Comparative Effectiveness of Behavioral Interventions for MUS: A Meta-Analysis

Justin M. Kimber

University at Albany, State University of New York

Dissertation Proposal

Chair: Lisa M. McAndrew, PhD

Committee Member: Michael V. Ellis, PhD

Committee Member: Rachel E. Brenner, PhD

**Introduction**

 Medically unexplained symptoms and syndromes (MUS), also termed persistent physical symptoms, are symptoms without a known cause or symptoms common across multiple physical and mental health conditions (e.g., fatigue). MUS are common with an estimated 25 to 50% of patients seen by primary care providers presenting with MUS (Burton, 2003; Escobar et al., 2010; Hilderink et al., 2013; olde Hartman et al., 2009). Common MUS include fibromyalgia, chronic fatigue syndrome and irritable bowel syndrome.

 MUS are problematic and negatively impact patient quality of life, physical/health-related functioning, and psychological well-being (Duddu, Hussain, & Dickens, 2008). Patients with MUS report higher levels of anxiety and depression, and poorer health-related outcomes than patients with medically explained health conditions (Duddu, Hussain, & Dickens, 2008). MUS patients also have as much as double the healthcare utilization as compared to patients with medically explained conditions (Barsky, Orav, & Bates, 2005; El-Serag, Olden, & Bjorkman, 2002; Maxion-Bergemann et al., 2006). This increased healthcare utilization contributes to an estimated annual one-billion-dollars in healthcare costs in the United States of America (Canavan, West, & Card, 2014; Inadomi, Fennerty, Bjorkman, 2003).

 Behavioral interventions, defined as non-pharmacological interventions targeted at reducing patient-reported symptom severity and improving patient quality of life, are recommended as the first line treatment for MUS (olde Hartman et al., 2017). These treatments, which include cognitive behavioral therapy, yoga, graded exercise, mindfulness/meditation, and acceptance commitment therapy,target various behavioral, cognitive, physiological and emotional factors that are thought to maintain MUS. Multiple behavioral treatments have efficacy for reducing symptoms (i.e., decreased pain severity), improving quality of life and functioning, and increasing psychological well-being (i.e., decreased depression/anxiety; Lackner, Mesmer, Morley, Dowzer, & Hamilton, 2004; Laird et al., 2017; Sánchez et al., 2017).

 Cognitive behavioral therapy (CBT) is the best researched treatment for MUS. In CBT, patients are taught strategies to reduce the cognitive (e.g., catastrophizing) and behavioral (e.g., activity avoidance) factors thought to maintain MUS. Previous meta-analytic reviews report moderate to large effect sizes (i.e., .32 to .80) and odds ratios between 6.01 to 12.00) for cognitive behavioral therapy versus treatment as usual/waitlist control to treat irritable bowel syndrome, somatoform disorders, and chronic fatigue syndrome (Castell, Kazantis, Moss-Morris, 2011; Garg et al., 2016; Malouff et al., 2008) on multiple outcomes including quality of life, health functioning, symptom severity, and psychological well-being (Lackner et al., 2004; Laird et al., 2017; Sánchez et al., 2017). However, effect sizes range substantially, suggesting that heterogeneity of cognitive behavioral interventions might influence the efficacy of treatments for MUS. To examine this, Henrich et al. (2015) conducted a meta-analysis of the techniques most associated with outcomes for CBT and other psychological treatments for irritable bowel syndrome. They found that theoretically targeted-based interventions that include self-monitoring, problem-solving, utilizing general support, and linking symptoms and cognitions are more effective than interventions that did not focus on theoretically driven targets.

 Meta-analyses also find moderate efficacy for multiple other behavioral interventions for MUS. The most common are meditation and mindfulness-based strategies (MMBS). MMBS are defined as mindfulness-meditation and yoga techniques aimed to improve patient depression, anxiety, and somatization. Providers who implement MMBS teach patients to develop metacognitive awareness to address rumination and experiential avoidance that maintains MUS. MMBS have effect sizes ranging from .20 to .62 at improving psychological distress and quality of life (Courtois, Cools, & Calsius, 2015; Lakhan & Schofield, 2013; Lauche et al., 2013). Graded exercise therapy (GET) is also effective at improving outcomes-related to MUS. GET , which teaches patients to slowly increase activity to reduce activity avoidance and to address physical deconditioning that maintains MUS, has moderate to large effect sizes for clinically relevant (i.e., physical functioning & psychological well-being) outcomes for MUS (Castell, Kazantis, & Moss-Morris, 2011; Hayden et al., 2005; Lopez-de-Uralde-Villanueva et al., 2016; Marinko et al., 2011; Smith, Littlewood, & May, 2014; Wang et al., 2012; White et al., 2011). Finally, acceptance commitment therapy (ACT), teaches patients to increase psychological flexibility by connecting to the present moment and to engage in behavior that serves personal values (Hayes, Stroshal, & Wilson, 2009). Meta-analyses examining the efficacy of ACT for MUS show small to moderate effect sizes to treat clinically relevant (i.e., pain severity & psychological well-being) outcomes (Veehof, Oskam, Schreurs, & Bohlmeijer, 2011; Veehof, Trompetter, Bohlmeijer, & Schreurs, 2016). Meta-analyses also find initial support for brief dynamic therapy (effect size= 0.58-0.78; Abbass, Kisely, & Kroenke, 2009), meditative moment therapies (e.g., yoga; effect size=0.49 to 0.66; Langhorst et al., 2013), self-care/management (effect size=0.58 to 0.66; van Gils et al., 2016), music therapy (effect size= 0.55-0.82; Garza-Villareal et al., 2017), and acupuncture (effect size= 0.2 to 0.5; Vickers et al., 2018).

 These meta-analyses suggest that behavioral treatments are efficacious for the treatment of MUS but they do not answer the question of which active behavioral treatments are most efficacious for the treatment of MUS related health outcomes. There have been a few reviews that compared the effectiveness of treatments for MUS. Of these, we are aware of three meta-analyses which compared the efficacy of behavioral treatments to pharmaceutical interventions, and find similar effect sizes for each (Ford et al. 2014; Ford et al. 2009; Lacy et al., 2016). We are aware of two meta-analyses that compared behavioral treatments to each other and these studies show similar efficacy across behavioral interventions. Laird et al. (2017) completed a comparative effectiveness meta-analysis of the relative efficacy of CBT to other psychotherapies for IBS, and found that CBT was equally efficacious as hypnosis, psychodynamic therapy, and relaxation techniques to improve psychological well-being and daily functioning CBT was significantly more efficacious than the relaxation condition at improving daily functioning; Laird cautioned against interpreting the difference between CBT and relaxation due to limited power. Castell et al., (2011) conducted a meta-analysis comparing CBT to graded exercise for chronic fatigue syndrome and found similar effect sizes. Together the similar effect sizes found across meta-analysis for behavioral treatments and the few differences found in comparative effectiveness meta-analyses, suggest that multiple behavioral treatments for MUS may have equivalent efficacy, however, this has never been a comprehensive comparative effectiveness meta-analysis comparing all evidence-based behavioral treatments for MUS.

 The fear-avoidance model (FAM) may explain why seemingly disparate behavioral treatments for MUS have similar efficacy (Lethem, Slade, Troup, & Bentley, 1983). The FAM is a widely used theoretical model to understand how acute symptoms become chronic. FAM suggests that symptoms are created and maintained by a cognitive-behavioral-affective cycle, and that any disruption to the cycle should lead to patient improvement. The cycle begins when a person has fear and a negative appraisal (e.g., catastrophizing) in response to a somatic symptom. Patients’ fear leads to hypervigilance, causing patients to focus their attention on bodily sensations or signals related to MUS. It also results in patient’s avoiding activities that may exacerbate their physicals, reducing quality of life and functioning. Overtime this causes physiological deconditioning and ultimately more symptoms.

 The FAM suggests that intervening anywhere in the cycle will lead to a cascade that improves outcomes. That is, patients can disrupt the cycle by reducing catastrophizing (e.g., mindfulness), increasing activity (e.g., graded exercise, cognitive behavioral therapy), reducing physiological deconditioning (e.g., graded exercise), or increasing functioning (e.g., ACT). By intervening on any of the cognitive emotional, physiological, or behavioral factors maintaining MUS – patients can experience improvements. That is, the FAM suggests that multiple behavioral treatments for MUS may have comparative efficacy.

