RESEARCH DESIGN AND EFFECT SIZE:

A META-ANALYSIS OF MOOD DISORDER EXPERIMENTAL TRIALS

by

LESLEY DHALIWAL

B.A., The University of British Columbia, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Measurement, Evaluation and Research Methodology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2015

Abstract

The design of experimental studies can have a significant influence on effect size; however, this influence is rarely given enough consideration during the interpretation and comparison of research results. This paper examines whether there is a significant difference between the effect sizes from placebo-controlled versus treatment-controlled trials. This issue was studied by conducting a meta-analysis of approximately 37 RCTs of mood disorder therapies. The results of this methodological investigation confirmed that there is a statistically significant difference between the weighted effect sizes from the two groups of studies that were compared. These results support the claim that the type of control group is an important factor to consider in the design and interpretation of experimental studies. This analysis is a methodological contribution as it addresses how the type of control group in a RCT impacts the outcome of a study, and more specifically the effect size. The outcome of this research also challenges the effectiveness of treatments that have been tested against only one type of control in experimental studies.

Preface

This thesis is original, unpublished and independent work by the author Lesley Dhaliwal.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I am very grateful to the faculty and staff at the University of British Columbia. I owe a special thanks to Dr. Bruno Zumbo, Dr. Nand Kishor and Dr. Sandra Mathison whose expertise, understanding and patience, added greatly to this thesis and my graduate experience.

Thank you to my parents and husband, who have supported and encouraged me throughout my years of education.

To my Mother

1   Introduction

There has been an ongoing debate in research methodology literature as to whether a placebo or therapy is a better choice for a control group in experimental studies and a strong case has been made for the use of both (Castro, 2007). This meta-analysis will investigate whether there is a significant difference in the effect size between placebo-controlled and treatment-controlled randomized trials (RCTs). The issue of control type is especially controversial when testing the efficacy of treatments for depression and bipolar disorder and, as such, mood disorder research has been chosen as the context for exploring this methodological issue (Ellenberg & Temple, 2000). More specifically, experimental studies of mood disorder treatments will be examined to address this research question. According to the Diagnostic and Statistical Manual of Mental Disorders, mood disorders are the following: major depressive episode, major depressive disorder, dysthymic disorder, bipolar episode, bipolar disorder and other psychiatric conditions (4th ed., text rev.; DSM-IV-TR; American Psychiatric Association, 2000).

It is hypothesized that there will be significant difference in the effect sizes from the two groups of studies being examined and more specifically, that the effect sizes from trials using a placebo control will be significantly higher than those with a treatment control.

1.1   Importance of Research

According to Sneed et al., "Randomized controlled trials (RCTs) are the experimental gold standard for establishing evidence of treatment efficacy" (2008, pg. 65). As clinicians and researchers synthesize data from RCTs to make conclusions about the effectiveness of medications and interventions, it is important that resources are spent on ensuring that RCTs are conducted according to best practices (Rutherford, Sneed & Roose, 2009a). Although various methodology guides exist, the application of principles is not straightforward and researchers

still disagree about the situations in which each control group type is acceptable (Ellenberg & Temple, 2000). The controversy around the use of control type is mostly based on ethical concerns and the potential impact of inactive treatments on patient health, however, the issue of how control type may impact research outcomes has not been addressed as a concern. This meta-analysis will investigate this impact and by doing so will shed more light on one of the ways in which research design influences outcomes in experimental research.

There is a volume of research that examines how research design and methodology can influence statistical power and effect size in randomized controlled trials (RCTs). It has been shown that research design aspects such as the dosing schedule, patient gender ratio and number of treatment arms are associated with treatment response outcomes (Khan, Kolts, Thase, Krishnan, & Brown, 2004). Researchers have also shown that the type of control used in RCTs can moderate the treatment effect size. It is common for subjects in treatment-controlled trials to experience a greater response to anti-depressant treatment than subjects in placebo-controlled trials (Rutherford et al., 2009b; Sneed et al., 2008).

All treatments include nonspecific factors in addition to the specific effects of drugs; non-specific factors refer to shared aspects of most psychotherapies that include but are not limited to the therapeutic alliance, experimental competence and adherence to treatment protocols (Chatoor & Krupnick, 2001). It is thought that study design may influence patient outcomes via nonspecific factors (Rutherford et al, 2009b). One of these non-specific factors is patient expectations; subjects in treatment-controlled trials have a higher expectation of experiencing improvement and also expect to experience a higher magnitude of improvement than those in placebo-controlled trials (Rutherford et al., 2009b).

This analysis will provide valuable methodological insights around the design and use of control groups in experimental research in the area of mood disorder research. Specifically, showing that there is a significant difference in the effect size between experimental studies that use different controls would support the idea that consistency in the control group treatment is important for the valid comparison of effect sizes among different RCTs of the same drug and also for the valid comparison of different treatments for mood disorders.

Rutherford et al. conducted a similar meta-analysis looking only at anti-depressant clinical trials (2009a). This research showed that that anti-depressant trials with active controls resulted in significantly higher odds of response and remission. One of the goals of this research is to provide support for the findings of research conducted by Rutherford et al. and to improve upon the methodology by widening the scope of this meta-analysis to all mood disorders treatment trials and including non-pharmaceutical trials. By doing this, the results will contend the notion that the impact of control group type is exaggerated in anti-depressant pharmaceutical trials. The aforementioned methodological changes and including studies from a more recent time frame in this analysis will ensure that there is not overlap with the meta-analysis conducted by Rutherford et al. It should be noted that the trade-off of widening the scope of the research and improving the generalizability of the results, is that there is higher variability between the studies included in the meta-analysis.

This research may also provide support for the notion that when there are considerable differences in the design and methodological characteristics of the research, using the effect size as the only indicator of treatment strength may be problematic (Fern & Monroe, 1996).

Subsequent research in this area that establishes the connection between research design and treatment response and that identifies patient recovery expectations as the link between

design and outcome may lead to the development of methods and best practices used to control for subject expectations. Such methods may include the design and use of reliable and valid pre-treatment measures of recovery expectations (Krell, Leuchter, Morgan, Cook, & Abrams, 2004; Mondloch, 2001). If subjects in treatment-controlled trials are found to experience a greater response to treatment, it will provide support for a revaluation of current informed consent practices in placebo-controlled studies. More specifically, it will show that there is a need for subjects in placebo-controlled trials to be made aware of the risk of experiencing a weaker response to treatment than subjects in treatment-controlled trials.

1.2   Potential Outcomes and Significance of Research

A meta-analysis will be conducted to address how control group types influence the analysis and comparison of experimental studies in the area of mood disorder treatment research. If the results show that there is no significant difference in the weighted effect size between studies that administer a treatment control versus those that administer a placebo control, then it will provide support for the argument that a consistent control group type is not an important factor when comparing effect sizes. In this case, this research will help show that control group type does not significantly skew the results or influence the validity of research in this area.

Alternatively, if the research shows a significant difference in the weighted effect sizes between those studies with an treatment and placebo control as hypthesized, it will suggest that the type of control group has a significant influence on the effect size of experimental research that tests mood disorder treatments. Not only will this outcome provide support for the recommendation that there should be consistency in the type of control group when comparing the results from randomized trials, but it will also dispute the validity of trials in which significant effect sizes have been achieved using only one type of control. The implications of

additional meta-analyses in this area that support these results could result in a shift in focus from effect size to alternative and non-traditional qualitative and quantitative research being used as indicators of treatment strength.

1.3   Background

There are numerous aspects of research design that are important to consider when interpreting and comparing results. It has become the goal of some researchers to uncover whether there is an advantage to using a treatment control versus a placebo control and if the use of one over the other improves the validity of a study (Quitkin, Rabkin, Gerald, Davis, & Klein, 2000). Research in this area shows support for the idea that there is a strong connection between research design and effect size and suggests that to improve the validity of a study, it may be advantageous to utilize particular research design and methodological practices (Quitkin et al., 2000; Thase, 1999).

1.3.1   Treatment-Controlled Trials and Placebo-Controlled Trials

There are many studies that examine the validity of randomized trials and the best practices for research design and methodology. Unfortunately, some researchers do not make it a priority to adhere to these best practices particularly during sample selection. According to Thase (1999, pg 23), "too little attention is given to ensuring the reliability of diagnoses and dependent measures, [and] sample sizes are seldom large enough to detect modest yet honestly significant differences [between the treatment and control group]". When researchers do not make it a priority to follow best practices and fully understand the dose-response characteristics of the treatment, it can result in too many trials being pursued (Thase, 1999).

The inclusion of a control group in the research design is outlined as a best practice in the International ethical guidelines for biomedical research involving human subjects (2002). Castro

(2007) suggests that there are scientific as well as ethical reasons for including a control group in experimental trials and provides support for why best available therapy should be administered to the control group when testing new therapies or when testing interventions against existing therapies. Some may argue that there are societal benefits that outweigh the risks associated with delaying treatment, however, there is likelihood that the illness may exacerbate or result in death when subjects are withheld from receiving treatment (Castro, 2007; Kim, 2003). Even the most active supporters of placebos recognize that the use of placebos is unethical if they result in an increased risk of death or worsening of the illness (Kim, 2003). In the case of major mental illness, a lack of a treatment can have detrimental effects that are not limited to the risk of suicide (Kim, 2003). The state of being depressed or psychotic itself can cause immeasurable suffering and as such, the ethical concerns associated with delaying or not providing treatment should not be considered insignificant (Castro, 2007; Kim 2003; Thase, 1999). As stated by the World Medical Association, "a placebo controlled trial should only be used if there is an absence of existing established therapies" (Castro, 2007).

