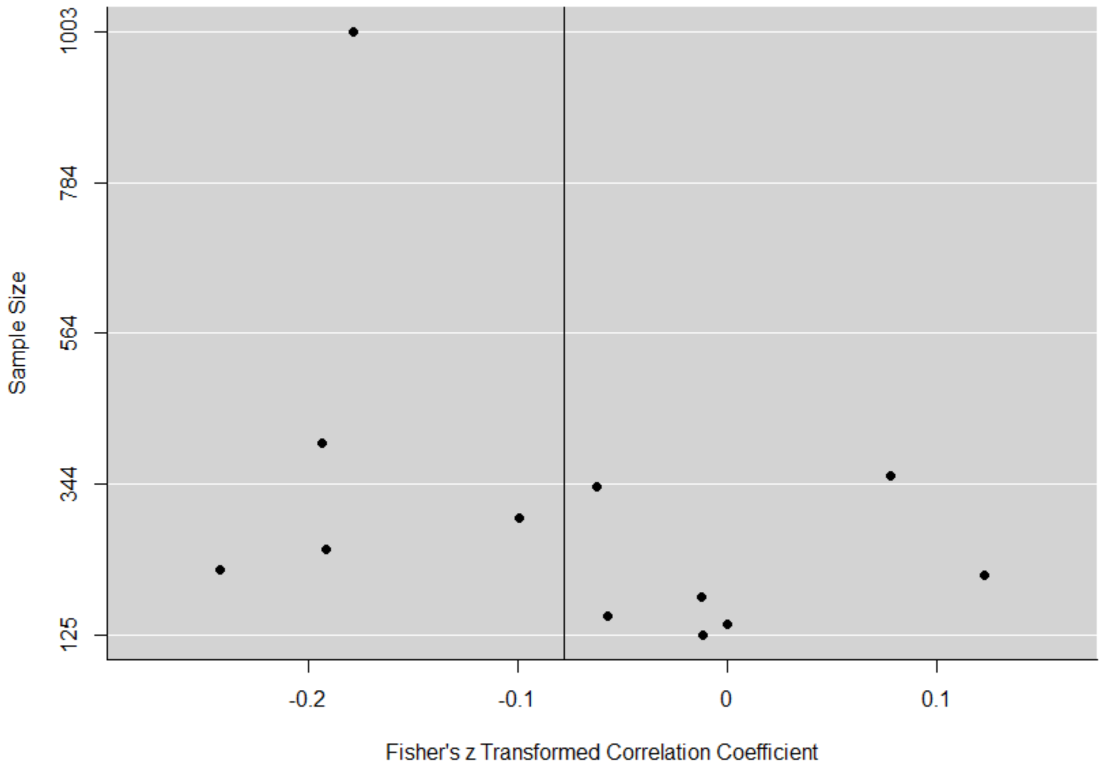
Fail Safe N was computed but is not reported because Card (2012) recommends against using Fail Safe N when there is substantial heterogeneity and a random-effects model is used. This is because the use of Fail Safe N has only been studied in fixed-effect models and not in random-effects models; therefore, there is not much information about the use of Fail Safe N when random-effects meta-analyses are done (Card, 2012).

The correlations from the studies used in the nine meta-analyses in this study were corrected for measurement error. Card (2012) said that when sample size is not perfectly related to effect size, it is useful to use the study weights (the study weights are the inverse of the study variance estimates) when creating funnel plots. For the meta-analyses in this study, funnel plots were created both with the standard error and with sample size.

Card (2012) also said that asymmetry of funnel plots can be examined in a less subjective way than just looking at the funnel plots by "regressing effect sizes onto sample sizes" (p. 266), and if there is a correlation between sample size and effect size, this is evidence that there is publication bias. This gives a statistical test of the asymmetry of funnel plots (Card, 2012). However, these statistical tests of asymmetry frequently have inadequate power, and Card (2012) gave rough guidelines for how many studies are needed for these statistical tests of asymmetry to have adequate power, but Card warned that these guidelines are preliminary and should be used with caution. The number of studies necessary to have adequate power depends on the level of severity of the publication bias (Card, 2012). For detecting severe publication bias with about 80% power, Egger's linear regression method needs at least 17 studies and Kendall's rank correlation method needs at least 40 studies (Card, 2012). For detecting moderate publication bias with about 80% power, Egger's linear regression method needs at least 50 to 60 studies and Kendall's rank correlation method needs at least 150 studies (Card, 2012). In summary, Egger's linear regression method is more powerful than Kendall's rank correlation (Sutton, 2009). According to these guidelines, only three of the meta-analyses in this study (the relationships between depression and each of the subscales HMPS SOP, HMPS SPP, and FMPS PS) had adequate power to detect severe publication bias using Egger's linear regression, and none of the meta-analyses had enough power to detect severe publication bias using Kendall's rank correlation method, and none of the meta-analyses had enough power to detect moderate publication bias using either Egger's linear regression method or Kendall's rank correlation method (Card, 2012).

Figure 4

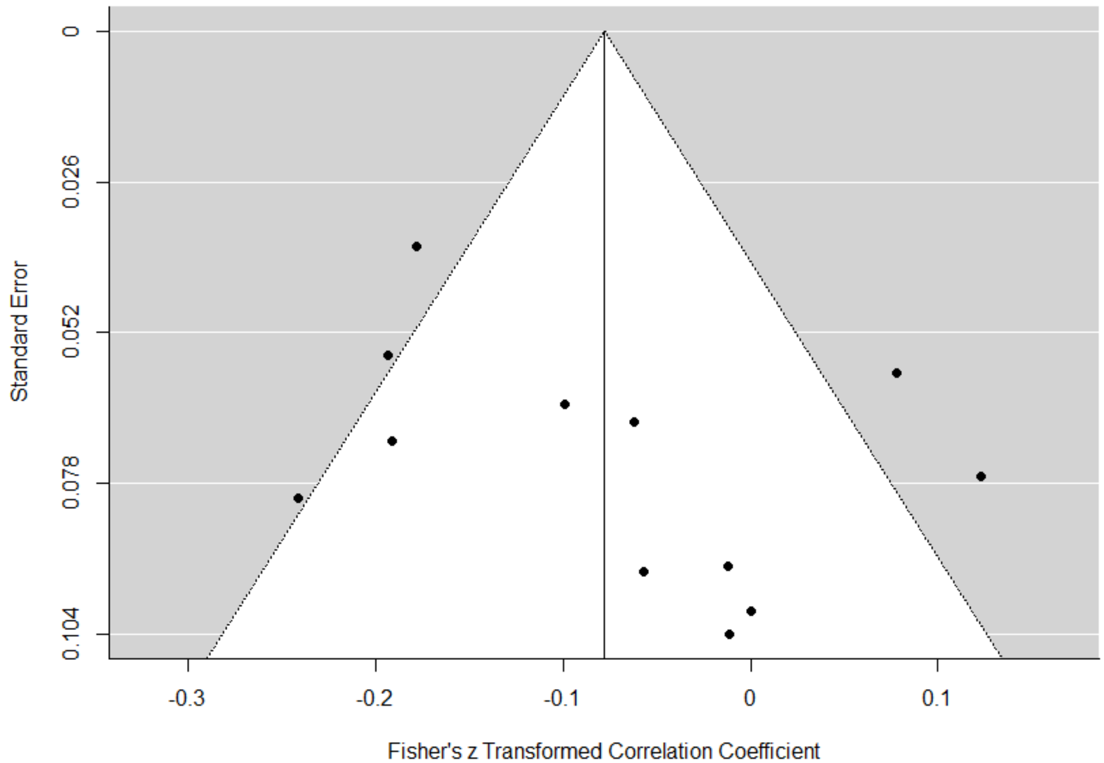
Funnel Plot Random-Effects HS_D using Sample Size



Because of the outlier study with an $N = 1{,}003$ in the funnel plot of Figure 4, it was difficult to determine visually whether there was asymmetry among the studies that have much smaller sample sizes because they are compressed “into a narrow range of the funnel plot” (Card, 2012, p. 265).

Figure 5

Funnel Plot HS_D after Trim & Fill Random-Effects Model



The Trim and Fill method is a two-step process (Card, 2012). First, studies are trimmed from the side of the funnel plot that has too many studies relative to the opposite side so that the funnel plot is symmetrical, and the mean effect size is estimated so that it is not biased by the asymmetry in the original funnel plot (Borenstein et al., 2009; Card, 2012). Second, the studies that were trimmed are added back into the funnel plot and the mirror image of those studies are imputed onto the opposite side of the plot in order to make the variance estimate correct because trimming the studies in the first step artificially reduces the variance, making the confidence intervals too narrow (Borenstein et al., 2009; Card, 2012). This is an iterative process that gives an estimate of the mean effect size and the between-studies variance that is corrected for the effects of publication bias (Borenstein et al., 2009; Card, 2012). Borenstein et al. said that the trim and fill method answers the question, “What is out best estimate of the unbiased effect size?” (p. 286). In Figure 5, if there were publication bias, the lower right part of the funnel plot would be expected to have fewer studies than the lower left part of the plot because the direction of the effect size is negative or to the left. In the R syntax for running this trim and fill analysis and for producing the funnel plot in Figure 5, the right side of the funnel plot was specified as the side where studies would be missing if there were publication bias. The results of this trim and fill analysis showed that the estimated number of missing studies on the right side was zero, and therefore, all the estimates for this meta-analysis were the same as the results above for the meta-analysis without correction for publication bias. However, this trim and fill analysis was run using a random-effects model, and Sutton (2009) recommended using fixed-effects models for the trim and fill analysis because smaller studies that are less precise are given relatively more weight in random-effects models than in fixed-effects models, and this can cause the results from meta-analyses done with random-effects models to be more influenced by publication bias. To determine if this issue with the relatively larger weights given to studies with small sample sizes affected the results of the trim and fill analysis, another trim and fill analysis was run using a fixed-effects model, and the result was still that the number of missing studies on the right side of the funnel plot was zero.

Egger’s regression test for asymmetry was not significant with $z = 1.20$ and $p = .23$, but there were not enough studies in this meta-analysis to even have adequate power to detect severe publication bias using this test (Card, 2012)

Figure 6

Cumulative Meta-Analysis for Random-Effects Model Starting with Largest Study

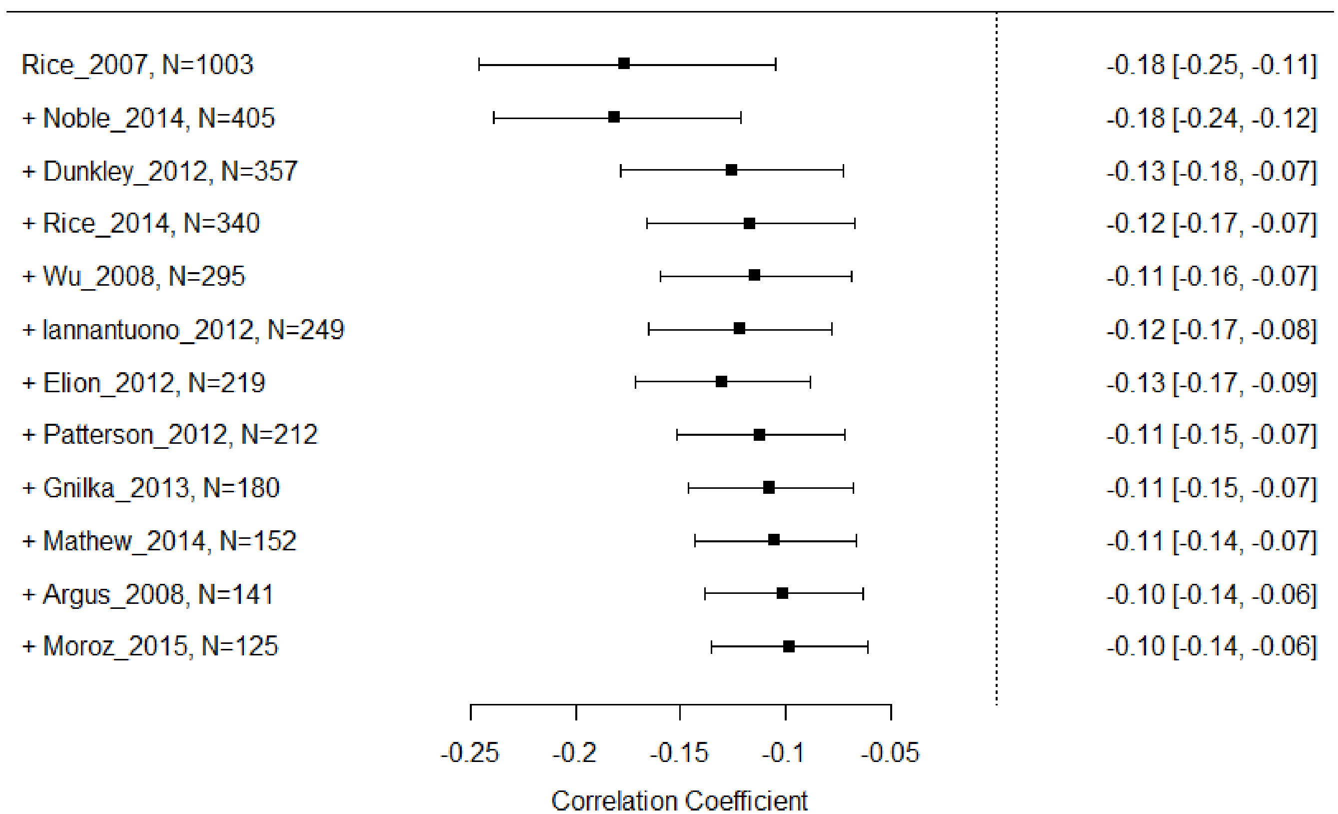
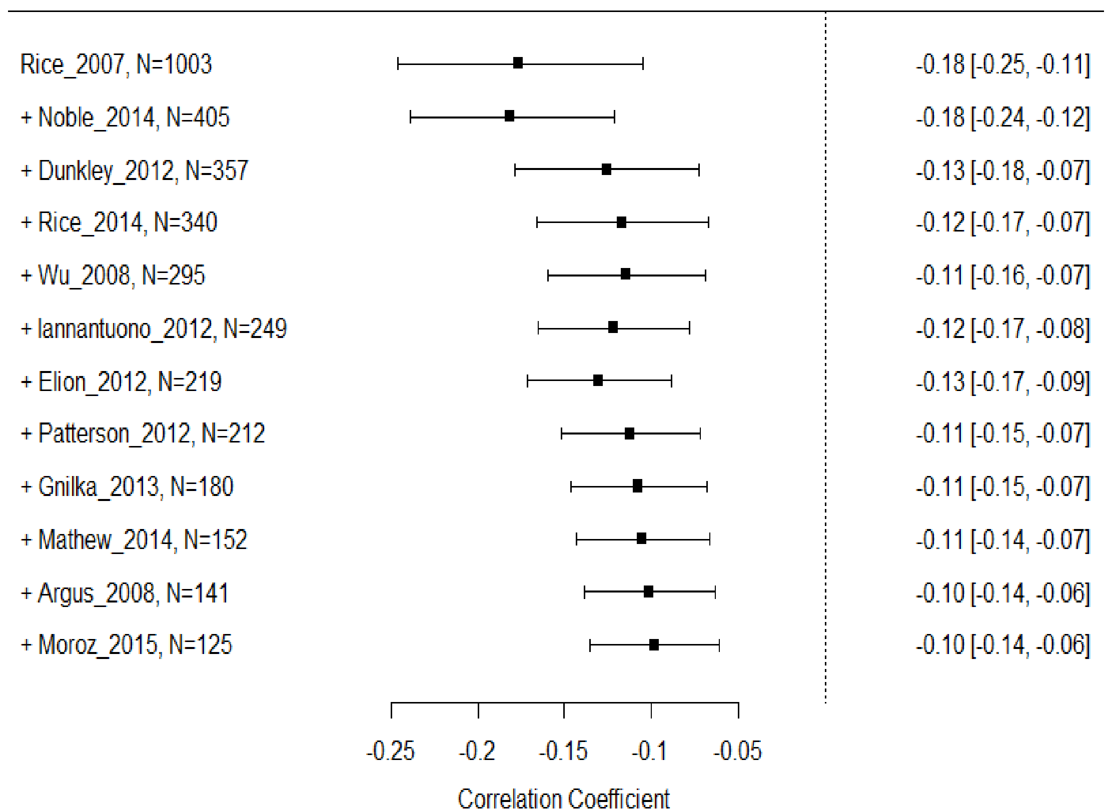


Figure 7

Cumulative Meta-Analysis for Fixed-Effects Model Starting with Largest Study



Figures 6 and 7 above give the forest plots for cumulative meta-analyses with Figure 6 using a random-effects model for the cumulative meta-analysis and Figure 7 using a fixed-effects model. The cumulative meta-analyses are specficied to run the meta-analysis first with only the study with the largest sample size (here Rice et al., 2007 with $N = 1{,}003$) and give an estimate of the effect size based on only the study with the largest sample size (Borenstein et al., 2009). Then the cumulative meta-analysis re-runs the meta-analysis again adding in the study with the second biggest sample size (here Noble et al., 2014 with $N = 405$), and it gives an estimate of the effect size based on the two studies with the two largest sample sizes ($r = -.18$, for the fixed-effect analysis), and then it re-runs the meta-analysis again adding in the study with the third largest sample size (here Dunlkley et al., 2012 with $N = 357$), and gives an estimate of the effect size based on a meta-analysis with only the three studies with the three largest sample sizes ($r = -.13$ in the fixed-effects cumulative meta-analysis; Borenstien et al., 2009). The cumulative meta-analysis keeps re-running the meta-analysis adding the one study with the next smallest sample size until it has included all the studies (Borenstein et al., 2009). When a cumulative meta-analysis shows the estimate of the mean correlation shifting to the right or left after the addition of studies with smaller sample sizes rather than stabilizing, this indicates that there is a relationship between sample size and effect size and this relationship between sample size and effect size might be due to publication bias, but the estimate of the effect size above such a shift in the cumulative meta-analysis gives an unbiased estimate of the effect size (Borenstein et al., 2009). Therefore, a cumulative meta-analysis is a transparent way to obtain an unbiased estimate of the mean correlation, and this unbiased estimate is not thrown off by a few studies with outlier effect sizes the way the unbiased estimate for the effect size from a trim and fill analysis might be (Borenstein et al., 2009). In Figures 6 and 7 above, the direction of the effect is negative and the estimate of the effect size shifts in the opposite direction as studies with increasingly smaller sample sizes are added and the meta-analysis is re-run with the addition of each study (Borenstein et al., 2009). Because random-effects models give relatively more weight to smaller studies and less weight to larger studies compared to fixed-effects models, the shift in the estimate of the effect size as studies with smaller sample sizes are added to the meta-analysis is more apparent in fixed-effects models, so a cumulative meta-analysis based on a fixed-effects model probably gives a better estimate of the unbiased effect size (Borenstein et al., 2009). In Figures 6 and 7 the absolute value of the correlation becomes smaller as studies with increasingly smaller sample sizes are added one by one as the cumulative meta-analysis is re-run. Looking at the estimated effect size just above this shift in effect size in the forest plot for the cumulative meta-analysis gives an unbiased estimate of the effect size for the meta-analysis (Borenstein et al., 2009). In the fixed-effects cumulative meta-analysis in Figure 7, there appears to be a relationship between sample size and effect size, which may be evidence of publication bias.

Second Meta-Analysis—APS-R Discrepancy and Depression

The second of nine meta-analyses estimated the mean correlation between the APS-R Discrepancy subscale and depression using a random-effects model and restricted maximum likelihood (REML). This meta-analysis had a total sample size of $N = 4{,}708$, and included 15 studies, one of which was an unpublished dissertation. The meta-analysis was conducted both with and without the unpublished dissertation. The correlations from the 15 individual studies were corrected for attenuation due to measurement error, and the Fisher's z_r variance was also corrected for the uncertainty that correcting the correlations for attenuation due to measurement error introduces (Card, 2012, p.131). For the analysis with the Garrison (2014) dissertation, the estimate of the mean effect size for the relationship between Discrepancy and depression using a random-effects model was $r = .56$, with a 95% CI [.51, .60], which by Cohen's (1988) guidelines is a large effect size. The estimate for τ^2, the between-studies variance or total heterogeneity was $T^2 = 0.0096$, and the estimated between-studies standard deviation τ was $T = 0.098$. The result of the test of heterogeneity was $Q(df = 14) = 47.33, p < .001$, and $I^2 = 69.67\%$. According to Higgins and Thompson's (2002, as cited in Shadish and Haddock, 2009) guidelines for interpreting the descriptive statistic I^2, the I^2 value of about 70% for this meta-analysis is almost a large amount of heterogeneity. For this meta-analysis, the approximate 95% credibility interval (*CrI*) had a lower bound of $r = .40$ and an upper bound of $r = .68$.

From running the same meta-analysis without the Garrison (2014) dissertation with 14 studies instead of 15 and a total $N = 3{,}963$, the estimate of the mean effect size for the relationship between Discrepancy and depression using a random-effects model was $r = .57$, 95% CI [.52, .61], which by Cohen's (1988) guidelines is a large effect size. These values were only slightly different than the same meta-analysis with all 15 studies. The estimate for τ^2, the between-studies variance or total heterogeneity with the Garrison (2014) study excluded was $T^2 = 0.0077$, and the estimated between-studies standard deviation τ was $T = 0.088$. The result of the test of heterogeneity was

$Q(df = 13) = 31.86, p < .0025$, and $I^2 = 62.13\%$. The I^2 value of about 62% for this meta-analysis is between a medium and a large amount of heterogeneity. The approximate 95% Credibility Interval was CrI $[.43, .68]$.

Figure 8 provides a forest plot of the 15 studies used in the meta-analysis that estimated the mean correlation between the APS-R Discrepancy subscale and depression with the effect sizes in Fisher's z_r transformed correlation coefficients. Figure 9 provides the same type of plot as Figure 8 except that the effect sizes are raw correlations rather than the Fisher's z_r transformed correlation coefficients in Figure 9. Figure 10 provides a funnel plot with the Fisher's z_r transformed correlation coefficients on the x-axis and the standard error on the y-axis. Figure 11 provides another funnel plot that has Fisher's z_r transformed correlation coefficients on the x-axis and sample size, instead of the standard error, on the y-axis. All figures for this meta-analysis include all 15 studies. Forest plots were created for this meta-analysis without the Garrison (2014) dissertation, but the values for the correlations and confidence intervals for all the other studies were the same—just the values for the summary effect and its confidence interval were slightly different. Figure 12 provides the funnel plot that resulted from doing a trim and fill analysis to obtain an unbiased estimate of the mean correlation between the APS-R Discrepancy subscale and depression. Figures 13 is a forest plot for a cumulative meta-analysis with all 15 studies that was done with a random-effects model. Figure 14 is a forest plot for a cumulative meta-analysis using the same 15 studies with a fixed-effects model.

In summary, the mean effect size with the Garrison (2014) dissertation excluded from the analysis was .01 less than without it in the analysis, and the confidence intervals for the two analyses are slightly different. Excluding the Garrison (2014) dissertation increased the lower bound of the confidence interval from .51 to .52 and increased the upper bound of the confidence interval from .60 to .61.

Figure 8

Forest Plot Dis_D Fisher's Zr Correlations Corrected for Attenuation Random-Effects Model

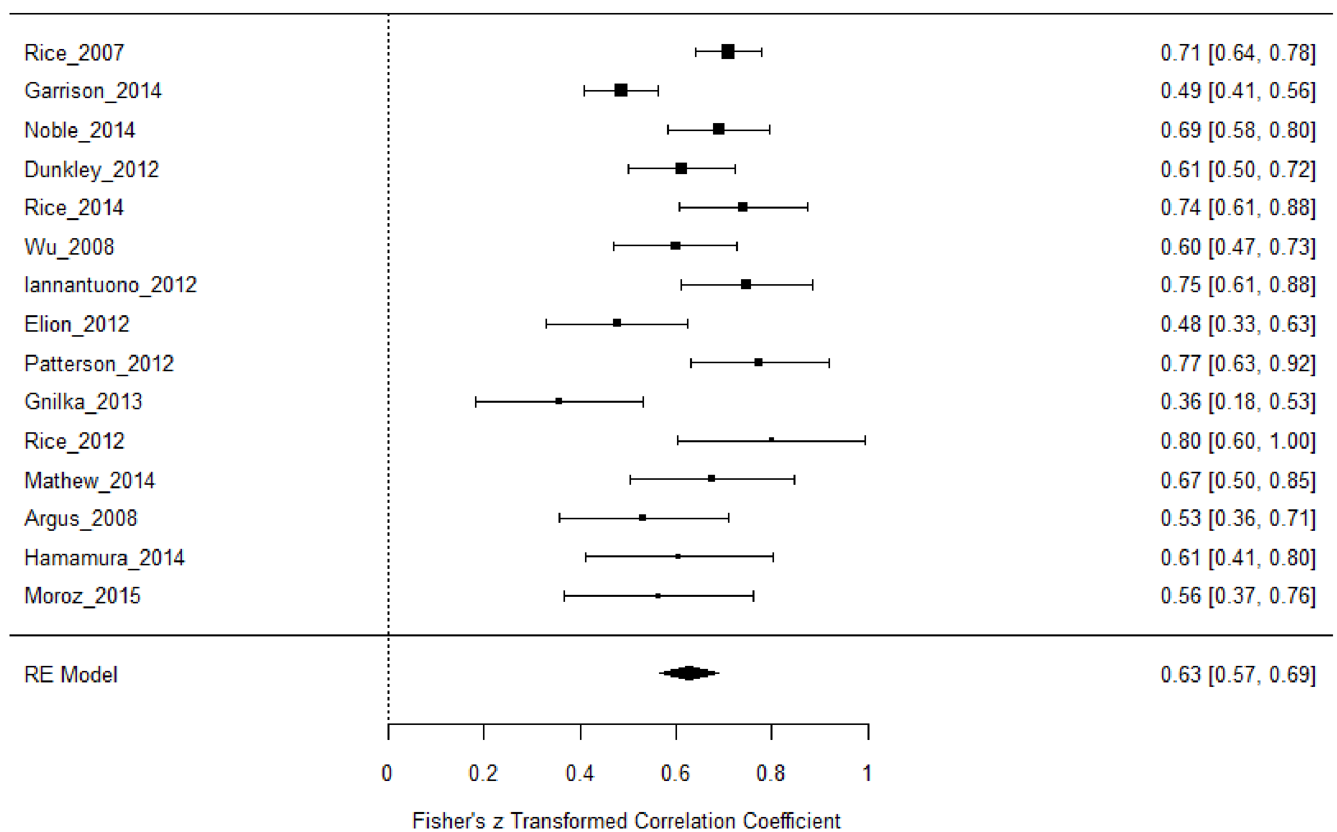
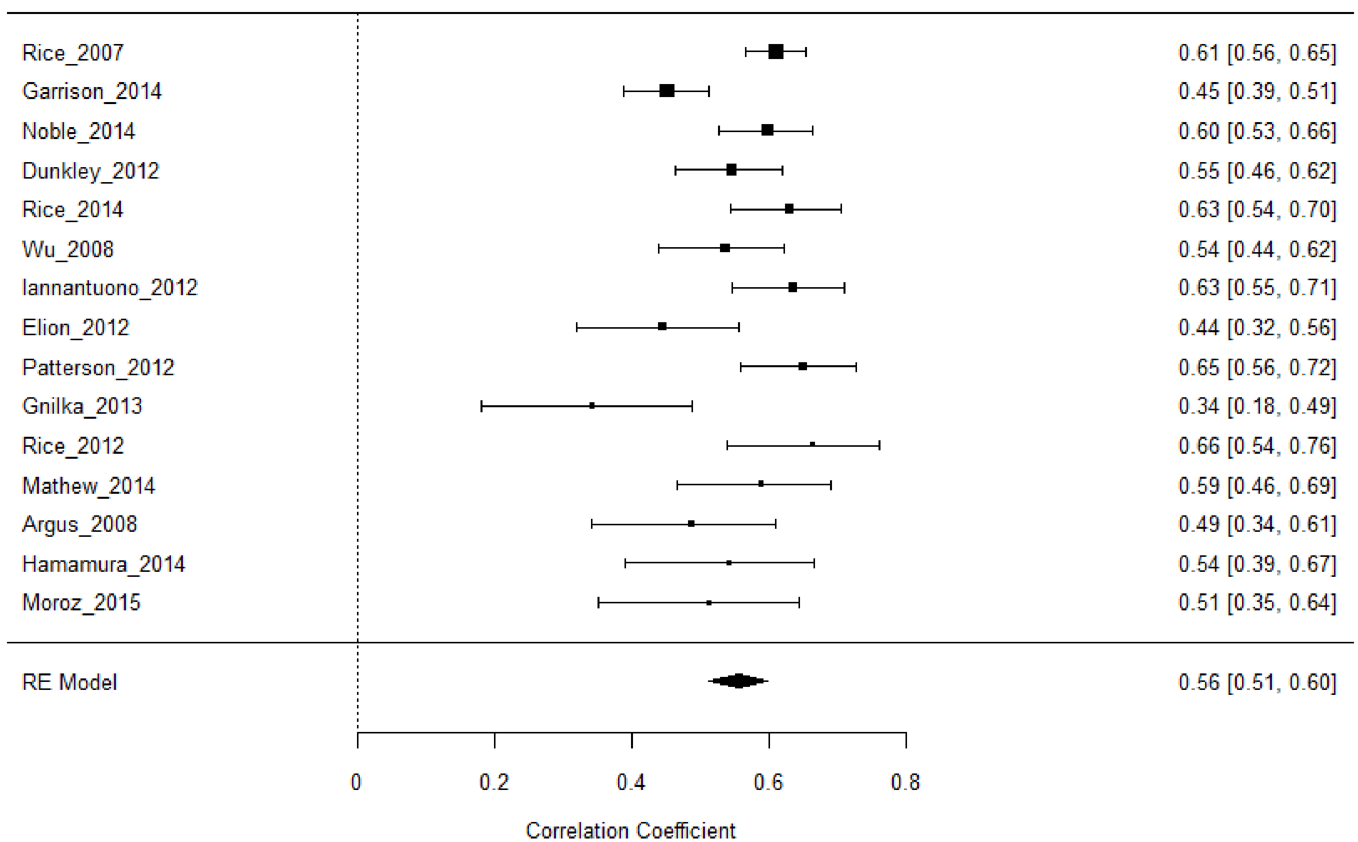


Figure 9

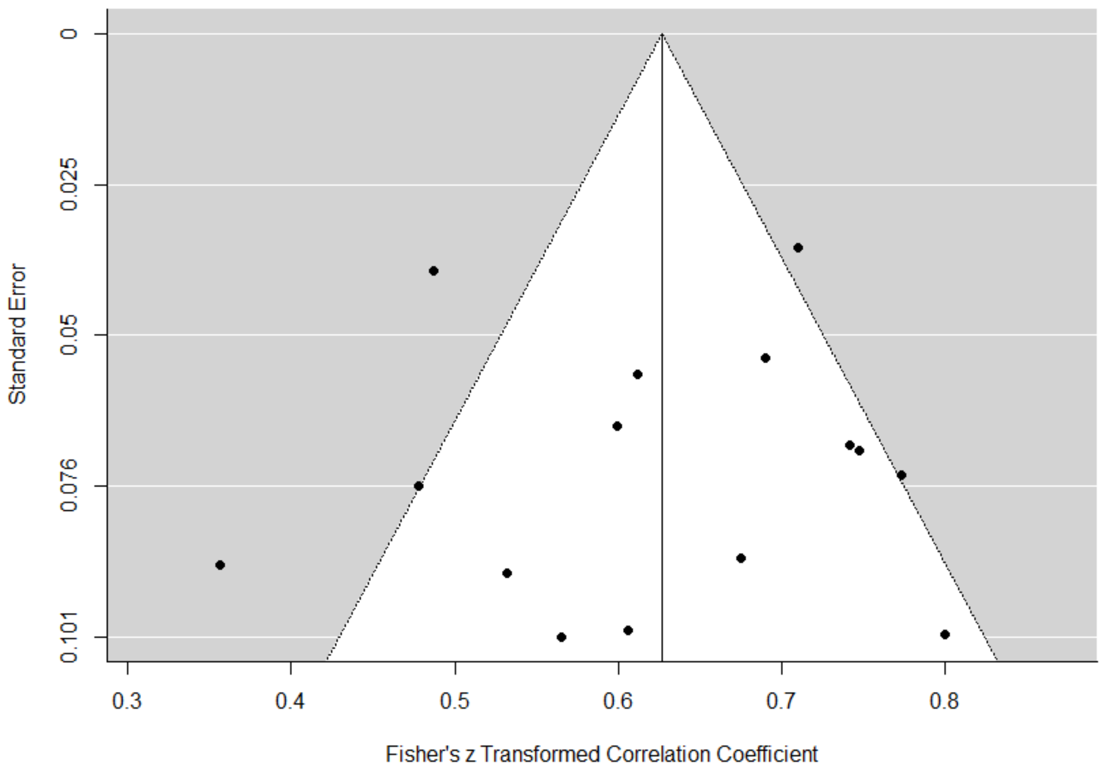
Forest Plot Dis_D Correlations Corrected for Attenuation Random-Effects Model



Figures 8 and 9 (above) include the dissertation by Garrison (2014).

Figure 10

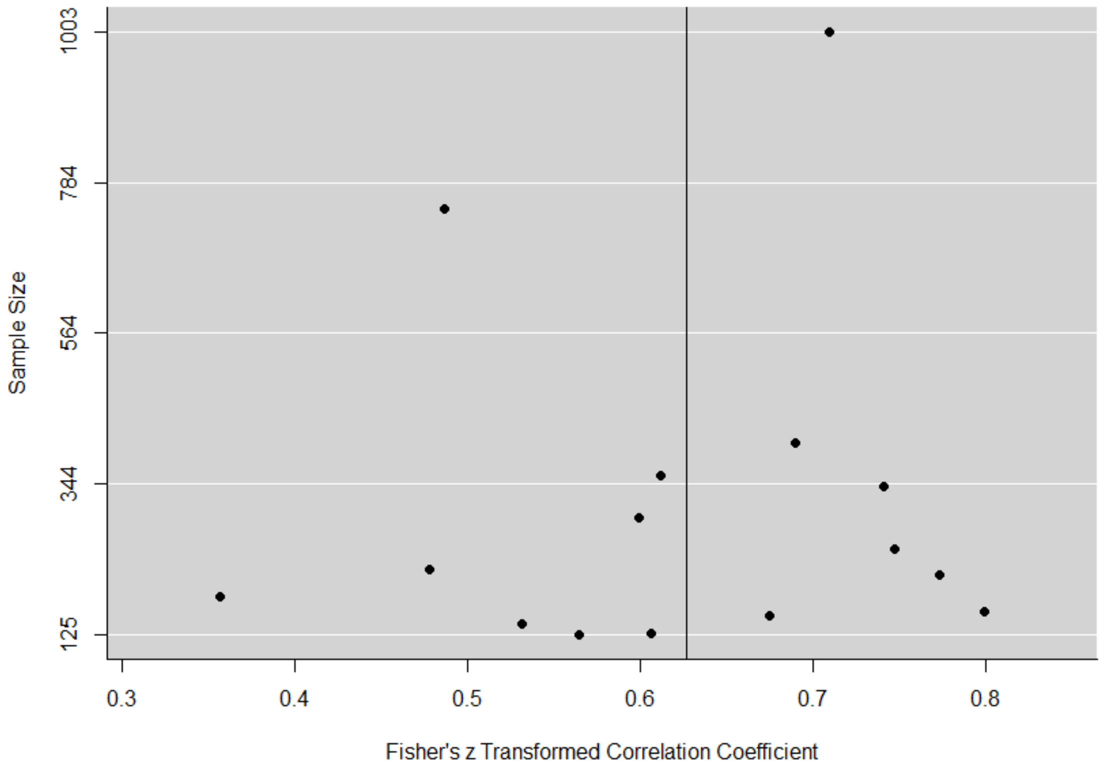
Funnel Plot Random-Effects Dis_D using Standard Error



The funnel plot (above) has all 15 studies.

Figure 11

Funnel Plot Random-Effects Dis_D using Sample Size



The funnel plot (above) has all 15 studies.

Figure 12

Funnel Plot Dis_D after Trim & Fill Random-Effects Model

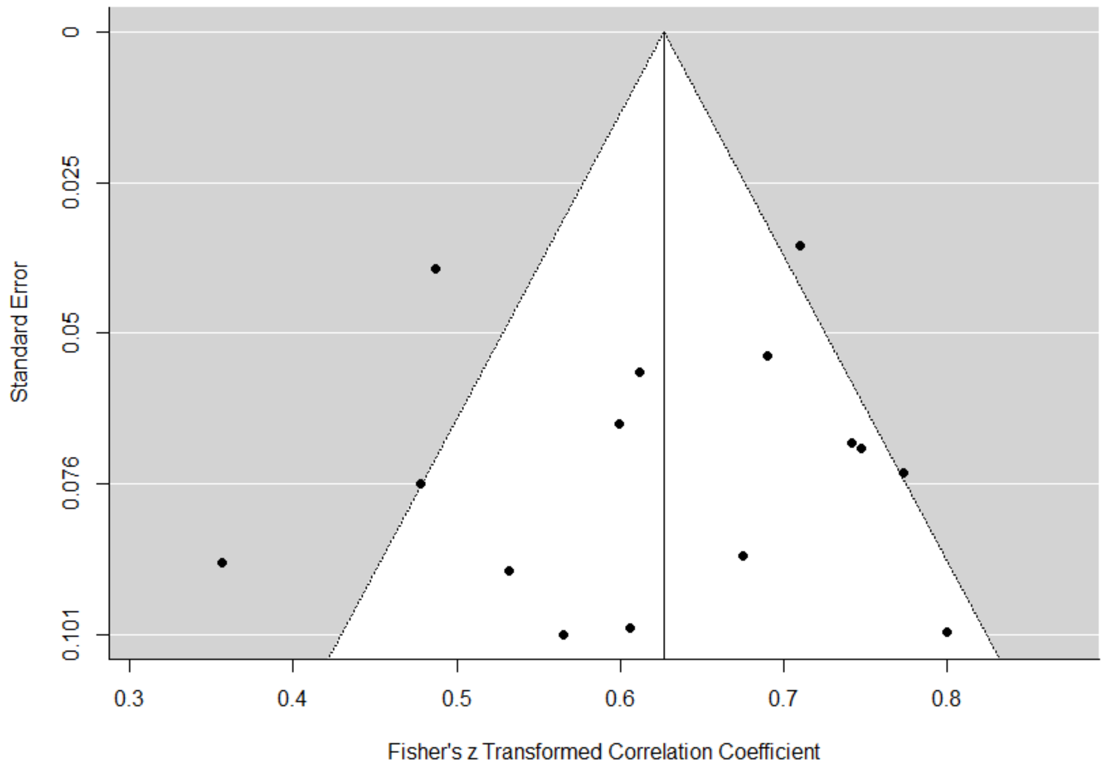


Figure 12 provides a funnel plot of the results of the trim and fill method for estimating what the effect size would be if it were corrected for publication bias. Figure 12 contains all 15 studies (including the dissertation by Garrison, 2014). The results of the trim and fill method showed that there were zero studies missing from the left side, and all the estimates from the trim and fill method were the same as those from the analysis that used all 15 studies (including the dissertation by Garrison, 2014), indicating little or no publication bias.

