Journal of Evidence-Based Psychotherapies,
Vol. 14, No. 2, September 2014, 105-124.

Invited article

SALVAGING PSYCHOTHERAPY RESEARCH: A MANIFESTO

James C. COYNE[*1,2] *& Robin N. KOK*[3,4]

[1]Department of Health Psychology, University of Groningen, University Medical Center, Groningen, the Netherlands
[2]Institute for Health, Health Care Policy and Aging Research, Rutgers, The State University of New Jersey, USA
[3]Department of Clinical Psychology and the EMGO institute for Health and Care Research, Faculty of Psychology and Education, VU University Amsterdam, The Netherlands
[4]Centre for Mental Health Research, The Australian National University, Canberra, Australia

Abstract

Recognition of the unreliability of findings in the biomedical literature, and especially pharmaceutical trials has led to a number of reforms. These include reporting preregistration of protocols for clinical trials and meta-analysis, reporting standards and making data available to others for all clinical trials, as well as recognition of the influence of conflicts of interest. These improvements are only partially and inconsistently reflected in the literature evaluating psychotherapies. The psychotherapy literature is currently of too poor quality to provide a reliable guide to clinicians, consumers, and policymakers. The literature is dominated by underpowered trials with high risk of bias producing positive effects at a statistically improbable rate. Meta-analyses that are poorly conducted with undisclosed conflicts of interest compound these problems. A number of reforms are proposed. These include accelerating adoption of those already occurring in the pharmaceutical literature. Additionally, psychotherapy research should parallel the orderly sequence of treatment development seen in the pharmaceutical literature. Phase III trials providing the effect sizes of treatments should not be conducted until the acceptability of treatment and the feasibility of accruing sufficient numbers of patients are established. The role of investigator allegiance as a potential and potent source of conflict of interest needs to be recognized. Yet, enforcement of existing standards could counter many of the deficiencies of the current literature,

[*] Correspondence concerning this article should be addressed to:
E-mail: jcoynester@gmail.com

but such enforcement may only come with pressures emanating from outside the field of psychotherapy.

Keywords: CONSORT, reporting standards, confirmatory bias, investigator allegiance, meta-analysis, conflict of interest.

In landmark papers that were enlightening to some and infuriating to others, John Ioannidis postulated that most positive findings in biomedical journals are false (Ioannidis, 2005) and that most 'breakthrough' discoveries are either exaggerated or fail to replicate (Ioannidis, 2008). In support, Ben Goldacre (2012) has raised important issues about the integrity and credibility of the literature concerning the efficacy of pharmaceuticals, including selective publication of positive trial data. He documented that:

> *"Drugs are tested by the people who manufacture them, in poorly designed trials, on hopelessly small numbers of weird, unrepresentative patients, and analysed using techniques which are flawed by design, in such a way that they exaggerate the benefits of treatments. Unsurprisingly, these trials tend to produce results that favour the manufacturer. When trials throw up results that companies don't like, they are perfectly entitled to hide them from doctors and patients, so we only ever see a distorted picture of any drug's true effects." (Goldacre, 2012, p. x)*

These revelations have added to widespread skepticism about claims of the efficacy of pharmaceutical treatment as being based on results of poorly designed studies, synthesized in meta-analyses that can only be as reliable as the studies they integrate and that often introduce additional biases. For example, entering results of unpublished studies that were reported to the US Food and Drug Administration (FDA) into meta-analyses with published studies have led to substantially lowered estimates of the efficacy of antidepressants (e.g., Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008).

Considerable efforts at reform are underway, involving;

1. Preregistration of clinical trials so that it is more difficult to hide negative trials or alter analytic plans after results are known (Zarin, Tse, Williams, Califf, & Ide, 2011);
2. Reporting standards that ensure more transparent and detailed papers so that results of trials can be independently validated (CONSORT; Schulz, Altman, & Moher, 2010).
3. Reporting standards (PRISMA; Liberati et al., 2009) and pre-registration of plans for meta-analyses and systematic reviews (PROSPERO; Booth et al., 2011);
4. Making trial data available for independent reanalysis (Peat et al., 2014).

The literature concerning the efficacy of psychotherapy is, in some respects, modelled after the literature concerning the evaluation of pharmaceuticals in randomized controlled trials (RCTs). However, psychotherapy literature often lags in adopting reforms such as recognition of risk of bias,

reporting standards, and preregistration of trials. For instance, in comparison with biomedical journals, CONSORT was adopted later and less consistently by psychological journals. The Association for Psychological Science has yet to adopt CONSORT. An example of lapses in reporting can be found in its journal *Psychological Science*, where a null clinical trial is presented as positive, with no clear labeling as a clinical trial (Kok et al., 2013). Investigator allegiance (Thase, 2006) and conflicts of interest are still given little attention in evaluating results of psychotherapy studies as compared to pharmacological trials. However, as a conflict of interest, investigator allegiance poses the same risk of bias that is more easily and readily recognized in psychopharmacological studies (Khurana, Henderson, Walter, & Martin, 2012; Perlis et al., 2005; Shimazawa & Ikeda, 2014).

Although preregistration of psychotherapy trials is now encouraged – indeed required by some journals – enforcement is lax, with some high profile trials (Morrison et al., 2014) exposed as registered after initiation of collection of data (Coyne & van Linschoten, 2014). Key features of pre-registered design and analyses are simply ignored in the subsequent publication of results (Milette, Roseman, & Thombs, 2011). Results of some trials, which could potentially alter public perceptions of the efficacy of particular therapies, are left unpublished after being registered and completed (Klingberg et al., 2010) and linger in the file drawer. Unfortunately, preregistration of trials is inconsistent, so the full extent of this phenomenon cannot be determined as we have no knowledge of trials that were both unregistered and subsequently left unpublished. Ignoring or altering details of a trial's registration after trial completion negates the purpose of registration, but is given little attention in the process of peer reviewing reports of results for publication.

In this paper, we argue that the psychotherapy literature needs to at least catch up to the pharmaceutical literature in terms of transparency, accountability and reliability of findings. Of course, much remains to be done to ensure the quality and transparency of drug studies and to get all of the data into public view. But psychotherapy researchers would do well to adopt and enforce some of the positive steps that are now seen in the pharmaceutical literature.

