BMJ open

Selective reporting of outcomes in randomised controlled trials of Cystic Fibrosis

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-002709
Article Type:	Research
Date Submitted by the Author:	11-Feb-2013
Complete List of Authors:	Dwan, Kerry; The University of Liverpool, Biostatistics Kirkham, Jamie; The University of Liverpool, Biostatistics Williamson, Paula; The University of Liverpool, Biostatistics Gamble, Carrol; The University of Liverpool, Biostatistics
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Genetics and genomics
Keywords:	STATISTICS & RESEARCH METHODS, MEDICAL EDUCATION & TRAINING, GENETICS

Selective reporting of outcomes in randomised controlled trials of Cystic Fibrosis

Kerry Dwan[1§], Jamie J Kirkham[1], Paula R Williamson[1], Carrol Gamble[1]

1. Department of Biostatistics, University of Liverpool, Liverpool, England.

[§]Corresponding author

Email addresses:

KD: kerry.dwan@liverpool.ac.uk

JJK: jjk@liverpool.ac.uk

PRW: prw@liverpool.ac.uk

CG: c.gamble@liverpool.ac.uk

Abstract

Background

Outcome reporting bias (ORB) in randomised trials and systematic reviews has been identified as a threat to the validity of systematic reviews. Previous work highlighting this problem is limited to considering a single primary review outcome. Cystic fibrosis systematic reviews are often characterised by inclusion of small randomised trials and specify multiple review primary outcomes increasing concern about ORB. The aim of this study was to assess ORB across all efficacy outcomes in systematic reviews of cystic fibrosis.

Methods

Systematic reviews of interventions for cystic fibrosis published on the Cochrane Library by the Cochrane Cystic Fibrosis and Genetic Disorders Group before 2010 were assessed for differences in outcomes between review protocol and full review. ORB in eligible trials was also assessed for all review outcomes. Two authors independently classified each outcome using a nine point classification system developed by the ORBIT (Outcome Reporting Bias In Trials) study. These classifications were used to inform the assessment of the risk of bias for selective outcome reporting for each trial.

Results

Forty six Cochrane cystic fibrosis systematic reviews were included. The median number of primary outcomes, number of trials, and participants per trial in the reviews were 3 (IQR 2, 3), 4 (IQR 2,8) and 21 (IQR 14,41) respectively. Eighteen reviews (39%, 18/46) had a discrepancy in outcomes between protocol and full review. Thirty seven reviews were eligible to be included in the ORB assessment. When considering review primary outcomes and all review outcomes, outcome reporting bias was suspected in at least one trial in 86% and 100% respectively.

Conclusion

Systematic reviews with multiple primary and secondary efficacy outcomes are at greater risk of ORB.

ORB in trials is highly prevalent within systematic reviews of cystic fibrosis. This could be reduced by the development of a core outcome set for trials and systematic reviews in cystic fibrosis.

Article summary

Article focus

- Assessment of discrepancies in outcome selection between systematic review protocols and full reviews.
- Assessment of outcome reporting bias across all efficacy systematic review outcomes.
- Assessment of the overall risk of bias from selective reporting of outcomes of a trial within a systematic review.

Key messages

- Systematic reviews with multiple primary and secondary efficacy outcomes are at greater risk of ORB.
- Clearer guidance is needed on how to assess the 'overall' risk of bias as a result of ORB for each included trial within a systematic review, when considering multiple outcomes.
- The development of a core outcome set in Cystic Fibrosis would help reduce the problem.

Strengths and limitations

This is the first study to consider the assessment of outcome reporting bias in all efficacy review outcomes. However, this is limited to reviews of cystic fibrosis.

Background

The value of systematic reviews in establishing an evidence base is widely acknowledged with well conducted systematic reviews of randomised controlled trials being placed at the top of the hierarchy of evidence (Green and Byar, 1984). It is essential, when conducting systematic reviews, to consider the potential for bias and its impact on the review conclusions. Bias may be induced through the decisions and actions of the authors of the included clinical trials or systematic review authors.

Bias in a systematic review is frequently considered in relation to limitations of the search strategy. However, bias may also occur, for example, when outcomes are added, omitted or changed after a systematic review protocol is published if the decision to deviate from the protocol is based on the significance of the results. A study of an unselected cohort of Cochrane reviews revealed that over a fifth (64/288) of protocol/review pairings showed some discrepancy in at least one outcome measure with just 6% (4/64) describing the reason for the change in the review (Kirkham, 2010a). Results also indicated that outcomes promoted from primary to secondary between the protocol and the review were more likely to report statistically significant meta-analysis results in comparison to reviews where there was no discrepancy in outcome specification with the review protocol (relative risk 1.66 95% confidence interval (1.10, 2.49), p = 0.02).