 Our goal is to compare the effectiveness of behavioral treatment versus no treatment, behavioral treatment versus treatment control, and behavioral treatments (e.g., Cognitive Behavioral Therapy) versus other behavioral treatments (e.g., graded exercise) on MUS related outcomes. Our hypothesis is that multiple behavioral treatments have comparative efficacy for MUS. If true, providers will be able to recommend behavioral treatment based on availability and what is acceptable to patients. This is critical. Both patients with MUS and their providers report difficult relationships, poor satisfaction and often are unable to develop collaborative and acceptable treatment plans (Hubley, Uebelacker, & Eaton, 2016). One way providers may be able to address this is to have multiple efficacious treatment options that meets the needs of patients. Thus, understanding which treatments are most efficacious to treat MUS is important to provide optimal care to patients with MUS

**Research Hypotheses**

 H1. Participants using a behavioral treatment (i.e., CBT) versus no-treatment will have significantly improved psychological distress (e.g., anxiety/depression), disease state (e.g., symptom severity), and well-being/quality of life (e.g., functioning).

 H2. Participants using an behavioral treatment (i.e., CBT) versus control (i.e., TAU/WLC) will conjointly evidence significant improvement across psychological distress (e.g., anxiety/depression), disease state (e.g., symptom severity), and well-being/quality of life (e.g., functioning).

 H3. Participants using a behavioral treatment (i.e., CBT) versus another behavioral treatment (i.e., graded exercises, meditation/mindfulness, or ACT) will be conjointly and equally efficacious across psychological distress (e.g., anxiety/depression), disease state (e.g., symptom severity), and well-being/quality of life (e.g., functioning)

 Additional hypotheses to test specific treatment effects are unknown; however, any significant multivariate finding will be followed up to assess for relative contribution of treatment.

**Method**

**Design**

 The proposed study design will be a meta-analysis to examine the comparative effectiveness of behavioral interventions (i.e., independent variable) for treating commonly health-related outcomes (i.e., dependent variable) among patients with MUS. Based on the current literature of behavioral interventions for MUS, behavioral interventions is operationalized as any psychological or behavioral strategy used to improve the overall quality of life, psychological well-being, or disease state of a patient with MUS. Specific examples of behavioral interventions include: cognitive behavioral therapy, acceptance commitment therapy, graded exercise, yoga, and meditation and mindfulness based strategies. The current literature on MUS typically examines three specific health-related outcomes that include: disease state, psychological well-being, and quality of life/functioning. Disease state is operationalized as a measure that assesses the patient’s severity of symptoms or their overall quantity and frequency of their symptoms related to their MUS. Psychological well-being will be operationalized as stress, anxiety, depression, and other cognitive-affective measures related to a patient’s experience of their MUS. Lastly, quality of life/functioning is operationalized as the degree or extent to which a patient’s MUS interferes with daily activities such as physical, interpersonal, occupational, or daily functioning. Measures that qualify for the proposed study are defined and included in the inclusion and exclusion criteria for the proposed study (see below).

***A priori* Power Analysis***.* An *a priori* power analysis was conducted to determine the necessary number of studies needed to detect an effect (Cohen, 1990). Pairwise comparison alpha for the proposed study is *p=*.05 and beta for power is .80. It is estimated that approximately 50 studies are needed to detect an effect size of .38 per the recommendations of Cohen (1992). An effect size of .38 was determined based on previous literature that reported .38 as the smallest effect size for patients with MUS using behavioral interventions. Meanwhile, Cooper (1991) addresses the complexity of determining power for meta-analysis, mentioning that determining power evolves throughout the research process; thus, the proposed study will evaluate power throughout each step of the research phase, using inclusion, exclusion, and conditional criteria as a guide for research decision-making.As the research process evolves, I will use the power analysis for meta-analysis guidelines by Valentine, Pigott, & Rothstein (2010). Per the recommendations of Valentine, Pigott, & Rothstein, they recommend including at least 40 studies to detect an effect of .35. While the recommendations for sample size differ between Cohen (1992) and Valentine et al. (2010), the proposed study will employ a more conservative approach and aim to include at least 50 studies.

**Inclusion and Exclusion of Studies**

 The following inclusion, exclusion, and conditional criteria were selected based on commonly researched behavioral interventions for medically unexplained symptoms, commonly researched medically unexplained symptoms, and methodological and scientific rigor (Cooper, 2017; Sterne, Egger, & Smith, 2001).

**Inclusion Criteria*.***The proposed study will have the following inclusionary criteria (a) studies must investigate a medically unexplained symptom (e.g., irritable bowel syndrome, TMJ/D, fibromyalgia or chronic fatigue syndrome), (b) participants must be 18 years or older, (c) study design must be a randomized clinical trial or quasi experimental design, (d) behavioral treatment/intervention must be implemented with the intent of improving patients’ MUS across one of the following domains: disease state, psychological distress, and/or psychological well-being QOL, (e) self-report data, and (f) be peer-reviewed. The proposed inclusion criteria is based on the recommendations from the Cochrane Handbook (Higgins & Green, 2011).

**Exclusion Criteria.**The proposed study will have the following exclusionary criteria (a) not a randomized control trial or quasi experimental design, (b) population is chronic musculoskeletal pain, (c) study does not have a control group, (d) cannot be a medication manipulated study, (e) pediatric patient population, (f) the following inclusion criteria above is not met, (g) study is not in English, (h) a systematic or meta-analytic review, (i) an animal study, (j) control group must be treatment as usual, and (k) studies with only ANOVA, ANCOVA, MANOVA, MANCOVA, and multiple regression data because it will limit the usability of data for the proposed study’s analysis (Cooper, 1991, p. 225). Chronic musculoskeletal pain (CMP) will not be included in the initial search terms because CMP is often complex (i.e., acute or chronic pain) and the origin of pain is often identifiable; thus, not fitting as neatly into the current operationalization of MUS.

**Conditional Criteria.** If the proposed study’s current inclusion and exclusion criteria does not yield an appropriately large sample size, additional criteria will be added and include the following terms (a) chronic musculoskeletal pain, and/or (b) a control group can also include waitlist control and education intervention. The conditional criteria was selected because it will increase the sample size for the current study, resulting in increased power for statistical tests.

**Behavioral Interventions.**Broadly, behavioral interventions are operationalized as either a psychological intervention, (e.g., cognitive, behavioral, cognitive-behavioral, acceptance commitment therapy, psychodynamic) (Castell, Kazantis, Moss-Morris, 2011; Hayes, Stroshal, & Wilson, 2009), lifestyle behavior (e.g., diet/exercise) (White et al., 2011), and mindfulness and MMBS (Hayes & Plumb, 2007). Studies that pass the initial identification stage will then be reviewed for methodological rigor per the recommendations of Ellis (1991). A database will be provided at proposal date to display how identification, rating, and criteria will be presented to raters. A general presentation of such information can be found in Appendices A through H.

**Health-Related Outcomes.**Health outcomes were coded into three subcategories (i.e., disease state, psychological distress, and quality of life) representative of MUS outcome research. The proposed study will examine commonly studied health-related outcomes for MUS patients. To date, the MUS literature focuses on three primary health outcomes – disease state, psychological distress, and well-being/quality of life. Coders will identify and rate data related to outcome measures based on the following criteria: a) must be self-report data, b) data is not altered from original state (e.g., continuous to dichotomized), c) data must fall into one of the three outcome variable types (i.e., disease state, psychological well-being, and quality of life).

***Disease State.*** For the proposed study, disease state refers to a self-report questionnaire that measures the progression (i.e., improvement or worsening) of patient’s MUS. Examples of eligible questionnaires that measure disease state are: the IBS-Symptom Severity Scale (Francis, Morris & Whorwell, 1997); Visual Analogue Scale (Bengtsson, Ohlsson, & Ulander, 2007; Bengtsson, Persson, Sjölund, & Ohlsson, 2013). These measures are often adapted across various MUS and used to assess MUS progression throughout RCTs.

***Psychological Distress.*** For the proposed study, psychological distress refers to a self-report questionnaire that measures a person’s mental health symptoms such as depression, anxiety, and mood. To assess psychological distress the following measures are often used: Hamilton Depression Scale (Hamilton, 1986); Hamilton Anxiety Scale (Hamilton, 1959); PANAS (Watson, Clark, & Tellegen, 1988); Beck Depression Inventory (Beck, Steer, & Brown, 1996); Generalized Anxiety Scale-7 (Spitzer, Kroenke, Williams, & Löwe, 2006).

