Addiction
SSA SOCIETY FOR THE STUDY OF ADDICTION
METHODS AND TECHNIQUES
doi:10.1111/add.12799

Compared with what? An analysis of control-group types in Cochrane and Campbell reviews of psychosocial treatment efficacy with substance use disorders

Patrik Karlsson & Anders Bergmark

Department of Social Work, Stockholm University, 106 91 Stockholm, Sweden

ABSTRACT

Background and Aims A crucial, but under-appreciated, aspect in experimental research on psychosocial treatments of substance use disorders concerns what kinds of control groups are used. This paper examines how the distinction between different control-group designs have been handled by the Cochrane and the Campbell Collaborations in their systematic reviews of psychosocial treatments of substance abuse disorders. **Methods** We assessed Cochrane and Campbell reviews ($n = 8$) that were devoted to psychosocial treatments of substance use disorders. We noted what control groups were considered and analysed the extent to which the reviews provided a rationale for chosen comparison conditions. We also analysed whether type of control group in the primary studies influenced how the reviews framed the effects discussed and whether this was related to conclusions drawn. **Results** The reviews covered studies involving widely different control conditions. Overall, little attention was paid to the use of different control groups (e.g. head-to-head comparisons versus untreated controls) and what this implies when interpreting effect sizes. Seven of eight reviews did not provide a rationale for the choice of comparison conditions. **Conclusions** Cochrane and Campbell reviews of the efficacy of psychosocial interventions with substance use disorders seem to underappreciate that the use of different control-group types yields different effect estimates. Most reviews have not distinguished between different control-group designs and therefore have provided a confused picture regarding absolute and relative treatment efficacy. A systematic approach to treating different control-group designs in research reviews is necessary for meaningful estimates of treatment efficacy.

Keywords Active, absolute effects, control groups, inactive, randomized controlled trials, relative effects.

Correspondence to: Patrik Karlsson, Department of Social Work, Stockholm University, 106 91 Stockholm, Sweden. E-mail: patrik.karlsson@socarb.su.se
Submitted 14 January 2014; initial review completed 2 April 2014; final version accepted 28 October 2014

INTRODUCTION

An underappreciated issue in the addiction field concerns the types of control conditions that are used in experimental studies of psychosocial treatments of substance use disorders. An early exception is an article by Finney [1] that directly addresses the problem of variable comparison conditions. He writes: 'Modalities that have weaker competition in their relevant studies are more likely to be found to be effective than those that have been pitted against stronger competition' (p. 1493). Finney suggested a standardization of the control condition by always using motivational enhancement therapy (MET) as a control for psychosocial alcohol interventions. Magill & Longabaugh [2] have recently highlighted the importance of taking seriously the question of what the intervention group is compared with in treatment efficacy studies (see also [3]). Different control conditions imply different effect estimates, and recognizing this is crucial for drawing accurate conclusions concerning treatment efficacy.

The control group issue is also crucial in systematic reviews. Because systematic reviews aim to synthesize the available research of treatment efficacy, how they handle different control-group designs will affect the conclusions drawn. This paper focuses on how the control-group issue is treated by Cochrane Collaboration and Campbell Collaboration (in the following referred to as Cochrane and Campbell) in their systematic reviews of psychosocial treatments of substance use disorders and how this is related to the conclusions drawn regarding treatment efficacy. These two international collaborations are highly influential in guiding clinical choices globally, and Cochrane reviews

are considered to be of higher quality than reviews published through other channels [4–6]. Although our focus on these two review sources limits the generalizability of the study it is, for example, less sensitive to factors outside the review authors' control that may affect reporting (e.g. different accepted word counts in different journals). By focusing exclusively on Cochrane and Campbell we are, to a large extent, able to 'remove' the effects of such external factors and the additional variability in reporting across reviews that could stem from them. The study can therefore be assumed to include a less variable sample compared to if the entire addiction literature was sampled.

Two main research questions are addressed: (1) what types of control groups are included in Cochrane and Campbell systematic reviews of psychosocial treatments of substance use disorders; and (2) is the selection of studies related to the type of control group and is this discussed in relation to the conclusions that are drawn in the reviews?