To contend the notion that delaying or withholding treatment from patients with major depression can increase the risk of suicide or result in long-term toxic effects, several analyses of suicide rates have been conducted (Kim, 2003). The results show that, even when controlling for duration of exposure, there are no significant differences between the suicide or suicide attempt rates for those exposed to investigational drugs, active controls and placebos (Kim, 2003; Khan, Warner & Brown, 2000). It should be noted that these analyses were specific to depression and cannot be generalized to all fields of study.

In terms of treating an illness, the purpose of placebo-controlled trials and comparator trials differ. Placebo controlled trials are designed to determine if a medication is effective in

treating an illness, however, comparative trials are designed to determine if the medication under question is superior to the control treatment (Sneed et al., 2008). One of the challenges with treatment-controlled trials is that using the standard that requires the investigational drug to be superior than available treatments is too difficult and can result in some benefits not be realized (Kim, 2003). Such benefits could include higher efficacy in the long-run or fewer toxic effects. With this 'better than active control' standard in place, developing and approving new treatments becomes more difficult (Kim, 2003). Using a treatment-control instead of a placebo-control also has advantages as it gives researchers the ability to compare dose-response characteristics of the treatment under question instead of looking to other studies to make such comparisons. When the comparison of drug efficacy and dose-response characteristics is made across studies, there is a risk that there will be significant differences in the non-specific factors or limitations in the research design and methodology of one of the comparative studies. Such differences and limitations often have implications for the validity and reliability of the results. A recent study conducted by Jan and Shieh (2011) focused on uncovering the optimal sample sizes required for adequate statistical power. Their results confirmed that comparing means from studies with inadequate sample size is problematic which supports the notion that treatment outcomes should only be compared when the treatments have been tested using similar research design under the same conditions (Jan & Shieh, 2011).

Despite these and many other similar findings, best practices for research may not be adhered to for various reasons. It is often the case that ideal research design is not attainable as budgetary constraints must be taken into account during the research design process (Jan and Shieh, 2011). These types of cost concerns can complicate study design and often trade-offs must be made between aspects of study design such as study length, type of control and sample size

(Jan and Shieh, 2011). For example, under budgetary constraints, researchers that aim to study the effect of a particular pharmacotherapy may have to choose between selecting a smaller sample size or administering a less costly therapy to the control group. One of the benefits of placebo-controlled trials is that they have a higher statistical power and can therefore be conducted with fewer patients than active control trials (Castro, 2007). This can be explained by the high level of treatment separation in placebo-controlled studies, which results in an increased likelihood of detecting the effects of the treatment (Castro, 2007). According to Castro (2007, p. 571) the ethical implications of using a smaller sample size is that "fewer subjects are exposed to toxic or ineffective treatments" (Castro, 2007).

1.3.2 Relationship Between Design and Response

The study of the relationship between design and response has largely been in the area of mental illness and more specifically depression. Studies that examine the design of late-life (over 50 years of age) depression RCTs provide evidence of a relationship between study design and response rates. Results show that the antidepressant response rates in comparator trials were between 20% and 30% higher than antidepressant response rates in placebo-controlled trials (Roose & Schatzberg, 2005; Sneed et al., 2008). The results of one meta-analysis revealed significant variability in response rates of outpatients with late-life depression and 27% of this variability was attributed to the study type (placebo-controlled vs. treatment-controlled) (Sneed et al., 2008).

The differences in antidepressant response rates are not limited to the results of late-life depression RCTs. Rutherford et al. have conducted extensive research on the relationship between research design and treatment response in adult patients between 18 and 65 years of age (2009a). Their literature review examined randomized controlled trials of antidepressants and

was aimed at uncovering whether design aspects influence the response to the treatment being tested (Rutherford et al., 2009a). Using mixed effect logistic regression models, their aim was to determine whether the non-specific factors, such as the type of control and study duration, affected the response and remission rates in patients with major depression. All trials compared an antidepressant to a placebo or another antidepressant. Ninety trials were analyzed and it was determined that comparative studies, in which a drug was administered to the control group, resulted in significantly higher odds of response and remission. Although this literature review was very extensive in terms of the number of studies examined, some may argue that the impact of control group type is exaggerated in antidepressant trials.

One possible explanation for the difference in response rates between treatment-controlled trials and placebo-controlled trials are expectations (Sneed et al., 2008). The association between the type of control group and treatment outcome can be explained by subject expectations. In experimental trials, a subject's knowledge about the research design can influence their expectations about the response they will have to the treatment (Rutherford, 2009b; Krell et al., 2004). Subjects participating in a comparative study know that they will receive a treatment regardless of whether they are placed in the treatment group or control group (Sneed et al., 2008). However, subjects in placebo-controlled trials know that it is possible they will be administered a placebo. Sneed et al. point out that, although patients in placebo-controlled trials may experience hope for improvement, they will also experience doubt (2005). Doubt regarding whether one will receive an active medication can lead to lower treatment response expectations and in turn a decreased response to treatment (Sneed et al., 2008). The expectations about treatment outcome can also be attributable to one's personal attitudes regarding the condition being treated, personal experiences with alternate treatments and type of

treatment being tested (Krell et al., 2004). According to Krell et al., the condition or illness itself can result in changes to one's mood, personality and psyche that may shape expectations of treatment outcome (2004). Together, design and unique patient characteristics play a role in recovery expectations and may even interact with one another causing the influence of one to be more or less pronounced.   For example, subjects who have had positive experiences with treatments in the past when participating in placebo-controlled trials may attend less to research design aspects. In other words, positive expectations resulting from previous treatment success may offset the doubt experienced by those in placebo-controlled trials.

Research conducted by Rutherford et al. on the relationship between study design and participant expectations provides an explanation for the results of their literature review; why the type of control group affects the response and remission rates in patients. Rutherford et al. administered a questionnaire to thirty-seven undergraduate students in an introductory psychology course (2009b). The questionnaire described two hypothetical clinical trials that were testing the same drug for the treatment of a skin rash, but one of the trials was placebo-controlled and the other was treatment-controlled. Participants were asked to rate their expectations of the improvement they would experience if they participated in each trial without knowing what treatment group they would be assigned to. The results showed that participants in clinical trials have higher expectations of improvement when they know they could receive an active treatment in the control group as opposed to a placebo. Unlike much of the research on this topic which has been in the area of depression, this study provides support for the link between research design and patient expectations in a more general context.

In the attempts to directly test the link between study design and patient expectancy, Rutherford, Sneed, Eisenstadt, and Roose conducted an eight week RCT in which out-patients

with major depressive disorder (MDD) were randomized and placed in a placebo-controlled or comparator group (2010). Subjects were informed of their group assignment but were blinded to their specific treatment assignment. Those in the placebo-controlled group were told that they have a 50% chance of being administered an anti-depressant medication and a 50% of being administered a placebo. Subjects in the comparator group were told that they would be receiving an antidepressant drug whether they were assigned to the treatment or control group. Before and after randomization, subjects in each group reported on their expectations of the likelihood that they would experience improvements and their expectations of the magnitude of these improvements. A regression model was fit to the data in order to determine whether subjects in the comparator group reported greater expectations of improvement. Results from the regression analysis showed that the effect of group assignment on the magnitude of improvement were significant, however, there were no group differences in the expected likelihood of improvement. In other words, the expectation of how likely one was to experience an improvement in depression was the same among both groups, however, subjects in the comparator group expected a significantly higher magnitude of improvement. These results support the hypothesis that trial design influences participant expectations of improvement.

The predictive relationship between recovery expectations and health outcomes has been supported by many research findings and is exhibited by the placebo-effect (Mondloch, 2001). In other words, expectations have a central role in mediating this effect. Geers and Rose defined the placebo effect as "changes resulting from an individual's interpretations and expectations produced by therapeutic activities" (2011, p. 735). Expectations are future-oriented beliefs that facilitate one's response and adjustment to the environment (Geers & Rose, 2011). In patients, they can result in psychological responses that in turn influence health outcomes (Mondloch,

2001). These responses can vary from having motivation to experience improvement and gaining knowledge about the condition/illness (Mondloch, 2001). According to Flood, Lorence, Ding, McPherson and Black, there are five mechanisms by which expectations can affect health outcomes; "triggering of a physiologic response, acting to help motivate patients to achieve better outcomes, conditioning the patient psychologically to observe certain types of symptoms and ignore others, changing the understanding of the disease or acting in concert with anxiety to heighten or reduce symptoms" (1993, p. 1044). Expectations have the power to cause individuals to attend to or overlook particular sensations or symptoms, interpret sensations that may have been ambiguous in the past as being negative or positive and attribute these sensations to the treatment they are receiving (Barsky, Saintford, Rogers & Borus, 2001; Mondloch, 2001).