***Well-Being/Quality of Life****.* For the proposed study, well-being/quality of life refers to a patient’s perceived functioning within the context of their MUS. Well-being/quality of life is often used in MUS research, because it illuminates how frequently a patient’s MUS symptoms affect several patient life domains such as social, occupational, and economical contexts. Well-being/Quality of Life is assessed and collapsed into a composite construct of QOL by the following measures: IBS-QOL (Drossman, 2000), Health-Related Quality of Life (Guyatt, Feeny, & Patrick, 1993); Quality of Life (Juniper, Guyatt, Willan, & Griffith, 1994); and EuroQol (Group ,1990).

**Handling Multiple Reports.**For studies that publish multiple articles using the same data in each study, the primary publication will be used for coding procedures. Reasoning for this decision is to eliminate analyzing the same data more than once. Analyzing the same data more than once threatens the study’s overall internal validity.

**Handling Studies with Insufficient Data.**If studies do not contain sufficient information to judge eligibility (e.g., lacking information about study design), or lack sufficient information for analysis (e.g., did not report numerical data about those outcomes), corresponding authors of the study will be contacted. Upon contact, I will request relevant data and study information necessary to fulfill the aims of the proposed study.

**Information Sources**

**Search procedures.**Studies will be identified from Cochrane Central Register of Controlled Trials (CCRCT) (Higgins & Green, 2011), Medline, PubMed, PsycInfo, Web of Science, and Proquest Dissertations and Clarivate Analytics databases. Keywords that will be included in the proposed study are based on the inclusion and exclusion criteria for anticipated articles. These databases were selected based on the Cochrane Handbook (Higgins & Green, 2011), which states that incorporating a wide range of databases helps minimize publication bias and the file drawer effect.

**Contacting Study Authors.**The proposed study will keep a log of dates when corresponding authors were contacted. It is unlikely that the proposed study will contact corporate sponsors or mailing and distribution lists (Appendix G).

**Study Selection**

**Raters***.* Approximately six counseling psychology graduate students (e.g., doctoral/masters) will serve to screen potential eligible articles based on inclusion criteria to be included in the final sample of articles. Each article will be reviewed to screen for inclusion by two coders to ensure quality data control. Raters will be trained to ensure that they understand protocol procedures until they reach 100% agreement on the study’s variables. Coders will use independent and dependent variable operationalizations to assess if a study qualifies for the proposed study; however, if the coder is unsure if the data qualifies, the coder inquire with coding team during weekly meetings. Effect sizes between each outcome variable and independent variable will be extracted and computed into a data entry document. For each outcome measure, central tendency (e.g., mean, median, mode), range, and internal consistency will be collected. Studies that pass the initial identification stage will then be reviewed for methodological rigor per the recommendations of Ellis (1991). A database will be provided at proposal date to display how identification, rating, and criteria will be presented to raters.

 **Resolving Disagreements.**If two coders disagree on study inclusion criteria or coding procedures for a single study, they will bring their disagreement to weekly coding meetings to discuss disagreement with team. Disagreements will be resolved using a majority vote among team members on whether or not data should be included or exclude. Each member will have the operationalization, inclusion, and exclusion criteria to refer to while voting.

**Study Selection.**Approximately six counseling psychology students (e.g., doctoral/masters) will serve to screen potential eligible articles based on inclusion criteria to be included in sample of articles (Appendix A). Each article will be reviewed to screen for inclusion by two coders to ensure quality data control. Discrepancies for coding will be discussed and resolved after weekly meetings with the dissertation writer (see above). Consensus ratings will be calculated to ensure coder reliability and quality.

 The first stage of the search procedure will include gathering eligible articles form academic databases, using pre-determined keyword search terms. It is expected that this stage will generate a sizeable sample size (~16,000), including duplicate articles and study designs that do not meet the proposed study’s criteria. Thus, the six coders will clean the generated database results by deleting duplicate articles and clearly ineligible studies (e.g., a study about rats) based on the title.

 After this initial phase, it is expected that approximately fifty percent of the articles will be removed. Each coder will then go through the remaining study titles and remove articles that clearly do not meet criteria (e.g., a qualitative review) for the current study (Appendix B). It is expected that approximately fifty percent of the articles will be removed after this data cleaning iteration. The remaining studies (abstracts and full texts) will be reviewed and removed if they do not meet the proposed study’s criteria.

 Articles that passed the initial screen were then rated by the six graduate student coders using the criteria detailed below. Each coded study variable was conceptualized and based off of previous research. Quality of research studies will be assessed using quality review details outlined by Ellis, Ladany, Krengel, & Schult (1996), who based his outline procedures from Cook and Campbell (1979) and Wampold, Davis, & Good (1990).

**Data collection**

 **Coder Information.**Coders will be the same six doctoral and master level counseling students asked to exclude duplicate studies and studies that do not meet inclusion and exclusion criteria based on the title, abstract, or body of the main paper (Appendix H).

**Data Collection Procedure.**The following information will be computed or tallied for the proposed study using Cooper’s (2017) proposed data extraction log of intervention data: 1) measure of effect size (e.g., *F*-statistic, cohen’s *d*, odds ratio, semi-partial correlation) between study variables (behavioral intervention or outcomes), 2) sample size, 3) the independent variables (e.g., psychotherapy) and 4) the dependent variables (e.g., disease state, psychological well-being, quality of life) for the observed effect size, 5) measure of internal consistency for dependent variable, 6) the statistical data for each reported statistical test, and 7) the number of statistical tests not reported (Appendix E). In addition, demographic variables (e.g., age, race, education) will be collected and coded to better generalize the proposed study’s findings. Per Cooper’s (2017, pgs. 227 and 228) data extraction recommendations, I will use this information to best estimate effect sizes for each research study. Additional data regarding setting characteristics, study characteristics, outcome measure data, study design, intervention and control type, data collection timepoint, and that will be collected and placed into a data entry sheet per the recommendation of Cooper (2017, p. 126). Behavioral interventions (i.e., independent variable) will only be included in the analyses if they meet the operationalization standards proposed prior to the start of analyses.

 **Assessing Coder Reliability & Resolving Coder Discrepancies***.* The proposed study will ensure data quality (i.e., test for rater drift) by using chi-square testing and inter-rater reliability among the six data coders. Data extraction being tested for quality will be from Ellis’ (1991) outlined meta-analysis data extraction procedures. By using such procedures, data quality is ensured in two specific ways: internal consistency across raters is determined and rater fatigue is controlled, resulting in increased data quality. The dissertation writer will conduct each chi-square test and inter-rater reliability analyses to determine data quality. A quality check will be done after each study elimination phase including; elimination based on title of abstract, the abstract, and careful read through of the study’s methodology and statistical reports.

**Methods for Assessing Validity**

**Minimizing Bias.**The proposed study aims to reduce publication bias (e.g., file drawer effect) by following the Cochrane and Meta-Analysis Reporting Standards (MARS) guidelines (Higgins & Green, 2009; Schalken & Rietbergen, 2017). Specifically, the first author of the proposed study will gather articles from the Cochrane Central Register of Controlled Trials, in addition to Proquest Dissertations and Clarivate Analytics (Higgins & Green, 2011). If selected studies for the data extraction phase do not report required data, the original study’s authors will be contacted for data required to fulfill the aim of the proposed study. To ensure HIPPA compliance, we will only request de-identified data (i.e., without participant numbers or names) to ensure confidentiality of previous participants. By following the Cochrane and MARS meta-analysis guidelines, internal validity of the meta-analysis increases because such procedures provide a methodologically sound framework to address the proposed study’s aims.

**Statistical Summary Measures**

**Effect Size Types.** Per the study aims both cohen’s *d* and squared semi-partials will be collected because they are most commonly reported by researchers. Multivariate analysis, using either cohen’s or squared semi-partials effect sizes based on availability, will be used to test the comparative effectiveness of behavioral interventions on outcomes and MUS. After analyses are conducted, effects will be interpreted in light of data transformation (i.e., cohen’s d and semi-partial) (Cooper, 2017). Furthermore, if *F*-statistics are reported in studies, they will be converted to cohen’s d and semi-partial effect sizes.

