

## •Commentary•

# Addressing risk of bias in trials of cognitive behavioral therapy

Katherine S. BUTTON<sup>1,\*</sup>, Marcus R. MUNAFÒ<sup>2</sup>

**Summary:** A recent network meta-analysis by Zhu and colleagues reported in the *Shanghai Archives of Psychiatry* compared two different comparators (psychological placebo and waitlist control) in trials assessing the effectiveness of cognitive behavioral therapy (CBT) for the treatment of generalized anxiety disorder (GAD). CBT was superior to both of these control conditions, but psychological placebo was superior to waitlist. However, we argue that the term 'psychological placebo' is a misnomer because the impossibility of effectively blinding participants to treatment allocation in CBT trials makes it impossible to control for placebo effects. This failure to blind participants and therapists – and the resultant high risk of bias – was the main reason Zhu and colleagues found that the overall quality of the evidence supporting the conclusion that CBT is effective for GAD is poor. This is a general problem in all psychotherapy trials, which suffer from well-documented methodological and conceptual problems that prevent adequate placebo control and undermine causal inference. We discuss these problems and suggest potential solutions. We conclude that, while it may be difficult to remove potential bias in randomized controlled trials of psychotherapy, we can improve on the status quo by integrating basic science within applied trials to adjust for these biases and, thus, improve the strength of the causal inferences.

**Keywords:** cognitive behavioral therapy; psychotherapy; generalized anxiety disorder; randomized controlled trial; meta-analysis; network meta-analysis; psychological placebo

[*Shanghai Arch Psychiatry*. 2015; 27(3): 144-148. Epub 2015 Apr 24. doi: <http://dx.doi.org/10.11919/j.issn.1002-0829.215042>]

## 1. Introduction

A recent network meta-analysis by Zhu and colleagues<sup>[1]</sup> compared two different comparators (psychological placebo and waitlist control) in trials assessing the effectiveness of cognitive behavioral therapy (CBT) for the treatment of generalized anxiety disorder (GAD). CBT was found to be superior to waitlist control and (to a lesser extent) to psychological placebo. Moreover, psychological placebo was also superior to waitlist control, which might be interpreted as evidence of the need to control for the non-specific effects of therapy, such as therapist and researcher contact time. However, we argue that 'psychological placebo' is a misnomer as it fails to meet key criteria for controlling for placebo effects.

Zhu and colleagues also identified problems with study quality in this literature. Using the GRADE criteria for study quality (<http://www.gradeworkinggroup.org/intro.htm>) developed by the Cochrane Collaboration, they found that the quality of the evidence supporting the conclusion that CBT was effective was poor.<sup>[2]</sup> Eight of the 12 studies they identified were classified as at 'high risk of bias' and the quality of evidence as a whole was classified as 'moderate', suggesting that the evidence for the effectiveness of CBT for GAD is not yet robust. This is surprising given the confidence many practitioners have in the effectiveness of CBT. However, the evidence-base for psychotherapy as a whole suffers from well-documented methodological and conceptual problems that prevent adequate placebo control and,

<sup>1</sup> School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom

<sup>2</sup> MRC Integrative Epidemiology Unit, School of Experimental Psychology, University of Bristol, Bristol, United Kingdom

\* correspondence: [Kate.Button@bristol.ac.uk](mailto:Kate.Button@bristol.ac.uk)

thus, undermine the strength of casual inferences.<sup>[3-5]</sup> This commentary discusses these problems and suggests potential solutions.

## 2. Causal inference, placebo effects, and the importance of blinding

Randomized controlled trials (RCTs) are the gold standard for causal inference, and underpin evidence-based practice.<sup>[6]</sup> Concealed random allocation ensures both measured and unmeasured confounders are randomly distributed across treatment arms. The control condition or comparator excludes changes that are not caused by the intervention, such as natural recovery over time, and placebo effects (i.e., differences in outcomes due to psychological factors such as expectancy of improvement). Placebo effects are powerful, producing clinically important improvements that correlate with expectations of improvement.<sup>[3]</sup> Blinding of participants, clinicians, and researchers to treatment allocation prevents the introduction of bias in the reporting and assessment of outcomes.