The role of expectations has been shown to be important to the experience and recovery of individuals with various illnesses and even disabilities (Cole et al., 2002).   Positive expectations are associated with better treatment outcome and this association is strongest in subjects that are suffering from pain (Krell et al., 2004; Mondloch, 2001. In patients who suffer from chronic pain, expectations about pain severity can moderate pain tolerance for future episodes of pain and expectations regarding recovery can influence the duration of pain (Cipher, 1997; Cole, Mondloch & Hogg-Johnson, 2002). Research shows that higher recovery expectations explain why patients in treatment-controlled trials have significantly higher odds of response and remission (Rutherford et al., 2009b). There is also evidence of negative patient expectations having an influence on treatment outcome. In their study on non-specific medication side effects, Barsky et al. found that patients who expected negative side effects before taking a medication were more likely to develop these side-effects (2001).

Patient expectations have not only been shown to correlate with clinical outcomes, but have also been shown to be a predictive factor of these outcomes (Vaz-Leal, 1989; Krell et al., 2004). In the case of major depressive disorder, Krell et al. describes the expectations of patients as "modulating the rate of response to antidepressant medication" (Krell et al., 2004, page 1174). To establish the association between expectation and response, Krell et al. conducted a single-blind experimental antidepressant treatment study; it was found that subject expectations were the only variable that accounted for a significant proportion of the variance in depression scores (according to the HAM-D scale) (2004). Analysis of the results also revealed a significantly higher level of response and lower depression scores among subjects who reported an expectation of treatment effectiveness versus those who reported an expectation that the treatment would be somewhat effective (Krell et al., 2004). The influence of recovery expectation on outcome and the strength of this association has been repeatedly been shown to be independent from the effects of psychological, physiological or social variables (Mondloch, 2001).

The response that subjects have to a drug is thought to be the result of three types of effects; pharmacological, conditioning and expectation (Volkow et al., 2006). According to Volkow et al., the response to drugs is affected by the pharmacological treatment itself as well as "expectation which is in turn sensitive to prior drug experiences" (2006, p. 1782). Their research shows that expectations not only affect responses to treatment but also to placebos. In their study, when subjects expected to receive a drug (for the treatment of Attention Deficit Disorder (ADD)) but instead received a placebo it was found that they experienced increased activity in the brain regions involved with emotional reactivity and reward. This effect was especially evident in subjects who had no prior experience with the drug they were expecting to be

administered. This research provides evidence that expectation modulates the reinforcing and therapeutic effects of drugs in subjects that may or may not have been administered the drug in the past (Volkow et al., 2006).

Although the research findings provide a strong case for the key role that patient expectations have in moderating response rates in placebo-controlled and treatment-controlled trials, there are several other possibilities that exist as well. There may also be differences in the design of placebo-controlled trials and treatment-controlled trials that contribute to a higher response rate among patients. Some of the design features that have been found to significantly differ between successful and unsuccessful trials are the dosing regimen, the number of treatment arms and patient gender (Khan et al., 2004). It may also be the case that in pharmaceutical clinical trials, the impact of design is stronger than it is in trials that test non-pharmaceutical treatments. Further research must be done to determine whether there are differences in these specific design features in studies with a placebo-control and a treatment-control and whether the type of treatment being tested or the illness being treated moderates this impact. Much of the research that has investigated the relationship between patient expectations and outcome has been in the context of depression.

Systematic reviews of trials provide evidence that many controlled trials are methodologically weak and that these weaknesses result in biased findings (Juni, Altman & Egger, 2001). The conscious and sub-conscious attitudes and expectations of patients, doctors and raters concerning patient improvement can also result in biases that contribute to differences between these trials that, in turn, affect response rates (Sneed et al., 2005). There are four types of biases in experimental trials that result in the distortion of statistical analysis: selection bias, performance bias, detection bias, and attrition bias (Juni, Altman &

Egger, 2001). It may be the case that trial design aspects like the type of control group or the severity of the participant illness in a study may indirectly pronounce any one of these biases. The method used to collect samples and allocate participants to treatment and control groups can result in a selection bias; this bias occurs when there are differences in the groups that are compared that result from errors in choosing the individuals or groups to take part in a study (Juni, Altman & Egger, 2001). This bias may account for some of the difference in response rates between placebo controlled trials and treatment controlled trials. When there is inadequate or unclear concealment of treatment allocation and researchers' conscious and subconscious attitudes can play a part in patient recruitment and in the allocation of participants to each group, the treatment effects can be exaggerated (Juni, Altman & Egger, 2001, Quitkin et al, 2000). Odds ratios from trials with inadequate or unclear concealment were on average 30% lower (more beneficial) than those from trials with adequate methodology (Juni, Altman & Egger, 2001). In the context of anti-depressant trials, the differences in the type and severity of depression among subjects in each type of study may also cause subjects to have different expectations about their response to treatment and in-turn affect actual response rates. It has also been shown that the severity of depressive symptoms can be significantly associated with a difference in response to antidepressants and placebos (Khan et al., 2004). For example, it is possible that the type of depression among subjects that sign up for and are chosen for comparator trials is on average more severe and recurrent than those who enter placebo-controlled trials (Sneed et al., 2005). Randomization protects against the selection bias; it is important that the sample for the study is randomly chosen and that participants are randomly assigned to the treatment and control groups.

Blinding of treatment recipients and providers is used to protect against the performance bias; this bias occurs when there are differences in the care received by patients (Juni, Altman &

Egger, 2001). The conclusions that are drawn from the results of a study may be incorrect if the performance bias is not taken into account. Some patients may experience improvements or symptoms that are a result of other treatments or therapies they are receiving outside of the trial (Juni, Altman & Egger, 2001; Vaz-leal, 1989)). It may be the case that the performance bias is more pronounced in placebo-controlled trials versus treatment-controlled trials, or vice versa and this may account for some of the difference in effect size between both types of studies. For example, patients who are participating in a placebo-controlled trial may be more apt to seek treatment outside of the study because they know they may be assigned to the placebo group and therefore, may not be receiving any treatment in the study.

The detection bias results when there are differences in the observed and reported incidences of the illness being treated among a certain group of subjects (Juni, Altman & Egger, 2001). Depending on the illness or condition, a particular gender or age group may be more inclined or have a higher tendency to report their symptoms (Cole et al., 2002). There could be differences in reporting of symptoms among participants in placebo-controlled and treatment-controlled trials. For example, participants who know they may have received a placebo may be less likely to report an improvement in their symptoms because they may be under the impression that any improvements they have experienced are a result of a placebo effect (Flood et al., 1993; Juni, Altman & Egger, 2001). Further research needs to be done to determine if the detection bias plays a role in this difference between the response rates among placebo and treatment controlled trials.

When patients are excluded from treatment after they have been allocated to treatment groups, an attrition bias may be introduced to the trial (Juni, Altman & Egger, 2001; Weissman et al., 2012)). There are several possible reasons that patients may be excluded from the trial,

these include deviations from protocol such as not adhering to treatment or treatment schedule and a loss to follow up after treatment because patients are not available after treatment (Heo, Papademetriou, & Meyers, 2009; Juni, Altman & Egger, 2001). An attrition bias occurs when the patient excluded from the trial is not representative of patients that followed study protocol and remained in the trial (Juni, Altman & Egger, 2001). It is often the case that those patients excluded from the study did not adhere to the study protocol due to their prognosis (Juni, Altman & Egger, 2001). Their prognosis can include severe side effects or increased severity of the illness. In clinical trials of patients with depression, high rates of attrition have been reported (Weissman et al., 2012). In a review of antidepressant randomized trials with elderly depressed patients, there were a total of 8, 385 subjects and 27.3% of these subjects were terminated early (Heo et al., 2009).

Research shows that various study-level and group-level design factors such as the severity of symptoms, use of a placebo arm, number of arms, unbalanced treatment allocation and study duration can account for significant differences in attrition rates (Heo at el., 2009; Weissman et al., 2012). Heo et al. have found that higher attrition rates are significantly associated with groups receiving an active antidepressant treatment as opposed to a placebo (2009). In order to correct for the attrition bias, patients should be included in the analysis of the study results despite not adhering to the study protocol (Juni, Altman & Egger, 2001).

Sneed et al. suggest that further research is required to determine if there may also be differences in key factors such as research design and prognosis of patients between studies conducted in North America and Europe (2008). Currently there is some literature that shows that there are key differences in the prognosis of study participants and outcomes of cross-cultural research (Strauss, 2007). Various factors such as genetics, environment and culture are

17

thought to be responsible for these differences (Struass, 2007. Not much research has been done to study these specific cross-cultural and social differences as researchers have made it a priority to first understand whether there are cross-cultural differences in the prognosis of disorders. The very fact that prognosis and measurement is often difficult when doing cross-cultural studies is evidence that there are indeed differences (Strauss, 2007).

The goal of the International Pilot Study of Schizophrenia was to determine whether a disorder like schizophrenia actually existed in other cultures or if there were cultural variations at the syndrome level (Strauss, 2007). Researchers encountered issues with sampling, epidemiology and patient functioning when trying to reach cross-cultural conclusions, indicating that future research in this area will not be straightforward. It was concluded that there are definite cross-cultural differences in the availability and efficacy of treatment and rehabilitation, recovery programs, employment options and housing possibilities for patients with schizophrenia (Strauss, 2007).

