Published in final edited form as:
J Clin Epidemiol. 2013 August ; 66(8): 874–882. doi:10.1016/j.jclinepi.2013.03.006.

The Oxford Implementation Index: A new tool for incorporating implementation data into systematic reviews and meta-analyses

P Montgomery[1], **K Underhill**[2], **F Gardner**[1], **D Operario**[3], and **E Mayo-Wilson**[4]

[1]Centre for Evidence-Based Intervention, University of Oxford, Barnett House, 32 Wellington Square, Oxford, UK

[2]Yale Law School / Yale Center for Interdisciplinary Research on AIDS, 127 Wall St, New Haven CT, USA

[3]Behavior and Social Sciences, Public Health Program, Brown University, Providence, Rhode Island, USA

[4]Centre for Outcomes Research and Effectiveness, Department of Clinical, Educational and Health Psychology, University College London, 1-19 Torrington Place, London, UK

Abstract

Objective—This article presents a new tool that helps systematic reviewers extract and compare implementation data across primary trials. Currently, systematic review guidance does not provide guidelines for the identification and extraction of data related to the implementation of underlying interventions.

Study Design and Setting—A team of systematic reviewers used a multi-staged consensus development approach to develop this tool. First, a systematic literature search on the implementation and synthesis of clinical trial evidence was performed. The team then met in a series of subcommittees to develop an initial draft index. Drafts were presented at several research conferences and circulated to methodological experts in various health-related disciplines for feedback. The team systematically recorded, discussed, and incorporated all feedback into further revisions. A penultimate draft was discussed at the 2010 Cochrane-Campbell Collaboration Colloquium to finalise its content.

Results—The Oxford Implementation Index provides a checklist of implementation data to extract from primary trials. Checklist items are organised into four domains: *intervention design*, actual *delivery* by trial practitioners, *uptake* of the intervention by participants, and *contextual factors*. Systematic reviewers piloting the index at the Cochrane-Campbell Colloquium reported that the index was helpful for the identification of implementation data.

Conclusion—The Oxford Implementation Index provides a framework to help reviewers assess implementation data across trials. Reviewers can use this tool to identify implementation data, extract relevant information, and compare features of implementation across primary trials in a

Correspondence to: paul.montgomery@spi.ox.ac.uk.
P Montgomery, Professor of Psycho-Social Intervention; K Underhill, Associate Research Scholar; F Gardner, Professor of Child and Family Psychology; D Operario, Associate Professor of Behavioural and Social Science; E Mayo-Wilson, Senior Research Associate.

Table S1. The Oxford Implementation Index (Full form with instructions as online supplement)

systematic review. The index is a work in progress, and future efforts will focus on refining the index, improving usability, and integrating the index with other guidance on systematic reviewing.

Keywords

systematic review; implementation; heterogeneity; generalisability; randomised controlled trial; reporting guideline

Introduction

Evidence-based practice encourages clinicians to look to systematic reviews of specific interventions, such as those produced and maintained by the Cochrane Collaboration, as the "gold standard" for measuring effects. Rigorous systematic reviews and meta-analyses are designed to minimise bias, to efficiently distil large amounts of information, and to provide information of value to clinicians [1]. However, systematic reviews and meta-analyses of medical or psychosocial interventions present methodological challenges [2]. Reviewers must exercise subjective judgment when deciding whether a statistical meta-analysis will be a reliable summary, and when describing how results may be applied to clinical practice. Reviewers are trained to appraise many sources of variation across trials, particularly characteristics related to the methodological quality of included trials. Systematic review guidance, such as the Cochrane Handbook and PRISMA statement, provide thorough instructions for the identification, extraction, and appraisal of information on methodological quality. These guidance documents, however, place less emphasis on intervention implementation, which can be an important source of variation across primary trials. Implementation encompasses information related to intervention design, delivery, and uptake in primary trials. To date, systematic review guidance has not provided clear direction for reviewers seeking to compare implementation across trials.

Distinguishing between Fidelity and Implementation

Individual trials use various terms for *implementation fidelity*, defined as the degree to which interventions are implemented according to design. These terms include adherence, treatment fidelity, treatment integrity, program integrity, and implementation quality. As used in primary trials, these definitions of fidelity are intended to capture practitioners' and participants' compliance with intervention protocols. The rigorous assessment of implementation fidelity can provide many benefits on the trial level, such as appraising internal validity, enabling replication, and conducting dose-response analyses. Prior analyses have identified a number of elements that comprise implementation fidelity, classifying these elements as core components or important moderators of fidelity. Common among these elements are adherence [3–17] (also called compliance[18]), intervention complexity [17], exposure, or dosage [3, 5, 9–12, 14], which are sometimes classified as part of adherence [17]), quality of delivery [3, 5, 9–12, 14], competence [4, 6–8, 13, 16–18], participant responsiveness [3, 5, 9–12, 14, 17] (also called receipt and enactment [19–22]), program differentiation [3–6, 10, 14, 15, 23], strategies to facilitate implementation [17], program delivery by staff [19–22], staff training [20–22], context [18], and operational definitions of treatment [24–26].

