Cochrane

CEU report

Review quality & methods

Toby Lasserson

Trusted evidence.
Informed decisions.
Better health.

Table of contents

1 Changes to CEU review screening: paper submitted to the Cochrane Governing Board

Prepared by:	Toby Lasserson; sponsored by David Tovey
Date:	16 March 2017
Operations or project?	Project
Strategy to 2020 target:	Yes
Status:	For decision

Executive summary

The paper outlines proposed changes to review screening, building on work carried out on the recent of review abstracts and commits to the development of a quality assurance checklist. It puts three proposed actions to management of review screening and briefly describes how the process will work on an operational basis.

We request that the Governing Board ratifies the proposed changes to the screening programme.

Background

This report outlines proposals to monitor and manage review quality on an ongoing basis. This takes account of our current approach to review screening, the recent abstract audit, and plans to pilot the 'fast track editorial service' and the separation of editorial from developmental functions by Cochrane Review Groups.

Pre-publication screening of reviews has evolved since it began in 2013. The work of the 'Screening Team' is valued by many in Cochrane who request input on reviews (three reviews per week in 2016 and about four per week in 2017 to date). In addition to these reviews that are referred to the Screening Team, we also assess all reviews selected for press releasing, reviews referred from the copy edit support service, and the Cochrane UK's ARGO meeting.

As we acknowledged in the CEU quality report for the Seoul Colloquium, the supportive nature of this approach is restricted to reviews that are unlikely to represent the average, making it challenging for us to monitor the quality of the "average" review. To identify the best way forward we conducted the abstract audit using a 'publication checklist'. The results of the abstract audit were informative but also demonstrated the limitations of the tool. This led us to discuss a triage of all reviews using a modified version of the checklist. We intend to triage reviews as they are signed off by CRGs before making further decisions about whether to check the review more fully. By providing a more structured approach we hope to make the checks more transparent and replicable at an earlier stage of the sign off process.

Proposal:

The CEU screening team will undertake the following:

Preserve current referral system

The referral system will continue in its current form, allowing CRGs, Copy Edit Support, and the Communications and External Affairs Department (CEAD) to seek the team's assistance when necessary. We will develop a formal referral mechanism so that we can record a clear rationale for each review that is referred, by whom and at what stage the review is currently at. Where we can feedback verbally to the

CRG teams we will seek to do this, as well as offering in time screening where this can be organised and resources permit.

Sample from signed off reviews

New or updated intervention reviews signed off on a weekly will be selected and assessed against a checklist that builds on the checklist from 2016. The tool we will use aims to triage reviews based on the abstracts, content of the Summary of Findings tables and results for key analyses. The current version of the tool is presented in Appendix 1. For purposes of equity we will sample from reviews signed off by all CRGs. The proposed process is outlined in Appendix 2.

The checklist is intended to be transparent and our piloting of the checklist indicate that it can take less than one hour to use for any given review. After this point, it should be possible to tell when a review may require a closer look from a screening editor or by the CRG. We are aware of variation in practices around the use of workflows around the sign off process and we intend to respect the way in which they are used by the CRGs. The tool is not intended to generate an aggregate score.

Develop and finalise a Review Quality Assurance Checklist

The screening process has considered several reviews against a subset of the MECIR standards. We intend to apply the same methods as we have been using up until now, but develop a QA Checklist that would be applied to:

1. Reviews sent to us by the referral process

2. Reviews identified by the Triage Tool as requiring closer attention

We plan to develop guidance that explains the deployment of the Triage Tool and the QA Checklist.

Measures of success:

We aim to oversee cyclical audits of published review abstracts, Summary of Findings, and main analyses to provide CRG specific and community wide comparative data on abstract quality over time. This will tell us how much an effect the tool has had over time.

Issues and strategic implications:

Strategy Implications:

This relates directly to Goal 1. Planned changes from the Structure and Function transformation programme could impact on the screening process. As networks form we expect there to be a structural alignment of the CEU editors to accommodate this.

