How to prove that your therapy is effective, even when it is not: a guideline

P. Cuijpers[1,2]* and I. A. Cristea[3,4]

[1] *Department of Clinical Psychology, VU University Amsterdam, The Netherlands*
[2] *EMGO Institute for Health and Care Research, The Netherlands*
[3] *Department of Clinical Psychology and Psychotherapy, Babeş-Bolyai University, Cluj-Napoca, Romania*
[4] *Clinical Psychology Branch, Department of Surgical, Medical, Molecular and Critical Pathology, University of Pisa, Pisa, Italy*

Aims. Suppose you are the developer of a new therapy for a mental health problem or you have several years of experience working with such a therapy, and you would like to prove that it is effective. Randomised trials have become the gold standard to prove that interventions are effective, and they are used by treatment guidelines and policy makers to decide whether or not to adopt, implement or fund a therapy.

Methods. You would want to do such a randomised trial to get your therapy disseminated, but in reality your clinical experience already showed you that the therapy works. How could you do a trial in order to optimise the chance of finding a positive effect?

Results. Methods that can help include a strong allegiance towards the therapy, anything that increases expectations and hope in participants, making use of the weak spots of randomised trials (risk of bias), small sample sizes and waiting list control groups (but not comparisons with existing interventions). And if all that fails one can always not publish the outcomes and wait for positive trials.

Conclusions. Several methods are available to help you show that your therapy is effective, even when it is not.

Introduction

Randomised controlled trials have become the gold standard to prove that therapies for mental health problems, are effective and have even been regarded as 'objective scientific methodology' (Kaptchuk, 2001) (p. 541). Treatment guidelines are using these randomised trials to advise professionals to use specific interventions and not others, and policy makers and health insurance companies use this evidence to decide whether or not to adopt and implement a particular intervention.

The experimental design of randomised trials is straightforward and not very complicated, but the results are deemed very strong from a scientific point of view and considered to be the strongest scientific evidence available. By splitting one group of participants randomly into two subgroups, with one receiving the therapy and the other a control or alternative intervention, differences between these subgroups must be caused by the therapy that the one subgroup received and the other did not. Logically there is no other explanation possible (Nezu & Nezu, 2008).

So, suppose you have developed a new and innovative therapy or you have been working for several years with a therapy you believe is effective. The patients receiving this therapy are satisfied and they tell you that this therapy has helped them a lot. So, you do not really need a trial, because based on a rich clinical experience and case studies you already know your therapy works. However, in order to get it into treatment guidelines you have to show in a trial that this therapy is effective. Then your therapy gets the tag of 'evidence-based' or 'empirically-supported' and that can help in getting it better implemented and disseminated.

If this were your starting position, how could you make sure that the randomised trial you do actually results in positive outcomes that your therapy is indeed effective? There are several methods you can use to optimise the chance that your trial will show

* Address for correspondence: P. Cuijpers, Professor of Clinical Psychology, Department of Clinical Psychology, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands.

(Email: p.cuijpers@vu.nl)

that the intervention works. Even when in reality it does not really work. The goal of this paper is to describe these 'techniques'.

Have a strong allegiance to your therapy

If you are the developer of the therapy or have worked with it for a long time, you have in fact already attained one important method to optimise the chance that the results of your trial will be favourable. There is a lot of research showing that when authors of randomised trial have a strong allegiance towards the intervention they examine, they obtain outcomes favouring that intervention (Luborsky *et al.* 1999; Munder *et al.* 2011, 2012, 2013). Other meta-analyses (Miller *et al.* 2008) have even shown that in direct comparisons between different therapies, controlling for researcher allegiance effectively eliminated any observed systematic differences between treatments. In fact, some researchers (Wampold, 2001) went as far as to state that '*allegiance to therapy is a very strong determinant of outcome in clinical trials*' (p. 168).

It is not clear, why this happens and exactly how this works. Possible mechanisms include the possibility that an investigator keen on a therapy might favour better training and supervision of therapists implementing it as opposed to a less preferred alternative (Leykin & DeRubeis, 2009). Or that simply therapists and investigators have better expertise and skill in implementing the preferred treatment, directly but 'honestly' contributing to its superior performance (Hollon, 1999). However, it is very well possible that allegiance in a researcher just indicates that this researcher is more inclined to use the other techniques that we will lay out below.

