Commentary


Sarah J Egan¹,²*, Tracey D Wade³*, Hunna J Watson²,⁴,⁵*, Lars-Göran Öst⁶*, Pim Cuijpers⁷*, Roz Shafran⁸*

Author Affiliations:
1. enAble Institute, Faculty of Health Sciences, Curtin University, Perth, Australia
2. Discipline of Psychology, School of Population Health, Faculty of Health Sciences, Curtin University, Perth, Australia.
4. Department of Psychiatry, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, USA.
5. Division of Paediatrics, School of Medicine, University of Western Australia, Perth, Australia.
6. Department of Psychology, Stockholm University, Sweden.
8. UCL Great Ormond Street Institute of Child Health, University College London, United Kingdom.

* all authors contributed equally to the manuscript. Please address correspondence to: Associate Professor Sarah Egan, enAble Institute and School of Population Health, Curtin University, GPO Box U1987, Perth, WA, 6845, Australia. Email: s.egan@curtin.edu.au.

Acknowledgements

We thank Danyelle Greene for comments on an early draft of the manuscript.
Commentary


Public significance statement

A recent meta-analysis by Smith et al. (2023) in Canadian Psychology made conclusions and generalisations regarding the efficacy of cognitive behaviour therapy for perfectionism (CBT-P). We provide a commentary on the article to outline numerous concerns with the research, including conflating dropout with treatment tolerance, insufficient statistical power, and too few studies for the multiple outcomes assessed. It is concluded that the findings from prior meta-analyses on the efficacy of CBT-P remain valid and that the public can have confidence in such an intervention.
Abstract

Perfectionism is associated with symptoms of multiple psychological disorders. In this commentary, we outline our numerous concerns regarding a recent meta-analysis by Smith et al. (2023) that examined the efficacy of cognitive behaviour therapy for perfectionism (CBT-P). To ensure healthcare and policy decisions are based on high-quality evidence, evidence summaries need to be held to high standards of accountability. The study did not systematically search the literature, and omitted previous studies included in the meta-analyses they sought to re-analyse. Additionally, there was insufficient statistical power to detect intervention effects with small numbers of studies and multiple outcomes, other statistical concerns (e.g., numbers-needed-to-treat analysis), and conflation of the issue of dropout and treatment tolerance. To ensure appropriate guidance for the healthcare sector, evidence summaries of intervention effects must uphold high standards of quality. CBT-P has demonstrated efficacy in addressing the risk factor of perfectionism, and preventing and decreasing symptoms of anxiety, depression, and eating disorders. Further systematic reviews and meta-analyses with rigorous methodology are encouraged.

Keywords: perfectionism; efficacy; meta-analysis; cognitive behaviour therapy; commentary.

Interventions that target perfectionism have been shown to reduce perfectionism and symptoms of anxiety, depression and eating disorders (see Egan, Shafran, & Wade, 2022 and Shafran et al., 2023; for summaries). This work is based on the demonstration of perfectionism as a transdiagnostic risk and maintaining process across a wide range of psychopathology (Egan et al., 2011; Limburg et al., 2017). Cognitive behaviour therapy, specifically tailored to perfectionism (CBT-P), can treat those aspects of perfectionism associated with distress and dysfunction (Egan, Shafran, & Wade, 2022; Shafran et al., 2023). A recent systematic review and meta-analysis of the efficacy of CBT-P for perfectionism and symptoms of anxiety, depression, and eating disorders, synthesised data from 15 prevention and treatment clinical trials, and provided supporting evidence (Galloway et al., 2022). Other systematic reviews and meta-analyses with various scopes and purposes, have provided similarly converging evidence for CBT-P (Lloyd et al., 2015; Robinson & Wade, 2021; Suh et al., 2019).