Resource implications:

Triaging and quality assurance work will be carried out by the team who run the screening process in the CEU. We propose to select reviews from the weekly sign offs to allow us to continue to accept reviews on a referrals basis, to work with colleagues in CEAD and LSD on dissemination and learning initiatives, and to ensure that we have adequate capacity to work on the pilot for separating editorial from developmental function.

For cyclical audits we would like to involve independent assessors, ideally from the CRG community, for the purpose of assessing review abstract quality.

Risks and dependencies:

Currently we are aware that due to variation in CRG processes, the alerts in the workflows system that we use to identify signed off reviews (stage E) can be misleading. As previously managed between 2013 and

2015, we plan to flag reviews that we intend to triage with the CRGs and CES directly to reduce disruption to the author and CRG editorial process.

Should there be a backlog of work created by other triaging work we will communicate the reason for delays on receipt of reviews that have been referred.

Impact and change management

Not applicable.

Timelines

This is an ongoing process.

Management Responsibility

Toby Lasserson from the CEU will have operational leadership of the QA process.

Consultation:

List the names and titles of the people involved in the preparation of the Board paper.

Nuala Livingstone, Newton Opiyo, David Tovey, Liz Bickerdike, Kerry Dwan

Recommendation(s):

We would like the board to endorse the proposals that we have outlined in relation to review screening and support for other activities.

Appendix 1. Current triage tool (10 March 2017)

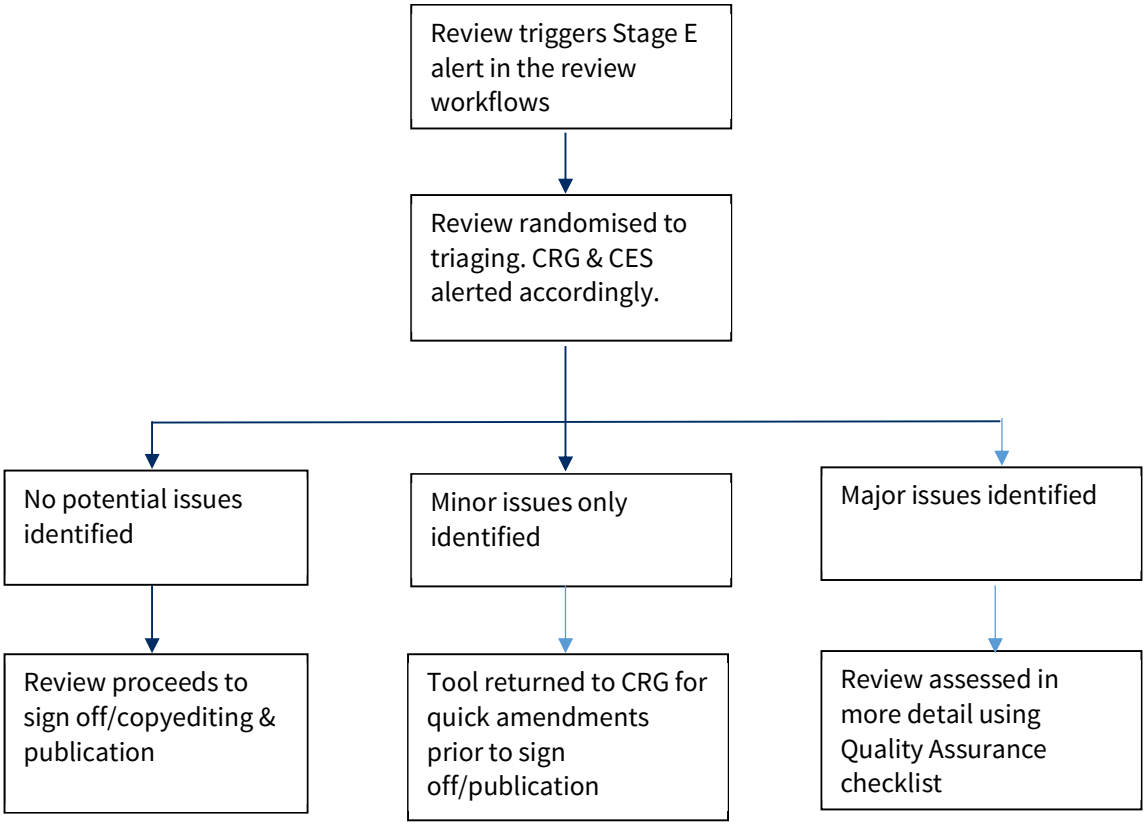
ABSTRACT		**SUMMARY OF FINDINGS TABLE**		**DATA AND ANALYSIS (for Critical and Important outcomes in Main comparison)**	
Item	**Response**	**Item**	**Response**	**Item**	**Response**
Title reflects the review question		SoF table presents main outcomes (both benefits and adverse effects) for main comparison		Analyses match the plan specified in the methods section (e.g. MDs or SMDs; fixed or random effects meta-analysis)	
Research question (PICO) is clear and the rationale for the review is well described		PICO (including Settings) presented and accurate		Data from non-standard designs (cluster, cross-over, etc.) appropriately incorporated where relevant (check 'Unit of analysis issues' in methods & footnotes in forest plots)	
Search date is less than 12 months from publication?		Outcomes fully defined (i.e. time of measurement, scale of measurement, range of scores specified)		Multiple measurements from multi-arm studies or subgroups handled appropriately (check for double counting of studies in Forest plot and adjustment of sample size in control groups)	
Direction, magnitude and confidence intervals of effects clearly described where appropriate		Assumed and Corresponding risks presented (where appropriate)		Outlying results acknowledged and explored appropriately	
Findings for all important outcomes reported for the main comparison(s), including information about harm? (i.e. consistent with the outcomes reported in the SoF table)		Clear and accurate summary of narrative results (where appropriate)		No unusually high or low mean/SD/count data (look at comparability of SDs for studies using same scale; check that sample sizes for same studies are similar across key outcomes; look at weights of individual studies relative to sample size)	
There an estimation of the certainty (or quality) of the body of evidence		Quality ratings presented for narrative results (where appropriate)			