Each of these elements is an important part of identifying the match between the interventions designed and those actually delivered in a randomized trial. But at the level of the systematic review, the concept of fidelity does not capture all the implementation-related data that may be necessary to interpret results across trials. This is because reviewers face several sources of variation that are not present in a primary study, including differences in the design of interventions, the use of different comparison conditions across trials, a wide

variety of intervention contexts, and the susceptibility of primary trial reports to bias [. For example, when reviewers look across many primary trials to draw conclusions, those trials may evaluate interventions that vary in key design features, such as dosage, duration of treatment, frequency of sessions, or sequence of intervention events. If the intervention itself differs from trial to trial, then fidelity alone (i.e., the match between intervention design and intervention delivery) is too narrow a concept to be of help to the systematic reviewer. Instead of relying solely on measures of fidelity, reviewers may wish to extract data on a wider range of intervention characteristics, including characteristics of intervention design, the actual components of intervention delivery, any reported information on participant uptake, and features of the context that may differ across trials. Similarly, reviewers may confront trials that use different control conditions, such as "usual care" that varies depending on the trial setting, which may contribute to varying effect sizes across trials. Moreover, systematic reviews depend on data provided by primary trials that may be differentially susceptible to bias; to use implementation data in the systematic review process, it may be important to identify the source and potential biases affecting the data.

Because existing literature on implementation focuses on the conduct and reporting of primary trials, there is a need for specific guidance to assist systematic reviewers seeking to compare intervention implementation across trials. Our objective was to identify a conceptual framework to assist systematic reviewers in the identification of data related to implementation in primary trials, and to develop a data extraction tool to enable reviewers to compare implementation information across trials. To assist reviewers in the use of the index, we also generated a list of ways in which implementation data may inform the systematic review process.

Methods

The Oxford Implementation Index was developed by a team of systematic reviewers at the Centre for Evidence Based Intervention at Oxford University. Team members are affiliated with the Cochrane and Campbell Collaborations, and independently conduct randomized controlled trials in areas of public health, psychiatry, and social welfare. A steering committee was established, and index development began in 2005.

Literature review

One reviewer undertook a systematic literature search on implementation, adherence, treatment fidelity, and research synthesis, in order to identify prior definitions of treatment fidelity and its associated components. Searches were conducted in MEDLINE, PsycINFO, EMBASE, CINAHL, the Cochrane Methodology Register, and abstracts from the Cochrane Collaboration Colloquium. This search retrieved 4309 references, of which 42 papers provided relevant guidance for identifying implementation characteristics, and 55 papers provided relevant guidance on the design and reporting of systematic reviews. Key papers on implementation included those cited above [3–26], and unpublished work conducted at Oxford by Tamayo [27]. This reviewer also examined guidance documents intended to improve the design, conduct, and reporting of research synthesis, as well as the reporting of primary studies, in order to identify how existing guidance handles implementation data. Initial searches were conducted in 2005–2006. As we revised the index over the subsequent years, we consulted additional guidance documents as they were released until 2012. These guidance documents included the CONSORT Statement [28–30] with extensions [31–39], the PRISMA and QUOROM Statements [40, 41], the TREND Statement [42], the MOOSE Statement [43], the STROBE Statement [44], the RE-AIM Statement [45–49], the Cochrane Handbook [50] and other Cochrane guidelines [51], the Campbell guidelines [52], the CRD guidance for reviews [53], the EPPI guidelines [54], ESRC guidelines for narrative synthesis [55], HTA assessment guidelines [56], the GRADE guidelines [57], the Agency for Health

Care Research and Quality (AHRQ) guidance on assessing risk of bias on individual studies in systematic reviews of health care [58], and other guidance documents [59–70]. Conceptual gaps in recommendations for assessing heterogeneity, comparability, and generalisability in reviews were identified and initially prioritised into six domains (source of implementation information, protocol, and delivery by intervention staff, participant experience, staff characteristics and training, locations/contexts).

Preliminary index development

The team then attended a series of meetings to discuss each domain; all meetings were moderated by PM or KU, and notes recording each meeting were kept. Based on these discussions, we consolidated the domains to four: intervention design, delivery by staff, uptake by participants, and contextual information. Based on pre-existing literature on implementation fidelity, we discussed specific implementation characteristics within each domain, including 27 elements of design (e.g., treatment manuals, program theory), 52 elements of delivery by staff (e.g., staff credentials, dosage delivered), 16 elements of participant uptake (e.g., attendance, materials used), and 78 elements of context (e.g., funding source, legal and policy mandates). Team members introduced, defined, and discussed each of these elements until we reached a unanimous decision on including the element in the index. We eventually consolidated the elements into a draft index. This work was informed by team members' expertise in reviews of HIV prevention interventions, behavioural treatments for anxiety, sleep therapy, and a drug trial for schizophrenia. We concluded meetings by discussing potential uses for implementation data in the design and conduct of the systematic review process.

