The Inflating Impact of Waiting-List Controls on Effect Size Estimates

Keith R. Laws [1]*, Luca Pellegrini [1,2,3], Jemma E. Reid [1], Lynne M. Drummond [1,4] and Naomi A. Fineberg [1,4,5,6]

[1] School of Life and Medical Sciences, University of Hertfordshire, Hatfield, United Kingdom, [2] Hertfordshire Partnership University NHS Foundation Trust, Welwyn Garden City, United Kingdom, [3] Centre for Psychedelic Research, Imperial College London, London, United Kingdom, [4] South West London and St George's NHS Trust, London, United Kingdom, [5] Hertfordshire Partnership University NHS Foundation Trust, Hatfield, United Kingdom, [6] University of Cambridge School of Clinical Medicine, Cambridge, United Kingdom

INTRODUCTION

The primary psychological intervention recommended for obsessive compulsive disorder (OCD) is cognitive behavioral therapy (CBT) with exposure and response prevention (ERP), in which individuals are taught to confront and tolerate conditions that provoke obsessions and compulsions and resist acting on them. A recently published systematic review and meta-analysis of 24 studies by Ferrando and Selai (1) reports a large mean effect size in favor of ERP for reducing OCD symptomatology [−0.75, (95% CI = −0.92 to −0.59)], with the authors endorsing ERP as "the treatment of choice for OCD." (p. 10). We highlight below a series of methodological issues that contextualize and temper such strong advocacy for ERP.

The overall effect size of ERP for OCD reported by Ferrando and Selai (1) is similar to that reported earlier by us (2) (g = 0.74: 95% CI = 0.51 to 0.97, k = 36). Any similarity, however, may be misleading since Ferrando and Selai determined effect sizes using between-group pre-post change-scores and a fixed-effects model, while we analyzed end-of-trial between-group effect sizes using a random effects approach and on a 50% larger sample of randomized controlled trials (RCT). Another key difference between the two meta-analyses concerns the importance of addressing the impact of factors known to moderate the effect sizes of psychological interventions in this population. In particular, we argue that the conclusions of Ferrando and Selai are overly optimistic because they did not examine the moderating effect of type of control group, and its associated impact on the heterogeneity in their results –and that had they done so, this would have substantially altered the reported efficacy of ERP for OCD.

The choice of control group (e.g., a waiting list, another form of psychological treatment, or treatment as usual) has repeatedly been shown to moderate the effect size of psychological treatments for OCD [e.g., (2–4)]. Although Ferrando and Selai state that ERP was superior to various controls, they did not compare ERP outcomes across different control groups. This point is important to clarify, because their abstract states that "Our review suggests that ERP was superior to the other groups, including *both neutral and active treatments*, in reducing OCD symptomatology." Similarly, they open their discussion with "Our main aim was to determine whether ERP-based therapy was more effective in reducing OCD symptoms *compared to no treatment or other psychotherapeutic or pharmaceutical interventions.*" Later in the discussion they similarly say "The efficacy of ERP therapy in reducing OCD symptoms was compared to a control group, *either waiting-list or placebo, or another therapy,* which included anxiety management, cognitive therapy, IBA, autogenic training, relaxation therapy, fluoxetine, sertraline and eye-movement desensitization and reprocessing" (our italics). These descriptions could be interpreted to mean that ERP was compared against each of these different controls, but their analysis only compares ERP with *all* control conditions collectively.

Using the data presented by Ferrando and Selai, we calculated effect sizes for ERP vs. waiting-list controls [−1.82 (−2.11 to −1.52; k = 8], which was significantly larger than the ERP comparisons vs. all remaining control conditions [−0.32 (−0.52 to −0.13; k = 15] (Q = 11.67, $p < 0.001$). Indeed, we found that whether the trials used waiting-list or other controls accounted for one third of the variance in the effect sizes [R^2 analog = 0.33; and removal of one extreme outlier (5) increased the R^2 analog = 0.41]. Thus, by pooling data across all control groups Ferrando and Selai do not address this important factor, with effect sizes for ERP being almost 6 times larger when compared to waiting-list rather than to other control conditions. Further, this re-analysis of their data accords with our own meta-analysis (2), which showed similar large effect sizes for ERP vs. waiting-list (g = 1.27: 95% CI 0.79–1.75, k = 8), but not when we compared to active psychological therapy controls (g= −0.05: 95% CI −0.27 to 0.16, k = 8).