using GRADE for each outcome reported in the abstract					
Absolute effects used to illustrate the relative effects where appropriate		GRADE ratings are clearly justified (supported by clear and appropriate quality assessment criteria in Footnotes)		Key findings consistent across the summary versions of the review (compare abstract, PLS, SoF table, Effects of interventions and Data tables)	
Reporting of results avoids emphasizing statistical significance to determine presence or absence of an effect					
Conclusions are an accurate reflection of the evidence presented in the GRADE SoF table(s) and do not make direct recommendations					

Time taken to Triage:
Decision (e.g. Proceed to full screen; Return to CRG for amendment; Proceed for publication):
Main Points of Note:

Cochrane Governing Board Paper

Appendix 2. Proposed workflow for triaging[1]



[1] 'Minor issues' are those that are easily explained and thus easily fixed (e.g., discrepancies between results in abstract and those in SoF tables, details omitted from the SoF table).

'Major issues' are those that are less easily explained and may require more guidance to fix (e.g. unit of analysis errors detected, conclusions accurately fail to reflect the evidence presented in the GRADE SoF table(s), discordance between abstract outcomes and those presented in the SoF tables, authors make recommendations, GRADE ratings are not clear and justifiable).

2 Audit report of published abstracts and 'Summary of findings' tables

Prepared by:	Toby Lasserson and Karla Soares-Weiser; sponsored by David Tovey
Date:	For Governing Board meeting, 5 April 2017
Decision or information	Information; request for the Board to note the contents of the paper
Strategy to 2020 target:	Yes
Status:	For decision

Executive summary

Cochrane Review abstracts provide a structured narrative summary of the review question, methods, results and conclusions. They are likely to be more widely read than the entire review, and may flag wider issues with the methods or interpretation of evidence in the full text of the review. One of the key objectives of the Structure and Function Review proposal approved by the Governing Board in Seoul in October 2016 was to develop and implement a rapid screening tool to evaluate reviews that had been signed off for publication by Cochrane Review Groups (CRGs). We assessed the current reporting quality in abstracts and explored whether this would be a feasible and effective way of screening all new and updated reviews.