Index revision

Drafts of the checklist were presented at the 2005 Cochrane Collaboration Colloquium and 2006 Campbell Collaboration Colloquium, and were circulated in February 2006 to methodological experts in the areas of public health, oncology, paediatrics, airways, and the allied professions. Two versions of the checklist were piloted with students enrolled in the Systematic Review class for the Evidence-based Social Intervention masters course at Oxford, and students' feedback was considered in revisions. Reviewers commented that the index identified useful information (e.g., “I am sure this system, if implemented, would improve the generalization of results of systematic reviews," "I found the issues raised by [the checklist] very important, interesting, and not well known… [There is] a need for reporting guidelines to ensure that authors provide description of the intervention intended but also the intervention actually administered."). Reviewers also commented, however, that the data extraction process was time-consuming, that implementation data are often unavailable, and that the relevant elements for assessing implementation may vary across research fields. Using a consensus method similar to that used to develop the original CONSORT Statement, all feedback was systematically recorded, discussed, and incorporated in further revisions. Namely, each concept was discussed in turn: participants considered the relevance, acceptability, and feasibility of inclusion in the Index for each individual item. After participants agreed on items for inclusion in the Index, the team then revised the Index based on participants’ views and transcribed notes from the meeting.

Final index

The revised checklist was then discussed at a meeting during the joint Cochrane-Campbell Collaboration Colloquium held in 2010. By way of content analysis of the transcripts of this meeting and by team discussion, we consolidated our prioritised list of four domains into the finalised index of implementation characteristics, which systematic reviewers can use to extract data from each arm of each primary trial.

Results

Domains of the Oxford Implementation Index

Each implementation domain encompasses a number of implementation characteristics, which may vary across trials assessing similar interventions. The Index highlights many of these characteristics, but reviewers should decide which aspects are most relevant to their topic areas. Broadly, the final domains are *intervention design*, the actual *delivery* by trial clinicians, the *uptake* of the intervention by participants, and *contextual factors*.

1. Intervention Design—To assess the comparability of different interventions, reviewers must consider whether the core components of each intervention are specified clearly, either in trial reports or in cited manuals. This applies to the core components of the control group as well. Even if two interventions accomplish the same task (e.g., aortic valve surgery), they may follow incomparable procedures (e.g., minimally invasive or conventional). Complex interventions may include unspecified core components, such as a nil-by-mouth period preceding surgery; the reviewer must decide which components are likely to influence the comparability of interventions across trials. Depending on the topic of the systematic review, reviewers might consider factors such as intended dosage, setting, materials, and training.

2. Delivery by Trialists and Staff—Intervention delivery often differs from intervention design. Investigating delivery by trial staff is therefore necessary to assess the comparability of interventions *as implemented*. Elements of delivery might include staff qualifications, the quality and use of materials, dosage administered, and efforts to monitor adherence or drift over time. Utilizing data from several sources (e.g., clinicians and independent observers) may increase confidence in the reliability of delivery information. If no adaptations of protocol are mentioned in a report, reviewers should seek confirmation from trialists that none occurred.

3. Participant Uptake—Participants in trials of identical interventions may take up different components to differing extents. ‘Participant uptake’ encompasses aspects of the participant experience, including the core components and dosage actually utilized (e.g., pills taken, counselling sessions attended) and whether participants sought treatment outside the trial protocol. These data can help determine whether participants in different trial arms actually experienced different interventions. Participant uptake may be especially variable in interventions that require behaviour change or self-administered treatment.

4. Context—Context refers to the broad setting in which trials occur, including socioeconomic characteristics, culture, geography, legal environment, and service structures (e.g., managed care or nationalized medicine). Contextual information is essential for clinicians to apply reviews to practice. Tharyan vividly highlighted this need at the 2005 Cochrane Collaboration Colloquium, where he discussed a Cochrane review of tap water for wound cleansing. The review concluded that tap water might be as effective as sterile or sterile saline water for infection and healing; however, a majority of included trials took place in countries with sanitary tap water, which diminished the review’s generalizability. Though they did discuss the availability of sterile tap water, the review authors did not state this limitation clearly. Hence, the results may have been inappropriately applied in settings without high-quality water supplies. Another example arises in the operation of syringe exchange programs for the prevention of HIV infection among persons who inject drugs. The implementation of these programs depends in part on local policing and laws regarding controlled substances and drug paraphernalia [71]. To allow the accurate application of

results, a systematic review of syringe exchange programs would need to incorporate information about the context of included trials.

Using the Oxford Implementation Index

To use the Oxford Implementation Index, a systematic reviewer would first identify the implementation characteristics for each domain that are relevant to the subject matter of the review. The Oxford Implementation Index then provides a systematic data extraction format to identify data from each trial (see Table 1). Once data are extracted, reviewers may analyze them for several purposes, as described below. The usefulness of this index depends on whether implementation information is provided by trialists. Previous reviews suggest that implementation data are often underreported or difficult to collect [3, 19, 21, 23, 72–74], which may undermine the utility of the Index. In these cases, reviewers are encouraged to examine any data that *are* reported, to contact trialists for additional information, and to consider and discuss the implications of missing data.