Systematic reviews and the appeal of randomized controlled trials (RCTs)

Systematic research reviews play a crucial role in summarizing the knowledge of the effects of interventions. Cochrane and Campbell are probably the most well-known organizations devoted to systematic reviews in the medical and social/educational field, respectively. Cochrane, however, also includes non-medical interventions, and several reviews are published both in Cochrane's and Campbell's libraries. The organizations have an explicitly international profile compared to, for example, governmentally commissioned reviews whose readership is typically more local. Consequently, their influence on clinical choices worldwide is potentially enormous. Both value experimental studies highly, even though non-experimental studies are also occasionally considered. Cochrane and Campbell set up rigorous standards for reviews that potentially may put the name 'Cochrane' or 'Campbell' to them. To facilitate the review process, Cochrane has developed a *Handbook* devoted to methodological and other issues [7]. The *Handbook* is also used by Campbell.

As has been said, the RCT is the best study design for detecting causal effects (see [8]). The treatment effect in RCTs is often estimated by contrasting post-treatment or follow-up outcomes for the intervention group with control-group outcomes. Differences in outcomes, given that there is a relatively large number of participants, are attributable to different treatments received in such between-group designs, as the distribution of all potentially confounding factors is balanced across experimental conditions by design [9]. In contrast, the potential of confounding confronts basically any non-experimental study. Modern advances in causal analyses highlight a series of additional analytical challenges to those already facing investigators using observational data. Indeed, attempts to obtain the *ceteris paribus* condition (all other things being equal) in observational studies through covariate adjustment can lead investigators seriously astray, as illustrated by the literature on causal graph theory and its empirical applications [10–13]. Although this does not preclude causal inference from observational studies, it is obviously more difficult than in experimental settings; hence the appeal of RCTs.

The importance of defining the causal contrast

However, there are crucial, often unrecognized, causal issues that the RCT design in itself cannot solve. One of the more important of these concerns the question of what the treatment group is compared with [14]. Estimates of treatment efficacy depend on a contrast being made between experimental conditions, so the definition of the contrast used is fundamental. The quality of this contrast cannot be resolved by only standardizing the intervention through, for example, the use of manuals; it is by contrasting outcomes for the intervention versus the control condition that the effect is estimated. As Holland argues: 'The effect of a cause is *always* relative to another cause… "A causes B" almost always means that A causes B relative to some other cause that includes the condition "not A"' ([15], p. 946).

The dominant approach to causal thinking in empirical research, the Neyman–Rubin model, views causal effects in terms of counterfactual or potential outcomes (see e.g. [16]), and defines the effect as the difference between the observed outcome and the outcome that would have been observed if the exposure status had been different for the units analysed. Thus, in experimental studies, a control group is used to approximate the counterfactual condition for the experimental group. As pointed out by Hernán, although in relation to observational studies, clarity regarding the counterfactual outcome is necessary for meaningful effect estimates: 'ill-defined counterfactuals question the existence of the causal effect itself' ([17], p. 619). Farrington [18], p. 53] underscores this when he says that 'it is important to specify the effect size—compared to what?'. Thus, comparisons of effects across studies with variable control conditions give rise to an apples-and-oranges problem; the counterfactuals are non-comparable. Reviews pooling different control conditions (e.g. waiting-list controls and another 'active' treatment) into a generic control thus provide blurred effect-size estimates and this may also obscure potential causal processes generating the effect (cf. [2]).

Different causal contrasts in active and inactive control-group designs

Following the Cochrane *Handbook* ([7], section 5.3), a main distinction can be drawn between inactive and active

control-group designs. The first provides absolute-effect estimates and the latter provides relative-effect estimates (not to be confused with Holland's terminology) (see [14] on this distinction). While all estimates may, to some extent, be relative or contextual (e.g. providers' therapeutic skills may differ across studies), this distinction is useful for separating two fundamentally different causal contrasts in efficacy research. The *Handbook* ([7[, section 5.3) states that it is important in reviews 'to specify the interventions of interest and the interventions against which these will be compared (comparisons). In particular, are the interventions to be compared with an inactive control intervention (e.g. placebo, no treatment, standard care, or a waiting list control), or with an active control intervention (e.g. a different variant of the same intervention, a different drug, a different kind of therapy)?'.