Furthermore, psychotherapy research would do well to adhere more to the orderly, progressive treatment development seen in pharmaceutical research: from proof of concept (Phase I) to demonstration of feasibility and acceptability (Phase II) to evaluation in randomized trials (Phase III). Clear identification of phases would allow distinguishing between what particular studies can, and possibly even more important, cannot tell us. Moreover, a systematic weeding out of interventions that are unacceptable to patients or unfeasible trials in Phases 1 and 2 would prevent many ill-conceived trials being interpreted as providing estimates of efficacy, when they are crippled by problems in recruitment and low statistical power.

As it now stands, the psychotherapy literature does not provide a dependable guide to policy makers, clinicians, and consumers attempting to assess the relative costs and benefits of choosing a particular therapy over others. If such stakeholders uncritically depend upon the psychotherapy literature to evaluate the evidence-supported status of treatments, they will be confused or misled.

The dire state of psychotherapy research

In the broader field of science, many RCTs are underpowered, yet consistently obtain positive results by redefining the primary outcomes after results are known (Chan, Hróbjartsson, Haahr, Gøtzsche, & Altman, 2004; Pinto et al., 2013). A psychotherapy trial with 30 or fewer patients in the smallest cell has less than a 50% probability of detecting a moderate size, significant effect, even if it is present (Coyne, Thombs, & Hagedoorn, 2010). Typical psychotherapy RCTs are small, methodologically flawed studies conducted by investigators with strong allegiances to one of the treatments being evaluated (Munder, Brütsch, Leonhart, Gerger, & Barth, 2013). Moreover, a lack of enforcement of trial preregistration allows for consistently positive results by redefining the primary outcomes after results are known, effectively cherry-picking the most positiveoutcomes from a battery of outcome measures.

Examination of the studies mustered for treatments being evidence supported by APA Division 12 (http://www.div12.org/empirically-supported-treatments/) indicates that many studies were of low methodological quality and too underpowered to be reliably counted as evidence of efficacy, yet were included without comment about these problems. Taking an overview, it is striking to see the extent to which the literature continues to depend on small, methodologically flawed RCTs conducted by investigators with strong allegiances to one of the treatments being evaluated. Yet, which treatment is preferred by investigators is a better predictor of the outcome of the trial than the specific treatment being evaluated (Luborsky et al., 2006).

Many positive findings in psychotherapy research are created by spinning outcomes, involving confirmatory bias, flexible rules of design and recruitment, data analysis, selective outcome reporting and significance chasing (Simmons, Nelson, & Simonsohn, 2011), as demonstrated in psychotherapy research on depression (Flint, Cuijpers, Horder, Koole, & Munafò, 2014). Many studies considered positive, including those that become highly cited, are basically null trials for which results for the primary outcome are ignored, and post-hoc analysis of secondary outcomes and subgroup analyses are emphasized (Bach & Hayes, 2002; Dimidjian et al., 2006; Morrison et al., 2014). In a recent re-analysis of individual patient data trials, no fewer than 35% of re-analyses led to conclusions that differed from the findings of the original paper (Ebrahim et al., 2014).

Confirmatory bias and the spinning of weak and null findings generally start in abstracts (Yavchitz et al., 2012). Although CONSORT has specific recommendations for what should be reported in abstracts (Hopewell et al., 2008), these standards are largely ignored in reports of its psychotherapy interventions. A recent review of studies of therapy for couples dealing with cancer found important discrepancies between the positive spin in abstracts and the actual results in the paper (Coyne, 2013; Cristea, Kafescioglu, & Coyne, 2013). It is likely this pattern is much more common, given the lack of attention of editors and reviewers to either CONSORT for abstracts or discrepancies between abstracts and results sections. Indeed, there seems to be a strong link between significant study results, subsequent publication bias and deviations from study protocol (Dwan, Gamble, Williamson, & Kirkham, 2013).

The bulk of psychotherapy RCTs involve comparisons between a single active treatment and an inactive control group such as wait list, no treatment, or "routine care". The latter is typically left undefined but often lacks exposure to any treatment of adequate quality and intensity (Hesser, Weise, Rief, & Andersson, 2011; Posternak & Miller, 2001). In many instances, there may be little difference between routine care and no treatment at all. At best, these studies with inactive control groups can tell us whether a treatment is better than doing nothing at all, or better than the possible nocebo effect of patients expecting treatment because they have enrolled in a trial but not getting it (Furukawa et al., 2014). In sum, the design of a clinical trial and its comparison groups have a substantial impact on effect sizes (Mohr et al., 2009) and this effect should not be ignored, particularly as a source of heterogeneity in any synthesis of the literature.

Meta-silliness?

Hans Eysenck reacted to the first meta-analysis of the effects of psychotherapy (Smith & Glass, 1977) with a dismissal of it as meta-silliness (Eysenck, 1978). Regardless, Smith and Glass' application of meta-analysis showed the potential power of integrating data from multiple independent trials into summary statistics. The trials that were available at that time were quite limited in quality, and their integration in a single meta-analysis was crude by contemporary standards. Nonetheless, the concept of using meta-analysis to integrate and interpret data from multiple clinical trials caught on beyond psychology.

However, psychotherapy researchers have generally lagged behind clinical epidemiology and biomedical meta-analyses in terms of adopting standards for conducting and reporting meta-analyses, as well as methodological and statistical innovations. Meta-analyses of psychotherapy, even in top-tier journals, often lack reproducible literature search strategies (e.g., Wolitzky-Taylor, Horowitz, Powers, & Telch, 2008), do not qualify conclusions by grade of evidence, ignore clinical and statistical heterogeneity, address investigator

allegiance inadequately or not at all (Dragioti, Dimoliatis, & Evangelou, 2014), downplay the domination of the analyses by small trials with statistically improbable rates of positive findings (Rücker, Carpenter, & Schwarzer, 2011) and ignore the extent to which positive effect sizes occur mainly in comparisons between active and inactive treatments.

Meta-analyses of psychotherapies are strongly biased toward concluding that treatments work, especially when conducted by those who have undeclared conflicts of interest and investigator allegiances. This includes developers and promoters of treatments that stand to gain financially or otherwise from their branding as "evidence-supported" (e.g., Hayes, Luoma, Bond, Masuda, & Lillis, 2006; Johnson, Hunsley, Greenberg, & Schindler, 2006; Nowak & Heinrichs, 2008; Sanders, Kirby, Tellegen, & Day, 2014), as well as meta-analyses organized by professional organizations (Hart et al., 2012) intent on demonstrating the availability of evidence supported treatments for dissemination and reimbursement (Coyne, 2012).