Of course, it would also be possible to let this therapy be tested by an independent group of researchers, and most regulating bodies do require independent testing before an intervention can be implemented. However, if you use the techniques described in this paper well, it may very well be possible that the effects you find are quite large. And when independent trials are done that do find smaller effects, still, later on, meta-analyses pooling all the trials examining the therapy end up producing higher mean effects because of the very large effects you have realised in these first trials. An aggravation of this is the so-called 'time lag bias'- the phenomenon in which studies with positive results get to be published first and dominate the field, until the negative, but equally important, studies are published (Ioannidis, 1998; eds Higgins & Green, 2011). So, by the time the negative results start to pile up, you can already count on quite a few trials with positive results (some with huge effect sizes too) and even some meta-analyses summarising

these trials and finding that the therapy is essentially effective.

Increase expectations in patients

One of the interesting characteristics of many mental health problems is that they respond quite well to placebo effects. That is not something unique for mental health problems, as it is also present in chronic medical illnesses with a fluctuating course that are associated with subjective distress (Khan & Brown, 2015), hypertension (Preston *et al.* 2000), osteoarthritis (Moseley *et al.* 2002) and Parkinson's disease (Ondo *et al.* 2007). It is not clear how placebos work, but it is assumed that they are the products of a general expectancy learning mechanism in which verbal, conditioned and social cues are centrally integrated to change behaviours and outcomes (Colagiuri *et al.* 2015). But an important consequence of the placebo effect is that many patients get better anyway, as long as they expect the therapy to work. Patients also typically think that they improved because of the treatment, even when this is the result of the placebo effect, hence nurturing their expectations about the efficiency of the therapy and furthering the placebo effect.

Consequently, users of many therapies are happy with the outcomes as they have improved, and the deliverers of the treatment are inclined to think that it is the intervention that caused the improvement. This could be an explanation why many interventions, including exotic ones, such as acupuncture (Wu *et al.* 2012; Boyuan *et al.* 2014; Rafiei *et al.* 2014; Errington-Evans, 2015), swimming with dolphins (Fiksdal *et al.* 2012) or other animal-assisted therapies (Kamioka *et al.* 2014*a*), horticultural therapy (Kamioka *et al.* 2014*b*) or dancing Argentine tango (Pinniger *et al.* 2012) can still be considered to be effective by patients and therapists. Jerome Frank suggested already in the 1950s (Frank & Frank, 1991) that the most important effects of psychotherapies were caused by the expectations of the patients, the decision they made to seek help, the suggestion and hope that the specialist who treated them was an expert really capable of helping them.

However, it is very well possible to strengthen expectations and hope in participants to the therapy. Just express your own belief to them, namely that this is the best therapy currently available. And you can advertise for your trial in the media explaining why your intervention is so innovative and unique and definitely the best among the available interventions. For instance, in the case of one popular anxiety treatment, cognitive bias modification (CBM), the authors recounted (Carey, 2012) that after one of their studies, still recruiting, was featured in a highly laudatory *Economist* expose about CBM interventions

(Psychiatry: Therapist-free therapy, 2011), featuring an enticing subtitle 'CBM may put the psychiatrist's couch out of business', they began to be flooded by participants eager to enter their trial. Not only that, but people who reported having learnt about the trial from the *Economist* piece all responded well, whether they were getting the CBM treatment or the placebo, as if, the authors noted, 'the article itself had some power of suggestion'. Ask a few participants to declare that they have benefited very much from the therapy and tell their personal stories. You can also go to conferences and testify about your clinical experiences with the therapy, give educational workshops on the therapy where you present successful case studies, and convince other deliverers of treatments that this is really something new, surprisingly effective and definitely worth trying on patients. If and when they start using it, they will also increase expectations and hope in the participants, who in their turn will indeed experience how good this therapy is. There are several examples of early trials that find extraordinary large effects of such new and innovative treatments, which resulted to be much smaller in later independent trials (Nezu, 1986; Schmidt *et al.* 2009).

Use the 'weak spots' of randomised trials

Another thing that you have to learn when you want to optimise the effects found for your therapy is that randomised trials have 'weak spots', also called 'risk of bias'. As indicated earlier, the logic of trials is quite straightforward, but there are several points in the design where the researchers can have an influence on the outcomes of the trial (Table 1).