Studies comparing an intervention group receiving a certain treatment modality with an untreated control group (inactive control) can show whether people benefit from receiving this treatment package compared with not receiving it, i.e. they estimate absolute effects. The counterfactual for the treated is no treatment at all, including an absence of the entire therapeutic context as well as specific therapeutic ingredients. Consequently, this design cannot show whether the effect is due to specific ingredients [e.g. cognitive re-structuring in cognitive–behavioural therapy (CBT)] or to elements common to different treatment modalities (common factors, e.g. empathic therapists) [14]. There is thus a confounding situation in inactive control-group designs; specific ingredients become mixed with common factors. However, common factors may need to be embedded in a coherent treatment model for the treatment being effective [19], so it may be difficult to separate the individual contribution of specific and common factors even in studies comparing active treatments.

In contrast, studies employing active control conditions estimate relative effects, showing whether a certain treatment modality is relatively more effective than another modality [14]. There is substantive evidence that different active substance abuse treatment modalities (bona fide treatments) are generally equally effective [20–23] and this pattern is also established within general psychotherapy research [14]. However, there are some exceptions. For example, behavioural couples therapy has been shown to be more effective than individual therapy for alcohol use disorders [24]

The active–inactive control-group terminology may, however, best be understood as end-points on a continuum; for example, there is a difference in the kind of effect estimates that are obtained when comparing the intervention group with treatment as usual (TAU) and with no-treatment controls [2]. TAU often includes some potentially effective components that may be absent in no-treatment controls (e.g. positive encouragement), so studies using the former as control condition estimate a different causal contrast.

In general, the more effective components that are entailed in the comparison condition the smaller the resulting effect estimates can be assumed to be. Failure to distinguish between different control-group types may therefore lead also to biased conclusions regarding the effect of entire treatment packages, including both specific and common factors. This latter problem can be severe, in that lower-cost treatments that in fact are effective in absolute terms may be discouraged and withheld from the treatment population if their relative effects are not superior to those of other treatments. Although the *Cochrane Handbook* stresses the importance of specifying the control conditions used, it remains to be seen how this issue is handled in the reviews.

METHODOLOGY

Search strategy

Our study includes reviews of psychosocial substance abuse treatment published in the Cochrane and Campbell libraries. Cochrane reviews were identified through reviews listed by the Cochrane Drugs and Alcohol Group and Campbell reviews were identified by reading the titles of all reviews published in the Campbell library. The search was conducted during the first weeks of September 2013.

Inclusion criteria

Any psychosocial treatment of substance use disorders was deemed eligible. Reviews, including meta-analyses and narrative reviews, had to focus on psychosocial treatments, but there were otherwise no inclusion criteria. If there were updates of previous reviews, we included the most recent version. We assessed eligibility criteria by reading through the summaries of all reviews published by the Cochrane Drug and Alcohol Group and by the Campbell Collaboration. Relevant or potentially relevant papers were retrieved in full text.

Coding procedure

We then created a file in which information concerning the reviews was included. This information was organized under the following headings: author(s), title, type(s) of comparisons considered in review and type(s) of comparisons discussed in relation to analysis/conclusions.

We applied a simple three-level coding for the last two headings. For type(s) of comparisons in review, the highest value (2) was assigned to reviews that stated

explicitly which comparison conditions in primary studies were considered in the review. Although all reviews received a code of (2) regarding this heading, code (1) would have been used if the comparison conditions were vague and (0) would have been used if the types of comparisons were not stated.

The rationale for including the last heading was to assess the extent to which the importance of control groups was discussed in the reviews in relation to conclusions drawn. Because the *Cochrane Handbook* specifies that it is important to explicate the kind of comparisons made, many reviews could be expected to include this information, but perhaps without actually discussing its implications. We applied a three-level coding concerning this heading as well. This coding was made concerning both the rationale for considered comparison conditions and regarding the extent to which the review authors discussed the importance of different control-group types when drawing conclusions regarding effects, or lack thereof, of the interventions. With regard to the rationale for chosen comparisons, code (2) meant that the reviews made a well-developed argument for considered comparison conditions and code (1) was assigned to reviews that discussed this only briefly. Reviews received (0) if they did not justify the choice of comparison conditions at all.