Figure 13

Cumulative Meta-Analysis for Random-Effects Model Starting with Largest Study

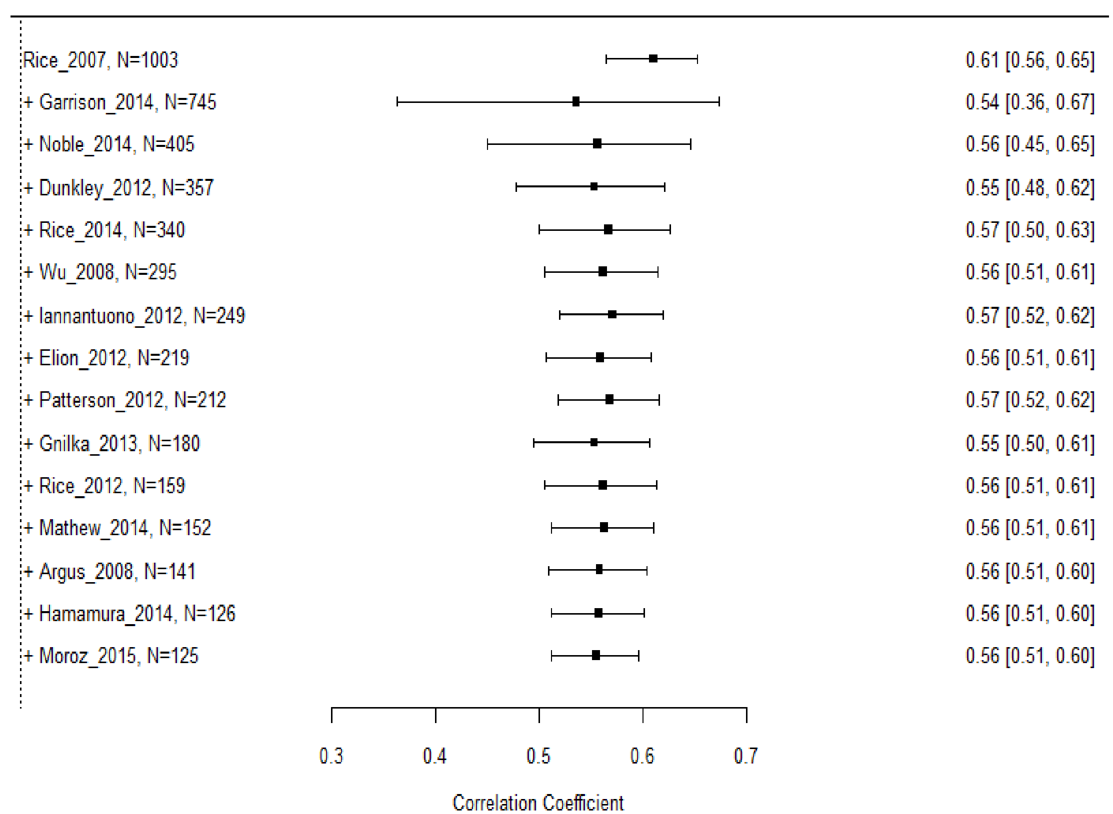
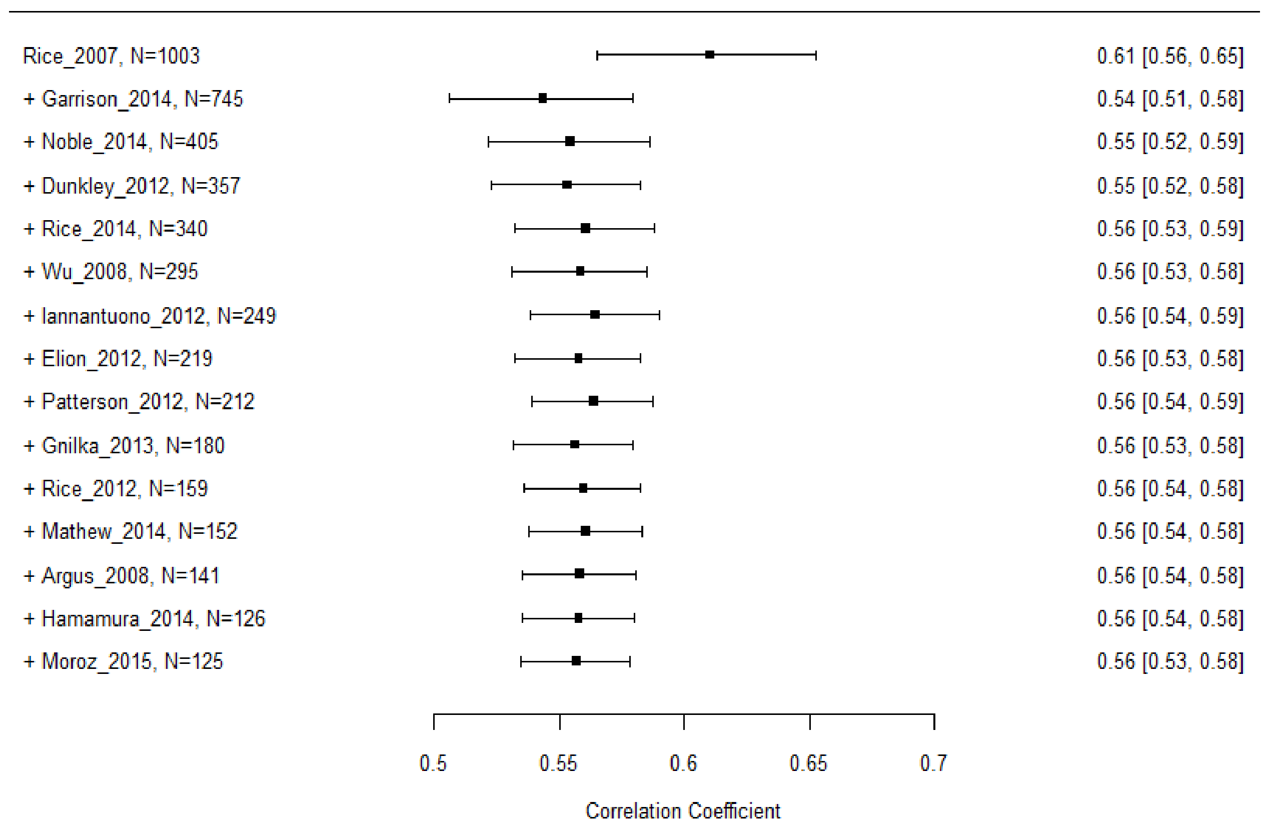


Figure 14

Cumulative Meta-Analysis for Fixed-Effects Model Starting with Largest Study



In the above two cumulative meta-analyses in Figures 13 and 14, the direction of the effect is positive or to the right, but the estimate of the effect size does not shift much to either direction, and the absolute value of the correlation does not change much as studies with smaller sample sizes are added one by one as the analysis is re-run, and the estimate of the mean correlation is stable. This is evidence against the existence of publication bias in this meta-analysis (Borenstein et al., 2009).

Egger's regression test for funnel plot asymmetry was not significant with $z = -.26$ and $p = .79$, but this meta-analysis did not have adequate power for this test to detect even severe asymmetry (Card, 2012).

Third Meta-Analysis—HMPS SOP Subscale and Depression

The third of nine meta-analyses estimated the mean correlation between the HMPS Self-Oriented Perfectionism (SOP) subscale and depression using a random-effects model and REML. This meta-analysis included 25 studies including an unpublished dissertation by Leventhal (2007) with a total sample size of $N = 5{,}581$ with all studies included. The meta-analysis was conducted both with and without the unpublished dissertation. The correlations from the 25 individual studies were corrected for attenuation due to measurement error, and the Fisher's z_r variance was also corrected for the uncertainty that correcting the correlations for measurement error introduces (Card, 2012). For the analysis with the Leventhal (2007) dissertation, the estimate of the mean effect size for the relationship between SOP and depression using a random-effects model was $r = .17$, with a $95\%\ CI$ [.11,.22], a small effect size. The estimate for τ^2, the between-studies variance or total heterogeneity was $T^2 = 0.012$, and the estimated between-studies standard deviation τ was $T = 0.11$. The result of the test of heterogeneity was $Q(df = 24) = 68.04$, $p < .0001$, and $I^2 = 66.38\%$. The I^2 value of about 66% for this meta-analysis is between a medium and large amount of heterogeneity. For this meta-analysis, the approximate 95% credibility interval (*CrI*) was $[-.05, .37]$.

From running the same meta-analysis without the Leventhal (2007) dissertation with 24 studies instead of 25 and a total $N = 5{,}436$, the estimate of the mean effect size for the relationship between SOP and depression using a random-effects model was $r = .17$, 95% CI $[.11, .22]$, between a small and medium effect size. The estimate for τ^2, the between-studies variance or total heterogeneity with the Leventhal (2007) study excluded was $T^2 = 0.013$, and the estimated between-studies standard deviation τ was $T = 0.11$. The result of the test of heterogeneity was $Q(df = 23) = 67.87$, $p < .001$, and $I^2 = 67.87\%$. The I^2 value of about 68% for this meta-analysis is between a medium and a large amount of heterogeneity. The approximate 95% Credibility Interval CrI $[-.06, .38]$. In the meta-analysis with the dissertation by Leventhal excluded, the values for Q, τ^2, τ, I^2, and the 95% CrI were slightly different than the same meta-analysis with all 25 studies, but the values for the mean correlation and 95% confidence interval were the same.

Figure 15 provides a forest plot of the 25 studies used in the meta-analysis that estimated the mean correlation between the HMPS SOP subscale and depression with the effect sizes in Fisher's z_r transformed correlation coefficients. Figure 16 provides the same type of plot as Figure 15 except that the effect sizes are raw correlations rather than Fisher's z_r transformed correlation coefficients. Figure 17 provides a funnel plot with the

Fisher's z_r transformed correlation coefficients on the x-axis and the standard error on the y-axis, and Figure 18 provides another funnel plot that has Fisher's z_r transformed correlation coefficients on the x-axis and sample size, instead of the standard error, on the y-axis. Figure 19 provides the funnel plot that resulted from doing a trim and fill analysis to obtain an unbiased estimate of the mean correlation between the HMPS SOP subscale and depression. Figures 20 is a forest plot for a cumulative meta-analysis with all 25 studies that was done with a random-effects model. Figure 21 is a forest plot for a cumulative meta-analysis using the same 25 studies with a fixed-effects model. All these figures for this meta-analysis include all 25 studies.

Figure 15

Forest Plot SOP_D Fisher's Zr Correlations Corrected for Attenuation Random-Effects

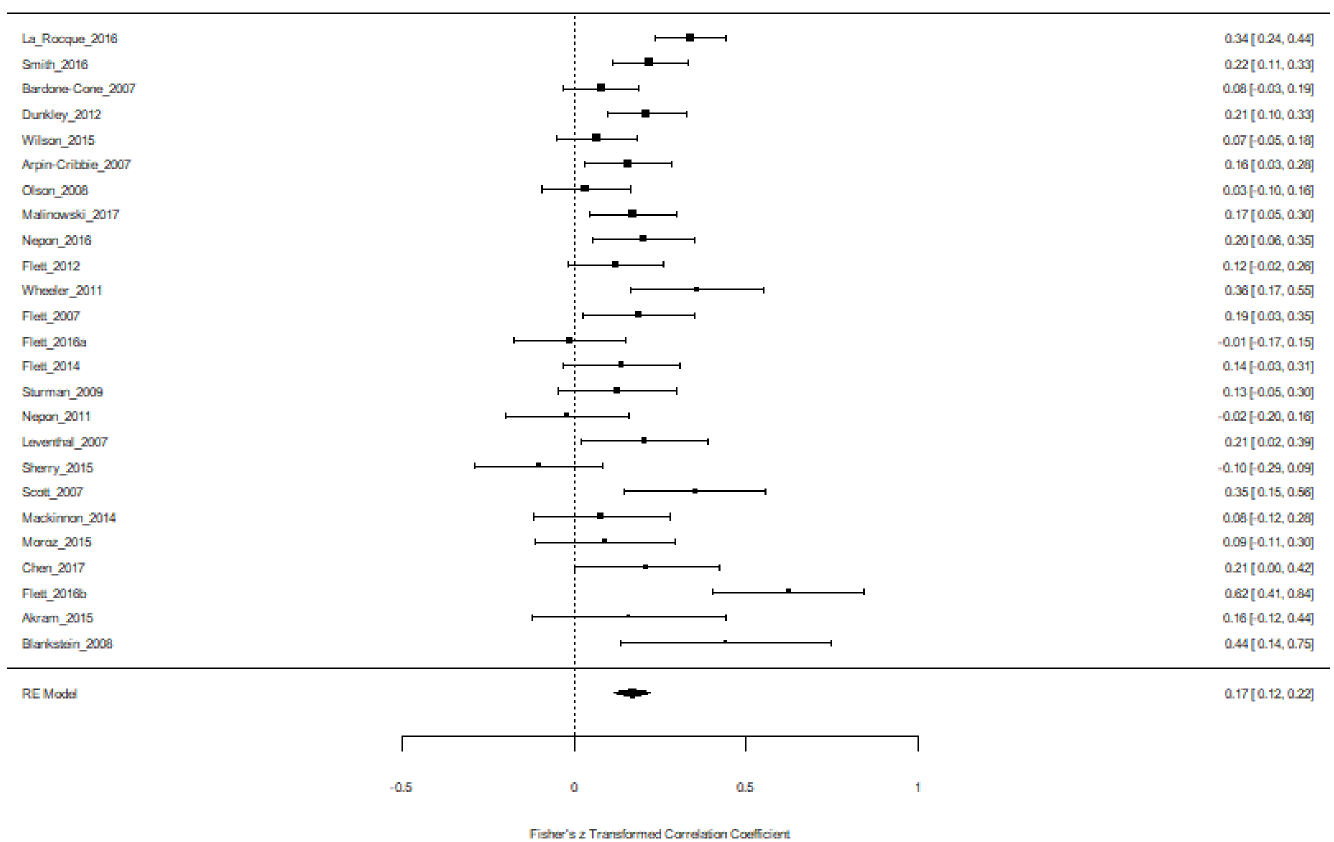
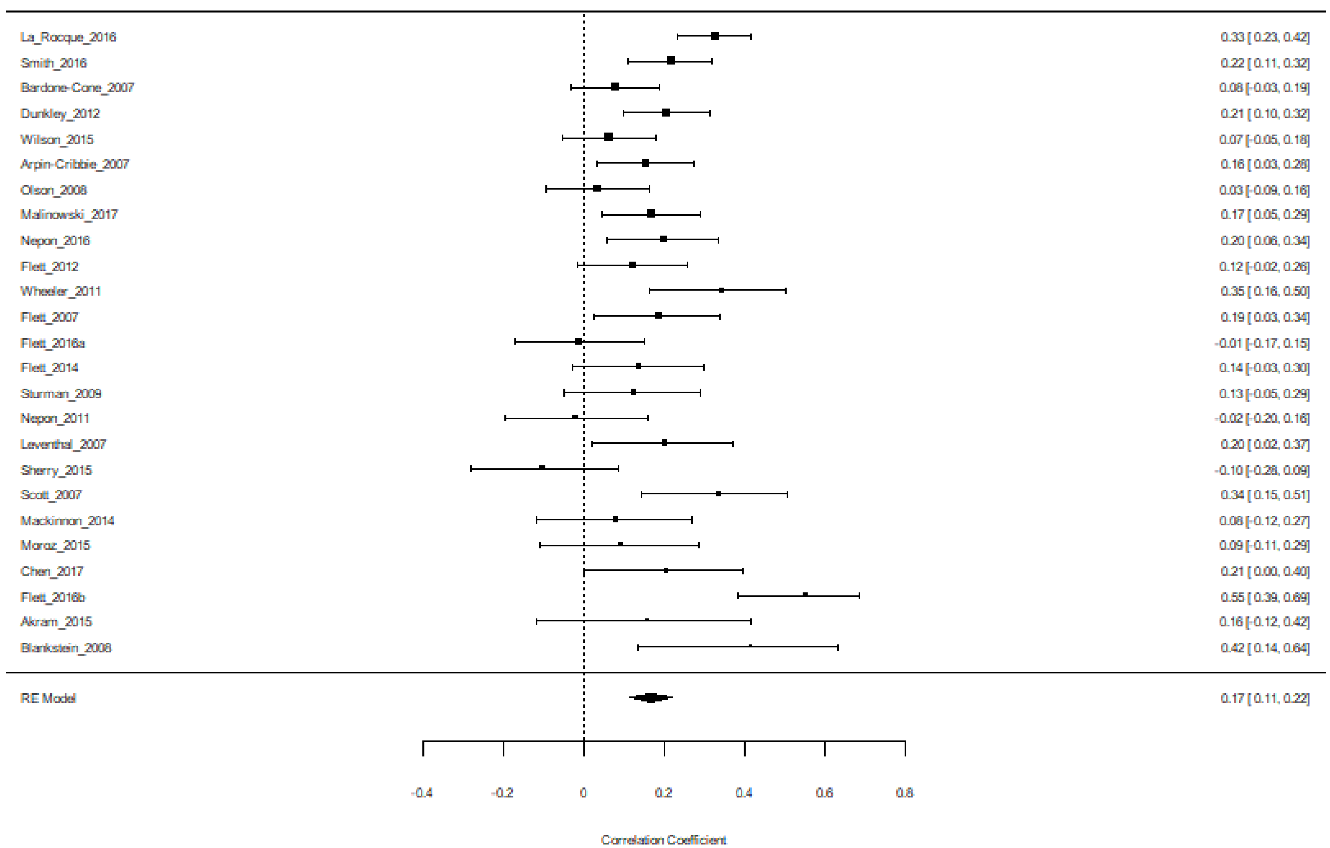


Figure 16

Forest Plot SOP_D Correlations Corrected for Attenuation Random-Effects Model



In Figures 15 and 16 (above) the studies are sorted by sample size with larger sample sizes on the top and smaller sample sizes on the bottom so that the forest plot can be inspected visually to determine if the effect size shifts as sample size decreases. The Flett et al. (2016b) study had an unusually high correlation between SOP and depression (not a coding error). Blankstein and Lumley (2008) also had a higher than usual value for the correlation between SOP and depression, but in this study, the results were reported separately for males and females, and the correlations for males were chosen because Blankstein and Lumley (2008) was the only study that provided any correlations between the relevant subscales and depression.

Figure 17

Funnel Plot Random-Effects SOP_D using Standard Error

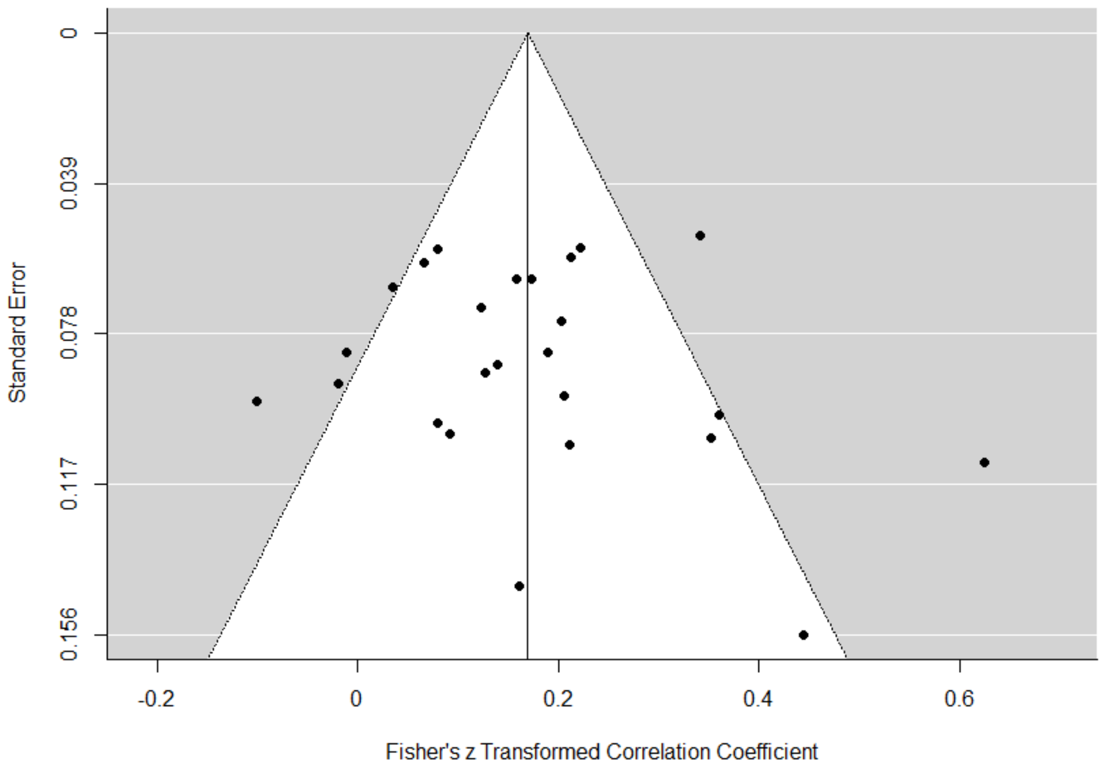


Figure 18

Funnel Plot Random-Effects SOP_D using Sample Size

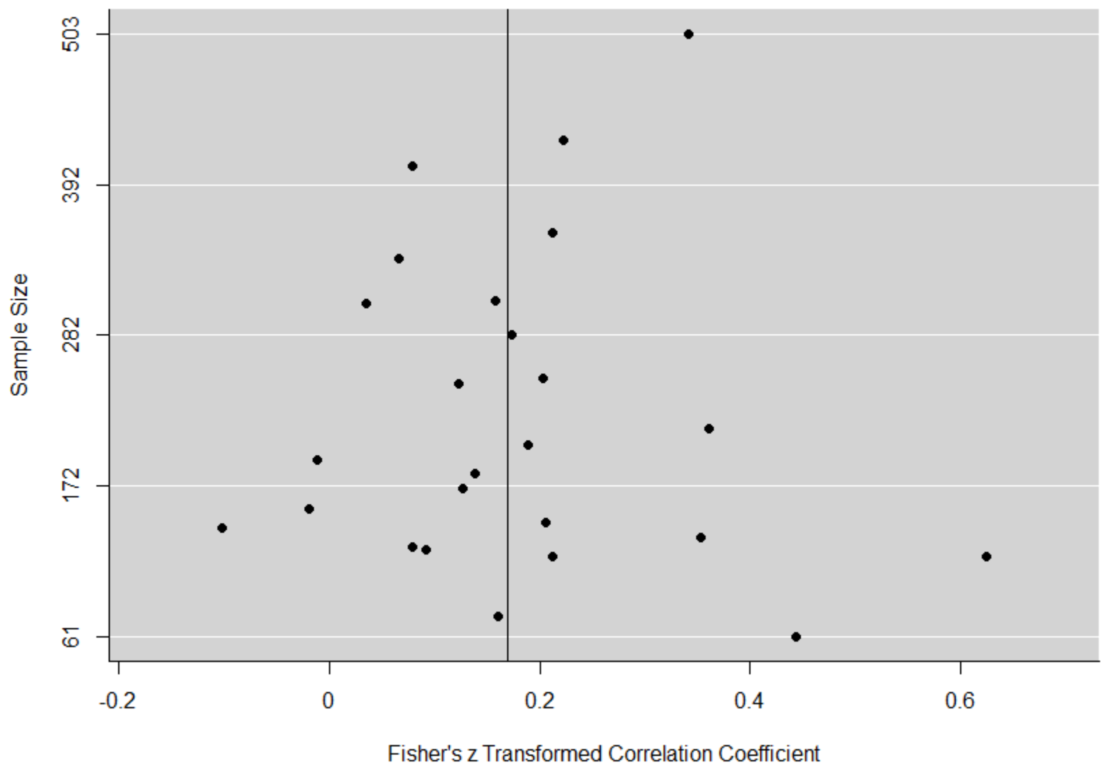


Figure 19

Funnel Plot SOP_D after Trim & Fill Random-Effects Model

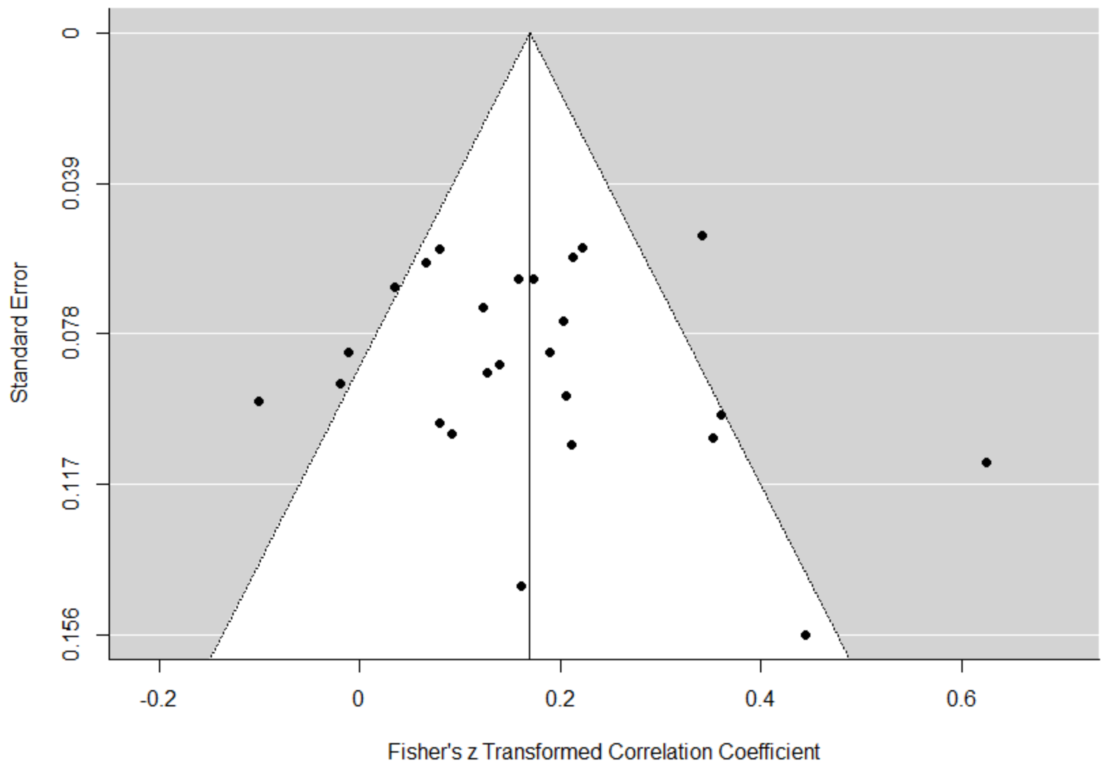


Figure 19 above is based on all 25 studies. If there were evidence of publication bias or a relationship between effect size and sample size, there would be missing studies on the lower left side of the funnel plot where studies should be that had small sample sizes and that found small effect sizes (Borenstein et al., 2009; Card, 2012). The results of the trim and fill method showed that there were zero studies missing from the left side, and all the estimates from the trim and fill method were the same as those from the analysis that used all 25 studies (including the dissertation by Leventhal, 2007), indicating no publication bias.

Figure 20

Cumulative Meta-Analysis for Random-Effects Model Starting with Largest Study

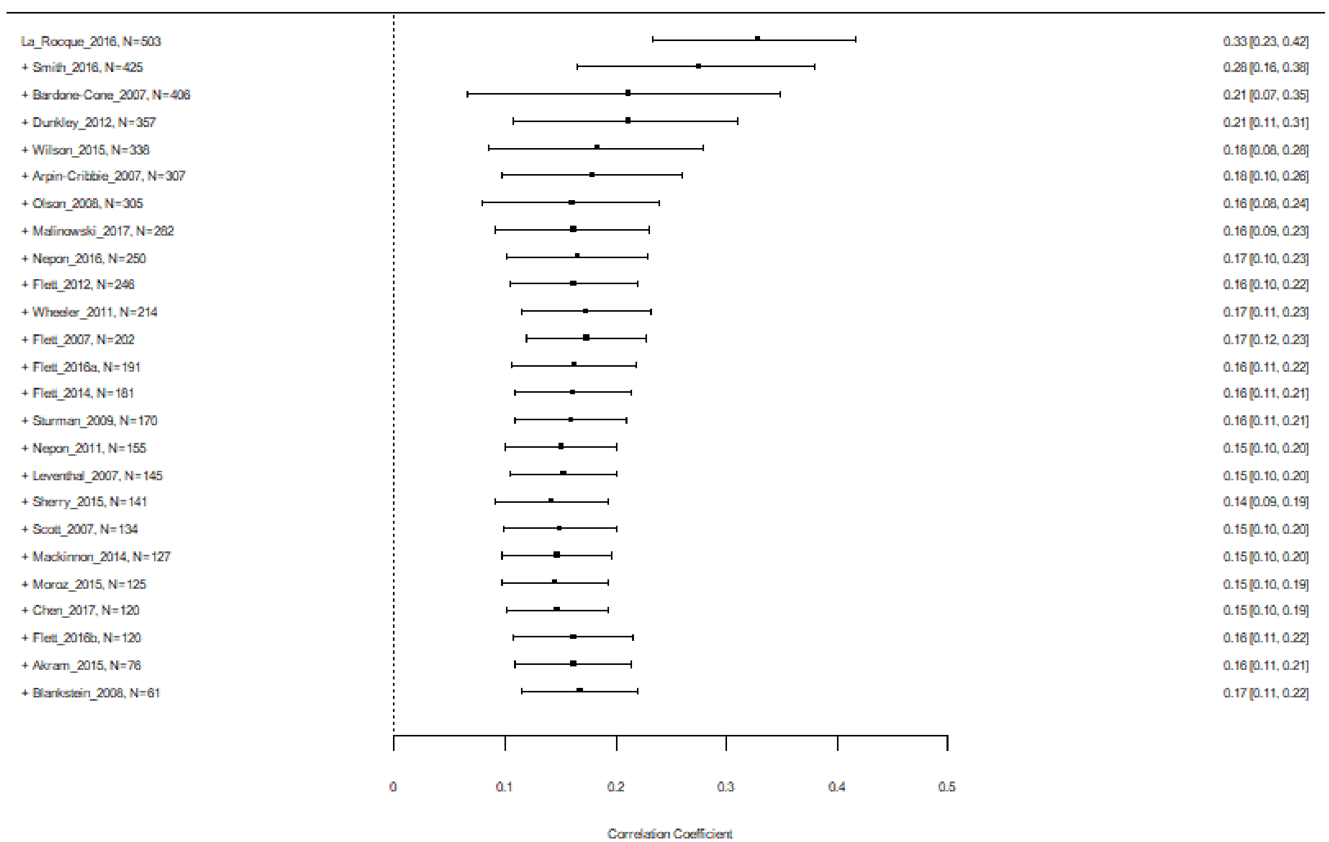


Figure 21

Cumulative Meta-Analysis for Fixed-Effects Model Starting with Largest Study

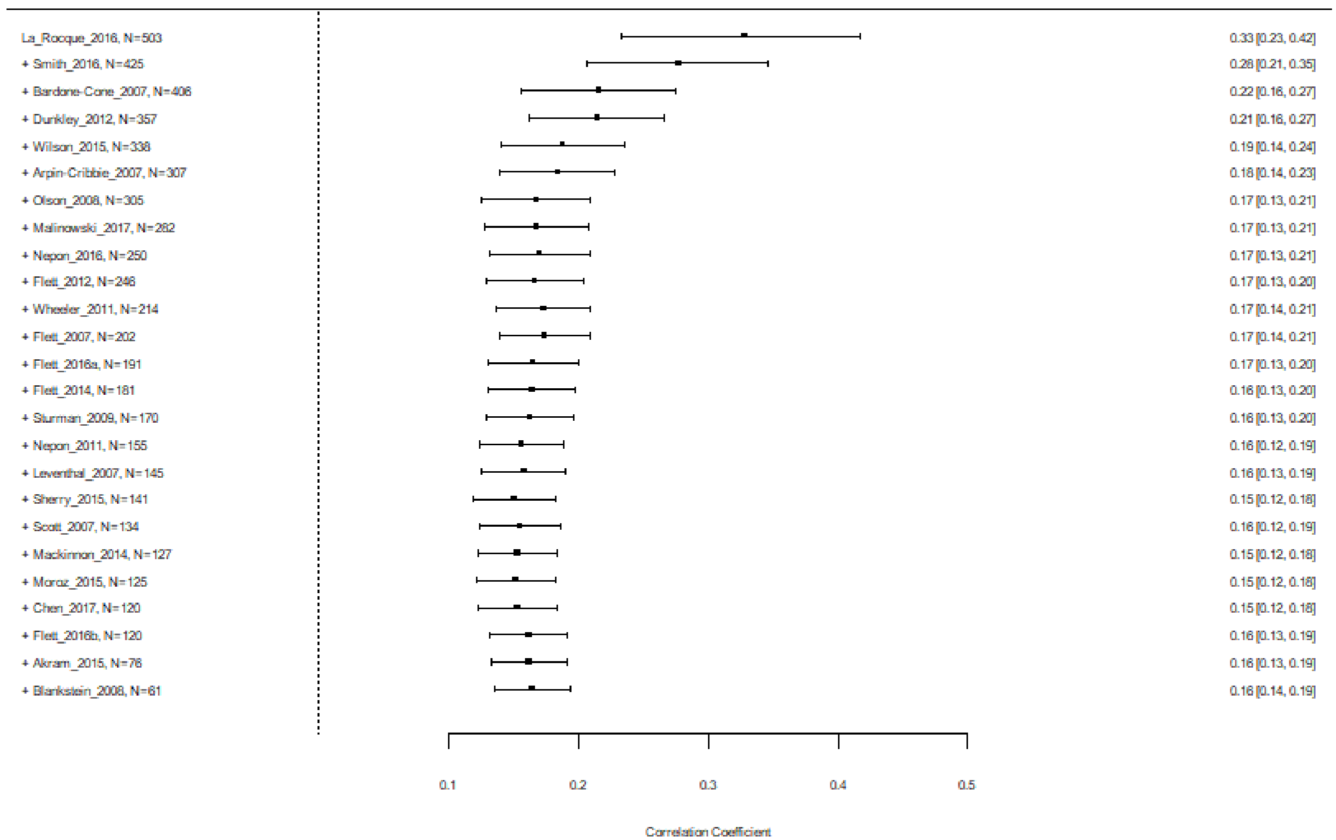


Figure 21 above includes all 25 studies. In Figures 20 and 21 the cumulative meta-analyses show a shift in the estimate of the effect size as studies with smaller sample sizes are added one by one to the analysis and the analysis is re-run, and this may be evidence of a relationship between sample size and effect size that may be due to publication bias (Borenstein et al., 2009). Also, as shown in these two cumulative meta-analyses, the estimated mean correlation was slightly larger in the random-effects analysis (with $r = .17$) than in the fixed-effects analysis (with $r = .16$).

Egger's regression test for funnel plot asymmetry was not significant with $z = 1.20$ and $p = .23$, but this meta-analysis only had enough studies to detect severe publication bias with this test (Card, 2012).

For the meta-analysis above for SOP and depression and the meta-analysis below for SPP and depression, the study by Mackinnon, Sherry, Pratt, Smith (2014) [Mackinnon_2014] combined a short 4-item version of the POMS and a 7-item version of the CES-D to form a composite score for depression. The psychometrically tested short form or the POMS (about which information is given in Table 2) has 8 items (Malouff, Schutte & Ramerth, 1985).

Fourth Meta-Analysis—The HMPS SPP Subscale and Depression

The fourth meta-analysis estimated the mean correlation between the HMPS Socially Prescribed Perfectionism (SPP) subscale and depression. It included a total of 26 studies with one of the studies being the unpublished dissertation by Leventhal (2007). The mean correlation between the HMPS SPP subscale and depression was estimated using a random-effects model and restricted maximum likelihood (REML). This meta including a dissertation by Leventhal (2007) had a total sample size of $N = 5{,}637$ with all studies included. The meta-analysis was conducted both with and without the unpublished dissertation. The correlations from the 26 individual studies were corrected for attenuation due to measurement error, and the Fisher's z_r variance was also corrected for the uncertainty that correcting the correlations for measurement error introduces (Card, 2012). For the analysis with the Leventhal (2007) dissertation, the estimate of the mean effect size for the relationship between SPP and depression using a random-effects model was $r = .45$, with a $95\%\ CI$ $[.41, .49]$, almost a large effect size. The estimate for τ^2, the between-studies variance or total heterogeneity was $T^2 = 0.0086$, and the estimated between-studies standard deviation τ was $T = 0.093$. The result of the test of heterogeneity was $Q(df = 25) = 61.72$, $p < .0001$, and $I^2 = 57.50\%$. The I^2 value of about 58% for this meta-analysis is about a medium amount of heterogeneity. For this meta-analysis, the approximate 95% credibility interval (*CrI*) was $[.29, .59]$.

From running the same meta-analysis without the Leventhal (2007) dissertation with 25 studies instead of 26 and a total $N = 5{,}492$, the estimate of the mean effect size for the relationship between SPP, the 95% *CI*, the approximate 95% *CrI* were the same, and the values for Q, T^2, and T were almost the same, and the value for I^2 increased to 59%.