We designed a checklist in SurveyMonkey, and CRGs were allocated to CEU Editors during December 2016. Results were collated in a spreadsheet, and cross-checked for discrepancies. Overall, we found that several CRGs performed well and the spread was perhaps less than we had anticipated. There was variation in performance on the different questions, but in most cases the PICO criteria were judged to be sufficiently clear, the methods used were appropriate to the review question, and the conclusions of the reviews avoided giving recommendations for practice or policy.

Background

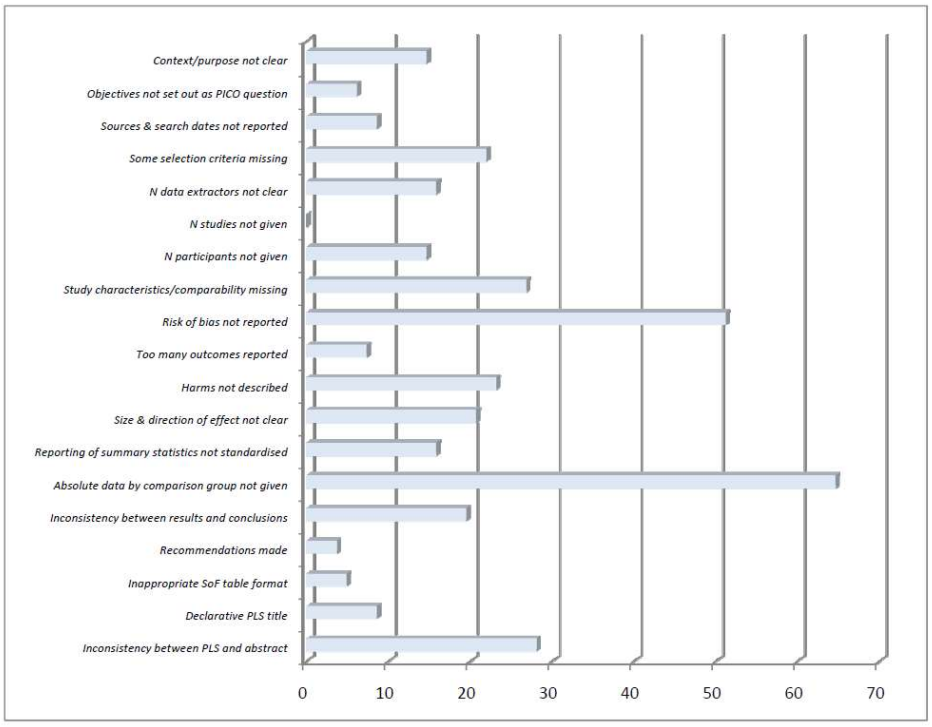
Why abstracts, why now?

Cochrane Review abstracts provide a structured narrative summary of the review question, methods, results and conclusions. They are likely to be more widely read than the entire review,1 and may flag wider issues with the methods or interpretation of evidence in the full text of the review. One of the key objectives of the Structure and Function Review proposal approved by the Governing Board in Seoul in October 2016 was to develop and implement a rapid screening tool to evaluate reviews that had been signed off for publication by Cochrane Review Groups (CRGs). We wanted to assess current reporting quality in abstracts and explore whether this was a feasible and effective way of screening all new and updated reviews.

This audit builds on previous work carried out within the Cochrane Editorial Unit (CEU) and contributes to a growing evidence base of systematic review abstract quality more generally. In 2011 an audit of the abstract, Plain language summary (PLS) and 'Summary of findings' tables (SoF) in 82 published Cochrane Reviews found a number of problems with abstracts. The main issues were inconsistency between the

abstract results and conclusions, omission of important information regarding selection criteria, lack of absolute effects, unclear search dates and risk of bias of included studies. See Figure 1 for summary of results of the audit.

Figure 1. Abstract audit results assessing reviews published in 2011. Note that the results are reported as items NOT met.



In 2013, the Preferred Reporting Items for Systematic Reviews and Meta-analyses for Abstracts (PRISMA-A) published guidance on how to write and present abstracts for systematic reviews and meta-analyses.[1] A number of audits of systematic review abstracts using PRISMA-A describe similar issues and raised concerns about the quality of published abstracts in non-Cochrane systematic reviews.[2-4]

We wanted to find out if quality of reporting of recently published Cochrane Review abstracts has improved and which areas remain problematic.

