Abstract

Objectives: This article presents a new tool that helps systematic reviewers to 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 the underlying interventions.

Study Design and Setting: A team of systematic reviewers used a multistaged 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 finalize its content.

Results: The Oxford Implementation Index provides a checklist of implementation data to extract from primary trials. Checklist items are organized 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 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. © 2013 Elsevier Inc. All rights reserved.

Keywords: Systematic review; Implementation; Heterogeneity; Generalizability; Randomized controlled trial; Reporting guideline

Conflict of interest: The Oxford Implementation Index was developed by a team of systematic reviewers at the Centre for Evidence-Based Intervention at Oxford University. The 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. P.M. is the author of over a dozen Cochrane Reviews and was, until recently, cochair of the Campbell Collaboration Social Welfare Group. He led this team and managed the project. P.M., E.M.-W., D.O., F.G., and K.U. contributed to the conception of the index and initial project planning. A systematic literature search was developed and conducted by K.U. and P.M. The index was developed by all the authors, led by P.M. and K.U. Piloting was led by P.M. and facilitated by F.G., E.M.-W., D.O., and K.U. K.U., E.M.-W., and P.M. initially evaluated the index using their own reviews. The Delphi panel was run by P.M. and K.U.; all the authors discussed feedback and, led by P.M., refined the index. All the authors contributed to the writing and revision of the article, led by K.U., E.M.-W., and P.M. The index was the focus of a doctoral dissertation by K.U.

Funding: We acknowledge the support of the Oxford Department of Social Policy and Intervention and the Centre for Evidence-Based Intervention. This research was supported in part by infrastructure and resources provided by the Brown University Alcohol Research Center on HIV/AIDS (NIH/NIAAA P01AA019072) and the Lifespan/Tufts/Brown Center for AIDS Research (NIH/NIAID P30AI042853). Ethical approval: This article was based entirely on secondary research. Ethical approval was not required. Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coiDisclosure.pdf and declared no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years; no other relationships or activities that could appear to have influenced the submitted work.

* Corresponding author. Tel.: 44 1865 270325; fax: 44 1865 270324. E-mail address: paul.montgomery@psych.ox.ac.uk (P. Montgomery).
1. Introduction

Evidence-based practice encourages clinicians to look at 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 minimize bias, efficiently distil large amounts of information, and 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 (Preferred Reporting Items for Systematic Reviews and Meta-analyses) 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.

2. Distinguishing between fidelity and implementation

Individual trials use various terms for implementation fidelity, defined as the degree to which interventions are implemented according to the 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 the 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 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.

3. 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 the index development began in 2005.

3.1. Literature review

One reviewer undertook a systematic literature search on implementation, adherence, treatment fidelity, and research synthesis, 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 4,309 references, of which 42 articles provided relevant guidance for identifying implementation characteristics and 55 articles provided relevant guidance on the design and reporting of systematic reviews. Key articles on implementation included those cited previously [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, and the reporting of primary studies, to identify how the existing guidance handles implementation data.

Initial searches were conducted in 2005–06. 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 (Consolidated Standards of Reporting Trials) Statement [28–30] with extensions [31–39], the PRISMA and QUOROM (Quality of Reporting of Meta-analyses) Statements [40,41], the TREND (Transparent Reporting of Evaluations with Nonrandomized Designs) Statement [42], the MOOSE (Meta-analysis of Observational Studies in Epidemiology) Statement [43], the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) Statement [44], the RE-AIM (Reach Effectiveness Adoption Implementation Maintenance) Statement [45–49], the Cochrane Handbook [50] and other Cochrane guidelines [51], the Campbell guidelines [52], the CRD (Centre for Reviews and Dissemination) guidance for reviews [53], the EPI (Evidence for Policy and Practice Information) guidelines [54], ESRC (Economic and Social Research Council) guidelines for narrative synthesis [55], HTA (Health Technology Assessment) assessment guidelines [56], the GRADE (Grades of Recommendation, Assessment, Development and Evaluation) guidelines [57], the Agency for Health Care Research and Quality 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 generalizability in reviews were identified and initially prioritized into six domains (source of implementation information, protocol, delivery by intervention staff, participant experience, staff characteristics and training, and locations/contexts).