**Methods of synthesis**

All analyses will be conducted using the newest version of SPSS. Syntax used to conduct analyses will be available to public upon request. We selected the following synthesis processes based on the proposed study’s aims:

**Random Effects Model.**A random effects model will be used for the proposed meta-analysis. A random effects model assumes that variables selected to be meta-analyzed are conceptually like each other, but that the variables selected are not completely identical. Thus, a random effects model versus a fixed effects model, accounts for the commonalities and differences between interventions. Given the proposed study’s aims, a random effects model appropriately fits the purpose and questions of the proposed study.

 **Descriptive Analysis of Effect Sizes.** Once all of the effect sizes have been calculated, an average effect size and confidence interval that estimates the same comparison or relationship will be conducted (Cooper, 2017, p. 229) for both between and within study. A vote counting procedure will also be used to provide insight into the normality, direction, and magnitude of the effect size data. Broadly, a vote counting procedure entails a comparison of the directionality and number of positive and negative effects found in a meta-analysis. However, these data will only be interpreted within the context of the additional statistical tests of the proposed study because vote counting procedures, though informative, can lead to an overestimation of study effects (Cooper, 1980).

 **Multivariate Analysis & Group Comparisons.**To test hypotheses one through three, MANOVA will be used to compare the effectiveness of behavioral interventions for dependent variables. To assess the relationship between behavioral intervention (X) and disease state (Y) a multivariate multiple regression analysis (i.e., MANOVA) will be used. Preliminary analyses will be conducted in accordance to the ethical guidelines for meta-analysis by Cooper (2017). I will test for outliers, normality, homoscedasticity, multicollinearity, order effects, and internal consistency of measures. All analytic procedures will be in accordance to the general linear model. Each behavioral intervention will have a unique dummy code (e.g., “1” equals “Cognitive Behavioral Therapy”) that will be used for multivariate analysis (e.g., MANOVA). All health outcomes (i.e., dependent variable) will be added to the dataset as continuous data. Additional hypotheses to test specific treatment effects are unknown; however, any significant multivariate finding will be followed up to assess for relative contribution of treatment based on the recommendations of Haase and Ellis (1987).

 **Test of Imprecision.**To test for imprecision of the data, I will estimate the amount of error around findings by using confidence intervals, both within and between studies. Confidence intervals help clarify the magnitude of an observed effect. The test of imprecision will help to interpret and infer findings in the proposed study.

 **Methodological Artifacts.**Hunter and Schmidt (2004) suggest ten distinct methodological artifacts that can shrink effect sizes including: sampling error, error of measurement in the DV, error of measurement in the IV, dichotomization of a continuous dependent variable, dichotomization of a continuous independent variable, range variation in the independent variable, attrition artifacts such as range variation in the dependent variable, deviation from perfect construct validity in the independent variable, deviation from perfect construct validity in the dependent variable, reporting or transcriptional error, and variance due to extraneous factors that affect the relationship. Hunter and Schmidt (2004) operationalize these artifacts as follows:

 ***Sampling Error.***The sampling error in each study is often determined by the study’s sample size, meaning that sample size can affect the direction and magnitude of an observed effect. Sampling error also affects the validity of each study because it creates variation from the population value. One way to minimize sampling error is to conduct a power analysis. The current study will measure sampling artifact by collecting study sample size data (Appendix F) and if a study conducted a power analysis.

***Error of Measurement in the IV and DV.***Errors found in the measurement of the IV and DV is best captured by the overall reliability of a measure. Specifically, if a measure is reliable, there is a higher likelihood that construct and internal validity is sounder than if reliability was low. Measures with internal consistency of .70 and above are considered to have adequate to excellent reliability (Cronbach, 1951). Therefore, the current study will collect data on reported reliabilities of measures to comment on the overall quality of the construct being measured in the proposed meta-analysis.

***Dichotomization.***For our purposes, dichotomization refers to when a researcher takes continuous data (e.g., any value between 1-10) and collapses the data into a restricted range (i.e., 1 equals values between 1-10). When researchers dichotomize data, the point biseral correlation is smaller for the dichotomized data than the correlation from the continuous data. Therefore, the current study will track whether or not researchers dichotomized data. This information will be used to drop effect sizes that are calculated from dichotomized data.

***Variation in the Range of IV.*** This artifact refers to the level of homogeneity among the independent variable. One way to correct range variation is to determine a reference group for the independent variable of the study. The current study will use waitlist control as a reference group against all active treatment groups.

***Attrition Artifacts.***Attrition is directly related to range variation in the dependent variable. This is relevant to the proposed study because if participants dropout of a treatment arm of a study, it can restrict the range of the dependent variable. It is important to conduct an attrition analysis to determine if there are significant differences between treatment completers and dropouts, because if significant differences exist, it might mean that variation in the dependent variable was affected. Therefore, the current study will track if study’s conducted attrition analysis on their data.

***Threat to Construct Validity in the IV and DV****.* Construct validity refers to how well a test or intervention measures what it claims to measure. For the current study, coders will collect data from potentially eligible studies on whether or not the operationalization of the independent and dependent variables (i.e., behavioral interventions and health; see above) matches the proposed study. This information will help ensure that construct validity is being maintained throughout all procedures of the study.

***Computational Errors in Data.***One of the most challenging artifacts to control for are computational errors in data. This refers to errors in the data related to inaccuracy in coding data, computational errors, errors in reading computer output, and typographical errors. Such errors often produce outliers in researcher’s datasets, thus emphasizing the importance of normality testing data. To prevent such errors, previous researchers (Gulliksen, 1986; Tukey, 1960; Wolins, 1962) suggest double checking data to ensure that errors are caught before analyzing and interpreting data. In the proposed study, researchers will track whether or not data extraction and coding was checked before analysis.

***Extraneous Factors Introduced by Study Procedure.***This artifact refers to outside variables that might affect the overall quality of the data. For the proposed study, therapist effects related to administering a behavioral intervention to treat MUS is not directly measured, which might affect potential findings. In the proposed study, I will collect therapist data to account for potential outside effects that contribute to health outcomes.

 To account for such artifacts, the author will collect potential artifact data during the coding and extraction period (Appendix F). This data will be used to report on overall study quality and rigor for studies in the proposed study.

 **Conditional Analyses.** If initial and/or conditional inclusionary criteria yields a disproportionate amount of studies (e.g., sample most representative of IBS patients over other MUS) per the proposed hypotheses, sub meta-analyses might be conducted for each type of MUS; reasoning for doing so is to minimize obscure findings; Cooper 2017, p. 197).

**Testing for****Publication bias and Selective Reporting**

 **Descriptive Data for Contacting Authors.**Data will be collected about how many authors were contacted for data, how many authors provided data, and how many authors that did not respond. A response-rate frequency will be calculated and reported with an overall percentage of response.

**Test for publication bias.**Publication bias affects the overall effects observed in a meta-analysis. That is, publication bias can over or under estimate the overall size of an effect size. To correct for this, a funnel plot will be used in the proposed study. A funnel plot typically displays skew within the data; however, it also allows researchers to see effect sizes against their standard errors (Light & Pillemer, 1984; Sterne & Egger, 2001). Testing for publication bias allows researchers to infer overall quality and inclusivity of studies in their dataset.

**Test for selective reporting.** Selective reporting is defined as research articles that include certain data – typically, data that supports their overall hypotheses. One way to assess for selective reporting is to construct a matrix that depicts outcomes that were recorded in the study and incomplete reporting. This type of matrix will clarify which study authors reported outcomes and how reporting characteristics might shed light on tests for publication bias.

**Sampling Error.**To test for sampling error (e.g., file-drawer effect), a homogeneity of variance analysis will be conducted for each effect size. Procedures used for this analysis will be per the recommendation of Rosenthal and Rubin (1982) and Hedges and Olkin (1986;1992) for the *d*-index effect sizes (Cooper, 2017, 243&244).

**Descriptive Data Selection Process and Included Studies
 Tracking Study Procedures.**At the commencement of the study, I will track the number of potential eligible citations (i.e., total results generated from search), number of unique studies included in the syntheses, reasons for excluding studies at each stage of screening, and creating a table with studies that met many but not all of the inclusion criteria (Appendix C). Furthermore, I will track the frequency of how many studies reported participant demographic data (e.g., sex, racial/ethnic background, age, country, education background, marital status, and physical/mental health comorbidity) and methodological variables such as study setting, method of sampling, and number of treatments. Collecting this information will strengthen overall external validity and improve generalizability of findings at the end of the study.