One of these weak spots is the randomisation of participants. The randomisation is the core of the trial, because if the assignment of participants to the groups is not random, the effects found later on are possibly not caused by the intervention, but by baseline differences between the groups. Until about 20 years ago, the method of randomisation was not described in most studies on therapies at all and it was usually only reported that participants were randomised. Since the concept of risk of bias was introduced (Higgins *et al.* 2011), the methods of randomisation are more often described in reports of therapies, but this is still not always done. For instance, for depression the percentage of psychotherapy trials with properly conducted randomisation is around 35% (Chen *et al.* 2014). Whether this is due to randomisation not having been conducted properly or simply not having been described in the report remains anyone's guess.

There are two important aspects of randomisation. The first is that the random numbers should be generated in the right way, for example by using a computerised random number generator or coin toss, instead of for example the date of admission, date of birth or clinic record number. The second aspect is the allocation concealment. Researchers conducting the trial or their assistants can assign participants, they expect to respond well to the intervention to the intervention group instead of to the control group. Therefore, it is important that allocation is done by an independent

Table 1. *Ten methods that can help prove that your intervention is effective (even when it is not)*

1. Express in all communications about the intervention that you as developer or expert believe it to be best intervention ever (helps to increase expectations in participants).

2. Do everything else that can increase expectations, such as writing books about the intervention, going to conferences to convince other professionals that this is the best intervention ever, giving interviews in the media showing your enthusiasm, preferably seasoned with some personal stories of participants who declare they have benefited very much from the intervention.

3. Use the 'weak spots' of randomised trials: let the assignment to conditions be done by research staff involved in the trial or do it yourself (not by an independent person not involved in the trial).

4. Do not conceal conditions to which participants were assigned to for the assessors of outcome.

5. Analyse only participants who completed the intervention and ignore those who dropped out from the intervention or the study (and do not examine all participants who were randomised).

6. Use multiple outcome instruments and report only the ones resulting in significantly positive outcomes for the intervention.

7. Use a small sample size in your trial (and just call it a 'pilot randomised trial').

8. Use a waiting list control group.

9. Do not compare the intervention to already existing ones (but do tell your colleagues that based on your clinical experiences you expect that this intervention is better than other existing ones (good for the expectations))

10. If the results are not positive, consider not publishing them and wait until one of the clinicians you have persuaded about the benefits of this intervention conducts a trial that does find positive outcomes

person not involved in the trial, or by sequentially numbered, opaque, sealed envelopes. In a meta-analysis of psychotherapies for adult depression clear indications were found that studies in which allocation was properly concealed resulted in significantly smaller effect sizes than studies in which this was not done (Cuijpers *et al.* 2010*b*).

Another weak spot you can use to influence the outcomes of your trial is to use non-blinded raters of clinical assessments of outcome. So, if the clinicians or research assistants know the condition to which the participant were assigned, they may be inclined to think that the participants getting the therapy score better simply because they received the therapy. So, if you want to make use of this, just inform the raters of the assigned condition, or don't say anything and hope that in the interaction between the rater and the participant it becomes clear whether the latter was in the intervention condition or not. But don't instruct raters and participants not to talk about it as doing this may result in smaller effects of the therapy. Predictably, studies with proper blinding also resulted in smaller effect sizes for psychotherapy for depression (Cuijpers *et al.* 2010*b*).

A further possibility to strengthen the outcomes of the trial in favour of the therapy lies in another weak point of the trial, namely the study drop-outs. Often the people who drop out from a study are the ones who do not respond to the intervention or who are experiencing side effects. It does not help them, it may even harm them, so why would they continue with it? What you can do is just ignore these drop-outs in the analyses of the outcomes, and look exclusively at completers, participants who actually stayed on in the therapy and in the study. That suggests the therapy has better outcomes than when you had also included individuals who dropped out.

The correct alternative would be to implement the *intent to treat* principle of the trial, meaning that all participants who are randomised should also be included in the final analyses. There are several techniques available for imputing the missing data from drop-outs such as using the last observation that is available, multiple imputation techniques or mixed models for repeated measurements (Siddiqui *et al.* 2009; Crameri *et al.* 2015). However, by ignoring these missing data, the effects of the therapy can be increased considerably.