Overall, meta-analyses too heavily depend on underpowered, flawed studies conducted by investigators with strong allegiances to a particular treatment or to finding that psychotherapy is in general efficacious. When controls are introduced for risk of bias or investigator allegiance, effects greatly diminish or even disappear (Cristea, Kok, & Cuijpers, in press; Munder et al., 2013; Staines & Cleland, 2007).

Conflicts of interest associated with authors having substantial financial benefits at stake are rarely disclosed in the studies that are reviewed or the meta-analyses themselves. Moreover, claims from these flawed meta-analyses are often backed up by invoking Fail-safe N (Orwin, 1983), or number of negative trials that would have to be published to bring the reported effect size to zero. Although there are a number of serious statistical, conceptual and practical issues with this statistic (Becker, 2005; Heene, 2010), this sometimes staggeringly high number is touted as evidence for the robustness of the effect size, presenting the results to unwitting readers as definitive 'evidence' and discouraging them from looking further. Fail-safe N is often conveniently used even if more sensitive and accurate techniques have been widely available for more than 15 years (e.g., Duval & Tweedie, 2000; Egger, Smith, Schneider, & Minder, 1997). Moreover, if publication bias is not reported, we should not simply assume it is not there – it often is, but it is not reported either erroneously or deliberately (Onishi & Furukawa, 2014). However, both journals and authors are to blame for the publication bias (Malički & Marušić, 2014) and although some solutions have been proposed (Smulders, 2013), these have not met with wide acceptance.

The scam of continuing education credit

Requirements that therapists obtain continuing education credit are intended to protect consumers from outdated, ineffective treatments and to

disseminate up-to-date, evidence-based treatment into the community. But there is inadequate oversight of the scientific quality of exactly what is offered and by whom. Continued education credit is offered for treatments that have unestablished efficacy or for which efficacy is established only in a narrow range of clinical problems and populations. The American Psychological Association (APA) suppresses discussion of the problem by prohibiting groups of members publicly protesting the quality of what is being offered. As a result, APA continues to allow education credits for bogus and unproven treatments like thought field therapy and somatic experiencing.

Providing opportunities for continuing education credit is lucrative for both accrediting bodies and sponsors. In the competitive world of workshops and trainings, entertainment value trumps evidence. Training in delivery of manualized evidence-supported treatments has little appeal when alternative trainings emphasize patient testimonials and dramatic displays of sudden therapeutic gain in carefully edited videotapes, often with actors rather than actual patients. Branding treatments as evidence supported is used to advertise workshops and trainings in which the particular crowd-pleasing interventions that are presented are not evidence supported.

For instance, clinicians attending Acceptance and Commitment (ACT) workshops may see videotapes where the presenter cries with patients. These clinicians should ask themselves: "Entertaining, moving perhaps, but is this an evidence supported technique?". Although claims are made for the superiority of clinical superiority of ACT over other therapies (Levin & Hayes, 2009), a recent meta-analysis reveals that despite its widespread acceptance, the evidence base for ACT is at best weak when compared to established treatments (Öst, 2014). More generally, it is commonplace for psychotherapies with modest support from evidence to be overenthusiastically advocated for conditions even though there is no evidence for their efficacy for that particular condition. What would be disallowed as "off label applications" for pharmaceuticals is routinely accepted in psychotherapy workshops and when under scrutiny, 'evidence-based' often turns out to mean 'evidence-assumed'. Yet the therapists sign off on the attendance list for continuing education credit, go home and the accrediting body gets its pay.

We know we can do better

Psychotherapy research has achieved considerable sophistication in design, analyses, and comprehensive statistical strategies to compensate for missing data such as multiple imputation and moderator analyses to elucidate mechanisms of change. Yet, psychotherapy research lags behind the rigors of pharmaceutical research and clinical epidemiology.

Psychotherapy research already has recommendations and requirements for trial preregistration, a comprehensive checklist for structured reporting of protocols of clinical trials (SPIRIT; Chan et al., 2013), including specification of

primary outcomes; completion of CONSORT checklists to ensure basic details of trials are reported; preregistration of meta-analyses and systematic reviews at sites like PROSPERO, as well as completion of the PRISMA checklist for adequacy of reporting of meta-analyses and systematic reviews (Liberati et al., 2009). Recently, a strong call for promoting replications in psychological science was issued (Koole & Lakens, 2012), but there is little echoing in the psychotherapy literature.

Declarations of conflicts of interest are rare and exposure of authors who routinely failed to disclose conflicts of interest is even rarer. At best, in the unusual cases where someone is found out, the original paper may receive a small, unnoticed erratum buried deep on the journal's website.

Departures from preregistered protocols in published reports of RCTs are common, and there is little checking of discrepancies in abstracts from results that were actually obtained, or promised in preregistration by authors. There is inconsistent and incomplete adherence to these requirements. There is little likelihood that noncompliant authors will be held accountable and a strong incentive to report positive findings in order for a study to be published in a prestigious journal such as the APA's *Journal of Consulting and Clinical Psychology (JCCP)*. Examining the abstracts of papers published in JCCP gives the impression that trials are almost always positive, even when seriously underpowered. A recent review of publication bias in JCCP even declared that psychotherapy research was only marginally affected by selective reporting of positive outcomes (Niemeyer, Musch, & Pietrowsky, 2013); despite recent evidence of excess significance in psychotherapy research (Flint et al., 2014).

Psychotherapy research is conducted and evaluated within a mutual admiration society in which members are careful not to disparage others' results or to embrace standards that they themselves might want relaxed when it comes to evaluation of their own research. There are rivalries between tribes like psychodynamic therapy and cognitive behavior therapy (Leichsenring & Rabung, 2011) that preclude acceptance of common methodological and statistical standards once debate becomes heated (Coyne, Bhar, Pignotti, Tovote, & Beck, 2011). However, within the tribes, there is muted criticism and strenuous efforts to create the appearance that tribe members only 'do what works' (Ioannidis, 2012).