**Study Sample Characteristics**

 **Study’s Sample Demographic Data.**A select group of approximately six counseling doctoral and master level trainees will extract demographic data from eligible studies (Appendix D). Demographic data of interest include: sex, racial/ethnic background, age, country, education background, marital status, and physical/mental health comorbidity. If this information is not reported, authors of the original study will be contacted for such data. All data will be coded in their reported format (i.e., continuous or percentages). Furthermore, central tendencies for the independent and dependent variables will be calculated and reported as sample characteristics.

**Strengths and Limitations**

Several strengths and limitations are present in the current study proposal. To date, and to the best of the author’s knowledge, the proposed study is the only study to examine the comparative effectiveness of common behavioral interventions for a range of MUS. The novelty of the current study, coupled with a meta-analytic design incorporating primarily randomized clinical trials, allows future MUS research to target and tailor current behavioral treatments for MUS patients. Additionally, the proposed study proposes an understanding of the efficacy of current behavioral interventions to improve MUS patient’s quality of life, disease state, and psychological well-being.

 While strengths of the proposed study exist, limitations are present in the current proposal. Specifically, the meta-analysis is subject to sampling error and bias. It’s plausible that potential fixed-effects are present in the current sample that are unable to be controlled due to meta-analytic sampling procedures. Additionally, despite inclusion of high quality research studies, the current proposal is limited to the theoretical and conceptual understanding of the primary researchers. Similarly, meta-analytic statistical procedures have a series of drawbacks that limit the interpretation and generalizability of findings.

**References**

Abbass, A., Kisely, S., & Kroenke, K. (2009). Short-term psychodynamic psychotherapy for somatic disorders. *Psychotherapy and psychosomatics*, *78*(5), 265-274.

Barsky, A. J., Orav, E. J., & Bates, D. W. (2005). Somatization Increases Medical Utilization and Costs Independent of Psychiatric and Medical Comorbidity. *Archives of General Psychiatry*, *62*(8), 903. doi:10.1001/archpsyc.62.8.903

Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck Depression Inventory–II. PsycTESTS Dataset. doi:10.1037/t00742-000

Bengtsson, M., Ohlsson, B., & Ulander, K. (2007). Development and psychometric testing of the Visual Analogue Scale for Irritable Bowel Syndrome (VAS-IBS). *BMC Gastroenterology*, *7*(1R. doi:10.1186/1471-230x-7-16

Bengtsson, M., Persson, J., Sjölund, K., & Ohlsson, B. (2013). Further Validation of the Visual Analogue Scale for Irritable Bowel Syndrome After Use in Clinical Practice. *Gastroenterology Nursing*, *36*(3), 188–198. doi:10.1097/sga.0b013e3182945881

Burton, C. (2003). Beyond somatisation: a review of the understanding and treatment of medically unexplained physical symptoms (MUPS). *British Journal of General Practice*, *53*(488), 231-239.

Canavan, C., West, J., & Card, T. (2014). Review article: the economic impact of the irritable bowel syndrome. *Alimentary Pharmacology & Therapeutics*, *40*(9), 1023–1034. doi:10.1111/apt.12938

Castell, B. D., Kazantzis, N., & Moss‐Morris, R. E. (2011). Cognitive behavioral therapy and graded exercise for chronic fatigue syndrome: A meta‐analysis. *Clinical Psychology: Science and Practice*, *18*(4), 311-324.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. doi:10.1037/0033-2909.112.1.155

Cohen, J. (1992). Things I have learned (so far). *Methodological Issues & Strategies in Clinical Research*., 315–333. doi:10.1037/10109-028

Cook, T.D., & Campbell, D.T. (1979). Quasi-experimentation: Design and analysis for field settings. Boston: Houghton Mifflin.

Cooper, H. M. (2017). *Research synthesis and meta-analysis: A step-by-step approach*.

Cooper, H. M. (1980). *The Literature review: elevating its status to scientific inquiry*. Center for Research in Social Behavior, University of Missouri-Columbia.

Courtois, I., Cools, F., & Calsius, J. (2015). Effectiveness of body awareness interventions in fibromyalgia and chronic fatigue syndrome: a systematic review and meta-analysis. *Journal of bodywork and movement therapies*, *19*(1), 35-56.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Drossman, D. (2000). Further validation of the IBS-QOL: a disease-specific quality-of-life questionnaire. *The American Journal of Gastroenterology*, *95*(4), 999–1007. doi:10.1016/s0002-9270(00)00733-4

Duddu, V., Husain, N., & Dickens, C. (2008). Medically unexplained presentations and quality of life: A study of a predominantly South Asian primary care population in England. *Journal of Psychosomatic Research*, *65*(4), 311–317. doi:10.1016/j.jpsychores.2008.05.002

Ellis, M. V., Ladany, N., Krengel, M., & Schult, D. (1996). Clinical supervision research from 1981 to 1993: A methodological critique. *Journal of Counseling Psychology*, *43*(1), 35.

Ellis, M. V. (1991). Conducting and reporting integrative research reviews: Accumulating scientific knowledge. *Counselor Education and Supervision*.

El-Serag, H. B., Olden, K., & Bjorkman, D. (2002). Health-related quality of life among persons with irritable bowel syndrome: a systematic review. *Alimentary Pharmacology and Therapeutics*, *16*(6), 1171–1185. doi:10.1046/j.1365-2036.2002.01290.x

Escobar, J. I., Cook, B., Chen, C.-N., Gara, M. A., Alegría, M., Interian, A., & Diaz, E. (2010). Whether medically unexplained or not, three or more concurrent somatic symptoms predict psychopathology and service use in community populations. *Journal of Psychosomatic Research*, *69*(1), 1–8. doi:10.1016/j.jpsychores.2010.01.001

Ford, A. C., Quigley, E. M., Lacy, B. E., Lembo, A. J., Saito, Y. A., Schiller, L. R., ... & Moayyedi, P. (2014). Effect of antidepressants and psychological therapies, including hypnotherapy, in irritable bowel syndrome: systematic review and meta-analysis. *The American journal of gastroenterology*, *109*(9), 1350.

Ford, A. C., Talley, N. J., Schoenfeld, P. S., Quigley, E. M., & Moayyedi, P. (2009). Efficacy of antidepressants and psychological therapies in irritable bowel syndrome: systematic review and meta-analysis. *Gut*, *58*(3), 367-378.

Garg, S. K., Wadhwa, V., Anugwom, C. M., Gupta, N., George, J., Sanaka, M. R., & Sultan, S. (2016). Mo1655 Effectiveness of Pharmacological and Non-Pharmacological Therapies for Irritable Bowel Syndrome: A Systematic Review and Bayesian Network Meta-Analysis. *Gastroenterology*, *150*(4), S744.

Garza-Villarreal, E. A., Pando, V., Parsons, C., & Vuust, P. (2017). Music-induced analgesia in chronic pain conditions: a systematic review and meta-analysis. *bioRxiv*, 105148.

Group, T. E. (1990). EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy*, *16*(3), 199–208. doi:10.1016/0168-8510(90)90421-9

Gulliksen, H. (1986). The increasing importance of mathematics in psychological research (Part 3). *The Score*, 9, 1-5.

Guyatt, G. H., Feeny, D. H., & Patrick, D. L. (1993). Measuring health-related quality of life. *Annals of internal medicine*, *118*(8), 622-629.

Haase, R. F., & Ellis, M. V. (1987). Multivariate analysis of variance. *Journal of Counseling Psychology*, *34*(4), 404.

Hamilton, M. (1986). The Hamilton Rating Scale for Depression. *Assessment of Depression*, 143–152. doi:10.1007/978-3-642-70486-4\_14

Hamilton, M. (1959). The Assessment of Anxiety States by Rating. *British Journal of Medical Psychology*, *32*(1), 50–55. doi:10.1111/j.2044-8341.1959.tb00467.x

Hayden, J. A., Van Tulder, M. W., Malmivaara, A. V., & Koes, B. W. (2005). Meta-analysis: exercise therapy for nonspecific low back pain. *Annals of internal medicine*, *142*(9), 765.

Hayes, S. C., & Plumb, J. C. (2007). Mindfulness from the bottom up: Providing an inductive framework for understanding mindfulness processes and their application to human suffering. *Psychological Inquiry*, *18*(4), 242-248.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2009). *Acceptance and commitment therapy*. American Psychological Association.