Figure 22 provides a forest plot of the 26 studies used in the meta-analysis that estimated the mean correlation between the HMPS SPP subscale and depression with the effect sizes in Fisher's z_r transformed correlation coefficients, Figure 23 provides the same type of plot as Figure 22 except that the effect sizes are raw correlations rather than Fisher's z_r transformed correlation coefficients, Figure 24 provides a funnel plot with the Fisher's z_r transformed correlation coefficients on the X-axis and the standard error on the Y-axis, and Figure 25 provides another funnel plot that has Fisher's z_r transformed correlation coefficients on the x-axis and sample size, instead of the standard error, on the y-axis. Figure 26 provides the funnel plot that resulted from doing a trim and fill analysis to obtain an unbiased estimate of the mean correlation between the HMPS SPP subscale and depression. Figure 27 provides the forest plot from a cumulative meta-analysis using a random-effects model, and Figure 28 provides another forest plot from a cumulative meta-analysis but using a fixed-effects model. All figures for this meta-analysis include all 26 studies.

Figure 22

Forest Plot SPP_D Fisher's Zr Correlations Corrected for Attenuation Random-Effects Models

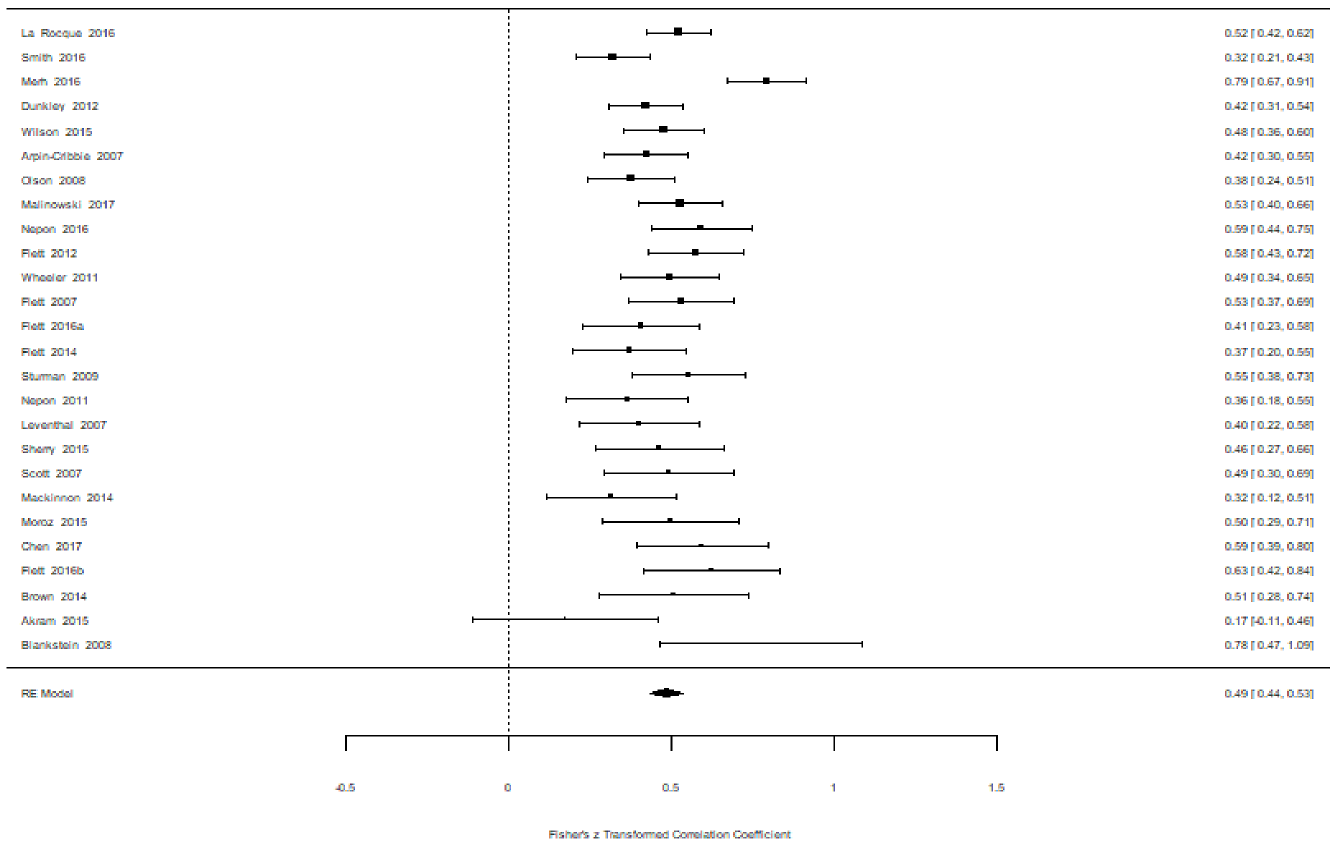


Figure 23

Forest Plot SPP_D Correlations Corrected for Attenuation Random-Effects Model

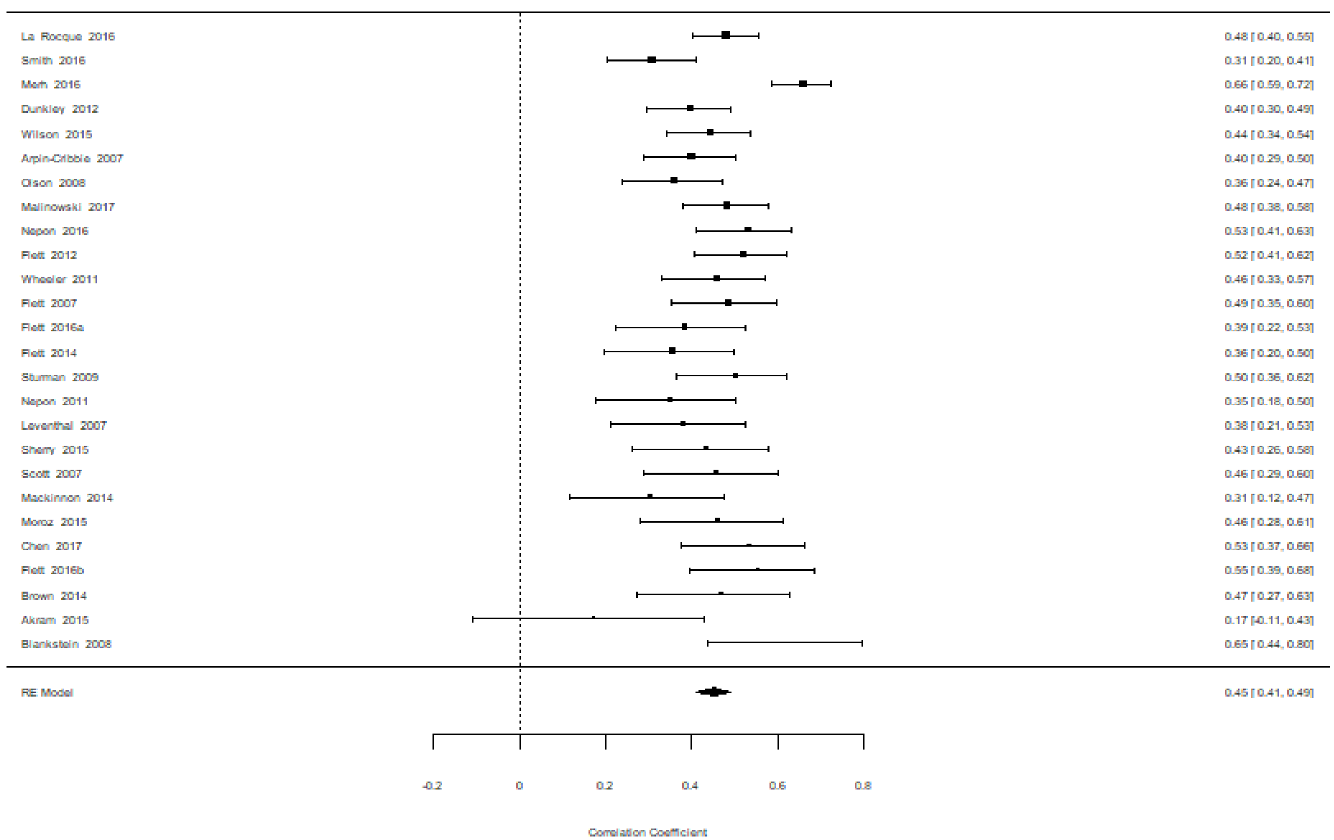


Figure 24

Funnel Plot Random-Effects SPP_D using Standard Error

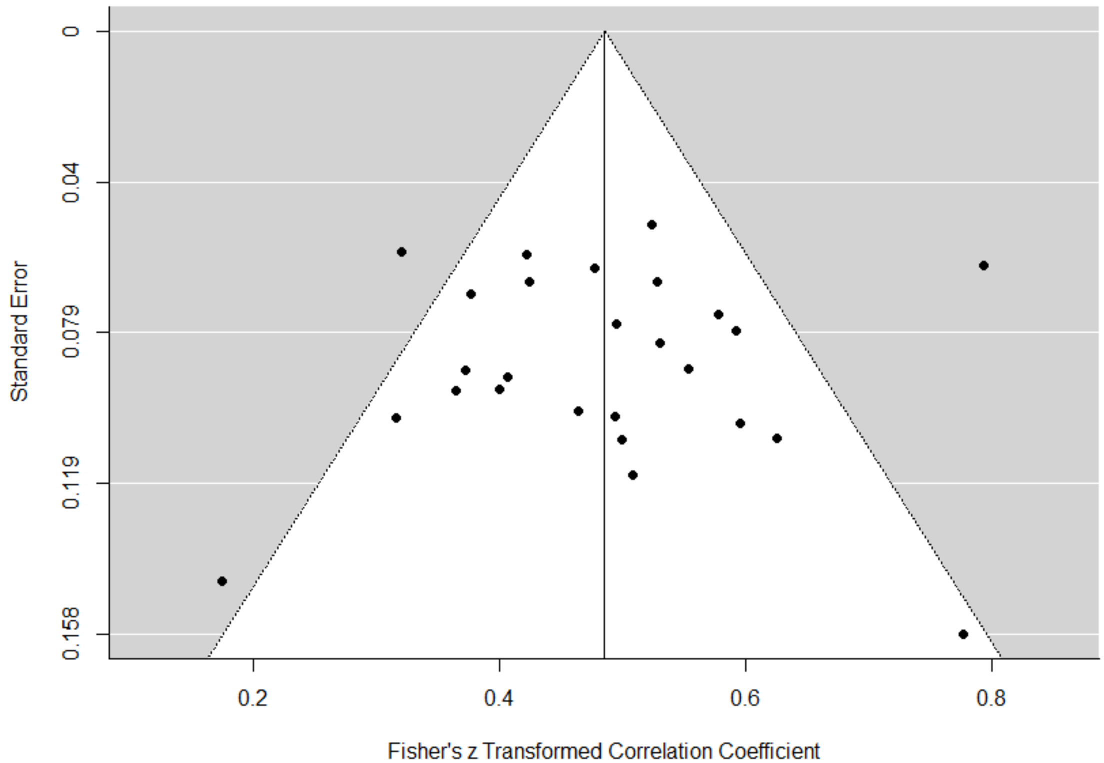


Figure 25

Funnel Plot Random-Effects SPP_D using Sample Size

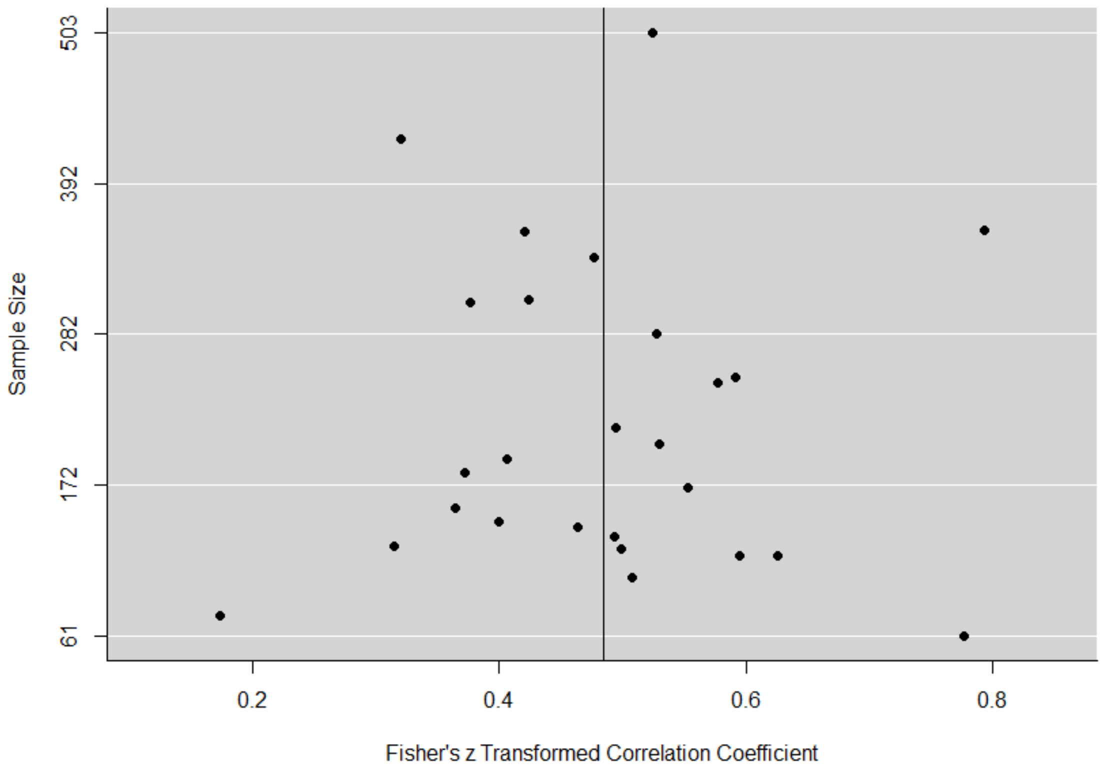


Figure 26

Funnel Plot SPP_D after Trim & Fill Random-Effects Model

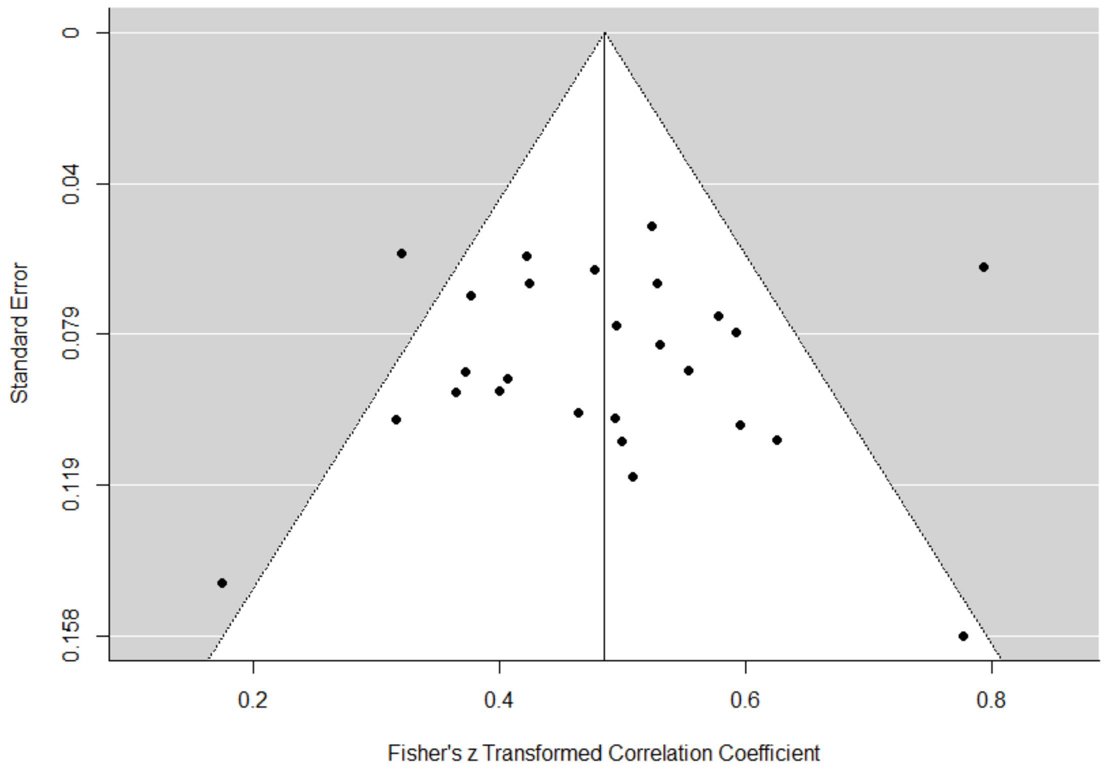


Figure 26 contains all 26 studies. If there were evidence of publication bias or a relationship between effect size and sample size, there would be missing studies on the lower left side of the funnel plot where studies should be that had small sample sizes and that found small effect sizes. The results of the trim and fill method showed that there were zero studies missing from the left side, and all the estimates from the trim and fill method were the same as those from the analysis that used all 26 studies (including the dissertation by Leventhal, 2007).

Figure 27

Cumulative Meta-Analysis for Random-Effects Model Starting with Largest Study

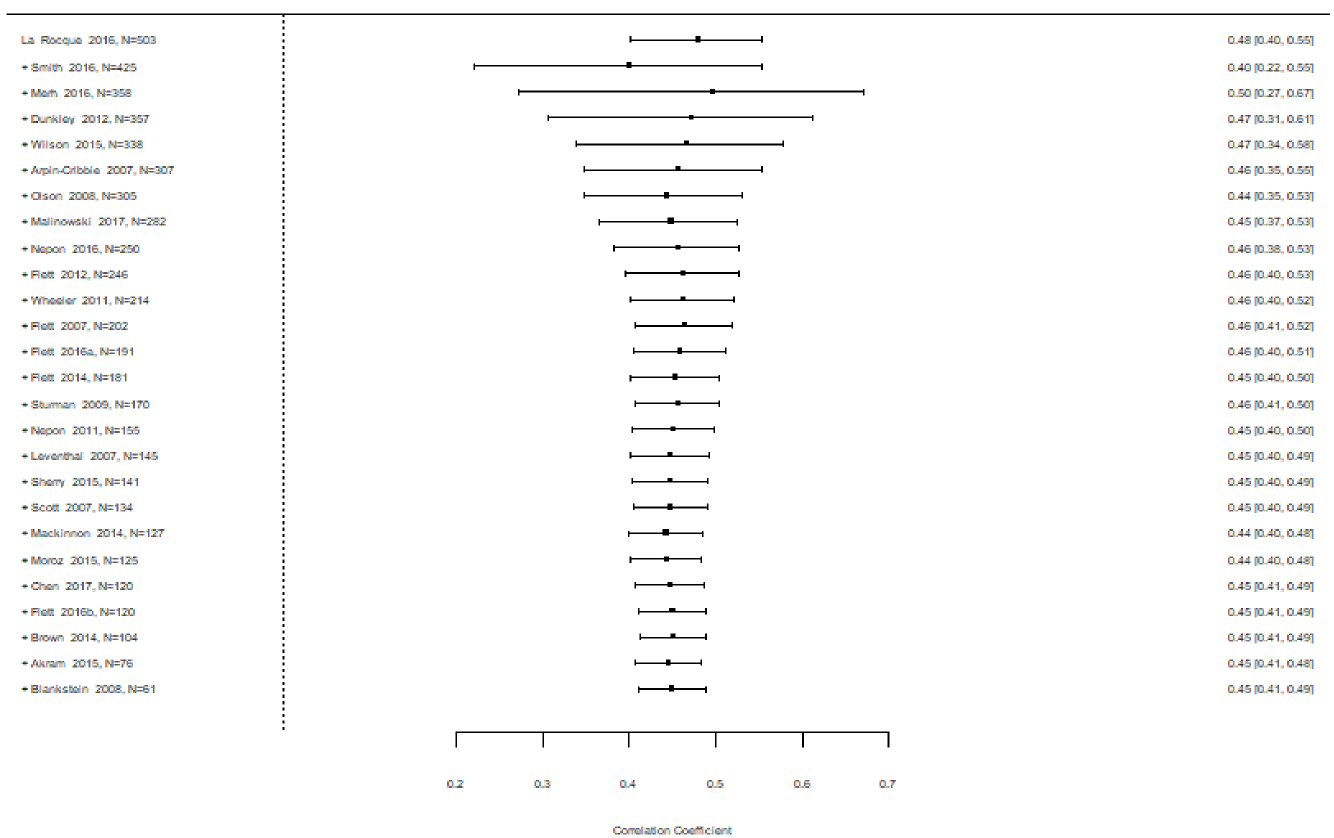
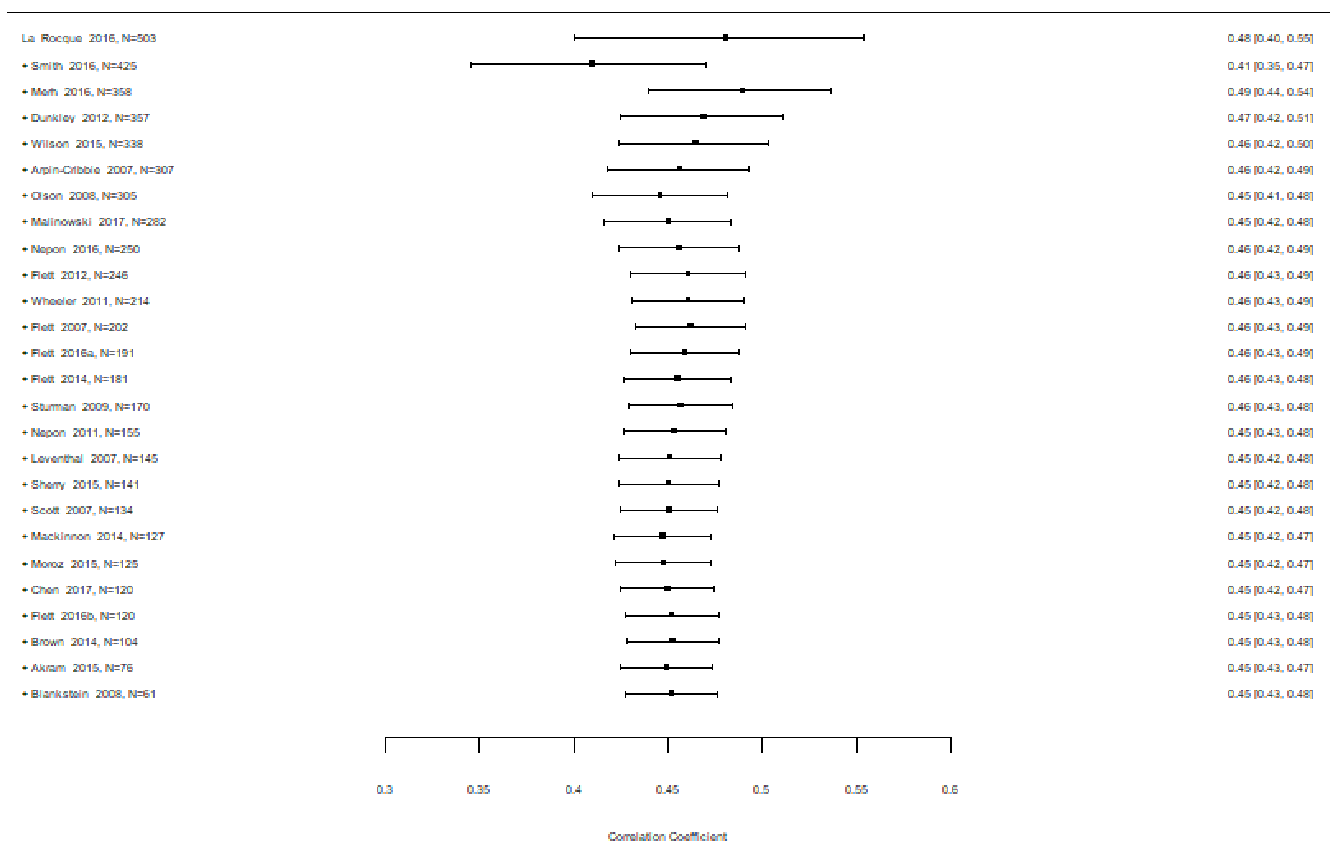


Figure 28

Cumulative Meta-Analysis for Fixed-Effects Model Starting with Largest Study



In Figures 27 and 28 above for the two cumulative meta-analyses, there was little shift in the estimate of the effect size as studies with increasingly smaller sample sizes were added to the analysis one by one and the analysis re-run, and this may be evidence against the existence of publication bias in this meta-analysis because there did not appear to be much of a relationship between sample size and effect size.

Egger's regression test for funnel plot asymmetry was not significant with $z = -0.16$ and $p = .88$, but this meta-analysis only had enough studies to detect severe publication bias with this test (Card, 2012).

Fifth Meta-Analysis—FMPS PS Subscale and Depression

The fifth meta-analysis estimated the mean correlation between the FMPS Personal Standards (PS) subscale and depression. It included a total of 17 studies with a total sample size of $N = 3{,}781$. The mean correlation between the FMPS PS subscale and depression was estimated using a random-effects model and restricted maximum likelihood (REML). The correlations from the 17 individual studies were corrected for attenuation due to measurement error, and the Fisher's z_r variance was also corrected for the uncertainty that correcting the correlations for measurement error introduces (Card, 2012). The estimate of the mean correlation coefficient effect size for the relationship between PS and depression using a random-effects model was $r = .08$, with a $95\%\ CI\ [.03, .14]$, a small effect size. The estimate for τ^2, the between-studies variance or total heterogeneity was $T^2 = 0.0064$ and the estimated between-studies standard deviation τ was $T = 0.080$. The result of the test of heterogeneity was $Q(df = 16) = 33.92$, $p < .006$, and $I^2 = 50.46\%$. The I^2 value of about 50% for this meta-analysis is about a medium amount of heterogeneity. For this meta-analysis, the approximate 95% credibility interval (*CrI*) was $[-.08, .24]$.

The PS dataset had four studies that had only women participants: Chang et al. (2011), Steele et al (2011), Sturman et al. (2009), and DiBartolo et al. (2008). This fifth meta-analysis was run both with and without those four studies that had only women participants. Without the four studies that had only women participants, there were a total of 13 studies with $N = 3{,}177$ The mean correlation coefficient from the meta-analysis without the four studies that had only women was $r = .07$, with a $95\%\ CI$ [.01, .14], less than a small effect size. The estimate for τ^2, the between-studies variance or total heterogeneity was $T^2 = 0.0067$ and the estimated between-studies standard deviation τ was $T = 0.082$. The result of the test of heterogeneity was $Q(df = 12) = 26.17, p < .010$, and $I^2 = 54.23\%$. The I^2 value of about 54% for this meta-analysis is about a medium amount of heterogeneity. For this meta-analysis, the approximate 95% credibility interval (*CrI*) was $[-.10, .24]$. Running the meta-analysis without the four studies that had only women participants reduced the mean correlation from $r = .08$ to $r = .07$, and it reduced the lower bound of the 95% confidence interval from $r = .03$ to $r = .01$, and but the upper bound of the 95% confidence interval stayed the same at $r = .14$, and the value for the Q statistic decreased slightly and the values for T^2, T, and I^2 increased slightly, and the 95% credibility interval became slightly wider.

Figure 29 provides a forest plot of the 17 studies used in the meta-analysis that estimated the mean correlation between the FMPS PS subscale and depression with the effect sizes in Fisher's z_r transformed correlation coefficients. Figure 30 provides the same type of plot as Figure 29 except that the effect sizes are raw correlations rather than Fisher's z_r transformed correlation coefficients. Figure 31 provides a funnel plot with the Fisher's z_r transformed correlation coefficients on the x-axis and the standard error on the y-axis. Figure 32 provides another funnel plot that has Fisher's z_r transformed correlation coefficients on the x-axis and sample size, instead of the standard error, on the y-axis. Figure 33 provides a funnel plot of the results of the trim and fill method for estimating what the effect size would be if it were corrected for publication bias. Figure 34 provides a forest plot for a cumulative meta-analysis using a random-effects model, and Figure 35 provides another forest plot for a cumulative meta-analysis but using a fixed-effects model. All figures for this meta-analysis include all 17 studies (including the four studies that only had women participants).

Figure 29

Forest Plot PS_D Fisher's Zr Correlations Corrected for Attenuation Random-Effects Model

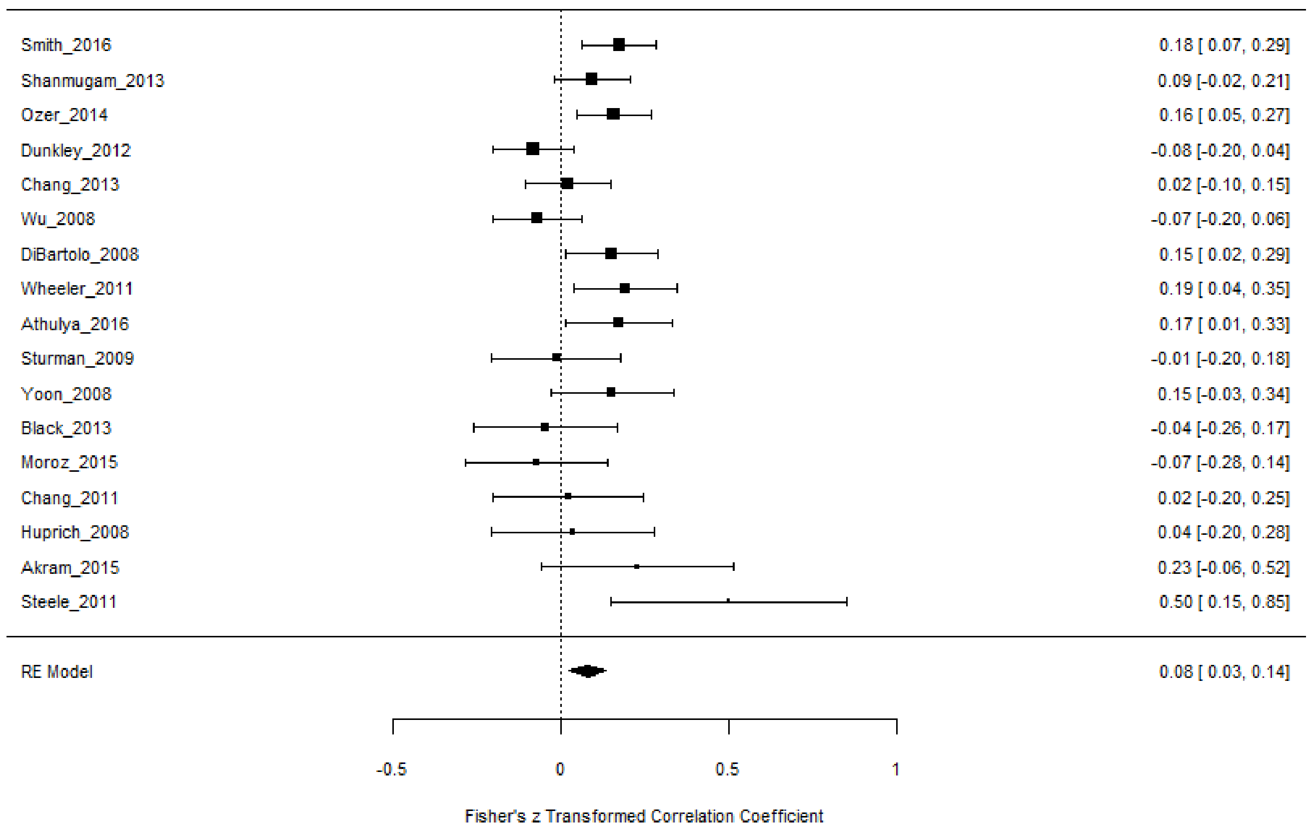
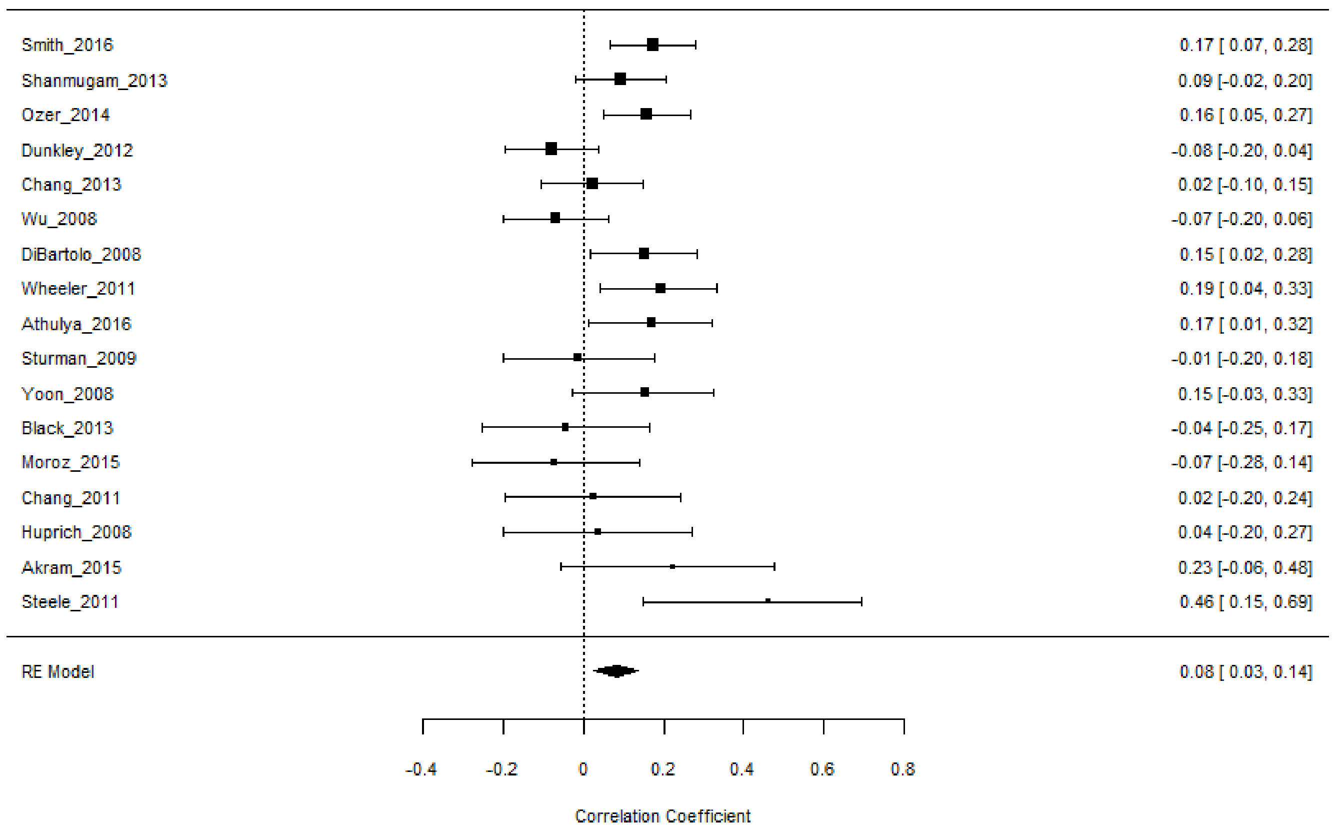


Figure 30

Forest Plot PS_D Correlations Corrected for Attenuation Random-Effects Model



Figures 29 and 30 (above) include the four studies that had only female participants: Chang et al. (2011), Steele et al. (2011), Sturman et al. (2009), and DiBartolo et al. (2008).

Figure 31

Funnel Plot Random-Effects PS_D using Standard Error

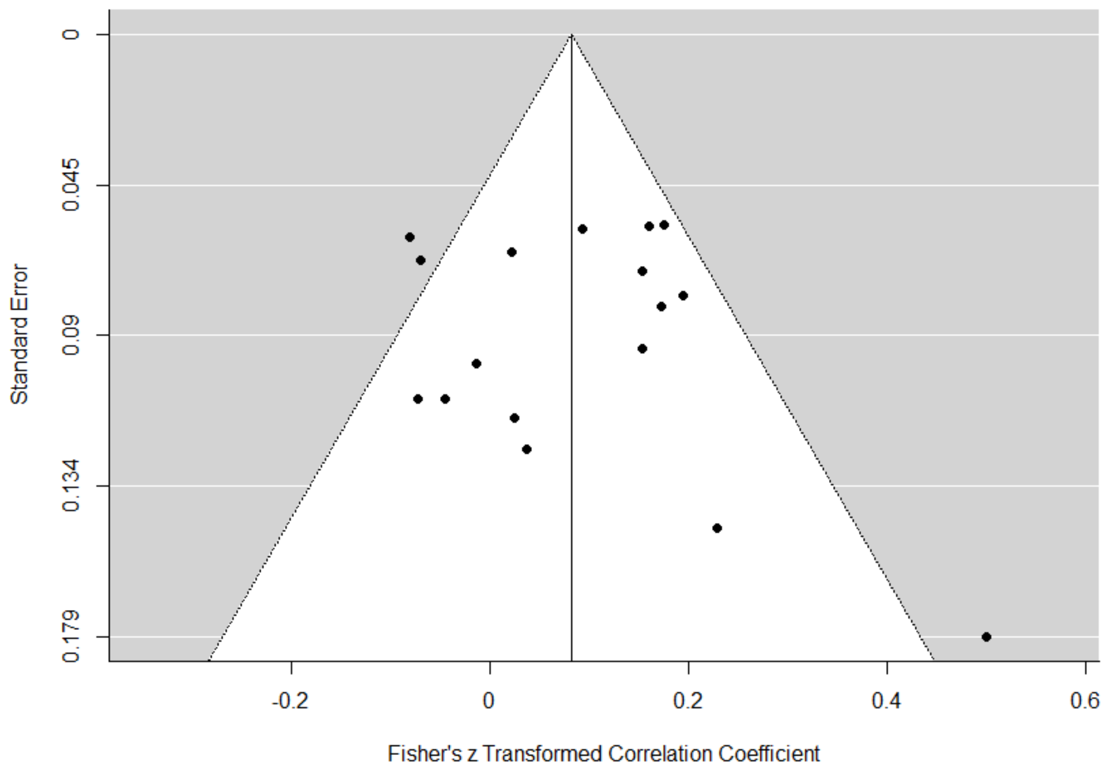


Figure 31 includes all 17 studies. The outlier in the lower right-hand corner is the study of 39 women with eating disorders by Steele et al. (2011).