Within the evidence-based pyramid, systematic reviews and meta-analyses represent the highest level of evidence. It therefore came as a surprise to find an article in Canadian Psychology (Smith et al., 2023), arguing against the most common metric of effect size (ES) for assessing change, stating “extant meta-analyses on CBT for perfectionism focus on standardized mean differences (SMDs), which indicate the difference between the treatment and control conditions in the degree of change in an outcome using standard deviations”. We outline in this paper our numerous concerns with Smith et al.’s (2023) analyses. In the interests of upholding standards for scientific evidence and considering the real-world implications of reviews for translating evidence-based interventions into policy and into the lives of individuals and communities, we detail our concerns as follows.
1. **Omission of studies that were included in the previous meta-analyses which were re-examined.**

Smith et al. (2023) state that their meta-analysis is a replication and extension of three previous meta-analyses (Galloway et al., 2022; Robinson & Wade, 2021; Suh et al., 2019), however, it is problematic that they omitted two previous randomised controlled trials (RCTs) (Pleva & Wade, 2007; Wilksch et al., 2008) that were included in these meta-analyses, despite their footnote that they did not exclude any studies. The three previous systematic reviews Smith et al. (2023) aimed to replicate (Galloway et al., 2022; Robinson & Wade, 2021; Suh et al., 2019) applied different inclusion and exclusion criteria across the reviews, searched for studies with different outcomes, searched in different databases, some but not all included the grey literature, and have different search dates. Therefore, simply equating the studies included in independent systematic reviews into one’s own review and meta-analysis is not a standardised strategy. Further, Smith et al. (2023) did not consider newer data from RCTs testing CBT-P that were published after the previous systematic reviews’ search dates (e.g., Mahmoodi et al., 2021; Oloidi et al., 2022; Osenk et al., 2022; Ward & Wheaton, 2022) or grey literature (Uwadiale, 2022). Additionally, Smith et al. (2023) extended beyond the outcomes reported in the previous meta-analyses, for example, by including life satisfaction and self-esteem. In not conducting an independent search of the literature, they did not consider life satisfaction and self-esteem data from RCTs that were protocol-ineligible from Galloway et al. (2022) and Robinson and Wade’s (2021) reviews (e.g., Zetterberg et al., 2019) and/or appeared after the prior reviews’ search dates (e.g., Uwadiale, 2022). Therefore, a major limitation of Smith et al.’s (2023) review was they did not pre-register their meta-analysis nor conduct a systematic search for the literature reporting additional novel outcomes. This resulted in no systematic identification of the relevant
literature in accordance with a pre-specified protocol, which limits conclusions and generalisability of their results.

2. Dropout and Treatment Tolerance

Smith et al. (2023) conflate dropout with whether treatment is “tolerated”. Lack of measure completion does not equate to poor treatment tolerance. If their interest is in toleration, then treatment completion should have been examined instead. The definition of dropout used i.e., “all randomized participants who did not complete post-treatment measures regardless of whether they started treatment or the reason for dropout” combines ‘decliners’ who verbally accept randomization but do not attend the first session (e.g., because they were randomized to a condition they did not like), ‘dropouts’ who had at least one session but fewer than the number of sessions defined as therapy completion, and ‘no-shows’ who completed treatment but for various reasons failed to come to post-assessment. While study dropout for any reason has been used in some meta-analyses, this reflects dropout, not tolerance, as Smith et al. (2023) incorrectly conclude.

For many mental health conditions (e.g., anxiety, depression, eating disorders), CBT is the gold standard prevention and treatment approach endorsed by expert and clinical guidelines and RCTs have reported dropout rates of 26.2% (Fernandez et al., 2015) as pointed out by Smith et al. (2023). By drawing the conclusion that CBT-P may not be well-tolerated, without any evidence, and at a rate (27%), strikingly similar to CBT in general which the authors describe as “well tolerated” (Fernandez et al., 2015), the public health risk is that policymakers and decision-makers may be affected by this messaging and overlook an efficacious intervention.

A minor further note pertaining to accuracy is that Smith et al. (2023) state “meta-analytic research suggests that during treatment dropout from CBT hinges on diagnosis and is
significantly higher for depressive disorders (36.4%) than anxiety disorders (19.6%).”

However, a reference is not provided for this statement, which is at odds with the largest meta-analysis on dropout (587 RCTs) by Swift and Greenberg (2014). They found a rate of 19.2% for depression, 15.2% for generalised anxiety disorder, 15.4% for panic disorder, 18.0% for social anxiety, 16.3% for obsessive compulsive disorder, and 21.0% for post-traumatic stress disorder (Swift & Greenberg, 2014).