Systematic reviews are only as valid as the trials they contain (Juni et al 2001), consequently much effort is given to assessing the risk of bias within the trials identified by assessing their methodological quality. However, it is also important to consider the content of trial reports in an assessment of bias. Outcome reporting bias (ORB) within a RCT is defined as the result-based selection of a subset of the original outcomes for publication (Williamson and

Hutton 2000). In a systematic empirical assessment of Cochrane reviews within which a single review primary outcome could be identified (Kirkham et al 2010b), ORB was suspected in at least one randomised controlled trial in more than a third of the systematic reviews that were examined (35%). This study may have underestimated this problem as review primary outcomes are chosen due to their clinical importance so are more likely to have been measured and reported in trials increasing concern regarding the prevalence and impact of outcome reporting bias in reviews where multiple primary outcomes are specified, or in secondary outcomes.

Systematic reviews in cystic fibrosis are characterised by inclusion of small randomised trials specifying multiple primary outcomes. Reporting standards for trials of cystic fibrosis have also been shown to be low (von Mosch and Dwan 2011). The aims of this current study were to

1. Examine the potential for bias created by review authors by identifying inconsistencies between outcomes published in review protocols and in the associated published reviews
2. Determine the prevalence of ORB in trials in systematic reviews of CF, extending previous work by considering all review efficacy outcomes (multiple primary and secondary).

Methods

A cohort of systematic reviews published by the Cochrane Cystic Fibrosis and Genetic Disorders (CFGD) group on the Cochrane Library before 2010 were identified (The Cochrane Library, 2009). Reviews were eligible for inclusion if they compared interventions for cystic fibrosis and identified one or more eligible RCTs. RCTs that had been excluded (in

the "characteristics of excluded studies" section) were also checked for any suggestion of outcome reporting bias. For example, if a review had excluded trials as a result of 'no relevant outcome data (NROD)', then these trials were also scrutinised for the presence of ORB and included in the assessment.

Changes in outcomes between systematic review protocol and full review

The numbers of primary and secondary outcomes per review were compared to the recommendations for the number of outcomes (no more than three primary outcomes and a limited number of secondary outcomes) to include in a review in the Cochrane Handbook (Higgins and Green 2011). If a review did not distinguish between primary and secondary outcomes, the first three outcomes listed were taken to be the primary outcomes and the rest were considered as secondary outcomes. Protocols of the systematic reviews were accessed and outcomes stated in the protocol were compared to those stated in the full review. Changes in outcomes were identified and categorised by one author (KD) as: primary outcome downgraded to secondary (*downgrade*); secondary outcome upgraded to primary (*upgrade*); a new outcome not stated in the protocol was added to the full review (*addition*) or an outcome stated in the protocol was omitted from the full review (*omission*). If there had been a change in outcomes, the section 'changes between protocol and review' was examined for a declaration and explanation of the changes.

Assessing trial reports for full ORB

For each eligible systematic review, all reports relating to included studies and studies excluded due to no relevant outcome data were obtained. Reviews were checked to see whether review authors had contacted trialists for further information or data for outcomes. Where this was not clear in the review, review authors were asked to clarify.

A nine-point classification system (Table 1) developed for missing or incomplete outcome reporting in randomised trials (Kirkham et al, 2010b) was used to make an assessment of the risk of bias. Table 1 also provides examples of outcomes that were not assessed because they had poor outcome definitions. An outcome matrix (Table 2) was created for each review using the ORBIT matrix generator (http://ctrc.liv.ac.uk/orbit/), with studies listed in the rows and review primary and secondary outcomes listed in the columns with the ORBIT classifications (Table 1) given for each review outcome that was not fully reported (e.g. not reported or partially reported e.g. $p>0.05$).

The outcomes listed or detailed in the method section and the outcomes reported in the results section were compared for all trial publications to determine whether each outcome of the systematic review was measured and analysed. In some instances it may be obvious that an outcome was measured given the other outcomes reported. For example, if cause-specific mortality is reported then overall mortality must have been measured, even if not reported. In other situations it may be that a battery of tests or measurements are usually undertaken together, for example FEV1 (forced expiratory volume in 1 second) and FVC (forced vital capacity). If FVC is reported but FEV1 is not, suspicion should be raised that the latter may have been selectively not reported. However, it is often difficult to assess whether an outcome was measured, and clinical judgment is required. The clinical lead for each review was contacted by email and asked for their input into the assessment of selective outcome reporting within the trials included in their review. An assessment of whether the review outcomes had been measured and reported within each trial using the classification system was completed. The clinical lead for the review and KD independently assessed the trials in the review and any disagreements were resolved through discussion and then checked with a third person (JJK or PRW).