Figure 32

Funnel Plot Random-Effects PS_D using Sample Size

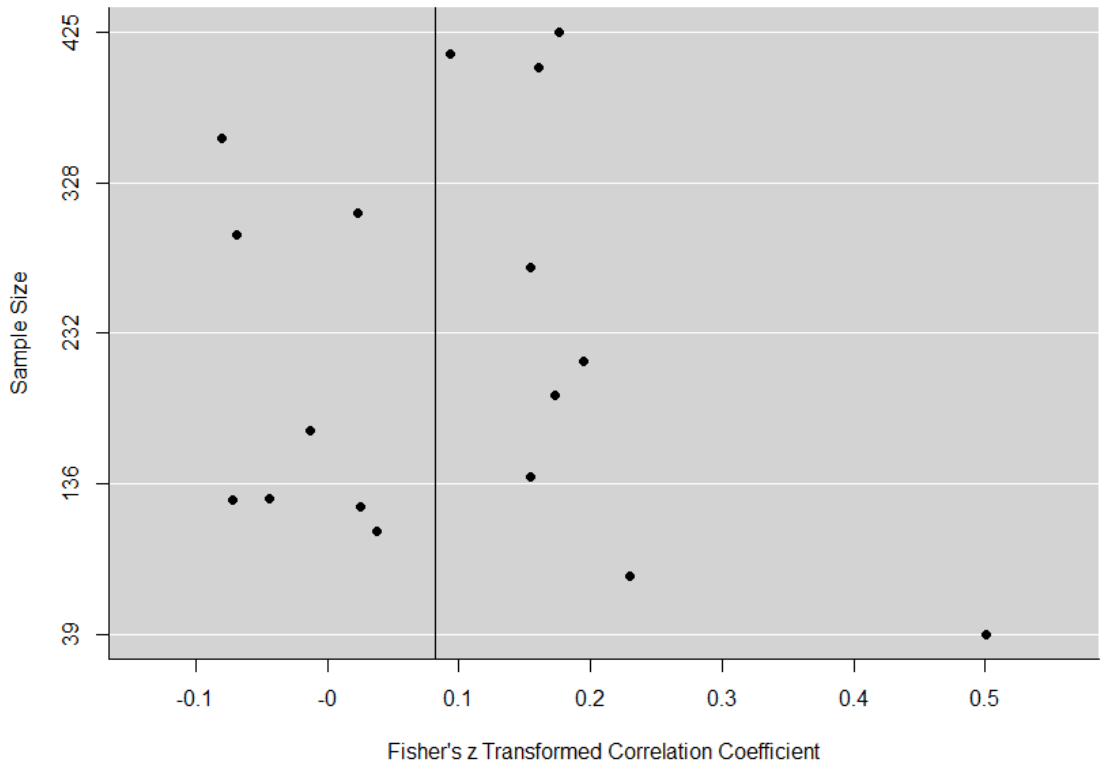


Figure 32 includes all 17 studies.

Figure 33

Funnel Plot PS_D after Trim & Fill Random-Effects Model

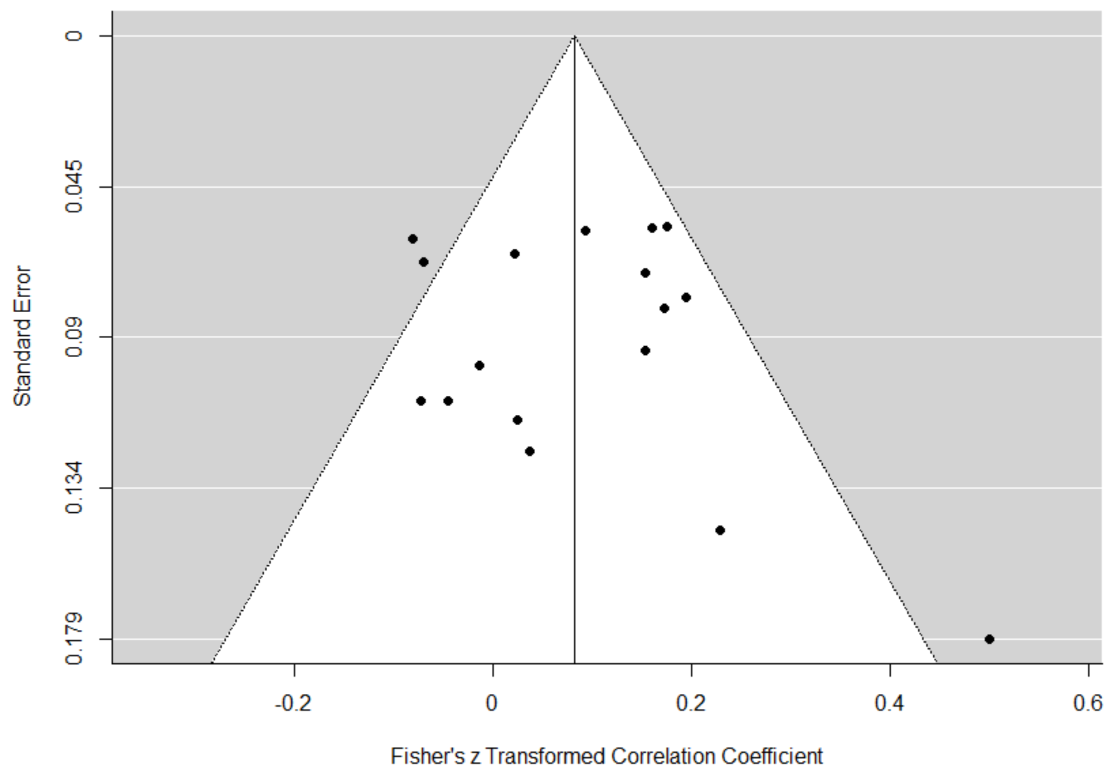


Figure 33 includes all 17 studies. The estimated number of missing studies on the lower left side is zero. Egger's regression test for funnel plot asymmetry was not significant with $z = 0.86$ and $p = .39$, but this meta-analysis only had enough studies to detect severe publication bias with this this statistical test for asymmetry (Card, 2012).

Figure 34

Cumulative Meta-Analysis for Random-Effects Model Starting with Largest Study

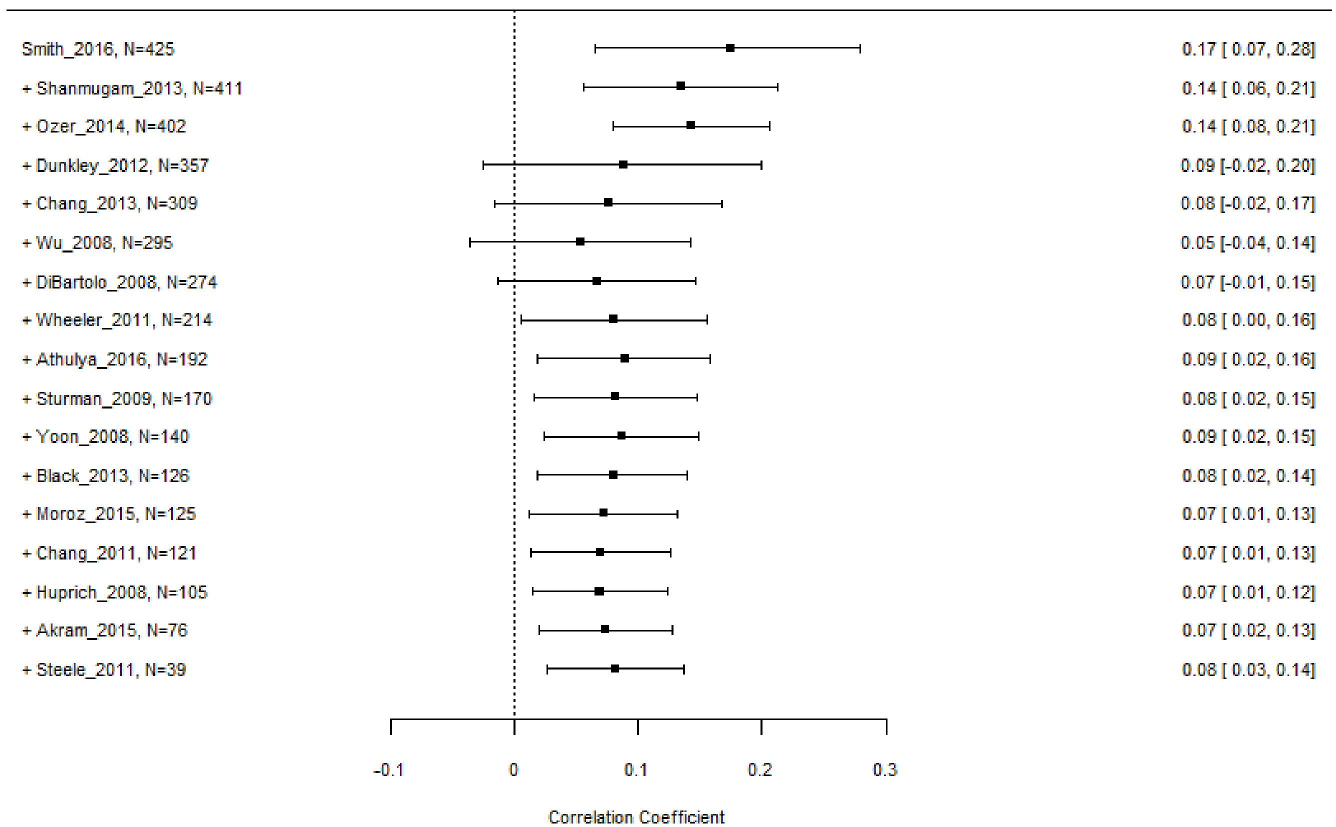
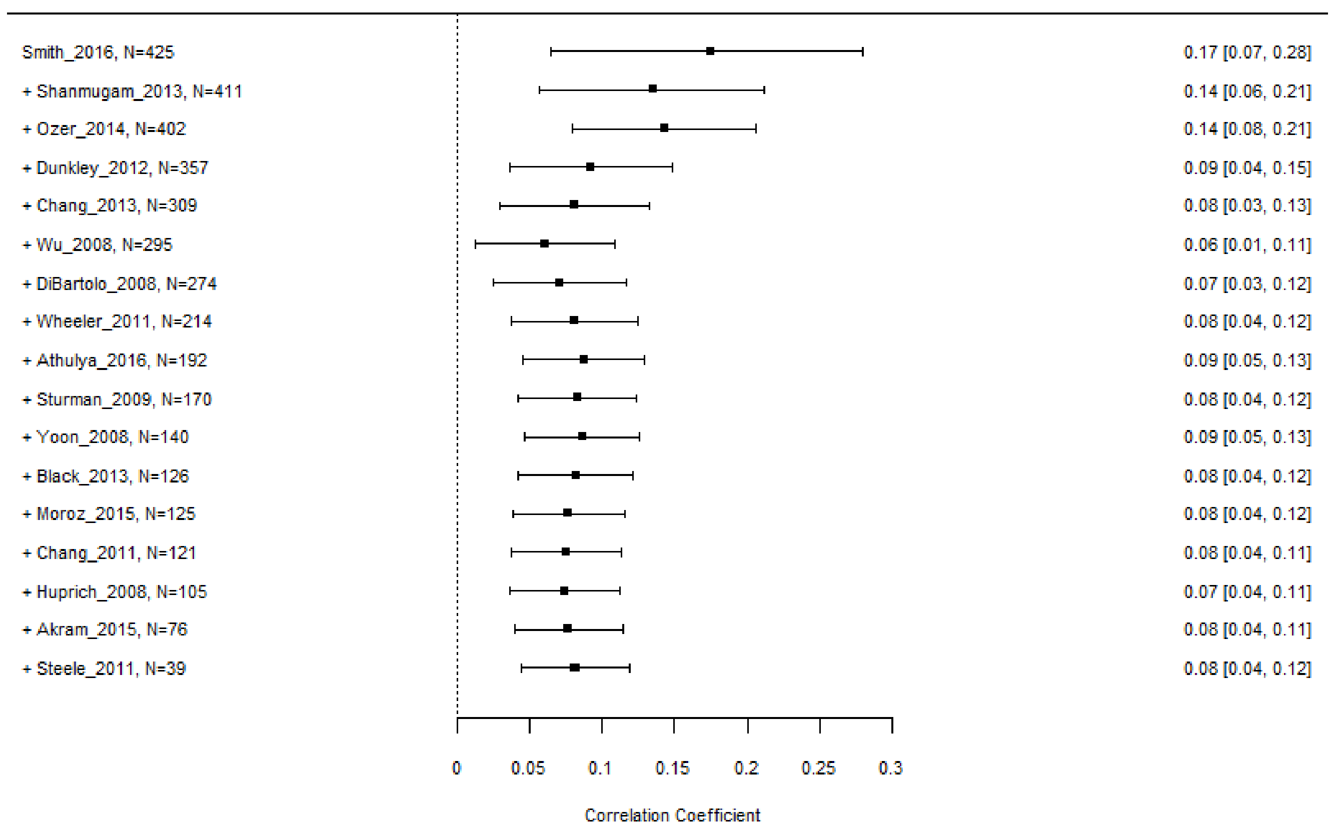


Figure 35

Cumulative Meta-Analysis for Fixed-Effects Model Starting with Largest Study



Figures 34 and 35 includes all 17 studies. In the cumulative meta-analyses in Figures 34 and 35, there is a shift in the estimate of the effect size as studies with smaller sample sizes are added one by one and the analysis is re-run. This is especially apparent in Figure 35, the fixed-effect cumulative meta-analysis, and this may be evidence of the existence of a relationship between sample size and effect size in this meta-analysis that may be due to publication bias (Borenstein et al., 2009)

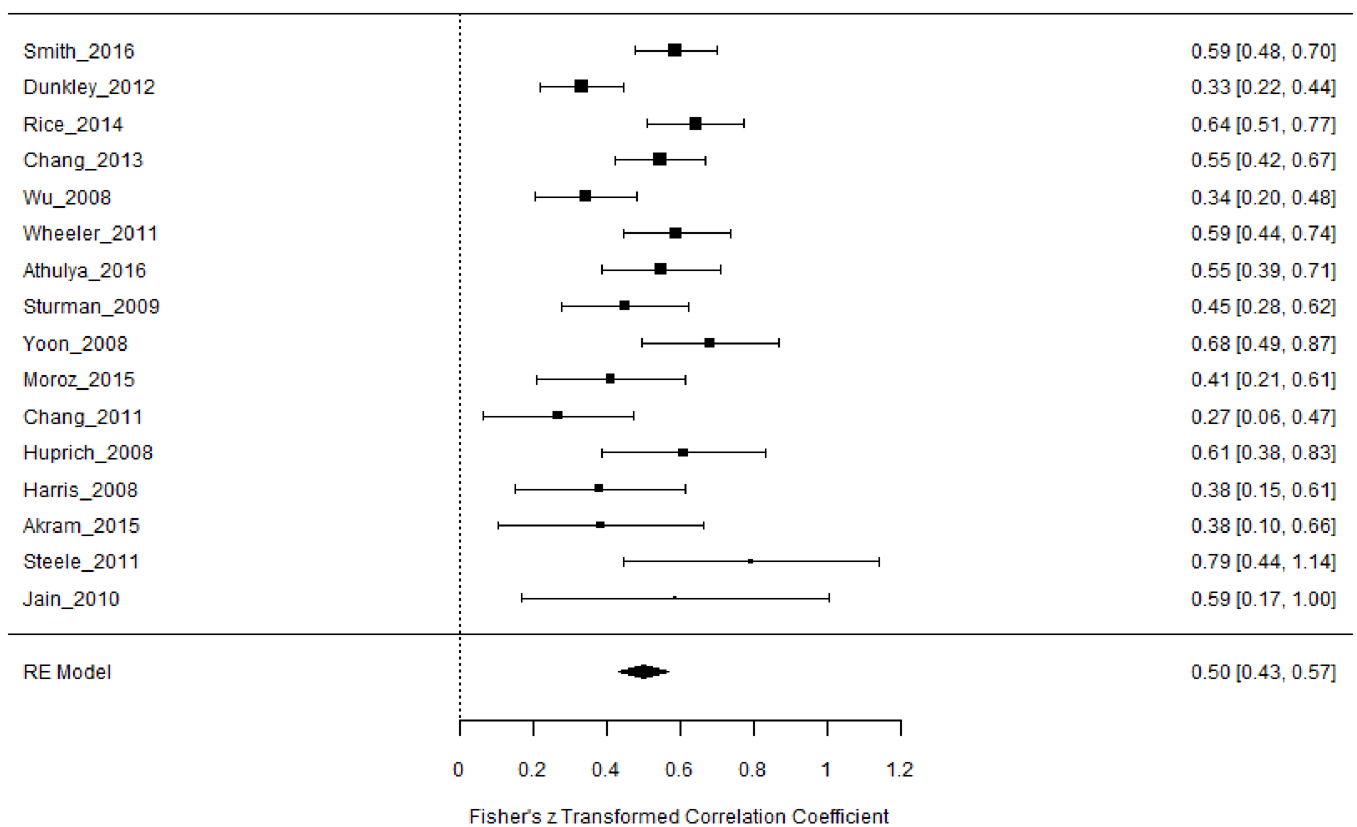
Sixth Meta-Analysis—FMPS CM Subscale and Depression

The sixth meta-analysis estimated the mean correlation between the FMPS Concern Over Mistakes (CM) subscale and depression. It included a total of 16 studies with a total sample size of $N = 3{,}034$ The mean correlation between the FMPS CM subscale and depression was estimated using a random-effects model and restricted maximum likelihood (REML). The correlations from the 16 individual studies were corrected for attenuation due to measurement error, and the Fisher's z_r variance was also corrected for the uncertainty that correcting the correlations for measurement error introduces (Card, 2012). The estimate of the mean correlation coefficient effect size for the relationship between CM and depression using a random-effects model was $r = .46$, with a 95% CI $[.41, .52]$, a large effect size. The estimate for τ^2, the between-studies variance or total heterogeneity was $T^2 = 0.0106$ and the estimated between-studies standard deviation τ was $T = 0.1028$. The result of the test of heterogeneity was $Q(df = 15) = 37.63$, $p = .001$, and $I^2 = 59.52\%$. The I^2 value of about 59% for this meta-analysis is about a medium amount of heterogeneity. For this meta-analysis, the approximate 95% credibility interval (*CrI*) was $[.28, .61]$.

Figure 36 provides a forest plot of the 16 studies used in the meta-analysis that estimated the mean correlation between the FMPS CM subscale and depression with the effect sizes in Fisher's z_r transformed correlation coefficients. Figure 37 provides the same type of plot as Figure 36 except that the effect sizes are raw correlations rather than Fisher's z_r transformed correlation coefficients. Figure 38 provides a funnel plot with the Fisher's z_r transformed correlation coefficients on the x-axis and the standard error on the y-axis, and Figure 39 provides another funnel plot that has Fisher's z_r transformed correlation coefficients on the x-axis and sample size, instead of the standard error, on the y-axis. Figure 40 provides a funnel plot of the results of the trim and fill method for estimating what the effect size would be if it were corrected for publication bias. Figure 41 provides the forest plot for a cumulative meta-analysis that used a random-effects model, and Figure 42 provides another forest plot for a cumulative meta-analysis that used a fixed-effects model. All figures for this meta-analysis include all 16 studies.

Figure 36

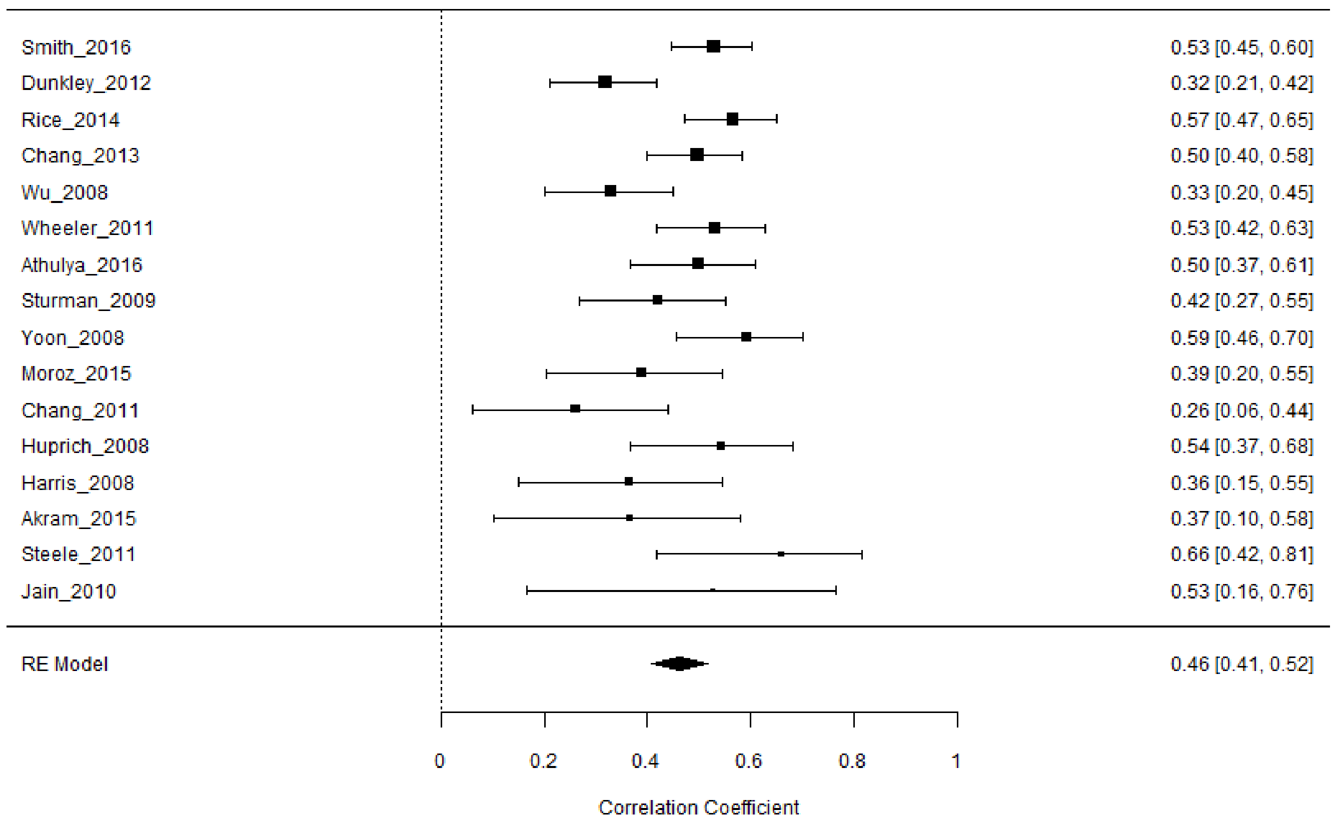
Forest Plot CM_D Fisher's Zr Correlations Corrected for Attenuation Random-Effects Model



.

Figure 37

Forest Plot CM_D Correlations Corrected for Attenuation Random-Effects Model



Figures 36 and 37 (above) have all 16 studies including outlier Steele et al. (2011).

Figure 38

Funnel Plot Random-Effects CM_D using Standard Error

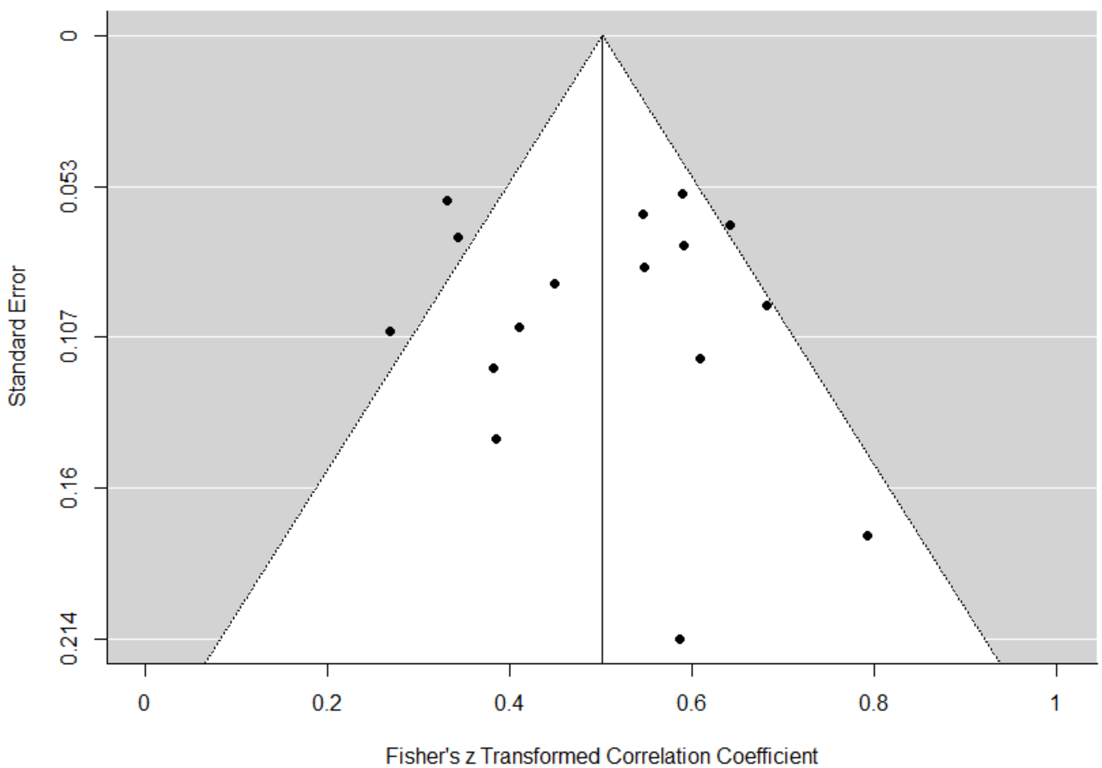


Figure 38 (above) has all 16 studies including the outlier Steele et al. (2011).

Figure 39

Funnel Plot Random-Effects CM_D using Sample Size

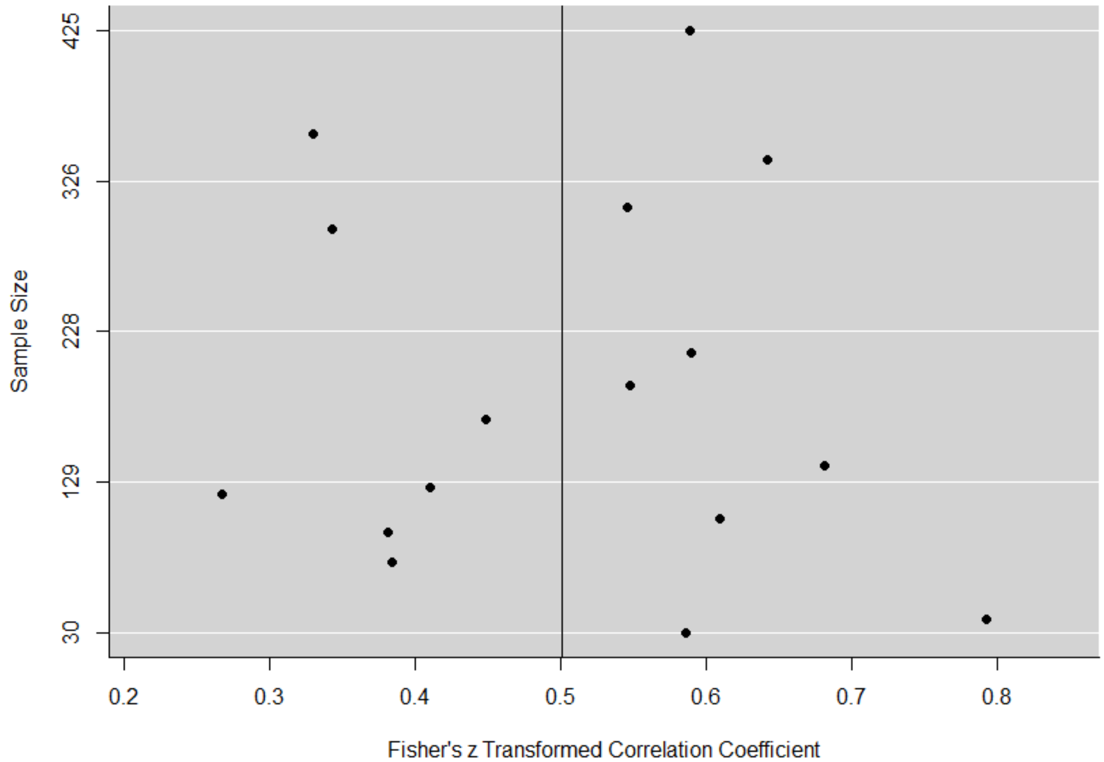


Figure 39 (above) has all 16 studies including outlier Steele et al. (2011).

Figure 40

Funnel Plot CM_D after Trim & Fill Random-Effects Model

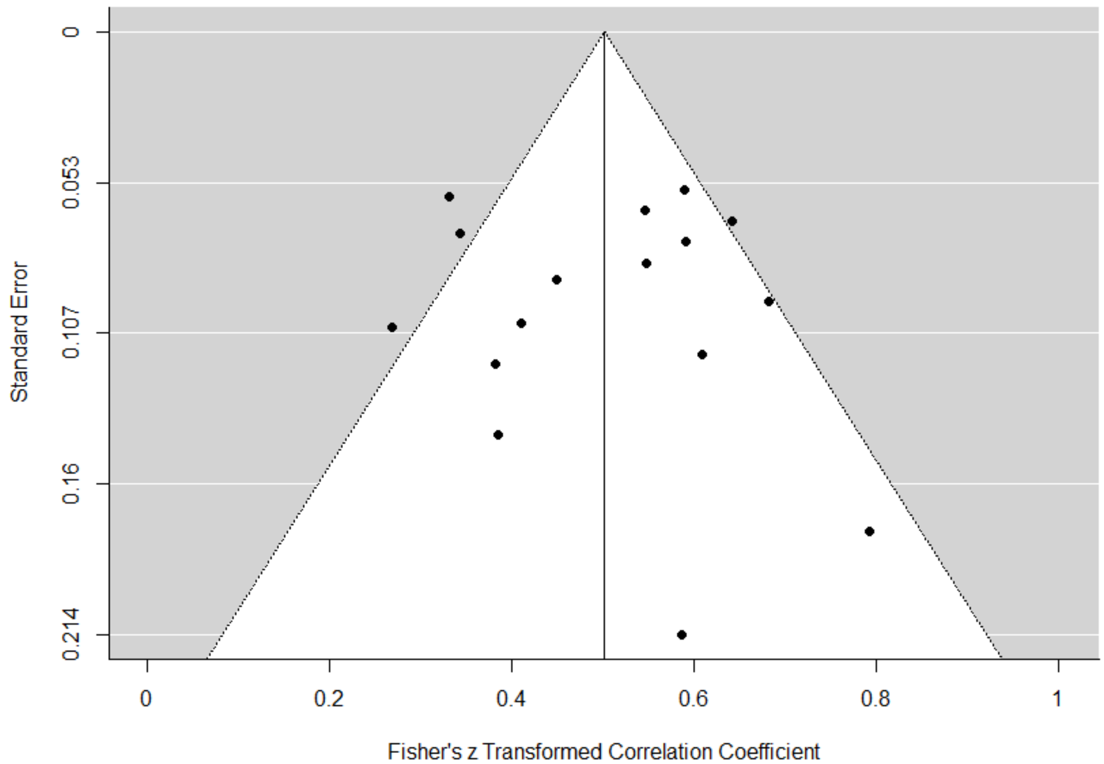


Figure 40 has all 16 studies including outlier Steele et al. (2011). The estimated number of missing studies on the left side is zero. Egger's regression test for funnel plot asymmetry was not significant with $z = 0.38$ and $p = .70$, but this meta-analysis did not have enough studies for this test to detect even severe publication bias (Card, 2012).

Figure 41

Cumulative Meta-Analysis for Random-Effects Model Starting with Largest Study

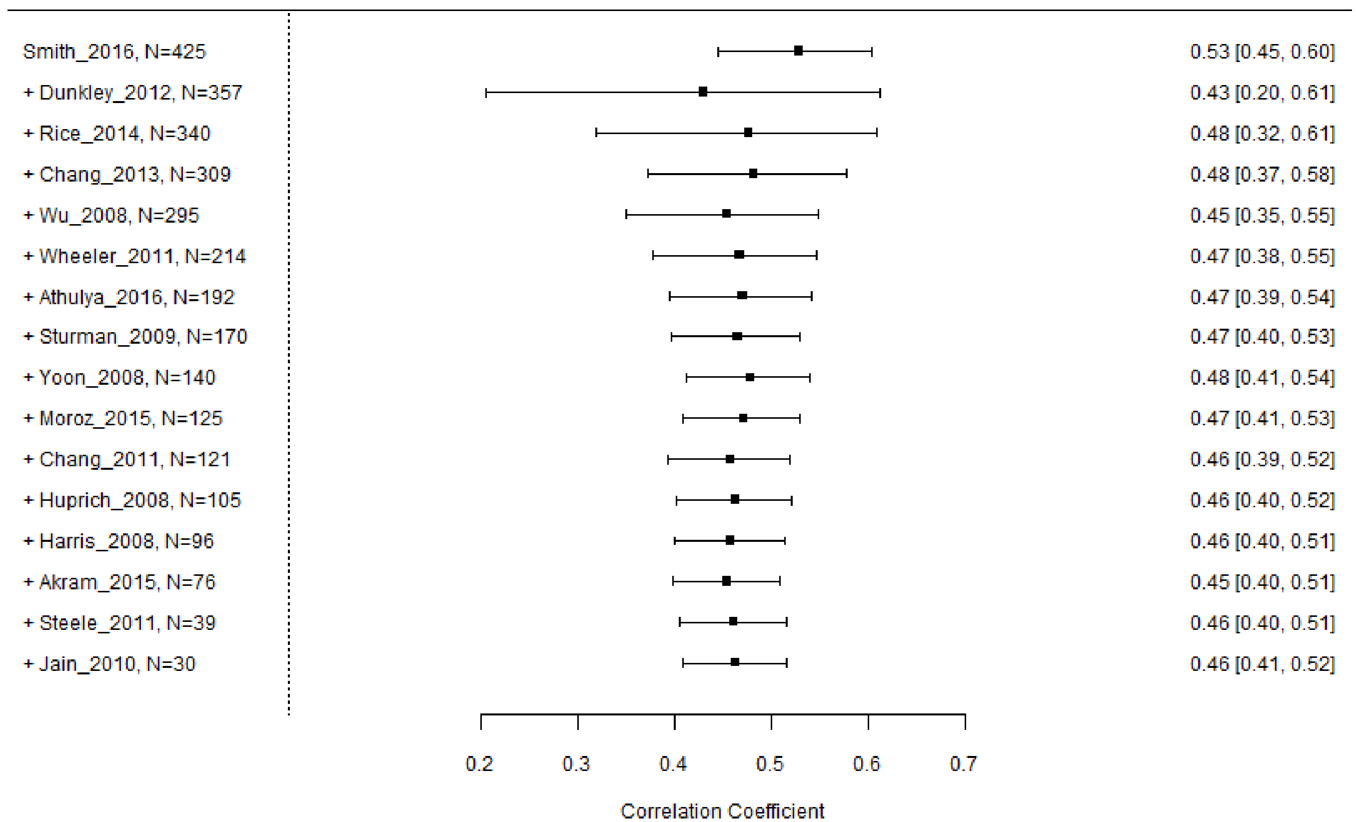
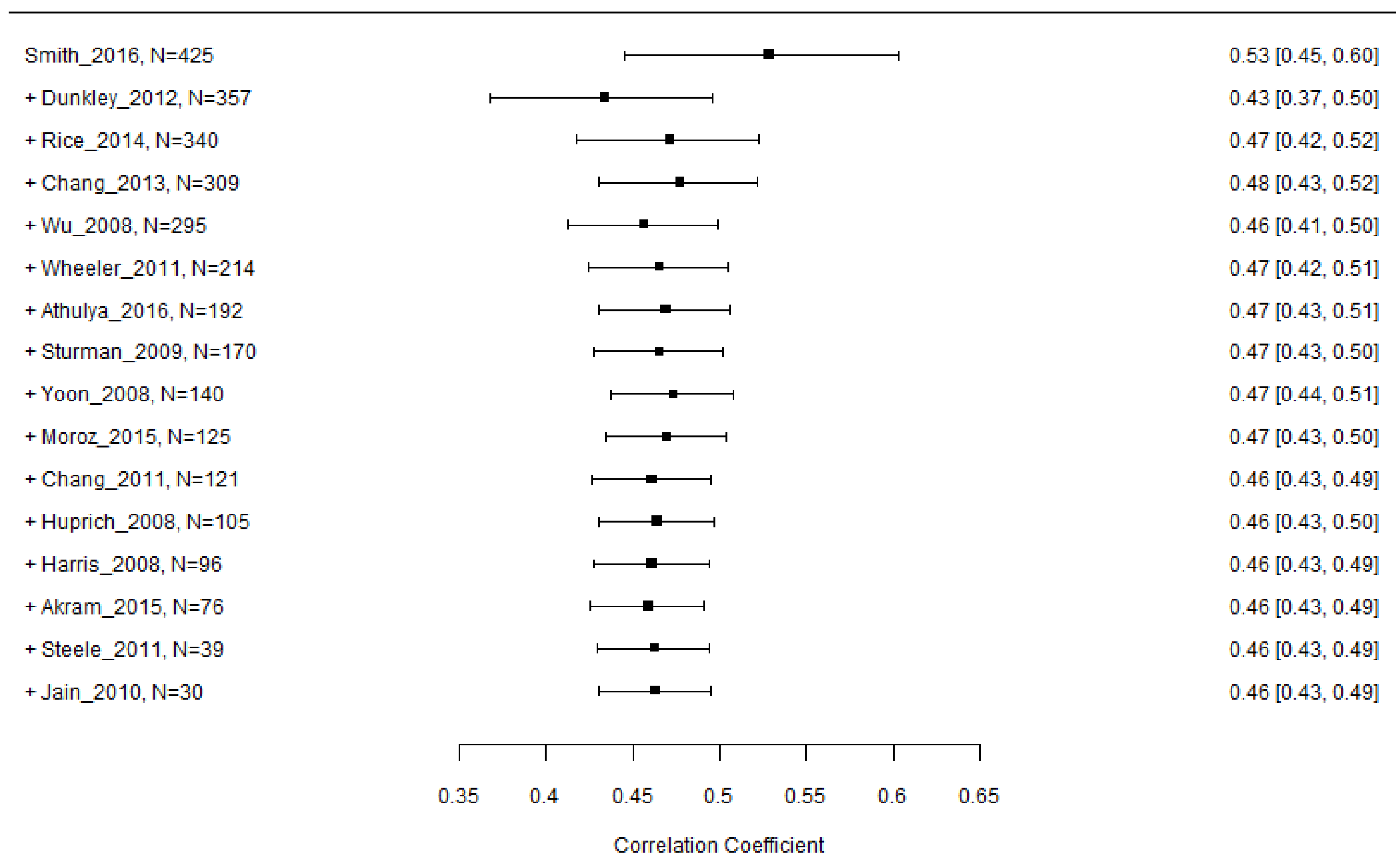


Figure 42

Cumulative Meta-Analysis for Fixed-Effects Model Starting with Largest Study



Figures 41 and 42 include all 16 studies including the outlier (Steele et al., 2011) which had a sample of 39 females in treatment for eatting disorders. The correlation for the Steele et al. study (2011) was higher than the rest of the studies on CM and PS. In Figures 41 and 42 for the cumulative meta-analyses, the effect size estimate does not shift in either direction as studies with increasingly smaller sample sizes are added one by one and the analysis re-run, and this might be evidence against the existence of publication bias in this meta-analysis because there does not appear to be much of a relationship between sample size and effect size. Except for the study with the strongest correlation, the absolute value of the correlation did not decrease much as studies with smaller sample sizes were included in the analysis.