Finally, another weak spot of trials is in the use of outcome measures. So, if you want to make use of this weakness, you should include multiple outcome measures, and then when you analyse the results you can simply look at which outcome measure has the best results. Then you present these outcomes in your reports and simply not mention the other measures or sweep them under the rug as secondary. This

'selective outcome reporting' is getting more and more difficult to realise because the protocol of trials are now more often published, which allows reviewers to verify whether the reported outcomes were also the ones that were planned. However, not all protocols of trials are published in trial registries, so this still remains an available option. Even in the cases where protocols are published, a number of problems easily go unnoticed (Coyne & Kok, 2014). Information in trial registries can be modified and even if in most registries these changes are saved and can potentially be browsed, reviewers, clinicians and patients seldom take on the painstaking operation of going through them. But even in the cases changes are minor or non-existent, many other issues may arise. In many cases, registration is not prospective, and it's done after the trial has started (sometimes even an year after the start of the trial), which presumably gives investigators ample time to observe the direction in which results are going and which outcomes are more readily affected by the intervention. Another equally serious problem regards discrepancies about the primary outcome between the trial protocol and subsequent published reports. Published articles can simply fail to mention the trial registration number, again not prompting readers to dig up the available protocol and check for possible selective outcome reporting. For instance, going back to the case of acupuncture, a recent systematic analysis documented almost all of these problems: only roughly 20% of the trials were registered before the start of the trial, and in around 40% of the published reports the trial registration number was available. But most disquietingly, in 45% of the cases where a comparison between registered and published primary outcomes could be carried out, there was evidence of inconsistency and in the overwhelming majority of these cases (over 70%) the discrepancy favoured statistically significant primary outcomes.

Design your trial in the right way: small samples, waiting list control groups but no comparative trials

If you really want the trial to show that your therapy is effective there are some other techniques you can use. First, you should use a small sample size. There is quite some evidence that small sample sizes in trials result in better outcomes than trials with large samples, not only in psychotherapy (Cuijpers *et al.* 2010*b*), but also for example in pharmacotherapy for depression (Gibertini *et al.* 2012). Small samples make it possible that there are systematic differences between groups given by the fact the numbers are not large enough to be able to accurately reproduce a chance distribution of these differences. For an intuitive example, just think of the basic statistics problem of the coin toss: doing it

ten times might lead you to wrongly assume that one of its sides is likely to appear with more frequency, but that would not happen if you tossed the coin 100 times. Small groups also render the influence of outliers very powerful and give you the possibility to 'play' with varying thresholds of excluding them from the analysis. However, it is also plausible that this 'small sample bias' is caused by the previously mentioned methods of increasing expectations in participants, as early pilot projects may attract a specific type of participant, who is willing to undergo a new treatment and expect strong effect from it.

Another important possibility is to use a waiting list control group. When participants are on such a waiting list they probably do nothing to solve their problems because they are waiting for the therapy (Mohr *et al.* 2009, 2014). If they had been assigned to a care usual control group, at least some of them would have possibly taken other actions to solve their problems. Perhaps patients are willing to be randomised to a waiting list control group only if they have high expectations for the therapy. Why else would they want to wait for getting treatment? Several meta-analytic studies have shown that waiting list control groups result in much larger effects for the therapy than other control groups (Mohr *et al.* 2009, 2014; Barth *et al.* 2013; Furukawa *et al.* 2014). In fact, a meta-analysis of psychotherapies for depression (Furukawa *et al.* 2014) even suggested waiting list might be a nocebo condition, performing worse than a simple no treatment one. So, in our case, given we want to see good effects of our treatment, a waiting list is definitely the best option.

What you certainly should not do is compare your new therapy to an already existing therapy for the same problem. Of course, you can say during presentations about your therapy that you think your therapy works better than the existing ones and that user reports are so positive that superior effects are very probable, but you should not examine that in your trial. The reason why you should not examine that is that your trial should be small (see above) and in order to show that your therapy is better than existing interventions, you will have to design a trial with a very large group of participants. Because you cannot expect that your therapy is notably better than existing therapies you must assume that the difference between them is small. But if you calculate how many participants are needed for finding a small effect you easily end up with a trial of several hundred or even a 1000 participants (Cuijpers & van Straten, 2011). Such a trial is not feasible, since it is very expensive and you run the risk that it does not support your original assumption (that your therapy is better than existing ones). Of course, the reason why a new therapy is needed in the first place is exactly that it should

be better than existing ones, or at least for some patients, either in terms of efficiency, side effects or of reduced costs (why else would we need a new intervention for the same problem). But in your trial you should not examine that.

Instead of showing that your intervention is superior to existing therapies you could also test whether your therapy is not unacceptably worse than a therapy already in use (Schumi & Wittes, 2011). Such non-inferiority trials are often done to show that a simpler or cheaper treatment is as good as an existing therapy. However in our case, it is better to avoid these trials because they typically need large sample sizes as well. Furthermore, we do not want to show that our treatment is equivalent to existing therapies, because we already know it is better.