Regarding the extent to which the reviews paid attention to different control-group types when presenting conclusions, reviews received code (2) if they discussed this point and its implications and (1) if they mentioned it only briefly, otherwise they received (0). Besides presenting the codes under the last two headings, we have included a short summary of each review in Table 1.

RESULTS

Considered comparison conditions in reviews

We identified eight systematic reviews. As shown in the first column of Table 1, reviews covered both specific psychosocial approaches and psychosocial treatments in general. Each review considered both active and inactive control groups (Table 1, third column). All reviews stated clearly which comparisons they considered.

Type(s) of comparisons discussed in relation to analyses and conclusions

There was mainly little justification for considered control-group types. Besides a short notice in [28], basically no rationale was given for considered control conditions in any of the reviews.

There was little discussion overall of the importance of the control-group issue regarding what kind of conclusions as to effects could be drawn. Of those seven reviews that found studies meeting inclusion criteria (no studies met inclusion criteria in Lui *et al.* [29]), four reviews basically did not address this issue [25,27,31,32]. The reviews by Lindstrøm *et al.* [28] and Smedslund *et al.* [30] were both explicit about what kind of effect estimates can be found in studies with active versus inactive control conditions. Smedslund *et al.* are clear in their 'plain language summary' that motivational interviewing (MI) is beneficial compared with no treatment, but state: 'it seems that other active treatments, treatment as usual and being assessed and receiving feedback can be as effective…' (p. 2). They further argue in the main text (pp. 27–28) that this may be due to MI sharing common factors with other treatment modalities (e.g. the therapeutic alliance). The other five reviews were silent about the distinction between active and inactive controls and the accompanying differences in effect estimates. Compared with assessment of different potential biases in the primary studies, little attention was paid to the implications of using different control groups in all reviews except Lindstrøm *et al.* [28] and Smedslund *et al.* [30]. Hesse *et al.* [26], when discussing planned subgroup analyses, noted shortly that case management may not be more effective than other psychosocial interventions, but they were silent on the implications of this point further in the review.

The review on Alcoholics Anonymous (AA)/12-Step treatment [25], based on primary studies (including Project MATCH) where several studies used active comparisons and none seems to have used untreated control groups (see p. 5), concluded in the abstract that: 'no experimental studies unequivocally demonstrated the effectiveness of AA or TSF [12-Step facilitation] approaches for reducing alcohol dependence or problems' (p. 2). This stand is in contrast to Project MATCH's own conclusion that different therapies for alcohol abuse are equally effective [20]. In Ferri *et al.* [25], the first sentence in the 'Authors' conclusions' claims that 'People considering attending AA or TSF programmes should be made aware that there is a lack of experimental evidence on the effectiveness of such programmes', although it is then noted that 'in the available studies all the interventions appeared to improve at least some of the outcomes considered' (p. 8). This mixture of claims as to the effect of AA/TSF is present in other parts of the report.

There were also some uncertainties as to what controls may entail. The review of psychosocial illicit drug treatments for pregnant women [32] contrasted the treatments with what was generally labelled 'controls' in the main text (it is also said that no treatment modalities were compared), but without further description in the main text of what this may entail. This could give the impression that the studies focused only on absolute

Table 1 Systematic reviews included.

