A Meta-Analysis of Expressive Writing on Positive Psychology Variables and Traumatic Stress

Abstract

Emotional expression has been shown to be beneficial for promoting both positive psychological and physical health outcomes. Unfortunately, inhibiting emotions is related to impairments in physical and psychological health. James Pennebaker showed that expressive writing is an effective form of emotional expression, and he and others have used expressive writing as an experimental manipulation to gauge its efficacy in treating a wide variety of health-related and psychological outcomes. While many studies have been conducted that examine the efficacy of expressive writing across such outcomes, a considerable amount of these studies tend to neglect necessary considerations such as different levels of symptomatology, power, and meaningfulness of respective effect sizes. Six previous meta-analyses have been conducted that examine expressive writing's effect on psychological outcomes. However, these studies focus on the experimental versus control group effect size. Thus, our meta-analysis sought to examine the efficacy of an expressive writing task on only the experimental conditions in studies measuring posttraumatic stress, posttraumatic growth, and quality of life using random effects models. Results indicated a small overall effect size for posttraumatic stress and negligible to small effect sizes for posttraumatic growth and quality of life. However, those studies requiring a diagnosis of PTSD exhibited a medium to large effect size. Implications for future research design and interpretation of published research are discussed.

Keywords: meta-analysis, posttraumatic stress, expressive writing

A Meta-Analysis of Expressive Writing on Positive Psychology Variables and Traumatic Stress

Emotional Expression

Emotional expression enhances both psychological and health-related outcomes (Esterling, Antoni, Kumar, & Schneiderman, 1990; Fawzy et al., 1993; Lieberman & Goldstein, 2006; Rachman, 1980; Scheff, 1979). Pennebaker & Beall (1986) first pioneered expressive writing, a form of emotional expression that involved writing about the thoughts and feelings associated with either a "stressful or traumatic" or neutral event. Further, the original protocol included 3-5 writing sessions, each lasting 15-20 minutes in length. In their seminal study employing expressive writing methodology in comparison to a control group, Pennebaker & Beall (1986) discovered that participants assigned to write about thoughts and feelings related to the stressful/traumatic event reported a reduction in health visits at the university health center. Termed written emotional disclosure (WED), this protocol has since been employed across varying contexts. Indeed, as of 2014, the expressive writing literature recognizes over 400 studies across different populations and outcome variables (Niles, Haltom, Mulvenna, Lieberman, & Stanton, 2014). For example, WED is efficacious for health-related outcomes, such as reduced doctor visits for those diagnosed with Type I diabetes (Bodor, 2002) or breast cancer (Stanton et al., 2002) and medication use in those suffering from chronic illness (i.e., asthma and rheumatoid arthritis; Smyth, Stone, Hurewitz, & Kaell, 1999). In regards to psychological outcomes, WED is efficacious for reducing depression symptoms (Gortner, Rude, & Pennebaker, 2006), posttraumatic stress (Di Blasio et al., 2015), and anxiety (Dean, Potts, & Barker, 2016). Whereas emotional expression via expressive writing is efficacious in producing favorable outcomes, a lack of emotional expression is problematic across the aforementioned outcomes and contexts.

Individuals having experienced a traumatic or stressful life event are more likely to repress thoughts and feelings about their experience compared to individuals who have not experienced such events, thereby subjecting them to potential negative outcomes related to a lack of emotional expression (Bodor, 2002). For example, Posttraumatic Stress Disorder (PTSD) diagnostic criteria are characterized by repeated attempts to cognitively or behaviorally avoid thoughts, feelings, or places related to a given trauma (American Psychiatric Association, 2013). Trauma patients who avoid intrusive thoughts or physiological sensations experience various forms of psychopathology, such as depression and trauma-related symptoms (Marx & Sloan, 2005), anxiety (Levitt, Brown, Orsillo, & Barlow, 2004), substance use (García-Oliva & Piqueras, 2016), and social concerns (Pennebaker, 1989; Pennebaker & Beall, 1986). Admittedly, the hypothetical nature of emotional inhibition makes it difficult to establish a causal relation between inexpression and the aforementioned symptoms. However, inhibiting thoughts or emotions is generally associated with impairments in physical and psychological health (Goldstein, Edelberg, Meier, & Davis, 1988; Gross & Levenson, 1997; Larson & Chastain, 1990). Although studies employing expressive writing have produced positive psychological and health-related outcomes, some of these studies neglect necessary considerations, the most important of which is whether or not the effects are meaningful (Smyth, 1998). For a more in-depth review of the efficacy of WED across contexts, the authors turn to previously-conducted meta-analyses.

Meta-Analytic Techniques

Meta-analyses allow researchers the opportunity to collectively examine the efficacy of different psychological interventions/tasks on outcome variables by calculating an overall, weighted, population effect (Borenstein, Hedges, & Rothstein, 2007; Glass, 1976; Hedges, 1982). The following meta-analyses delineate the efficacy of expressive writing across outcomes and warrant individual explanation: Smyth (1998); Frisina, Borod, & Lepore (2004); Frattaroli (2006); Mogk, Otte, Reinhold-Hurley, & Kröner-Herwig (2006); Van Emmerik, Reijntjes, & Kamphuis (2013); and Reinhold, Bürkner, & Holling (2018).

Smyth (1998) conducted the seminal meta-analysis examining the efficacy of expressive writing on psychological well-being, general health, and physical functioning. They included studies employing an expressive writing group and control group (i.e., neutral topic). In sum, thirteen studies/effect sizes were included, and the authors found an overall medium effect size, $d = 0.47$, for the experimental group compared to the control group. A later meta-analysis conducted by Frisina et al. (2004) expanded these analyses and included studies with clinical samples. This meta-analysis included nine studies and found an effect size of $d = 0.19$ for health-related outcomes and $d = 0.07$ for psychological outcomes. Mogk et al. (2006) conducted the next expressive writing meta-analysis to update the state of the literature regarding expressive writing. Studies employing Pennebaker's paradigm on experimental and control groups were included. Further, inclusion criteria were methodological techniques that included a four-week follow up and at least 10 participants. Thirty studies met inclusion criteria. Efficacy relating to somatic and psychological health outcomes were nonsigificant, corroborating findings from Frisina et al. (2004).

Frattaroli (2006) conducted perhaps the most notable meta-analysis to date examining the efficacy of emotional disclosure on the following constructs using only randomized and control conditions: psychological health, physiological functioning, reported health, health behaviors, and general functioning/life outcomes. Additionally, this meta-analysis was the first to employ random effects models, which estimate the mean of a proposed distribution of population effect sizes. Prior meta-analyses employed fixed effects models, which assume that all studies assess the same "true" population effect size. This assumption may be untenable across different populations (Borenstein et al., 2007). They included a wide range of studies, $N = 146$. Individual studies were again collapsed into one publication effect size, although these effects were also examined separately by health outcome. Overall, Frattaroli (2006) found $d = 0.16$ for all outcomes combined, which would be considered small. Additionally, they examined potential moderators and found larger effect sizes for the following samples: those with physical health problems, those with a history of having experienced traumatic or stressful events, samples not including college students, samples where expressive writing tasks were conducted at home and in private settings, paid participants, more male participants, and fewer participants (see Frattaroli, 2006 for a complete list of moderators). A recent analysis conducted by Van Emmerik et al. (2013) employing Pennebaker's paradigm included six eligible studies that compared treatment to control groups. In regards to inclusion criteria, they included studies where participants had a diagnosis of Acute Stress Disorder or PTSD. They found that those who participated in the expressive writing group experienced short-term reductions in PTS and comorbid depressive symptoms, combined d = 0.81.

The most recently published meta-analysis was conducted by Reinhold et al. (2018) and examined the efficacy of expressive writing on depression by randomizing participants to conditions (expressive writing vs. control). They included thirty-nine randomized controlled trials and excluded individuals with diagnoses of PTSD. This study did not support utilizing expressive writing for depression outcome measures for the specified sample, d = -0.09. Further, they found that expressive writing did not yield any type of long-term effect on depression outcomes. In sum, previous meta-analyses exhibit small to medium effect sizes for a brief, innocuous intervention and therefore individuals experiencing trauma have been shown to benefit from such interventions.

Posttraumatic Stress

Posttraumatic Stress Disorder is a disorder involving re-experiencing thoughts or events after a trauma. This generates a context where individuals are prone to affect-related deficiencies and maladaptive behaviors (American Psychiatric Association, 2013). DSM-5 criteria are based on twenty symptoms structured into four different subsets in those having experienced a traumatic event. These subsets are as follows: intrusion symptoms (i.e., re-experiencing), avoidance, negative alterations in cognition and mood, and increased arousal (Crespo & Gomez, 2016). While the renewed DSM-5 criteria are now increasingly utilized via structured clinical interviews, the current meta-analysis considers studies using DSM-IV criteria. DSM-IV criteria are similar and include the following: exposure to a traumatic event, intrusion, avoidance, and increased arousal (American Psychiatric Association, 2013). The studies employed in the current meta-analysis are divided according to these subsets (arousal, intrusion, and avoidance). Posttraumatic Stress Disorder affects a wide variety of populations, a few of which are sexual assault survivors (Klump, 2008), Iraq and Afghanistan war veterans (Gentes et al., 2014), and those exposed to natural disasters (Wang et al., 2000).

Research conducted on the efficacy of expressive writing on PTSD symptoms presents intriguing results. Sloan, Marx, Epstein, & Lexington (2007) examined individuals with at least moderate PTSD symptom severity and found that individuals assigned to an emotional expression writing condition reported fewer PTSD and depression symptoms during follow up. Sloan, Marx, & Greenberg (2011) found that PTSD symptoms decreased after a written emotional disclosure task, although this decrease was not significantly different than a control group change. Di Blasio et al. (2015) recruited women who had just given birth and assessed them a few days after experiencing childbirth along with a three-month follow-up. Results showed that women who had participated in the expressive writing task had lower depression and posttraumatic stress symptoms than the group assigned to a neutral writing condition. Additionally, regression models showed that expressive writing was significantly linked to a reduction of PTSD symptoms across different dimensional levels of symptom severity. Only 20 of the 113 women recruited for this study qualified for a diagnosis of PTSD, but those who reported mild symptomatology responded better to the task than those meeting criteria for PTSD. This limitation suggests that those with moderate distress could perhaps benefit more from an expressive writing task than those diagnosed with or meeting the qualifications for PTSD. It may also explain the differences in results in comparing to Sloan et al. (2011), as they found that those with a clinical diagnosis of PTSD did not respond to an emotional disclosure writing task. Perhaps it may be more advantageous to examine effect sizes separately for diagnoses of PTSD and subclinical symptoms.

Sloan, Marx, Bovin, Feinstein, & Gallagher (2012) adapted a writing protocol to focus primarily on the emotions, meaning, and "hot spots" associated with the trauma. They referred to this procedure as the written exposure therapy (WET) protocol, distinguishable from the paradigm adapted by Pennebaker & Beall (1986). In their seminal study examining the efficacy of WET for motor-vehicle accident related PTSD, they found that those in the WET condition experienced significant reductions in PTSD symptoms throughout the course of the study. Since then, a small number of other studies employing the WET procedure have been employed in those with PTSD. Indeed, Sloan, Marx, Lee, & Resick (2018) found that WET was noninferior (i.e., just as effective) as Cognitive Processing Therapy, considered first-line treatment for PTSD. Further, treatment gains were maintained at 24 and 36-week follow up. While studies employing this protocol will be included in the current review, the newness of this protocol does not allow exclusive examination using meta-analytic techniques.