2 Methods

A meta-analysis was done to compare the effect sizes from randomized controlled trials of treatments for mood disorders in which there was either a placebo or a comparative treatment administered to the control group. The dependent variable for the investigation was the group effect size which is a measure of treatment strength and an indicator of how important the results of the study are. An exhaustive search of UBC's online databases (specifically BioMed Central, Biomedical Reference Collection, CogNet, MEDLINE, PsychARTICLES, PyschINFO, PsychEXTRA, Public Library of Science and PubMed) was conducted using the search terms 'depression – therapy', 'depression – RCTs', 'mood disorder – therapy', 'mood disorder – RCTs', 'anti-depressant trials' and 'mood disorder – trials'. This search returned approximately 16,000 results which were reduced to 37 peer-reviewed journal articles using the following inclusion criteria: 1) English language articles, 2) published between year 2000 and 2013, 3) a comparison of a mood disorder treatment to a placebo or another treatment, 4) enrolment of adult outpatients (over 18 years of age), 5) patients diagnosed with a mood disorder according to DSM-III-R, DSM-IV, DSM-IV-TR, ICD-9, or ICD-10, 3, 6) an assessment of cognitive functioning, 7) reporting of essential statistics so that an effect size (Cohen's d) could be obtained (see Figure 1) . To reduce variation, studies were excluded if there were inpatients, subjects with psychosis or mania or pregnant women were enrolled. Several of these studies compared more than two independent sample groups resulting in multiple effect sizes.

Figure 2.1. Flow chart of included trials
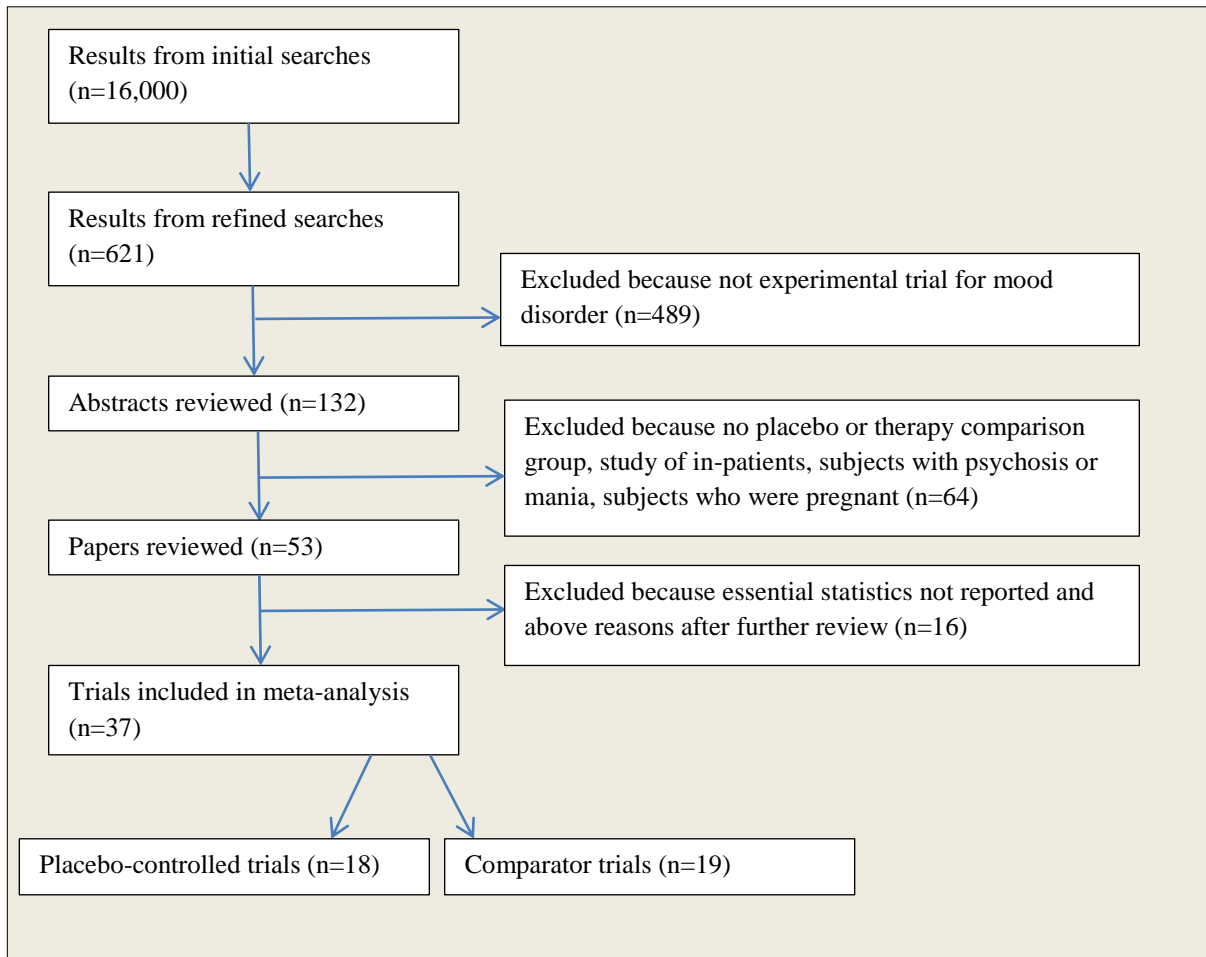
A coding manual was developed to outline the various descriptors that were used to code each research study (the detailed coding manual can be found in Appendix A). The articles were coded based on sample descriptors (sample size, age, gender, clinical characteristics), source descriptors (year of publication, funding source, type of study, number of groups), and most importantly design descriptors (treatment name, measurement tool), outcome data (test mean, control mean, standard deviations, group effect size) and quality (concealment of treatment allocation, blinding of outcome assessment). The most important code, in terms of the purposes of this meta-analysis, indicated whether or not the research design included a placebo or comparative control. A sample of the coding table has been provided in Appendix B.

All of the articles were coded by the author. For the purpose of calculating inter-coder reliability of the coding, a random 30% of the articles analyzed were selected using a formula in excel and coded by a second coder. The second coder was a fourth year undergraduate student from Simon Fraser University who was provided training prior to coding. The average of Cohen's Kappa (measure of the agreement between two judges which takes into account the agreement that is due to chance) for these 30% of the articles was 0.88, which indicates that inter-rater reliability was relatively high. Any discrepancies in coding were resolved by the author who reviewed and recoded the articles that were miscoded.

The weighted effect size average from those studies that used placebos were compared to that of those studies that administered treatments to the control group. The group effects of the treatments were coded, however, coding of the effects of time or group-by-time interactions were not. This is because the effects of time cannot be influenced by the presence or lack of a control group in the research design. Also, analyzing only the group effect size meant that the impact of dependent variable, could be isolated. However, the decision to code only group effects presented a challenge because many authors did not report this effect size (or the data required for calculation of effect size) if they were not statistically significant and as such many articles did not meet the criteria for coding. The group effect sizes from each study were averaged so that there was only one effect size per study (unless the study had more than two independent samples).

3 Results

The dependent variable for the investigation was the group effect size. This variable was calculated by taking an average of the group effect sizes from each study. The effect size was, in some cases converted to and expressed as Cohen's d. The mean effect size (d=0.3153), standard error (0.0228), and 95% confidence interval (0.2705-0.3600) was also calculated. Based on these statistics and inspection of a funnel plot (Figure 3.1) it was decided that one of the outliers, from the placebo group of trials, be removed from the analysis (d=4.40). The statistics for the studies included in the analysis have been summarized in Table 1. As shown in Table 3.1, the 37 placebo and comparator-controlled studies analyzed included 66 active treatment conditions and enrolled a total of 6940 participants.

Figure 3.1 Funnel plot of effect size by study size
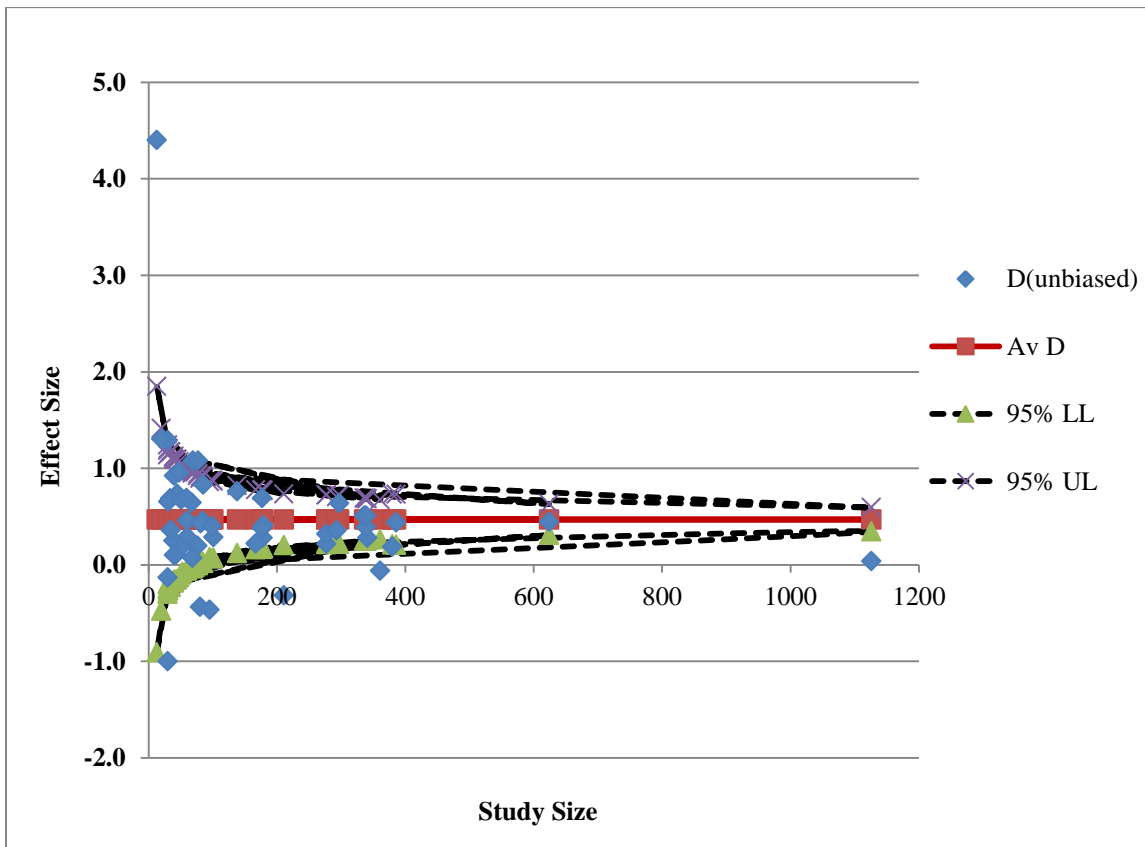
Table 3.1 Selected characteristics of included studies