If one or more of the outcomes for a trial was given a high risk classification according to Table 1, the trial was deemed at high risk of bias from selective reporting.

Analysis

Descriptive results are presented. The median and interquartile range for the number of review primary and secondary outcomes was calculated.

Data are tabulated and excerpts found in the trial reports relating to review outcomes are used to support decisions made regarding ORBIT classifications and the assessment of risk of bias.

Results

The CFGD group had 46 cystic fibrosis systematic reviews published as of 2010.

1. Changes in outcomes between systematic review protocol and full review

Protocols were available for all 46 systematic reviews. Nine protocols (20%) did not distinguish between primary and secondary outcomes. Table 3 shows the median number of primary and secondary outcomes for the 46 reviews and the changes in outcomes between protocol and full review.

Eighteen reviews (39%, 18/46) had a discrepancy in outcomes between protocol and full review. Between review protocol and full review, five (28%) listed all changes, two (11%) listed some changes and 11 reviews (61%) did not mention any change in outcomes. Of the seven reviews describing the changes between protocol and full review, three provided no reason for the changes, two stated that the changes in recommendations in the Cochrane Handbook to have a maximum of three primary outcomes were the reason for downgrading

outcomes and two reviews stated that they added clinically relevant outcomes that were discovered during the review process.

2. ORBIT classifications

Of the 46 published reviews, 38 were eligible to be assessed for outcome reporting bias (Figure 1).

One review was excluded at this stage as the outcomes could not be assessed for ORB due to the different ways the outcome definitions could be measured and reported. The primary outcomes were psychosocial outcomes, which included any objective measure with adequate psychometric properties and demonstrable reliability and validity quantifying psychological or social outcomes or both, including individual psychological adjustment, relational, social functioning and adaptation to life with cystic fibrosis.

Therefore 37 reviews were assessed for ORB, including 280 RCTs (278 included and 2 excluded due to no relevant outcome data but confirmed by review authors that they would have otherwise been included). The median number of trials per review was four (IQR 2, 8) and there was a median sample size of 21 (IQR 14, 41) per trial.

Review authors contacted trialists for missing outcome data in 33 reviews (89%), one stated that "trialists were not contacted but would be in updates of the review" and three reviews did not state if trialists were contacted for further data.

Lead authors of each review were contacted. The lead authors of twelve reviews assessed the included trials and gave classifications for each outcome. For thirteen reviews, authors gave

input on which outcomes they expected to be measured for trials in their review and which outcomes they expected to be measured in routine clinical practice but did not classify each outcome due to time restrictions. The authors of twelve reviews did not respond to our request.

For the twelve reviews where the authors assigned classifications, discussion was needed on all outcomes to come to an agreed classification. For the other 25 reviews it was difficult to assign a classification to all outcomes as some outcomes needed a large amount of clinical input in understanding the outcome and language used to describe the outcomes within the trial reports.

Due to the number and complexity of outcomes and lack of reviewer input on the majority of reviews, it was decided that the assessment of all primary outcomes listed in the full review that were well defined should take priority. Many outcomes were also split into sub outcomes or ill defined to maximise the ability of a trial to contribute data to the review. For example lung function was often split into FEV1 (Volume that has been exhaled at the end of the first second of forced expiration), FVC (Forced vital capacity), PEFR Peak expiratory flow rate), FEF_{25-75} (average expired flow over the middle half of the FVC manoeuvre) and these were assessed separately. FEV1 is the outcome most often considered for lung function due to its validity, repeatability and it is the outcome most understood by clinicians. However, the device used to measure FEV1 also measures the majority of other lung function outcomes. Therefore if FEV1 was reported in a trial, it was assumed that other lung function outcomes were also measured but not necessarily analysed (classification F) unless they were specifically stated as an outcome in the trial report. However, if FEV1 was not reported but

other lung function outcomes were then an E classification was given to FEV1. This was decided after discussion with clinical experts.

The ORBIT classifications for the review primary outcomes for the 280 RCTs are shown in Table 1. For the 12 reviews where reviewer input was obtained, classifications for 64 included trials for review secondary outcomes are also shown in Table 1. In addition to these classifications, a 'G, no events' classification (For example, mortality, were clinical judgement says it is likely to have been measured and it would have been reported had any deaths occurred. Therefore, it is assumed no deaths occurred during the trial.) was given to 109 trials for review primary outcomes and 22 trials for review secondary outcomes. We were unable to assess outcomes (including: adverse events, symptoms, complications, biochemical measures of glycaemic control, symptoms of sleep disordered breathing and measures of specific indices of strength, mass, effort and general fatigue) for 102 trials for review primary outcomes and 59 trials for review secondary outcomes.