Use the right publication strategy

It is possible that even when you applied all the techniques described in this paper, your trial finds still no significant effects of your therapy. In that case, you can always consider to simply not publish it. You can just wait until a new trial is done that does find positive results. And if you think this is unethical towards the participants and the funder of the trial, you can always tell yourself that so many other researchers do that too, so it must be an acceptable strategy in research. Several meta-analyses of psychological interventions find indirect evidence that the results of up to 25% of the trials are never published (Cuijpers *et al.* 2010*a*; Driessen *et al.* in press). There is also some direct evidence (based on NIH-funded trials) that a considerable number of trials on psychological therapies is never published (Driessen *et al.* in press). Negative findings are an important reason for not publishing trial data, and that is not only true for pharmacological interventions (Turner *et al.* 2008), but also for psychological interventions. And if you decide not to publish, then you can always blame journal editors who are often only interested in positive and significant results.

If you have been talking about the therapy at conferences and you have managed to convince other clinicians that this is such a good and innovative therapy, there is a good chance that some of them will do their own trials. So you can just wait until one of the clinicians you have persuaded about the benefits of this therapy conducts a trial that does find positive outcomes.

Conclusions

It was claimed that most published research findings are false (Ioannidis, 2005) and that up to 85% of biomedical research is wasted (Chalmers & Glasziou,

2009) despite attempts to pin down wrong attitudes of researchers (see, for example, www.alltrials.net) and guidelines on how to conduct randomised trials (Schulz *et al.* 2010). Research on the effects on therapies is no exception to this predicament. Many published research findings were found not to be true when other researchers tried to replicate these finding. When you want to show that your therapy is effective, you can simply wait until a trial is conducted and published that does find positive outcomes. And then you can still claim that your therapy is effective and evidence-based. The possibility that findings from that new trial are not true (but a chance finding) or an overestimation of the true effect is considerable. In fact, corroborating this, a meta-analysis (Flint *et al.* 2014) of psychotherapy for depression showed an excess of significant findings relative to what would have been expected given the average statistical power of the trials included. However, your goal was to show that the therapy is effective, not to find out what it does in reality, because you already knew from the start that the therapy worked and you only needed the trial to convince others.

In this paper, we described how a committed researcher can design a trial with an optimal chance of finding a positive effect of the examined therapy. There is an abundant literature for the interested reader wanting to learn more about conducting randomised trials (Akobeng, 2005; Schulz *et al.* 2010; Higgins & Green, 2011). We saw that a strong allegiance towards the therapy, anything that increases expectations and hope in participants, making use of the weak spots of randomised trials (the randomisation procedure, blinding of assessors, ignoring participants who dropped out, and reporting only significant outcomes, while leaving out non-significant ones), small sample sizes, waiting list control groups (but not comparisons with existing interventions) are all methods that can help to find positive effects of your therapy. And if all this fails you can always not publish the outcomes, and just wait until a positive trial shows what you had known from the beginning: that your therapy is effective anyway, regardless of what the trials say.

For those who think this is all somewhat exaggerated, all of the techniques described here are very common in research on the effects of many therapies for mental disorders.

Acknowledgements

Financial support

Conflict of Interest

Personal author note

The authors would like to stress that they do not intend to disqualify psychotherapy research or researchers in this field, and that they fully respect much of the work that is done in this field, as well as the effort of researchers to improve treatments for patients with mental health problems. They also want to stress that they have been involved in many randomised controlled trials in the field of psychotherapy themselves and have also used some of the methods described in this papers. This paper is only intended to point to practices in the field of psychotherapy that lead to overestimations of the effects of therapies and that are unfortunately far from being exceptions, but rather common practice.

References

Akobeng AK (2005). Understanding randomised controlled trials. *Archives of Disease in Childhood* **90**, 840–844.

Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, Jüni P, Cuijpers P (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *PLoS Medicine* **10**, e1001454.

Boyuan Z, Yang C, Ke C, Xueyong S, Sheng L (2014). Efficacy of acupuncture for psychological symptoms associated with opioid addiction: a systematic review and meta-analysis. *Evidence-Based Complementary and Alternative Medicine: eCAM* **2014**, 313549.

Carey B (2012). Feeling Anxious? Soon There Will Be an App for That. *The New York Times* 13 February.