Posttraumatic Growth

While the literature mostly discusses potentially harmful outcomes to traumatic events such as emotional distress, traumatic events also provide opportunities for personal growth (Aslam & Kamal, 2013). Traumatic events, either natural or human-inflicted, may lead to positive outcomes by allowing the individual to take a different perspective (Cobb, Tedeschi, Calhoun, & Cann, 2006; Taku, Calhoun, Cann, & Tedeschi, 2008). The relationship between positive growth after a traumatic event and symptom reduction is unclear, as it is a complex process. Thus, it is necessary to examine how expressive writing might influence each variable separately, which is one of the key goals of this meta-analysis (Slavin-Spenny, Cohen, Oberleitner, & Lumley, 2011). Models receiving empirical support within the last decade suggest that traumatic events offer opportunities for both negative and positive experiences (Tedeschi & Calhoun, 1995; Weiss, 2002). Posttraumatic Growth (PTG) is a positive experience after a traumatic event (Aslam & Kamal, 2013; Yilmaz & Zara, 2016). Specifically, PTG is classified as broad cognitive benefits that are seen after a traumatic experience. These benefits can be categorized into building closer relationships, examining new possibilities, appreciating life, recognizing personal strengths, and undergoing spiritual changes (Dursun, Steger, Bentele, & Schulenberg, 2016; Tedeschi & Calhoun, 2004).

PTG is associated with a variety of desired outcomes (Dursun et al., 2016). PTG has been studied in those experiencing natural disasters, war, and other harms such as sexual assault. Finally, PTG has been studied in those experiencing medical diagnoses such as different types of cancer and diseases. Although the relationship between PTG and symptom reduction is not yet fully understood, perhaps expressive writing allows the individual to fully comprehend the event. Pennebaker & Graybeal (2001) speculated that expressive writing allows an individual to feel more connected with his or her surroundings. Although this speculation does not directly explain positive outcomes after an expressive writing task, perhaps individuals gain a better appreciation for life after gaining a better sense of connectedness with that individual's surroundings. One might expect effect sizes to be larger for those studies requiring a diagnosis of PTSD, as such growth may not be possible in those with subclinical symptomatology.

Quality of Life

Quality of Life (QOL), according to Theofilou (2013) is an evaluation of the "goodness" that an individual experiences, separated into domains of reactions to life events, disposition, life fulfillment, and satisfaction with life experiences. More generally, QOL refers to an individual's attitude towards the target life situation (Costanza et al., 2007), delineated into objective and subjective components. Objectively, QOL refers to components outside of an individual and measurable by others, while subjective QOL is an individual's assessment of his or her own experiences (Costanza et al., 2007). The current meta-analysis will focus solely on the subjective components of QOL, as it is obtainable through questionnaires. Pennebaker & Graybeal (2001) suggested that expressive writing allows one to feel more connected with their surroundings. Further, they explain that expressive writing allows people to see things in a different way and better understand themselves. By understanding a traumatic or stressful event, one is said to see things differently and perhaps look at the situation with a more positive mindset. The changes that occur after expressive writing may also allow one to find meaning in the traumatic event, thereby increasing the QOL of that individual (Frankl, 1959). Higher QOL may be considered a type of PTG, which is why the current meta-analysis sought to examine the efficacy of studies utilizing expressive writing to improve QOL and PTG in the same study.

Current Meta-Analysis

The purpose of the current meta-analysis is to examine studies employing expressive writing procedures using Pennebaker's paradigm (WED) and the more recent WET protocol on variables relevant to the field of positive psychology (PTG and QOL) and PTS, with effect sizes separated by the paper's indication of PTSD diagnosis when sample sizes are large enough. Based on recently published literature regarding efficacy of expressive writing for different levels of PTSD symptoms, this diagnostic marker is an important facet to consider (Di Blasio et al., 2015; Reinhold et al., 2018; Sloan et al., 2011). No review has examined the efficacy of expressive writing on PTS separated by diagnosis. Additionally, no meta-analysis has been conducted that examines the efficacy of expressive writing on positive outcome variables such as PTG and QOL, in line with the field of positive psychology. The meta-analyses described sequentially above also focused on experimental versus control group effect sizes or *p*-values, rather than emphasizing change for the expressive writing group. This focus is likely because of the analyses provided in these publications, especially when using randomized controlled trial research designs. While this design is the gold standard for medicine, the current meta-analysis sought to examine the magnitude of change for participants who experienced an expressive writing task. For example, a comparison group may increase their quality of life scores by two points in a controlled study, while the experimental group increases their quality of life scores by four points; thus, creating a significant difference in change between the two groups. This information is valuable, but it does not tell the reader the magnitude of the change for the writing group, wherein four points might only be a small effect when examined within the group who received the writing task.

This analysis will also focus on changes across time for groups who received the expressive writing task to determine what size of effects one might expect given a specific measurement schedule (i.e., one to three months, three months to six months, etc.). Indeed, Sloan et al. (2018) discovered long-term gains for those in the WET condition. This analysis should present researchers with a renewed examination of the efficacy of expressive writing on the aforementioned variables using newer meta-analytic techniques. Newer methods of meta-analysis, including *p*-curve (Simonsohn, Nelson, & Simmons, 2014; Simonsohn, Simmons, & Nelson, 2015), *p*-uniform (Van Aert, Wicherts, & Van Assen, 2016), PET-PEESE (Stanley & Doucouliagos, 2014), selection models (Vevea & Hedges, 1995), and trim and fill methods (Carter & McCullough, 2014) allow for better estimation of meta-analytic effect sizes. These analyses would be best performed by examining each potential effect separately, rather than averaging effects of each publication into one study effect size (a common trend in the previously mentioned meta-analysis). In addition to an estimate of overall effect sizes using updated techniques, the current meta-analysis estimates power for effects on writing groups, as research has shown a consistent under powering of psychological studies, combined with a misunderstanding of the sample size needed for adequately powering one's work (Bakker, Hartgerink, Wicherts, & Van Der Maas, 2016).

Method

Data Collection

Studies were collected through online databases, such as PsycINFO and Google Scholar, using the following search terms and their combinations: *Posttraumatic Growth, PTG, Quality of Life, QOL, Posttraumatic Stress, PTS, Expressive Writing, Emotional*

Disclosure, Written Emotional Disclosure (WED), Written Exposure Therapy (WET). Within these articles, the change in outcome variables (PTS, PTG, QOL) from pre- to post-test was the dependent variable of interest. Generally, groups were separated into an experimental and control group and then examined at different time points. For purposes of this meta-analysis, only participants assigned to the experimental condition were examined due to having received the expressive writing task. If a study included multiple assessment time points, then these measurements were examined sequentially (i.e., time 1 to time 2, time 2 to time 3) to determine change across time for the dependent variable. The time variable was coded as the number of months between two comparison points. For example, if a study included three time points (baseline, one month, three months), two pairwise effect sizes would be calculated (baseline to one month, one month to three months) and the time variable would be one month for comparison one and two months for comparison two. If a study included multiple experimental conditions (i.e., different instructions or forms for WED), all experimental conditions were included in the dataset.

264 citations focusing on PTS, PTG, and QOL were identified through the literature search and previous meta-analyses. Citations for PTS were separated by diagnostic criteria (intrusions, avoidance, and hyperarousal), where possible. After screening these studies, 53 articles were retained for containing the appropriate information for this meta-analysis. This manuscript was written with *papaja* in *R* (Aust & Barth, 2017) with the analyses inline with the text. The complete set of data, excluded article list with reasoning, and other relevant information can be found at: https://osf.io/4mjqt. Generally, studies were included if they utilized WED or WET, included relevant numbers to compute an effect size, and included the relevant outcome variables. The questionnaire for each relevant outcome variable is coded in the online data provided on the Open Science Framework (link above). These varied across study, however, the nature of Cohen's *d* allows for different Likert-type scales, as it takes into consideration the study standard deviation in the denominator to create standardized scores for comparison across studies.

After having two reviewers independently code articles, 223 effect sizes were calculated. On average, each study represented $M = 4.21$, $SD = 3.31$ effects, ranging from 1 to 16 effects. 163 effects were calculated for PTS, 21 for PTG, and 37 for QOL. Studies were coded for PTSD diagnosis as no (not mentioned or not included), mixed (mentioned number of participants but all included), and yes (included as criteria). After examining the number of effects in each of these categories for each variable, only the PTS results will be split by PTSD diagnosis with 16 no mention, 16 in the mixed category, and 86 yeses.

Calculations for Effect Size, Variance, and Confidence Intervals

For our purposes, we used Cohen's (1988) standards for nomenclature for small (0.20), medium (0.50), and large (0.80) d values, although it is important to note that Cohen himself suggested that these values should be based on the area of study. Generally, however, these effect size criteria are used within the social sciences. Each study implemented a pre-test to post-test style repeated measures design, usually with paired t-tests, ANOVA, or regression analyses. The means, standard deviations, and N values were collected from each study. In general, Cohen's d values were calculated using the following formula for paired t using means and standard deviations for each time point:

$$d_{av} = \frac{M_1 - M_2}{\frac{SD_1 + SD_2}{2}}$$

This equation is described in detail in Cumming (2012) as an alternative to the traditional calculation of d for paired samples t, wherein the denominator is the standard deviation of the difference scores:

$$d_z = \frac{M_1 - M_2}{SD_{diff}}$$

This equation for d_{av} not only allows for calculations from published articles that do not include SD_{diff} (i.e., most articles included), but also has been shown to be less upwardly biased than d_z. Alternative formulas include controlling for r between paired levels, as described in Lakens (2013); however, these values were not available in the selected articles, and Lakens also recommends d_{av} as an effect size for paired designs. When only mean differences and standard deviation of the difference scores were available, the second equation for d_z was used.

We planned to use traditional and newer methods of meta-analysis, following guidelines from Cooper, Hedges, & Valentine (2009) and Borenstein et al. (2007), as well as Van Aert et al. (2016). Sampling variance of the effect sizes were estimated using the *escalc()* function from the *metafor* package in *R* (Viechtbauer, 2010). The variance formula was originally published in Morris & DeShon (2002) and is shown below:

$$v = \frac{1}{n}(\frac{n-1}{n-3})(1 + n * d^2) - \frac{d^2}{[c(n-1)]^2}$$

In this formula, n is the number of paired observations, d is the calculated effect size, and c is a correction factor, wherein df are $n - 1$ (Hedges, 1982):

$$c = 1 - \frac{3}{4 * df - 1}$$

We used the *metagen()* function in the *metafor* package to calculate both fixed and random effects models, which uses standard error of the effect to calculate overall estimates of an effect and their confidence intervals. Thus, we took the square root of the variance estimate for standard error. Given these calculations, the goal of this analysis was to calculate a combined effect size, along with a confidence interval for study planning and an assessment of the literature. A fixed effects model requires the assumption that there is a true population effect size across all studies. By including multiple measures of psychological outcomes, this assumption may be tenuous, and therefore, a random effects model was also calculated. In random effects models, the true effect is assumed to vary across studies (Borenstein et al., 2007). For a fixed effects model, the effect sizes are weighted by their inverse variance (v; Sánchez-Meca & Marín-Martínez, 2008), which is calculated automatically in *metafor* by:

$$w_i^{FE} = \frac{1}{v}$$

The advantage to this procedure is that analyses are weighted by their precision, that is, that studies with more information (often, larger samples), are given larger weights in the overall estimated effect size (Borenstein et al., 2007). Random effects models are also weighted by inverse variance, with an additional correction for variance between studies, τ^2_{DL}, as described by DerSimonian & Laird (1986):

$$w_i^{RE} = \frac{1}{v + \tau^2_{DL}}$$

Confidence intervals were calculated in two ways for this study. Cumming (2012), Kelley (2007), and Smithson (2001) have shown that the distribution of d values are non-normal, and thus, CIs should be estimated using the non-centrality parameter and a non-normal distribution. These values were calculated using the functions in the *MOTE* library which iteratively estimates the appropriate non-centrality parameter and converts back to d values (i.e., non-centrality parameter divided by the square root of n; Buchanan, Valentine, & Scofield, 2017; Smithson, 2001, 2003). However, the *metafor* package in *R* uses central distributions to estimate CIs for each study and overall effect sizes. Therefore, we present both sets of values for the interested reader, as meta-analytic procedures have not implemented non-central distributions of effect sizes.