Fewer, better randomized controlled trials

One striking contrast between the psychotherapy literature and the literature concerning the evaluation of pharmaceuticals is the utter dearth of psychotherapy studies explicitly labeled as Phase II studies in the predominance of underpowered, under-resourced Phase III trials. In the development of pharmaceutical research, a distinct Phase II trial is necessary, in which basic data are collected in terms of the acceptability and tolerability of a treatment, including

dosage required to increase acceptability and decrease negative side effects. At this point, the drug is not assumed to have any therapeutic effect whatsoever. Parameters set in Phase II trials are used to guide the design of Phase III trials so that these later studies can be assured of accruing appropriate numbers of patients and retaining them for follow-up. In psychotherapy literature, in contrast, there is frequently a direct move from an interesting concept or variation on an existing treatment to an underpowered efficacy trial that is nonetheless claimed to generate large effect sizes. These trials are often under-resourced and when viewing odd, small numbers of patients in each cell (e.g., why 14 patients rather than 17?), this undoubtedly reflects a convenience sample and an inability to reach a larger sample size with available resources and within an acceptable timeframe (for an example, see Seppälä et al., 2014).

Without preregistration, there is no indication of why the investigators stopped when they did, but such small sample sizes would hardly be justified by any formal power analysis. We believe that this phenomenon reflects investigators simply charging ahead and continuing accrual until resources are exhausted, or worse, a monitoring of incoming data reveals a positive effect and so recruitment is stopped, informed by the results that have been achieved. Formal stopping rules and rules for interim analyses (SPIRIT item 21b, Chan et al., 2013, p. 4) are rarely described in psychotherapy research protocols. Preregistration may take time to be enforced as a requirement for publication, but editors and reviewers can certainly be stopped from accepting small underpowered trials and thereby stop entering spurious effect sizes into the literature. The negative correlation between effect size and sample size and its commensurate overabundance of p-values only just crossing the border of statistical significance cannot be ignored (Kühberger, Fritz, & Scherndl, 2014).

Small trials may be used to generate information about basic parameters in the design of larger trials, but should not be used to generate effect sizes. Indeed, the accepted practice of investigators disguising what are more appropriately Phase II trials as Phase III trials generating statistically improbable effect sizes leads to gross overestimates of the efficacy of particular treatments and discouragement of more adequately resourced and powered trials.

Keeping up pressures for reform from without

Journals and their editors have often resisted changes such as adoption of CONSORT, structured abstracts, and preregistration of trials. The Communications and Publications Board of the American Psychological Association made APA one of the last major holdout publishers to endorse CONSORT and initially provided an escape clause, stating that CONSORT only applied to articles explicitly labeled as a randomized trial (Coyne, Cook, Palmer, & Rusiewicz, 2004), even though implementation of CONSORT has resulted in a measurable improvement in trial reporting elsewhere (Pandis, Shamseer, Kokich,

Fleming, & Moher, 2014; Plint et al., 2006). The board also blocked a push by the Editor of *Health Psychology* for structured abstracts (Hopewell et al., 2008) that reliably report details needed to evaluate what had actually been done in trials and the results were obtained. In both instances, the committee was most concerned about the implications for the major outlet for clinical trials among its journals, the *Journal of Consulting and Clinical Psychology* (2013 impact factor: 5.2).

Although generally not an outlet for psychotherapy trials, the journals of the Association for Psychological Science (APS) show signs of even being worse offenders in terms of ignoring standards and commitment to confirmatory bias. For instance, it takes a reader a great deal of probing to discover that a high-profile paper (Kok et al., 2013) was actually a randomized trial and it needs even further detective work to discover that it was actually a null trial, despite its positive abstract. There is no indication that a CONSORT checklist was ever filed for the study. And despite this, results of the *Psychological Science* article are cited to promote workshops (Mentorcoach, 2014) even though the published article explicitly denies any conflict of interest for the authors (Kok et al., 2013, p. 1131).

There will undoubtedly be a struggle between APS and APA clinical journals for top position in the hierarchy, by publishing only papers that are attention grabbing, even if flawed, while leaving to other journals that are considered less prestigious the publishing of negative and null trials and failed replications. Standards for pre-registering and reporting clinical trials are likely to continue to be sacrificed in the struggle.

If there is to be reform, pressures must come from outside the field of psychotherapy, from professionals without vested interests in promoting particular treatments or the treatments offered by members of professional organizations. Pressures must also come from skeptical external review by consumers and policymakers equipped to understand the games that psychotherapy researchers play in creating the appearance that all treatments work, but the dodo bird is dead.

We need to raise stakeholders' levels of skepticism, disseminate critical appraisal skills widely and provide for their application in evaluating exaggerated claims and methodological flaws in articles published in prestigious, high impact journals. Bad science in the evaluation of psychotherapy must be recognized as the current norm, not an anomaly. Unfortunately, much psychotherapy research is not just bad science – as this is easily spotted by the trained eye – but cargo-cult science: bad science posturing as legitimate science while hiding its shortcomings and inflating its positive results.

We could get far by enforcing rules that we already have

We need to continually expose journals' failures to enforce accepted standards about preregistration, disclosure of conflicts of interest, and

discrepancies between published clinical trials and their preregistration. This is especially the case for meta-analyses, as modern meta-analytical software has made it trivial to produce meta-analyses. With carefully chosen selection criteria, these meta-analyses can cherry-pick primary data to support a certain viewpoint (Egger & Smith, 1998) while at the same time obfuscating the methodological deficiencies of the primary studies. As the primary articles sink into oblivion, all that remains is the inflated effect size reported in the meta-analysis, cleared from all its methodological flaws.

There are too many blatant examples of investigators failing to deliver what they promised in the preregistration, registering after trials have started to accrue patients or openly changing trial protocols after recruitment, and reviewers apparently not ever checking if authors deliver on the primary outcomes and analyses promised in trial registration.

Implementing reform

Editors should:

- Require an explicit statement of whether the trial has been registered and where. SPIRIT for trial protocols should be disseminated and enforced. Trials without prior registration should be rejected outright;
- Insist that reviewers consult trial registration, including modifications made during recruitment, and comment on any deviation; especially with regards to sample size and outcome measures. Deviations from protocol should be explicitly mentioned in the manuscript;
- Explicitly label registration and alterations dated after patient accrual has started.

CONSORT for abstracts and outcome papers should be disseminated and enforced. A lot of hype and misrepresentation in the media starts with authors' own spin in the abstract. Editors should insist that main analyses for the preregistered primary outcome be presented in the abstract and highlighted in any interpretation of results to prevent minor, chance significant findings on secondary outcomes from being hyped.