Authors	*Title*	*Type(s) of comparisons considered in review (code)*[a]	*Type(s) of comparisons discussed in relation to analysis/conclusions (codes)*[b]
*Ferri *et al.* [25]	Alcoholics Anonymous and other 12-Step programmes for alcohol dependence	'No treatment', 'other psychological interventions' , 'Twelve-Step programme variants' (p. 4) (2)	No rationale for chosen comparisons (0,0)
*Hesse *et al.* [26]	Case management for persons with substance use disorders	'"Treatment as usual', standard community treatment, other psychosocial interventions or waitlist controls' (p. 4) (2)	No rationale for chosen comparisons Some points regarding control-group types discussed in planned moderator analyses (e.g. that case management may not be more effective than other psychosocial interventions, p. 5) Did statistical analyses to test whether different control groups may imply different effects (p. 6) Excluded studies comparing different versions of CM but without any other control condition (p. 7) (0,1)
*Knapp *et al.* [27]	Psychosocial interventions for cocaine and psychostimulant amphetamine-related disorders	'Other psychosocial treatment', ' pharmacotherapy alone or in combination with psychosocial intervention', 'placebo', 'non-intervention (untreated control groups)' (p. 3) (2)	No rationale for chosen comparisons Some problems regarding the possibility of comparing different intervention groups are noted (p. 10) Some problems with identifying effective ingredients also noted (0,0)
**Lindstrøm *et al.* [28]	Brief strategic family therapy (BSFT) for young people in treatment for non-opioid drug use	'No intervention', 'waitlist controls' 'alternative interventions including treatment as usual (TAU)' (p. 24) (2)	Choice of comparisons justified briefly by an interest in both 'absolute and relative effects' (p. 24) Acknowledges that 'absolute' effects could not be estimated due to no such studies in review [#] (1,1)
*Lui *et al.* [29]	Psychosocial interventions for women enrolled in alcohol treatment during pregnancy	'Other psychosocial treatment, placebo, non-intervention, pharmacological treatment and pharmacological treatment in association with psychosocial treatment' (p. 4) (2)	No rationale for chosen comparisons No studies found that fulfilled inclusion criteria (0, not applicable)
***Smedslund *et al* [30]	Motivational interviewing for substance abuse	'no intervention, waiting list control, placebo psychotherapy or other active therapy' (p. 7) (2)	No rationale for chosen comparisons, but the importance of control group-type discussed extensively in the review (0,2)

(***Continues***)

Table 1. (Continued)

Authors	*Title*	*Type(s) of comparisons considered in review (code)*[a]	*Type(s) of comparisons discussed in relation to analysis/conclusions (codes)*[b]
*Smith *et al* [31]	Therapeutic communities for substance-related disorder	'Pharmacological maintenance treatments, detoxification treatments, psychosocial treatments, placebo or no treatment group and another therapeutic community that differed in duration of treatment or programme or care offered' (p. 3) (2)	No rationale for chosen comparisons No explicit discussion on the importance of type of control group, but it is noted that there is not enough evidence to show that this intervention is more effective than other residential interventions or that there are differences between different versions of therapeutic communities (p. 8) (0,1)
*Terplan & Lui [32]	Psychosocial interventions for pregnant women in out-patient illicit drug treatment programmes compared to other interventions	'Pharmacological intervention or placebo or no intervention or a different psychosocial intervention' (p. 3) (2)	No rationale for chosen comparisons Lack of comparisons with other active treatments in the primary studies noted No definition of what 'controls' may entail in the main text (0,0)

[a]Coding scheme: 2, explicit description of comparison conditions considered (regardless of whether the primary studies actually covered all these comparisons)
[b]Coding scheme first digit, 2: well-developed rationale for considered comparison conditions; 1, short rationale for considered comparison conditions; 0, no rationale for considered comparison conditions. Coding scheme second digit: 2, well-developed discussion of importance of control-group types regarding conclusions, 1: brief discussion of importance of control-group types regarding conclusions, 0: no or virtually no discussion of importance of control-group types regarding conclusions, na: not applicable *Listed in Cochrane library only; **Listed in Campbell library only; ***Listed in both Cochrane and Campbell libraries.
[#]Uncertainties as to whether the control in one of the studies should be regarded as active or passive is also discussed (pp. 37-38)

effects. However, several studies were conducted among patients in methadone maintenance treatment. This indicates that the original calculations largely concerned the potential benefit of including additional psychosocial components to existing services (an additive design). For example, one of the primary studies included — Silverman *et al*. [33], on contingency management (CM)—comprised only participants who already were in a treatment programme designed for pregnant substance abusers and where only patients with ongoing methadone maintenance treatment were eligible for participation. Participants in the study by Elk *et al*. [34], also included in the review [32], were all part of a treatment package entailing, for example, drug counselling, prenatal care and education on nutritional and prenatal issues (but no methadone treatment). Given that obstetrical and neonatal outcomes together with substance use were the primary outcomes in the review, this fact becomes crucial when interpreting the effects.