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

Hedges, L., & Olkin, I. (1986). Book Review: Meta Analysis: A Review and A New View. *Educational Researcher*, *15*(8), 14–16. doi:10.3102/0013189x015008014

Henrich, J. F., Knittle, K., De Gucht, V., Warren, S., Dombrowski, S. U., & Maes, S. (2015). Identifying effective techniques within psychological treatments for irritable bowel syndrome: a meta-analysis. *Journal of psychosomatic research*, *78*(3), 205-222.

Higgins, J. P., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Vol. 4). John Wiley & Sons.

Higgins JPT, Green S, eds. (2009). Cochrane handbook for systematic reviews of interventions. Cochrane Collaboration

Hilderink, P. H., Collard, R., Rosmalen, J. G. M., & Oude Voshaar, R. C. (2013). Prevalence of somatoform disorders and medically unexplained symptoms in old age populations in comparison with younger age groups: A systematic review. *Ageing Research Reviews*, *12*(1), 151–156. doi:10.1016/j.arr.2012.04.004

Hubley, S., Uebelacker, L., & Eaton, C. (2016). Managing medically unexplained symptoms in primary care: A narrative review and treatment recommendations. *American journal of lifestyle medicine*, *10*(2), 109-119.

Hunter, J., & Schmidt, F. (2004). *Methods of Meta-Analysis*. doi:10.4135/9781412985031

Inadomi, J. M., Fennerty, M. B., & Bjorkman, D. (2003). The economic impact of irritable bowel syndrome. *Alimentary pharmacology & therapeutics*, *18*(7), 671-682.

Juniper, E. F., Guyatt, G. H., Willan, A., & Griffith, L. E. (1994). Determining a minimal important change in a disease-specific quality of life questionnaire. *Journal of clinical epidemiology*, *47*(1), 81-87.

 Lackner, J. M., Mesmer, C., Morley, S., Dowzer, C., & Hamilton, S. (2004). Psychological treatments for irritable bowel syndrome: a systematic review and meta-analysis. *Journal of consulting and clinical psychology*, *72*(6), 1100.

Lacy, B. E., Mearin, F., Chang, L., Chey, W. D., Lembo, A. J., Simren, M., & Spiller, R. (2016). Bowel disorders. *Gastroenterology*, *150*(6), 1393-1407.

Lakhan, S. E., & Schofield, K. L. (2013). Mindfulness-based therapies in the treatment of somatization disorders: a systematic review and meta-analysis. *PloS one*, *8*(8), e71834.

Langhorst, J., Klose, P., Dobos, G. J., Bernardy, K., & Häuser, W. (2013). Efficacy and safety of meditative movement therapies in fibromyalgia syndrome: a systematic review and meta-analysis of randomized controlled trials. *Rheumatology international*, *33*(1), 193-207.

Laird, K. T., Tanner-Smith, E. E., Russell, A. C., Hollon, S. D., & Walker, L. S. (2017). Comparative efficacy of psychological therapies for improving mental health and daily functioning in irritable bowel syndrome: A systematic review and meta-analysis. *Clinical Psychology Review*, *51*, 142–152. doi:10.1016/j.cpr.2016.11.001

Lauche, R., Langhorst, J., Dobos, G., & Cramer, H. (2013). A systematic review and meta-analysis of Tai Chi for osteoarthritis of the knee. *Complementary therapies in medicine*, *21*(4), 396-406.

Lethem, J., Slade, P. D., Troup, J. D. G., & Bentley, G. (1983). Outline of a fear-avoidance model of exaggerated pain perception—I. *Behaviour research and therapy*, *21*(4), 401-408.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, Mass: Harvard University Press.

López-de-Uralde-Villanueva, I., Munoz-Garcia, D., Gil-Martinez, A., Pardo-Montero, J., Munoz-Plata, R., Angulo-Diaz-Parreno, S., ... & La Touche, R. (2016). A systematic review and meta-analysis on the effectiveness of graded activity and graded exposure for chronic nonspecific low back pain. *Pain Medicine*, *17*(1), 172-188.

Malouff, J. M., Thorsteinsson, E. B., Rooke, S. E., Bhullar, N., & Schutte, N. S. (2008). Efficacy of cognitive behavioral therapy for chronic fatigue syndrome: a meta-analysis. *Clinical psychology review*, *28*(5), 736-745.

Marinko, L. N., Chacko, J. M., Dalton, D., & Chacko, C. C. (2011). The effectiveness of therapeutic exercise for painful shoulder conditions: a meta-analysis. *Journal of shoulder and elbow surgery*, *20*(8), 1351-1359.

Michie, S., & Prestwich, A. (2010). Are interventions theory-based? Development of a theory coding scheme. *Health psychology*, *29*(1), 1.

Olde Hartman, T. C., Rosendal, M., Aamland, A., van der Horst, H. E., Rosmalen, J. G., Burton, C. D., & Lucassen, P. L. (2017). What do guidelines and systematic reviews tell us about the management of medically unexplained symptoms in primary care? *BJGP Open*, *1*(3), BJGP–2016–0868. doi:10.3399/bjgpopen17x101061

Olde Hartman, T. C., Borghuis, M. S., Lucassen, P. L. B. J., van de Laar, F. A., Speckens, A. E., & van Weel, C. (2009). Medically unexplained symptoms, somatisation disorder and hypochondriasis: Course and prognosis. A systematic review. *Journal of Psychosomatic Research*, *66*(5), 363–377. doi:10.1016/j.jpsychores.2008.09.018

Maxion-Bergemann, S., Thielecke, F., Abel, F., & Bergemann, R. (2006). Costs of Irritable Bowel Syndrome in the UK and US. *Pharmaco Economics*, *24*(1), 21–37. doi:10.2165/00019053-200624010-00002

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, *74*(2), 166–169. doi:10.1037/0022-0663.74.2.166

Sebastián Sánchez, B., Gil Roales-Nieto, J., Ferreira, N. B., Gil Luciano, B., & Sebastián Domingo, J. J. (2017). New psychological therapies for irritable bowel syndrome: mindfulness, acceptance and commitment therapy (ACT). *Revista Española de Enfermedades Digestivas*, *109*. doi:10.17235/reed.2017.4660/2016

Schalken, N., & Rietbergen, C. (2017). The Reporting Quality of Systematic Reviews and Meta-Analyses in Industrial and Organizational Psychology: A Systematic Review. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.01395

Smith, B. E., Littlewood, C., & May, S. (2014). An update of stabilisation exercises for low back pain: a systematic review with meta-analysis. *BMC musculoskeletal disorders*, *15*(1), 416.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder. *Archives of Internal Medicine*, *166*(10), 1092. doi:10.1001/archinte.166.10.1092

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of clinical epidemiology*, *54*(10), 1046-1055.

Sterne, J. A., Egger, M., & Smith, G. D. (2001). Investigating and dealing with publication and other biases in meta-analysis. *Bmj*, *323*(7304), 101-105.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In I. Olkin, J. G. Ghurye, W. Hoeffding, W.G. Madoo, & H Mann (Eds.), *Contributions to probability and statistics*. Stanford, CA: Stanford University Press.

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215-247.

Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The Study Design and Implementation Assessment Device (Study DIAD). *Psychological methods*, *13*(2), 130.

van Gils, A., Schoevers, R. A., Bonvanie, I. J., Gelauff, J. M., Roest, A. M., & Rosmalen, J. G. (2016). Self-help for medically unexplained symptoms: a systematic review and meta-analysis. *Psychosomatic medicine*, *78*(6), 728-739.

Veehof, M. M., Trompetter, H. R., Bohlmeijer, E. T., & Schreurs, K. M. G. (2016). Acceptance-and mindfulness-based interventions for the treatment of chronic pain: a meta-analytic review. *Cognitive behaviour therapy*, *45*(1), 5-31.

Veehof, M. M., Oskam, M. J., Schreurs, K. M., & Bohlmeijer, E. T. (2011). Acceptance-based interventions for the treatment of chronic pain: a systematic review and meta-analysis. *PAIN®*, *152*(3), 533-542.

Vickers, A. J., Vertosick, E. A., Lewith, G., MacPherson, H., Foster, N. E., Sherman, K. J., ... & Acupuncture Trialists' Collaboration. (2018). Acupuncture for chronic pain: update of an individual patient data meta-analysis. *The Journal of Pain*, *19*(5), 455-474.