The Oxford Implementation Index facilitates the extraction of implementation information for use in reviews, but reviewers must draw on clinical expertise to determine which data are relevant and the extent to which data are vulnerable to bias. As the Cochrane Handbook notes [50], “guidelines are not a substitute for good judgement.” When extracting implementation data across trials, a reviewer must evaluate its source (i.e., trialist, clinician, participant, independent observer) for potential bias, since different sources may produce differing reports. For example, blood tests, pill counts, electronic monitoring, patient reports, and clinician reports are known to give different estimates of dosage during drug trials [75–81].

Uses of Implementation Data in Systematic Reviews

During consensus meetings and the solicitation of expert feedback, the team identified multiple opportunities for reviewers to use implementation data to improve the design, conduct, and reporting of a systematic review. Reviewers must decide which of these uses are appropriate.

During the protocol stage, reviewers could specify certain implementation characteristics as inclusion or exclusion criteria for a systematic review; this would require the use of implementation data during the process of screening trials for inclusion. If data on delivery, uptake, and context are unavailable in this circumstance, a reviewer would need to decide whether it is appropriate to appraise eligibility solely on the basis of intervention design.

Reviewers may also decide *a priori* to use implementation data during the analysis process. That is, reviewers could pre-specify strategies for grouping trials for analysis based on aspects of implementation, or they could plan to use key implementation characteristics for the investigation of heterogeneity, mediators, and/or moderators. These analyses have the potential to identify intervention components causing benefit or harm, or contextual factors that may influence effectiveness. Mihalic [11] refers to specific cases where reviewers have examined the influence of implementation on trial findings within meta-analyses. Because implementation may vary most widely in reviews that examine complex or diverse interventions without a single standardised protocol (e.g., behavioral interventions for back pain), these reviews may be most amenable to *a priori* decisions to separate trials based on implementation characteristics. But even when reviews examine well-defined and manualized treatments (e.g., bell-and-pad treatments for nocturnal enuresis), reviewers may plan analyses based on anticipated differences across trials in delivery, participant uptake, or context. If reviewers do not plan these analyses in advance, it may nevertheless be appropriate to conduct exploratory post hoc analyses on the basis of implementation data in

order to generate hypotheses for future research [82–84]. All post-hoc analyses, however, must be clearly identified as such to avoid bias.

Even if reviewers decide to refrain from using implementation data in inclusion criteria or analyses, they may still find the data useful for descriptive purposes. Describing the available implementation data for included trials can provide information of key relevance to review users, such as practitioners seeking to gauge the feasibility of implementing an intervention in their own setting. Identifying differences across trials in implementation can help reviewer's comments on intervention adaptability or core components, particularly when trials achieve similar effect sizes despite variations in implementation.

When communicating the conclusions of a review, data on program implementation can assist the reviewer in appraising the generalizability of findings to new contexts, as well as in acknowledging sources of uncertainty affecting review findings. Examining variations in implementation across trials can aid a reviewer in making concrete suggestions for future research, as well as disseminating results to practitioners seeking the data necessary to select and/or replicate a program of interest. If a manualized intervention underwent adaptation across trials, reviewers should discuss these adaptations. Finally, if implementation data are missing from individual trials, reviewers should note this as a source of uncertainty. Prior research suggests that implementation information is not always reported in primary trials, but integrating what data there is into reviews can highlight deficiencies and should encourage more comprehensive reporting in the future.

The Oxford Implementation Index does not make these decisions on behalf of reviewers; systematic reviewers using the Index will make their own decisions about how implementation data can best contribute to their review. But identifying these potential uses for implementation data may help reviewers consider new possibilities for using this information at every stage during the review process.

Discussion

The Oxford Implementation Index fills an important gap in current guidelines for conducting systematic reviews and meta-analyses. It provides an explicit and systematic method of assessing implementation data across trials, which can help reviewers combine trials appropriately, explain heterogeneity within reviews, and critically appraise the generalisability of reviews to clinical practice. By encouraging reviewers to focus more explicitly and carefully on implementation data, we hope that the Index will increase the applicability of systematic reviews and encourage their uptake in practice and policy. Reviews and meta-analyses have long incurred criticism for aggregating dissimilar trials [85, 86], and reviewers may be better prepared to meet these challenges if they incorporate implementation data explicitly and methodically.

In its current form, the Oxford Implementation Index needs further development and broader consensus. Future work is needed to refine the Index, improve speed and ease of use, consider possibilities for scoring the completeness of reporting, integrate the Index with other guidelines for systematic reviewing, and identify circumstances in which each use of implementation data may be appropriate. Like the CONSORT Statement, the Index will improve through feedback from reviewers, clinicians, and trialists in different fields. Further development is also needed to determine the inter-rater reliability of data extraction using the Index; prior studies on inter-rater reliability for coding the reporting quality of primary trials have found strong agreement, in the range of 78–100% [87–92]. Another important limitation is the underreporting of implementation data by primary trials, which may limit opportunities for reviewers to use the Oxford Implementation Index. This limitation,

however, may diminish as implementation research and reporting gain significance in primary research. In recent years, reporting guidelines for primary studies, such as the CONSORT and STROBE Statements, have been revised or extended to incorporate additional details related to intervention implementation. For example, the CONSORT extension to randomized trials of nonpharmacologic treatments includes a new item entitled "implementation of intervention," which requests "details of the experimental treatment and comparator as they were implemented" [35]. As implementation information becomes more routinely reported for randomized trials and other primary study designs, systematic reviewers will have greater opportunity to use these data in research synthesis. The Index is also immediately useful for identifying the absence of key data on implementation, which may help reviewers appraise the generalizability of results and refine recommendations for future research.