Since wait-list comparisons comprised a substantial proportion (one-third) of all ERP trials included by Ferrando and Selai, it is not surprising that they report extremely high levels of heterogeneity. From their reported Q = 185.66, we estimate an I^2 of 87%, which would represent a "severe" level of heterogeneity. Ferrando and Selai state that "Our test for heterogeneity was significant (Q-statistic = 185.66, $p < 0.01$), indicating that one study could be driving the results as there is some unexplained heterogeneity between studies. Cochrane guidelines, used throughout this piece, report that when using a fixed-effects model the heterogeneity result can be ignored (6)." However, Cochrane rightly state that fixed-effect meta-analyses ignore heterogeneity, not that heterogeneity can be ignored. Heterogeneity is extremely high and this is neither a reason for using a fixed-effects analysis nor for assuming that a fixed-effects analysis negates concerns about heterogeneity. We would question the selection of a fixed-effects rather than a random-effects model. It seems implausible that the included samples are derived from the same general population and estimating the same underlying fixed treatment effect. This is underscored by the extreme heterogeneity for individual study effect sizes reported by Ferrando and Selai, which ranges massively from + 1.03− −16.01.

Ferrando and Selai did conduct a leave-1-out analysis, where the effect size is recalculated after removing each individual study; and they report it always remained significant. In this context, they identified three trials (7–9) as outliers and ran the meta-analysis removing these three outliers. They report that their "… original findings were unaffected by the removal of these studies {effect size = −0.97, significant at $p < 0.01$ [95% CI = (−1.18 to −0.76); SE = 0.11, $z = −8.94$]}. This, combined with the leave 1 out test, demonstrates that our original results can be considered valid." Although the leave-1-out test remains significant following each removal, it is important to examine the range of effect sizes that emerge. Our re-assessment of this analysis suggests that effect sizes vary considerably–between −0.64 (leaving out Gomes et al.) to −0.96 (leaving out Belloto-Silva et al.). In other words, these individual studies have a large

impact on the reported effect size. Removal of these three studies increased the overall effect size by ∼30% from −0.75 to −0.97 (−1.18 to −0.76). We note that the new lower 95% CI of −0.76 exceeds the mean across the original sample (−0.75). Finally, we note that these 3 studies undoubtedly influence the overall results having the largest samples and their weighting being greatest in the analysis. In fact, these 3 trials comprise over 30% of the participants across all 24 RCTs analyzed by Ferrando and Selai. In this context, it is worth noting that larger trials would be expected to produce the most accurate effect size estimates–and these three larger trials alone produce a much smaller pooled effect size (−0.41, 95% CI −0.68 to −0.15), which is less than half of the −0.97 estimated by Ferrando and Selai (after their removal).

The authors do not present funnel plots or check for possible publication bias; however, our re-analysis of their data (using their own fixed effects model) indicates some funnel plot asymmetry; and a trim and fill analysis points to 4 potentially missing studies that would reduce the effect size to −0.65 (95% CI −0.81 to −0.49). This possible publication bias is underscored by Begg and Mazumdar rank correlation, which is significant (z for tau 1.77, $p = 0.03$); and Egger's intercept −3.52 ($p < 0.01$) also pointing to possible presence of publication bias. Inspection of their forest plot also identifies evidence of extreme outliers. For example, Khodarahimi (5) has an effect size of −16.01 (95%CI −20.98 to −11.03); however, removal of this study does not reduce the I^2 value (which is still extremely high at I2 = 86.67%). Clearly, multiple studies have produced unusual effect sizes. We also note that Simpson et al. appears to be missing from the overall effect size calculation (see forest plot).

To summarize, we argue that the effect size reported by Ferrando and Selai exaggerates the efficacy of ERP at reducing OCD symptoms. As with our own meta-analysis (2), our re-analysis of the data here shows that waiting list control groups inflate effect sizes [see (10, 11)]. As noted by Leichsenring and Steinert (12) "When examining efficacy, a treatment may be compared with different comparators, that is, with an established treatment, treatment as usual, a placebo, or a waiting list, with *decreasing strictness of the empirical test.*" (p. 1,323, our italics). Several factors might lead to the exaggerated difference for waiting-list controls. It may be that only some patients are willing to be randomized to waiting-list control conditions e.g., if they have particularly high expectations for the therapy (13)–such factors might both inflate the intervention and deflate the control. Thus, we question the conclusion that ERP should be considered the treatment of choice for OCD. Only comparisons with *no treatment* indicate large effects, while those with other types of control groups identify much more modest evidence for ERP efficacy.