Wampold, B. E., Davis, B., & Good, R. H. (1990). Hypothesis validity of clinical research. *Journal of Consulting and Clinical Psychology*, *58*(3), 360–367. doi:10.1037/0022-006x.58.3.360

Wang, X. Q., Zheng, J. J., Yu, Z. W., Bi, X., Lou, S. J., Liu, J., ... & Shen, H. M. (2012). A meta-analysis of core stability exercise versus general exercise for chronic low back pain. *PloS one*, *7*(12), e52082.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. doi:10.1037/0022-3514.54.6.1063

White, P. D., Goldsmith, K. A., Johnson, A. L., Potts, L., Walwyn, R., DeCesare, J. C., ... & Bavinton, J. (2011). Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. *The Lancet*, *377*(9768), 823-836.

Wolins, L. (1962). Responsibility for raw data. *American Psychologist*, 17, 657-658.

**Appendix A**

**Data Extraction/Coding for Initial Screening of Article for Relevance\***

Directions: As you collect the following data, please input the appropriate value (bold number) into the database. If bold value is not applicable (i.e., coding sheet states ‘specify’), please write your finding into the space provided.

1. What is the report ID number? \_\_\_\_\_\_\_\_\_\_\_

2. What is the screener’s name? \_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. What is the date of this screening? \_\_\_\_\_\_\_\_\_\_\_\_

4. What is the first author’s last name? \_\_\_\_\_\_\_\_\_\_\_\_

5. In what year did the document appear? \_\_\_\_\_\_\_\_\_\_\_\_\_\_

6. What type of document is this?
 a. Journal Article **(1)**

 b. Book or Book Chapter **(2)**

 c. Dissertation **(3)**

 d. MA thesis **(4)**

 e. Private report **(5)**

 f. Government Report **(6)**

 g. Conference Paper **(7)**

 h. Other (specify) **(8)** \_\_\_\_\_\_\_\_\_

 i. Cannot tell **(9)**

7. What type of information is contained in this document?

 a. Background **(1)**

 b. Empirical Evidence **(2)**

 c. Both **(3)**

 d. This document is irrelevant **(4)**

8. If empirical, what type of empirical evidence does this document contain?

 a. Descriptive **(1)**

 b. Association or experimental **(2)**

 c. Both **(3)**

 d. Other (specify) **(4)** \_\_\_\_\_\_\_\_\_\_

9. If background, what type of background information does this document contain? (Place a 1 for each item that applies, 0 for each item that does not apply)

 a. Descriptions of program variations

 b. Issues in program implementation

 c. Arguments for and/or against

 d. Review of previous research

 e. Other (specify)

\*Coding sheet recommendation by Cooper (2017)

**Appendix B**

**Report Characteristics\***

Directions: As you collect the following data, please input the appropriate value (bold number) into the database. If bold value is not applicable (i.e., coding sheet states ‘specify’), please write your finding into the space provided.

1. What is the report ID number?
2. What was the first author’s last name? (Enter ? if you can’t tell)
3. What was the year of appearance of the report or Publication? (Enter ? if you can’t tell.)
4. What type of report was this?
5. Journal article **(1)**
6. Book or book chapter **(2)**
7. Dissertation **(3)**
8. MA thesis **(4)**
9. Private report **(5)**
10. Government report (federal, state, country, city) **(6)**
11. Conference paper **(7)**
12. Other (specify) **(8)** \_\_\_\_\_\_\_\_\_
13. Can’t tell **(9)**
14. Was this a peer-review document?
15. Not peer reviewed **(1)**
16. Peer reviewed **(2)**
17. Can’t tell **(3)**
18. What type of organization produced this report?
19. University (Specify)
20. Government entity (Specify)
21. Contract research firm (Specify)
22. Other (Specify)
23. Can’t tell **(99)**

7a. Was this research conducted using funds from a grant or other sponsor?

1. No **(0)**
2. Yes **(1)**
3. Can’t tell **(2)**

7b. If yes, who was the funder?

1. Federal government (specify) \_\_\_\_\_\_\_
2. Private foundation (specify) \_\_\_\_\_\_\_\_\_\_\_
3. Other (specify) \_\_\_\_\_\_\_\_\_\_\_\_

\*Coding sheet adapted and recommended by Cooper (2017)

**Appendix C**

**Descriptive Coding Sheet of Primary Papers Reporting Demographic Data**

Directions: As you collect the following data, please input the appropriate value (bold number) into the database.

1. Did the primary study report sex characteristics of their sample?

 **(0) No**

 **(1) Yes**

2. Did the primary study report gender characteristics of their sample?

 **(0) No**

 **(1) Yes**

 3. Did the primary study report racial/ethnic background information of their sample?

 **(0) No**

 **(1) Yes**

 4. Did the primary study report age mean, age standard deviation, or age range?

 **(0) No**

 **(1) Yes**

 5. Did the primary study report what country the study was conducted?

 **(0) No**

 **(1) Yes**

 6. Did the primary study report the education background of their sample?

 **(0) No**

 **(1) Yes**

 7. Did the primary study report the marital status of their sample?

 **(0) No**

 **(1) Yes**

 8. Did the primary study report physical health comorbidities of their sample?

 **(0) No**

 **(1) Yes**

 9. Did the primary study report mental health comorbidities of their sample?

 **(0) No**

 **(1) Yes**

 10. Did the primary study report the percentage of their sample that had a MUS?

 **(0) No**

 **(1) Yes**

**Appendix D**

**Participant and Sample Characteristics**

Directions: As you collect the following data, please input the appropriate value (bold number) into the database. If bold value is not applicable (i.e., coding sheet states ‘specify’), please write your finding into the space provided.

1. Which of the following medically unexplained symptoms were reported in the study? (Place a 1 in each column that applies, 0 if not, ? if not reported).
2. Fibromyalgia
3. Irritable bowel syndrome
4. Chronic fatigue syndrome
5. TMJ/D
6. Other (specify):

2. What was the reported sex of the sample?

 a. Male **(report %)**

b. Female **(report %)**

c. Intersex **(report %)**

3. What was the reported gender of the sample?

 a. Male **(report %)**

b. Female **(report %)**

c. Transgender **(report %)**

d. Other **(report % and specify)**

4. What was the racial/ethnic background of the sample?

 a. White **(report %)**

b. Black/African American **(report %)**

c. Asian **(report %)**

d. Latino/Latina **(report %)**

e. Other **(report % and specify)**

5. Collect the following age demographic data for the sample:

 a. Mean

 b. Standard deviation

 c. Range

6. What is the highest level of education for the sample?

 a. High school **(report %)**

b. Associates **(report %)**

c. Bachelors **(report %)**

d. Masters **(report %)**

 e. Doctoral/Professional Degree **(report %)**

f. Other **(report % and specify)**

7. What is the current relationship status of the sample?

 a. Single **(report %)**

b. Long-term relationship **(report %)**

c. Married **(report %)**

d. Cohabitating **(report %)**

 e. Widow/Widower **(report %)**

f. Other **(report % and specify)**

8. What is the percentage of current physical health comorbidities?

 a. Report % and specify

9. What is the percentage of current mental health comorbidities?

 a. Report % and specify

**Appendix E**

**Statistical & Effect Size Coding Procedures**

Directions: As you collect the following data, please input the appropriate value (bold number) into the database. If bold value is not applicable (i.e., coding sheet states ‘specify’), please write your finding into the space provided.

1. What was the direction of the effect of behavioral intervention on the achievement measure?
2. Positive **(1)**
3. Negative **(2)**
4. **Non-significant**
5. Information about each experimental group (Note: Leave blank if not reported. *M*= Mean. *SD =* standard deviation.)