Despite its current limitations, this version of the Index offers benefits to several audiences. For reviewers, a general framework for appraising implementation across trials can improve judgments about inclusion, analysis, and generalizability. For clinicians, the Index can aid in the critical appraisal of reviews and meta-analyses, and it can encourage the production of more clinically relevant reviews. For trialists, the Index can encourage more comprehensive reporting about implementation, and reviewers using the Index may identify implementation components that merit further exploration in primary trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Chalmers I. Trying to do more good than harm in policy and practice: the role of rigorous, transparent, up-to-date, replicable evaluations. Ann Am Acad Polit Soc Sci. 2003; 589(1):22–40.
2. Jackson N, Waters E. for the Guidelines for Systematic Reviews of Health Promotion and Public Health Interventions Taskforce. The challenges of systematically reviewing public health interventions. J Public Health. 2004; 26(3):303–307.
3. Dane A, Schneider B. Program integrity in primary and early secondary prevention: Are implementation effects out of control? Clin Psychol Rev. 1998; 18:23–45. [PubMed: 9455622]
4. Dobson KS, Singer AR. Definitional and practical issues in the assessment of treatment integrity. Clin Psychol Sci Prac. 2005; 12:384–387.
5. Dusenbury L, Brannigan R, Falco M, Hansen W. A review of research on fidelity of implementation: implications for drug abuse in school settings. Health Educ Res. 2003; 18(2):237–256. [PubMed: 12729182]
6. Flannery-Schroeder E. Treatment integrity: implications for training. Clin Psychol Sci Prac. 2005; 12:388–390.
7. Forgatch M, Patterson G, DeGarmo D. Evaluating fidelity: Predictive validity for a measure of competent adherence to the Oregon model of parent management training. Behav Ther. 2004; 36(1): 3–13. [PubMed: 16718302]
8. Huppert J, Bufka L, Barlow D, Gorman J, Shear M, Woods S. Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. J Consult Clin Psych. 2001; 69(5):747–755.
9. Hutchings, J.; Gardner, F.; Lane, E. Making evidence based interventions work. In: Farringdon, D.; Sutton, C.; Uttig, D., editors. Support from the Start: Working with Young Children and their Families to Reduce the Risks of Crime and Antisocial Behavior. London: Department for Education and Skills; 2004. p. 69-81.
10. Lynch, S.; O' Donnell, C., editors. "Fidelity of implementation" in implementation and scale-up research designs: applications from four studies of innovative science curriculum materials and diverse populations; Montreal, Canada. AERA Annual Meeting; 2005.

11. Mihalic S. The importance of implementation fidelity. Blueprints. 2002
12. Mihalic S. Successful program implementation: Lessons from Blueprints. OJJDP Juvenile Justice Bulletin. 2004 Jul.:1–11.
13. Perepletchikova F, Kazdin A. Treatment integrity and therapeutic change: issues and research recommendations. Clin Psychol Sci Prac. 2005; 12:365–383.
14. Ruiz-Primo, M., editor. A multi-method and multi-source approach for studying fidelity of implementation; Montreal, Canada. AERA Annual Meeting; 2005.
15. Schoenwald S, Henggeler S, Brondino M, Rowland M. Multisystemic therapy: Monitoring treatment fidelity. Fam Proc. 2000; 39:83–104.
16. Waltz J, Addis M, Koerner K, Jacobson N. Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. J Consult Clin Psych. 1993; 61(4):620–630.
17. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. Implement Sci. 2007; 2:40. [PubMed: 18053122]
18. Fixsen, DL.; Naoom, SF.; Blase, KA.; Friedman, RM.; Wallace, F. Implementation research: A synthesis of the literature. Tampa, FL: National Implementation Research Network, University of South Florida; 2005. Contract No.: Document Number|
19. Lichstein KL, Riedel BW, Grieve R. Fair tests of clinical trials: A treatment implementation model. Adv Behav Res Ther. 1994; 16:1–29.
20. Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, et al. Enhancing Treatment Fidelity in Health Behavior Change Studies: Best Practices and Recommendations From the NIH Behavior Change Consortium. Health Psychol. 2004; 23(5):443–451. [PubMed: 15367063]
21. Borrelli B, Sepinwall D, Ernst D, Bellg A, Czajkowski S, Breger R, et al. A New Tool to Assess Treatment Fidelity and Evaluation of Treatment Fidelity Across 10 Years of Health Behavior Research. J Consult Clin Psychol. 2005; 73(5):852–860. [PubMed: 16287385]
22. Resnick B, Bellg A, Borrelli B, DeFrancesco C, Breger R, Hecht J, et al. Examples of implementation and evaluation of treatment fidelity in the BCC studies: where we are and where we need to go. Ann Behav Med. 2005; 29(supplement):46–54. [PubMed: 15921489]
23. Moncher F, Prinz R. Treatment fidelity in outcome studies. Clin Psychol Rev. 1991; 11:247–266.
24. Gresham FM, Gansle KA, Noell GH. Treatment integrity in applied behavior analysis with children. J Appl Behav Anal. 1993; 26(2):257–263. [PubMed: 8331022]
25. Gresham F, Gansle K. Treatment integrity of school-based behavioral intervention studies: 1980–1990. School Psych Rev. 1993; 22(2):254–272.
26. Peterson L, Homer A, Wonderlich S. The integrity of independent variables in behavior analysis. J Appl Behav Anal. 1982; 15(4):477–492. [PubMed: 7153187]
27. Tamayo, S. Orange data: Implications of treatment fidelity for systematic reviewing [M.Sc. Thesis in Evidence-Based Social Work]. University of Oxford; 2004.
28. Moher D, Schulz KF, Altman D. for the CONSORT Group. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet. 2001; 357(9263):1191–1194. [PubMed: 11323066]
29. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol. 2010 Aug; 63(8):e1–e37. [PubMed: 20346624]
30. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol. 2010 Aug; 63(8):834–840. [PubMed: 20346629]
31. Piaggio G, Elbourne D, Altman D, Pocock S, Evans S. for the CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT Statement. JAMA. 2006; 294(10):1152–1160. [PubMed: 16522836]
32. Campbell MK, Elbourne DR, Altman DG. for the CONSORT Group. CONSORT statement: Extension to cluster randomised trials. BMJ. 2004; 328(7441):702–708. [PubMed: 15031246]
33. Ioannidis JPA, Evans SJW, Gotszche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: An extension of the CONSORT statement. Ann Intern Med. 2004; 141:781–788. [PubMed: 15545678]