Assessment of risk of bias from selective reporting

Eighteen reviews (49%) had not yet assessed the risk of bias for selective outcome reporting as they had not been updated since the new Cochrane guidance on the risk of bias had been introduced and prior to this study. Seventeen reviews (46%) had assessed the risk of bias for all included trials and two reviews (5%) assessed this for some of their included trials.

As we were unable to assess secondary outcomes for ORB for all reviews, the risk of bias assessments were made based on classifications of primary outcomes in order to be consistent across reviews. Table 4 shows the risk of bias for selective outcome reporting as defined in

this study and also as assessed within the published reviews for the 280 trials assessed for ORB based on the consideration of review primary outcomes only.

Only five (14%) of the 37 reviews had no trials at high risk of bias based on the review primary outcomes only.

Table 5 shows the risk of bias for selective outcome reporting based on the consideration of review primary and secondary outcomes separately for the 12 reviews (64 trials) were reviewers also provided classifications. This was to see if decisions regarding risk of bias would change if we considered all outcomes. Only four (6%) of the 64 RCTs had a low risk of bias when considering all outcomes.

Discrepancies in the risk of bias when considering all outcomes arose in 34 (53%) trials; 31 were at low risk when considering review primary outcomes only but high risk of bias (excluding G classifications: 13, G classification only: 18) when considering all outcomes; 3 were at high risk (G classifications only) when considering review primary outcomes only but high risk (excluding G classifications) when considering all outcomes. This often occurred in reviews were there was only one or two primary outcomes and a large number of secondary outcomes.

Based on all review outcomes, none of the 12 reviews had all included trials at low risk of bias.

Discussion

This is the first study to consider all review efficacy outcomes in an ORB assessment which has allowed us to make practical recommendations on assessing the risk of bias of selective reporting for systematic reviews at both the review and trial level. Over a third of Cochrane cystic fibrosis reviews (39%) examined had a discrepancy in outcomes between the review protocol and full review. This compares to 22% of reviews (64/288) that contained a discrepancy in at least one outcome measure in the main ORBIT study which looked at reviews covering all 50 Cochrane review groups (Kirkham et al 2010a). However, this is confounded by the times in which the reviews were published. Furthermore, for the cystic fibrosis reviews outcome reporting bias was suspected in at least one randomised controlled trial in 86% of reviews when considering all review primary outcomes. The prevalence of reviews containing at least one trial with high suspicion of outcome reporting bias from ORBIT, when only a single primary outcome was considered was substantially lower at 34% (96/283) (Kirkham et al 2010b). While this study is limited only to CF trials, it is clear that the problem of outcome reporting bias is much larger when considering more than just the single primary review outcome of importance that was used in the ORBIT study.

Use of the ORBIT classification system offered a robust methodology for assessing the risk of bias for trials included within a systematic review. When considering the 64 trials in the 12 reviews were it was possible to assess both primary and secondary outcomes, when basing the risk of bias assessment on review primary outcomes, 45% of trials were at high risk of bias and when using all outcomes in the assessment, 94% were at high risk of bias. Using the current selective reporting item of the current Cochrane risk of bias tool, 69% of trials included in CF reviews were assessed by reviewers as ‘unclear’ risk of bias indicating the need for more informed guidance on assigning risk of bias in the systematic review process for all outcomes within a review.

The ORBIT classification system has already been validated as part of the original project. Sensitivity results for predicting that the outcome had been measured (G-classification) was 92% (23/25, 95% CI 81% to 100%), while the specificity for predicting that the outcome had not been measured (H-classification) was 77% (23/30, 95% CI 62% to 92%). With the additional requirement to assess all outcomes in this project, there were an increasing number of outcomes that were not mentioned in the trials reports and therefore clinical judgement was needed as to whether the outcome of interest was likely to have been measured in a particular trial. Many review authors did not respond to our request, therefore only primary outcomes were assessed within the majority of reviews due to the clinical complexity of many of the secondary outcomes.

Reviewers should ensure that changes between protocol and reviews are listed and justifications provided to enhance the validity of these decisions. Eligible trials should not be excluded on the basis of "No relevant outcome data" because although an outcome was not reported it may have been measured and contact with the authors is advised. Reviewers should be encouraged to consider trials that have not reported an outcome of interest and to assess whether selective reporting has occurred for all review outcomes. They should consider the amount of missing data from their meta-analysis (i.e. the sample sizes of the studies that would have been eligible to be included in the meta-analysis but no outcome data reported) and this information should be included along with the pooled effect estimate. If appropriate, a sensitivity analysis should be applied to assess the robustness of the conclusions of the review (Dwan et al 2010).