Chalmers I, Glasziou P (2009). Avoidable waste in the production and reporting of research evidence. *Lancet (London, England)* **374**, 86–89.

Chen P, Furukawa TA, Shinohara K, Honyashiki M, Imai H, Ichikawa K, Caldwell DM, Hunot V, Churchill R (2014). Quantity and quality of psychotherapy trials for depression in the past five decades. *Journal of Affective Disorders* **165**, 190–195.

Colagiuri B, Schenk LA, Kessler MD, Dorsey SG, Colloca L (2015). The placebo effect: from concepts to genes. *Neuroscience* **307**, 171–190.

Coyne J, Kok RN (2014). Salvaging psychotherapy research: a manifesto. *Journal of Evidence-Based Psychotherapies* **14**, 105–124.

Crameri A, von Wyl A, Koemeda M, Schulthess P, Tschuschke V (2015). Sensitivity analysis in multiple imputation in effectiveness studies of psychotherapy. *Frontiers in Psychology* **6**, 1042.

Cuijpers P, van Straten A (2011). New psychotherapies for mood and anxiety disorders: necessary innovation or waste

of resources? *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie* **56**, 251; author reply 251–252.

Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G (2010*a*). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *British Journal of Psychiatry: The Journal of Mental Science* **196**, 173–178.

Cuijpers P, van Straten A, Bohlmeijer E, Hollon SD, Andersson G (2010*b*). The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychological Medicine* **40**, 211–223.

Driessen E, Hollon SD, Bockting CLH, Cuijpers P, Turner EH (2015). Does Publication Bias Inflate the Apparent Efficacy of Psychological Treatment for Major Depressive Disorder? A Systematic Review and Meta-analysis of US National Institutes of Health-Funded Trials. *Plos One*, in press.

Errington-Evans N (2015). Randomised controlled trial on the use of acupuncture in adults with chronic, non-responding anxiety symptoms. *Acupuncture in Medicine: Journal of the British Medical Acupuncture Society* **33**, 98–102.

Fiksdal BL, Houlihan D, Barnes AC (2012). Dolphin-assisted therapy: claims versus evidence. *Autism Research and Treatment* **2012**, 839792.

Flint J, Cuijpers P, Horder J, Koole SL, Munafò MR (2014). Is there an excess of significant findings in published studies of psychotherapy for depression? *Psychological Medicine*, **45**, 439–446.

Frank JD, Frank JB (1991). *Antidepressants Versus Placebo in Major Depression: an Overview*. John Hopkins University Press: Baltimore.

Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, Chen P, Hunot V, Churchill R (2014). Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta psychiatrica Scandinavica* **130**, 181–192.

Gibertini M, Nations KR, Whitaker JA (2012). Obtained effect size as a function of sample size in approved antidepressants: a real-world illustration in support of better trial design. *International Clinical Psychopharmacology* **27**, 100–106.

Higgins JPT, Green S (eds) (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011]. The Cochrane Collaboration. Available from www.cochrane-handbook.org

Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JAC, Cochrane Bias Methods Group, Cochrane Statistical Methods Group (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* **343**, d5928–d5928.

Hollon SD (1999). Allegiance effects in treatment research: a commentary. *Clinical Psychology: Science and Practice* **6**, 107–112.

Ioannidis JPA (1998). Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* **279**, 281–286.

Ioannidis JPA (2005). Why most published research findings are false. *PLoS Medicine* **2**, e124.

Kamioka H, Okada S, Tsutani K, Park H, Okuizumi H, Handa S, Oshio T, Park S-J, Kitayuguchi J, Abe T, Honda T, Mutoh Y (2014*a*). Effectiveness of animal-assisted therapy: a systematic review of randomized controlled trials. *Complementary Therapies in Medicine* **22**, 371–390.

Kamioka H, Tsutani K, Yamada M, Park H, Okuizumi H, Honda T, Okada S, Park S-J, Kitayuguchi J, Abe T, Handa S, Mutoh Y (2014*b*). Effectiveness of horticultural therapy: a systematic review of randomized controlled trials. *Complementary Therapies in Medicine* **22**, 930–943.

Kaptchuk TJ (2001). The double-blind, randomized, placebo-controlled trial: gold standard or golden calf? *Journal of Clinical Epidemiology* **54**, 541–549.

Khan A, Brown WA (2015). Antidepressants versus placebo in major depression: an overview. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*.