34. Davidson KW, Goldstein M, Kaplan RM, Kaufmann PG, Knatterud GL, Orleans CT, et al. Evidence-based behavioral medicine: What is it and how do we achieve it? Ann Behav Med. 2003; 26(3):161–171. [PubMed: 14644692]
35. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med. 2008 Feb 19; 148(4):295–309. [PubMed: 18283207]
36. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012; 345:e5661. [PubMed: 22951546]
37. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ. 2008; 337:a2390. [PubMed: 19001484]
38. Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C. Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement. Ann Intern Med. 2006 Mar 7; 144(5):364–367. [PubMed: 16520478]
39. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. PLoS Med. 2008 Jan 22.5(1):e20. [PubMed: 18215107]
40. Moher D, Cook D, Eastwood S, Olkin I, Rennie D, Stroup D. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. Lancet. 1999; 354:1896–1900. [PubMed: 10584742]
41. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 2009 Oct; 62(10):e1–e34. [PubMed: 19631507]
42. Des Jarlais D, Lyles C, Crepaz N, Group T. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. Am J Public Health. 2004; 94(3):361–366. [PubMed: 14998794]
43. Stroup D, Berlin J, Morton S, Olkin I, Williamson G, Rennie D, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. JAMA. 2000; 283:2008–2012. [PubMed: 10789670]
44. Altman, D.; Egger, M.; Gotszche, P.; Pocock, S.; Vandenbroucke, J.; Von Elm, E. [cited 2007 17 Jan] for the STROBE Group. STROBE Statement: STrengthening the Reporting of OBservational studies in Epidemiology. STROBE Group. 2007. [updated 2007;]; Available from: www.strobe-statement.org
45. Dzewaltowski DA, Estabrooks PA, Klesges LM, Glasgow RE. TREND: An important step, but not enough. Am J Public Health. 2004; 94(9):1474. [PubMed: 15333294]
46. Glasgow RE. Evaluating the public health impact of health promotion interventions: The RE-AIM framework. Am J Public Health. 1999; 89(9):1322–1327. [PubMed: 10474547]
47. Glasgow RE, Klesges LM, Dzewaltowski DA, Bull SS, Estabrooks PA. The future of health behavior change research: What is needed to improve translation of research into health promotion practice? Ann Behav Med. 2004; 27(1):3–12. [PubMed: 14979858]
48. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice: Rethinking the efficacy-to-evidence transition. Am J Public Health. 2003; 93(8):1261–1267. [PubMed: 12893608]
49. Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: Issues in external validation and translation methodology. Eval Health Prof. 2006; 29:126. [PubMed: 16510882]
50. Higgins, J.; Green, S., editors. Cochrane handbook for systematic reviews of interventions 5.1.0. Chichester, UK: Wiley-Blackwell; 2011. [updated March 2011]
51. Jackson, N.; Waters, E. for the Guidelines for Systematic Reviews of Health Promotion and Public Health Interventions Taskforce. Guidelines for systematic reviews of health promotion and public health interventions. Cochrane Health Promotion and Public Health Field. 2005. [updated 2005; cited]; Available from: http://www.vichealth.vic.gov.au/cochrane/activities/Guidelines%20for%20HPPH%20reviews.pdf