 *Intervention Group*

1. Pretest *M* on outcome (if any) \_\_\_\_\_\_\_\_
2. Pretest *SD (*if any) \_\_\_\_\_\_\_\_
3. Posttest *M* on outcome \_\_\_\_\_\_\_\_\_\_
4. Posttest *SD\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*
5. Sample size \_\_\_\_\_\_\_\_\_\_\_

 *Control Group*

1. Pretest *M* on outcome (if any) \_\_\_\_\_\_\_\_
2. Pretest *SD (*if any) \_\_\_\_\_\_\_\_
3. Posttest *M* on outcome \_\_\_\_\_\_\_\_\_\_
4. Posttest *SD\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*
5. Sample size \_\_\_\_\_\_\_\_\_\_\_
6. Total sample size (if not given for each group separately) \_\_\_\_\_\_\_\_
7. Information about null hypothesis significant tests
8. Value of independent *t*-statistic (or square root of *F-*test in one factor ANOVA) \_\_\_\_\_\_\_\_
9. Degrees of freedom for test (in the denominator) \_\_\_\_\_\_\_\_
10. *p*-value from test \_\_\_\_\_\_\_
11. Dependent *t-*statistic \_\_\_\_\_\_\_\_\_\_\_\_
12. Degrees of freedom for test (in denominator) \_\_\_\_\_\_\_\_
13. *p-*value from test \_\_\_\_\_\_\_\_\_\_\_\_\_
14. *F­-*statistic (when included in a multifactored ANOVA)\_\_\_\_\_\_\_\_
15. Degrees of freedom for denominator of *F*-test \_\_\_\_\_\_\_\_\_\_\_\_\_
16. *p*-value from *F-*test \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
17. # of variables in multifactored ANOVA \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
18. Effect Size Estimate
19. What is the metric of the effect size (*d, r, OR, RR, other) \_\_\_\_\_\_\_\_\_\_*
20. Was an effect size calculator used to calculate this effect?

 i. No

 ii. Yes

If yes, what calculator was used?

\*Coding sheet adapted and recommended by Cooper (2017)

**Appendix F.**

**Artifact Data Coding\***

Directions: As you collect the following data, please input the appropriate value (bold number) into the database.

Study ID Number: \_\_\_\_\_\_\_\_\_\_\_\_

*Artifacts relevant to Internal and Construct Validity*

1.1 Were the participants treated in a way that is consistent with the definition of the intervention?

 **(0) No**

 **(1) Yes**

1.1.1 To what extent does the intervention reflect commonly held or theoretically derived characteristics about what it should contain?

 **(0) Not at all**

 **(1) Somewhat**

 **(2) Largely**

 **(3) Fully**

1.1.2 Was the intervention described at a level of detail that would allow its replication by other implementers?

 **(0) No**

 **(1) Yes**

1.1.3 Was there evidence that the group receiving the intervention might also have experienced a changed expectancy, novelty and/or disruption effect not also experienced by the comparison group (or vice versa)?

 **(0) No**

 **(1) Yes**

1.1.4 Was there evidence that the intervention was implemented in a manner similar to the way it was defined?

 **(0) No**

 **(1) Yes**

1.2 Were the outcomes measured in a way that is consistent with the proposed effects of the intervention?

 **(0) No**

 **(1) Yes**

1.2.1 Do items on the outcome measure appear to represent the content of interest to this synthesis (i.e., have face validity)?

 **(0) No**

 **(1) Yes**

1.2.2 Were the scores on the outcome measure acceptably reliable?

 **(0) No**

 **(1) Yes**

1.2.3 Was the outcome measure properly aligned to the intervention condition?

 **(0) No**

 **(1) Yes**

*Artifacts relevant to sampling error and methodology*

2.1 Were the participants in the group receiving the intervention comparable to the participants in the comparison group?

 **(0) No**

 **(1) Yes**

2.1.1 Was random assignment used to place participants into conditions? (If no, answer the next questions [Question 2.1.1a].)

 **(0) No**

 **(1) Yes**

2.1. 1a For quasi-experiments: Were adequate equating procedures used to recreate the selection model?

 **(0) No**

 **(1) Yes**

 **(-99) N/A**

2.1.2 Was there differential attrition between intervention and comparison groups after equating occurred?

 **(0) No**

 **(1) Yes**

2.1.3 Was there severe overall attrition after equating occurred?

 **(0) No**

 **(1) Yes**

2.2 Was the study free of events that happened at the same time as the intervention that confused its effect?

 **(0) No**

 **(1) Yes**

2.2.1 Was there evidence of a local history event?

 **(0) No**

 **(1) Yes**

2.2.2 Were the intervention and comparison groups drawn from the same local pool? (If yes, answer the next question [Question 2.2.2a].)

 **(0) No**

 **(1) Yes**

2.2.3 Did the description of the study give any other indication of the strong plausibility of other intervention contaminants?

 **(0) No**

 **(1) Yes**

*Artifacts relevant to range variation*

3.1 Did the study include variation on participants, settings, and outcomes representative of the intended beneficiaries?

 **(0) No**

 **(1) Yes**

3.1.1 Did the sample contain participants with the necessary characteristics to be considered part of the target population?

 **(0) No**

 **(1) Yes**

3.1.2 To what extent did the sample capture variation among participants on important characteristics of the target population?

 **(0) Not at all**

 **(1) Limited**

 **(2) Reasonable range**

 **(3) Fully**

3.1.3 To what extent did the study include variation on important characteristics of the target setting?

 **(0) Not at all**

 **(1) Limited**

 **(2) Reasonable range**

 **(3) Fully**

3.1.4 To what extent were important classes of outcome measures included in the study?

 **(0) Not at all**

 **(1) Limited**

 **(2) Reasonable range**

 **(3) Fully**

3.1.5 Did the study measure the outcome at a time appropriate for capturing the intervention’s effect?

 **(0) No**

 **(1) Yes**

3.1.6 Was the study conducted during a time frame appropriate for extrapolating to current conditions?

 **(0) No**

 **(1) Yes**

3.2 Was the intervention tested for its effect within important subgroups of participants, settings and outcomes?

 **(0) No**

 **(1) Yes**

3.2.1 To what extent was the intervention tested for effectiveness within important subgropus of participants?

 **(0) Not at all**

 **(1) Limited**

 **(2) Reasonable range**

 **(3) Fully**

3.2.2 To what extent was the intervention tested for effectiveness within important subgroups of settings?

 **(0) Not at all**

 **(1) Limited**

 **(2) Reasonable range**

 **(3) Fully**

3.2.3 Was the intervention tested for its effectiveness across important classes of outcomes?

 **(0) No**

 **(1) Yes**

*Artifacts relevant to reporting error and statistical reporting*

4.1 Were effect sizes and their standard errors accurately estimated?

 **(0) No**

 **(1) Yes**

4.1.1 Was the assumption of the independence met, or could dependence (including dependence arising from clustering) be accounted for in estimates of effect sizes and their standard errors?

 **(0) No**

 **(1) Yes**

4.1.2 Did the statistical properties of the data (e.g., distributional and variance assumptions, presence of outliers) allow for valid estimates of the effect sizes?

 **(0) No**

 **(1) Yes**

4.1.3 Were the sample sizes adequate to provide sufficiently precise estimates of effect sizes?

 **(0) No**

 **(1) Yes**

4.1.4 Were the outcome measures sufficiently reliable to allow adequately precise estimates of the effect sizes?

 **(0) No**

 **(1) Yes**

4.2 Were statistical tests adequately reported?

 **(0) No**

 **(1) Yes**

4.2.1 To what extent were sample sizes reported (or estimable) from statistical information presented?

 **(0) Rarely**

 **(1) Largely**

 **(2) Fully**

4.2.2 To what extent could direction of effects be identified for important measured outcomes?

 **(0) Rarely**

 **(1) Largely**

 **(2) Fully**

4.2.3a To what extent could effect sizes be estimated for important measured outcomes?

 **(0) Rarely**

 **(1) Largely**

 **(2) Fully**

4.2.3b Could estimates of effect sizes be computed using a standard formula (or algebraic equivalent)?

 **(0) No**

 **(1) Yes**

 **(-99) N/A**

\*The following measure is the DIAD which was created by Valentine & Cooper (2008) and addresses methodological artifacts by Hunter & Schmidt (2004)

**Appendix G**

**Contact Primary Author Log\***

Use this log to track correspondences with primary authors about requests for data.

1. Study ID number that data is being requested: \_\_\_\_\_\_\_\_\_\_\_\_\_

2. Who was contacted (researchers name)? \_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. Contact information of researcher (e.g., email)? \_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. Date first contact sent: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

5. Date reply received by contact author? \_\_\_\_\_\_\_\_\_\_\_\_\_\_

6. Nature of reply by author: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**APPENDIX H**

**Coder and Coding Characteristics**

Please use the following information whenever you start to code, input, or extract data for any study in the final sample.

1. What is your coder ID number? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. On what date did you complete coding this study? MM/DD/YYYY

3. In minutes, how long did it take you to code? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. Did you code the article first or second? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Notes (provide below any notes about the study or concerns you had regarding your codes):

\*Coding sheet adapted and recommended by Cooper (2017)