Individuals conducting systematic reviews need to address explicitly the issue of missing outcome data for their review to be considered a reliable source of evidence. Extra care is required during data extraction, reviewers should identify when a trial reports that an outcome was measured but no results were reported or events observed, and contact with trialists should be encouraged. Contacting authors is encouraged by the CRG and is standard practice within CFGD reviews which is reflected in our results as 89% of reviews stated that they contacted authors for extra information on outcomes.

Reviewers also need to ensure that outcomes are well defined. Lung function was specified as the first primary outcome in nineteen reviews (50%), as the second or third primary outcome in 11 reviews (29%), as a secondary outcome in six reviews (16%) and it was not included as an outcome in only one review (5%). However, it is often split into 'suboutcomes', including FEV1, FVC, mid forced expiratory flow (FEF), residual volume (RV), total lung capacity (TLC), Lung clearance index (LCI) and maximum expiratory flow (MEF). These outcomes can then be analysed and reported in different ways such as**:** % predicted, litres, litres/second and post treatment, absolute change from baseline, relative change from baseline or annual rate of change. Therefore there is a large scope for selective reporting. One solution is the development of a core outcome set for cystic fibrosis (Ramsey and Boat 1994 , Sinha et al 2008, Clarke 2007). It is recommended that review authors ensure that they limit the number of outcomes in the review and define them clearly. This will allow easier assessments of selective reporting, which can be done during data extraction of the included trials as long as a knowledgeable clinical person is involved.

Unanswered questions and future research

Work is needed to consider what the best method is to assess the impact of ORB on the results of the meta-analysis when there are multiple outcomes. Multivariate meta-analysis has been suggested by Kirkham et al 2012 and a model based correction has been suggested by Copas et al (2012).

Conclusion

Systematic reviews need to clearly state the primary and secondary outcomes that they will consider and be consistent between review protocol and full review.

Outcome reporting bias is a major problem for systematic reviews and more guidance needs to be included in the Cochrane handbook to allow assessment of this important item within the risk of bias tool. We recommend that an outcome matrix be completed during the production of a review to allow an ORB assessment for all review outcomes which can then inform the risk of bias assessment.

A core set of outcomes should be agreed upon for cystic fibrosis which in turn will have a positive impact on systematic reviews.

Abbreviations

CFGD	Cystic Fibrosis and Genetic Disorders
CRG	Cochrane Review Group
COMET	Core Outcome Measures in Effectiveness Trials
NROD	No Relevant Outcome Data
ORB	Outcome Reporting Bias
ORBIT	Outcome Reporting Bias In Trials
RCT	Randomised Controlled Trial

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing Interests

There are no financial or non-financial competing interests.

Authors' contributions

KD drafted the protocol, completed the ORB assessments and wrote the manuscript.

JJK completed the ORB assessments and commented on the manuscript.

CG commented on the manuscript.

PRW commented on the protocol and the manuscript and commented on the ORB assessments.

Acknowledgements

The authors would like to thank the Cochrane CFGD group reviewers for their assistance with clinical input for the ORB assessments of the included trials in their reviews.

Data Sharing

The outcome matrix for each systematic review included is available from the contact author upon request.

References

Clarke, M. Standardising Outcomes for Clinical Trials and Systematic Reviews. *Trials; 2007,* 8**,** 39.

Copas J, Dwan K, Kirkham JJ, Williamson PR. A model based correction for outcome reporting bias in meta-analysis. (submitted).

Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, et al. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. PLoS ONE 2008, 3(8): e3081.

Dwan K, Gamble C, Kolamunnage-Dona R, Mohammed S, Powell C, Williamson PR. Assessing the potential for outcome reporting bias in a review: A tutorial. *Trials 2010,* **11:**52 doi:10.1186/1745-6215-11-52.

Green, S. B. & Byar, D. P. Using observational data from registries to compare treatments: the fallacy of omnimetrics. *Stat Med, 1984;* 3**,** 361-73.

Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001; 323: 42-46.

Kirkham JJ, Altman DG, Williamson PR. Bias Due to Changes in Specified Outcomes during the Systematic Review Process. PLoS ONE 2010a; 5(3): e9810.doi:10.1371/journal.pone.0009810

Kirkham JJ, Dwan K, Dodd S, Altman DG, Smyth R, Jacoby A, Gamble C, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ 2010b; 340: c365

Kirkham JJ, Riley RD and Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. Statistics in Medicine 2012*; 31:20, 2179-2195.*

Ramsey BW, Boat TF (1994) Outcome measures for clinical trials in cystic fibrosis. Summary of a Cystic Fibrosis Foundation consensus conference. J Pediatr 1994; 124: 177–192.