### 3. Statistical Concerns

The Smith et al. (2023) review contains numerous statistical concerns. First, the authors conducted a very large number of statistical tests, more than 200 tests. This very large number of analyses which are presented as independent outcomes is a concern with only 16 RCTs. Related to this issue of over testing, the measures described for assessing perfectionism are not independent measures, since they are provided by the same participants within each study. From Table 1 in Smith et al. (2023) we can see that the number of primary measures extracted from each study varies from 2-9 (mean 3.8) and the number of secondary measures from 1-5 (mean 2.5). For a meta-analysis extracting more than one measure per study a methodological guidance paper by Pigott and Polanin (2020) says: “Computing separate analyses by outcome increases the probability that a meta-analysis will be subject to multiplicity problems, or issues with conducting too many statistical tests” (page 35) (see also Polanin & Pigott, 2015). **We note that the issue of many statistical tests on a small number of studies applies to other meta-analyses of treatment efficacy, a concern across the literature.**

This is especially important because of the lack of power of most subgroup analyses.

Second, there were very small numbers of studies for many of the outcomes reported. Meta-analyses with few studies (e.g., around four or less) are fraught with the risk that wrong conclusions will be drawn because they possess too low power (Jackson & Turner, 2017). This impacts 8 of the 15 (53%) results reported in Table 2 for the combined studies, 8 of the
15 (53%) self-help studies, and all the face-to-face studies and follow-up studies. Regarding the analyses, Smith et al. (2023) comment that at least two studies are needed for a meta-analysis, but typically many more studies are needed depending on the number of trials, participants per trial, expected outcome, heterogeneity and quality or risk of bias of the trials, in order to generate reliable estimates of intervention effect (Cuijpers, 2016). While it is usually assumed that a meta-analysis is a way to increase statistical power, with too few studies, the power of a meta-analysis is lower than the average power of the individual studies within the meta-analysis, even if there is no between-study heterogeneity, as demonstrated statistically by Jackson and Turner (2017). Moreover, if the between-study heterogeneity is large then the confidence intervals for the mean effect will be too wide and contain the null, and the hypothesis test will not be able to reject the null hypothesis (Jackson & Turner, 2017). Most of the meta-analyses presented by Smith and colleagues (2023) are based on few studies and followed this pattern of wide confidence intervals. Those with more studies were statistically significant in 12 out of 14 cases (86%), while those fewer than five studies were not significant in most cases, only in 7 out of 28 (25%) cases. Jackson and Turner (2017) state “not only is statistical inference under the random-effects model challenging when there are very few studies but also less worthwhile in such cases” (page 290). Smith and colleagues’ (2023) present overly pessimistic conclusions about the efficacy of CBT-P, especially regarding impact on trait perfectionism and follow-up effects which are based on a small number of studies. There are insufficient numbers of studies which have compared CBT-P to an active treatment at follow-up to analyse if change is sustained at follow-up yet in a meta-analysis. Our suggestion is that although many authors of meta-analyses do calculate effects on a small number of studies, we argue authors are best to conduct analyses when there are sufficient numbers of studies, and to refrain from drawing conclusions based on a small number of studies (i.e., less than five per outcome), especially
when there is risk of bias and the quality of the studies is suboptimal. Doing too many subgroup analyses is a concern, a particularly important point because of the low power of these analyses.

Third, the authors calculated the pooled weighted event rates to determine reliable improvement, but they do not describe what, if any, transformation was used in this procedure. The Comprehensive Meta Analysis, version 3, was used in Smith et al. (2023), which uses logit transformation. However, Barendregt et al. (2013) and Barker et al. (2021) recommend using double arcsine transformation since it solves one problem that the logit transformation does not.

Fourth, Smith et al. (2023) state “Crucially, for studies that did not report reliable improvement or reliable deterioration, we imputed estimates using the means, standard deviations, and $N$ at pre-treatment, post-treatment, and follow-up to determine the number of participants scoring above or below a cut-off assuming a normal distribution.” However, there is no information on how many of the 16 studies provided this information and for how many studies imputed estimates had to be used. Is it reasonable to assume that this measure is normally distributed, within the 16 studies?

Fifth, Smith et al. (2023) used Klauer’s (2001) correction for potential pre-treatment differences between treatment and control conditions, where pre-post-tests are not treated as repeated measures but as independent data, thus not providing the most accurate estimate. The need for such corrections was not examined on a case-by-case basis but applied indiscriminately.