Additional Meta-Analytic Techniques

p-Curve and p-Uniform. We used *p*-curve.com to conduct a *p*-curve analysis (Simonsohn et al., 2014). The purpose of this type of analysis is to detect true effects. Specifically, *p*-curve is used to reveal possible *p*-hacking in published literature in order to decipher whether or not a true effect exists. Broadly, *p*-hacking occurs when researchers use questionable research practices to create significant results by manipulating dependent variables or covariates. Additionally, authors may add participants if the initial findings are not significant (Bruns & Ioannidis, 2016). Researchers may also decide to exclude participants for final analyses if that exclusion leads to a significant difference (John, Loewenstein, & Prelec, 2012). Thus, it is necessary to distinguish between true and false effects in order to effectively interpret effect sizes corresponding to those p-values. p-curve accomplishes this task by examining the distributions of the published p-values. If an effect exists, or rather the results should be interpreted as presented, the distribution of p-values will be positively skewed (Simonsohn et al., 2014). If, however, no effect exists, then the distribution of p-values will be flat. p-curve analyses ultimately provide evidence of p-hacking in groups of studies and has become an important tool for interpreting meta-analyses. In order to accurately estimate effect sizes because of scrutiny associated with effect size estimation of p-curve, we also conducted p-uniform. p-uniform analyses, too, are interpreted by examining the distribution of p-values in a set of studies (Van Aert et al., 2016). However, it is assumed that the population effect size equals the effect size from the dataset. Because of this assumption, the population effect size is referred to as uniform. This analysis also examines for publication bias and presents the researcher with a corrected effect size. Publication bias occurs when only select studies are published, usually only significant studies, although many factors can bias a study's publication (McShane, Böckenholt, & Hansen, 2016). p-uniform was calculated from code provided by Van Aert (2017) on GitHub.

PET-PEESE. Originally, meta-analyses relied on the calculation of Egger's regression test which examined the relationship of the standard error (predictor) to the effect size estimates (criterion). In this regression, the intercept values were used to determine if effect size measures were different than zero, by providing a meta-analytic estimate (Egger, Davey Smith, Schneider, & Minder, 1997; Stanley, 2005). PET-PEESE analyses examine for publication bias by adapting parts from Egger's traditional regression tests: PET (Precision Effect Test) and PEESE (Precision Effect Estimate with Standard Error, Carter & McCullough, 2014). PET is a more reliable test of publication bias with effect size estimates of zero, $b_0 = 0$, while PEESE is more accurate with non-zero effect size estimates, $b_0 \neq 0$

(Stanley & Doucouliagos, 2014). PET-PEESE was calculated using Hilgard's (2016) code provided on GitHub.

Selection Models. Selection model analyses provide the researcher with a test of publication bias and effect size estimates using maximum likelihood estimation (Vevea & Hedges, 1995; Vevea & Woods, 2005). Using selection models, researchers are able to discover effect size estimates as well as evidence of publication bias (McShane et al., 2016) by using a mixed general linear model to estimate these values. Selection models were calculated with the *weightr* package in *R* (Coburn & Vevea, 2017).

Trim and Fill. Trim and Fill analyses, in contrast to PET-PEESE, regress standard error (criterion) and effect size estimates (predictor). Specifically, the purpose of Trim and Fill techniques is to examine whether or not publication bias may influence the regression equation (Carter & McCullough, 2014). Effect sizes and standard error terms are graphically displayed on x and y-axes, respectively, in a funnel plot. If this graphical representation indicates asymmetry, considered a gap of missing data points in the lower center area of the plot, the study set can be assumed to have studies that are both non-significant and small in sample size (Van Assen, Van Aert, & Wicherts, 2015). This funnel is then trimmed until symmetry is achieved. Missing studies from the symmetrical graph are imputed (filled) while maintaining the given symmetry (Duval & Tweedie, 2000). The meta-analytic effect size is then estimated from the trimmed and filled funnel plot. Trim and fill analyses, as well as funnel plots included below, were calculated with the *metafor* package.

Results

Posttraumatic Stress

Overall Effect Size. As described above, both fixed effects and random effects models with centralized confidence intervals are presented in Table 1. Studies were examined for potential outliers using the *metafor* package in *R*. This package calculates traditional regression influence values, such as Cook's and hat values (Cohen, 1988). These values indicate change in overall meta-analytic model with and without the effect; thus, determining their impact on the pooled effect size (Viechtbauer, 2010). Because published studies likely represent the range of the sampling distribution of effect sizes, we included the analyses with and without outliers to present evidence for both paths a researcher might take when examining an overall effect.

2 outliers were detected with this procedure, all showing very large effect sizes, average $d = 2.81$. The fixed and random effects estimates without these points are also included in Table 1. Figures 1, 2, 3, and 4 portray the effect sizes for PTS studies, separated by intrusions, avoidance, hyperarousal, and total scores for easier viewing (i.e., over 100+ effect sizes did not fit easily on one combined graph). Although these categories are not reflective of updated DSM-5 criteria, researchers have not yet conducted enough studies using expressive writing on PTS with updated PTSD criteria to warrant a meta-analysis. Name acronym coding can be found in the data online. This forest plot includes the non-centralized confidence interval calculated from the *MOTE* library (Buchanan et al., 2017). Shape size indicates study weight, and these values were taken from the overall random effects meta-analysis and normalized by dividing by the mean weight. The dashed lines indicate the average non-weighted lower and upper confidence interval limit for the non-centralized estimates. Overall, PTS studies include a small effect size that appears to be significantly greater than zero across all estimate types (fixed, random, with or without outliers).

We further calculated the overall effect sizes by PTSD diagnosis category using a random effects model. Studies only including individuals with a diagnosis of PTSD exhibited a medium effect size (before and after outlier exclusion): with outliers $d = 0.64$ [0.48, 0.79]; without outliers $d = 0.55$ [0.41, 0.69], while studies not requiring (or listing) a PTSD diagnosis showed a small to medium effect size: $d = 0.32$ [0.24, 0.40]. Similarly, the mixed category showed a small to medium effect size : $d = 0.35$ [0.16, 0.54]. Complete estimates of all the following analyses split by diagnosis are included online at https://osf.io/4mjqt/, and their pattern of results is similar to the overall pattern here.

Homogeneity. A prerequisite for newer meta-analytic techniques includes the assessment of homogeneity of the effects (Van Aert et al., 2016). Using the *metafor* package in *R*, we calculated the Q-statistic and the I^2 index (Cochran, 1954; Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). Significant values imply inconsistencies across the variable or variables of interest and are represented by Q. In contrast, I^2 indicates the percentage of heterogeneity along with a 95% CI. Both can, however, be biased with a small number of experiments included for analyses (Higgins, Thompson, Deeks, & Altman, 2003; Huedo-Medina et al., 2006). Thus, we sought to calculate an overall level of heterogeneity after examining each variable separately before and after excluding outliers. For PTS studies including outliers, we found significant heterogeneity, $Q(162) = 776.74$, $p < .001$ and $I^2 = 79.1$, 95% CI[75.9 - 81.9]. These values were reduced slightly with the exclusion of outliers, $Q(160) = 677.98$, $p < .001$ and $I^2 = 76.4$, 95% CI[72.6 - 79.7].

Power. Power was calculated in two different ways using the *pwr* package in *R* (Champely, 2016). *Post hoc* power was first calculated using sample size and effect size statistics from each individual study. Additionally, we calculated power using the study sample size and estimated overall effect size from the random effects model with and without outliers, as explained by Francis (2012) and Francis (2014). The first estimate indicates the likelihood of finding an effect from our sample statistics, while the second indicates the likelihood of finding the true population effect size. If each study had been conducted on only the change in the experimental group, 46.6% of studies would have been considered significant at $\alpha < .05$. The average power of these studies based on their original study characteristics was .48 ($SD = .36$). Power for the random-effects meta-analytic effect size with outliers was .52 ($SD = .25$) and without outliers was .49 ($SD = .25$). Therefore, power consistently was around 40-50% for studies examining PTS, regardless of outlier effects. In these studies, only 28.8% achieved recommended 80% power for their found effect size, a smaller 24.5% for the random-effect outlier effect size, and even smaller 20.2% for power calculations on the random-effect size without the outliers.

Other Meta-Analytic Estimates. As noted in Van Aert et al. (2016), p-curve and p-uniform analyses are upwardly biased when heterogeneity is high. Therefore, we use caution when interpreting these analyses on PTS outcomes. As seen in Table 1, the estimates for p-uniform were higher than other techniques, likely because of the focus on significant p-values and the great degree of heterogeneity described earlier. P-curve pictures can be found at https://osf.io/4mjqt/ online, and this analysis indicated evidentiary value at $p < .001$. Additionally, the p-uniform analysis indicated that there was likely no publication bias present, Z = -5.71, p = 1.000. When examining the PET analysis, we found that the intercept was significant, which indicated that PEESE was likely a better estimator of the meta-analytic effect size. PEESE estimates were lower than the original meta-analytic estimate, but confidence intervals indicated that the effect is small to medium, and still larger than zero. Selection models indicated a larger effect size, especially with the random-effects models, and these effects were influenced by the outliers found in the published studies. Trim and fill models are shown in Table 1, and figures are included online. Nineteen missing studies were imputed for both models with and without outliers. Across all these effect size estimates, we found that expressive writing was likely to decrease PTS symptoms in a small to moderate way. The correlation of effect size with time between measurement times was $r = -.01$, 95% CI $[-.17, .14]$, $t(161) = -0.15$, $p = .879$, and $r = -.08$, 95% CI $[-.23, .08]$, $t(159) = -1.00$, $p = .320$ without outliers. This result indicated that the effect of expressive writing slightly decreased across time.

Postraumatic Growth

Overall Effect Size. Both fixed and random effects models with centralized confidence intervals for PTG are presented in Table 2. When examining expressive writing on PTG, no outliers were detected. Fixed and random effects estimates are included in Table 2, while Figure 5 shows effect sizes for PTG studies where shape size indicates the normalized weight of the study. Dashed lines indicate non-weighted lower and upper confidence intervals for non-centralized estimates. Overall, PTG studies indicated a negligible to small effect size across both random and fixed effects models, and the non-centralized confidence intervals indicated an effect that crossed zero.