Developing the Abstract Checklist

Checklist development started immediately following Governing Board approval in October 2016. Our initial tool comprised 10 questions (each awarded 1 or 0 points depending upon whether criteria were present or absent), and, after consideration of MECIR reporting standards and testing by a single CEU Editor on 15 abstracts, was modified to the final version. See Appendix 1 for details of the initial abstract checklist.

Applying the checklist

A modified version of the initial checklist was applied to the five most recently published Cochrane Reviews of each of the 52 CRGs (including HIV and Fertility Regulation CRGs).

The checklist was designed in SurveyMonkey, and CRGs allocated to CEU Editors during December 2016. Results were collated in a spreadsheet, and cross-checked for discrepancies. After initial assessment, we realised that for some reviews not all responses were available (for example, information on results and interpretation for 'empty reviews'), and we decided to use 'not applicable' (NA) as a possible response in these circumstances. Details and guidance for the checklist can be seen in Table 1.

Table 1. Modified abstract checklist applied to all CRGs

#	Item	Question	Scoring system (0 to 12 points)
1	Title	Does the title reflect the review question?	Yes (1 point) No (0 points)
2	Background and objectives	Is the research question (PICO) clear and the rationale for the review well described?	Yes (1 point) No (0 points)
3	Search methods	Is the search date less than 12 months from publication?	Yes (1 point) No (0 points)
4	Search methods	Does the abstract indicate that trials registers were searched?	Yes (1 point) No (0 points)
5	Selection criteria	Are the eligible study designs described in the abstract appropriate to the review question?	Yes (1 point) No (0 points)
6	Data collection and analysis	Are the direction, magnitude and confidence intervals of effects clearly described where appropriate?	Yes (1 point) No (0 points)
7	Main results and SoF table	Are the findings for all important outcomes reported for the main comparison(s), and does this include information about harm (i.e. consistent with the outcomes reported in the SoF table)?	Yes/NA (1 point) No (0 points)
8	Main results	Is there an estimation of the certainty (or quality) of the body of evidence using GRADE for each outcome reported in the abstract?	Yes/NA (1 point) No (0 points)
9	Main results	Have absolute effects been used to illustrate the relative effects where appropriate?	Yes/NA (1 point) No (0 points)
10	Main results	Does the reporting of results avoid emphasizing statistical significance to determine presence or absence of an effect?	Yes/NA (1 point) No (0 points)
11	Authors' conclusions	Are the conclusions an accurate reflection of the evidence presented in the GRADE SoF table(s)?	Yes (1 point) No (0 points)
12	Authors' conclusions	Do the authors avoid making recommendations?	Yes (1 point) No (0 points)

Results

Table 2 shows the final score for each one of the five most recent reviews, per CRG. Figure 2 shows the overall results for the five reviews (0-60 points; 5 reviews, 0-12 points per review) per CRG, ranked according to higher scores.

Table 2: Audit of the five most recently published reviews in the Cochrane Library, per CRG, as of 15 December 2016

Cochrane Review Group	Review 1 (0-12 points)	Review 2 (0-12 points)	Review 3 (0-12 points)	Review 4 (0-12 points)	Review 5 (0-12 points)
Airways	9	11	8	12	10
Anaesthesia	9	12	9	9	11
ARI	9	12	12	9	10
Back and Neck	8	8	9	9	9
Bone, Joint & Muscle Trauma	11	12	9	11	7
Breast Cancer	11	10	10	11	11
Childhood Cancer	11	7	8	8	11
CIDG	10	11	11	11	12
Colorectal Cancer	9	9	10	10	10
Common Mental Disorders	9	10	11	9	9
Consumers	9	10	10	10	10
Cystic Fibrosis	3	6	11	7	12
Dementia	10	11	11	12	12
Development	8	11	10	10	12
Drugs & Alcohol	8	10	9	9	11
ENT	10	12	11	10	11
Epilepsy	7	8	9	5	9
EPOC	9	10	9	10	10
Eyes & Vision	12	12	12	11	12
Fertility Regulation	7	9	9	10	10
Gynaecological Cancer	12	10	9	11	10
Gynaecology	11	12	12	12	12
Haematological Malignancies	12	12	12	12	12
Heart	8	10	10	10	11
Hepato-biliary	8	9	12	9	8
HIV	8	9	10	12	10
Hypertension	9	12	8	11	11
IBD	8	9	11	9	8
Incontinence	8	8	10	7	8
Injuries	9	10	10	10	11
Kidney	6	10	6	7	7
Lung Cancer	6	8	12	6	10
Metabolic & Endocrine	10	10	11	12	12
Movement Disorders	8	11	12	8	11
Multiple Sclerosis	10	8	9	10	12
Musculoskeletal	8	10	11	12	12
Neonatal	10	8	10	9	11