	Placebo-controlled Trials	Comparator Trials	All Trials
Studies	19	18	37
Active Treatment Groups	22	44	66
Patients	4549	2391	6940
Mean ES	0.296	0.344	0.313
Standard Error	0.029	0.038	0.023
95% Confidence Interval	0.240-0.352	0.270-0.419	0.269-0.358

Note. Mean ES= Weighted Effect Size expressed as Cohen's D.

It was expected that the effect sizes from those studies that included a placebo control group in their research design would be significantly different than the effect sizes from those studies that include active controls. In order to determine this, the mean effect size for each group of studies was calculated and expressed as Cohen's d (Table 3.1).

Next, the homogeneity analysis of the data was conducted; homogeneity was rejected (Q_T=188.22, df=51, $p<0.05$) and it was concluded that the variability across the data exceeded the sampling error. A fixed effects model was then used to analyze the heterogeneous distribution of effect sizes in which no assumption is required. The justification for using a fixed effects model was that that the random effects assumption could not be met; this is the assumption that the unobserved characteristics that may influence effect size, such as patient expectations or concealment of treatment group, are not correlated with the moderator of interest which is the type of control group (Clarke, Crawford, Steele & Vignoles, 2010). The results of this analysis showed that grouping variable, which is the type of control used (active control or placebo-control) in the research design, was responsible for significant variability between effect sizes (Q_B=12.85, df=1, $p<0.05$). Although the results from the analog to the one-way ANOVA showed that the grouping variable of interest has significant moderating effects, there was still a

significant amount of variability in the data that was unaccounted for (Q_W=175.36, df=50,

p<0.05).

4   Conclusion

4.1 Purpose and Potential Applications of Research

The purpose of this meta-analysis was to investigate whether there is a significant difference in the effect size between placebo-controlled and treatment-controlled randomized trials, and mood disorder trials was used as the context for this investigation. The results of this analysis were consistent with the initial hypotheses which stated that effect sizes were significantly different in comparator trials versus placebo-controlled trials. It can be concluded that control group type was a significant moderator of the variability observed in the effect sizes in the mood disorder treatment trials of this meta-analysis. As stated previously, much of the research examining the relationship between control group type and effect size has been in the context of depression and has been limited to trials that test moderating effect of control group type on effect size is restricted to randomized controlled trials for depression. By expanding this meta-analysis to include non-pharmaceutical mood disorder trials, these results not only support previous research but also build on it, which serves as the major strength of this analysis.

This analysis also provides important practical insights around the design, use of control groups in experimental research and comparison of interventions. It supports the notion that consistency in research design should be a key consideration for the valid comparison of treatments and effect sizes from randomized controlled trials and that the effectiveness of a treatment cannot be determined by comparing against only one type of control or under one type of research design. When there are considerable differences in the design and methods utilized in randomized controlled trials, using effect size as the sole indicator of treatment strength in the comparison of treatments, could be an issue. This research may lead to potential changes in standards for experimental research design and a potential shift in focus from effect size to

alternative indicators of treatment strength. It also suggests that the very definition of RCTs and our consideration of them as the gold standard for experimental research may need to be reassessed as RCTs can vary significantly in design.

Previous research has shown that the impact of research design on effect size is moderated by patient expectations and although the results of this analysis cannot be said to provide direct evidence of this role, it is important to note that these results do not contradict it. Higher expectations of recovery have been shown to explain why patients in treatment-controlled trials have significantly higher odds of response and remission (Rutherford et al., 2009a). The weighted effect size from the treatment-controlled group of trials in this meta-analysis was significantly higher than that of the placebo-controlled trials and therefore supports the hypothesis of Rutherford et al. about the role of expectations in recovery. These results support the need for the development of methods and best practices to control for subject expectations, which could include pre-treatment measures of recovery expectations (Krell et al., 2004). The difficulty in controlling for patient expectation may in part explain why such methods have not yet been developed.

The implications of this research may also extend to the procedures for informed consent in mood disorder trials. The results of this meta-analysis support numerous studies which have shown that subjects in treatment-controlled trials experience a greater response to treatment. This suggests that an assessment of and potential modifications to current informed consent procedures is required. In addition to being made aware of the risks associated with not receiving treatment in placebo-controlled trial, there may be a need for patients to be informed of the risk of experiencing a weaker response to treatment when enrolled in a placebo-controlled trial versus a treatment-controlled trial.

Although the results of this meta-analysis showed to be significant, it should be mentioned that there was a small difference between the mean effect sizes of the studies being compared. This closeness in effect sizes speaks to the strength of subject expectations in placebo-controlled trials and as such warrants an alternative interpretation of the results; the expectation of improvement alone can essentially have a similar impact as treatment controls have on patient response.

4.2   Future Research

Additional research is required to further understand the mechanism by which control group type impacts effect size. As discussed previously, it may be found that comparator trials and placebo-controlled trials differ in various unidentified design aspects that have an impact on the effectiveness of treatments for mood disorders. Rutherford et al. has gone as far to suggest that "medication accounts for a minority of the change observed in clinical trials" (2009a, p. 177). Further research is also required to build on the current understanding of how patient expectations play a role in this relationship. When subjects are made aware, before participation, that they are entering a placebo-controlled or treatment-controlled study, research shows that there is an impact on expectations of recovery (Rutherford et al., 2010). Improving patient expectations may show to be an important method for improving the response to treatments, however, as mentioned before, requirements will need to be introduced for monitoring and controlling for patient expectations when testing and comparing treatments.

Future research may show that the moderating effect of research design on effect size differs by the type of treatment being tested. For example, it may be found that in pharmaceutical trials, the impact of research design, and more specifically, patient expectations on treatment response could be different than in the testing of behavioral interventions. There is also a need

for additional research to understand the link between research design and effect size outside the context of depression and to determine whether and to what extent patient expectations play a role in modulating this in other contexts. Research has yet to confirm whether the type of illness being treated is a factor in the relationship between research design and response. It is understood that perceptions of recovery differ by illness, so it is not implausible that the type of illness impacts patient expectations of treatment efficacy and in turn, the response to treatment.

Until now, much of the research in this area has suggested that patient expectations have a crucial role in the relationship between research design and response, however, further research is needed to uncover the presence and influence of other moderating factors. There may be disparities in the design of placebo-controlled trials and treatment-controlled trials that contribute to differences in patient response or that make one design more vulnerable to biases or cross-contamination which in turn have an impact on effect size. Cross-contamination in the context of a RCT can result in the diffusion of treatments which occurs when there is an exchange of information between the treatment group and control group (Campbell, 1963). This results in similarity in dependent variable between both groups. Further investigation is required to determine whether the diffusion of treatments varies by research design.

Although it was not the focus of this research, subsequent research should consider the impact of medical ghost writers on treatment response. Medical ghostwriting involves professional writers and other parties who are paid, most commonly by pharmaceutical and medical device companies, to produce or assist in the writing of scientific publications, educational materials and other communication materials (Langdon-Neuner, 2008). What makes an author a ghost writer is the absence of their name on pieces of work that they have contributed to. When a writer's name is omitted from an article, so is the name of the funding source for the

assistance that is provided by the writer. It is believed by some that the when writers who are paid by the pharmaceutical industry, it results in bias (Langdon-Neuner, 2008). This strength and effect of this bias is still to be determined.   Assessing the impact of using ghost writers on clinical trial outcomes is problematic due to the challenge of accurately identifying them.

4.3 Limitations

There are several limitations that should be considered when interpreting the results of this meta-analysis. First, this research synthesis was limited to clinical trial of mood disorders. Further research is required if generalizations are to be made about the relationship between control group type and effect size in other contexts outside of psychology. As mentioned before, it may be found that the impact of control group type on effect size is moderated by the type of disorder or symptoms being treated.

Second, the number of studies examined in this meta-analysis was a limitation. The small sample size was a result of many authors only reporting significant effect sizes or main effects of time which leads to the third limitation; the reporting of significant results only. Many studies only reported significant results so it follows that effect size for the sample of trials included in this meta-analysis may not be representative of the all mood disorder clinical trials in the specified time frame.

Another limitation of this meta-analysis is that the studies included may be affected by the publication bias, however, this influence is unclear. Many studies are not published as a result of not having significant results. It could be found that there are more unpublished studies with placebo controls than treatment controls or vice a versa, and if so this would impact the inclusion of studies and in turn the results of this meta-analysis.