Homogeneity. Using the *metafor* package in *R*, we calculated both a Q statistic and I^2 index. Since PTG studied did not contain any outliers, we did not calculate two separate analyses to examine heterogeneity both with and without outliers. We did not find significant heterogeneity across PTG studies, $Q(20) = 14.18$, $p = .821$ and $I^2 = 0.0$, 95% CI[0.0 - 25.3].

Power. First, we calculated *post hoc* power using both sample and effect size statistics from individual studies. Individual studies examining change in experimental groups showed that 9.5% of studies would have been considered significant at $\alpha < .05$. Average power of PTG studies was .15 ($SD = .16$). 0.0% achieved recommended 80% power for their found effect size. Additionally, we calculated power using study sample size and estimated effect size from our random effects model. Power for the true effect size was .08 ($SD = .02$). Again, 0.0% achieved recommended 80% power.

Other Meta-Analytic Estimates. Due to no heterogeneity across PTG studies, we can use both p-curve and p-uniform analyses with more confidence. A pictorial representation of p-curve can be found at https://osf.io/4mjqt/. This analysis did not indicate evidentiary value, $p = .75$, as only two of the results would be considered significant at $\alpha < .05$. p-uniform estimates are presented in Table 2. Specifically, these analyses indicated that there was no publication bias present, $Z = 0.70$, $p = .243$. The p-uniform estimates of the effect size for PTG were negative, in contrast to the fixed and random effects overall model. The confidence interval for this analysis indicates a wide range of possible effects. In examining PET-PEESE analyses, we did not find a significant intercept, indicating that PET is most likely a better effect size estimator. PET analyses indicated that the effect size is negligible to small, with our confidence interval crossing zero. These results corroborated our original effect size calculations. Selection models indicated negligible to small effect sizes, again wherein the confidence interval includes zero effect. Trim and fill models are shown in Table 2, and figures are included online. Zero studies were imputed for our model, and thus, the effect size estimate is the same as the overall model. Across techniques, we found that expressive writing has little to no effect on PTG. The correlation of effect size across measurement times in PTG studies at subsequent time points was r = .09, 95% CI [−.36, .50], $t(19) = 0.38$, p = .707, and no change over time was found.

Quality of Life

Overall Effect Size. Finally, for QOL, both fixed and random effects models with centralized confidence intervals are presented in Table 3. Two outliers were detected with this procedure, average d = -0.07. While the average effect of these outliers indicates a small number, it is important to note that these two outliers were the largest positive and negative effects found from the Possemato, Ouimette, & Geller (2010) study. Fixed and random effects estimates without these points are also included in Table 3, while Figure 6 shows effect sizes for QOL studies. Overall, QOL studies indicated a negligible to small effect that showed a non-significant decrease in quality of life as a result of expressive writing.

Homogeneity. For QOL studies including outliers, we found significant heterogeneity from our random effects model, $Q(36) = 200.09$, $p < .001$ and $I^2 = 82.0$, 95% CI[75.9 - 86.5]. After excluding outliers, our random effects model still indicated heterogeneity,$Q(34) = 93.18$, $p < .001$ and $I^2 = 63.5$, 95% CI[47.6 - 74.6].

Power. In conducting *post hoc* power using sample and effect size statistics from individual studies, we found that 21.6% of studies would have been considered significant at $\alpha < .05$. Average power based on actual study characteristics was .33 (SD = .32). Power for the random effects meta-analytic effect size with outliers was .05 (SD = .00) and without outliers was .05 (SD = .00). Unfortunately, power was around 5% for both random effects models with and without outliers. In these studies, 18.9% achieved adequate power of 80% on their found effect size, while 0.0% achieved 80% power for our random effects model with outliers. Finally, without outliers, 0.0% achieved 80% power.

Other Meta-Analytic Estimates. We exert caution in interpreting p-curve and p-uniform analyses on QOL outcomes with and without outliers due to heterogeneity. As seen in Table 1, p-uniform estimates were stronger and positive than other techniques because of the high degree of heterogeneity recently described. p-curve pictures can be found at the following OSF Link: https://osf.io/4mjqt. Eight studies were significant at $\alpha < .05$, and the studies indicated evidentiary value, $p = .004$. p-uniform analyses did not indicate publication bias, Z = -2.75, $p = .997$. In PET-PEESE analyses, we found that the intercept was not significant, and therefore, PET was a better estimator of the meta-analytic effect. Table 1 indicates that both of these analyses estimate the effect size around zero, with a confidence interval that includes zero. Selection models correspondingly show small effects crossing zero, except for random effects models with outliers, that appear to be heavily influenced by the outliers. Trim and fill models are shown in Table 3, and figures are included online. No studies were imputed for these analyses, and therefore, the effect size estimates match the original meta-analysis. Overall, these results appear to point to no effects, ranging across zero with several negative estimates. Interestingly, the correlation of effect sizes across measurement times with outliers was $r = -.37$, 95% CI $[-.62, -.05]$, $t(35) = -2.33$, $p = .026$ and $r = -.64$, 95% CI $[-.80, -.39]$, $t(33) = -4.75$, $p < .001$ without outliers. The effect of expressive writing appears to be positive at short time intervals and decreases into negative effects at longer time intervals.

Discussion

In examining pre- to post-test comparisons across each variable separately, we found that PTS studies indicated a small effect size across all meta-analytic estimates. As mentioned, PTS is operationally defined as re-experiencing thoughts and feelings associated with a traumatic event and subsequently seeking to avoid these thoughts and feelings. DSM-IV criteria for a PTSD diagnosis include exposure to a traumatic event, intrusions, avoidance, and hyperarousal. Interestingly, those studies requiring a diagnosis of PTSD for inclusion resulted in a medium effect size, while those studies not requiring a PTSD diagnosis resulted in a small to medium effect size. These results suggest that those with clinical symptoms of PTSD may benefit more from expressive writing interventions. Further, these results are in contrast to recently-conducted studies, which suggest that those with subclinical symptoms benefit more from expressive writing tasks (Di Blasio et al., 2015; Sloan et al., 2011). Both QOL and PTG studies indicated a negligible to small effect size using random effects models. Although the PTG effect in our overall meta-analysis estimate was significant, other methods indicate this small effect is likely not different from zero. These results should be considered within the context of the intervention. Given that expressive writing is an innocuous, easy-to-administer intervention, even small effect sizes should be considered important when interpreting these results. While small, these effect sizes exhibit a profound impact of expressive writing on PTS.

Additionally, our analyses focus on the change for the experimental group across time, rather than an experimental group to a control group. This focus allowed us to estimate the changes for individuals who received a WED/WET intervention, therefore estimating the impact on participants who used written expression. Potentially, these effects could be contributed to other factors (such as the simple passage of time), but we demonstrate here that for both PTS and PTG, there was no relationship between effect size and time. For QOL studies, a medium to large negative correlation was found. A negative relationship between time and effect size implies that writing tasks were more effective in the initial time points, and effects decreased over longer time spans.

The authors note several limitations. Generally, ineffective emotional expression may be a contributing factor. If participants/clients are not deeply engaged with the material, an expressive writing task may not be effective, as Pennebaker & Graybeal (2001) imply that connectedness is an important factor for the task. However, it may be difficult to implement a check for engagement in these types of research designs. Doing so may also set a context that will inhibit emotional processing and general responses. Research on expressive writing has found a wide range of outcomes for different variables (Frattaroli, 2006), and these various results may explain the large heterogeneity found in this study. Encouragingly, we did not find much evidence of publication bias, and therefore, these estimates may represent a true population effect size. Regardless, methodology of expressive writing studies is variable, as it is applied in different forms across different contexts. Ideally, it would be possible to control for these varied instructions and protocols. However, this is simply not feasible, as most studies do not use measures that examine how engaged an individual is with the material. As such, this current meta-analysis sought to provide readers with a global effect of expressive writing on the aforementioned outcome variables. More studies are needed to examine potential moderating effects of participant engagement.

The authors also note limitations in regards to the specific outcome variables. The nature of the construct of PTG makes it difficult to analyze rigorously. For example, on the Posttraumatic Growth Inventory (commonly used to study PTG), one could respond 0 to the item "I have a greater appreciation for the value in my own life" because they already had a high level of appreciation in their life (i.e., ceiling effect). This conceptual issue may account for the non-effect of expressive writing on PTG. Logically, it would be difficult to determine whether or not an individual experiences growth from trauma without having experienced trauma. In conducting the literature search for the present meta-analysis, an insufficient number of studies requiring a diagnosis of PTSD employed PTG as an outcome variable. Thus, it was difficult to determine whether participants in the studies employed had experienced trauma in line with DSM-IV criteria. For PTS, studies not specifying whether or not participants had a diagnosis of PTSD were included. It is possible that studies included in the subclinical symptom category did in fact include participants without PTSD diagnosis (perhaps it was simply not assessed by means of a structured clinical interview). It is also crucial to consider mainstream issues not specific to expressive writing and the outcome variables utilized in the present study.

The psychological scientific community has shifted focus to reproducibility and research design in the last several years (Nelson, Simmons, & Simonsohn, 2018), and much of this discussion has focused on adequately powering studies for publication (Bakker et al., 2016; Maxwell, Lau, & Howard, 2015). Maxwell et al. (2015) and Open Science Collaboration (2015) have shown that the "replication crisis" may be attributed to low power in published studies. The power found in the current meta-analysis was very poor, with very few studies reaching the suggested 80% criterion to adequately power their study. This result was the same when considering individual study characteristics or the estimate true population effect size. Research by Bakker et al. (2016) indicates that researchers' intuitions about power are particularly poor, and many studies could benefit from more informed power analyses. Although, personnel and time required to conduct an expressive writing study is high. While the expressive writing task itself is relatively easy to administer, screening multiple participants and collecting data at multiple time points is time consuming. Anderson, Kelley, & Maxwell (2017) recently published a primer on power, with an online application to help with sample size planning for many types of research designs. Additionally, we encourage researchers to report power analyses of studies in order to better understand methodology for replication and reproducibility.

Meta-analyses, while useful tools to pool for population effect sizes, contain various limitations to their usefulness (Van Elk et al., 2015). As mentioned previously, these analyses can be affected by high heterogeneity, which was found in this study (Van Aert et al., 2016). Selection models have been criticized when using a smaller number of studies (Van Assen et al., 2015), and trim and fill analyses may not always estimate accurate confidence intervals and funnel plots may be biased with heterogeneity (Terrin, Schmid, Lau, & Olkin, 2003). When focusing on improving the psychological sciences, Van Elk et al. (2015) suggest that the reliability and size of effects may be best elucidated by conducting large preregistered studies. This suggestion will also improve the outlook for power in published studies, and projects such as Many Labs and the Psychological Science Accelerator can aide in subsidizing large samples (Klein et al., 2014; Moshontz et al., 2018). For example, studies can be proposed to the Psychological Science Accelerator and labs across the globe can be recruited to improve sample size for a study, which is a similar procedure to the Many Labs projects. Distributed networks of research teams can solve the problems with power that are present across all types of psychological research (Bakker et al., 2016). Even with limitations, meta-analyses allow researchers to examine the state of a research area, and we find potential with expressive writing on reducing PTS symptoms, and an overall need for better sample size and power planning for studies.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (Fifth.). Washington, DC: American Psychiatric Association. doi:10.1176/appi.books.9780890425596.744053

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, *28*(11), 1547–1562. doi:10.1177/0956797617723724

Aslam, N., & Kamal, A. (2013). Gender differences in distress responses, rumination patterns, perceived social support and posttraumatic growth among flood affected individuals. *Journal of Pakistan Psychiatric Society*, *10*, 86–90.