Sinha I, Jones L, Smyth RL, Williamson PR. A Systematic Review of Studies That Aim to Determine Which Outcomes to Measure in Clinical Trials in Children . PLoS Med 2008; 5(4): e96. doi:10.1371/journal.pmed.0050096

Smyth AR,Walters S. Prophylactic anti-staphylococcal antibiotics for cystic fibrosis. *Cochrane Database of Systematic Reviews* 2003, Issue 3. Art. No.: CD001912. DOI: 10.1002/14651858.CD001912.

The Cochrane Library, Issue 4, 2009. Chichester: Wiley.

Williamson, P. R., Gamble, C., Altman, D. G. & Hutton, J. L. Outcome selection bias in meta-analysis. *Stat Methods Med Res, 2005;* 14**,** 515-24.

von Mosch S, Dwan K. Reporting of harm and efficacy outcomes of randomised clinical trials in cystic fibrosis before and after the publication of the revised CONSORT (CONsolidated Standards Of Reporting Trials) statement in 2001 – a comparison of trial reports published from 1994 and from 2008. The University of Liverpool (thesis) 2011.

Figure 1: Study flow diagram

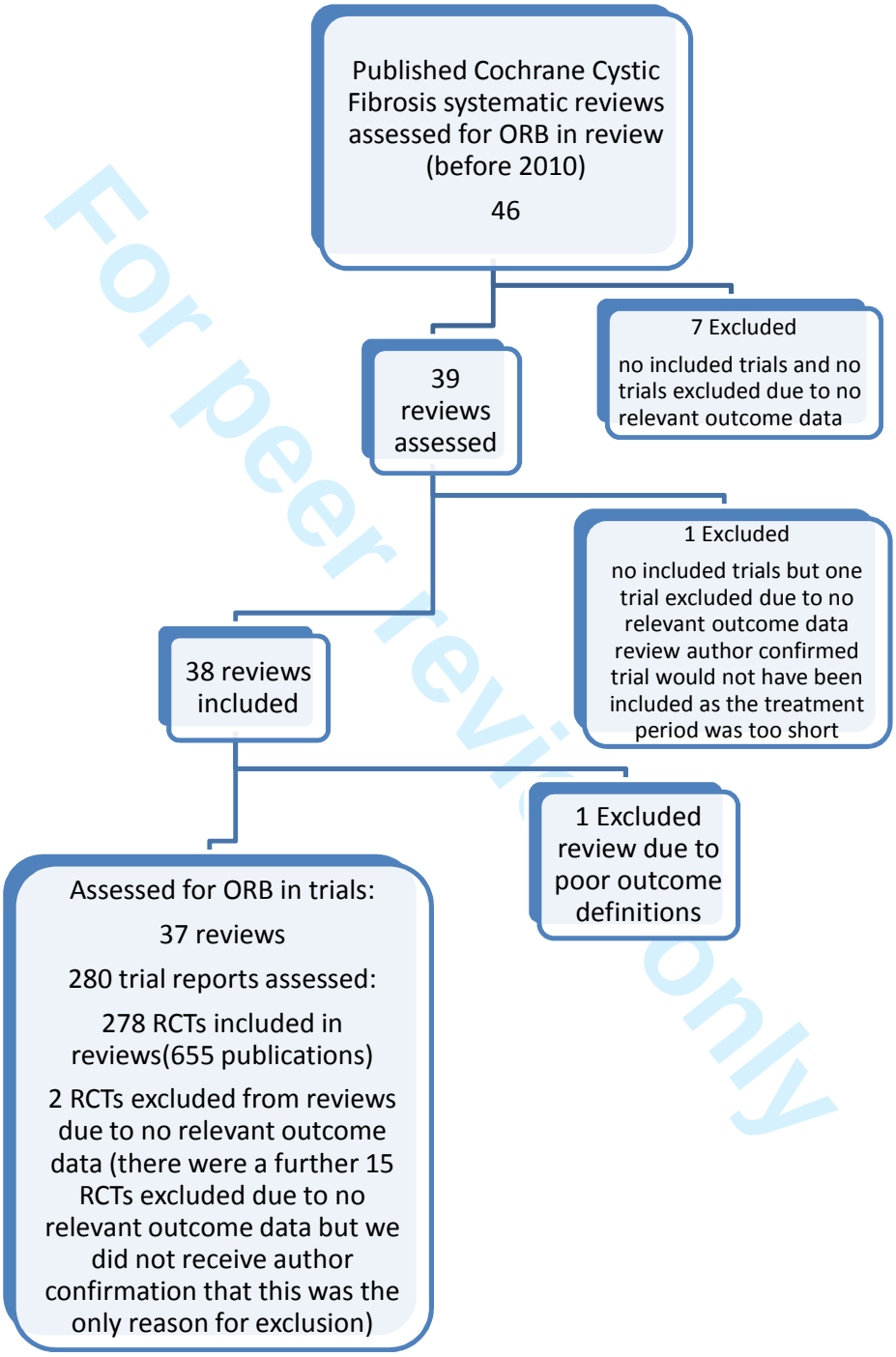


Table 1: ORBIT classifications

Classification	Description	Level of reporting	Level of suspicion of ORB	Primary outcome classifications	Secondary outcome classifications
				Number of trials	Number of trials
Clear that the outcome was measured and analysed					
A	States outcome analysed but only reported that result not significant (typically stating p-value >0.05).	Partial	High risk	75	12
B	States outcome analysed but only reported that result significant (typically stating p-value <0.05).	Partial	Low risk	13	2
C	States outcome analysed but insufficient data presented to be included in meta-analysis or to be considered to be fully tabulated.	Partial	Low risk	53	15
D	States outcome analysed but no results reported.	None	High risk	0	0
Clear that the outcome was measured					
E	Clear that outcome was measured but not necessarily analysed.	None	High risk	59	26
F	Clear that outcome was measured but not necessarily analysed.	None	Low risk	110	15
Unclear that the outcome was measured					
G	Not mentioned but clinical judgment says likely to have been measured and analysed.	None	High risk	195	197
H	Not mentioned but clinical judgment says unlikely to have been measured.	None	Low risk	141	256
Clear that the outcome was NOT measured					
I	Clear that outcome was not measured.	N/A	No risk	0	0

The ORBIT classifications for review primary outcomes for the 280 RCTs. For the 12 reviews where reviewer input was obtained, classifications for 64 included trials for review secondary outcomes are also shown.

Table 2: Example of review outcome matrix for 6 of 17 outcomes in a review of Prophylactic anti-staphylococcal antibiotics for cystic fibrosis (Smyth and Walters, 2003).