Cochrane Review Group	Review 1 (0-12 points)	Review 2 (0-12 points)	Review 3 (0-12 points)	Review 4 (0-12 points)	Review 5 (0-12 points)
Neuromuscular	7	7	11	9	11
Oral Health	11	10	11	10	11
PaPaS	10	8	8	12	12
Pregnancy & Childbirth	9	11	11	8	12
Public Health	6	7	7	8	11
Schizophrenia	7	9	8	9	9
Skin	9	10	11	10	12
STI	7	10	11	12	12
Stroke	8	9	6	9	10
Tobacco Addiction	8	8	8	8	10
Upper GI	5	7	9	10	10
Urology	7	9	10	11	8
Vascular	12	12	11	10	12
Work	12	8	9	9	11
Wounds	12	12	12	11	12

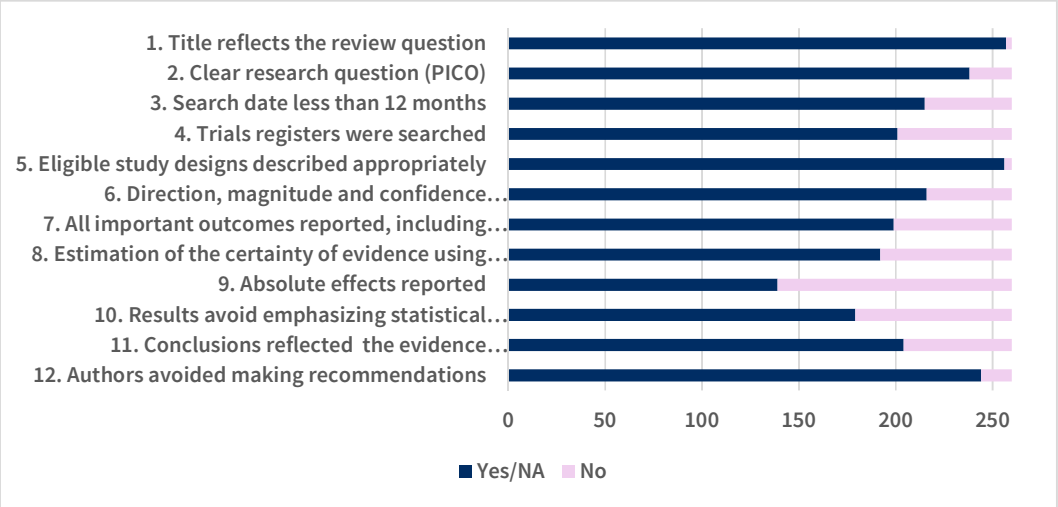
Overall, we found that several CRGs performed well on the audit and the spread was perhaps less than we had anticipated. In addition, some of the CRGs who had previously been identified as being at high risk performed creditably – perhaps due to changes in the editorial process and quality assurance system within the CRGs and possibly due to direct input from the CEU team.

There was variation in performance on the different questions. In most cases the PICO criteria were judged to be sufficiently clear, the methods used were appropriate to the review question, and the conclusions of the reviews avoided giving recommendations for practice or policy. In contrast, the following features were most likely to be associated with lost points (Figure 3).