Seventh Meta-Analysis—FMPS DA Subscale and Depression

The seventh meta-analysis estimated the mean correlation between the FMPS Doubts about Actions (DA) subscale and depression. It included a total of 14 studies with a total sample size of $N = 2{,}915$. The mean correlation between the FMPS DA subscale and depression was estimated using a random-effects model and restricted maximum likelihood (REML). The correlations from the 14 individual studies were corrected for attenuation due to measurement error, and the Fisher's z_r variance was also corrected for the uncertainty that correcting the correlations for measurement error introduces (Card, 2012). The estimate of the mean correlation coefficient effect size for the relationship between DA and depression using a random-effects model was $r = .55$, with a 95% CI [.48, .61], a large effect size. The estimate for τ^2, the between-studies variance or total heterogeneity was $T^2 = 0.0225$ and the estimated between-studies standard deviation τ was $T = 0.1501$. The result of the test of heterogeneity was $Q(df = 13) = 55.25$, $p < .001$, and $I^2 = 75.97\%$. The I^2 value of about 76% for this meta-analysis is a large amount of heterogeneity. For this meta-analysis, the approximate 95% credibility interval (*CrI*) was $[.29, .73]$.

Figure 43 provides a forest plot of the 14 studies used in the meta-analysis that estimated the mean correlation between the FMPS DA subscale and depression with the effect sizes in Fisher's z_r transformed correlation coefficients. Figure 44 provides the same type of plot as Figure 42 except that the effect sizes are raw correlations rather than Fisher's z_r transformed correlation coefficients. Figure 45 provides a funnel plot with the Fisher's z_r transformed correlation coefficients on the x-axis and the standard error on the y-axis, and Figure 46 provides another funnel plot that has Fisher's z transformed correlation coefficients on the x-axis and sample size, instead of the standard error, on the y-axis. Figure 47 provides a funnel plot of the results of the trim and fill method for estimating what the effect size would be if it were corrected for publication bias. Figure 48 provides the forest plot for a cumulative meta-analysis that used a random-effects model, and Figure 49 provides another forest plot for a cumulative meta-analysis that used a fixed-effects model. All figures for this meta-analysis include all 14 studies.

Figure 43

Forest Plot DA_D Fisher's Zr Correlations Corrected for Attenuation Random-Effects Model

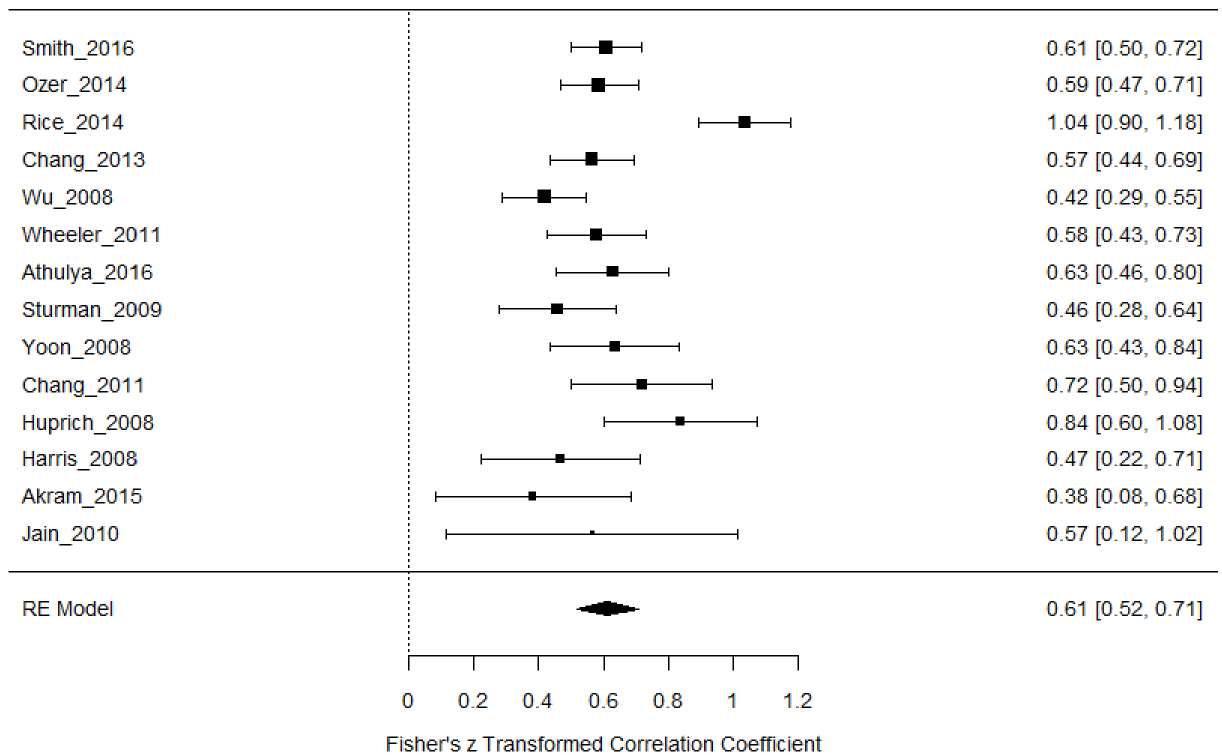


Figure 44

Forest Plot DA_D Correlations Corrected for Attenuation Random-Effects Model

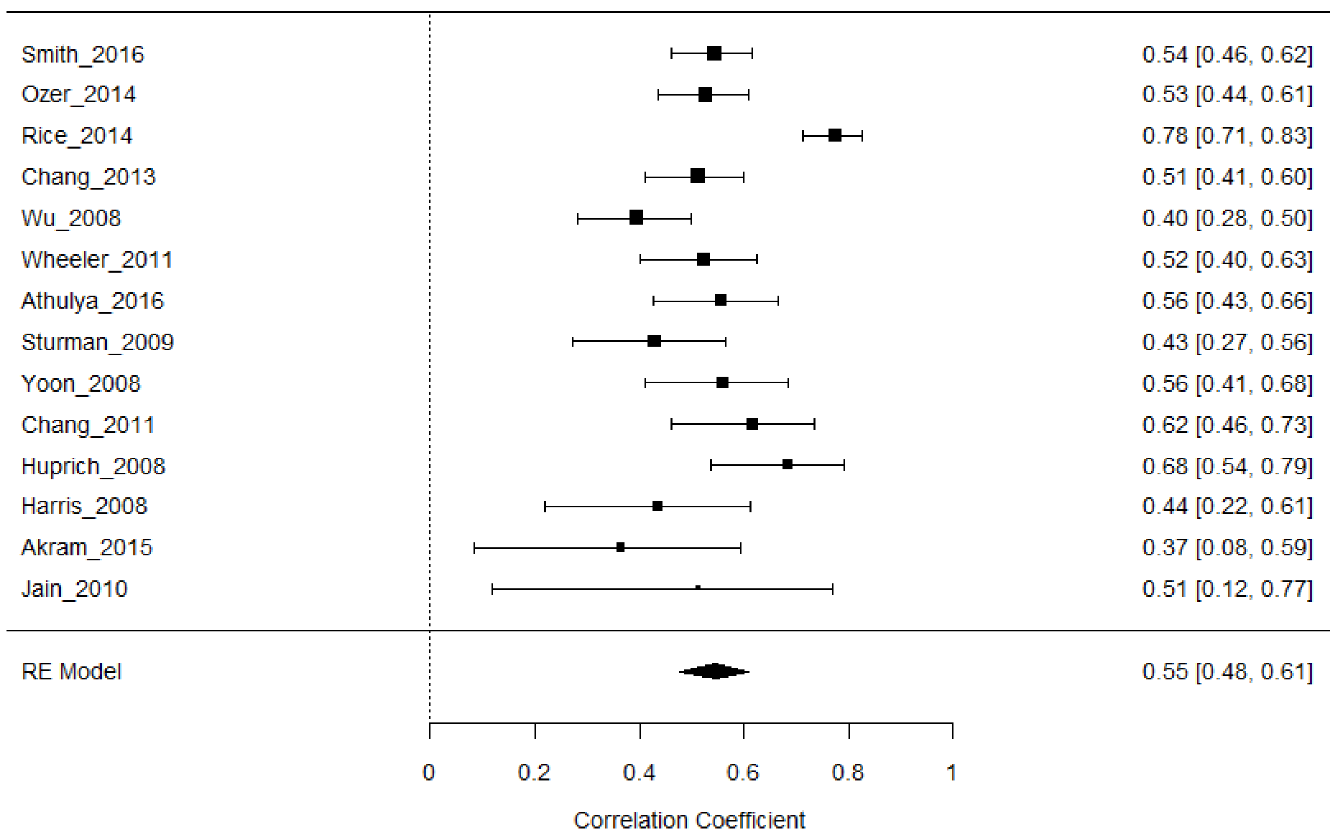


Figure 45

Funnel Plot Random-Effects DA_D using Standard Error

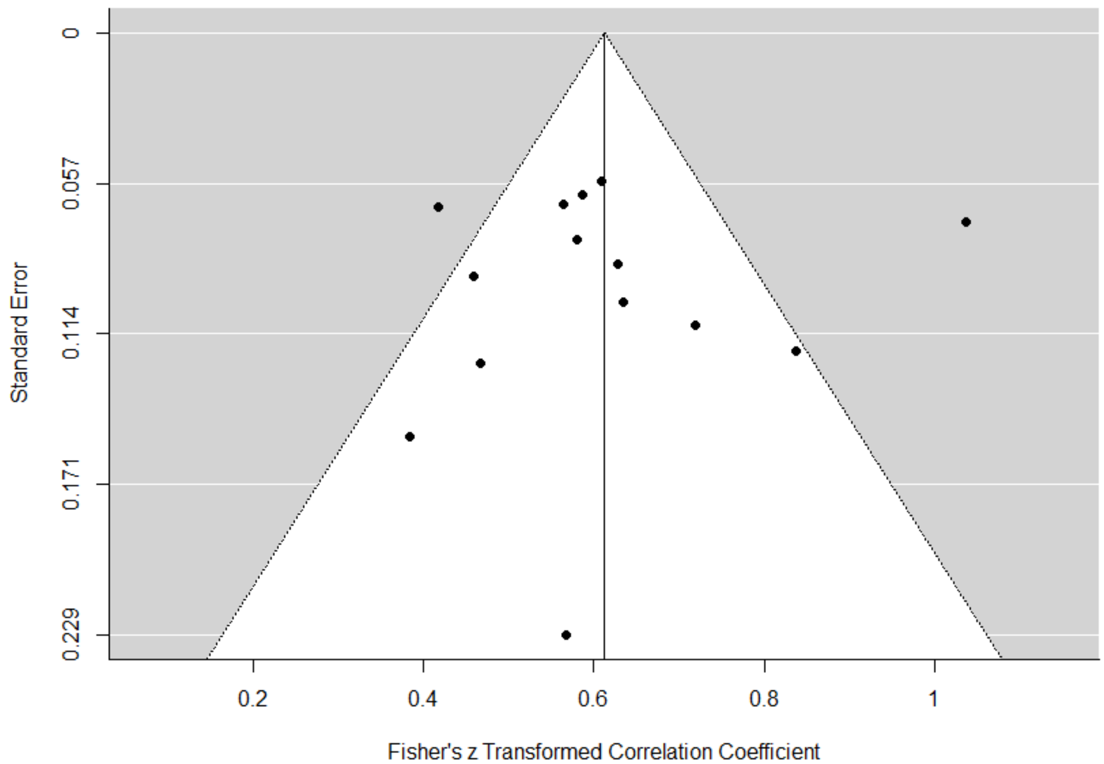


Figure 46

Funnel Plot Random-Effects DA_D using Sample Size

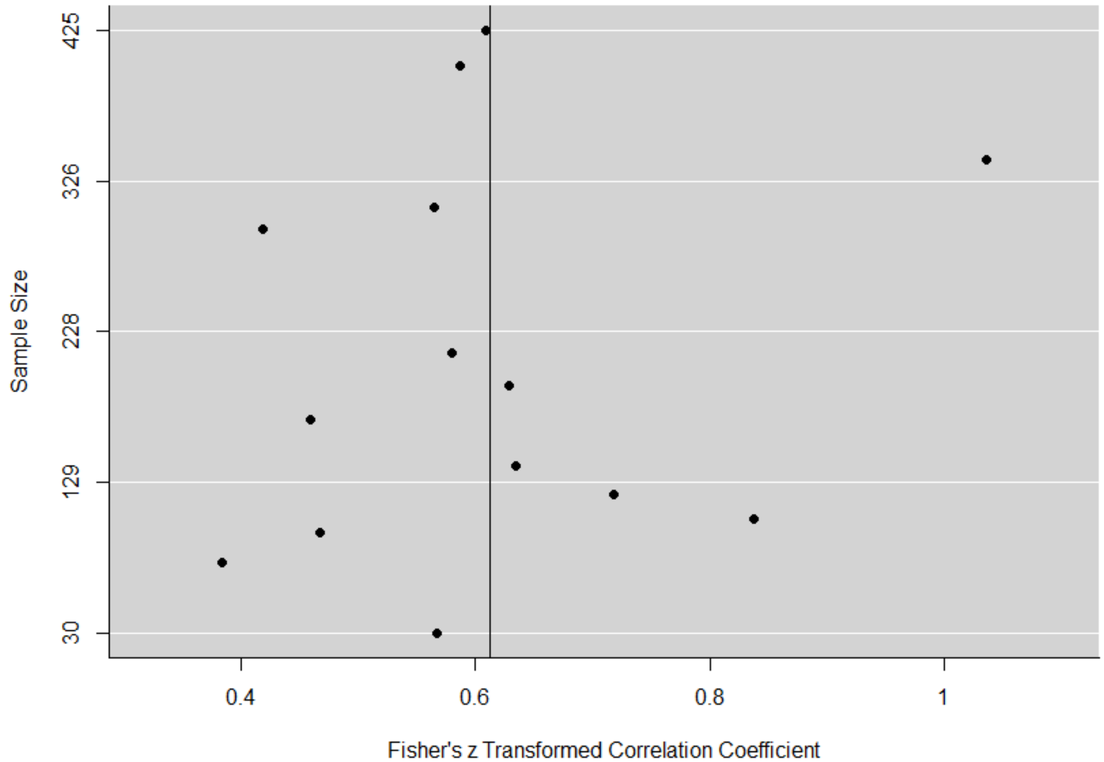
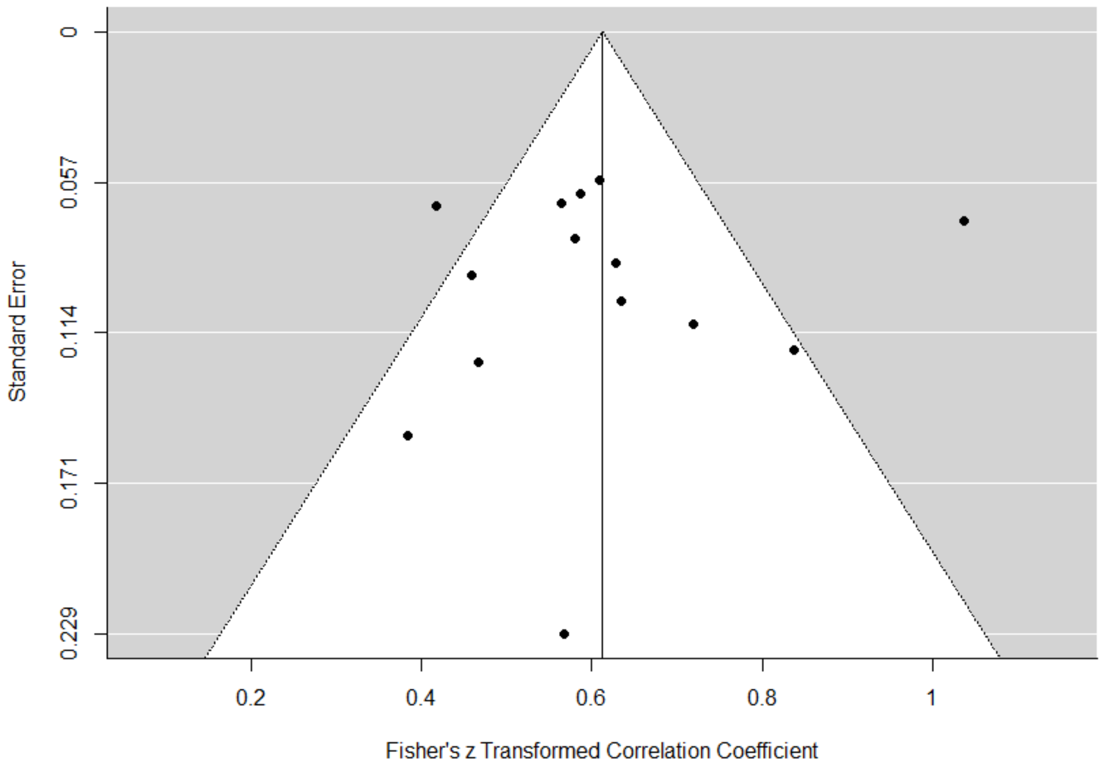


Figure 47

Funnel Plot DA_D after Trim & Fill Random-Effects Model



The estimated number of missing studies on the left is zero. Egger's regression test for funnel plot asymmetry was not significant with $z = -0.40$ and $p = .69$, but this meta-analysis did not have enough studies for this test to detect even severe publication bias (Card, 2012).

Figure 48

Cumulative Meta-Analysis for Random-Effects Model Starting with Largest Study

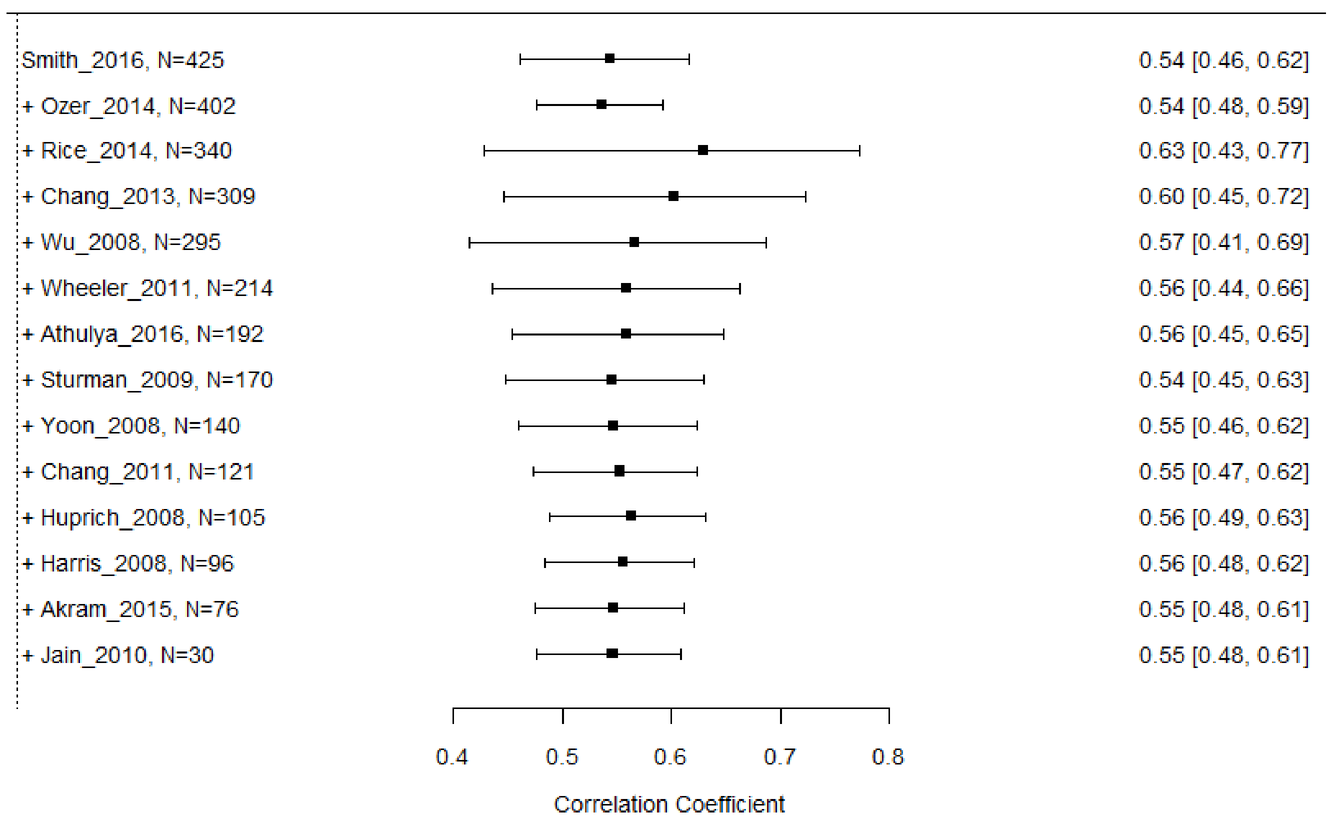
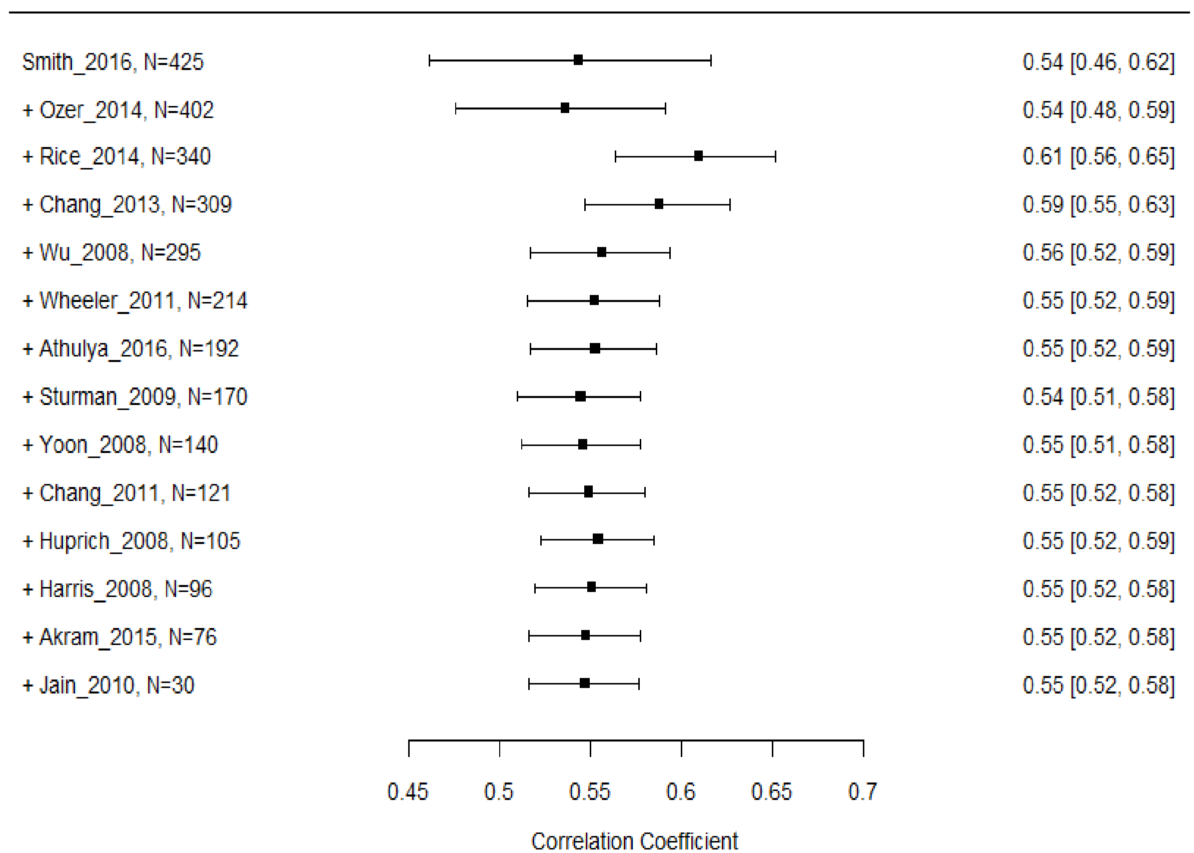


Figure 49

Cumulative Meta-Analysis for Fixed-Effects Model Starting with Largest Study



In Figures 48 and 49 for the random-effects and fixed-effects cumulative meta-analyses, respectively, the absolute value of the estimate of the effect size did not decrease as studies with smaller sample sizes are added one by one to the analysis and the analysis re-run, and this may be evidence against the existence of publication bias in the meta-analysis for the relationship between depression and this perfectionism subscale because there does not appear to be much of a relationship between sample size and effect size.

Eighth Meta-Analysis—FMPS PE Subscale and Depression

The eighth meta-analysis estimated the mean correlation between the FMPS Parental Expectations (PE) subscale and depression. It included a total of only six studies with a total sample size of $N = 1{,}017$. The mean correlation between the FMPS PE subscale and depression was estimated using a random-effects model and restricted maximum likelihood (REML). The correlations from the six individual studies were corrected for attenuation due to measurement error, and the Fisher's z_r variance was also corrected for the uncertainty that correcting the correlations for measurement error introduces (Card, 2012). The estimate of the mean correlation coefficient effect size for the relationship between PE and depression using a random-effects model was $r = .26$, with a 95% CI $[.17, .35]$, almost a medium effect size. The estimate for τ^2, the between-studies variance or total heterogeneity was $T^2 = 0.0055$ and the estimated between-studies standard deviation τ was $T = 0.074$. The result of the test of heterogeneity was $Q(df = 5) = 7.84$, $p < .165$, and $I^2 = 39.07\%$. The test for heterogeneity was not significant, but that was probably because the test had low statistical power because this meta-analysis only had six studies. The I^2 value of about 39% for this meta-analysis is between a small and medium amount of heterogeneity. For this meta-analysis, the approximate 95% credibility interval (*CrI*) was $[.09, .42]$.

Figure 50 provides a forest plot of the six studies used in the meta-analysis that estimated the mean correlation between the FMPS PE subscale and depression with the effect sizes in Fisher's z_r transformed correlation coefficients. Figure 51 provides the same type of plot as Figure 50 except that the effect sizes are raw correlations rather than Fisher's z_r transformed correlation coefficients. Figure 52 provides a funnel plot with the Fisher's z_r transformed correlation coefficients on the x-axis and the standard error on the y-axis, and Figure 53 provides another funnel plot that has Fisher's z transformed correlation coefficients on the x-axis and sample size, instead of the standard error, on the y-axis. Figure 54 provides a funnel plot of the results of the trim and fill method for estimating what the effect size would be if it were corrected for publication bias. Figure 55 provides the forest plot for a cumulative meta-analysis that used a random-effects model, and Figure 56 provides another forest plot for a cumulative meta-analysis but that used a fixed-effects model. All figures for this meta-analysis include all six studies.

Figure 50

Forest Plot PE_D Fisher's Zr Correlations Corrected for Attenuation Random-Effects Model

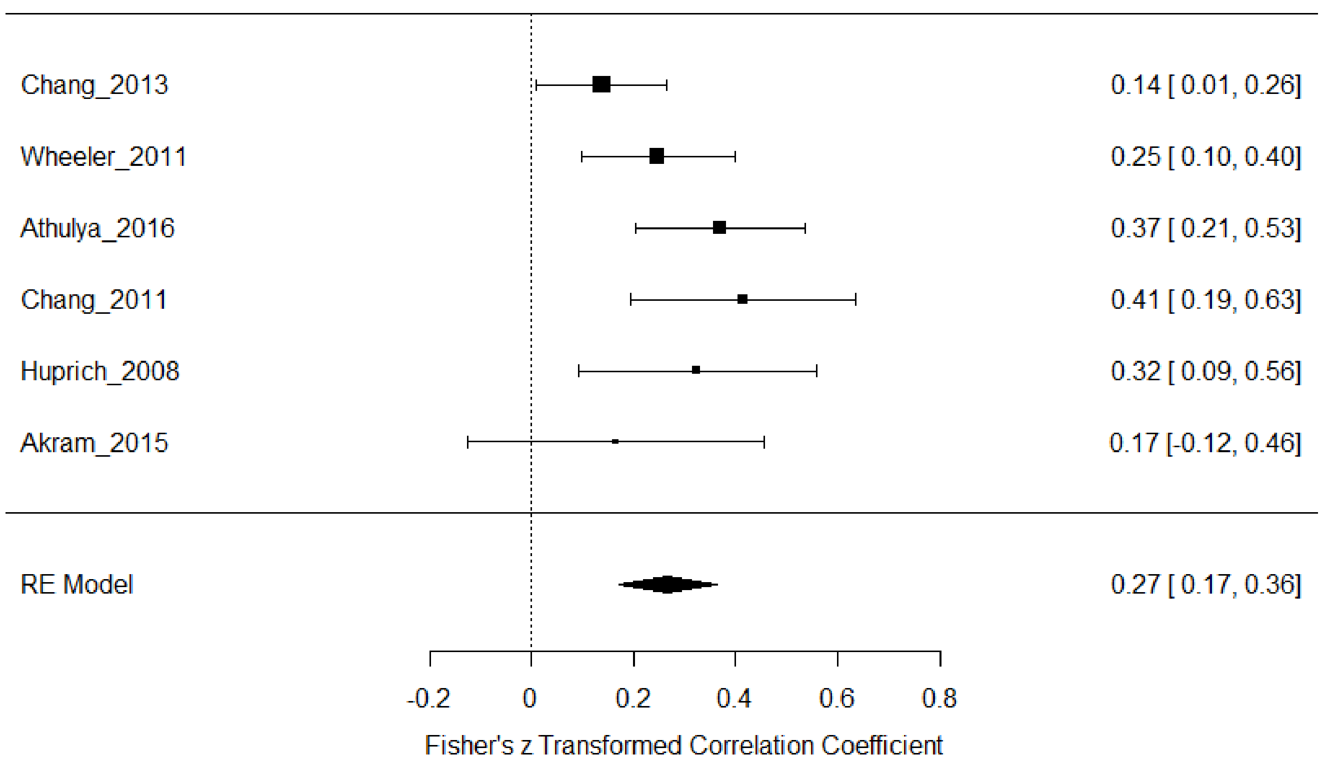


Figure 51

Forest Plot PE_D Correlations Corrected for Attenuation Random-Effects Model

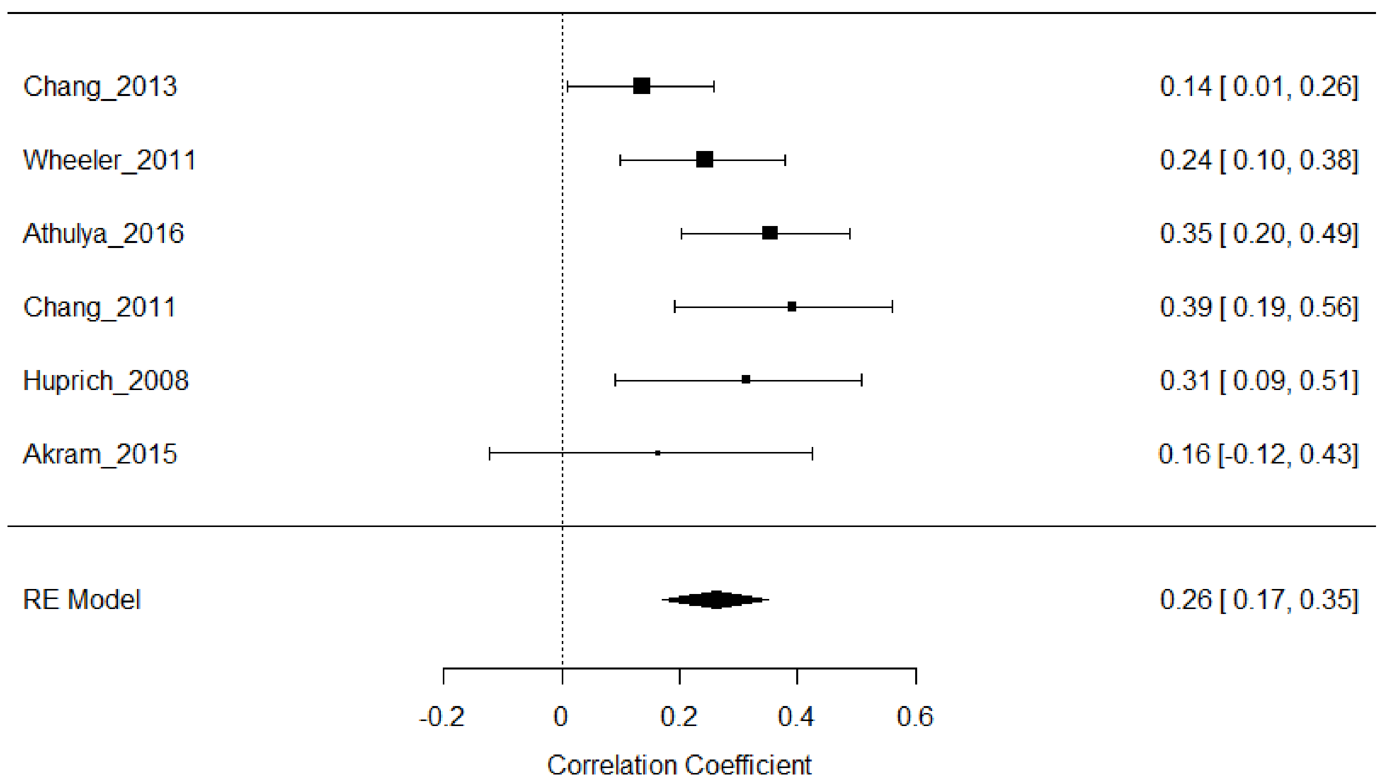


Figure 52

Funnel Plot PE_D Random-Effects using Standard Error

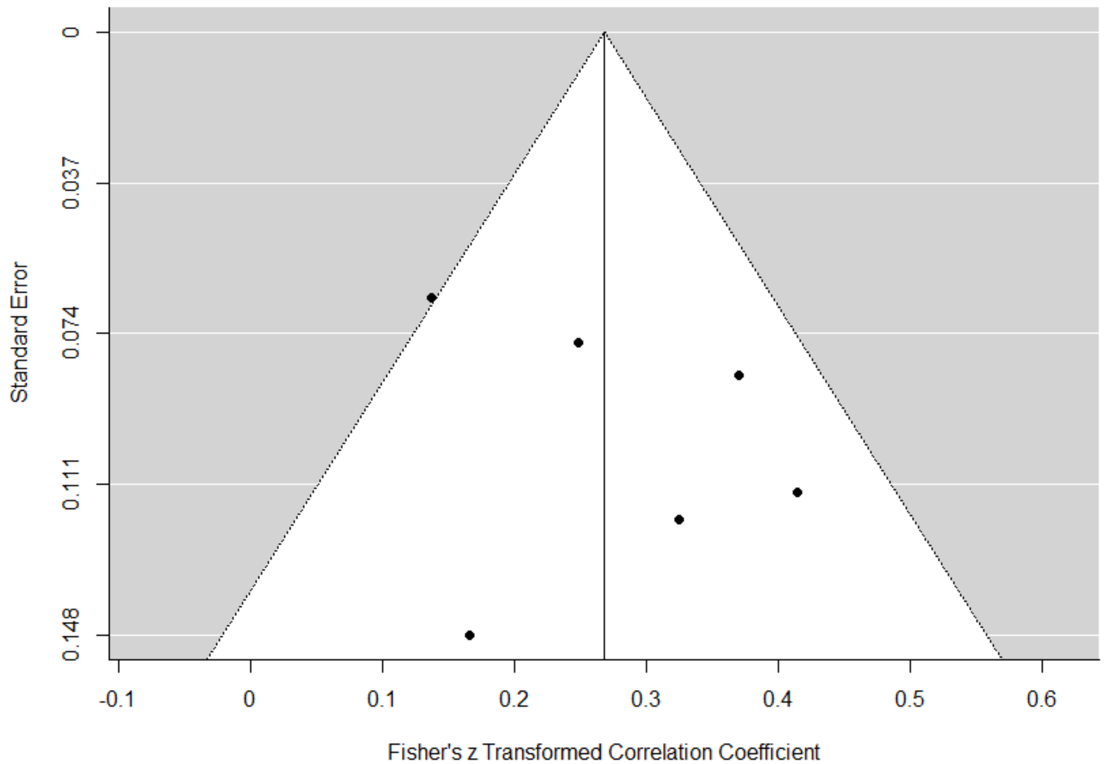


Figure 53

Funnel Plot PE_D Random-Effects using Sample Size

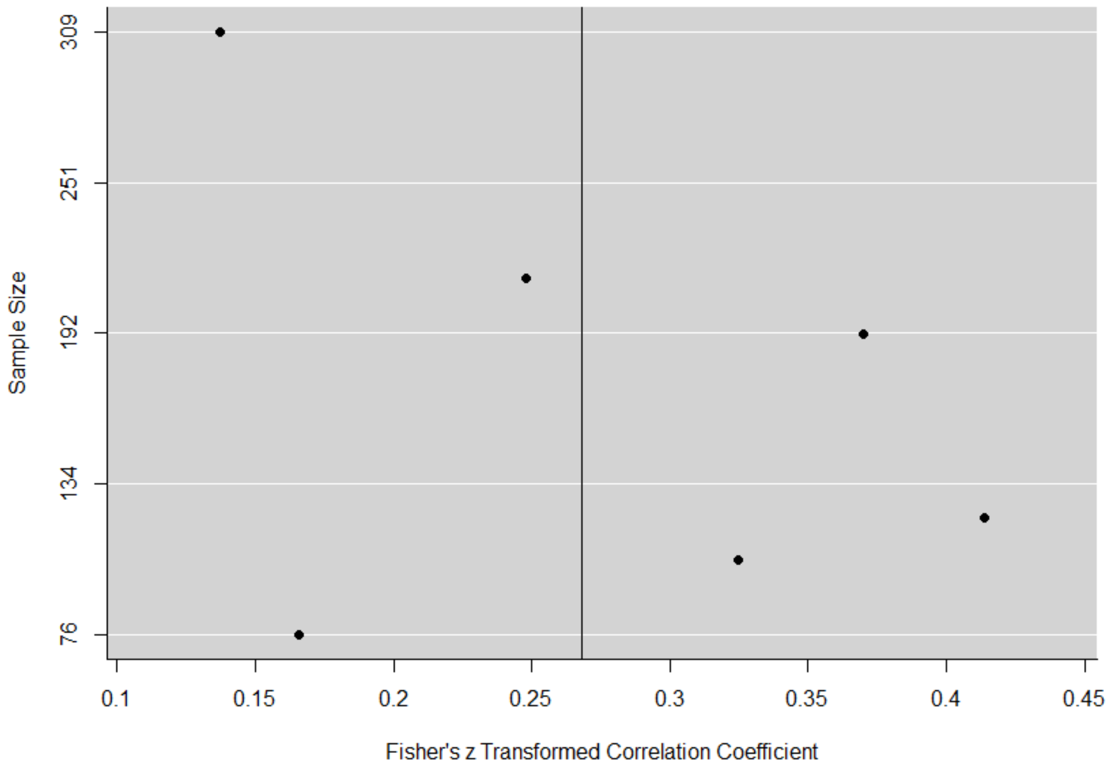
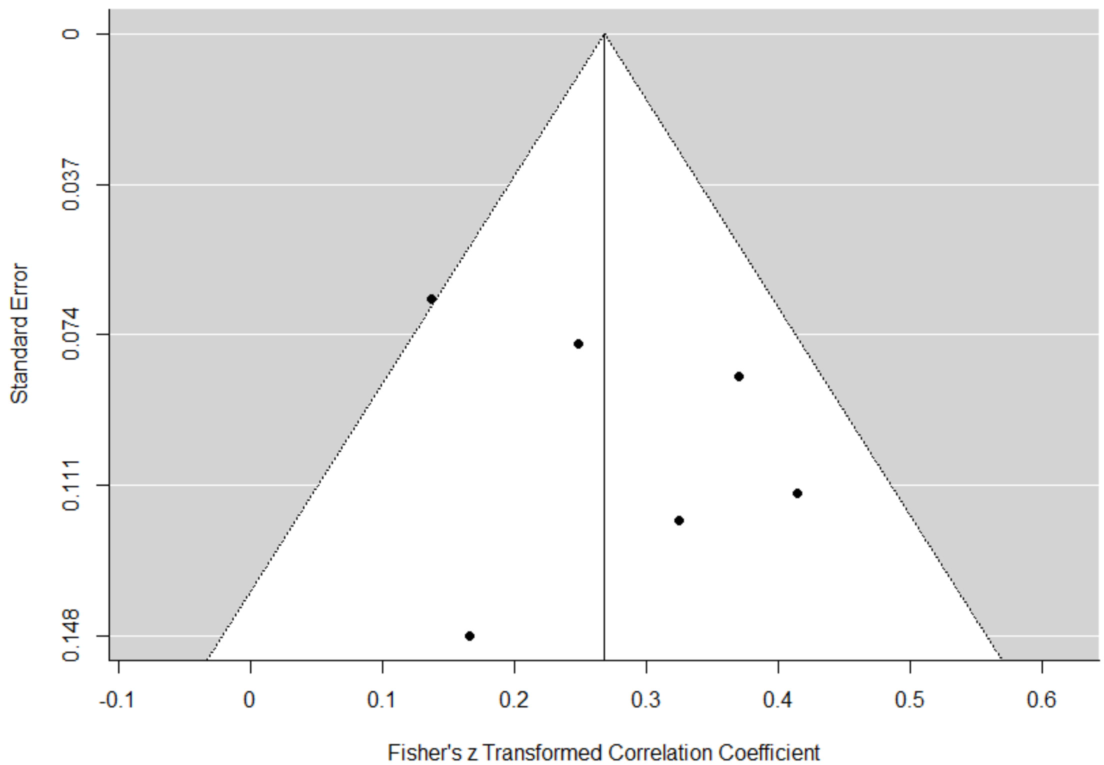


Figure 54

Funnel Plot PE_D after Trim & Fill Random-Effects Model



The estimated number of missing studies on the left side is zero. Egger's regression test for funnel plot asymmetry was not significant with $z = 0.73$ and $p = .47$, but this meta-analysis did not have enough studies for this test to detect even severe publication bias (Card, 2012).