Hesse *et al*. [26] was the only review that statistically modelled the effect of control-group type (TAU versus active control), but in those cases where it was possible to conduct such moderator analysis (drug use and linkage outcomes) they found no significant associations. They did not provide an interpretation of this result, but this may be due to poor statistical power, as few studies were included in the analysis. However, the lack of this kind of analysis in the other reviews cannot be interpreted as an

example of lack of attention of the control-group issue, because most reviews were narrative.

DISCUSSION

Our conclusion from the examination of how the Cochrane and the Campbell Collaborations treat types of control group in their reviews of psychosocial treatment of substance use disorders is straightforward: little attention is paid to the difference between designs including active and inactive control groups and the differential effect estimates that are obtained from these. Although there were exceptions, the bulk of the reviews mainly ignored this distinction.

Cochrane and Campbell are both central organizations in the context of evidence-based practice. They are organized by a worldwide network of researchers and the initiatives for conducting reviews are taken mainly by researchers, not governments or agencies with vested interests. Their credibility is high and thus their potential to inform clinical choices concerning psychosocial treatments of substance use disorders is substantial. We believe that the underappreciation of the importance of distinguishing between different control-group designs in the reviews is a matter of concern. Although researchers with methodological schooling may be equipped to draw the right conclusions from the reviews, the same may not hold for clinicians.

Recommendations

We advocate a more transparent approach to reporting in systematic reviews regarding absolute and relative effects. This does not, of course, apply only to Cochrane and Campbell, but also holds true for all investigators conducting research reviews in the addiction field. We provide the following recommendations:

- A strict approach to sorting studies based on control conditions is essential in research reviews. Specifically, researchers should avoid categorizing active and inactive controls into a generic control group. Separating these different control conditions is paramount to achieve meaningful estimates of treatment efficacy.
- Investigators should be explicit about what the choice of contrast implies when publishing reviews on the efficacy of psychosocial treatments. Failure do this can lead clinicians and other actors to make flawed interpretations of the evidence, potentially resulting in harmful consequences.
- When using inactive controls, researchers should be careful to state what kind of treatment efficacy estimates can be extracted from this design. While inactive control-group designs can show whether an entire treatment package is effective in absolute terms—an important issue, for example, for users and providers of treatment—investigators should be cautious with attributing treatment effects to specific ingredients. Caution is also needed in making inferences to the relative efficacy of a treatment compared to another treatment based on studies using inactive control-group designs. Of course, superiority of a bona fide treatment compared to no-treatment established in several studies may suggest that this treatment is more effective than another bona fide treatment where there is less support of its absolute effects. However, this does not provide direct evidence for the former treatment being relatively more effective than the latter.
- When using active controls investigators should be careful not to understate the total effect of treatments. It should be emphasized that the lack of differences between experimental conditions in active control-group designs in itself cannot be interpreted as lack of treatment efficacy; this design does not estimate absolute effects.
- In studies testing treatment mechanisms (e.g. using dismantling designs) researchers should ensure that potential differences in outcomes between experimental conditions is not confounded by other factors before attributing the effect to specific ingredients. For example, the effect may be due to different levels of exposure rather than to the component itself.
- When planning primary studies or reviews it is crucial to determine the effects of main interest, i.e. to determine the causal contrast to be made. We believe that reviews (and primary studies) in the protocol stage should specify which effects will be estimated and base the choice of control-group design on this. Researchers should state whether they are interested in absolute effects, relative effects, or both. Procedures pertaining to the sorting of different control groups should be described and justified.
- It might be wise to try to follow the advice from Finney [1] in order to try to standardize the alternative interventions. Such a standardizing effort could also include at least some aspects of TAU such as, for example, length and intensity of such interventions. TAU varies across studies, and there seem to be inconsistencies regarding what kind of control it constitutes. In Ferri *et al.* ([25], p. 44), comparisons with TAU is said to yield relative effects whereas the Cochrane Handbook ([7], section 5.3) treats TAU (standard care) as an inactive control. A standardization of TAU may resolve some inconsistencies. At least, it should be clear in primary studies what TAU entails in order to facilitate an accurate sorting of different control-group designs in reviews.
- To properly estimate relative effects, carefully planned head-to-head designs should be a top priority. In our view, the option that comes as closely as possible to the