Aust, F., & Barth, M. (2017). papaja: Create APA manuscripts with R Markdown. Retrieved from https://github.com/crsh/papaja

Bakker, M., Hartgerink, C. H. J., Wicherts, J. M., & Van Der Maas, H. L. J. (2016). Researchers' intuitions about power in psychological research. *Psychological Science*, *27*(8), 1069–1077. doi:10.1177/0956797616647519

Bodor, N. Z. (2002). *The health effects of emotional disclosure for individuals with Type 1 diabetes* (PhD thesis No. 10-B).

Borenstein, M., Hedges, L. V., & Rothstein, H. (2007). Meta-analysis fixed effect vs. random effects. Retrieved from https://www.meta-analysis.com/downloads/Meta-analysis fixed effect vs random effects 072607.pdf

Bruns, S. B., & Ioannidis, J. P. A. (2016). p-Curve and p-Hacking in observational research. *PLOS ONE*, *11*(2), e0149144. doi:10.1371/journal.pone.0149144

Buchanan, E. M., Valentine, K. D., & Scofield, J. E. (2017). MOTE. Retrieved from https://github.com/doomlab/MOTE

Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in*

Psychology, *5*(July), 1–11. doi:10.3389/fpsyg.2014.00823

Champely, S. (2016). pwr: Basic functions for power analysis. R package version 1.2-0. Retrieved from https://cran.r-project.org/package=pwr

Cobb, A. R., Tedeschi, R. G., Calhoun, L. G., & Cann, A. (2006). Correlates of posttraumatic growth in survivors of intimate partner violence. *Journal of Traumatic Stress*, *19*(6), 895–903. doi:10.1002/jts.20171

Coburn, K. M., & Vevea, J. L. (2017). Weightr. Retrieved from https://cran.r-project.org/web/packages/weightr/index.html

Cochran, W. G. (1954). Some methods for strengthening the common χ 2 tests. *Biometrics*, *10*(4), 417–451. doi:10.2307/3001616

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Earlbaum.

Cooper, H., Hedges, L. V., & Valentine, J. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.

Costanza, R., Fisher, B., Ali, S., Beer, C., Bond, L., Boumans, R., ... Snapp, R. (2007). Quality of life: An approach integrating opportunities, human needs, and subjective well-being. *Ecological Economics*, *61*(2-3), 267–276. doi:10.1016/j.ecolecon.2006.02.023

Crespo, M., & Gomez, M. M. (2016). Diagnostic concordance of DSM-IV and DSM-5 posttraumatic stress disorder (PTSD) in a clinical sample. *Psicothema*, *28*(2), 161–166. doi:10.7334/psicothema2015.213

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Dean, J., Potts, H. W., & Barker, C. (2016). Direction to an internet support group compared with online expressive writing for people with depression and anxiety: A randomized trial. *Journal of Medical Internet Research*, *3*(2), e12.

doi:10.2196/mental.5133

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188. doi:10.1016/0197-2456(86)90046-2

Di Blasio, P., Camisasca, E., Caravita, S. C. S., Ionio, C., Milani, L., Valtolina, G. G., … Valtolina, G. G. (2015). The effects of expressive writing on postpartum depression and posttraumatic stress symptoms. *Psychological Reports*, *117*(3), 856–882. doi:10.2466/02.13.PR0.117c29z3

Dursun, P., Steger, M. F., Bentele, C., & Schulenberg, S. E. (2016). Meaning and posttraumatic growth among survivors of the September 2013 Colorado floods. *Journal of Clinical Psychology*, *72*(12), 1247–1263. doi:10.1002/jclp.22344

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. doi:10.1111/j.0006-341X.2000.00455.x

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. doi:10.1136/bmj.316.7129.469

Esterling, B. A., Antoni, M. H., Kumar, M., & Schneiderman, N. (1990). Emotional repression, stress disclosure responses, and Epstein-Barr viral capsid antigen titers. *Psychosomatic Medicine*, *52*, 397–410. doi:10.1097/00006842-199007000-00002

Fawzy, N. W., Fawzy, N. W., Hyun, C. S., Elashoff, R., Guthrie, D., Fahey, J. L., & Morton, D. L. (1993). Malignant melanoma. Effects of an early structured psychiatric intervention, coping, and affective state on recurrence and survival 6 years later. *Archives of General Psychiatry*, *50*(9), 681–689. doi:10.1001/archpsyc.1993.01820210015002

Francis, G. (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin & Review*, *19*(6), 975–991. doi:10.3758/s13423-012-0322-y

Francis, G. (2014). The frequency of excess success for articles in Psychological Science.

Psychonomic Bulletin & Review, *21*(5), 1180–1187. doi:10.3758/s13423-014-0601-x

Frankl, V. (1959). *Man's search for meaning* (3rd ed.). Boston, MA: Beacon Press.

Frattaroli, J. (2006). Experimental disclosure and its moderators: A meta-analysis. *Psychological Bulletin*, *132*(6), 823–865. doi:10.1037/0033-2909.132.6.823

Frisina, P. G., Borod, J. C., & Lepore, S. J. (2004). A meta-analysis of the effects of written emotional disclosure on the health outcomes of clinical populations. *The Journal of Nervous and Mental Disease*, *192*(9), 629–634. doi:10.1097/01.nmd.0000138317.30764.63

García-Oliva, C., & Piqueras, J. A. (2016). Experiential avoidance and technological addictions in adolescents. *Journal of Behavioral Addictions*, *5*(2), 293–303. doi:10.1556/2006.5.2016.041

Gentes, E. L., Dennis, P. A., Kimbrel, N. A., Rissling, M. B., Beckham, J. C., & Calhoun, P. S. (2014). DSM-5 posttraumatic stress disorder: Factor structure and rates of diagnosis. *Journal of Psychiatric Research*, *59*(1), 60–67. doi:10.1016/j.jpsychires.2014.08.014

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3–8. doi:10.3102/0013189X005010003

Goldstein, H. S., Edelberg, R., Meier, C. F., & Davis, L. (1988). Relationship of resting blood pressure and heart rate to experienced anger and expressed anger. *Psychosomatic Medicine*, *50*(4), 321–329. doi:10.1097/00006842-198807000-00001

Gortner, E. M., Rude, S. S., & Pennebaker, J. W. (2006). Benefits of expressive writing in lowering rumination and depressive symptoms. *Behavior Therapy*, *37*(3), 292–303. doi:10.1016/j.beth.2006.01.004

Gross, J. J., & Levenson, R. W. (1997). Hiding feelings: The acute effects of inhibiting negative and positive emotion. *Journal of Abnormal Psychology*, *106*(1), 95–103. doi:10.1037/0021-843X.106.1.95

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments.

Psychological Bulletin, *92*(2), 490–499. doi:10.1037/0033-2909.92.2.490

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557–560. doi:10.1136/bmj.327.7414.557

Hilgard, J. (2016). PETPEESE. GitHub. Retrieved from https://github.com/Joe-Hilgard/PETPEESE

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, *11*(2), 193–206. doi:10.1037/1082-989X.11.2.193

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. doi:10.1177/0956797611430953

Kelley, K. (2007). Confidence intervals for standardized effect sizes. *Journal of Statistical Software*, *20*(8), 1–24. doi:10.18637/jss.v020.i08

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., … Nosek, B. A. (2014). Investigating variation in replicability. *Social Psychology*, *45*(3), 142–152. doi:10.1027/1864-9335/a000178

Klump, M. C. (2008). Posttraumatic stress disorder and sexual assault in women. *Journal of College Student Development*, *8225*(May 2014), 37–41. doi:10.1300/J035v21n02

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00863

Larson, D. G., & Chastain, R. L. (1990). Self-concealment: Conceptualization, measurement, and health implications. *Journal of Social and Clinical Psychology*, *9*(4), 439–455. doi:10.1521/jscp.1990.9.4.439

Levitt, J. T., Brown, T. A., Orsillo, S. M., & Barlow, D. H. (2004). The effects of acceptance versus suppression of emotion on subjective and psychophysiological response to carbon dioxide challenge in patients with panic disorder. *Behavior Therapy*, *35*(4), 747–766. doi:10.1016/S0005-7894(04)80018-2

Lieberman, M. A., & Goldstein, B. A. (2006). Not all negative emotions are equal: The role of emotional expression in online support groups for women with breast cancer. *Psycho-Oncology*, *15*(2), 160–168. doi:10.1002/pon.932

Marx, B. P., & Sloan, D. M. (2005). Peritraumatic dissociation and experiential avoidance as predictors of posttraumatic stress symptomatology. *Behaviour Research and Therapy*, *43*(5), 569–583.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist*, *70*(6), 487–498. doi:10.1037/a0039400

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis. *Perspectives on Psychological Science*, *11*(5), 730–749. doi:10.1177/1745691616662243

Mogk, C., Otte, S., Reinhold-Hurley, B., & Kröner-Herwig, B. (2006). Health effects of expressive writing on stressful or traumatic experiences - a meta-analysis. *Psychosocial Medicine*, *3*, Doc06.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105–125. doi:10.1037/1082-989X.7.1.105

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., … Chartier, C. R. (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, 251524591879760. doi:10.1177/2515245918797607

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, *69*(1), 511–534. doi:10.1146/annurev-psych-122216-011836

Niles, A. N., Haltom, K. E., Mulvenna, C. M., Lieberman, M. D., & Stanton, A. L. (2014).