- No mention of whether trials registers had been searched in the abstract[1]
- Absence of any attempt to report or estimate absolute effects
- Failure to import GRADE ratings into the narrative text of the abstract
- Over emphasis on statistical significance in the reporting of results, frequently leading to phrasing that equated non-statistical significance with no effect
- Failure to report all important outcomes, including harm

[1] Note that in this audit we did not check this against the Methods section of the review, so that in some cases points were deducted despite the authors having searched registers.

Figure 2. Overall responses for each one of the 12 questions (260 abstracts of published reviews evaluated; 1 point awarded for each question that received a positive answer)



Implications of the results

Our findings show that there are areas for improvement in a number of abstracts. The proportion of published abstracts that overlook harms remains around 25%. The proportion of abstracts that convey information about absolute effects stands at just over 53% compared with 35% in the 2011 cohort. Increasing uptake of GRADE and inclusion of absolute effects in 'Summary of findings' tables could explain this increase. We also saw that 73% of abstracts include GRADE ratings for important outcomes.

Experience of using the checklist has shown that it needs to be modified before it can be used as a screening tool. Further items relevant to review conduct may need to be incorporated to improve its ability to identify quality issues beyond the summary versions of the review. However, recognising that abstracts, along with PLS, are the most widely read sections of the reviews, the checklist is a useful gauge of the state of abstracts in Cochrane Reviews. Our checklist was intended to cover key processes of the review reported in abstracts. In retrospect we think that reporting searches of trials registers might not be an essential element of the abstract, notwithstanding their importance in searches for the review.

When creating the audit tool we assigned equal weight to each item. This may have overlooked varying degrees of importance attached to different criteria according to MECIR. Selective outcome reporting, especially of harms, for example, is a more serious source of bias than failing to include an estimation of absolute effects.

We wanted to identify examples of substandard reporting, so for empty reviews we scored the reviews positively for responses that were judged as 'not applicable', i.e. the reporting of results. This will have inflated the scores for CRGs that included empty reviews, and renders cross-CRG comparisons somewhat unreliable. Owing to issues of feasibility, most reviews were only scored by one editor and inter-rater differences would have affected the scores for individual items. We attempted to limit these by having regular discussions between the assessing editors, and also by validating scores independently for abstracts that had scored poorly. In the latter case, the inter-assessor reliability was not perfect, but we judged it to be reasonable.

We believe that the audit provides a useful snapshot of the quality of reporting of abstracts across Cochrane Reviews published in the last year or so across all CRGs. There are many examples of excellent practice, as well as clear areas for improvement that would make the reviews easier to interpret by readers and probably improve their impact and utility.

We will describe how the results of the audit have influenced our proposals to change the screening process in a separate paper.

CEU team involved in the abstract audit

Liz Bickerdike (Editor)

Kerry Dwan (Statistical Editor)

Nuala Livingstone (Editor)

Jane Marjoribanks (Editor)

Newton Opiyo (Editor)

Elizabeth Royle (Copy Edit Support Manager)

Sera Tort (Clinical Editor)

Helen Wakeford (Editor)

Toby Lasserson (Senior Editor)

Karla Soares-Weiser (Deputy Editor in Chief)

David Tovey (Editor in Chief

References

1. Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, et al. PRISMA for Abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Med*. 2013;10(4):e1001419.

2. Beller EM, Glasziou PP, Hopewell S, Altman DG. Reporting of effect direction and size in abstracts of systematic reviews. *JAMA*. 2011;306(18):1981-2.

3. Bigna JJ, Um LN, Nansseu JR. A comparison of quality of abstracts of systematic reviews including meta-analysis of randomized controlled trials in high-impact general medicine journals before and after the publication of PRISMA extension for abstracts: a systematic review and meta-analysis. *Syst Rev*. 2016;5(1):174.

4.Tsou AY, Treadwell JR. Quality and clarity in systematic review abstracts: an empirical study. *Res Synth Methods*. 2016;7(4):447-58.

Figure 3. Audit of the five most recently published reviews in the Cochrane Library, per CRG, as of 15 December 2016 (Each review scored 0-12 points, totalling a maximum of 60 points per CRG)