52. Campbell Collaboration. [cited 2006 7 February] Guidelines for the preparation of review protocols. 2001. (Version 1.0). [updated 2001;]; Available from: http://www.campbellcollaboration.org/Fraguidelines.html
53. Centre for Reviews and Dissemination. [cited 2006 7 February] Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for those Carrying Out or Commissioning Reviews. 2001. [updated 2001;]; Available from: http://www.york.ac.uk/inst/crd/report4.htm
54. Evidence for Policy and Practice Information and Co-ordinating Centre. [cited 28 May] Methods. 2007. [updated 2007;]; Available from: http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=89
55. Popay, J.; Roberts, H.; Sowden, A.; Petticrew, M.; Arai, L.; Rodgers, M., et al. Guidance on the conduct of narrative synthesis in systematic reviews. 2006. Version 1: ESRC Methods Programme; Contract No.: Document Number|.
56. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic reviews of trials and other studies. Health Technology Assessment. 1998; 2(19):1–276. [PubMed: 10347832]
57. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol. 2011 Dec; 64(12):1303–1310. [PubMed: 21802903]
58. Viswanathan M, Ansari M, Berkman N, Chang S, Hartling L, McPheeters L, et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews. 2012; 12 (EHC047-EF).
59. MacPherson H, White A, Cummings M, Jobst KA, Rose K, Niemtzow RC. for the STRICTA Group. Standards for Reporting Interventions in Controlled Trials of Acupuncture: The STRICTA Recommendations. J Alt Complement Med. 2002; 8(1):85–89.
60. Higashida RT, Furlan AJ, Roberts H, Tomsick T, Connors B, Barr J, et al. Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke. Stroke. 2003; 34:e109–e137. [PubMed: 12869717]
61. Rowan M, Huston P. Qualitative research articles: Information for authors and peer reviewers. Can Med Assoc J. 1997; 157(10):1442–1446. [PubMed: 9371080]
62. Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. Am J Epidemiol. 2005; 161(3):280–288. [PubMed: 15671260]
63. Medical Research Council. [cited 2006 7 February] A framework for development and evaluation of RCTs for complex interventions to improve health. 2000. [updated 2000;]; Available from: http://www.mrc.ac.uk/pdf-mrc_cpr.pdf
64. Centre EfPaPIaC-o. [cited 28 May] Promoting Health After Sifting the Evidence (PHASE): 12 questions to help you make sense of a process evaluation. 2007. [updated 2007;]; Available from: http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=/hp/reports/phase/phase_process.htm
65. Newman M, Elbourne DR. Guidelines for the REPOrting of primary empirical research Studies in Education (The REPOSE Guidelines): Draft for consultation. EPPI-Centre Social Science Research Unit. 2005 [updated 2005;]; Available from.
66. Cooper H, Hedges L. The handbook of research synthesis. 1994
67. Deeks, J.; Glanville, J.; Sheldon, T. Undertaking systematic reviews of research on effectiveness: CRD guidlines for those carrying out or commissioning reviews. York, UK: Centre for Reviews and Dissemination, York Publishing Services Ltd; 1996. Contract No.: Document Number|
68. Bailey KR. Inter-study differences – how should they influence the interpretation and analysis of results. Stat Med. 1987; 6:351–360. [PubMed: 3616288]
69. Furberg CT, Morgan TM. Lessons from overviews of cardiovascular trials. Stat Med. 1987; 6:295–303. [PubMed: 3616285]
70. Lipsey M, Wilson D. Practical meta-analysis. 2001
71. Beletsky L, Grau LE, White E, Bowman S, Heimer R. The roles of law, client race and program visibility in shaping police interference with the operation of US syringe exchange programs. Addiction. 2011 Feb; 106(2):357–365. [PubMed: 21054615]
72. Andrews DA, Dowden C. Managing correctional treatment for reduced recidivism: A meta-analytic review of programme integrity. Legal Criminol Psychol. 2005; 10(2):173–187.