Randomized controlled trial of expressive writing for psychological and physical health: The moderating role of emotional expressivity. *Anxiety, Stress and Coping*, *27*(1), 1–17. doi:10.1080/10615806.2013.802308

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. doi:10.1126/science.aac4716

Pennebaker, J. W. (1989). Confession, inhibition, and disease. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 22, pp. 211–244). Academic Press. doi:10.1016/S0065-2601(08)60309-3

Pennebaker, J. W., & Beall, S. K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology*, *95*(3), 274–281. doi:10.1037//0021-843X.95.3.274

Pennebaker, J. W., & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science*, *10*(3), 90–93. doi:10.1111/1467-8721.00123

Possemato, K., Ouimette, P., & Geller, P. (2010). Internet-based expressive writing for kidney transplant recipients: Effects on posttraumatic stress and quality of life. *Traumatology*, *16*(1), 49–54. doi:10.1177/1534765609347545

Rachman, S. (1980). Emotional processing. *Behaviour Research and Therapy*, *18*(1), 51–60. doi:10.1016/0005-7967(80)90069-8

Reinhold, M., Bürkner, P. C., & Holling, H. (2018). Effects of expressive writing on depressive symptoms—A meta-analysis. *Clinical Psychology: Science and Practice*, *25*(1). doi:10.1111/cpsp.12224

Sánchez-Meca, J., & Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods*, *13*(1), 31–48. doi:10.1037/1082-989X.13.1.31

Scheff, T. J. (1979). *Catharsis in healing, ritual, and drama*. Los Angeles: University of

California Press.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547. doi:10.1037/a0033242

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better p-curves: Making p-curve analysis more robust to errors, fraud, and ambitious p-hacking, a reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General*, *144*(6), 1146–1152. doi:10.1037/xge0000104

Slavin-Spenny, O. M., Cohen, J. L., Oberleitner, L. M., & Lumley, M. A. (2011). The effects of different methods of emotional disclosure: Differentiating posttraumatic growth from stress symptoms. *Journal of Clinical Psychology*, *67*(10), 993–1007. doi:10.1002/jclp.20750

Sloan, D. M., Marx, B. P., & Greenberg, E. M. (2011). A test of written emotional disclosure as an intervention for posttraumatic stress disorder. *Behaviour Research and Therapy*, *49*(4), 299–304. doi:10.1016/j.brat.2011.02.001

Sloan, D. M., Marx, B. P., Bovin, M. J., Feinstein, B. A., & Gallagher, M. W. (2012). Written exposure as an intervention for PTSD: A randomized clinical trial with motor vehicle accident survivors. *Behaviour Research and Therapy*, *50*(10), 627–635. doi:10.1016/j.brat.2012.07.001

Sloan, D. M., Marx, B. P., Epstein, E. M., & Lexington, J. M. (2007). Does altering the writing instructions influence outcome associated with written disclosure? *Behavior Therapy*, *38*(2), 155–168. doi:10.1016/j.beth.2006.06.005

Sloan, D. M., Marx, B. P., Lee, D. J., & Resick, P. A. (2018). A brief exposure-based treatment vs cognitive processing therapy for Posttraumatic Stress Disorder. *JAMA Psychiatry*. doi:10.1001/jamapsychiatry.2017.4249

Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurement*, *61*(4), 605–632.

doi:10.1177/00131640121971392

Smithson, M. (2003). *Confidence intervals*. Thousand Oaks, CA: Sage.

Smyth, J. M. (1998). Written emotional expression: Effect sizes , outcome types, and moderating variables. *Journal of Consulting and Clinical Psychology*, *66*(1), 174–184. doi:10.1037/0022-006X.66.1.174

Smyth, J. M., Stone, A. A., Hurewitz, A., & Kaell, A. (1999). Effects of writing about stressful experiences on symptom reduction in patients with asthma or rheumatoid arthritis: A randomized trial. *JAMA: The Journal of the American Medical Association*, *281*(14), 1304–1309. doi:10.1001/jama.281.14.1304

Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys*, *19*(3), 309–345. doi:10.1111/j.0950-0804.2005.00250.x

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. doi:10.1002/jrsm.1095

Stanton, A. L., Danoff-Burg, S., Sworowski, L. A., Collins, C. A., Branstetter, A. D., Rodriguez-Hanley, A., . . . Austenfeld, J. L. (2002). Randomized, controlled trial of written emotional expression and benefit finding in breast cancer patients. *Journal of Clinical Oncology*, *20*(20), 4160–4168. doi:10.1200/JCO.2002.08.521

Taku, K., Calhoun, L. G., Cann, A., & Tedeschi, R. G. (2008). The role of rumination in the coexistence of distress and posttraumatic growth among bereaved Japanese University students. *Death Studies*, *32*(5), 428–444. doi:10.1080/07481180801974745

Tedeschi, R. G., & Calhoun, L. G. (1995). *Trauma & transformation: Growing in the aftermath of suffering.* Thousand Oaks, CA: Sage Publications.

Tedeschi, R. G., & Calhoun, L. G. (2004). Posttraumatic growth: Conceptual foundations and empirical evidence. *Psychological Inquiry*, *15*(1), 1–18. doi:10.1207/s15327965pli1501

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, *22*(13), 2113–2126. doi:10.1002/sim.1461

Theofilou, P. (2013). Quality of life: Definition and measurement. *Europe's Journal of Psychology*, *9*(1), 150–162. doi:10.5964/ejop.v9i1.337

Van Aert, R. C. M. (2017). P-uniform. GitHub. Retrieved from https://github.com/RobbievanAert/puniform

Van Aert, R. C. M., Wicherts, J. M., & Van Assen, M. A. L. M. (2016). Conducting meta-analyses based on p-values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science*, *11*(5), 713–729. doi:10.1017/CBO9781107415324.004

Van Assen, M. A. L. M., Van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, *20*(3), 293–309. doi:http://dx.doi.org/10.1037/met0000025

Van Elk, M., Matzke, D., Gronau, Q. F., Guan, M., Vandekerckhove, J., & Wagenmakers, E.-J. (2015). Meta-analyses are no substitute for registered replications: A skeptical perspective on religious priming. *Frontiers in Psychology*, *6*, 1365. doi:10.3389/fpsyg.2015.01365

Van Emmerik, A. A. P., Reijntjes, A., & Kamphuis, J. H. (2013). Writing therapy for posttraumatic stress: A meta-analysis. *Psychotherapy and Psychosomatics*, *82*(2), 82–88. doi:10.1159/000343131

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419–435. doi:10.1007/BF02294384

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*(4), 428–443. doi:10.1037/1082-989X.10.4.428

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal* *of Statistical Software*, *36*(3), 1–48. doi:10.18637/jss.v036.i03

Wang, X., Gao, L., Shinfuku, N., Zhang, H., Zhao, C., & Shen, Y. (2000). Longitudinal study of earthquake-related PTSD in a randomly selected community sample in North China. *American Journal of Psychiatry*, *157*(8), 1260–1266. doi:10.1176/appi.ajp.157.8.1260

Weiss, T. (2002). Posttraumatic growth in women with breast cancer and their husbands – An intersubjective validation study. *Journal of Psychosocial Oncology*, *20*(2), 65–80. doi:10.1300/J077v20n02_04

Yilmaz, M., & Zara, A. (2016). Traumatic loss and posttraumatic growth: The effect of traumatic loss related factors on posttraumatic growth. *Anatolian Journal of Psychiatry*, *17*(1), 5–11. doi:10.5455/apd.188311

Table 1

Effect Size Estimates for PTS Results

Model	Fixed Effects	Random Effects	Fixed No Outliers	Random No Outliers
Overall Effects	0.36 [0.34, 0.39]	0.42 [0.35, 0.49]	0.36 [0.33, 0.38]	0.40 [0.33, 0.46]
Z Values	24.64, $p < .001$	12.35, $p < .001$	23.97, $p < .001$	12.38, $p < .001$
p-Uniform	0.63 [0.54, 0.72]	-	0.61 [0.52, 0.70]	-
PET	0.09 [0.01, 0.18]	-	0.14 [0.06, 0.22]	-
PEESE	0.24 [0.20, 0.29]	-	0.26 [0.22, 0.31]	-
Selection Models	0.33 [0.28, 0.37]	0.45 [0.33, 0.57]	0.29 [0.24, 0.33]	0.39 [0.27, 0.50]
Trim and Fill	0.28 [0.25, 0.31]	0.28 [0.21, 0.36]	0.28 [0.25, 0.31]	0.28 [0.21, 0.35]

Note. [] indicates the 95 percent confidence interval for each effect size estimate.

Table 2

Effect Size Estimates for PTG Results

Model	Fixed Effects	Random Effects
Overall Effects	0.10 [0.02, 0.17]	0.10 [0.02, 0.17]
Z Values	2.45, p = .014	2.45, p = .014
p-Uniform	-0.11 [-1.43, 0.42]	-
PET	0.06 [-0.20, 0.32]	-
PEESE	0.08 [-0.04, 0.20]	-
Selection Models	0.09 [-0.01, 0.18]	0.09 [-0.03, 0.20]
Trim and Fill	0.10 [0.02, 0.17]	0.10 [0.02, 0.17]

Note. [] indicates the 95 percent confidence interval for each effect size estimate.

Table 3

Effect Size Estimates for QOL Results

Model	Fixed Effects	Random Effects	Fixed No Outliers	Random No Outliers
Overall Effects	-0.01 [-0.07, 0.05]	-0.01 [-0.16, 0.13]	-0.01 [-0.07, 0.05]	-0.01 [-0.11, 0.09]
Z Values	-0.33, p = .745	-0.18, p = .860	-0.25, p = .805	-0.20, p = .838
p-Uniform	0.79 [0.33, 1.61]	-	0.62 [0.10, 0.96]	-
PET	0.05 [-0.26, 0.36]	-	0.05 [-0.29, 0.38]	-
PEESE	0.00 [-0.17, 0.17]	-	0.00 [-0.19, 0.19]	-
Selection Models	-0.06 [-0.12, 0.01]	0.51 [-0.09, 1.12]	-0.04 [-0.11, 0.03]	0.05 [-0.15, 0.24]
Trim and Fill	-0.01 [-0.07, 0.05]	-0.01 [-0.16, 0.13]	-0.01 [-0.07, 0.05]	-0.01 [-0.11, 0.09]

Note. [] indicates the 95 percent confidence interval for each effect size estimate.