No more should underpowered exploratory / pilot / feasibility studies be passed off as full-fledged RCTs when they achieve positive results. An orderly sequence of treatment development should occur before conducting what are essentially Phase III randomized trials. Researchers should be well aware of the possibilities and limitations of feasibility and pilot trials (Arain, Campbell, Cooper, & Lancaster, 2010; Leon, Davis, & Kraemer, 2011).

Here, as elsewhere in reforming psychotherapy research, there is something to be learned from drug trials – both in what to do, and what not to do. A process of intervention development ought to include establishing the feasibility (or proof-of-concept) and basic parameters of clinical trials to proceed

as Phase III randomized trials, but cannot be expected to substitute for Phase III trials or to provide effect sizes for the purposes of demonstrating efficacy or comparison to other treatments.

Use of wait list, no treatment, and ill-defined routine care as control groups should be discouraged. No-treatment conditions are unethical if there are well-established treatments and should be rejected outright by any Institutional Review Board or Ethics Committee principally and by journals in the case of a non-reviewed trial. Disconcertingly, these studies still appear to be published (e.g., Jensen & Ramasamy, 2009).

For clinical conditions for which there are well-established treatments, head-to-head comparisons should be conducted, as well as including control groups that might elucidate mechanisms of action. A practical example of the latter would be structured, supportive therapy that controls for attention and positive expectation. These trials should be conducted in a properly powered non-inferiority design with of necessity a large number of participants (D'Agostino, Massaro, & Sullivan, 2003; Nutt, Allgulander, Lecrubier, Peters, & Wittchen, 2008), of course with the appropriate CONSORT-extension for non-inferiority designs followed (Piaggio, Elbourne, Altman, Pocock, & Evans, 2006). Preferably, analyses of outcomes should be conducted blinded by an independent statistician.

There is little to be gained by a further accumulation of small, underpowered studies in which the efficacy of the preferred treatment is assured beforehand by comparing it to a lamed control group that lacks any conceivable element of affective care. The winner is always a foregone conclusion in these trials. There are enough pharmaceutical trials where the new drug is compared to a sub-therapeutic dose of the competing drug – we see easily through this trick in drug trials, why not in psychotherapy?

Evaluations of treatment effects should take into account prior probabilities suggested by the larger literature concerning comparisons between two active, credible treatments – not just previous trials with inflated effect sizes from the same researchers. The well-studied treatment of depression literature suggests some parameters: effect sizes associated with a treatment are greatly reduced when comparisons are restricted to credible, active treatments; with better quality studies; and when controls are introduced for investigator allegiance (Flint et al., 2014).

It is unlikely that initial claims about a breakthrough new treatment exceeding the efficacy of existing treatments will be sustained in larger studies conducted by investigators independent of developers and promoters. However, these initial claims are often eagerly accepted by high-impact journals and serve as a basis for statistically powering subsequent trials and replications. After all, trials are powered based on an expected effect size. As such, an initially hyped treatment with a high effect size will lead later researchers to underpower their studies as they expect to find an effect size roughly equivalent to the initial study.

Such errors in powering can cascade through generations of underpowered trials and label a new treatment as 'promising' for decades where just a few well-powered rigorous trials would spell the end for this 'promising' new treatment. Additionally, it could be argued that to conduct a number of these underpowered, small-scale and inconclusive trials is not only a waste of resources but also unethical to the trial participants, whose time, efforts and inconveniences contributed to the trial are ultimately futile. These underpowered studies can have negative effects on clinical services and social policy (Coyne & Kwakkenbos, 2013) that can be costly from a societal perspective.

Disclosure of conflicts of interest should be enforced and nondisclosure identified in correction statements and further penalized. Apart from reporting of conflicts of interest, reviewers and meta-analysts should consider investigator allegiance when assessing risk of bias (Coyne, 2013a) and should themselves proactively and voluntarily declare any conflicts of interest, financial or otherwise (Viswanathan et al., 2014 propose some useful questions for this issue). Developers of treatments and persons with significant financial gain from a treatment being declared "evidence-supported" should be discouraged from conducting meta-analyses of their own treatments (e.g., Sanders et al., 2014).

Trials should be conducted with sample sizes adequate to detect at least moderate effects based on realistic – not hyped – prior effect sizes. When large positive findings from underpowered studies are published, readers should scrutinize the literature for similarly underpowered trials that achieve similarly positive effects and be wary of any claims made.

Meta-analyses of psychotherapy should incorporate techniques to detect significance chasing in primary studies (Kühberger et al., 2014), such as p-hacking detection techniques (Simonsohn, Nelson, & Simmons, 2014) and tests of excess significance (Ioannidis & Trikalinos, 2007; Ioannidis, 2013) to evaluate the likelihood that patterns of significant findings exceeds likely probability. Publication bias (Duval & Tweedie, 2000; Egger et al., 1997; Sterne & Egger, 2001), significant heterogeneity (Ioannidis, Patsopoulos, & Evangelou, 2007; Thompson & Sharp, 1999) and small-study effects in meta-analyses (Harbord, Egger, & Sterne, 2006) should also be explored and, when present, adequately addressed or controlled for.

Adverse events and harms should routinely be reported (Vaughan, Goldstein, Alikakos, Cohen, & Serby, 2014), including estimates of lost opportunity costs such as failure to obtain more effective treatment.

We need to shift the culture of doing and reporting psychotherapy research. We need to shift from praising exaggerated claims about treatment and faux evidence generated to promote opportunities for therapists and their professional organizations. Instead, it is much more praiseworthy to provide robust, sustainable, reproducible, generalizable (even if more modest) claims and to call out hype and hokum in ways that preserve the credibility of psychotherapy.

The alternative is to continue protecting psychotherapy research from stringent criticism and enforcement of standards for conducting and reporting research. We can simply allow the branding of psychotherapies as "evidence supported" to fall into appropriate disrepute. We have the tools and the knowledge, now we need the consensus, cooperation and persistence.

REFERENCES

Arain, M., Campbell, M. J., Cooper, C. L., & Lancaster, G. A. (2010). What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Medical Research Methodology*, *10*, 67. doi:10.1186/1471-2288-10-67

Bach, P., & Hayes, S. C. (2002). The use of acceptance and commitment therapy to prevent the rehospitalization of psychotic patients: a randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *70*(5), 1129–39. doi:10.1037//0022-006X.70.5.1129

Becker, B. J. (2005). *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. (H. R. Rothstein, A. J. Sutton, & M. Borenstein, Eds.) (pp. 111–125). Wiley.