In medical trials, placebo effects are controlled for by means of an identical but 'non-active' comparator, such as a sugar pill for antidepressants or sham surgery for deep brain stimulation. The patient can be blinded as to whether they are in the active or placebo condition, and expectancy of improvement can be held constant across both treatment arms. Placebo effects can then be confidently excluded as a potential cause of any observed treatment effect. However, RCTs can yield biased results if they fail to meet these standards for blinding.<sup>[6]</sup>

Results of RCTs may be biased by psychological factors, such as a desire among participants or researchers for the trial to 'work' (demand characteristics) influencing outcome assessment,<sup>[4]</sup> or when participants self-report outcomes in a manner which they think desirable to the trial team. Researchers may treat patients in the active and comparator arms differently, or complete the outcome assessment in a biased way. Blinding of patients and researchers is therefore crucial to ensure unbiased results.<sup>[7]</sup> The extent and success of blinding often depends on the type of comparator; for example, it is relatively easy to blind all participants and researchers when a sugar pill is used as the comparator, but it may not be possible to blind the surgical team when comparing an actual surgery and a sham surgery.

### 2.1 The problems with psychological placebos

The purpose of a placebo is to control for the psychological effects of administering a treatment.<sup>[3]</sup> To achieve this, the placebo must appear identical to the active treatment but be missing the 'active ingredient', and the patient must be blinded. These conditions do not map well to psychotherapy. CBT targets psychological processes, so disentangling placebo effects from treatment effects is challenging. It is difficult to create an 'inactive' but

comparable psychological control, as the more similar the psychological placebo appears to CBT the more 'active' it is likely to be. Relaxation therapy or supportive consultation are often used as control conditions for CBT trials, but these may still have therapeutic effects. Finally, and critically, it is impossible to blind participants and therapists, as CBT involves active engagement with its theoretical rationale.<sup>[4]</sup> We concur with others who suggest that the concept of a 'psychological placebo' is both methodologically and conceptually flawed.<sup>[3]</sup>

Lack of blinding is arguably the main cause of the high risk of bias in CBT trials. None of the 12 studies identified in Zhu and colleagues' review had blinded the treating therapists or the participants to their treatment allocation so the risk of bias for these items was rated as 'high'. All of the studies had blinded assessors of various outcome measures, but in 6 of the 12 studies the main outcome measure was based on the results of a self-reported scale completed by the participant (i.e., it was not assessed by the blinded assessor). It is possible, although not straightforward, to blind outcome assessment. However, even when the assessors are effectively blinded to treatment allocation, bias could still occur due to demand characteristics if patients are aware of their treatment allocation and answer the blinded assessor's questions in a biased manner. Thus failure to control for differences in patient expectations, and demand characteristics, coupled with subjective self-reported outcome assessment, introduces a high risk of bias.

In the studies reviewed by Zhu and colleagues, psychological placebo was operationalized as either one-to-one supportive consultation or general group discussions of psychological problems.<sup>[1]</sup> While 'psychological placebo' is a misnomer due to the inability to blind trial participants to their treatment allocation, controlling for non-specific factors such as therapist contact time and general supportive discussions is useful when assessing the effectiveness of the specific components of CBT such as exposure therapy and reappraisal. However, such studies – which may underestimate the beneficial effects of CBT – should be relabeled as trials comparing different psychological interventions rather than as 'placebo-controlled' trials.

Given these difficulties, it is unsurprising that the majority of CBT trials use comparators such as no treatment, waitlist, or treatment as usual. Of these, treatment as usual is arguably the most informative (and conservative), because it addresses the important pragmatic question "Is CBT more effective than current best treatment?". However, the majority of studies reviewed by Zhu and colleagues used waitlist control. In a recent network meta-analysis patients randomized to waitlist control were found to do worse than those randomized to no treatment,<sup>[8]</sup> suggesting that waitlist may act as a nocebo (i.e., having an adverse effect on patients). Patients randomized to waitlist may have negative expectations of improvement as they are placed in a state of stasis rather than actively receiving

treatment. Furthermore, the sense of ‘waiting’ might discourage seeking treatment elsewhere. The implication is that treatment effects from comparison with waitlist are likely to be inflated, and thus provide inaccurate estimates.