double-blind design entails the comparison between two bona fide interventions that are delivered by advocates for those interventions. The latter condition is intended to control for the substantial influence of therapist allegiance on the outcome of psychosocial interventions [35].

- When estimating treatment efficacy, investigators should routinely provide effect size estimates in addition to *P*-values to gauge the magnitude of treatment effects. This should facilitate a more precise understanding of the absolute and relative effects of different treatments.

In summary, there are several reasons for stringent handling and interpretation of different control-group designs. Thus, review authors should take great care when sorting primary studies with different control conditions. This would advance our understanding of absolute and relative treatment efficacy of psychosocial interventions.

Declaration of interests

None.

References

1. Finney J. Limitations in using existing alcohol treatment trials to develop practice guidelines. *Addiction* 2000; **95**: 1491–1500.
2. Magill M., Longabaugh R. Efficacy combined with specified ingredients: a new direction for empirically supported addiction treatment. *Addiction* 2013; **108**: 874–81.
3. Bergmark A. Specific and contextual treatment mechanisms. *Nord Stud Alcohol Dr* 2008; **25**: 277–85.
4. Delaney A., Bagshaw S. M., Ferland A., Laupland K., Manns B., Doig C. The quality of reports of critical care meta-analyses in the Cochrane Database of Systematic Reviews. *Crit Care Med* 2007; **35**: 589–94.
5. Jadad A. R., Cook D. J., Jones A., Klassen T. P., Tugwell P., Moher M. *et al*. Methodology and reports of systematic reviews and meta-analyses. *JAMA* 1998; **280**: 278–80.
6. Moseley A. M., Elkins M. R., Herbert R. D., Maher C. G., Sherrington C. Cochrane reviews used more rigorous methods than non-Cochrane reviews: survey of systematic reviews in physiotherapy. *J Clin Epidemiol* 2009; **62**: 1021–30.
7. Higgins J. P. T., Green S. editors. *Cochrane Handbook for Systematic Reviews of Interventions*, version 5.1.0 (updated March 2011). The Cochrane Collaboration 2011. Available at: www.cochrane-handbook.org (accessed 15 August 2013).
8. Campbell D. T., Stanley J. C. Experimental and Quasi-Experimental Designs For Research. Boston: Houghton Mifflin Company; 1963.
9. Shadish W. R., Cook T. D., Campbell D. T. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin Company; 2002.
10. Greenland S., Pearl J., Robins J. M. Causal diagrams for epidemiological research. *Epidemiology* 1999; **10**: 37–48.
11. Shrier I., Platt R. W. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol* 2008; **8**:70.
12. Elwert F. Graphical causal models. In: Morgan S. L., editor. Handbook of Causal Analysis for Social Research. New York: Springer; 2013, pp 245–73.
13. Glymour M. M., Weuve J., Chen J. T. Methodological challenges in causal research on racial and ethnic patterns of cognitive trajectories: measurement, selection, and bias. *Neuropsychol Rev* 2008; **18**: 194–213.
14. Wampold B. The Great Psychotherapy Debate—Models, Methods and Findings. Mahwah, New Jersey: Lawrence Erlbaum Associates; 2001.
15. Holland P. W. Statistics and causal inference. *J Am Stat Assoc* 1986; **81**: 945–60.
16. Rubin D. Estimating causal effects of treatment in randomized and nonrandomized studies. *J Educ Psychol* 1974; **66**:688–701.
17. Hernan M. A. Invited commentary: hypothetical interventions to define causal effects—afterthought or prerequisite? *Am J Epidemiol* 2005; **162**: 618–20.