Booth, A., Clarke, M., Ghersi, D., Moher, D., Petticrew, M., & Stewart, L. (2011). An international registry of systematic-review protocols. *Lancet*, *377*(9760), 108–9. doi:10.1016/S0140-6736(10)60903-8

Chan, A., Tetzlaff, J. M., Gøtzsche, P. C., Altman, D. G., Mann, H., Berlin, J. A., … Moher, D. (2013). SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ (Clinical Research Ed.)*, *346*, e7586. doi:10.1136/bmj.e7586

Chan, A., Hróbjartsson, A., Haahr, M. T., Gøtzsche, P. C., & Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA : The Journal of the American Medical Association*, *291*(20), 2457–65. doi:10.1001/jama.291.20.2457

Coyne, J. C. (2013a). No Title. Retrieved from https://jcoynester.wordpress.com/tag/bill-silverman-prize/

Coyne, J. C. (2013b). Shhh! Keeping quiet about the sad state of couples interventions for cancer patients research. Retrieved September 15, 2014, from http://blogs.plos.org/mindthebrain/2013/04/09/shhh-keeping-quiet-about-the-sad-state-of-couples-interventions-for-cancer-patients-research/

Coyne, J. C., Bhar, S. S., Pignotti, M., Tovote, K. A., & Beck, A. T. (2011). Missed Opportunity to Rectify or Withdraw a Flawed Metaanalysis of Longer-Term Psychodynamic Psychotherapy. *Psychotherapy and Psychosomatics*, *80*(1), 53–54. doi:10.1159/000317537

Coyne, J. C., Cook, J. M., Palmer, S. C., & Rusiewicz, A. (2004). Clarification of clinical trial standards. *Monitor on Psychology*, *35*(11), 4–8.

Coyne, J. C., & Kwakkenbos, L. (2013). Triple P-Positive Parenting programs: the folly of basing social policy on underpowered flawed studies. *BMC Medicine*, *11*(1), 11. doi:10.1186/1741-7015-11-11

Coyne, J. C., Thombs, B. D., & Hagedoorn, M. (2010). Ain't necessarily so: review and critique of recent meta-analyses of behavioral medicine interventions in health

psychology. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, *29*(2), 107–16. doi:10.1037/a0017633

Coyne, J. C., & van Linschoten, H. S. (2014). Cognitive therapy for patients with schizophrenia. *Lancet*, *384*(9941), 400. doi:10.1016/S0140-6736(14)61273-3

Cristea, I. A., Kafesciogly, N., & Coyne, J. C. (2013). The worrisome state of research concerning couples interventions for cancer patients: A systematic review. *Psychology & Health*, *28 Suppl 1*, 2–335. doi:10.1080/08870446.2013.810851

Cristea, I. A., Kok, R. N., & Cuijpers, P. (in press). The efficacy of cognitive bias modification interventions in anxiety and depression: a meta-analysis. *British Journal of Psychiatry*. doi:10.1192/bjp.bp.114.146761

D'Agostino, R. B., Massaro, J. M., & Sullivan, L. M. (2003). Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Statistics in Medicine*, *22*(2), 169–86. doi:10.1002/sim.1425

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., … Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, *74*(4), 658–70. doi:10.1037/0022-006X.74.4.658

Dragioti, E., Dimoliatis, I., & Evangelou, E. (2014). EPA-0834 – Documentation of therapeutic investigator allegiance in meta-analyses of randomised controlled trials. *European Psychiatry*, *29*, 1. doi:10.1016/S0924-9338(14)78172-6

Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, *56*(2), 455–463. doi:10.1111/j.0006-341X.2000.00455.x

Dwan, K., Gamble, C., Williamson, P. R., & Kirkham, J. J. (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PloS One*, *8*(7), e66844. doi:10.1371/journal.pone.0066844

Ebrahim, S., Sohani, Z. N., Montoya, L., Agarwal, A., Thorlund, K., Mills, E. J., & Ioannidis, J. P. (2014). Reanalyses of Randomized Clinical Trial Data. *JAMA*, *312*(10), 1024. doi:10.1001/jama.2014.9646

Egger, M., & Smith, G. D. (1998). Meta-analysis bias in location and selection of studies. *British Medical Journal*, *316*(7124), 61–66. doi:10.1136/bmj.316.7124.61

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. doi:10.1136/bmj.315.7109.629

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist*, *33*(5), 517–517. doi:10.1037/0003-066X.33.5.517.a

Flint, J., Cuijpers, P., Horder, J., Koole, S. L., & Munafò, M. R. (2014). Is there an excess of significant findings in published studies of psychotherapy for depression? *Psychological Medicine*, 1–8. doi:10.1017/S0033291714001421

Furukawa, T. A., Noma, H., Caldwell, D. M., Honyashiki, M., Shinohara, K., Imai, H., … Churchill, R. (2014). Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatrica Scandinavica*, 1–12. doi:10.1111/acps.12275

Goldacre, B. (2012). *Bad Pharma*. London: HarperCollins.

Harbord, R. M., Egger, M., & Sterne, J. A. C. (2006). A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine*, *25*(20), 3443–57. doi:10.1002/sim.2380

Hart, S. L., Hoyt, M. A., Diefenbach, M., Anderson, D. R., Kilbourn, K. M., Craft, L. L., … Stanton, A. L. (2012). Meta-analysis of efficacy of interventions for elevated depressive symptoms in adults diagnosed with cancer. *Journal of the National Cancer Institute*, *104*(13), 990–1004. doi:10.1093/jnci/djs256

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: model, processes and outcomes. *Behaviour Research and Therapy*, *44*(1), 1–25. doi:10.1016/j.brat.2005.06.006

Heene, M. (2010, October 12). A brief history of the Fail Safe Number in Applied Research. Retrieved from http://arxiv.org/abs/1010.2326

Hesser, H., Weise, C., Rief, W., & Andersson, G. (2011). The effect of waiting: A meta-analysis of wait-list control groups in trials for tinnitus distress. *Journal of Psychosomatic Research*, *70*(4), 378–84. doi:10.1016/j.jpsychores.2010.12.006

Hopewell, S., Clarke, M., Moher, D., Wager, E., Middleton, P., Altman, D. G., & Schulz, K. F. (2008). CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Medicine*, *5*(1), e20. doi:10.1371/journal.pmed.0050020