### 3. What can be done to improve the evidence-base for CBT?

Does it matter that the evidence for CBT is not robust according to the GRADE criteria? More importantly, does it matter that we cannot make causal inferences about the effectiveness of CBT because of the possibility that observed effects are due to placebo (and/or nocebo) effects, or other biases arising from lack of blinding? Pragmatically, the evidence that CBT is superior to treatment as usual, irrespective of how, may be sufficient to guide the decision to support a treatment. However, the evidence reviewed by Zhu and colleagues is not ideal for addressing either of these points because it did not compare CBT to treatment as usual; comparisons with waitlist likely overestimate effectiveness, while comparisons with psychological controls likely underestimate effectiveness.

The evidence for CBT from traditional RCT designs may never fully satisfy the most robust GRADE criteria. However, causal inferences about the effectiveness of CBT could be strengthened in at least three ways. First, by going back to the laboratory to understand the basic science questions of how CBT and placebo effects work. Second, by including measures of bias in controlled clinical trials to allow post-hoc adjustment for potential confounding in treatment effects. Third, by using sophisticated multi-arm analyses to assess the relative effects of components of interventions, and meta-regression to explore how the level of risk of biases influences treatment effect.

#### 3.1 A call for basic science

CBT is a complex intervention, comprising a range of behavioral and cognitive techniques. Elucidating which are ‘active ingredients’ and which are redundant is an important basic science question, with clear clinical application. Understanding the mechanisms of change can lead to improved and more efficient therapies by focusing on those components which work and removing those that do not, and can lead to new avenues for treatment, potentially increasing the psychological treatment toolkit.

One way to address basic questions of mechanism is to examine each component of CBT in isolation in the laboratory.<sup>[9]</sup> For example, the protocol for exposure therapy, a key behavioral technique used in CBT for anxiety disorders, was developed in the 1960s from the basic science of fear-extinction learning. The mechanism by which exposure to the fear stimulus (e.g., spiders, social situations) extinguishes the fear response and reduces anxiety is now well understood, and basic research continues to inform the optimization

of related psychological therapies.<sup>[10]</sup> Mechanism studies could be routinely embedded within clinical trials, by incorporating measures of the psychological processes CBT is thought to target. This would be an efficient and cost-effective way to understand mechanisms. Basic science can also delineate the causal mechanisms by which placebos produce their effects.<sup>[3]</sup> The experimental pursuit of understanding the mechanisms of placebo effects will inform both our understanding of the active ingredients of current therapies and contribute to basic knowledge that can be used to maximize placebo effects to improve patient outcomes. More detailed understanding of placebo effects would also lead to the development of more precise measures of placebo effects which could be used to statistically adjust for these effects in RCTs (discussed below).

#### 3.2 A sticking plaster (bandaid) for clinical trials

Can we control for bias introduced by inadequate blinding statistically? Repeated measures of expectancy and beliefs about the demands of the research throughout the trial would provide a means by which potential bias could be adjusted for. If such data were routinely collected, summary effect estimates could be adjusted in an evidence syntheses to account for these potential biases.

There are theoretical reasons why we might expect that treatment effects for CBT would persist longer than placebo effects. CBT targets thinking styles, so adopting longer-term follow-up assessments (i.e., after the active treatment is completed) as the primary outcome may increase confidence in the efficacy of CBT. Similarly, research into placebo effects of antidepressants suggests that placebo effects vary with severity, being greatest for mild depression and decreasing as the severity of the depressive symptoms of participants increases.<sup>[11]</sup> If this also holds for GAD, subgroup analyses based on the severity of GAD might indirectly inform causal inference. The more we understand about the mechanisms underlying placebo and CBT effects, the better we can design pragmatic clinical trials to overcome the limitations associated with lack of blinding.