Finally, numbers-needed-to-treat (NNTs) for all Standard Mean Differences (SMDs) using the formula provided by Kraemer and Kupfer (2006) is not appropriate for the non-clinical study samples. Further, use of NNT in meta-analysis is likely to produce unreliable results and is not advised (Kraemer & Kupfer, 2006). When there are response and remission
rates NNT can be calculated directly without an estimate as the NNT is the inverse of the risk difference.

4. Population

Smith et al. (2023) challenged current research on the basis of including prevention/early intervention studies and thus including participants with modest levels of perfectionism. However, many advantages come from testing the efficacy of universal, selective, and indicated preventions, as well as assessing the efficacy of a single intervention at two or more levels. Universal prevention is used for the general population, regardless of risk level, while selective prevention is targeted at high-risk groups, such as individuals screening higher on the risk factor of perfectionism in this context. Indicated prevention is for those with symptoms of the illness, such as anxiety, depression, and eating disorders, but not meeting the criteria for a clinical diagnosis. Finally, an intervention can be used as an acute treatment approach for those with a clinical diagnosis. Rather than being seen as a limitation, this variety in the literature shows its strength and focus on the whole community, encompassing all aspects of healthcare, from health promotion and prevention to early intervention and treatment.

Conclusion

We encourage good scientific debate and collaboration. However, we discourage science that has a negative impact on the populations that we seek to help. We consider Smith et al. (2023) as an example of science which may have a negative impact on individuals seeking to engage in CBT-P. Our concerns, in summary, include that their meta-analysis omitted two RCTs that were included in the previous meta-analyses they sought to replicate, they did not conduct a systematic search of the literature, combined studies from reviews with different objectives and search methodologies as if they were homogenous, used
multiple statistical tests on a small number of studies, introduced novel outcomes, and
presented their findings selectively.

**Future directions**

We urge clinicians to read the original studies and meta-analyses that repeatedly find
a positive impact of CBT-P on perfectionism and related disorders in multiple populations,
across settings and formats. There is a need to consider a constructive way forward to work
together as researchers in perfectionism, with the shared goal of improving treatment and
outcomes for those with perfectionism. For the interested reader we point them to the ideas
outlined in Shafran et al. (2023), which provide a critique and future research agenda in this
area.

Shafran et al. (2023) outline that a key area for development is comparing CBT-P to
active treatments. There are currently further studies in progress comparing CBT-P to active
treatment comparisons (e.g., Buhrman et al., 2020), adding to a small number of comparisons
(e.g., Mahmoodi et al., 2021; Shu et al., 2019). Shafran at al. (2023) highlighted that
numerous more recent evaluations have been conducted independently of Shafran, Egan and
Wade appearing increasingly in the literature (e.g., Mahmoodi et al., 2021; Ward & Wheaton,
2022). Smith et al. (2023) criticised that many studies have been conducted by these authors
and therefore lack independence, so this recent independent evaluation is important. A key
strength Shafran et al. (2023) outlined is the scalability of CBT-P, which has been examined
across ages, different populations, internet-delivered, and in traditional self-help book
formats. We believe dissemination is a critical aim for psychotherapy and therefore it is a
strength that most studies of CBT-P have focused on approaches which require minimal or no
therapist time, given there will never be enough therapists available to reach those in need
with perfectionism. Shafran et al. (2023) outlined that future advancements in CBT-P may
occur through single session approaches (Schleider et al., 2022), developing treatment for children, and co-designed treatment with lived experience experts (see Egan, Wade, Fitzallen, O’Brien, & Shafran, 2022; O’Brien et al., 2022; Wade et al., 2021). Shafran et al. (2023) concluded that CBT-P has demonstrated clinical utility. Our hope is that research in the treatment of perfectionism can move forward by constructive work together as researchers, with the shared goal of reducing distress associated with perfectionism.
References


https://doi.org/10.1037/ccp0000044

https://doi.org/10.1080/16506073.2021.1952302

https://doi.org/10.1080/16506073.2001.1952302


https://doi.org/10.1016/j.biopsych.2005.09.014


https://doi.org/10.1017/S1352465814000162

Mahmoodi, M., Bakhtiyari, M., Arani, A.M., Mohmadi, A., & Isfeedvajani, M.S. (2021). The comparison between CBT focused on perfectionism and CBT focused on emotion regulation for individuals with depression and anxiety disorders and dysfunctional