18. Farrington D. P. Methodological Quality Standards for Evaluation Research. *ANN Am Acad Polit SS* 2003; **587**: 49–68.
19. Elkins D. N. Toward a common focus in psychotherapy research. *Psychotherapy* 2012; **49**: 450–54.
20. Babor T., Del Boca F. Treatment Matching in Alcoholism. Cambridge: Cambridge University Press; 2003.
21. Babor T. When prophecy fails: Flying saucers, the second coming,and the treatment matching hypothesis. *Nord Stud Alcohol Dr* 2008; **25**: 294–96.
22. Orford J. Asking the right questions in the right way: the need for a shift in research on psychological treatments for addiction. *Addiction* 2008; **103**: 875–85.
23. Imel Z. E., Wampold B. E., Miller S. D., Fleming R. R. Distinctions without a difference: direct comparisons of psychotherapies for alcohol use disorders. *Psychol Addict Behav* 2008; **22**: 533–43.
24. Fals-Stewart W., O'Farell T. J., Feehan M., Birchler G. R., Tiller S., McFarlin S. K. Behavioral couples therapy versus individual-based treatment for male substance-abusing patients. An evaluation of significant individual change and comparison of improvement rates. *J Subst Abuse Treat* 2000; **18**: 249–54.
25. Ferri M., Amato L., Davoli M. Alcoholics anonymous and other 12-step programmes for alcohol dependence. *Cochrane Database of Systematic Reviews* 2006, Issue 3. Art. No.: CD005032.
26. Hesse M., Vanderplasschen W., Rapp R., Broekaert E., Fridell M. Case management for persons with substance use disorder. *Cochrane Database of Systematic Reviews* 2007, Issue 4. Art. No.: CD006265.
27. Knapp W. P., Soares B., Farrell M., Silva de lima M. Psychosocial interventions for cocaine and psychostimulant amphetamines related disorders. *Cochrane Database of Systematic Reviews* 2007, Issue 3. Art. No.: CD003023.
28. Lindstrøm M., Saidj M., Kowalski K., Filges T., Rasmussen P.S., Jørgensen A-M K. Brief Strategic Family Therapy (BSFT) for young people in treatment for non-opioid drug use: a systematic review. *Campbell Syst Rev* 2013 :07.
29. Lui S., Terplan M., Smith E. J. Psychosocial interventions for women enrolled in alcohol treatment during pregnancy. *Cochrane Database of Systematic Reviews* 2008, Issue 3. Art. No.: CD006753.
30. Smedslund G., Berg R. C., Hammerstrøm K. T., Steiro A., Leiknes K. A., Dahl H. M. *et al*. Motivational interviewing for substance abuse. *Cochrane Database of Systematic Reviews* 2011, Issue 5. Art. No.: CD008063.

31. Smith L. A., Gates S., Foxcroft D. Therapeutic communities for substance related disorder. *Cochrane Database of Systematic Reviews* 2006, Issue 1. Art. No.: CD005338.
32. Terplan M., Lui S. Psychosocial interventions for pregnant women in outpatient illicit drug treatment programs compared to other interventions. *Cochrane Database of Systematic Reviews* 2007, Issue 4. Art. No.: CD006037.
33. Silverman K., Svikis D., Robles E., Stitzer M. L., Bigelow G. E. A reinforcement-based therapeutic workplace for the treatment of drug abuse: six-month abstinence outcomes. *Exp Clin Psychopharmacol* 2001; **9**: 14–23.
34. Elk R., Mangus L., Rhoades H., Andres R., Grabowski J. Cessation of cocaine use during pregnancy: effects of contingency management interventions on maintaining abstinence and complying with prenatal care. *Addict Behav* 1998; **23**: 57–64.
35. Luborsky L., Diguer L., Seligman D., Rosenthal R., Karuse E., Johnson S.*et al.* The researcher's own therapy allegiances: A 'wild card' in comparisons of treatment efficacy. *Clin Psychol Sci Pract* 1999; **6**: 95–106.