#### 3.3 Network meta-analysis and evidence synthesis

Treatment effects are relative, determined as much by the comparator as by the intervention. Network-meta analysis provides a powerful tool for comparing how treatment effects vary when different comparators are employed. To retain power in pairwise comparisons in traditional meta-analysis, different comparators (which can have very different effects on treatment effects) tend to be combined into a single control group, or separated as subgroups. Network meta-analysis goes beyond the traditional meta-analysis approach by additionally allowing indirect comparisons between the various control conditions, which can be quite useful. For example, this approach helped identify the nocebo

effects of waitlist control, which indirectly suggests that the treatment effects reported from trials using waitlist controls likely over-estimate the effectiveness of CBT.<sup>[8]</sup>

One way to overcome issues of expectancy and lack of blinding is to compare the same therapy with one or more components added or removed, to assess whether or not the component(s) directly causes the treatment target. This approach should control for most forms of bias (with the possible exception of therapist blinding), as patients would expect to improve similarly in both arms (assuming they are naive to the missing ingredient). However, these types of comparison trials would require large numbers of patients to have sufficient statistical power to detect the potentially small treatment effects attributable to single techniques. Alternatively, if different trials used slightly different CBT protocols, network meta-analytic approaches might provide a means of assessing the relative effectiveness of these protocols and, thus, assess the effects of the CBT components that are different between the trials. Finally, network meta-analysis could also be used to estimate the direction and extent of bias caused by lack of blinding and to make appropriate adjustments in the final estimates of the effectiveness of CBT.<sup>[12]</sup>

#### 4. Closing remarks

In most treatment studies of CBT (and other psychotherapeutic interventions) strong causal inferences about effectiveness are not justified because of the possibility that observed effects are due to placebo (and/or nocebo) effects or other biases arising from lack of blinding. Addressing the basic science questions of how CBT and placebos work will provide a better understanding of what is causal in complex interventions. Such knowledge could lead to the development of more efficient or novel psychological interventions and to improved trial design. The better we understand placebo effects, the better we can assess and quantify them, allowing more precise

measures of the placebo effect to be incorporated into clinical trials and adjusted for in statistical analyses. Similarly, we can include measures of the psychological processes thought to be targeted by CBT in clinical trials and assess how these measures change during the course of CBT treatment. Therefore, while it may be virtually impossible to remove potential bias in RCTs of psychotherapy that arise from the placebo effect and the lack of effective blinding, we can improve on the status quo by integrating basic science within applied trials to adjust for these biases and, thus, make stronger causal inferences.

#### Acknowledgements

The authors wish to thank Dr. Deborah Caldwell for her helpful comments. KSB is funded by the National Institute for Health Research School for Primary Care Research (NIHR SPCR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health of the United Kingdom. The NIHR SPCR is a partnership between the Universities of Birmingham, Bristol, Keele, Manchester, Nottingham, Oxford, Southampton, and University College London. MRM is a member of the UK Centre for Tobacco Control Studies, a UKCRC Public Health Research Centre of Excellence. Funding from the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

#### Conflict of interest

The authors report no conflict of interest related to this manuscript.

#### Funding

No funding was received to prepare this commentary.

## 解决认知行为治疗研究中存在的偏倚风险

Button KS, Munafò MR

**概述：**《上海精神医学》杂志最近刊登了一篇由朱智佩及其同事撰写的网状 meta 分析文章。该 meta 分析荟萃比较了采用两种不同的对照组（心理安慰剂组和等候治疗对照组）评估认知行为治疗 (cognitive behavioral therapy, CBT) 对广泛性焦虑障碍 (generalized anxiety disorder, GAD) 疗效的研究。CBT 优于这两种对照，但心理安慰剂优于等候治疗。然而，我们认为“心理安慰剂”一词不恰当，因为在 CBT 研究中受试者不可能完全不知道治疗分配，所以研究不可能真正控制安慰剂效应。无法使受试者和治疗师双盲以及因此而产生的高偏倚风险正是朱智佩及其同事发现支持 CBT 治疗 GAD 有效的证据整体质量差的主要原因。上述问题在所有的心理