3.2. Preliminary index development

The team then attended a series of meetings to discuss each domain; all the meetings were moderated by P.M. or K.U., 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 preexisting literature on implementation fidelity, we discussed specific implementation characteristics within each domain, including 27 elements of design (e.g., treatment manuals and program theory), 52 elements of delivery by staff (e.g., staff credentials and dosage delivered), 16 elements of participant uptake (e.g., attendance and materials used), and 78 elements of context (e.g., funding source and 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, behavioral 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.

3.3. 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, pediatrics, airways, and the allied professions. Two versions of the checklist were pilot-tested 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, implementation data are often unavailable, and 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.

3.4. 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 prioritized list of four domains into the finalized index of implementation characteristics, which systematic reviewers can use to extract data from each arm of each primary trial.

4. Results

4.1. 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.

4.1.1. Intervention design

To assess the comparability of different interventions, reviewers must consider whether the core components of each intervention are specified clearly, in either trial reports or 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.

4.1.2. 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. Using 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.

4.1.3. 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 used (e.g., pills taken and counseling sessions attended) and whether participants sought treatment outside the trial protocol. These data can help to determine whether participants in different trial arms actually experienced different interventions. Participant uptake may be especially variable in interventions that require behavior change or self-administered treatment.

4.1.4. 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, in which 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. Although 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.

4.2. 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 (Table 1; also see the Appendix [available on the journal’s website at www.jclinepi.com]). Once data are extracted, reviewers may analyze them for several purposes, as described in the following. 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, contact trialists for additional information, and 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 judgment.” When extracting implementation data across trials, a reviewer must evaluate its source (i.e., trialist, clinician, participant, and independent observer) for potential bias because 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].

4.3. 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

<table>
<thead>
<tr>
<th>Table 1. The Oxford Implementation Index—short form</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention characteristics</strong></td>
</tr>
<tr>
<td><strong>Active</strong></td>
</tr>
<tr>
<td>Core components, sequence of intervention components</td>
</tr>
<tr>
<td>Proscribed or incompatible activities</td>
</tr>
<tr>
<td>Technology, materials, and technical requirements/support staff</td>
</tr>
<tr>
<td>Dosage: number, frequency, and duration of sessions</td>
</tr>
<tr>
<td>Delivery method (e.g., phone, group, injection)</td>
</tr>
<tr>
<td>Staff characteristics, training, and supervision</td>
</tr>
<tr>
<td>Non-specific intervention components (e.g., therapeutic alliance)</td>
</tr>
<tr>
<td>Contact/meetings among staff, trialists, and program developers</td>
</tr>
<tr>
<td>Steps to promote staff and participant compliance</td>
</tr>
<tr>
<td>Intervention adaptation by trialists and staff</td>
</tr>
<tr>
<td>Other differences between and/or within trial arms</td>
</tr>
<tr>
<td>Contamination and/or uptake of treatments outside the trial context</td>
</tr>
</tbody>
</table>

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

878

could prespecify 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 in which 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 standardized 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 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 and 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 and 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 are 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 to consider new possibilities for using this information at every stage during the review process.

5. 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 to combine trials appropriately, explain heterogeneity within reviews, and critically appraise the generalizability 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 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 interrater reliability of data extraction using the index; prior studies on interrater 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 to
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 triallists, 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.

Acknowledgments

The authors thank the late Leonard Gibbs, Janet Harris, and graduate students in the Oxford Evidence-Based Intervention course for helping to refine the index in its early stages. They are grateful to Isabelle Boutron, Jane Dennis, Janet Harris, Alan Kazdin, Steve Milan, Francheska Perepletchikova, Prathap Tharyan, and Denise Thomson for their comments on this article. They also would particularly like to thank Sean Grant for his help in the latter stages of preparing this article.

Appendix

Supplementary data/material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jclinepi.2013.03.006.

References