AUTHOR CONTRIBUTIONS

KL and NF conceived of the initial study and created an initial draft. LP, LD, and JR commented on and contributed to the initial draft. All authors approve the current version of the paper.

REFERENCES

1. Ferrando C, Selai C. A systematic review and meta-analysis on the effectiveness of exposure and response prevention therapy in the treatment of obsessive-compulsive disorder. *J Obsessive Compuls Relat Disord.* (2021) 31:100684. doi: 10.1016/j.jocrd.2021.100684

2. Reid JE, Laws KR, Drummond L, Vismara M, Grancini B, Mpavaenda D, et al. Cognitive behavioural therapy with exposure and response prevention in the treatment of obsessive-compulsive disorder: a systematic review and meta-analysis of randomised controlled trials. *Compr Psychiatry.* (2021) 152223. doi: 10.1016/j.comppsych.2021.152223

3. Olatunji BO, Davis ML, Powers MB, Smits JAJ. Cognitive-behavioral therapy for obsessive-compulsive disorder: a meta-analysis of treatment outcome and moderators. *J Psychiatr Res.* (2013) 47:33–41. doi: 10.1016/j.jpsychires.2012.08.020

4. Pellegrini L, Laws KR, Albert U, Reid J, Fineberg NA. Letter to the editor on Cognitive behavior therapy for health anxiety: systematic review and meta-analysis of clinical efficacy and health economic outcomes. *Expert Rev Pharmacoecon Outcomes Res.* (2020) 20:683–4. doi: 10.1080/14737167.2020.1829480

5. Khodarahimi S. Satiation therapy and exposure response prevention in the treatment of obsessive compulsive disorder. *J Contemp Psychother.* (2009) 39:203–7. doi: 10.1007/s10879-009-9110-z

6. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions.* John Wiley and Sons (2011).

7. Belotto-Silva C, Diniz JB, Malavazzi DM, Val'erio C, Fossaluza V, Borcato S, et al. Group cognitive-behavioral therapy versus selective serotonin reuptake inhibitors for obsessive-compulsive disorder: a practical clinical trial. *J Anxiety Disord.* (2012) 26, 25–31. doi: 10.1016/j.janxdis.2011.08.008

8. Visser HA, van Megen H, van Oppen P, Eikelenboom M, Hoogendorn AW, Kaarsemaker M, et al. Inference-based approach versus cognitive behavioral therapy in the treatment of obsessive-compulsive disorder with poor insight: a 24-session randomized controlled trial. *Psychother Psychosom.* (2015) 84:284–93. doi: 10.1159/000382131

9. Gomes JB, Cordioli AV, Bortoncello CF, Braga DT, Gonçalves F, Heldt E. Impact of cognitive-behavioral group therapy for obsessive-compulsive disorder on family accommodation: a randomized clinical trial. *Psychiatry Res.* (2016) 246:70–6. doi: 10.1016/j.psychres.2016.09.019

10. Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatr Scand.* (2014) 130:181–92. doi: 10.1111/acps.12275

11. Cunningham JA, Kypri K, McCambridge J. Exploratory randomized controlled trial evaluating the impact of a waiting list control design. *BMC Med Res Methodol.* (2013) 13:150. doi: 10.1186/1471-2288-13-150

12. Leichsenring F, Steinert C. Is cognitive behavioral therapy the gold standard for psychotherapy?: the need for plurality in treatment and research. *JAMA.* (2017) 318:1323–4. doi: 10.1001/jama.2017.13737

13. Mohr DC, Ho J, Hart TL, Baron KG, Berendsen M, Beckner V, et al. Control condition design and implementation features in controlled trials: a meta-analysis of trials evaluating psychotherapy for depression. *Transl Behav Med.* (2014) 4:407–23. doi: 10.1007/s13142-014-0262-3