Study ID (author, date of publication)	Review primary outcomes			Review secondary outcomes			Other study outcomes
	Lung function FEV1	Lung function FVC	Number of people with one or more isolates of S. aureus	Growth	Survival	Quality of life	Serum levels of IgG
Chatfield 1991	o (A)[1]	o (A)[1]	✓	✓	✓	× (H)[2]	✓
Schlesinger 1984	× (H)[3]	× (H)[3]	✓	o (C)[4]	✓	× (H)[2]	✓
Stutman 2002	✓	✓	✓	✓	✓	× (H)[2]	×
Weaver 1994	× (H)[3]	× (H)[3]	✓	o (C)[4]	✓	× (H)[2]	×

1. Reasons for A classifications: 'no significant difference' reported in the text.
2. Reason for H classifications for quality of life: clinical judgement says it is unlikely to have been measured in these trials.
3. Reason for H classifications for lung function tests: both trials involve young children and these tests are not usually carried out on young children.
4. Reason for C classifications for Growth: trial reports give means but no standard deviations and also present the data in a graph.

✓ indicates full reporting of results for treatment comparison of interest
× indicates no reporting
o indicates partial reporting

Table 3: Changes in outcomes between review protocol and publication

		Primary outcomes	Secondary outcomes
Total number of outcomes included in the review (Median, IQR, range)		3 (IQR 2, 3 and range 1,8)	7 (IQR 5, 9 and range 2,13)
Reviews with any discrepancy in outcomes between protocol and full review	Protocol distinguished outcomes (n=37)[1]	14 (38%)	
	Protocol did not distinguish outcomes (n=9)[2]	4 (44%)	
Reviews which have *upgraded* at least one outcome from secondary in the protocol to primary in the full review (number of outcomes; minimum per review; maximum per review)	Protocol distinguished outcomes (n=37)[1]	3 (8%) (3 outcomes)	
	Protocol did not distinguish outcomes (n=9)[2]	0	
Reviews which have *downgraded* at least one outcome from primary in the protocol to secondary in the full review (number of outcomes; minimum per review; maximum per review)	Protocol distinguished outcomes (n=37)[1]	9 (24%) (16 outcomes; min 1, max 5)	
	Protocol did not distinguish outcomes (n=9)[2]	1 (11%) (2 outcomes)	
Reviews which have *added* a new outcome in the full review which was not included in the protocol (number of outcomes; minimum per review; maximum per review)	Protocol distinguished outcomes (n=37)[1]	2 (5%) (3 outcomes)	2 (5%) (4 outcomes; min 1, max 3)
	Protocol did not distinguish outcomes (n=9)[2]	1(11%) (1 outcome)	2 (22%) (2 outcomes)
Reviews which have *excluded* an outcome from the full review which was included in the protocol (number of outcomes; minimum per review; maximum per review)	Protocol distinguished outcomes (n=37)[1]	2 (5%) (10 outcomes; min 1, max 9)	3 (8%) (5 outcomes; min 1; max 2)
	Protocol did not distinguish outcomes (n=9)[2]	0	0