Figure 55

Cumulative Meta-Analysis for Random-Effects Model Starting with Largest Study

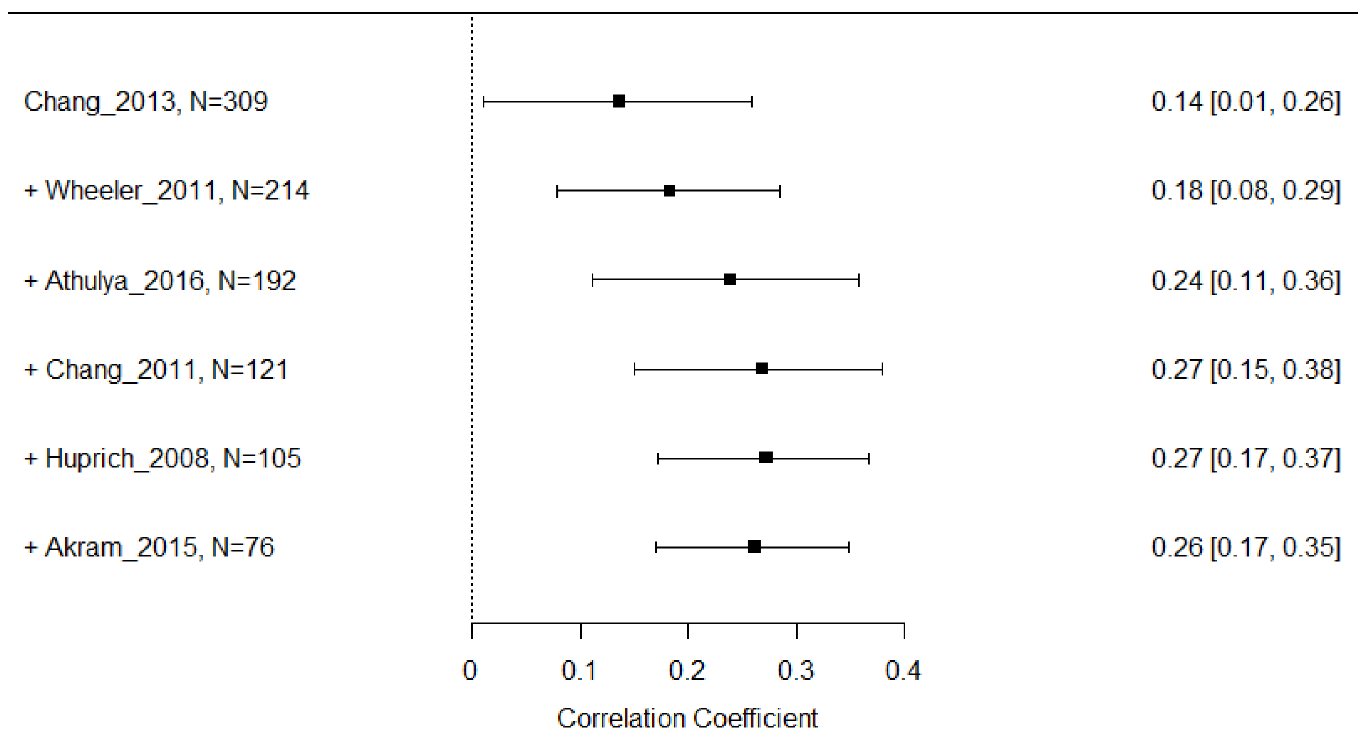
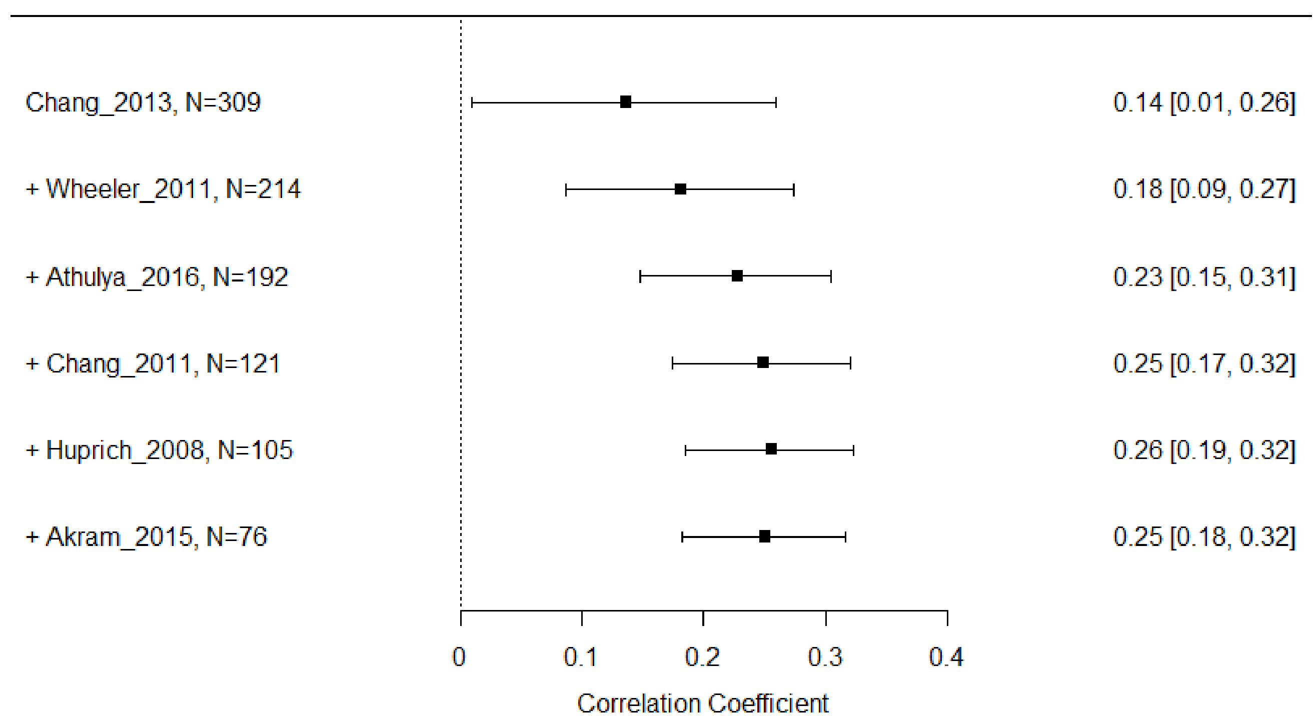


Figure 56

Cumulative Meta-Analysis for Fixed-Effects Model Starting with Largest Study



In Figures 55 and 56 above for the random-effects and fixed-effects cumulative meta-analyses, respectively, the absolute value of the correlation increased as studies with increasingly smaller sample sizes were added one by one and the analysis re-run. Also, the estimate of the mean correlation was slightly greater in the random-effects cumulative meta-analysis ($r = .26$) than in the fixed-effect cumulative meta-analysis ($r = .25$). The estimate for the mean correlation might be greater in the random-effects cumulative meta-analysis because the estimate of the effect size from the random-effects model is not very accurate because it is based on only six studies. Borenstein et al. (2009) explained the following about the effect of having only a small number of studies when using a random-effects model:

> Unlike the fixed-effect analysis, where the estimate of the error is based on sampling theory (and therefore reliable), in a random-effects analysis, our estimate of the error may itself be unreliable. Specifically, when based on a small number of studies, the estimate of the between-studies variance (T^2), may be substantially in error. (p. 363)

And poorly estimated between-studies variance affects all aspects of the random-effects analysis (Borenstein et al., 2009). Also, there appears to be a relationship between sample size and effect size that could be evidence of publication bias.

Ninth Meta-Analysis—FMPS PC and Depression

The ninth meta-analysis estimated the mean correlation between the FMPS Parental Criticism (PC) subscale and depression. It included a total of only eight studies with a total sample size of $N = 1{,}187$. The mean correlation between the FMPS PC subscale and depression was estimated using a random-effects model and restricted maximum likelihood (REML). The correlations from the eight individual studies were corrected for attenuation due to measurement error, and the Fisher's z_r variance was also corrected for the uncertainty that correcting the correlations for measurement error introduces (Card, 2012). The estimate of the mean correlation coefficient effect size for the relationship between PC and depression using a random-effects model was $r = .40$, with a 95% CI $[.31, .49]$, a medium effect size. The estimate for τ^2, the between-studies variance or total heterogeneity was $T^2 = 0.0127$ and the estimated between-studies standard deviation τ was $T = 0.1127$. The result of the test of heterogeneity was $Q(df = 7) = 15.11$, $p < .035$, and $I^2 = 56.39\%$. The I^2 value of about 56% for this meta-analysis is a medium amount of heterogeneity. For this meta-analysis, the approximate 95% credibility interval (*CrI*) was $[.18, .59]$.

Figure 57 provides a forest plot of the eight studies used in the meta-analysis that estimated the mean correlation between the FMPS PC subscale and depression with the effect sizes in Fisher's z_r transformed correlation coefficients. Figure 58 provides the same type of plot as Figure 57 except that the effect sizes are raw correlations rather than Fisher's z_r transformed correlation coefficients. Figure 59 provides a funnel plot with the Fisher's z_r transformed correlation coefficients on the x-axis and the standard error on the y-axis, and Figure 60 provides another funnel plot that has Fisher's z transformed correlation coefficients on the x-axis and sample size, instead of the standard error, on the y-axis. Figure 61 provides a funnel plot of the results of the trim and fill method for estimating what the effect size would be if it were corrected for publication bias. Figure 62 provides the forest plot for a cumulative meta-analysis using a random-effects model, and Figure 63 provides another forest plot for a cumulative meta-analysis but using a fixed-effects model. All figures for this meta-analysis include all eight studies.

Figure 57

Forest Plot PC_D Fisher's Zr Correlations Corrected for Attenuation Random-Effects Model

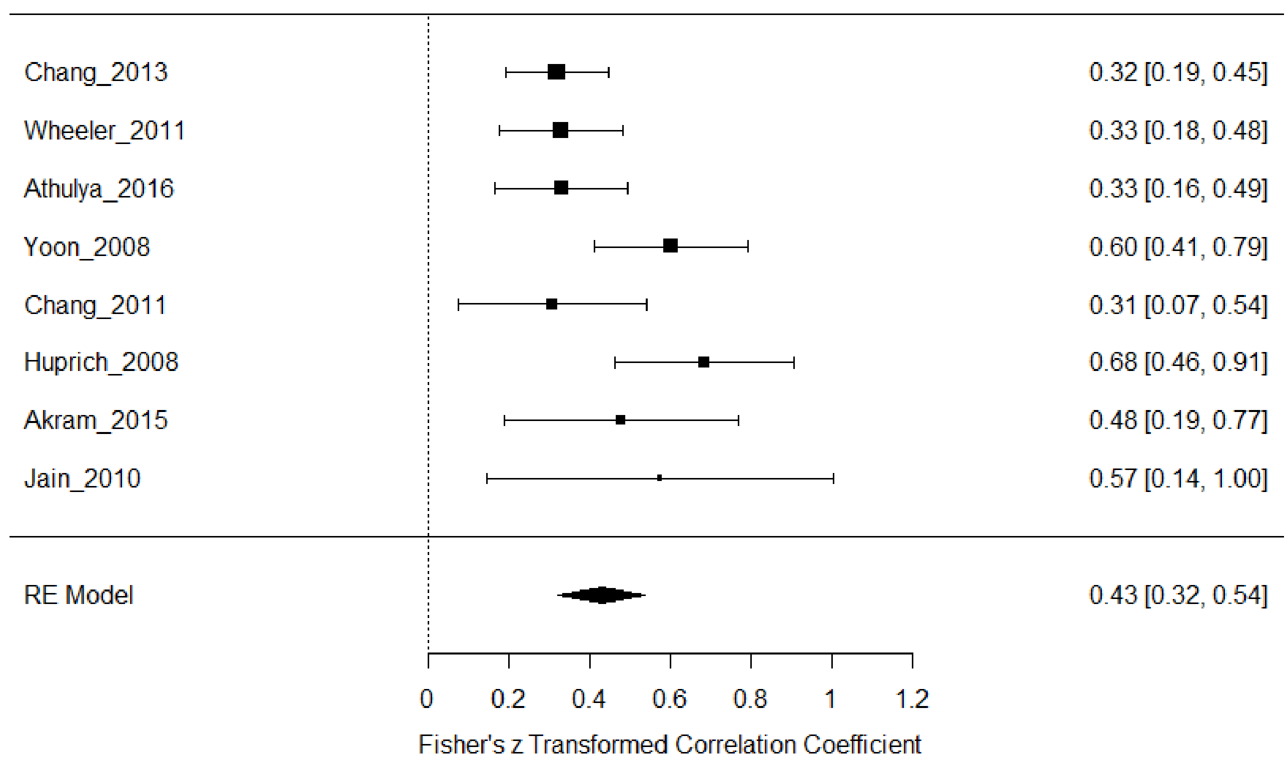


Figure 58

Forest Plot PC_D Correlations Corrected for Attenuation Random-Effects Model

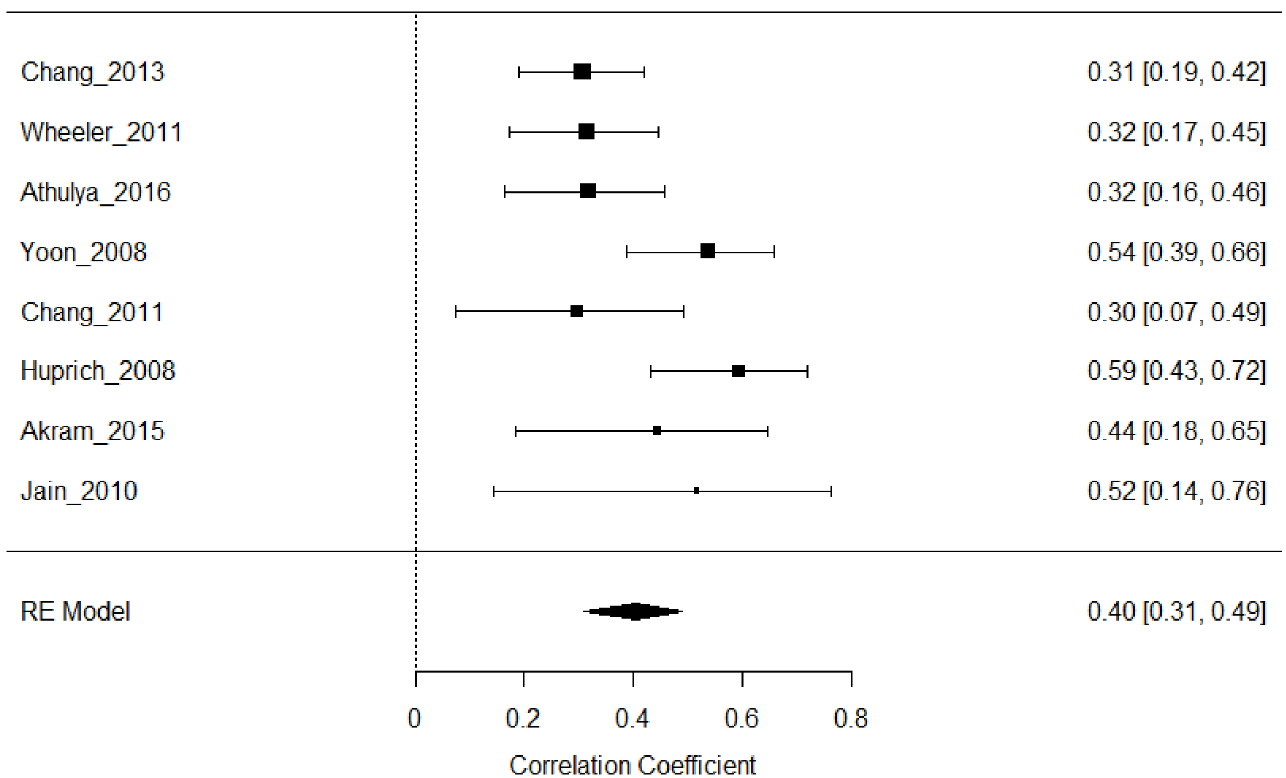


Figure 59

Funnel Plot PC_D Random-Effects using Standard Error

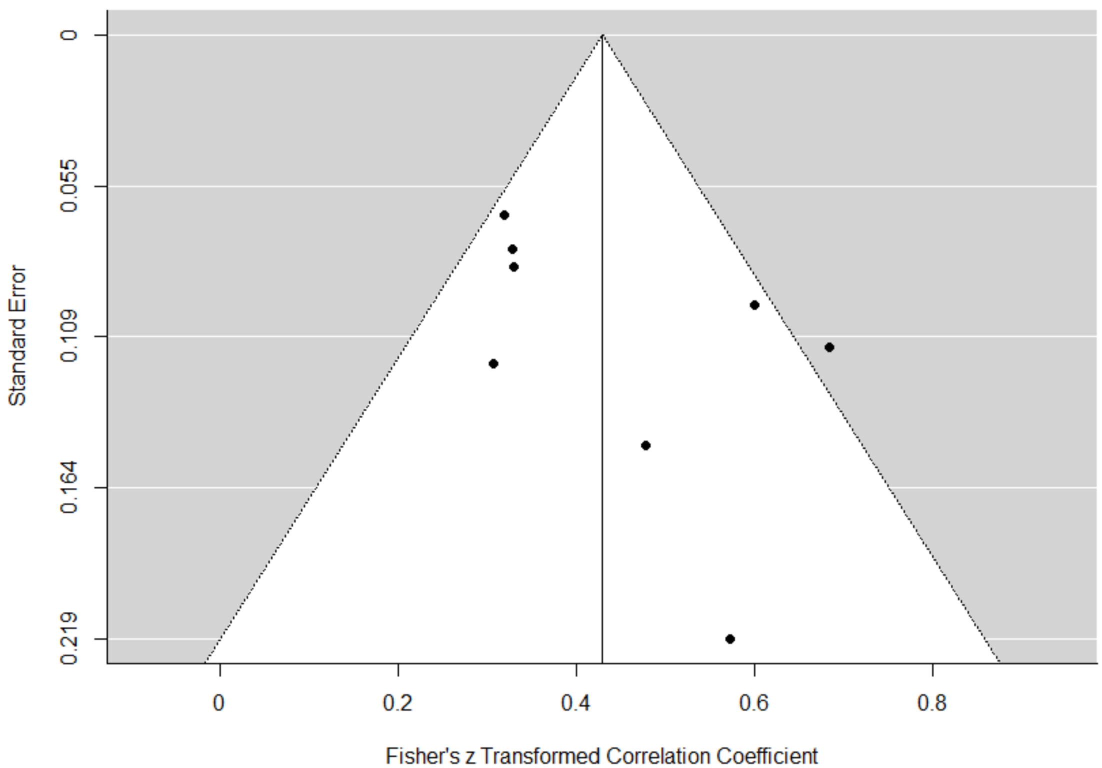


Figure 60

Funnel Plot PC_D Random-Effects using Sample Size

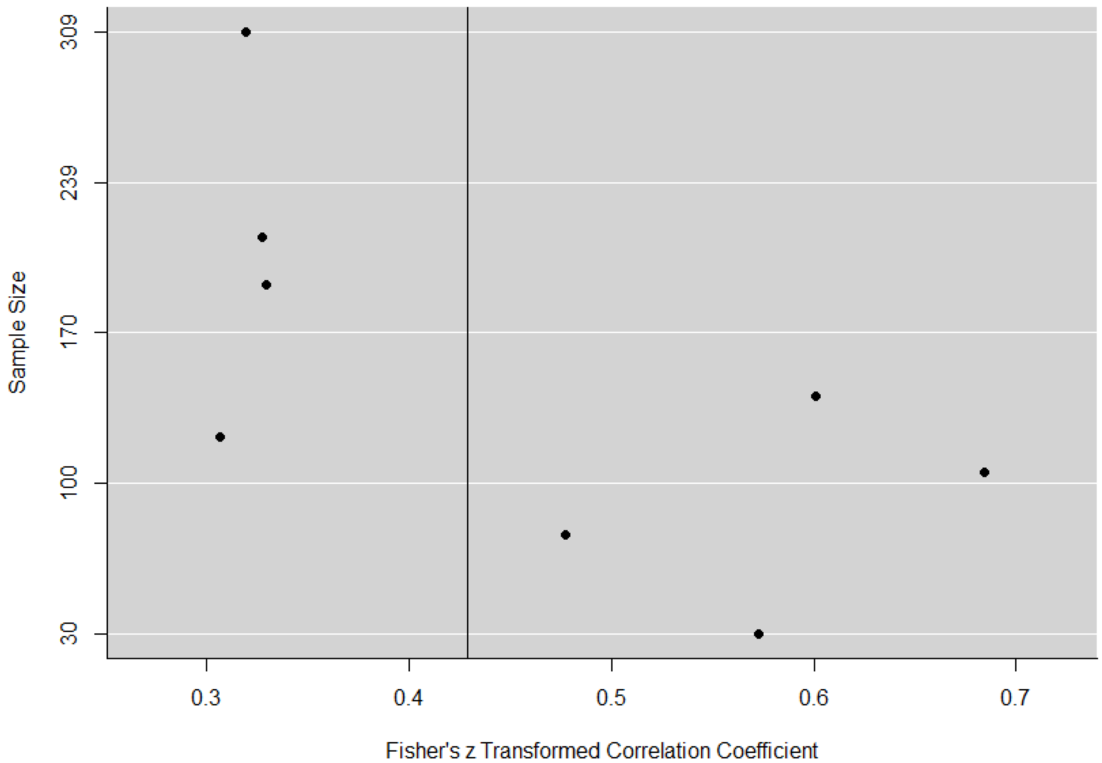
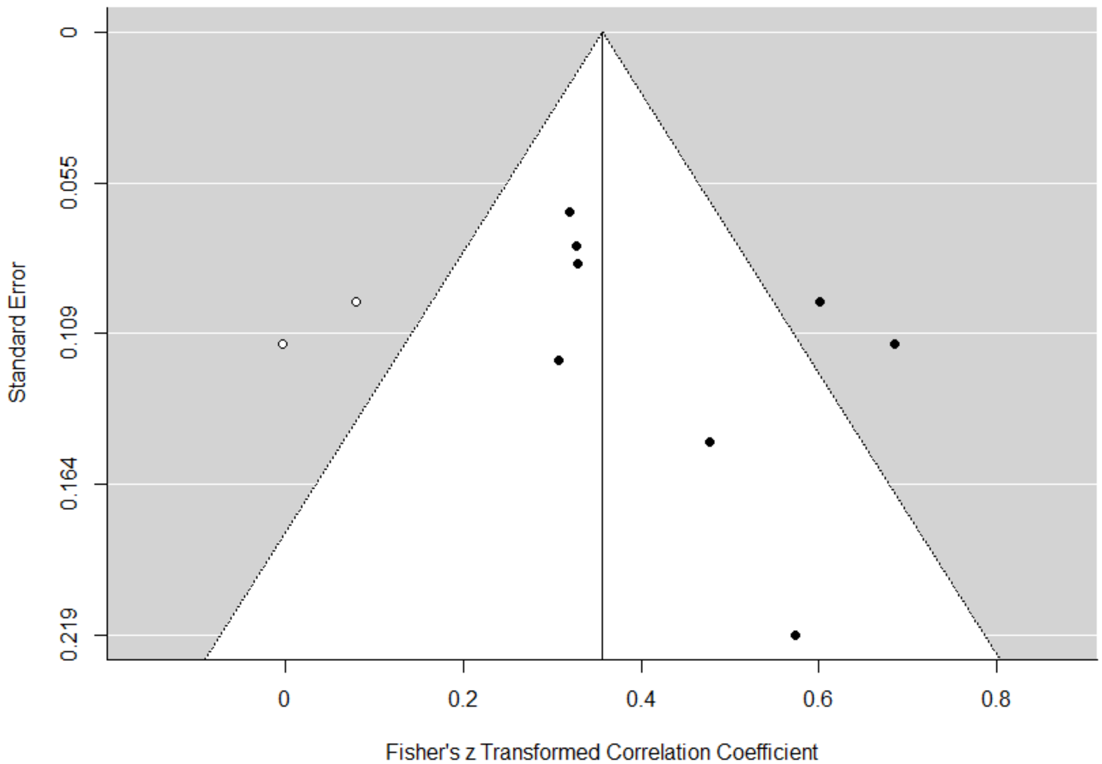


Figure 61

Funnel Plot PC_D after Trim & Fill Random-Effects Model



From the trim and fill analysis in Figure 61, two studies were estimated as missing from the left side. After the trim and fill procedure, “If the asymmetry is due to bias” (Borenstein et al., 2009, p. 291) there were 10 studies, and the estimate of the mean correlation coefficient effect size for the relationship between Parental Criticism and depression using a random-effects model was $r = .34$, with a 95% CI $[.22, .46]$, a medium effect size. The estimate for τ^2, the between-studies variance or total heterogeneity was $T^2 = 0.0344$ and the estimated between-studies standard deviation τ was $T = 0.1854$. The result of the test of heterogeneity was $Q(df = 9) = 34.81, p < .001$, and $I^2 = 77.55\%$. The I^2 value of about 76% for this meta-analysis is a large amount of heterogeneity. For this meta-analysis, the approximate 95% credibility interval (*CrI*) was $[-.03, .63]$.

Egger's regression test for funnel plot asymmetry was not significant with $z = 1.38$ and $p = .17$, but this meta-analysis did not have enough studies for this statistical test to detect even severe publication bias (Card, 2012).

Figure 62

Cumulative Meta-Analysis for Random-Effects Model Starting with Largest Study

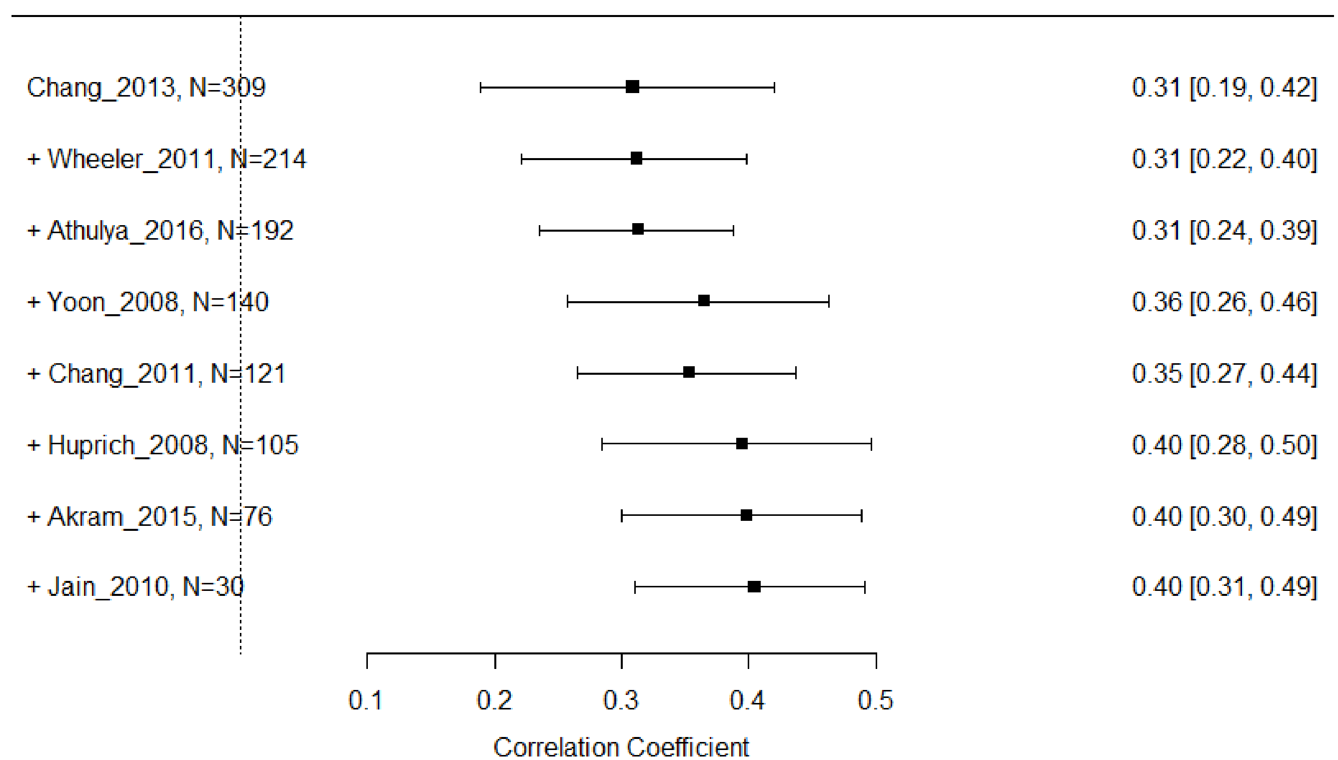
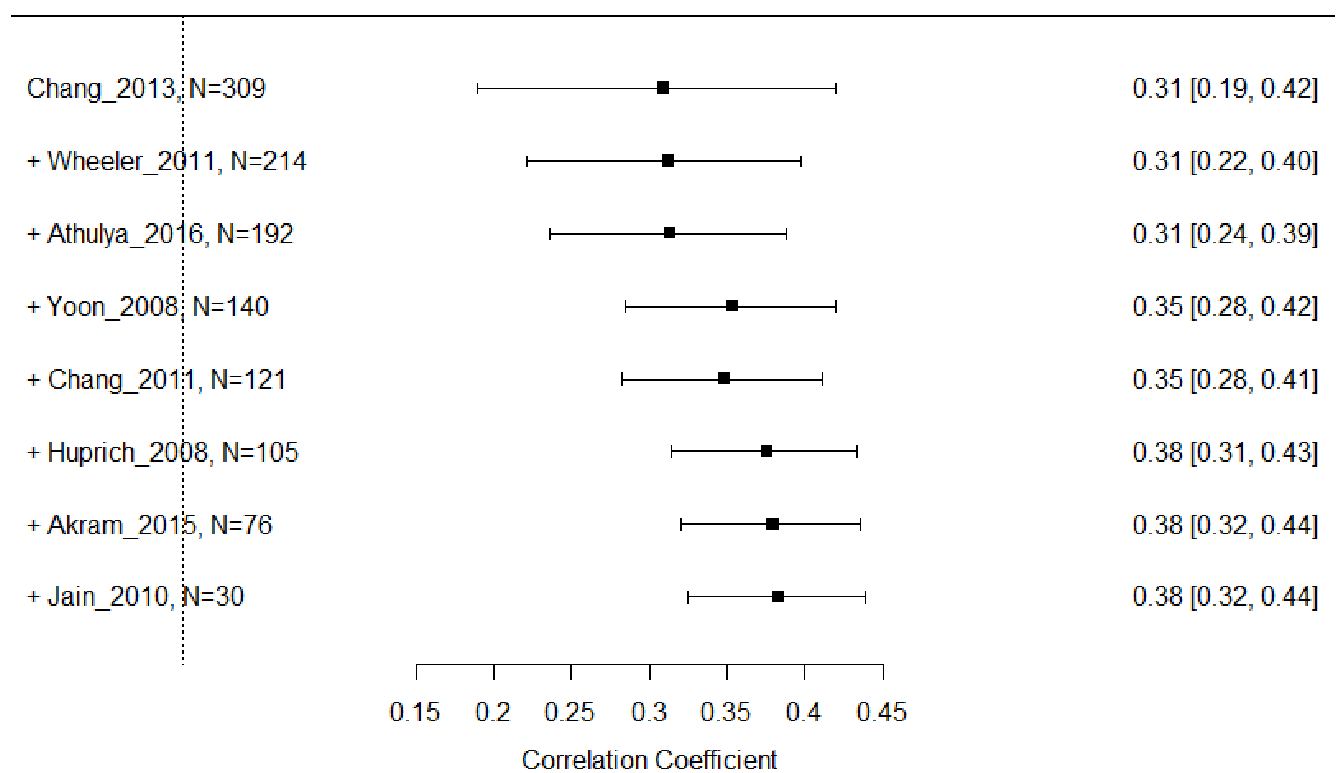


Figure 63

Cumulative Meta-Analysis for Fixed-Effects Model Starting with Largest Study



The cumulative meta-analyses in Figures 62 and 63 show the effect size actually getting larger as studies with increasingly smaller sample sizes are added and the analysis is re-run, and this might be evidence that there was a relationship between sample size and effect size.

As in the previous meta-analysis (the eighth meta-analysis between FMPS PE and depression), the random-effects and fixed-effects cumulative meta-analyses in Figures 62 and 63 show the random-effects analysis as having a slightly greater mean correlation ($r = .40$) than the fixed-effects analysis ($r = .38$), and the absolute value of the effect size increases as studies with increasingly smaller sample sizes are added one by one and the analysis is re-run. The estimate of the effect size being larger for the random-effects analysis than for the fixed-effects analysis is probably, again, the result of there being too few studies (only eight studies) so that the between-studies variance is poorly estimated which throws off the results for the whole random-effects analysis. Table 4 provides a summary of the results of the nine meta-analyses.