Leykin Y, DeRubeis RJ (2009). Allegiance in psychotherapy outcome research: separating association from bias. *Clinical Psychology: Science and Practice* **16**, 54–65.

Luborsky L, Diguer L, Seligman DA, Rosenthal R, Krause ED, Johnson S, Halperin G, Bishop M, Berman JS, Schweizer E (1999). The researcher's own therapy allegiances: a "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice* **6**, 95–106.

Miller S, Wampold B, Varhely K (2008). Direct comparisons of treatment modalities for youth disorders: a meta-analysis. *Psychotherapy Research: Journal of the Society for Psychotherapy Research* **18**, 5–14.

Mohr DC, Spring B, Freedland KE, Beckner V, Arean P, Hollon SD, Ockene J, Kaplan R (2009). The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychotherapy and Psychosomatics* **78**, 275–284.

Mohr DC, Ho J, Hart TL, Baron KG, Berendsen M, Beckner V, Cai X, Cuijpers P, Spring B, Kinsinger SW, Schroder KE, Duffecy J (2014). Control condition design and implementation features in controlled trials: a meta-analysis of trials evaluating psychotherapy for depression. *Translational Behavioral Medicine* **4**, 407–423.

Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, Hollingsworth JC, Ashton CM, Wray NP (2002). A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *New England Journal of Medicine* **347**, 81–88.

Munder T, Gerger H, Trelle S, Barth J (2011). Testing the allegiance bias hypothesis: a meta-analysis. *Psychotherapy Research: Journal of the Society for Psychotherapy Research* **21**, 670–684.

Munder T, Flückiger C, Gerger H, Wampold BE, Barth J (2012). Is the allegiance effect an epiphenomenon of true efficacy differences between treatments? a meta-analysis. *Journal of Counselling Psychology* **59**, 631–637.

Munder T, Brütsch O, Leonhart R, Gerger H, Barth J (2013). Researcher allegiance in psychotherapy outcome research: an overview of reviews. *Clinical Psychology Review* **33**, 501–511.

Nezu AM (1986). Efficacy of a social problem-solving therapy approach for unipolar depression. *Journal of Consulting and Clinical Psychology* **54**, 196–202.

Nezu AM, Nezu (2008). *Evidence-based Outcome Research; A Practical Guide to Conducting Randomized Controlled Trials for Psychosocial Interventions.* Oxford University Press: New York.

Ondo WG, Sethi KD, Kricorian G (2007). Selegiline orally disintegrating tablets in patients with Parkinson disease and 'wearing off' symptoms. *Clinical Neuropharmacology* **30**, 295–300.

Pinniger R, Brown RF, Thorsteinsson EB, McKinley P (2012). Argentine tango dance compared to mindfulness meditation and a waiting-list control: a randomised trial for treating depression. *Complementary Therapies in Medicine* **20**, 377–384.

Preston RA, Materson BJ, Reda DJ, Williams DW (2000). Placebo-associated blood pressure response and adverse effects in the treatment of hypertension: observations from a Department of Veterans Affairs Cooperative Study. *Archives of Internal Medicine* **160**, 1449–1454.

Psychiatry: Therapist-free therapy (2011). *The Economist.* March 5th 2011, downloaded from: http://www.economist.com/node/18276234.

Rafiei R, Ataie M, Ramezani MA, Etemadi A, Ataei B, Nikyar H, Abdoli S (2014). A new acupuncture method for management of irritable bowel syndrome: a randomized double blind clinical trial. *Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences* **19**, 913–917.

Schmidt NB, Richey JA, Buckner JD, Timpano KR (2009). Attention training for generalized social anxiety disorder. *Journal of Abnormal Psychology* **118**, 5–14.

Schulz KF, Altman DG, Moher D, CONSORT Group (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ (Clinical Research)* **340**, c332.

Schumi J, Wittes JT (2011). Through the looking glass: understanding non-inferiority. *Trials* **12**, 106.

Siddiqui O, Hung HMJ, O'Neill R (2009). MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. *Journal of Biopharmaceutical Statistics* **19**, 227–246.

Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine* **358**, 252–260.

Wampold BE (2001). *The Great Psychotherapy Debate: Models, Methods and Findings.* Lawrence Erlbaum Associates: Mahwah, NJ.

Wu J, Yeung AS, Schnyer R, Wang Y, Mischoulon D (2012). Acupuncture for depression: a review of clinical applications. *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie* **57**, 397–405.