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. doi:10.1371/journal.pmed.0020124

Ioannidis, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology (Cambridge, Mass.)*, *19*(5), 640–8. doi:10.1097/EDE.0b013e31818131e7

Ioannidis, J. P. (2012). Scientific inbreeding and same-team replication: type D personality as an example. *Journal of Psychosomatic Research*, *73*(6), 408–10. doi:10.1016/j.jpsychores.2012.09.014

Ioannidis, J. P. (2013). Clarifications on the application and interpretation of the test for excess significance and its extensions. *Journal of Mathematical Psychology*, *57*(5), 184–187. doi:10.1016/j.jmp.2013.03.002

Ioannidis, J. P., Patsopoulos, N. A., & Evangelou, E. (2007). Uncertainty in heterogeneity estimates in meta-analyses. *BMJ (Clinical Research Ed.)*, *335*(7626), 914–6. doi:10.1136/bmj.39343.408449.80

Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials (London, England)*, *4*(3), 245–53. doi:10.1177/1740774507079441

Jensen, A. M., & Ramasamy, A. (2009). Treating Spider Phobia Using Neuro Emotional Technique™: Findings from a Pilot Study. *The Journal of Alternative and …*, *15*(12), 1363–74. doi:10.1089/acm.2008.0595

Johnson, S. M., Hunsley, J., Greenberg, L., & Schindler, D. (2006). Emotionally Focused Couples Therapy: Status and Challenges. *Clinical Psychology: Science and Practice*, *6*(1), 67–79. doi:10.1093/clipsy.6.1.67

Khurana, G., Henderson, S., Walter, G., & Martin, A. (2012). Conflict of interest and disclosure policies in psychiatry and medicine: a comparative study of peer-reviewed journals. *Academic Psychiatry : The Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, *36*(1), 17–22. doi:10.1176/appi.ap.09120248

Klingberg, S., Wittorf, A., Meisner, C., Wölwer, W., Wiedemann, G., Herrlich, J., … Buchkremer, G. (2010). Cognitive behavioural therapy versus supportive therapy for persistent positive symptoms in psychotic disorders: the POSITIVE Study, a multicenter, prospective, single-blind, randomised controlled clinical trial. *Trials*, *11*, 123. doi:10.1186/1745-6215-11-123

Kok, B. E., Coffey, K. A., Cohn, M. A., Catalino, L. I., Vacharkulksemsuk, T., Algoe, S. B., … Fredrickson, B. L. (2013). How positive emotions build physical health: perceived positive social connections account for the upward spiral between positive emotions and vagal tone. *Psychological Science*, *24*(7), 1123–32. doi:10.1177/0956797612470827

Koole, S. L., & Lakens, D. (2012). Rewarding Replications: A Sure and Simple Way to Improve Psychological Science. *Perspectives on Psychological Science*, *7*(6), 608–614. doi:10.1177/1745691612462586

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication Bias in Psychology: A Diagnosis Based on the Correlation between Effect Size and Sample Size. *PloS One*, *9*(9), e105825. doi:10.1371/journal.pone.0105825

Leichsenring, F., & Rabung, S. (2011). Double standards in psychotherapy research. *Psychotherapy and Psychosomatics*, *80*(1), 48–51; author reply 53–4. doi:10.1159/000315365

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, *45*(5), 626–9. doi:10.1016/j.jpsychires.2010.10.008

Levin, M., & Hayes, S. C. (2009). Is Acceptance and Commitment Therapy Superior to Established Treatment Comparisons? *Psychotherapy and Psychosomatics*, *78*(6), 380–380. doi:10.1159/000235978

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine*, *6*(7), e1000100. doi:10.1371/journal.pmed.1000100

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., … Schweizer, E. (2006). The Researcher's Own Therapy Allegiances: A "Wild Card" in Comparisons of Treatment Efficacy. *Clinical Psychology: Science and Practice*, *6*(1), 95–106. doi:10.1093/clipsy.6.1.95

Malički, M., & Marušić, A. (2014). Is there a solution to publication bias? Researchers call for changes in dissemination of clinical research results. *Journal of Clinical Epidemiology*, *67*(10), 1103–10. doi:10.1016/j.jclinepi.2014.06.002

Mentorcoach, L. (2014). Love 2.0: the master class. Retrieved from http://www.mentorcoach.com/love2/. (Archived by WebCite® at http://www.webcitation.org/6Sbs8qQRb)

Milette, K., Roseman, M., & Thombs, B. D. (2011). Transparency of outcome reporting and trial registration of randomized controlled trials in top psychosomatic and behavioral health journals: A systematic review. *Journal of Psychosomatic Research*, *70*(3), 205–17. doi:10.1016/j.jpsychores.2010.09.015

Mohr, D. C., Spring, B., Freedland, K. E., Beckner, V., Arean, P., Hollon, S. D., … Kaplan, R. (2009). The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychotherapy and Psychosomatics*, *78*(5), 275–84. doi:10.1159/000228248

Morrison, A. P., Turkington, D., Pyle, M., Spencer, H., Brabban, A., Dunn, G., … Hutton, P. (2014). Cognitive therapy for people with schizophrenia spectrum disorders not taking antipsychotic drugs: a single-blind randomised controlled trial. *Lancet*, *383*(9926), 1395–403. doi:10.1016/S0140-6736(13)62246-1

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*(4), 501–11. doi:10.1016/j.cpr.2013.02.002

Niemeyer, H., Musch, J., & Pietrowsky, R. (2013). Publication bias in meta-analyses of the efficacy of psychotherapeutic interventions for depression. *Journal of Consulting and Clinical Psychology*, *81*(1), 58–74. doi:10.1037/a0031152

Nowak, C., & Heinrichs, N. (2008). A comprehensive meta-analysis of Triple P-Positive Parenting Program using hierarchical linear modeling: effectiveness and moderating variables. *Clinical Child and Family Psychology Review*, *11*(3), 114–44. doi:10.1007/s10567-008-0033-0

Nutt, D., Allgulander, C., Lecrubier, Y., Peters, T., & Wittchen, U. (2008). Establishing non-inferiority in treatment trials in psychiatry: guidelines from an Expert Consensus Meeting. *Journal of Psychopharmacology (Oxford, England)*, *22*(4), 409–16. doi:10.1177/0269881108091068

Onishi, A., & Furukawa, T. A. (2014). Publication bias is underreported in systematic reviews published in high-impact-factor journals: metaepidemiologic study. *Journal of Clinical Epidemiology*. doi:10.1016/j.jclinepi.2014.07.002

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*(2), 157–159.