Table 4

Summary of Results from the Nine Meta-Analyses

Meta-Analysis with all studies	Mean Correlation with Depression	Perfectionistic Strivings (PS) or Perfectionisti	95% *CI*	Number of Studies	Total N	T^2	95% *CrI*	I^2	Evidence of Publication Bias?
1st APS-R HS	-.08	PS	[-.14, -.01]	12	3,678	.0088	[-.27, .12]	66%	Yes
2nd APS-R Discrepancy	.56	PC	[.51, .60]	15	4,708	.0096	[.40, .68]	70%	No*
3rd HMPS SOP	.17	PS	[.11, .22]	25	5,581	.012	[-.05, .37]	66%	Yes
4th HMPS SPP	.45	PC	[.41, .49]	26	5,637	.0086	[.29, .59]	58%	No*
5th FMPS PS	.08	PS	[.03, .14]	17	3,781	.0064	[-.08, .24]	50%	Yes
6th FMPS CM	.46	PC	[.41, .52]	16	3,034	.0106	[.28, .61]	60%	No*
7th FMPS DA	.55	PC	[.48, .61]	14	2,915	.0225	[.29, .73]	76%	No*
8th FMPS PE	.26	PC	[.17, .35]	6	1,017	.0055	[.09, .42]	39%	Yes
9th FMPS PC	.40	PC	[.31, .49]	8	1,187	.0127	[.18, .59]	56%	Yes

For Table 4 (above) the results for the subscales that form the Perfectionistic Strivings higher-order factor are shaded grey and results from the subscales that form the Perfectionistic Concerns higher order factor are not shaded, and the asterisk (*) indicates that the meta-analysis did not have enough studies for there to be adequate power to detect moderate publication bias with Egger's Regression Test for Funnel Plot Asymmetry (Card, 2012)

Summary of Answers to Research Questions

The research questions that guided this study were:

1) Does the pattern of correlations for the association of depression with Perfectionistic Strivings (PS) and Perfectionistic Concerns (PC) dimensions of perfectionism differ enough to give evidence that these two types of perfectionism are distinct constructs?
    a) Are all the dimensions of Perfectionistic Concerns (PC) positively and significantly correlated with depression?

For the six dimensions of Perfectionistic Concerns perfectionism, the correlations with depression were $r = .56$ for the APS-R Discrepancy subscale, $r = .55$ for the FMPS Doubts about Actions subscale, $r = .46$ for the FMPS Concern over Mistakes subscale, $r = .45$ for the HMPS Socially Prescribed Perfectionism subscale, and $r = .40$ for the FMPS Parental Criticism subscale, and $r = .26$ for the FMPS Parental Expectations subscale. All correlations were significantly and positively correlated with depression.

    b) Are all the dimensions of Perfectionistic Strivings either not significantly correlated with depression or inversely correlated with depression?

The APS-R High Standards subscale was significantly inversely correlated with depression with $r = -.08$. However, the FMPS Personal Standards subscale was significantly positively correlated with depression with $r = .08$, and the HMPS Self-Oriented Perfectionism subscale was also significantly positively correlated with depression with $r = .17$.

2) Are the two possibly opposite types of perfectionism differentially related to depression?

   a) How strong is the association between the negative (maladaptive) dimensions of perfectionism that comprise Perfectionistic Concerns (PC) and severity of depression?

The correlations between the negative dimensions of perfectionism ranged from $r = .26$ to $r = .56$. All the correlations were significant, but the magnitude varied from small to large in size.

      i) Which of the Perfectionistic Concerns (PC) dimensions of perfectionism is most strongly associated with depression?

The APS-R Discrepancy dimension of Perfectionistic Concerns had the strongest positive correlation with depression with $r = .56$.

      ii) Are the associations between the Perfectionistic Concerns (PC) dimensions of perfectionism and depression stronger for women than for men?

It could not be determined if the correlations between the Perfectionistic Concerns dimensions of perfectionism were more strongly correlated with depression for women than for men because there were not enough studies with only women and only men to conduct a moderator analysis.

iii) As the research on perfectionism and depression indicates, are the Discrepancy subscale from the APS-R and the Socially Prescribed Perfectionism (SPP) subscale from the HMPS the two dimensions of Perfectionistic Concerns (PC) that are most strongly and positively associated with depression?

The APS-R Discrepancy subscale was the dimension of Perfectionistic Concerns that was most strongly and positively correlated with depression with $r = .56$. However, the HMPS Socially-Prescribed Perfectionism (SPP) subscale was not among the two Perfectionistic Concerns dimensions most strongly and positively correlated with depression because the correlation between SPP and depression was $r = .45$, whereas the correlation between depression and the FMPS Doubts about Actions was $r = .55$, and the correlation between depression and FMPS Concern over Mistakes was $r = .46$.

b) How strong is the association between the positive (adaptive) dimensions of perfectionism that comprise Perfectionistic Strivings (PS) and depression?

The correlation between the APS-R High Standards subscale and depression was $r = -.08$. The correlation between the FMPS Personal Standards subscale and depression was $r = .08$. The correlation between the HMPS Self-Oriented Perfectionism subscale and depression was $r = .17$. While all of these correlations were significant, they were varied in direction and were substantially lower in magnitude than were the correlations with the negative dimensions of perfectionism.

i) Are any of the Perfectionistic Strivings (PS) dimensions of perfectionism significantly positively correlated with depression?

The HMPS Self-Oriented subscale was significantly positively correlated with depression with $r = .17$. The FMPS Personal Standards subscale was significantly positively correlated with depression with $r = .08$.

ii) Are any of the Perfectionistic Strivings (PS) dimensions of perfectionism significantly *negatively* correlated with depression?

The APS-R High Standards subscale was significantly negatively correlated with depression with $r =. -08$.

c) Are the negative dimensions of perfectionism that comprise Perfectionistic Concerns (PC) perfectionism more strongly associated with severity of depression than the positive or neutral dimensions of perfectionism that comprise Perfectionistic Strivings (PS) perfectionism?

The negative dimensions of perfectionism that comprise Perfectionistic Concerns had stronger positive correlations with severity of depression than did the positive or neutral dimensions of perfectionism that comprise Perfectionistic Strivings.

Chapter 4: Discussion

Discussion of Answers to Research Questions

The pattern of correlations found in this study for the association of depression with Perfectionistic Strivings (PS) and Perfectionistic Concerns (PC) dimensions of perfectionism differed and offered evidence that these two types of perfectionism are distinct constructs. All of the Perfectionistic Concerns dimensions of perfectionism were directly and significantly correlated with depression, and they were more strongly correlated with depression than were the Perfectionistic Strivings dimensions of perfectionism. From the Perfectionistic Concerns group, the Discrepancy and Doubts about Actions dimensions had large correlations, and the Concern over Mistakes, Socially-Prescribed Perfectionism, and Parental Criticism subscales had at least medium correlations, and only the Parental Expectations subscale had less than a medium correlation (with between a small and medium correlation). The Discrepancy subscale from the Perfectionistic Concerns group had the strongest positive correlation with depression.

The results of this study support the findings in the literature on depression and perfectionism that the Perfectionistic Concerns dimensions of perfectionism are maladaptive because they have significant and usually at least moderate correlations with depression (Dunkley, Blankstein et al., 2006; Lo & Abbott, 2013). Also, the idea that the APS-R Discrepancy subscale is a measure of purely maladaptive perfectionism that can distinguish between positive or healthy perfectionism and negative or unhealthy perfectionism was supported by the results of this study since Discrepancy had the strongest positive correlation with depression (Flett & Hewitt, 2002). Socially-Prescribed Perfectionism was also said to be a defining aspect of negative perfectionism, but in this study not only Discrepancy from the APS-R, but also the Concern over Mistakes and Doubts about Actions subscales from the FMPS had stronger positive correlations with depression than did SPP. Contrary to some research (Blankstein & Dunkley, 2002; Hill, McIntire, & Bacharach, 1997), socially-prescribed perfectionism does not appear to be the most maladaptive aspect of perfectionism. The High Standards subscale from the APS-R had the only negative correlation with depression, and the Discrepancy subscale, also from the APS-R, had the strongest positive correlation with depression. This finding supports the literature about the APS-R where these two subscales are said to be independent of each other and are said to measure and distinguish between two opposite types of perfectionism (Flett & Hewitt, 2002; Slaney et al., 2002). The results for the High Standards and Discrepancy subscales from the APS-R support the finding in the literature that positive perfectionism is not correlated with negative psychological characteristics when the negative dimensions of perfectionism are either absent in the perfectionistic person or are statistically controlled for because Discrepancy and High Standard were further shown to be independent (Slaney et al., 2002). This was because Discrepancy had the strongest positive correlation with depression and High Standards was the only subscale that had a negative correlation with depression.

There were only four studies with only female participants and there was only one study with correlations for only male participants, so it was not possible to do a moderator analysis comparing the correlations between depression and perfectionism dimensions for males and females to see if females had stronger direct correlations between perfectionism and depression.

Even though the correlations between the Perfectionistic Concerns dimensions of perfectionism were more strongly positively related to depression than the Perfectionistic Strivings dimensions of perfectionism, only the APS-R High Standards subscale from the Perfectionistic Strivings group had a significant negative correlation with depression, and that correlation did not even reach the level of a small correlation. Also, the two other dimensions of Perfectionistic Strivings, the FMPS High Standard subscale and the HMPS Self-Oriented Perfectionism subscale, had significant positive correlations with depression, and the SOP correlation with depression was between small and medium. These findings indicate that contrary to some of the literature on positive perfectionism, the Perfectionistic Strivings dimensions of perfectionism are not entirely adaptive (Frost et al., 1993; Lo & Abbott, 2013; Stoeber & Otto, 2006). Some perfectionism researchers believe that there are two opposite types of perfectionism, one healthy and the other unhealthy (Slaney et al., 2001; Stoeber & Otto, 2006), but other perfectionism researchers believe that perfectionism is only a negative and unhealthy characteristic (Shafran et al., 2002). The fact that only one of the nine total dimensions of perfectionism investigated in this study had a negative relationship with depression and that the other eight dimensions had significant positive correlations with depression is evidence that perfectionism might be mainly a maladaptive or unhealthy trait. This lends credibility to Shafran et al.'s (2002) claim that perfectionism is better conceptualized as a unidimensional construct rather than a multidimensional construct. With all but one of the nine dimensions of perfectionism having a positive correlation with depression, it is likely that perfectionism is unidimensional and is a mostly maladaptive character trait. If there is a healthy type of perfectionism, it would be best measured by the APS-R High Standards subscale, which was the only subscale that had a negative correlation with depression. Perfectionistic Strivings perfectionism might only mean that a person has high standards and it might not truly be a form of perfectionism. Whether or not Perfectionistic Strivings is a form of perfectionism, it might be best measured with only the High Standards subscale of the APS-R because that was the only subscale that had a negative correlation with depression. The Personal Standards subscale from the FMPS and the Self-Oriented Perfectionism subscale of the HMPS might not be measures of a healthy type of perfectionism because those two subscales had significant positive correlations with depression. They might be two of a total of eight subscales that measure unhealthy or maladaptive perfectionism. One of the nine subscales might just measure high standards rather than perfectionism because people could have high standards without being perfectionists. The other eight dimensions might all be measures of unhealthy perfectionism, and perfectionism might be unidimensional, and it might be an inherently unhealthy or maladaptive personality trait. However, this conclusion might be partly the result of the subscales or dimensions of perfectionism that were used in this study because the different ways that perfectionism is measured affects the empirical results of perfectionism research (Sirois & Molnar, 2016). Rather than it being the case that that positive perfectionism is mainly associated with positive psychological characteristics and outcomes when the negative aspects of perfectionism are controlled for statistically or are absent (Stoeber & Otto, 2006), there might not be a positive or healthy type of perfectionism.

Maybe the reason that the APS-R High Standards subscale was the only perfectionism subscale that was negatively correlated with depression is because the other two subscales that are a part of the supposedly positive Perfectionistic Strivings Perfectionism were not created to theoretically be independent of the other subscales in their respective multidimensional measures of perfectionism (Frost et al., 1990; Hewitt & Flett, 1991b).

Since the APS-R High Standards subscale is theoretically independent from the APS-R Discrepancy (Slaney et al., 2002), the High Standards is a case where negative perfectionism is statistically controlled for and positive perfectionism is supposed to be correlated with healthy characteristics and outcomes when negative perfectionism is statistically controlled for or absent in the perfectionistic person (Blankstein & Dunkley, 2002; Stoeber & Otto, 2006), and the High Standards subscale was the only perfectionism subscale that had a negative correlation with depression.

In this study the perfectionism subscales were referred to as *dimensions*, but a confirmatory factor analysis would be necessary to determine how many different dimensions of perfectionism there actually are.

In conclusion, even though the Perfectionistic Concerns dimensions of perfectionism had a stronger direct correlation with depression than did the Perfectionistic Strivings dimensions of perfectionism, the fact that all but one of the Perfectionistic Strivings dimensions of perfectionism were also significantly positively correlated with depression indicates that, in general, perfectionism is a maladaptive and unhealthy character trait. Furthermore, it is unlikely that the supposedly positive aspects of perfectionism are truly correlated with positive personality traits when the negative aspects of perfectionism are absent or statistically controlled for, as Stoeber and Otto (2006) claim, because only one of the nine dimensions of perfectionism had a negative correlation with depression, and the negative correlation that the APS-R High Standard subscale had with depression was very small.

Publication Bias

Evidence of publication bias was examined with funnel plots, the Trim and Fill method, cumulative meta-analyses, and Egger's Regression Test for Funnel Plot Asymmetry. A relationship between effect size and sample size was found in the meta-analyses for the mean correlation of depression with High Standards (the first meta-analysis), Self-Oriented Perfectionism (the third meta-analysis), the FMP Personal Standards (the fifth meta-analysis), Parental Expectations (the eighth meta-analysis) and FMPS Parental Criticism (the ninth meta-analysis), but only the meta-analyses for the mean correlation of depression with HMPS Self-Oriented Perfectionism subscale, HMPS Socially-Prescribed Perfectionism subscale, and FMPS Personal Standard subscale had enough studies to detect even severe publication bias with Egger's Regression Test for Funnel Plot Asymmetry (Card, 2012), and none of those meta-analyses detected significant funnel plot asymmetry with Egger's test. Fail Safe *N* was calculated but not reported because Card (2012) recommended against using Fail Safe *N* when random-effects models are used in a meta-analysis. Kendall's rank correlation test was not used because none of the nine meta-analyses had enough studies to have adequate power to detect even severe publication bias with this test (Card, 2012). Based on the fixed-effects cumulative meta-analyses and a synthesis of the other evidence for a relationship between effect size and sample size the second (APS-R Discrepancy), fourth (HMPS SPP), sixth (FMPS CM), and seventh (FMPS DA) meta-analyses had stable estimates of the mean correlation that did not show much of a relationship between sample size and effect size and that showed stable estimates of the mean correlations, but the eighth (FMPS PE), and ninth (FMPS PC) showed evidence that the mean correlations for these meta-analyses might have been overestimated and the first (APS-R HS), third (HMPS SOP,) and fifth (FMPS PS) showed evidence that the mean correlations for these meta-analyses might have been underestimated. In aggregate, results suggest some evidence of publication bias.

When creating the funnel plots to visually examine whether there was evidence of publication bias, the standard error and sample size were used as the measures of precision for the y-axis, but a better choice for the y-axis for the funnel plots would have been the study weights $(1/SE^2)$ because there was not a perfect relationship between the standard error and the effect size because the correlations were corrected for attenuation due to measurement error (Card, 2012).

Choosing Between a Fixed-Effects and a Random-Effects Model

If there is heterogeneity in the effect sizes in a meta-analysis but the researcher only wants to make inferences about the specific studies used in the meta-analysis, then the fixed-effects model is appropriate (Hedges & Vevea, 1998). When researchers start with a fixed-effects model, then test for homogeneity, and conclude by using a random-effects model because of significant homogeneity between effect sizes, Hedges and Vevea (1998) call this a *conditional random-effects* analysis. According to Borenstein, Hedges, Higgins and Rothstein (2010), some researchers start their meta-analysis using a fixed-effects model, and then use the significance test from the *Q* statistic to determine if the fixed-effects model will suffice or if a random-effects model should be used because of significant between-studies variance, but this is inappropriate. A fixed-effects meta-analysis allows inferences to other studies that are only different from the studies used in the meta-analysis because of using different research participants (Shadish & Haddock, 2009). A random-effects meta-analysis allows inferences to other studies that differ in more than just the participants used in the study, such as different treatments, different measures, or other differences in study characteristics (Shadish & Haddock, 2009). When there is significant heterogeneity in effect sizes from studies used in a meta-analysis, the random-effects model is more conservative than the fixed-effects model (Shadish & Haddock, 2009). Shadish and Haddock (2009) cited Hedges and Vevea (1998) as saying that the type of model used for a meta-analysis should be determined by the inferences the meta-analysis researcher wants to make. According to Borenstein et al. (2010), if any model is going to be used as the default model for a meta-analysis, the random-effects model should be used rather than the fixed-effects model because if T^2, the estimate of the between-study variance, is zero, then the two models are the same and give the same result for the summary effect. Also, when researchers use a random-effects model for a meta-analysis, they should use credibility intervals in order to describe the distribution of effect sizes (Schmidt & Hunter, 2015).

Brown et al. (2003) said that the initial thorough review of a proportion of the relevant studies is an appropriate time for making other methodological decisions and decisions about the model used in the meta-analysis.

According to Petitti (2001) many think that the choice of model for a meta-analysis should be based on the research question that the meta-analyst is trying to answer.

The difference between running random-effects meta-analysis vs. fixed-effects meta-analysis can be seen in the forest plots for the cumulative meta-analyses. Usually effect size estimates for the random-effects meta-analyses were smaller or the same as for the fixed-effect meta-analyses (as revealed in the cumulative meta-analyses forest plots). Also, the confidence intervals for the random-effects analyses were wider than those for the fixed-effect models, but Borenstein et al. (2009) said that is usually the case because the random-effects models have the added between-studies variance component. Three exceptions to this were the third, eighth and ninth meta-analyses. For these meta-analyses, the random-effects analyses gave slightly higher mean correlations than the fixed-effects analyses. But the eighth and ninth meta-analyses had the fewest number of studies out of all nine meta-analyses and Borenstein et al. (2009) said that having too few studies in a random-effects meta-analysis causes the estimate of the between-studies variance and the standard error to be inaccurate, so that may be why these two meta-analyses had different results than the other six meta-analyses. Also, the eighth meta-analysis was the only meta-analysis that did not show significant heterogeneity, but it had only six studies, and it was the meta-analysis with the fewest number of studies. The power to detect heterogeneity was probably too low for this meta-analysis even though the critical value for the test of heterogeneity was set at $\alpha = .10$ for all of the meta-analyses in order to increase the power of the tests for heterogeneity as suggested by Borenstein et al. (2009). In the third meta-analysis, the random-effects analysis also gave a slightly larger effect size for the correlation between HMPS SOP and depression than did the fixed-effects analysis, and since this meta-analysis was based on 25 studies, this result was probably not due to there being too few studies the way it was for the eighth and ninth meta-analyses.

Directions for Future Research

These nine meta-analyses have summarized the information in the sample of 52 studies, and the tentative conclusion was reached that perfectionism appears to be a negative unidimensional construct, but there is not a way to test this conclusion using current meta-analysis techniques. New meta-analysis techniques are needed to answer questions like this one: Is perfectionism really a unidimensional factor? Also, a new meta-analysis technique is needed that could determine if there is a significant difference between the correlation between depression and Parental Expectations ($r = .26$) and the correlation between depression and Self-Oriented Perfectionism ($r = .17$) A meta-analysis technique that could answer this question would help determine if perfectionism consists of two higher-order factors, such as Perfectionistic Strivings and Perfectionistic Concerns, or if perfectionism is unidimensional because it would be testing if the set of correlations between depression and the Perfectionistic Concerns subscales are separate from and significantly different than the set of correlations between depression and the Perfectionistic Strivings subscales.

Future researchers might want to exclude the Parental Expectations and Parental Criticism subscales from the FMPS among the negative dimensions of perfectionism because out of the six negative dimensions of perfectionism studied in these meta-analyses, the PE and PC subscales form the FMPS were the least strongly correlated with depression, so it might be more parsimonious to leave those two subscales out of the maladaptive higher-order factor of perfectionism. The Evaluative Concerns (EC) conception of perfectionism put forth by Dunkley, Blankstein et al. (2006), which was mentioned in the review of the perfectionism literature, seems to be a better combination of perfectionism subscales than the Perfectionistic Concerns construct used in this study because the EC perfectionism conceptualization because does not use the FMPS Parental Expectations or Parental Criticism subscales, which were the two negative dimensions of perfectionism that were least strongly correlated with depression in this study. Also, there were few studies available that gave correlations between depression and Parental Expectations and Parental Criticism, so the test for heterogeneity for the eighth meta-analysis of the FMPS Parental Expectations subscale and perfectionism did not have enough power to detect significant heterogeneity if it existed, so if these scales are going to be used, there needs to be more studies that report correlations for them.

The three multidimensional perfectionism scales that were the focus of this study have been translated into numerous languages other than English, so future research should look at how these measures of perfectionism work in other languages and cultures, and if the perfectionism subscales that were the focus of this study are correlated with depression in the same way when administered in languages other than English.

Future researchers should also look at the relationship between these three multidimensional measures and depression for the entire time-period that these measures have been in use. This study only looked at a time period that was less than half as long as the length of time that the FMPS and HMPS have been available for use in research since they were created in 1990 and 1991, respectively. The APS-R has been available for use in research since it was created in 2001. Doing meta-analyses with studies from the whole time these three measures have been available would give more accurate estimates of the relationships between depression and the subscales of these three multidimensional measures of perfectionism.

Future researchers are advised to conduct studies in which the entire sample of participants are all female as well as all male or conduct studies that report correlations separately for males and females so that moderator analyses are possible to determine if the correlations between depression and the dimensions of perfectionism are different for males and females, and to determine if females have stronger correlations between the Perfectionistic Concerns dimensions (the maladaptive dimensions) of perfectionism and depression since women have higher rates of depression than do men. Also, future researchers should conduct studies that focus on individual ethnic groups, or report correlations separately for separate ethnic groups so that moderator analyses could be done to determine if different ethnic groups have different associations between depression and the dimensions of perfectionism, and to determine if there is a strong correlation between depression and perfectionism for all ethnic group or for just some ethnic groups.

Finally, the dataset used in this study is available by request by emailing the author (gabriel.hottinger@hotmail.com)

Implications for Clinicians

Because the correlation between the APS-R Discrepancy subscale and depression was the strongest direct correlation, and because Discrepancy measures a type of black-and-white thinking, and because black-and-white thinking is a key component of maladaptive perfectionism, clinicians might focus on helping clients or patients who have perfectionism and depression overcome their black-and-white thinking. If a person thinks that he/she must achieve an absolute perfect standard in everything they do or else they are a complete failure, they potentially may become or be very depressed.

Clinicians are advised to ask their clients or patients who have depression to complete the four subscales that were found in this study to have the strongest direct correlations with depression (Concern over Mistakes and Doubts about Actions from the FMPS, Socially-Prescribed Perfectionism from the HMPS, and Discrepancy form the APS-R), and if their patients/clients have high scores on some or all of these four subscales, try to help reduce their clients'/patients' perfectionism as a way to help decrease the clients'/patients' depression. Because perfectionism is a transdiagnostic risk factor not only for depression, but also for other forms of psychopathology, helping patients to be less perfectionistic may help reduce their depression. Clinicians who treat depression should be informed about the relationship between depression and perfectionism so that they might better help clients/patients whose depression is exacerbated by their perfectionistic tendencies. Clinicians should be informed about how perfectionism can cause problems in psychotherapy and can impede progress in psychotherapy (Blatt & Zuroff, 2002).

Clinicians might help their clients or patients understand that a person can have high standards without being a perfectionist. That is, a person can have high standards without seeing those standards in an all-or-nothing way in which the person views him/herself as a complete failure if his/her standards are not completely met. Clinicians might also teach client or patients to have flexible standards rather than absolute standards in which any discrepancy between the high standards and actual performance is focused on and seen as failure, which is what the APS-R Discrepancy subscale measures. Because the APS-R Discrepancy measures rigid high standards and because this study found it to be the subscale that was most highly correlated with depression out of the nine subscales investigated in this study, clinicians should choose the Discrepancy subscale to screen their depressed patients or clients for perfectionism if they can only use one subscale for the purpose of screening for perfectionism.

Clinicians may want to focus on the High Standards and Discrepancy subscales from the APS-R, and use these two subscales to measure perfectionism in their clients/patients. If their clients/patients have high scores on the High Standards subscale, that is not a problem because High Standards was inversely correlated with depression. However, if clients/patients score high on Discrepancy, that is a problem because Discrepancy was the subscale that was most strongly and positively correlated with depression. Clinician might help their clients/patients by helping them decrease their scores on the Discrepancy subscale. Discrepancy is like a measure of black-and-white or all-or-nothing thinking because an unhealthy perfectionist has to do things perfectly or else they view their performance as a failure (Slaney et al., 2001; Tangney, 2002), and all-or-nothing thinking is rigid and inflexible, and negative or unhealthy perfectionism is correlated with all-or-nothing thinking (Blankstein & Dunkley, 2002).

Limitations of This Study

This study had some limitations. The estimates of the between-studies variance has poor precision in the random-effects meta-analyses that had a small number of studies because when conducting a meta-analysis using a random-effects model, increasing the precision of the estimated mean effect size depends not only on the sample size of each study included but also the total number of studies included in the meta-analysis (Borenstein et al., 2009). There is English language bias because only studies written in English were used and only English language databases were searched (Borenstein et al., 2009). The sample of studies was thorough but not exhaustive, and an exhaustive literature search is the best way to prevent publication bias (Borenstein et al., 2009).

References

Abramson, L. Y., Metalsky, G. I., & Alloy, L. B. (1989). Hopelessness depression: A theory-based subtype of depression. *Psychological Review*, *96*(2), 358-372.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Psychological Association. (n.d.) PsycEXTRA fact sheet. Retrieved June 12, 2017 form http://www.apa.org/pubs/databases/psycextra/psycextra-printable-fact-sheet.pdf

American Psychological Association (August 2012). Record Structure for APA Databases. Retrieved March 11, 2017 from www.apa.org/pubs/databases/training/record-structure.pdf

Anhedonia. (2015). In C. Schwarz, *The Chambers Dictionary*. London, UK: Chambers Harrap. Retrieved from http://du.idm.oclc.org/login?url=http://search.credoreference.com/content/entry/chambdict/anhedonia/0

Arbisi, P. A. (2001). Test review of the Beck Depression Inventory-II. In B. S. Plake, & J. C. Impara (Eds.), *The fourteenth mental measurements yearbook* [electronic version]. Retrieved from the Burros Institute's Mental Measurements Yearbook online database.

Bauer, D. J. (2006). Neurotransmitters. In N. Salkind (Ed.), *Encyclopedia of human development*. Thousand Oaks, CA: Sage Publications. Retrieved from http://du.idm.oclc.org/login?url=http://search.credoreference.com/content/entry/sagehd/neurotransmitters/0

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II* San Antonio, TX: Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561-571.

Békés, V., Dunkley, D. M., Taylor, G., Zuroff, D. C., Lewkowski, M., Foley, J. E., Myhr, G., & Westreich, R. (2015). Chronic stress and attenuated improvement in depression over 1 year: The moderating role of perfectionism. *Behavior Therapy, 46,* 478-492.

Blankstein, K. R. & Dunkley, D. M. (2002). Evaluative concerns, self-critical, and personal standard perfectionism: A structural equation modeling strategy. In G. L. Flett & P. L. Hewitt (Eds.), *Perfectionism: Theory, research, and treatment* (pp. 285-315). Washington, DC: American Psychological Association.

Blatt, S. J. (1995). The destructiveness of perfectionism: Implications for the treatment of depression. *American Psychologist, 50,* 1003-1020.

Blatt, S. J., D'Affitti, J. P., & Quinlan, D. M. (1976). Experiences of depression in normal young adults. *Journal of Abnormal Psychology, 85*(4), 383-389. doi:10.1037/0021-843X.85.4.383

Blatt, S. J. & Zuroff, D. C. (2002). Perfectionism in the therapeutic process. In G. L. Flett & P. L. Hewitt (Eds.), *Perfectionism: Theory, research, and treatment* (pp. 393-406). Washington, DC: American Psychological Association.

Bobko, P. (2001) *Correlation and regression: Applications for industrial organizational psychology and management* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Borenstein, M. (2009). Effect sizes for continuous data. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 221-235). New York, NY: Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Hoboken, NJ: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97-111.

Brown, S. A., Upchurch, S. L., & Acton, G. J. (2003). A framework for developing a coding scheme for meta-analysis. *Western Journal of Nursing Research, 25*(2), 205-222. doi:10.1177/0193945902250038

Burgess, A. M. & DiBartolo, P. M. (2016). Anxiety and perfectionism: Relationships, mechanisms, and conditions. In F. M. Sirois & D. S. Molnar (Eds.), *Perfectionism, health, and well-being* (pp. 177-203). Cham, Switzerland: Springer.

Burns, D. D. (1980). The perfectionist's script for self-defeat. *Psychology Today*, 34-52.

Card, N. A. (2012). *Applied meta-analysis for social science research*. New York, NY: Guilford Press.

Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Cole, J. C., Rabin, A. S., Smith, T. L., & Kaufman, A. S. (2004). Development and Validation of a Rasch-Derived CES-D Short Form. *Psychological Assessment, 16*(4), 360-372

Cooper, H. (2009) Hypothesis and problems in research synthesis. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 19-35). New York, NY: Russell Sage Foundation.

Cooper, H. & Hedges, L. V. (2009). Research synthesis as a scientific process. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 3-16). New York, NY: Russell Sage Foundation.

Cox, B. J., Enns, M. W., Clara, I. P. (2002). The multidimensional structure of perfectionism in clinically distressed and college student samples. *Psychological Assessment*, *14*(3), 365-373.

Derogatis, L.R. (1983). SCL-90: Administration, scoring and procedures manual-II. Towson, MD: Clinical Psychometric Research.

Derogatis, L. R., Rickels, K., & Rock, A. F. (1976). The SCL-90 and the MMPI: A step in the validation of a new self-report scale. *British Journal of Psychiatry*, *128,* 280-289.

Dunkley, D. M., Berg, J-L., & Zuroff, D. C. (2012). The role of perfectionism in daily self-esteem, attachment, and negative affect. *Journal of Personality*, *80*(3), 633-663.

Dunkley, D. M., Blankstein, K. R., Masheb, R. M., & Grilo, C. M. (2006). Personal standards and evaluative concerns dimensions of “clinical” perfectionism: A reply to Shafran et al. (2002, 2003) and Hewitt et al. (2003). *Behaviour Research and Therapy, 44*(1), 63-84. doi:10.1016/j.brat.2004.12.004

Dunkley, D. M., Mandel, T., & Ma, D. (2014). Perfectionism, neuroticism, and daily stress reactivity and coping effectiveness 6 months and 3 years later. *Journal of Counseling Psychology, 61*(4), 616-633.

Dunkley, D. M., Solomon-Krakus, S., & Moroz, M. (2016). Personal standards and self-critical perfectionism and distress: Stress, coping, and perceived social support as mediators and moderators. In F. M. Sirois & D. S. Molnar (Eds.), *Perfectionism, health, and well-being* (pp. 157-176). Cham, Switzerland: Springer.

Dunkley, D. M., Zuroff, D. C. & Blankstein, K. R. (2006). Specific perfectionism components versus self-criticism in predicting maladjustment. *Personality and Individual Differences, 40,* 665-676.

Duval, S., & Tweedie, R. (2000a). A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*, 89-98.

Duval, S, & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455-463.

EBSCO Help (n.d.) Wildcard and truncation symbols. Retrieved March 18, 2017 from http://support.ebsco.com/help/index.php?help_id=137

Egan, S. J., Wade, T. D., & Shafran, R. (2011). Perfectionism as a transdiagnostic process: A clinical review. *Clinical Psychology Review, 31*, 203-212.

Egan, S. J., Wade, T. D., & Shafran, R. (2012). The transdiagnostic process of perfectionism. *Revista de Psicopatología y Psicología Clínica*, *17*(3), 279-294.

Egan, S. J., Wade, T. D., Shafron, R. & Antony, M. M. (2014). *Cognitive-behavioral treatment of perfectionism*. New York, NY: Guildford Press.

Eichman, W. J. (1978). Test review of the Profile of Mood States. In O.K. Buros (Ed.), *The eighth mental measurements yearbook* [electronic version]. Retrieved from the Burros Institute's Mental Measurements Yearbook online database.

Enns, M. W. & Cox, B. J. (2002). The nature and assessment of perfectionism: A critical analysis. In G. L. Flett & P. L. Hewitt (Eds.), *Perfectionism: Theory, research, and treatment* (pp. 33-62). Washington, DC: American Psychological Association.

Farmer, R. F. (2001). Test review of the Beck Depression Inventory-II. In B. S. Plake & J. C. Impara, (Eds.), *The fourteenth mental measurements yearbook* [electronic version]. Retrieved from the Burros Institute's Mental Measurements Yearbook online database.

Fernandez, E. (1998) Test review of the Hamilton Depression Inventory. In J. C. Impara & B. S. Plake (Eds.), *The thirteenth mental measurements yearbook* [electronic version]. Retrieved from the Burros Institute's Mental Measurements Yearbook online database.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *User's guide for the Structured Clinical Interview for DSM-IV Axis I disorders, clinical version*. Washington, DC: American Psychiatric Association.

Flett, G. L., & Hewitt, P. L. (2002). Perfectionism and maladjustment: An overview of theoretical, definitional, and treatment issues. In G. L. Flett & P. L. Hewitt (Eds.), *Perfectionism: Theory, research, and treatment* (pp. 5-31). Washington, DC: American Psychological Association.

Flett, G. L. & Hewitt, P. L. (2015). Measures of perfectionism. In G. J. Boyle, D. H. Saklofske, & G Matthews (Eds.) *Measures of personality and social psychology constructs* (pp. 595-618). New York, NY: Academic Press.

Flett, G. L., Hewitt, P. L., Blankstein, K. R., & Gray, L. (1998) Psychological distress and the frequency of perfectionistic thinking. *Journal of Personality and Social Psychology, 75*(5), 1363-1381.

Flett, G. L., Hewitt, P. L., Besser, A., Su, C., Vaillancourt, T., Boucher, D.,…Gale, O. (2016). The Child–Adolescent Perfectionism Scale: Development, psychometric properties, and associations with stress, distress, and psychiatric symptoms. *Journal of Psychoeducational Assessment, 34*(7), 634-652.

Flett, G. L., Hewitt, P. L. & Heisel, M. J. (2014). The destructiveness of perfectionism revisited: Implications for the assessment of suicide risk and the prevention of suicide. *Review of General Psychology, 18*(3), 156-172.

Frost, R. O., Glossner, K., & Maxner, S. (2010). Social anxiety disorder and its relationship to perfectionism. In S. G. Hofmann & P. M. DiBartolo (Eds.) *Social anxiety: Clinical, developmental, and social perspectives*, pp. 119-145. Saint Louis, MO: Elsevier Science.

Frost, R. O., Heimberg, R. G., Holt, C. S., Mattia, J. I., & Neubauer, A. L. (1993). A comparison of two measures of perfectionism. *Personality and Individual Differences, 14*(1), 119-126. doi:10.1016/0191-8869(93)90181-2

Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research, 14*(5), 449-468. doi:10.1007/BF01172967

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3-8.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Boston, MA: Allyn and Bacon.

Gliner, J., A., Morgan, G. A., Leech, N. L. (2009). *Research methods in applied settings: An integrative approach to design and analysis* (2nd ed.). New York, NY: Routledge.

Goldberg, D. (2016). The current status of the diagnosis of depression. In J. C. Wakefield & S, Demazeuz (Eds.), *Sadness or depression? International perspectives on the depression epidemic and its meaning* (pp. 17-27). Dordrecht, Netherlands: Springer.

Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Pike, C. T. & Kessler, R. C. (2015). The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *Journal of Clinical Psychiatry*, *76*(2), 155-162.

Hames, J. L., Hagan, C. R., & Joiner, T. E. (2013). Interpersonal processes in depression. *The Annual Review of Clinical Psychology, 9*, 355-377.

Hedges, L. V. (1992). Meta-analysis. *Journal of Educational Statistics, 17*(4), 279-296.

Hedges, L. V. (2009). Statistical considerations. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 37-47). New York, NY: Russell Sage Foundation.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486-504.

Hewitt, P. L. & Flett, G. L. (1991a). Dimensions of perfectionism in unipolar depression. *Journal of Abnormal Psychology, 100*, 98-101.

Hewitt, P. L., & Flett, G. L. (1991b). Perfectionism in the self and social contexts: Conceptualization, assessment, and association with psychopathology. *Journal of Personality and Social Psychology, 60*(3), 456-470. doi:10.1037/0022-3514.60.3.456

Hewitt, P. L., & Flett, G. L. (1993). Dimensions of perfectionism, daily stress, and depression: A test of the specific vulnerability hypothesis. *Journal of Abnormal Psychology, 102*, 58-65.

Hewitt, P. L., Flett, G. L., Besser, A., Sherry, S. B., & McGee, B. (2003). Perfectionism is multidimensional: A reply to Shafran, Cooper and Fairburn (2002). *Behaviour Research and Therapy, 41*, 1221-1236

Hewitt, P. L., Flett, G. L., & Ediger, E. (1996). Perfectionism and depression: Longitudinal assessment of a specific vulnerability hypothesis. *Journal of Abnormal Psychology, 105*, 276-280.

Hill, R. W., McIntire, K. & Bacharach, V. R. (1997). Perfectionism and the Big Five factors. *Journal of Social Behavior and Personality, 12*(1), 257-270.

Ingram, R. (2012). Depression. In *encyclopedia of human behavior* (2nd ed., pp.682-689). Lawrence, KS: University of Kansas.

Kandel, D. B., & Davies, M. (1982). Epidemiology of depressive mood in adolescents: An empirical study. *Archives of General Psychiatry, 39*, 1205-1212.