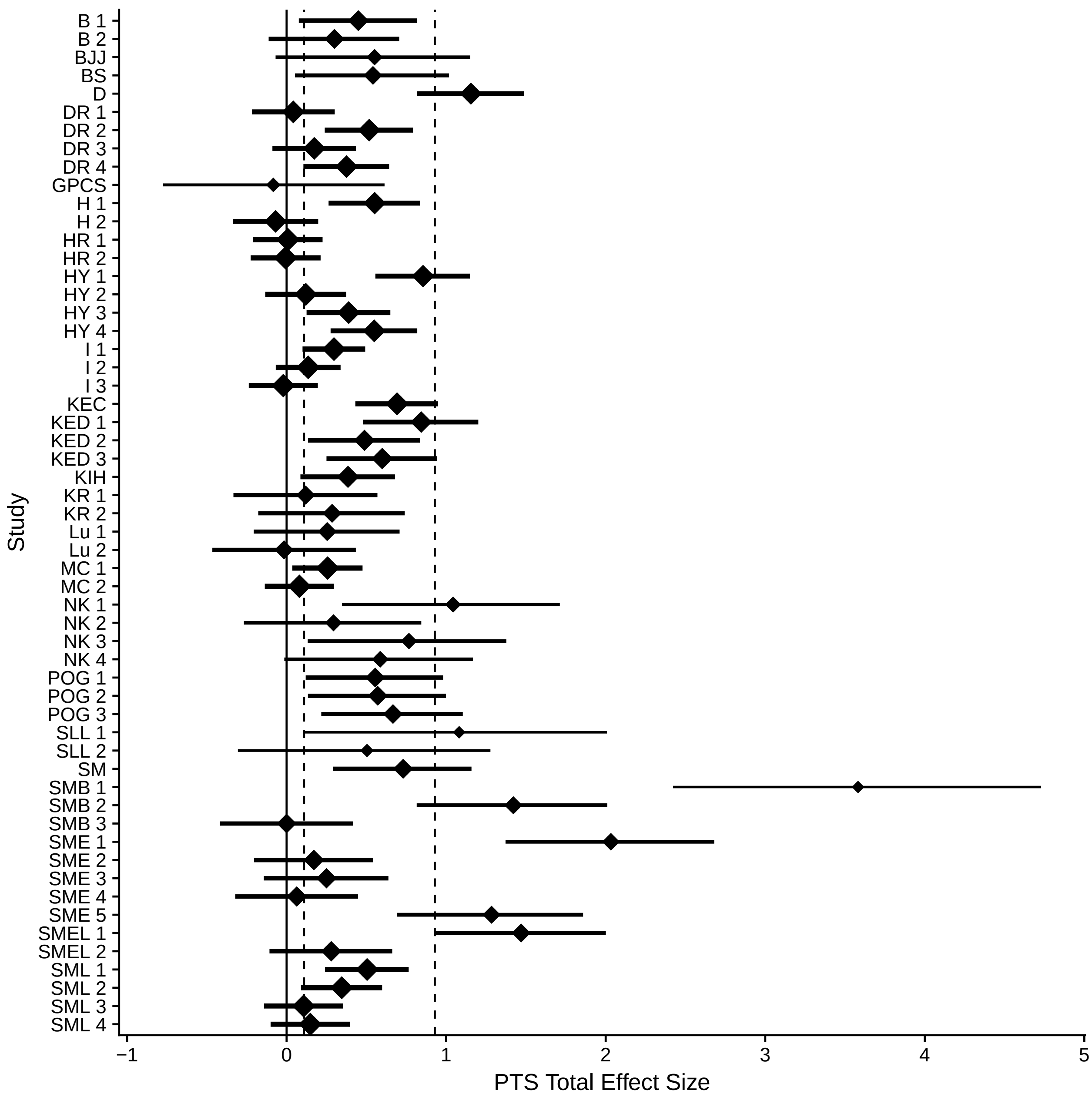

Figure 1. Effect sizes and their non-centralized confidence interval for PTS total scores. Dashed lines indicated average non-weighted lower and upper confidence interval limits. Diamond size indicates normalized study weight from a random effects model. Y-axis labels indicate citation and pairwise time combination, these labels can be matched to the exact data by viewing the provided data online. Table 1 includes meta-analytic effect size for PTS overall.

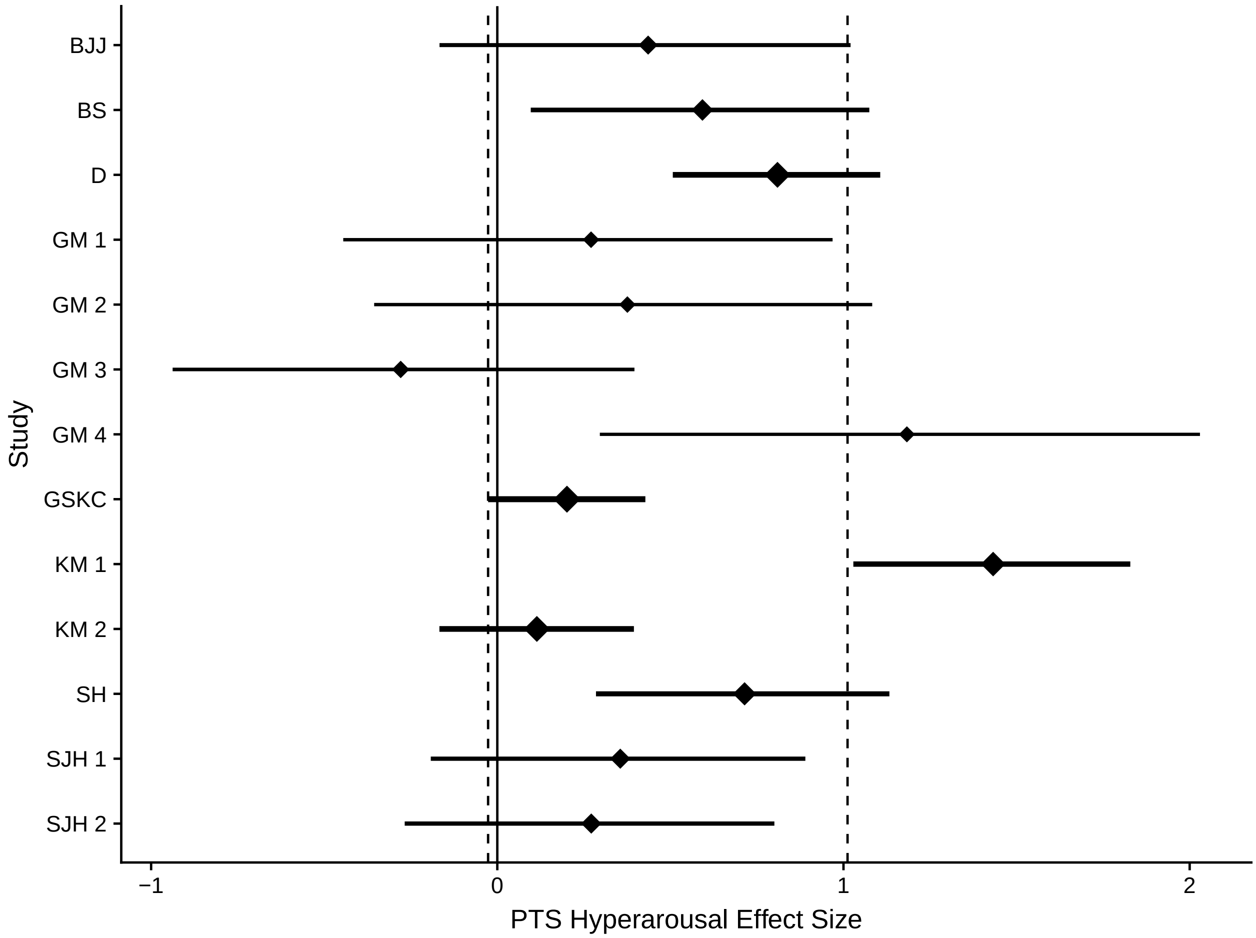

Figure 2. Effect sizes and their non-centralized confidence interval for PTS Hyperarousal. Dashed lines indicated average non-weighted lower and upper confidence interval limits. Diamond size indicates normalized study weight from a random effects model. Y-axis labels indicate citation and pairwise time combination, these labels can be matched to the exact data by viewing the provided data online. Table 1 includes meta-analytic effect size for PTS overall.

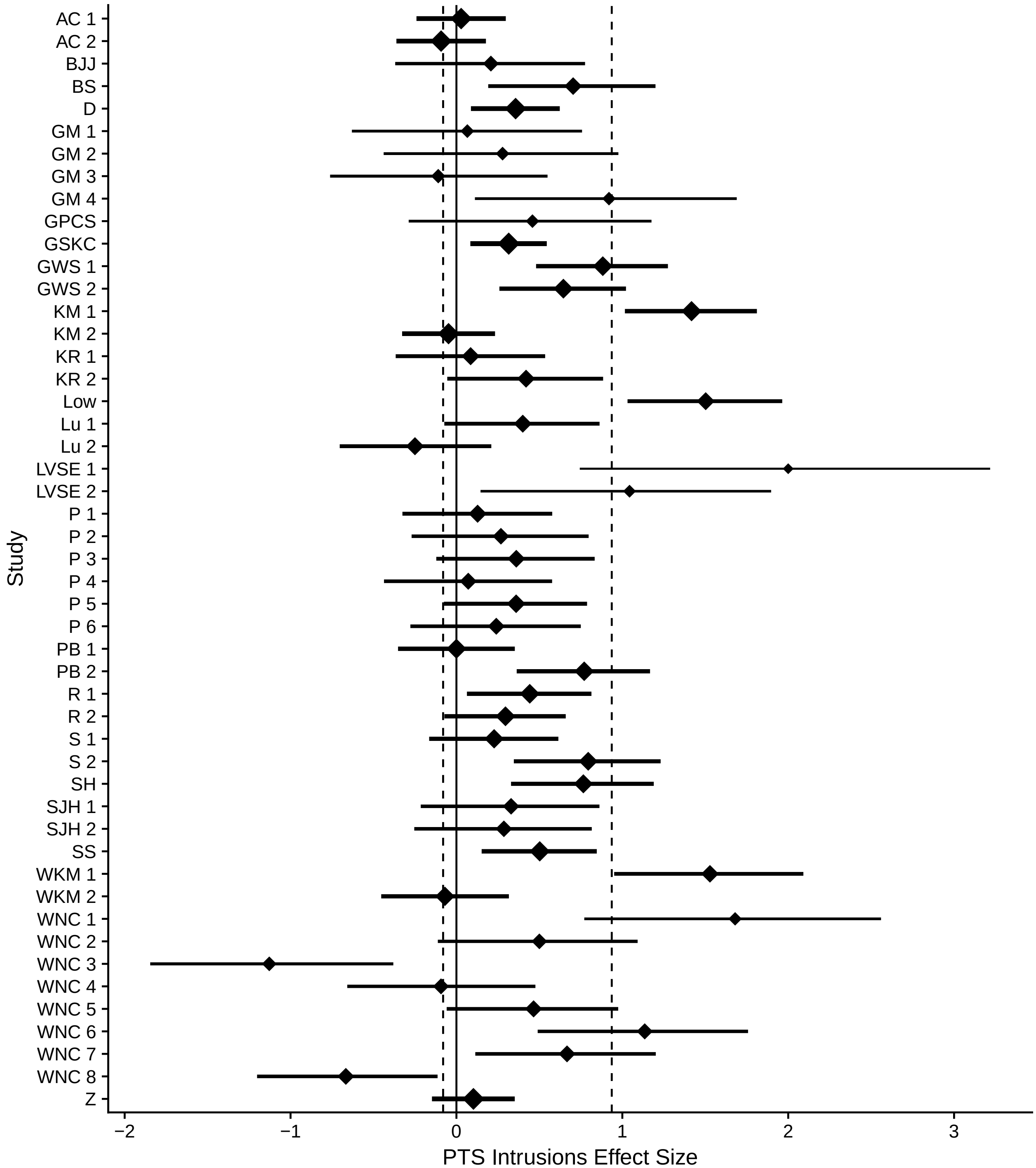

Figure 3. Effect sizes and their non-centralized confidence interval for PTS Intrusion scores. Dashed lines indicated average non-weighted lower and upper confidence interval limits. Diamond size indicates normalized study weight from a random effects model. Y-axis labels indicate citation and pairwise time combination, these labels can be matched to the exact data by viewing the provided data online. Table 1 includes meta-analytic effect size for PTS overall.