Öst, L.-G. (2014). The efficacy of Acceptance and Commitment Therapy: An updated systematic review and meta-analysis. *Behaviour Research and Therapy*, (2014). doi:10.1016/j.brat.2014.07.018

Pandis, N., Shamseer, L., Kokich, V. G., Fleming, P. S., & Moher, D. (2014). Active implementation strategy of CONSORT adherence by a dental specialty journal improved randomized clinical trial reporting. *Journal of Clinical Epidemiology*, *67*(9), 1044–8. doi:10.1016/j.jclinepi.2014.04.001

Peat, G., Riley, R. D., Croft, P., Morley, K. I., Kyzas, P. A., Moons, K. G. M., … Hemingway, H. (2014). Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. *PLoS Medicine*, *11*(7), e1001671. doi:10.1371/journal.pmed.1001671

Perlis, R. H., Perlis, C. S., Wu, Y., Hwang, C., Joseph, M., & Nierenberg, A. A. (2005). Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. *The American Journal of Psychiatry*, *162*(10), 1957–60. doi:10.1176/appi.ajp.162.10.1957

Piaggio, G., Elbourne, D. R., Altman, D. G., Pocock, S. J., & Evans, S. J. W. (2006). Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA*, *295*(10), 1152. doi:10.1001/jama.295.10.1152

Pinto, R. Z., Elkins, M. R., Moseley, A. M., Sherrington, C., Herbert, R. D., Maher, C. G., … Ferreira, M. L. (2013). Many randomized trials of physical therapy interventions are not adequately registered: a survey of 200 published trials. *Physical Therapy*, *93*(3), 299–309. doi:10.2522/ptj.20120206

Plint, A. C., Moher, D., Morrison, A., Schulz, K., Altman, D. G., Hill, C., & Gaboury, I. (2006). Does the CONSORT checklist improve the quality of reports of

randomised controlled trials? A systematic review. *MEDICAL JOURNAL OF AUSTRALIA*, *185*(5), 263–267.

Posternak, M. A., & Miller, I. (2001). Untreated short-term course of major depression: a meta-analysis of outcomes from studies using wait-list control groups. *Journal of Affective Disorders*, *66*(2-3), 139–46.

Rücker, G., Carpenter, J. R., & Schwarzer, G. (2011). Detecting and adjusting for small-study effects in meta-analysis. *Biometrical Journal. Biometrische Zeitschrift*, *53*(2), 351–68. doi:10.1002/bimj.201000151

Sanders, M. R., Kirby, J. N., Tellegen, C. L., & Day, J. J. (2014). The Triple P-Positive Parenting Program: a systematic review and meta-analysis of a multi-level system of parenting support. *Clinical Psychology Review*, *34*(4), 337–57. doi:10.1016/j.cpr.2014.04.003

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine*, *152*(11), 726–32. doi:10.7326/0003-4819-152-11-201006010-00232

Seppälä, E. M., Nitschke, J. B., Tudorascu, D. L., Hayes, A., Goldstein, M. R., Nguyen, D. T. H., … Davidson, R. J. (2014). Breathing-based meditation decreases posttraumatic stress disorder symptoms in u.s. Military veterans: a randomized controlled longitudinal study. *Journal of Traumatic Stress*, *27*(4), 397–405. doi:10.1002/jts.21936

Shimazawa, R., & Ikeda, M. (2014). Conflicts of interest in psychiatry: strategies to cultivate literacy in daily practice. *Psychiatry and Clinical Neurosciences*, *68*(7), 489–97. doi:10.1111/pcn.12201

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–66. doi:10.1177/0956797611417632

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology. General*, *143*(2), 534–47. doi:10.1037/a0033242

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*(9), 752–760. doi:10.1037/0003-066X.32.9.752

Smulders, Y. M. (2013). A two-step manuscript submission process can reduce publication bias. *Journal of Clinical Epidemiology*, *66*(9), 946–7. doi:10.1016/j.jclinepi.2013.03.023

Staines, G. L., & Cleland, C. M. (2007). Bias in meta-analytic estimates of the absolute efficacy of psychotherapy. *Review of General Psychology*, *11*(4), 329–347. doi:10.1037/1089-2680.11.4.329

Sterne, J. a, & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*(10), 1046–55. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11576817

Thase, M. E. (2006). What Is the Investigator Allegiance Effect and What Should We Do About It? *Clinical Psychology: Science and Practice*, *6*(1), 113–115. doi:10.1093/clipsy.6.1.113

Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, *18*(20), 2693–708. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10521860

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *The New England Journal of Medicine*, *358*(3), 252–60. doi:10.1056/NEJMsa065779

Vaughan, B., Goldstein, M. H., Alikakos, M., Cohen, L. J., & Serby, M. J. (2014). Frequency of reporting of adverse events in randomized controlled trials of psychotherapy vs. psychopharmacotherapy. *Comprehensive Psychiatry*, *55*(4), 849–55. doi:10.1016/j.comppsych.2014.01.001

Viswanathan, M., Carey, T. S., Belinson, S. E., Berliner, E., Chang, S. M., Graham, E., … White, C. M. (2014). A proposed approach may help systematic reviews retain needed expertise while minimizing bias from nonfinancial conflicts of interest. *Journal of Clinical Epidemiology*. doi:10.1016/j.jclinepi.2014.02.023

Wolitzky-Taylor, K. B., Horowitz, J. D., Powers, M. B., & Telch, M. J. (2008). Psychological approaches in the treatment of specific phobias: a meta-analysis. *Clinical Psychology Review*, *28*(6), 1021–37. doi:10.1016/j.cpr.2008.02.007

Yavchitz, A., Boutron, I., Bafeta, A., Marroun, I., Charles, P., Mantz, J., & Ravaud, P. (2012). Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study. *PLoS Medicine*, *9*(9), e1001308. doi:10.1371/journal.pmed.1001308

Zarin, D. A., Tse, T., Williams, R. J., Califf, R. M., & Ide, N. C. (2011). The ClinicalTrials.gov results database--update and key issues. *The New England Journal of Medicine*, *364*(9), 852–60. doi:10.1056/NEJMsa1012065