Keller, M. B., Lavori, P. W., Friedman, B., Nielsen, E., Endicott, J., McDonald-Scott, P., & Andreasen, N. C. (1987). The Longitudinal Interval Follow-Up Evaluation: A comprehensive method for assessing outcome in prospective longitudinal studies. *Archives of General Psychiatry, 44*, 540-548.

Kempke, S., Van Houdenhove, B., Claes, S., Luyten, P. (2016). The role of perfectionism in chronic fatigue syndrome. In F. M. Sirois and D. S. Molnar (Eds.) *Perfectionism, health, and well-being,* (pp. 101-118). Cham, Switzerland: Springer.

Konstantopoulos, S. (2013). Meta-analysis. In T. Teo (Ed.), *Handbook of quantitative methods for educational research,* (pp. 231-246). Boston, MA: Sense Publishers.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49). Thousand Oaks, CA: Sage publications.

Lo, A., & Abbott, M. J. (2013). Review of the theoretical, empirical, and clinical status of adaptive and maladaptive perfectionism. *Behaviour Change, 30*(2), 96-116.

Malouff, J. M., Schutte, N. S., & Ramerth, W. (1985). Evaluation of a short form of the POMS-Depression scale. *Journal of Clinical Psychology, 41*(3), 389-390.

Martin, W. E. (2003). Test review of the Hospital Anxiety and Depression Scale. In B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), *The fifteenth mental measurements* *yearbook* [electronic version]. Retrieved from the Burros Institute's Mental Measurements Yearbook online database.

Mathers, C. D. & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine, 3*(11), 2011-2030.

Matt, G., & Cook, T. D. (2009). Threats to the validity of generalized inferences. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 537-560). New York, NY: Russell Sage Foundation.

McInnis, M. G., Riba, M., & Greden, J. F. (2014). Depressive disorders. In R. E. Hales, S. C. Yudofsky, & L. W. Roberts (Eds.), *The American Psychiatric Publishing textbook of psychiatry* (6th ed., pp. 353-389). Arlington, VA: American Psychiatric Publishing.

McNair, D. M., Lorr, M., & Droppleman, I. F. (1971). *Profile of Mood States.* San Diego: Educational and Industrial Testing Service.

Mineka, S., Pury, C. L., & Luten, A. G. (1995). Explanatory style in anxiety and depression. In G. M. Buchanan & M. E. P. Seligman (Eds.), *Explanatory style*, pp.135-158. New York, NY: Routledge.

Molnar, D. S., & Sirois, F. M. (2016). Perfectionism, health, and well-being: Epilogue and future directions. In F. M. Sirois and D. S. Molnar (Eds.) *Perfectionism, health, and well-being,* (pp. 285-302). Cham, Switzerland: Springer.

Molnar, D. S., Sirois, F. M., & Methot-Jones, T. (2016). Trying to be perfect in an imperfect world: Examining the role of perfectionism in the context of chronic illness. In F. M. Sirois and D. S. Molnar (Eds.) *Perfectionism, health, and well-being*, (pp. 69-99). Cham, Switzerland: Springer.

Nahas, Z. (2016). Depression: Current conceptual trends. In J. A. Camprodon, S. L. Rauch, B. D. Greenberg & D. D. Dougherty (Eds.) *Psychiatric neurotherapeutics: Contemporary surgical and device-based treatments* (pp. 1-21). New York, NY: Humana Press.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*(2), 157-159.

Orwin, R. G., & Vevea, J. L. (2009) Evaluating coding decisions. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 177-203). New York, NY: Russell Sage Foundation.

Payne, R. W. (1985). Test review of the Symptom Checklist-90-R. In J. V. Mitchell, Jr. (Ed.), *The ninth mental measurements yearbook* [electronic version]. Retrieved from the Burros Institute's Mental Measurements Yearbook online database.

Petitti, D. B. (2001). Approaches to heterogeneity in meta-analysis. *Statistics in Medicine, 20*, 3625-3633.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401.

Raudenbush, S. W. (2009) Analyzing effect sizes: Random-effects models. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research* *synthesis and meta-analysis* (2nd ed., pp. 295-315). New York, NY: Russell Sage Foundation.

Reynolds, W. M., & Kobak, K. A. (1995). Reliability and validity of the Hamilton Depression Inventory: A paper-and-pencil version of the Hamilton Depression Rating Scale Clinical Interview. *Psychological Assessment, 7*(4), 472-483.

Rice, K. G., Richardson, C. M. & Tueller, S. (2014) The Short Form of the Revised Almost Perfect Scale. *Journal of Personality Assessment, 96*(3), 368-379.

Richards, D. (2011). Prevalence and clinical course of depression: A review. *Clinical Psychology Review, 31*(2011), 1117-1125.

Rittberg, B. R. (2016). Major depressive disorder. In S. H. Fatemi and P, J. Clayton (Eds.) *The medical bias of psychiatry* (pp. 79-90). New York, N. Y.: Springer.

Ritschel, L. A., Gillespie, C. F., Arnarson, E. O. & Craighead, W. E. (2013). Major depressive disorder. In W. E. Craighead, D. J. Milkowitz, & L. W. Craighead (Eds.) *Psychopathology: History, diagnosis, and empirical foundations* (2nd ed.; pp. 285-333). Somerset, N.J.: Wiley.

Rosenthal, R. (1979). The 'File drawer problem' and tolerance for null results. *Psychological Bulletin, 86,* 638-641.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage Publications.

RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N.,…Keller, M. B. (2003). The 16-item Quick Inventory of Depressive

Symptomatology, clinician rating, and self-report: A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*, 573-583.

Schmidt, F. L. & Hunter, J. E. (2015) *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Los Angeles, CA: Sage.

Schmidt, F. L., Le, H., & Oh, I. (2009). Correcting for the distorting effects of study artifacts in meta-analysis. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 317-333). New York, NY: Russell Sage Foundation.

Schwarzer, G. (2007). meta: An R package for meta-analysis, *R News, 7*(3), 40-45.

Shadish, W. R., & Haddock, C. K. (2009). Combining estimates of effect size. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 257-277). New York, NY: Russell Sage Foundation.

Shafran, R., Cooper, Z., & Fairburn, C. G. (2002). Clinical perfectionism: A cognitive-behavioural analysis. *Behaviour Research and Therapy, 40*, 773-791.

Sherry, S. B., Gautreau, C. M., Mushquash, A. R., Sherry, D. L., & Allen, S. L. (2014). Self-critical perfectionism confers vulnerability to depression after controlling for neuroticism: A longitudinal study of middle aged, community-dwelling women. *Personality and Individual Differences, 69*, 1-4.

Sherry, S. B., Richards, J. E., Sherry, D. L., & Stewart, S. H. (2014). Self-critical perfectionism is a vulnerability factor for depression but not anxiety: A 12-month, 3-wave longitudinal study. *Journal of Research in Personality, 52*, 1-5.

Sirois, F. M. (2016). Perfectionism and health behaviors: A self-regulation resource perspective. In F. M. Sirois and D. S. Molnar (Eds.) *Perfectionism, health, and well-being,* (pp. 45-67). Cham, Switzerland: Springer.

Sirois, F. M. & Molnar, D. S. (2016). Conceptualizations of perfectionism, health, and well-being: An introductory overview. In F. M. Sirois and D. S. Molnar (Eds.*)* *Perfectionism, health, and well-being*, (pp. 1-21). Cham, Switzerland: Springer.

Slaney, R. B., Rice, K. J., & Ashby, J. S. (2002). A programmatic approach to measuring perfectionism: The Almost Perfect Scales. In G. L. Flett & P. L. Hewitt (Eds.), *Perfectionism: Theory, research, and treatment* (pp. 63-88). Washington, DC: American Psychological Association.

Slaney, R. B., Rice, K. G., Mobley, M., Trippi, J., & Ashby, J. S. (2001). The Revised Almost Perfect Scale. *Measurement and Evaluation in Counseling and Development*, *34*(3), 130-145.

Smith, M. M., Sherry, S. B., Rnic, K., Saklofske, D. H., Enns, M., & Gralnick, T. (2016). Are perfectionism dimensions vulnerability factors for depressive symptoms after controlling for neuroticism? A meta-analysis of 10 longitudinal studies. *European Journal of Personality, 30*, 201-212.

Stoeber, J., & Otto, K. (2006). Positive conceptions of perfectionism: Approaches, evidence, challenges. *Personality and Social Psychology Review, 10*(4), 295-319. doi:10.1207/s15327957pspr1004_2

Sutton, A. J. (2009). Publication bias. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 435-452). New York, NY: Russell Sage Foundation.

Tangney, J. P. (2002). Perfectionism and the self-conscious emotions: Shame, guilt, embarrassment, and pride. In G. L. Flett & P. L. Hewitt (Eds.), *Perfectionism: Theory, research, and treatment* (pp. 199-215). Washington, DC: American Psychological Association.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48. Retrieved from http://www.jstatsoft.org/v36/i03/

Vanchu-Orosco, M. (2012). *A meta-analysis of testing accommodations for students with disabilities: Implications for high-stakes testing disabilities: Implications for high-stakes testing* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (3549773)

Wakefield, J. C. & Demazeux, S. (2016). Introduction: Depression, one and many. In J. C. Wakefield & S. Demazeux (Eds.), *Sadness or depression? International perspectives on the depression epidemic and its meaning* (pp. 1-15). Dordrecht, Netherlands: Springer.

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite Model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology, 104*(1), 3-14.

Wilson, D. B. (2009). Systematic coding. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 159-176). New York, NY: Russell Sage Foundation.

Wood, J. A. (2008). Methodology for dealing with duplicate study effects in a meta-analysis. *Organizational Research Methods, 11*(1), 79-95.

World Health Organization (1993). *The ICD-10 classification of mental and behavioral disorders: Diagnostic criteria for research*. Geneva, Switzerland: Author.

World Health Organization (April 2016). Depression fact sheet. Retrieved from http://www.who.int/mediacentre/factsheets/fs369/en/

Wryobeck, J. M., Haines, M. E., Wynkoop, T. F., & Swanson, M. M. (2013). Depressive Disorders. In C. A. Noggle & R. S. Dean (Eds.), *The neuropsychology of psychopathology* (pp. 201-220). New York, NY: Springer.

Yeaton, W. H. & Wortman, P. M. (1993). On the reliability of meta-analytic reviews: The role of intercoder agreement. *Evaluation Review, 17*(3), 292-309.

Zigmond, A. S. and Snaith, R. P. (1983), The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica, 67*, 361–370. doi:10.1111/j.1600-0447.1983.tb09716.x

Appendices

Appendix A: Codebook

No.	Name	Values or Codes	Description
1	ID	Text	Identification number assigned to every study
2	Year	Text	
3	First Author	Text	
4	Study Type	code=0	Published
		code=1	Unpublished study or Dissertation/Thesis
5	Duplicate Sample/Data		When there are more than one study uses the same participant data, only code the earliest study out of the set. Put “N/A” if the study being coded does not use the same participant data as a previous study, but if the study being coded uses the same participant data as a previously study, put the following info from the previous study which the study being coded duplicates: first author’s last name from the previous study and the year of previous study (so that I know what previous study the study being coded duplicates).
6a	Age Range	Number range	Range of age of sample—code “NR” if Not Reported
6b	Age	Number	Mean age for the entire sample—code “NR” if not reported
7	Population	Text	Code a few words that describe the population that the study participants represent, such as “females with eating disorders,” or “clinical depressed inpatients” or “college students”—if nothing is specified, type “NR” for not reported
8	Sample size	Number	Final total sample used in the study

Codebook continued			
9	Sample by Gender	Gender description of the sample	
		Number or percentage	Number or percentage of males in the sample (if percentage, please include the sign “%”)
		Number or percentage	Number or percentage of females in the sample (if percentage, please include the sign “%”)
		NR	Not reported
10	Nationality of Sample	Text	Please specify Nationality or Nationalities of Study Participants if other than U.S.
11	Language Scales Administered in	dichotomous	0=English; 1=Other than English
12	**Depression Measure**		
12a	Name of Depression Measure		Types of instruments used to measure depression
		BDI	BDI=Beck Depression Inventory
		BDI-2	BDI-2=Beck Depression Inventory-2
		CES-D	CES-D=Center for Epidemiologic Studies-Depression scale
		SRDS	Self-Rating Depression Scale
		HDRS	Hamilton Depression Rating Scale (HDRS)
		HADS	Hospital Anxiety and Depression Scale (HADS)
		DASS-D	Depression, Anxiety and Stress Scale (Depression subscale)
		HDI	Hamilton Depression Inventory
		MASQ	Mood and Anxiety Symptom Questionnaire (MASQ)
		POMS-D	Profile of Mood States POMS-D (Depression subscale)
		Other	Specify name of depression measure
12b	Any Modification to the measure of depression	code=0	shortened version
		code-1	translated version or other modification
12c	Rel_Depr	Estimate	Internal consistency (Cronbach's alpha) of the instrument—record only if estimate based on the study's sample is reported

Codebook continued			
13	**Perfectionism Measures**		
13a	Name(s) of Perfectionism Measure(s)		Names of the 3 relevant multidimensional measures of perfectionism
		0=FMPS	FMPS=Frost et al (1990) Multidimensional Perfectionism Scale
		1=HMPS	HMPS=Hewitt & Flett (1991) Multidimensional Perfectionism Scale
		2=APS-R	APS-R=Almost Perfect Scale-Revised by Slaney et al. (2001)
13b	Any Modification of the Three Relevant Perfectionism Scales	code=0	shortened version
		code=1	translated version or other modification
14	**Reliability_FMPS Scale**		Internal consistency of FMPS subscales. Reported as Cronbach's alpha, record only if an estimate based on the study's sample is reported
14a	Rel_PC	Estimate	Internal consistency of Parental Criticism scale
14b	Rel_PE	Estimate	Internal consistency of Parental Expectation scale
14c	Rel_CM	Estimate	Internal consistency of Concern of Mistakes scale
14d	Rel_DA	Estimate	Internal consistency of Doubts of Action scale
14e	Rel_PS	Estimate	Internal consistency of Personal Standards scale
		Empty	Not reported
15	**FMPS: Correlation Coefficients**		Relation between FMPS subscale scores with Depression scores. It is reported as a bivariate correlation coefficient for the sample size specified in No. 8 above
15a	rPC_FMPS~Dep	Estimate	Relationship between Parental Criticism and Depression
15b	rPE_FMPS~Dep	Estimate	Relationship between Parental Expectation and Depression
15c	rCM_FMPS~Dep	Estimate	Relationship between Concern of Mistakes and Depression
15d	rDA_FMPS~Dep	Estimate	Relationship between Doubts of Action and Depression

Codebook continued			
15e	rPS_FMPS~Dep	Estimate	Relationship between Personal Standards and Depression
		Empty	Not reported
16	**Reliability_HMPS scale**		Internal consistency of HMPS subscales. Reported as Cronbach's alpha, record only if an estimate based on the study's sample is reported
16a	Rel_SPP	Estimate	Internal consistency of Socially Prescribed Perfectionism scale
16b	Rel_SOP	Estimate	Internal consistency of Self-Oriented Perfectionism scale
		Empty	Not reported
17	**HMPS: Correlation coefficients**		Relation between HMPS subscale scores with Depression scores. It is reported as a bivariate correlation coefficient for the sample size specified in No. 8 above
17a	rSPP_HMPS~Dep	Estimate	Relationship between Socially Prescribed Perfectionism and Depression
17b	rSOP_HMPS~Dep	Estimate	Relationship between Self-Oriented Perfectionism and Depression
		Empty	Not reported
18	**Reliability_APS-R scale**		Internal consistency of APS-R subscales. Reported as Cronbach's alpha, record only if an estimate based on the study's sample is reported
18a	Rel_HS	Estimate	Internal consistency of High Standard scale
18b	Rel_Dis	Estimate	Internal consistency of Discrepancy scale
		Empty	Not reported
19	**APS-R: Correlation coefficients**		Relation between APS-R subscale scores with Depression scores. It is reported as a bivariate correlation coefficient for the sample size specified in No.8 above
19a	rHS_APS-R~Dep	Estimate	Relationship between High Standards and Depression
19b	rDis_APS-R~Dep	Estimate	Relationship between Discrepancy and Depression
		Empty	Not reported

Appendix B: 53 Studies Included in The Nine Meta-Analyses

Akram, U., Ellis, J. G., Barclay, N. L. (2015). Anxiety mediates the relationship between perfectionism and insomnia symptoms: A longitudinal study. *PLoS ONE, 10*(10), 1-14. doi:10.1371/journal.pone.0138865

Argus, G. & Thompson, M. (2008). Perceived social problem solving, perfectionism, and mindful awareness in clinical depression: An exploratory study. *Cognitive Therapy and Research, 32,* 745-757. doi:10.1007/s10608-006-9102-1

Arpin-Cribbie, C. A., & Cribbie, R. A. (2007) Psychological correlates of fatigue: Examining depression, perfectionism, and automatic negative thoughts. *Personality and Individual Differences, 43,* 1310-1320. doi:10.1016/j.paid.2007.03.020

Athulya, J., Sudhir, P. M., & Philip, M. (2016). Procrastination, perfectionism, coping and their relation to distress and self-esteem in college students. *Journal of Indian Academy of Applied Psychology, 42*(1), 82-91.

Bardone-Cone, A. M. (2007). Self-oriented and socially prescribed perfectionism dimensions and their associations with disordered eating. *Behaviour Research and Therapy, 45*, 1977-1986. doi:10.1016/j.brat.2006.10.004

Black, J. & Reynolds, W. M. (2013). Examining the relationship of perfectionism, depression, and optimism: Testing for mediation and moderation. *Personality and Individual Differences, 54,* 426-431. doi:10.1016/j.paid.2012.10.012

Blankstein, K. R. & Lumley, C. H. (2008). Multidimensional perfectionism and ruminative brooding in current dysphoria, anxiety, worry, and anger. *Journal of*

Rational-Emotive Cognitive-Behavior Therapy, 26(3), 168-193. doi:10.1007/s10942-007-0068-z

Brown, J. R. & Kocovski, N. L. (2013). Perfectionism as a predictor of post-event rumination in a socially anxious sample. *Journal of Rational-Emotive Cognitive-Behavior Therapy, 32,*150-163. doi:10.1007/s10942-013-0175-y

Chang, E. C. (2013). Perfectionism and loneliness as predictors of depressive and anxious symptoms in Asian and European Americans: Do self-construal schemas also matter? *Cognitive Therapy and Research, 37*, 1179-1188. doi:10.1007/s10608-013-9549-9

Chang, E. C., Hirsch, J. K., Sanna, L. J., Jeglic, E. L., & Fabian, C. G. (2011). A preliminary study of perfectionism and loneliness as predictors of depressive and anxious symptoms in Latinas: A top-down test of a model. *Journal of Counseling Psychology, 58* (2), 441-448. doi:10.1037/a0023255

Chen, C., Hewitt, P. L. & Flett, G. L. (2017). Ethnic variations in other-oriented perfectionism's associations with depression and suicide behaviour. *Personality and Individual Differences, 104*, 504-509. doi:10.1016/j.paid.2016.09.021

DiBartolo, P. M., Li, C. Y., & Frost, R. O. (2008). How do the dimensions of perfectionism relate to mental health? *Cognitive Therapy and Research, 32*, 401-417. doi:10.1007/s10608-007-9157-7

Dunkley, D. M., Blankstein, K. R., & Berg, J.-L. (2012). Perfectionism dimensions and the five-factor model of personality. *European Journal of Personality, 26*, 233-244. doi:10.1002/per.829

Elion, A. A., Wang, K. T., Slaney, R. B., & French, B. H. (2012). Perfectionism an African American students: Relationship to racial identity, GPA, self-esteem, and depression. *Cultural Diversity and Ethnic Minority Psychology, 18*(2) 118-127. doi:10.1037/a0026491

Flett, G. L., Besser, A., Hewitt, P. L. & Davis, R. A. (2007). Perfectionism, silencing the self, and depression. *Personality and Individual Differences, 43*, 1211-1222. doi:10.1016/j.paid.2007.03.012

Flett, G. L., Galfi-Pechenkov, I., Molnar, D. S., Hewitt, P. L. & Goldstein, A. L. (2012). Perfectionism, mattering and depression: A mediational analysis. *Personality and Individual Differences, 52*, 828-832. doi:10.1016/j.paid.2011.12.041

Flett, G. L., Besser, A. & Hewitt, P. L. (2014). Perfectionism and interpersonal orientations in depression: An analysis of validation seeking and rejection sensitivity in a community sample of young adults. *Psychiatry, 77*(1), 67-85. doi:10.1521/psyc.2014.77.1.67

Flett, G. L., Nepon, T., Hewitt, P. L. & Fitzgerald, K. (2016a). Perfectionism, components of stress reactivity, and depressive symptoms. *Journal of Psychopathology and Behavioral Assessment, 38*, 645-654. doi:10.1007/s10862-016-9554-x

Flett, G. L., Nepon, T., Hewitt, P. L, Molnar, D. S. & Zhao, W. (2016b) Projecting perfection by hiding effort: Supplementing the perfectionistic self-presentation scale with a brief self-presentation measure. *Self and Identity, 15*(3), 245-261. doi:10.1080/15298868.2015.1119188

Garrison, A. M. (2014). *Testing a model of maladaptive perfectionism and depression symptoms: The roles of emotional disclosure, emotion regulation strategies, adult attachment, and shame* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3582759)

Gnilka, P. B., Ashby, J. S. & Noble, C. M. (2013). Adaptive and maladaptive perfectionism as mediators of adult attachment styles and depression, hopelessness, and life satisfaction. *Journal of Counseling and Development, 91*, 78-86. doi:10.1002/j.1556-6676.2013.00074.x

Hamamura, T., & Laird, P. G. (2014). The effect of perfectionism and acculturative stress on levels of depression experienced by East Asian international students. *Journal of Multicultural Counseling and Development*, *42*, 205-217. doi:10.1002/j.2161-1912.2014.00055.x

Harris, P. W., Pepper, C. M., & Maack, D. J. (2008). The relationship between maladaptive perfectionism and depressive symptoms: The mediating role of rumination. *Personality and Individual Differences, 44*, 150-160. doi:10.1016/j.paid.2007.07.011

Huprich, S. K., Porcerelli, J., Keaschuk, R., Binienda, J., & Engle, B. (2008). Depressive personality disorder, dysthymia, and their relationship to perfectionism. *Depression and Anxiety, 25,* 207-217. doi:10.1002/da.20290

Iannantuono, A. C., & Tylka, T. L. (2012). Interpersonal and intrapersonal links to body appreciation in college women: An exploratory model. *Body Image, 9*(2), 227-235. doi:10.1016/j.bodyim.2012.01.004

Jain, M. Sudhir, P. M. (2010). Dimensions of perfectionism and perfectionistic self-presentation in social phobia. *Asian Journal of Psychiatry, 3*(4), 216-221. doi:10.1016/j.ajp.2010.08.006

La Rocque, C. L., & Lee, L., & Harkness, K. L. (2016). The role of current depression symptoms in perfectionistic stress enhancement and stress generation. Journal of Social and Clinical Psychology, 35(1), 64-86. doi:10.1521/jscp.2016.35.1.64

Leventhal, N. (2007). *Perfectionism, self-concealment, and attitudes toward mental health treatment* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3333088)

Mackinnon, S. P., Sherry, S. B., Pratt, M. W., & Smith, M. M. (2014). Perfectionism, friendship intimacy, and depressive affect in transitioning university students: A longitudinal study using mixed methods. *Canadian Journal of Behavioural Science, 46*(1), 49-59. doi:10.1037/a0033764

Malinowski, A. S., Veselka, L., & Atkinson, B. (2017). An investigation of vulnerability factors for depression. *Personality and Individual Differences, 107*(1), 126-130. doi:10.1016/j.paid.2016.11.049

Mathew, J., Dunning, C., Coats, C., & Whelan, T. (2014). The mediating influence of hope on multidimensional perfectionism and depression. *Personality and Individual Differences, 70*, 66-71. doi:10.1016/j.paid.2014.06.008

Mehr, K. E., & Adams, A. C. (2016) Self-compassion as a Mediator of maladaptive perfectionism and depressive symptoms in college students. *Journal of College Student Psychotherapy, 30*(2), 132-145. doi:10.1080/87568225.2016.1140991

Moroz, M., & Dunkley, D. M. (2015). Self-critical perfectionism and depressive symptoms: Low self-esteem and experiential avoidance as mediators. *Personality and Individual Differences, 87,* 174-179. doi:10.1016/j.paid.2015.07.044

Nepon, T., Flett, G. L., & Hewitt, P. L. (2016). Self-image goals in trait perfectionism and perfectionistic self-presentation: Toward a broader understanding of the drives and motives of perfectionists. *Self and Identity, 15*(6), 683-701. doi:10.1080/15298868.2016.1197847

Nepon, T., Flett, G. L., Hewitt, P. L., & Molnar, D. S. (2011). Perfectionism, negative social feedback, and interpersonal rumination in depression and social anxiety. *Canadian Journal of Behavioural Science, 43*(4), 297-308. doi:10.1037/a0025032

Noble, C. L., Ashby, J. S., & Gnilka, P. B. (2014). Multidimensional perfectionism, coping, and depression: Differential prediction of depression symptoms by perfectionism type. *Journal of College Counseling, 17*(1), 80-94. doi:10.1002/j.2161-1882.2014.00049.x

Olson, M. L. & Kwon, P. (2008). Brooding perfectionism: Refining the roles of rumination and perfectionism in the etiology of depression. *Cognitive Therapy and Research, 32*(6), 788-802. doi:10.1007/s10608-007-9173-7

Ozer, B. U., O'Callaghan, J., Bokszczanin, A., Ederer, E., & Essau, C. (2014). Dynamic interplay of depression, perfectionism and self-regulation on procrastination. *British Journal of Guidance and Counseling, 42*(3), 309-319. doi:10.1080/03069885.2014.896454

Patterson, R. E., Wang, K. T., & Slaney, R. B. (2012). Multidimensional perfectionism, depression and relational health in women with eating disturbances. *Eating Behaviors, 13*(3), 226-232. doi:10.1016/j.eatbeh.2012.03.004

Rice, K. G. & Ashby, J. S. (2007). An efficient method for classifying perfectionists. *Journal of Counseling Psychology, 54*(1), 72-85. doi:10.1037/0022-0167.54.1.72

Rice, K. G., Choi, C.-C., Zhang, Y., Morero, Y. I., & Anderson, D. (2012). Self-critical perfectionism, acculturative stress and depression among international students. *The Counseling Psychologist, 40*(4), 575-600. doi:10.1177/0011000011427061

Rice, K. G., Richardson, C. M., & Tueller, S. (2014). The Short Form of the Revised Almost Perfect Scale. *Journal of Personality Assessment, 96*(3), 368-379. doi:10.1080/00223891.2013.838172

Scott, J. (2007). The effect of perfectionism and unconditional self-acceptance on depression. *Journal of Rational-Emotive and Cognitive-Behavior Therapy, 25*(1), 35-64. doi:10.1007/s10942-006-0032-3

Shanmugam, V., Jowett, S., & Meyer, C. (2013) Eating psychopathology amongst athletes: The importance of relationships with parents, coaches and teammates. *International Journal of Sport and Exercise Psychology, 11*(1), 24-38. doi:10.1080/1612197X.2012.724197

Sherry, D. L., Sherry, S. B., Hewitt, P. L., Mushquash, A., & Flett, G. L. (2015). The existential model of perfectionism and depressive symptoms: Tests of incremental validity, gender differences, and moderated mediation. *Personality and Individual Differences, 76*, 104-110. doi:10.1016/j.paid.2014.12.002

Smith, M. M., Saklofske, D. H., Yan, G., Sherry, S. B. (2016). Cultural similarities in perfectionism: Perfectionistic strivings and concerns generalize across Chinese and Canadian groups. *Measurement and Evaluation in Counseling and Development*, *49*(1), 63-76. doi:10.1177/0748175615596785

Steele, A. L., O'Shea, A., Murdock, A., & Wade, T. D. (2011). Perfectionism and its relationship to overvaluation of weight and shape and depression in an eating disorder sample. *International Journal of Eating Disorders, 44*(5), 459-464.

Sturman, E. D., Flett, G. L., Hewitt, P. L., & Rudolph, S. G. (2009). Dimensions of perfectionism and self-worth contingencies in depression. *Journal of Rational-Emotive Cognitive-Behavior Therapy, 27*, 213-231. doi:10.1007/s10942-007-0079-9

Wheeler, H. A., Blankstein, K. R., Antony, M. M., McCabe, R. E., & Bieling, P. J. (2011). Perfectionism in anxiety and depression: Comparisons across disorders, relations with symptom severity, and role of comorbidity. *International Journal of Cognitive Therapy, 4*(1), 66-91. doi:10.1521/ijct.2011.4.1.66

Wilson, C., Hunter, S. C., Rasmussen, S. & McGowan, A. (2015). They made you perfect: A test of the social reaction model of perfectionism. *Aggressive Behavior, 41*, 421-431. doi:10.1002/ab.21572

Wu, T.-F., & Wei, M. (2008). Perfectionism and negative mood: The mediating role of validation from others versus self. *Journal of Counseling Psychology*, *55*(2), 276-288. doi:10.1037/0022-0167.55.2.276

Yoon, J., & Lau, A. S. (2008). Maladaptive perfectionism and depressive symptoms among Asian American college students: Contributions of interdependence and parental relations. *Cultural Diversity and Ethnic Minority Psychology*, *14*(2), 92-101. doi:10.1037/1099-9809.14.2.92

One Duplicate Study

Akram, U., Ellis, J. G., Myachykov, A., Chapman, A. J., & Barclay, N. L. (2017). Anxiety mediates the relationship between multidimensional perfectionism and insomnia disorder. *Personality and Individual Differences*, *104*, 82-86. doi:10.1016/j.paid.2016.07.042

Two Relevant Missed Studies

Black, J., & Reynolds, W. M. (2013). Examining the relationship of perfectionism, depression, and optimism: Testing for mediation and moderation. *Personality and Individual Differences*, *54*(3), 426-431. doi: 10.1016/j.paid.2012.10.012

Rice, K. G., Ashby, J. S., & Slaney, R. B. (2007). Perfectionism and the Five-Factor Model of Personality. *Assessment, 14*(4), 385-398. doi:10.1177/1073191107303217

Appendix C: IRB Determination Letter

UNIVERSITY of DENVER

OFFICE OF RESEARCH AND SPONSORED PROGRAMS

July 14, 2016

Gabriel Hottinger, Ph.D. Candidate
RE: Determination of Proposed Project
Project Title: The Relationship Between Depression and Perfectionism: A Meta-Analysis

Dear Gabriel,

Thank you for submitting the IRB Determination Form, dated 06/28/16, to the University of Denver Institutional Review Board for evaluation to determine if the above-referenced project qualifies as human subject research. Based on the information provided, it has been determined that the proposed project does not require IRB review. This determination is based on whether this proposed project is research with human subjects defined by the federal regulations.

The IRB Determination Form was evaluated and it was assessed that the proposed project will involve collecting data from published and unpublished research studies and set of meta-analyzes. The data will be used to determine the magnitude of the relationship between each of the 7 dimensions of perfectionism and depression. This proposed project does not meet the regulatory definition of research with human subjects.

<u>The Regulatory Definition of Research and Human Subject</u>
Federal research regulations define **research** as *"a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge."*

During the review of this proposed project, it was noted that the project will collect data from publicly available sources with the intent to generate knowledge that can be applied broadly. This project meets the definition of research.

Per the regulations, **Human subject** means a living individual about whom an investigator (whether professional or student) conducting research obtains 1) data through intervention or interaction with the individual, or 2) identifiable private information. This project will not involve interacting with individuals or obtaining identifiable information. This project does meet the definition of involving human subjects.

In order for a project to require IRB review, the proposed research must qualify under **both** definitions of being research and involving human subjects. This research project does NOT fulfill the regulatory definition of research but does involve human subjects per the federal regulation definition.

My evaluation, based only on the information provided, determined that the proposed project does not require IRB review.

Mary Reed Building, 222 | 2199 S. University Blvd. | Denver, CO 80208-4820 | Main: 303.871.2121 | www.du.edu/orsp

If you have questions regarding this determination or believe that this proposed project does qualify as human subject research, please feel free to contact me directly at 303-871-4049 or via e-mail at: mary.travis@du.edu.

Sincerely,

Mary Travis

Mary Travis
Director, Research Integrity & Education
Office of Research and Sponsored Programs
University of Denver

Appendix D: Characteristics of the Participants in the 52 Studies

Authors	Age Range	Mean Age	Population	Sample Size	Females	Males
Akram, Ellis, & Barclay	19-64	25.3	college students and non-student adults	76	NR	NR
Argus, & Thompson	18-76	43.06	clinically depressed inpatients	141	96	45
Arpin-Cribbie, & Cribbie	18-52	19.73	Undergraduate college students	307	187	120
Athulya, Sudhir, & Philip	NR	21.22	Asian Indian students	192	132	0.6
Bardone-Cone	17-25	18.58	Undergraduate women	426	426	0
Black & Reynolds	NR	27.7	college student and general adults	126	98	28
Blankstein & Lumley	NR	NR	undergraduate college students	61	0	61
Brown & Kocovski,	NR	18.58	socially anxious students	104	72	32
Chang, Hirsch, Sanna, Jeglic, & Fabian	NR	19.78	Latina undergraduates	121	121	0
Chang	NR	NR	European American College students	309	212	97
Chen, Hewitt, & Flett	NR	NR	undergraduate students	120	NR	NR

Authors	Age Range	Mean Age	Population	Sample Size	Females	Males
DiBartolo, Li, & Frost	18-53	20.09	students from an undergraduate women's college	274	274	0
Dunkley, Blankstein, & Berg	NR	20	undergraduate students	357	218	139
Elion, Wang, Slaney, & French,	18-43	NR	African American undergraduate college student	219	105	114
Flett, Besser, Hewitt, & Davis	NR	21.59	undergraduate students (Third year)	202	102	100
Flett, Galfi-Pechenkov, Molnar, Hewitt, & Goldstein	NR	20.3	first year university students at York University	246	155	91
Flett, Besser, & Hewitt	NR	24.23	young community adults	181	91	92
Flett, Nepon, Hewitt, & Fitzgerald	NR	20.5	undergraduate students	191	106	85
Flett, Nepon, Hewitt, Molnar, & Zhao	NR	20.4	undergraduate students	120	86	34
Garrison	NR	19.9	students from a large public university	745	480	262
Gnilka, Ashby, & Noble	18-48	21.2	undergraduate students	180	134	40

Authors	Age Range	Mean Age	Population	Sample Size	Females	Males
Hamamura & Laird	18-46	21.6	college students	126	NR	NR
Harris, Pepper, & Maack	NR	21.21	undergraduate college students	96	67	29
Huprich, Porcerelli, Keaschuk, Binienda, & Engle	18-27	19.64	Undergraduate college students	105	63	42
Iannantuono & Tylka	18-28	19.1	college women	249	249	0
Jain & Sudhir	NR	28.9	clinical sample of Asian Indians diagnosed with social anxiety disorder	30	28	2
La Rocque, Lee, & Harkness	NR	NR	undergraduate university students	503	425	78
Leventhal	18-29	19.9	undergraduate students	145	101	44
Mackinnon, Sherry, Pratt, & Smith	18-25	18.31	first year undergraduate student	127	98	29
Malinowski, Veselka, & Atkinson,	NR	NR	adults & undergraduates	282	NR	NR
Mathew, Dunning, Coats, & Whelan	18-54	21.51	undergraduate students	152	127	25
Merh & Adams	18-32	18.8	college students	358	255	102

Authors	Age Range	Mean Age	Population	Sample Size	Females	Males
Moroz & Dunkley	NR	39.02	community adults	125	85	40
Nepon, Flett, Hewitt, & Molnar	NR	20.7	undergraduate students	155	125	29
Nepon, Flett, & Hewitt	NR	19.9	university students	250	168	82
Noble, Ashby, & Gnilka	NR	NR	undergraduate students	405	270	135
Olson & Kwon	17-28	18.6	undergraduate students	305	227	62
Ozer, O'Callaghan, Bokszczanin, Ederer, & Essau	18-51	22.9	university and college students	402	286	115
Patterson, Wang, & Slaney	18-55	NR	university students and women from an eating disorder treatment center	212	212	0
Rice & Ashby	18-24	NR	undergraduate students	1003	NR	NR
Rice, Choi, Zhang, Morero, & Anderson	20-43	23.37	Asian Indian sample of International graduate students	159	33	133
Rice, Richardson, & Tueller	18-46	19.4	undergraduate university students	340	264	67
Scott	NR	30.1	undergraduate students and community sample	134	104	30

Authors	Age Range	Mean Age	Population	Sample Size	Females	Males
Shanmugam, Jowett, & Meyer	NR	20.95	British athletes	411	252	159
Sherry, Sherry, Hewitt, Mushquash, & Flett	NR	Males 19.26 Females 19.55	undergraduate university students	141	0	141
Smith, Saklofske, Yan, & Sherry	NR	19.8	undergraduate students	425	316	109
Steele, O'Shea, Murdock, & Wade	NR	25.2	females with eating disorders	39	39	0
Sturman, Flett, Hewitt, & Rudolph	NR	19.9	female university students	170	170	0
Wheeler, Blankstein, Antony, McCabe, & Bieling	NR	37	clinical sample and nonclinical volunteers	214	148	66
Wilson, Hunter, Rasmussen, & McGowan	18-65	28.35	college students and non-student adults	338	183	154
Wu & Wei	NR	19.49	undergraduate university students	295	182	113
Yoon & Lau	NR	19.8	Asian American undergraduate students	140	111	29

NR=Not Reporte

Appendix E: VeriCite Dissertation Plagiarism Review Results

Hottinger, Gabriel - G_Hottinger_s_Dissertation_11-2-2017 https://ui.vericite.com/printReport/652bce09-d2c5-4b06-b281-e

VERICITE®

User:	Hottinger, Gabriel
User ID:	11878
Paper ID:	149958519559935
Submitted:	Nov 3, 2017, 9:08 AM
Assignment:	dissertation/proposal plagiarism review
Site:	149958519031609
Site Title:	Independent Research
Plagiarism Score:	10
Matched:	253 sources
Exclude Quotes:	Yes
Exclude Self-Plagiarism:	Yes
Store In Index:	Yes
Report Date:*	Nov 3, 2017, 11:29 AM

G_Hottinger_s_Dissertation_11-2-2017