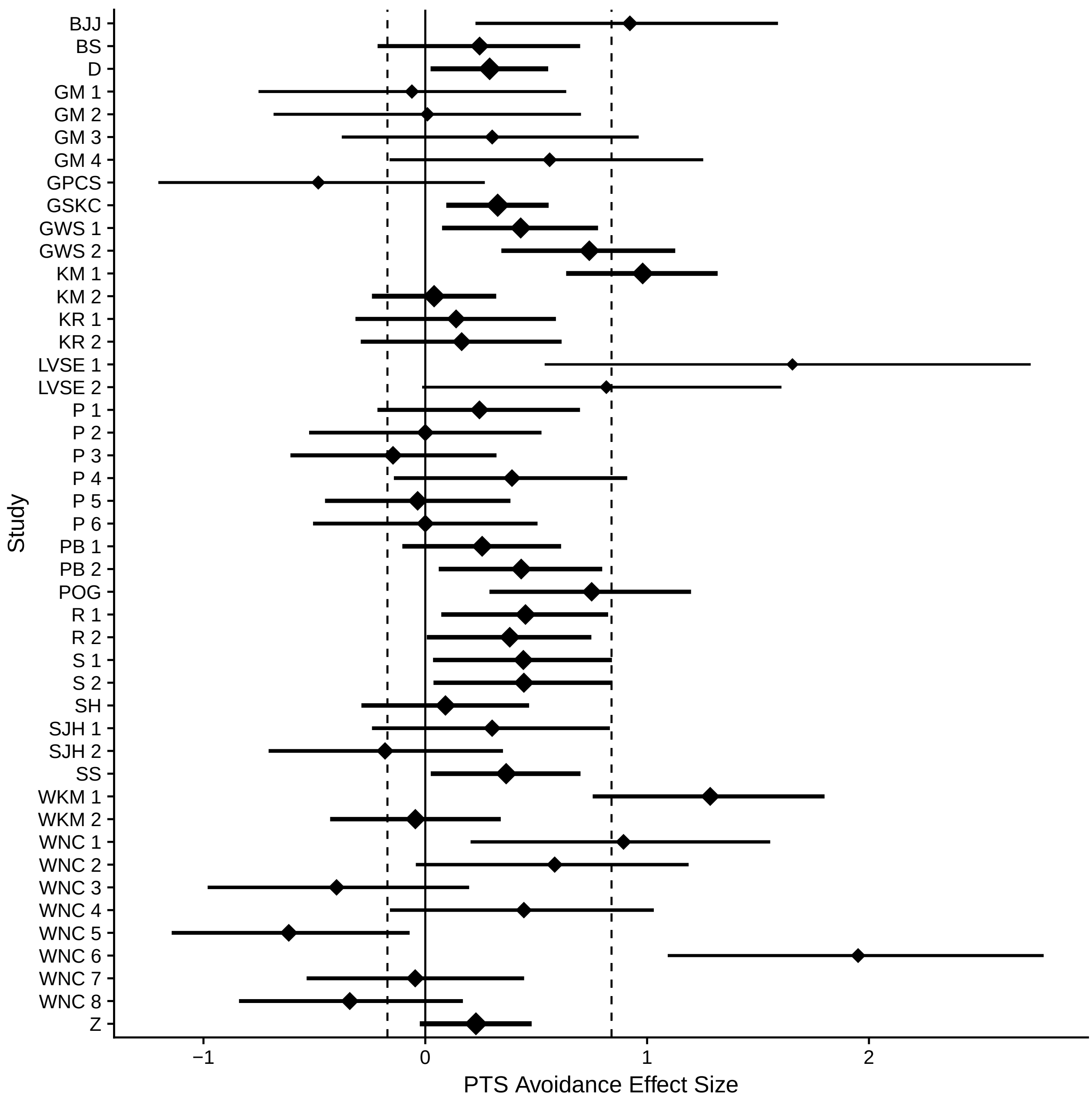

Figure 4. Effect sizes and their non-centralized confidence interval for PTS Avoidance Scores. Dashed lines indicated average non-weighted lower and upper confidence interval limits. Diamond size indicates normalized study weight from a random effects model. Y-axis labels indicate citation and pairwise time combination, these labels can be matched to the exact data by viewing the provided data online. Table 1 includes meta-analytic effect size for PTS overall.

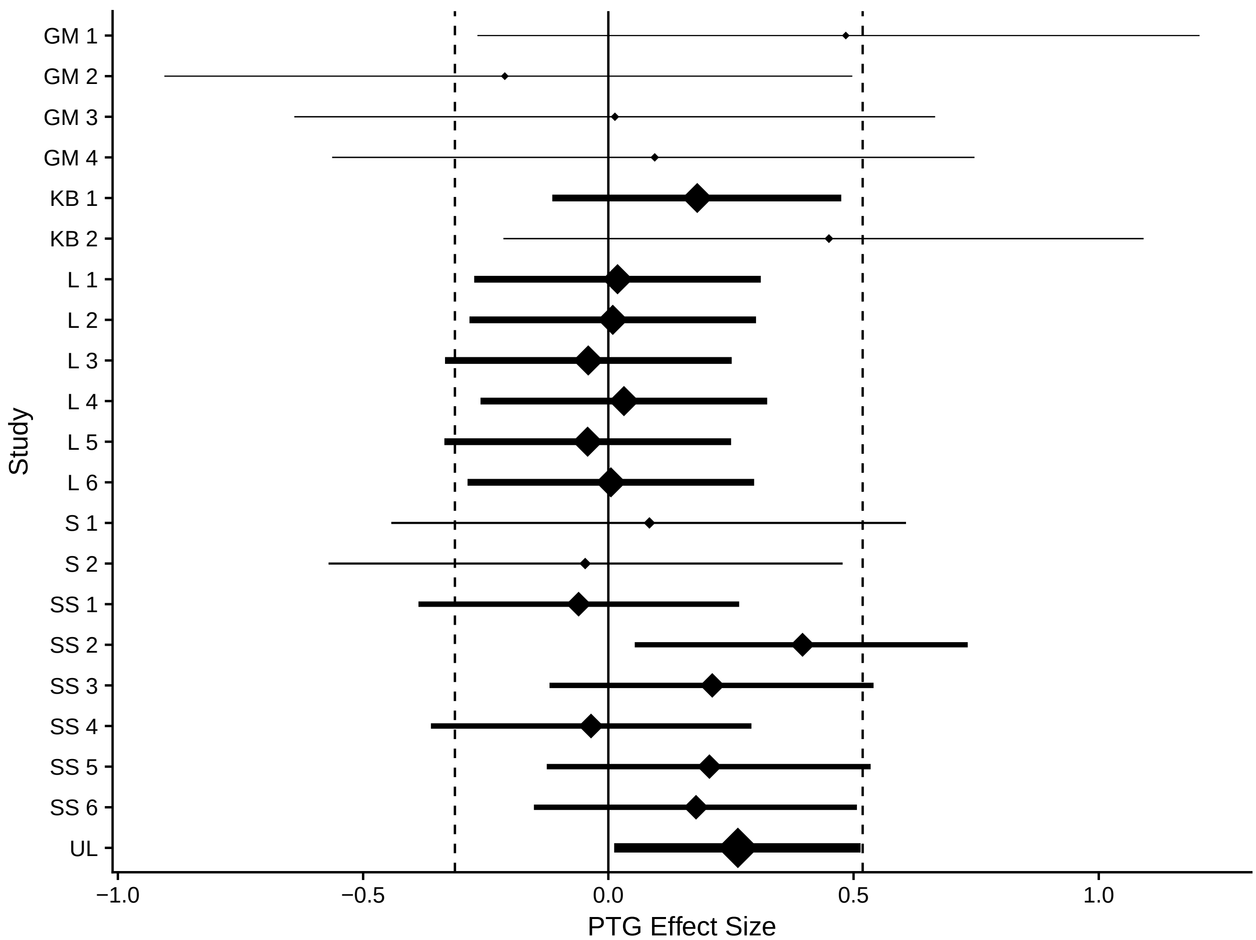

Figure 5. Effect sizes and their non-centralized confidence interval for PTG outcome variables. Dashed lines indicated average non-weighted lower and upper confidence interval limits. Diamond size indicates normalized study weight from a random effects model. Y-axis labels indicate citation and pairwise time combination, these labels can be matched to the exact data by viewing the provided data online. Table 2 includes meta-analytic effect size for PTG.

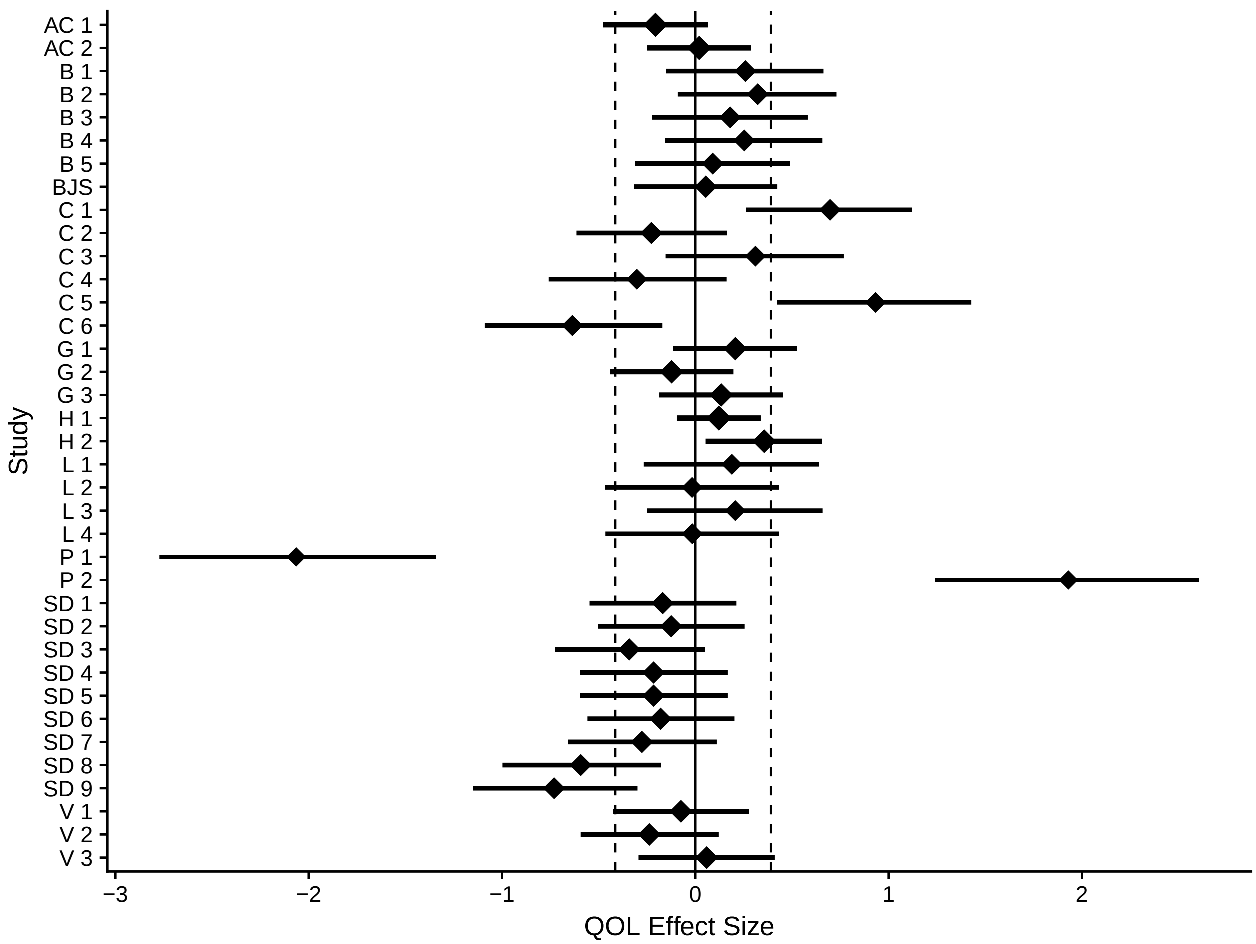

Figure 6. Effect sizes and their non-centralized confidence interval for QOL outcome variables. Dashed lines indicated average non-weighted lower and upper confidence interval limits. Diamond size indicates normalized study weight from a random effects model. Y-axis labels indicate citation and pairwise time combination, these labels can be matched to the exact data by viewing the provided data online. Table 3 includes meta-analytic effect size for QOL.