Although the results of this analysis confirmed that there was indeed a statistically significant variability between the placebo-controlled and treatment-controlled studies, it should be noted that there was a large amount of variability within groups that was unaccounted for by the grouping variable. This leads to the question of whether the between group variability (Q_B=12.85), although statistically significant, is large enough to be meaningful given the large amount of variability that is unaccounted for (Q_W=175.36). According to Huedo-Medina, Sanchez-Meca, Marin-Martinez and Botella, the weakness of the Q statistic is that it has poor power to detect heterogeneity when sample sizes are small, however, the definition of what constitutes a small sample size in the context of a meta-analysis is not discussed (2006). Until more attention has been given to this subject, the amount of variability between groups can only be considered a potential limitation of this meta-analysis.

Lastly, the results of this meta-analysis may be influenced by the variation among subjects who participated in the included trials, the different study types and study durations. An analysis of participant demographic and clinical characteristics would have to be done in order to assess this potential limitation.

References

References marked with an asterisk indicate studies included in the meta-analysis. The in-text citations to studies selected for meta-analysis are not preceded by asterisks.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). doi:10.1176/appi.books.9780890423349

*Allen, J. J. B., Schnyer, R. N., Chambers, A. S., Hitt, S. K., Moreno, F. A., & Manber, R. (2006). Acupuncture for depression: a randomized control trial. *Journal of Clinical Psychiatry, 67,* 1665-1673. Retrieved November 28, 2014, from http://www.psychiatrist .com/JCP/article/Pages/2006/v67n11/v67n1101.aspx

*Antonioli, C., & Reveley, M. A. (2005). Randomised controlled trial of animal facilitated therapy with dolphins in the treatment of depression. *British Medical Journal, 331*(7527), 1231. doi: 10.1136/bmj.331.7527.1231

Barsky A. J., Saintfort R., Rogers M. P. & Borus, J. F. (2002). Nonspecific medication side effects and the nocebo phenomenon. *JAMA, 287*, 622–627. doi: 10.1177/ 0004867412464717

*Berk, M., Dean, O. M., Cotton, S. M., Gama, C. S., Kapczinski, F., Fernandes, B., … Malhi, G. S. (2012). Maintenance N-acetyl cysteine treatment for bipolar disorder: A double-blind randomized placebo controlled trial. *BMC Medicine, 10*(91), 1-11. Retrieved November 22, 2014, from http://www.biomedcentral.com/1741-7015/10/91

*Berman, R M., Narasimhan, M., Sanacora, G., Miano, A. P., Hoffman, R. E., Hu, X. S., … Boutros, N. N. (1999). A randomized clinical trial of repetitive transcranial magnetic stimulation in the treatment of major depression. *Biol Psychiatry, 47*, 332-337. Retrieved from http://dx.doi.org/10.1016/S0006-3223(99)00243-7

*Binneman, B., Geltner, D., Kolluri, S., Shi, Y., Qiu, R., & Stiger, T. (2008). A 6-Week
Randomized, Placebo-Controlled Trial of CP-316,311 (a Selective CRH1 Antagonist)
in the Treatment of Major Depression. *American Journal of Psychiatry, 165*, 617-620.
doi: 10.1176/appi.ajp.2008.07071199

*Bodkin, J. A., & Amsterdam, J. D. (2002). Transdermal Selegiline in Major Depression: A
Double-Blind, Placebo-Controlled, Parallel-Group Study in Outpatients. *American
Journal of Psychiatry, 159,* 1869-1875. Retrieved from http://dx.doi.org/10.1176/appi.
ajp.159.11.1869

*Boggio, P. S., Rigonatti, S. P., Ribeiro, R. B., Myczkowski, M. L., Nitsche, M. A., Pascual-
Leone, A., & Fregni, F. (2008). A randomized, double-blind clinical trial on the efficacy
of cortical direct current stimulation for the treatment of major depression. *Int J
Neuropsychopharmacol, 11*(2), 249-254. doi:10.1017/S1461145707007833

*Bowden, C. L., Calabrese, J. R., McElroy, S. L., Gyulai, L., Wassef, A., Petty, F., … Wozniak,
P. J. (2000). A randomized, placebo-controlled 12-month trial of divalproex and lithium
in treatment of outpatients with bipolar I disorder. *Arch Gen Psychiatry, 57*, 481-489.
doi:10.1001/archpsyc.57.5.481

*Brunoni, A. R., Valiengo, L. Baccaro, A., Zanao, T. A., De Oliveira, J. F., Goulart, A., …
Fregni, F. (2013). The Sertraline vs Electrical Current Therapy for Treating Depression
Clinical Study. *Jama Psychiatry, 70*(4), 383-391. doi:10.1001/2013.jamapsychiatry.32

*Calabrese, J. R., Keck, P. E., Macfadden, W., Minkwitz, M., Ketter, T. A., Weisler, R. H., …
Mullen, J. (2005). A randomized, double-blind, placebo-controlled trial of quetiapine in
the treatment of bipolor I or II depression. *American Journal of Psychiatry, 162*, 1351-
1360. Retreived from http://ajp.psychiatryonline.org/doi/abs/10.1176/appi.ajp.162.7.1351

Campbell, D. T. (1963). Social attitudes and other acquired behavioural dispositions.

 Psychology: A Study of Science, 6, 94-172. doi: http://dx.doi.org/10.1037/10590-003

*Castle, D., White, C., Chamberlin, J., Berk, M., Berk, L., Lauder, S., … Gilbert, M. (2010).

 Group-based psychosocial intervention for bipolar disorder: randomised controlled trial.

 The British Journal of Psychiatry, 196, 383-388. doi: 10.1192/bjp.bp.108.058263

Castro, M. (2007). Placebo versus Best-Available-Therapy Control Group in Clinical Trials for

 Pharmacologic Therapies. *The American Thoracic Society, 4,* 570-573. doi: 10.1513/

 pats.200706-073JK

*Chalder, M., Wiles, N. J., Campbell, J., Hollinhurst, S. P., Hasse, A. M., Taylor, A. H., …

 Lewis, G. (2012). Facilitated physical activity as a treatment for depressed adults:

 randomised controlled trial. *British Medical Journal, 344,* e2758. doi: 10.1136/bmj.e2758

Chatoor, I. & Krupnick J. (2001). The role of non-specific factors in treatment outcome of

 psychotherapy studies. *European Child & Adolescent Psychiatry, 10*(1), 19-25.

 doi: 10.1007/s007870170004

*Ciechanowski, P., Wagner, E., Schmaling, K., Schwartz, S., Williams, B., Diehr, P., …

 LoGerfo, J. (2004). Community integrated home based depression treatment in older

 adults: a randomized Controlled Trial. *JAMA, 291,* 1569-1577. doi:10.1001/jama.291.

 13.1569

Cipher, D. J., & Fernandez, E. (1997). Expectancy variables in predicting tolerance and

 avoidance of pain in chronic pain patients. *Behavior Research and Therapy*, 35, 437–444.

 doi: 10.1016/S0005-7967(96)00121-0

Clarke, P., Crawford, C., Steele, F., & Vignoles, A. (2010). The Choice Between Fixed and

 Random Effects Models: Some Considerations for Educational Research. *Institute for the*

Study of Labor. Retrieved November 23, 2014, from http://www.iza.org/en/webcontent/
publications/papers/viewAbstract? dp_id=5287

Cole, D. C., Mondloch, M.V., & Hogg-Johnson, S. (2002). Listening to injured workers: how
recovery expectations predict outcomes: a prospective study. *Canadian Medical
Association Journal*, *166*, 749–754. doi: 10.1007/s00586-009-0954-4

Council for International Organizations of Medical Sciences. (2002). International ethical
guidelines for biomedical research involving human subjects. *Council for International
Organizations of Medical Sciences.* Retrieved May 4, 2014, from http://www.cioms.ch/
publications/guidelines/guidelines_nov_2002_blurb.htm

Delen, M. (1979). A New Design for Randomized Clinical Trials. *The New England Journal of
Medicine*, *300*, 1242-1245. doi: 10.1007/s10985-004-4769-7

*Dimidjian, S., Dobson, K. S., Kohlenberg, R. J., Gallop, R., Markley, D. K., Atkins, D. C., …
Jacobson, N. S. (2006) Randomized Trial of Behavioral Activation, Cognitive Therapy
and Antidepressant Medication in the Acute Treatment of Adults with Major Depression.
Journal of Consulting and Clinical Psychology, 74(4), 658-670. doi: 10.1037/0022-
006X.74.4.658

Ellenberg, S. S. & Temple, R. (2000) Placebo-Controlled Trials and Active-Control Trials in
the Evaluation of New Treatments. Part 2: Practical Issues and Specific Cases. *Annals of
Internal Medicine*, *133*(6), 464-470. doi:10.7326/0003-4819-133-6-200009190-00015

*Erkkila, J., Punkanen, M., Fachner, J., Ala-Ruona, E., Pontio, I., Tervaniemi, M., … Gold, C.
(2011). Individual music therapy for depression: randomised controlled trial. *The British
Journal of Psychiatry, 199*, 132-139. doi: 10.1192/bjp.bp.110.085431

Fern, E. F. & Monroe, K. B. (1996). Effect-Size Estimates: Issues and Problems in