73. Arai L, Roen K, Roberts H, Popay J. It might work in Oklahoma but will it work in Oakhampton? Context and implementation in the effectiveness literature on domestic smoke detectors. Inj Prev. 2005 Jun; 11(3):148–151. [PubMed: 15933405]
74. Jayaraman S, Rieder M, Matsui D. Compliance assessment in drug trials: has there been improvement in two decades? Can J Clin Pharmacol. 2005; 12(3):e251–e253. [PubMed: 16278498]
75. Bangsberg D, Hecht F, Clague H, Charlebois E, Ciccarone D, Chesney M, et al. Provider assessment of adherence to HIV antiretroviral therapy. JAIDS. 2001; 26(5):435–442. [PubMed: 11391162]
76. Craig H. Accuracy of indirect measures of medication compliance in hypertension. Res Nurs Health. 1985; 8:61–66. [PubMed: 3846318]
77. Lee J, Kusek J, Greene P, Bernhard S, Norris K, Smith D, et al. Assessing medication adherence by pill count and electronic monitoring in the African American Study of Kidney Disease and hypertension (AASK) pilot study. Am J Hypertens. 1996; 9:719–725. [PubMed: 8862216]
78. Matsui D, Hermann C, Klein J, et al. Critical comparison of novel and existing methods of compliance assessment during a clinical trial of an oral iron chelator. J Clin Pharmacol. 1994; 34:944–949. [PubMed: 7983239]
79. Paes A, Bakker A, Soe-Agnie C. Measurement of patient compliance. Pharm World Sci. 1998; 20:73–97. [PubMed: 9584340]
80. Gilbert J, Evans C, Haynes RB, et al. Predicting compliance with a regimen of digoxin therapy in family practice. Can Med Assoc J. 1980; 123:119–122. [PubMed: 7260749]
81. Waterhouse D, Calzone K, Mele C, et al. Adherence to oral tamoxifen: a comparison of patient self-report, pill counts, and microelectric monitoring. J Clin Oncol. 1993; 11:1189–1197. [PubMed: 8501505]
82. Kraemer H, Wilson T, Fairburn C, Agras W. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry. 2002; 59:877–883. [PubMed: 12365874]
83. Thompson S. Systematic review: why sources of heterogeneity in meta-analysis should be investigated. BMJ. 1994; 309:1351–1355. [PubMed: 7866085]
84. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med. 1992 Jan 1;116(1):78–84. [PubMed: 1530753]
85. Eysenck HJ. Systematic reviews: meta-analysis and its problems. BMJ. 1994; 309:789–792. [PubMed: 7950571]
86. Lau J, Ioannidis J, Schmid C. Summing up evidence: one answer is not always enough. Lancet. 1998 Jan 10.351:123–127. [PubMed: 9439507]
87. Han C, Kwak KP, Marks DM, Pae CU, Wu LT, Bhatia KS, et al. The impact of the CONSORT statement on reporting of randomized clinical trials in psychiatry. Contemp Clin Trials. 2009 Mar; 30(2):116–122. [PubMed: 19070681]
88. Ladd BO, McCrady BS, Manuel JK, Campbell W. Improving the quality of reporting alcohol outcome studies: effects of the CONSORT statement. Addict Behav. 2010 Jul; 35(7):660–666. [PubMed: 20207490]
89. Moberg-Mogren E, Nelson DL. Evaluating the quality of reporting occupational therapy randomized controlled trials by expanding the CONSORT criteria. Am J Occup Ther. 2006 Mar-Apr;60(2):226–235. [PubMed: 16596926]
90. Stinson JN, McGrath PJ, Yamada JT. Clinical trials in the Journal of Pediatric Psychology: applying the CONSORT statement. J Pediatr Psychol. 2003 Apr-May;28(3):159–167. [PubMed: 12654939]
91. Naleppa MJ, Cagle JG. Treatment fidelity in social work intervention research: A review of published studies. Res Soc Work Pract. 2003; 20:674–681.
92. Perry AE, Weisburd D, Hewitt C. Are criminologists describing randomized controlled trials in ways that allow us to assess them? Findings from a sample of crime and justice trials. J Exp Criminol. 2010; 6:245–262.

What's New?

- **Key findings**
- The Oxford Implementation Index is a new tool that helps systematic reviewers extract, appraise, and use data from primary trials describing how the intervention was implemented.
- **What this adds to what is known**
- Existing guidance for systematic reviews acknowledges the potential for variation across trials in the design and implementation of interventions; guidance documents, however, do not assist reviewers in examining these data. The Oxford Implementation Index provides a conceptual framework and data extraction tool to fill this gap.
- **What is the implication, what should change now**
- Reviewers should be alert to the possibility that included studies are in fact comparing highly disparate interventions, either due to differences in intervention design, delivery, participant uptake, and/or context.
- Without systematically assessing data on implementation in included studies, reviewers run the risk of drawing misleading conclusions across studies, overlooking important limitations, or misjudging the acceptability of review results to practice. The Oxford Implementation Index provides explicit guidance to assist reviewers in this process.

Table 1

The Oxford Implementation Index – Short Form

Intervention characteristics	**Design (page ref)**		**Delivery (page ref)**		**Participant Uptake (page ref)**	
	Active	***Control***	***Active***	***Control***	***Active***	***Control***
Core components, sequence of intervention components						
Proscribed or incompatible activities						
Technology, materials, and technical requirements/support staff						
Dosage: number, frequency, and duration of sessions						
Delivery method (e.g., phone, group, injection)						
Staff characteristics, training, and supervision						
Non-specific intervention components (e.g., therapeutic alliance)						
Contact/meetings among staff, trialists, and program developers						
Steps to promote staff and participant compliance						
Intervention adaptation by trialists and staff						
Other differences between and/or within trial arms						
Contamination and/or uptake of treatments outside the trial context						

Contextual factors	**Design (page ref)**	**Actual (page ref)**
Setting(s), geographic location(s), and date/time		
Participant characteristics		
Characteristics of the delivering organisation and wider service environment		
Unique ethical considerations		
Significant external events occurring at the time of intervention		