治疗研究中普遍存在，方法学问题以及概念性问题限制了安慰剂对照的作用，并削弱了对因果关系的推断作用。本文讨论了这些问题并提出可能的解决方案。我们的结论是，虽然在心理治疗的随机对照研究中可能难以完全消除潜在的偏倚，但是可以在进行研究时整合基础科学知识，来校正这些偏倚，改进现状，从而提高因果推论的强度。

**关键词：**认知行为治疗；心理治疗；广泛性焦虑障碍；随机对照研究；meta 分析；网状 meta 分析；心理安慰剂

本文全文中文版从 2015 年 08 月 06 日起在  
<http://dx.doi.org/10.11919/j.issn.1002-0829.215042> 可供免费阅读下载



## References

1. Zhu Z, Zhang L, Jiang J, Li W, Cao X, Zhou Z, et al. Comparison of psychological placebo and waiting list control conditions in the assessment of cognitive behavioral therapy for the treatment of generalized anxiety disorder: a meta-analysis. *Shanghai Arch Psychiatry*. 2014; **26**(6): 319-331. doi: <http://dx.doi.org/10.11919/j.issn.1002-0829.214173>
2. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008; **336**(7650): 924-926. doi: <http://dx.doi.org/10.1136/bmj.39489.470347.AD>
3. Kirsch I. Placebo psychotherapy: Synonym or oxymoron? *J Clin Psychol*. 2005; **61**(7): 791-803. doi: <http://dx.doi.org/10.1002/jclp.20126>
4. Borkovec TD, Sibrava NJ. Problems with the use of placebo conditions in psychotherapy research, suggested alternatives, and some strategies for the pursuit of the placebo phenomenon. *J Clin Psychol*. 2005; **61**(7): 805-818. doi: <http://dx.doi.org/10.1002/jclp.20127>
5. Flint J, Cuijpers P, Horder J, Koole SL, Munafo MR. Is there an excess of significant findings in published studies of psychotherapy for depression? *Psychol Med*. 2015; **45**: 439-446
6. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010; **340**: c332. doi: <http://dx.doi.org/10.7326/0003-4819-152-11-201006010-00232>
7. Juni P, Altman DG, Egger M. Systematic reviews in health care - Assessing the quality of controlled clinical trials. *Brit Med J*. 2001; **323**(7303): 42-46
8. 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**(3): 181-192. doi: <http://dx.doi.org/10.1111/acps.12275>
9. Holmes EA, Craske MG, Graybiel AM. Psychological treatments: A call for mental-health science. *Nature*. 2014; **511**(7509): 287-289. doi: <http://dx.doi.org/10.1038/511287a>
10. Graham BM, Milad MR. The study of fear extinction: implications for anxiety disorders. *Am J Psychiatry*. 2011; **168**(12): 1255-1265. doi: <http://dx.doi.org/10.1176/appi.ajp.2011.11040557>
11. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med*. 2008; **5**(2): e45. doi: <http://dx.doi.org/10.1371/journal.pmed.0050045>
12. Dias S, Welton NJ, Marinho VCC, Salanti G, Higgins JPT, Ades AE. Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. *J R Stat Soc A Stat*. 2010; **173**: 613-629. doi: <http://dx.doi.org/10.1111/j.1467-985X.2010.00639.x>

(received, 2015-03-23; accepted, 2015-04-05)



*Katherine Button is a National Institute for Health Research School for Primary Care Post-doctoral Fellow at the University of Bristol in the United Kingdom. Her research focuses on the interface between neuroscience and psychiatry, where she aims to translate insights from the neuroscience of social cognitive mechanisms of anxiety and depression to clinical application in primary care. Katherine read neuroscience as an undergraduate at Fitzwilliam College, University of Cambridge, before moving to the University of Bristol to complete a MEd in Psychology of Education. In January 2013 she was awarded a PhD in Psychiatry from the School of Social and Community Medicine, University of Bristol, where she was a Medical Research Council Centenary Award Fellow until starting her current fellowship in April 2014.*