Interpretation. *The Journal of Consumer Research, 23*(2)*,* 89-105. doi: 10.1086/209469

*Fitzgerald, P. B., Hoy, K., McQueen, S., Maller, J. J., Herring, S., Segrave, R., … Daskalakis, Z. J. (2009). A randomized trial of rTMS targeted with MRI based neuro-navigation in treatment-resistant depression. *Neuropsychopharmacology, 34*, 1255-1262. doi:10.1038/npp.2008.233

Flood, A. B., Lorence, D. P., Ding, J., McPherson, K., & Black, N. A. (1993). The role of expectations in patients' reports of post-operative outcomes and improvements following therapy. *Med Care, 31*, 1043-1056. doi: 10.1097/00005650-199311000-00006

*Forman, E. M., Herbert, J. D., Moitra, E., Yeomans, P. D., & Geller, P. A. (2007). A Randomized Controlled Effectiveness Trial of Acceptance and Commitment Therapy and Cognitive Therapy for Anxiety and Depression. *Sage Publications, 31*(6), 772-799. doi: 10.1177/0145445507302202

Geers, A. & Rose, J. (2011). Treatment Choice and Placebo Expectation Effects. *Social and Personality Psychology* Compass, *5*(10), 734–750. doi: 10.1111/j.1751-9004.2011.00385.x

*Gerhard, A., Bergstrom, J., Hollandare, F., Carlbring, P., Viktor, K., & Ekselsius, L. (2005). Internet-based self-help for depression: randomised control trial. *The British Journal of Psychiatry, 187*, 456-461. doi: 10.1192/bjp.187.5.456

*Grunhaus, L., Schreiber, S., Dolberg, O. T., Polak, D., & Dannon, P. N. (2003). A Randomized Controlled Comparison of Electroconvulsive Therapy and Repetitive Transcranial Magnetic Stimulation in Severe and Resistant Nonpsychotic Major Depression. *Biological Psychiatry, 53*(4), 324-331. doi: 10.1016/j.jad.2008.01.001.

Heo, M., Papademetriou, E., & Meyers, B. S. (2009). Design characteristics that influence

attrition in geriatric antidepressant trials: meta-analysis. *International Journal Of Geriatric Psychiatry*, *24*(9), 990-1001. doi: 10.1002/gps.2211

Huedo-Medina, T., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing Heterogeneity in Meta-Analysis: Q Statistic or I2 Index? *CHIP Documents, 19*. doi:10.1037/1082-989X.11.2.193. 193

Jan, S., & Shieh, G. (2011). Optimal sample sizes for Welch's test under various allocation and cost considerations. *Behavior Research Methods*, *43*(4), 1014-1022. doi: 10.3758/s13428-011-0095-7

Juni, P., Altman, D. G., & Egger, M. (2001). Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*, *323*(7303), 42-46. doi: http://d_x.doi.org/ 10.1136/bmj.323.7303.42

*Keitner, G. I., Garlow, S. J., Ryan, C. E., Ninan, P. T., Soloman, D. A., Nemeroff, C. B., & Keller, M. B. (2009). A randomized, placebo-controlled trial of risperidone augmentation for patients with difficult-to-treat unipolar, non-psychotic major depression. *J Psychiatr Res. 43*(3), 205-214. doi: 10.1016/j.jpsychires.2008.05.003

Kelly, M. P., & Moore, T. A. (2011). Methodological, theoretical, infrastructural, and design issues in conducting good outcome studies. *Research on Social Work Practice*, *21*(6), 644-653. doi: 10.1177/1049731511405070

*Kennedy, S. H., & Emsley, R. (2006). Placebo-controlled trial of agomelatine in the treatment of major depressive disorder. *European Neuropsychopharmacology, 16*, 93-100. doi:10.1016/ j.euroneuro.2005.09.002

Khan, A., Kolts, R. L., Thase, M. E., Krishnan, R. R., & Brown, W. (2004). Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. *American Journal of Psychiatry, 161*, 2045–2049. doi: 10.1186/2193-1801-3-740

Khan, A., Warner, H., Brown, W. (2000). Symptom reduction and suicide risk in patients treated

with placebo in antidepressant clinical trials. *Arch Gen Psychiatry, 57*, 311–317.

doi:10.1001/archpsyc.57.4.311

Kim, S. Y. H. (2003) Benefits and Burdens of Placebos in Psychiatric Research.

Psychopharmacology, 171(1), 13-18. doi: 10.1007/s00213-003-1458-2

Krell, H. V., Leuchter, A. F., Morgan, M., Cook, I. A., & Abrams, M.. (2004). Subject

Expectations of Treatment Effectiveness and Outcome of Treatment with Experimental

Antidepressant. *Journal of Clinical Psychiatry, 65*(9), 1174-1179. Retrieved from

http://dx.doi.org/10.4088/JCP.v65n0904

*Lam, R. W., Levitt, A. J., Levitan, R. D., Enns, M. W., Morehouse, R., Michalak, E. E., & Tam,

E. M. (2006). The Can-SAD Study: A Randomized Controlled Trial of the Effectiveness

of Light Therapy and Fluoxetine in Patients With Winter Seasonal Affective Disorder.

American Journal of Psychiatry, 163, 805-812. Retrieved from

http://ajp.psychiatryonline.org/doi/full/10.1176/ajp.2006.163.5.805

Langdon-Neuner, E. (2008). Medical Ghost Writing. *Mens Sana Monograph, 6*(1), 257-273.

doi: 10.4103/0973-1229.33006

*Loebel, A., Cucchiaro, J., Silva, R., Kroger, H., Hsu, J., Sarma, K., & Sachs, G. (2014).

Lurasidone Monotherapy in the Treatment of Bipolar I Depression: A Randomized,

Double-Blind, Placebo-Controlled Study. *American Journal of Psychiatry, 171,* 160-168.

doi: 10.1176/appi.ajp.2013.13070984

*McIntyre, R. S., Cohen, M., Zhao, J., Alphs, L., Macek, T. A., & Panagides, J. (2010).

Asenapine in the treatment of acute mania in bipolar I disorder: A randomized, double-

blind, placebo-controlled trial. *Journal of Affective Disorders, 122,* 27-38.

doi:10.1016/j.jad.2009.12.028

*Miklowitz, D. J., Otto, M. W., Frank, E., Reilly-Harrington, N. A., Wisniewski, S. R. … Sachs, G. S. (2007). Psychosocial Treatments for Bipolar Depression A 1-Year Randomized Trial From the Systematic Treatment Enhancement Program. *Arch Gen Psychiatry, 64,* 419-427. doi:10.1001/archpsyc.64.4.419

*Miranda, J., Chung, J. Y., Green, B. L., Krupnick, J., Siddique, J., Revicki, D. A., & Belin, T. (2003). Treating depression in predominantly low-income young minority women. *JAMA, 290*(1), 57-65. Retrieved from http://jama.jamanetwork.com/

Mondloch, M. V. (2001) Does how you do depend on how you think you'll do? A systematic review of the evidence for a relation between patients' recovery expectations and health outcomes. *Canadian Medical Association Journal, 165*(2), 174-179. doi: 10.1111/j.1744-1609.2008.00090.x.

*Murrough, J. W., Iosifescu, D. V., Chang, L. C., Al Jurdi, R. K., Green, C. E., Perez, A. M., … Mathew, S. J. (2013). Antidepressant efficacy of ketamine in a treatment-resistant major depression: a two-site randomized controlled trial. *American Journal of Psychiatry, 170*, 1134-1142. doi: 10.1176/appi.ajp.2013.13030392

*Perich, T., Manicavasagar, V., Mitchell, P. B., Ball, J. R., & Hazdi-Pavlovic, D. (2013)., A randomized controlled trial of mindfulness-based cognitive therapy for bipolar disorder. *Acta Psychiatrica Scandinavica, 127,* 333-343. doi: 10.1111/acps.12033

Quitkin, F. M., Rabkin, J. G., Gerald, J., Davis, J. M. & Klein, D. F. (2000). Validity of clinical trials of antidepressants. *American Journal of Psychiatry*, *157*, 327 -337. Retrieved at http://ajp.psychiatryonline.org/doi/full/10.1176/appi.ajp.157.3.327

Roose, S. P., Schatzberg, A. F. (2005). The efficacy of antidepressants in the treatment of late-life depression. *Journal of Clinical Psychopharmacology*, *25*(Supplement 1), S1–S7. doi: 10.1097/01.jcp.0000162807.84570.6b

*Rea, M. M., Tompson, M. C., Miklowitz, D. J., Goldstein, M. J., Hwang, S., Mintz, J. (2003). Family-Focused Treatment Versus Individual Treatment for Bipolar Disorder: Results of a Randomized Clinical Trial. *Journal of Consulting and Clinical Psychology, 71*(3), 482-492. doi: 10.1037/0022-006X.71.3.482

*Rush, A. J., Marangell, L.B., Sackeim, H. A., George, M. S., & Brannan, S. K., Davis, S. M., … Cooke, R. G. (2005). Vagus nerve stimulation for treatment-resistant depression: a randomized, controlled acute phase trial. *Biological Psychiatry,* 58(5), 347-354. doi: 10.1007/s00702-012-0908-6

Rutherford, B. R., Sneed, J. R., & Roose, S. P. (2009a) Does study design influence outcome? The effects of placebo control and treatment duration in antidepressant trials. *Psychotherapy and Psychosomatics, 78*(3), 172-181. doi: 10.1159/000209348\

Rutherford, B. R., Sneed, J. R., & Roose, S. P. (2009b) Study Design Affects Participant Expectations: A Survey. *Journal of Clinical Psychopharmacology, 29*, 197-181. doi: 10.1097/JCP.0b013e31819a9181

Rutherford, B. R., Sneed, J. R. , Eisenstadt, R., & Roose, S. P. (2010). Antidepressant Study Design Affects Patient Expectations: Initial Results from a Randomized Controlled Trial. *Psychological Medicine*, In Press. doi: 10.1017/S0033291709991085

*Savitha, E., Mogg, A., Pluck, G., Landau, S., Purvis, R., & Howard, R. (2007). A Randomized, Controlled Trial with 6-Month Follow-Up of Repetitive Transcranial Magnetic Stimulation and Electroconvulsive Therapy for Severe Depression. *American Journal of Psychiatry, 164*, 73-81. Retrieved from http://ajp.psychiatryonline.org/doi/abs/10.1176/ajp.2007.164.1.73

*Shelton, R. C., Keller, M. B., Gelenberg, A., Dunner, D. I., Hirschfeld, R., Thase, M. E., …

Halbreich, U. (2001). Effectiveness of st john's wort in major depression. *JAMA, 285*(15), 1979-1987. doi: 10.1002/14651858.CD000448.pub3

Sneed, J. R., Rutherford, B. R., Rindskopf, D., Lane, D. T., Sackeim, H. A., & Roose, S. P. (2008). Design makes a difference: A meta-analysis of antidepressant response rates in placebo-controlled versus comparator trials in late-life depression. *American Journal of Geriatric Psychiatry, 16*(1), 65-73. doi: 10.1097/JCP.0b013e31819a9181

Strauss, J. (2007). Prognosis in the Individual, the Environment, the Disease, or What? *Schizophrenia Bulletin*, *34*(2), 245–246. doi: 10.1093/schbul/sbm125.

*Titov, N., Andrews, G., Davies, M., McIntyre, K., Robinson, E., Solley, K. (2010). Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. *PLoS ONE, 5*(6), e10939. doi:10.1371/journal.pone.0010939

Thase, M. E. (1999). How should efficacy be evaluated in randomized clinical trials of treatments for depression? *Journal of Clinical Psychiatry*, *60*, 23-32. Retrieved May 20, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/10086480

*Tohen, M., Vieta, E., Calabrese, J., Ketter, T., Sachs, G., … Breier, A. (2003). Efficacy of Olanzapine and Olanzapine-Fluoxetine Combination in the Treatment of Bipolar I Depression. *Arch Gen Psychiatry, 60*, 1079-1089. doi:10.1001/archpsyc.60.11.1079

Vaz-leal, F. J. (1989). Expectations of the patient and attitudes toward treatment as predictive factors in the clinical course. *Actas Luso Esp Neurol Psiquiatr Cienc Afines*, *17*(3), 215-222. Retrieved at http://europepmc.org/abstract/med/2773695

Volkow, N. D., Wang, G. J., Ma, Y., Fowler, J. S., Wong, C., Jayne, M. et al. (2006). Effects of expectation on the brain metabolic responses to methylphenidate and to its placebo in non-drug abusing subjects. *Neuroimage, 32*, 1782–92. doi: 10.1371/journal.pone. 0002017.

Weissman, J., Flint, A., Meyers, B., Ghosh, S., Mulsant, B., Rothschild, A., & Whyte, E. (2012).

    Factors associated with non-completion in a double-blind randomized controlled trial of

    olanzapine plus sertraline versus olanzapine plus placebo for psychotic depression.

    Psychiatry Research, *197*(3), 221-226. doi: 10.1016/j.psychres.2012.02.015

*Wollmer, M. A., Boer, C., Kalak, N., Beck, J., Gotz, T., Schmidt, T. … Kruger, T. H. C.

    (2012). Facing depression with botulinum toxin: A randomized controlled trial. *Journal*

    of Psychiatric Research, 46, 574-581. doi:10.1016/j.jpsychires.2012.01.027

*Zarate, C. A., Brutsche, N. E., Ibrahim, L., Franco-Chaves, J., Diazgranados, N., Cravchik, A.,

    … Luckenbaugh, D. A. (2011). Replication of ketamine's antidepressant efficacy in

    bipolar depression: a randomized controlled add-on trial. *Biol Psychiatry, 71*, 939-946.

    doi:10.1016/j.biopsych.2011.12.010

*Zarate, C. A., Singh, J. B., Carlson, P. J., Brutsche, N. E., Ameli, R., Luckenbaugh, D. A., …

    Manji, H. K. (2006). A Randomized Trial of anN-methyl-D-aspartate Antagonist in

    Treatment-Resistant Major Depression. *Arch Gen Psychiatry, 63*(8), 856—864. doi:

    10.1001/archpsyc.63.8.856

Appendices

Appendix A: Coding Manual

1) Author Name
2) Study Name
3) Year of Publication
4) Funding Source
5) Type of Study
    a. Dissertation
    b. Investigator Initiated
    c. Other
    d. Not stated
6) Number of Treatment Groups
7) Sample Size
8) Mean Age
9) Gender Split
10) Clinical Characteristics
    a. Major Depression/Depressive Disorder
    b. Unipolar Depression or Major Depressive Disorder
    c. Depressed Phase
    d. Bipolar Disorder (includes Bipolar I, Bipolar II and Bipolar NOS Disorders)
    e. Acute Bipolar Mania
    f. Mania
    g. Hypomania
    h. Mixed Affective Disorder
    i. Dysthymia (all types)
11) Test Treatment
12) Measurement Tool (primary and secondary)
    a. Beck Depression Inventory (BDI)
    b. Hamilton Depression Rating Scale (HDRS/HAM-D)
    c. Montgomery-Asberg Depression Rating Scale (MADRS)
    d. Brief Psychiatric Rating Scale (BPRS)
    e. Young Mania Rating Scale (YMRS)
    f. Mania Rating Scale (MRS)
    g. Hospital Anxiety and Depression Scale (HADS-A)
    h. Global Assessment of Functioning (GAF)
    i. Depressive Syndrome Scale (DDS)
    j. Global Assessment Scale (GAS)
    k. Hopkins Symptom Checklist Depression Scale (HSCL-D-20)
    l. DSM-IV-TR

       m. Clinical Global Impressions (CGI)

       n. Clinical Severity Score for Glabellar Frown Lines (CSS-GFL)

13) Control Group Type

       a. Comparator

       b. Placebo

14) Test Mean(s)

15) Test Standard Deviation(s)

16) Control Mean(s)

17) Control Standard Deviation(s)

18) Group Effect Size(s)

19) Standard Deviation

20) Concealment of Treatment Allocation

       a. Yes

       b. No

       c. Not Stated

21) Blinding of Outcome Assessment

       a. Yes

       b. No

       c. Not Stated

Appendix B: Sample of Coding Table

Author Name(s)	Wollmer, M.A. et al.	Zarate, C. A. et al.	Keitner, G. I. et al.	Erkilla, J. et al.
Study Name	Facing depression with botulinum toxin: A randomized controlled trial	A Randomized Trial of anN-methyl-D-aspartate Antagonist in Treatment-Resistant Major Depression	A randomized, placebo-controlled trial of risperidone augmentation for patients with difficult-to-treat unipolar, non-psychotic major depression	Individual music therapy for depression: randomised controlled trial
Year of Publication	2012	2006	2013	2011
Funding Source	A private foundation that supports medical research.	National Institute of Mental Health	Janssen Pharmaceutical	New and Emerging Science and Technology programme of the European Commission and the programme for Centres of Excellence in research, Academe of Finland
Type of Study	Not stated	Not stated	Investigator Initiated Study	Not stated
Number of Groups	2	2	2	2
Sample Size	30	18	97	79
Mean Age	50.57	46.7	45	35.5
Gender Split	Not stated	12 female, 6 male	58.5% female, 41.5% male	78% female, 22% male
Clinical Characteristics	Major depressive disorder	Major depressive disorder	Major depressive disorder	Unipolar depression
Test Treatment	Vistabel Botox Cosmetic, Allergen	NMDA antagonist ketamine hydrochloride	Risperidone Augmentation	Music
Measurement Tool	Primary Outcome: HDRS. Secondary Outcomes: BDI, CGI, CSS-GFL	Primary Outcomes: HDRS. Secondary Outcomes: BDI. BPRS, YMRS	Primary Outcomes: MADRS. Secondary Outcomes: HDRS, CGI	Primary Outcomes: MADRS. Secondary Outcomes: HADS-A, GAF
Control Group Type	Placebo	Placebo	Placebo	Comparator
Test Mean	-10.07	24.9	230.3	3 month: 16.43 , 6 month: 14.74
Test Mean SD	Not stated	6.9	346.9	3 month: 9.33, 6 month: 10.65
Control Mean	-1.73	24.4	111.4	3 month: 14.10 , 6 month: 14.48
Control Mean SD	Not stated		159.3	3 month: 8.77, 6 month: 9.60
Group Effect Size	$F(1,28)=12.30$	$F(1,16)=10.44$	$F=3.25$	3 month t-test=2.29, 6 month t-test=1.53
Concealment of Treatment Allocation	Yes	Yes	Yes	No
Blinding of Outcome Assessment	Not stated	Not stated	Not stated	Not stated