1. Protocol distinguished primary from secondary outcomes
2. Protocol did not distinguish primary from secondary outcomes

Table 4: Risk of bias of RCTs based on review primary outcomes only

		As assessed in review			Total
		High risk	Low risk	Unclear risk/ Not assessed	
As assessed in this study on the primary outcomes of the review only	High risk excluding G	10	18	50	78 (28%)
	High risk (based on G classifications only)	3	17	64	84 (30%)
	Low risk	14	24	80	118 (42%)
Total		27 (10%)	59 (21%)	194 (69%)	280

Note that 'As assessed in this study on the primary outcomes of the review only' is split into three categories: high risk excluding G; high risk (based on G classifications only) and low risk. This is because G classifications, although high risk of bias, are subjective as they are given based on clinical judgment only when there are no details mentioned in the trial report. However, as shown in the original ORBIT study (Kirkham et al, 2010b) the sensitivity and specificity of assigning G and H classifications was high.

Table 5: Risk of bias of RCTs based on review primary and secondary outcomes

		Risk of bias based on review primary outcomes only			Total
		High risk excluding G	High risk (based on G classifications only)	Low risk	
Risk of bias based on review primary and secondary outcomes	High risk excluding G	13	3	13	29 (45%)
	High risk (based on G classifications only)	0	13	18	31 (49%)
	Low risk	0	0	4	4 (6%)
Total		13 (20%)	16 (25%)	35 (55%)	64

Table 6: Risk of bias table for selective outcome reporting.

SELECTIVE OUTCOME REPORTING **Are reports of the study free of suggestion of selective outcome reporting? [Short form: *Free of selective reporting*?]**	
Criteria for a judgement of 'YES' (i.e. low risk of bias).	Any of the following: The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way; The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon).
Criteria for the judgement of 'NO' (i.e. high risk of bias).	Any one of the following: Not all of the study's pre-specified primary outcomes have been reported; One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; The study report fails to include results for a key outcome that would be expected to have been reported for such a study.
Criteria for the judgement of 'UNCLEAR' (uncertain risk of bias).	Insufficient information to permit judgement of 'Yes' or 'No'. It is likely that the majority of studies will fall into this category.

Selective reporting of outcomes in randomised controlled trials in systematic reviews of Cystic Fibrosis

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-002709.R1
Article Type:	Research
Date Submitted by the Author:	02-Apr-2013
Complete List of Authors:	Dwan, Kerry; The University of Liverpool, Biostatistics Kirkham, Jamie; The University of Liverpool, Biostatistics Williamson, Paula; The University of Liverpool, Biostatistics Gamble, Carrol; The University of Liverpool, Biostatistics
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Genetics and genomics, Respiratory medicine, Research methods
Keywords:	STATISTICS & RESEARCH METHODS, MEDICAL EDUCATION & TRAINING, GENETICS

Selective reporting of outcomes in randomised controlled trials in systematic reviews of Cystic Fibrosis

Kerry Dwan[1§], Jamie J Kirkham[1], Paula R Williamson[1], Carrol Gamble[1]

1. Department of Biostatistics, University of Liverpool, Liverpool, England.

[§]Corresponding author

Email addresses:

KD: kerry.dwan@liverpool.ac.uk

JJK: jjk@liverpool.ac.uk

PRW: prw@liverpool.ac.uk

CG: c.gamble@liverpool.ac.uk

Abstract

Background

Outcome reporting bias (ORB) in randomised trials has been identified as a threat to the validity of systematic reviews. Previous work highlighting this problem is limited to considering a single primary review outcome. The aim of this study was to assess ORB across all efficacy outcomes in Cochrane systematic reviews of cystic fibrosis.

Methods

Systematic reviews of interventions for cystic fibrosis published on the Cochrane Library by the Cochrane Cystic Fibrosis and Genetic Disorders Group before 2010 were assessed for discrepancies in outcomes between review protocol and full review. ORB in eligible trials was also assessed for all efficacy review outcomes. Two authors independently classified each outcome using a nine point classification system developed by the ORBIT (Outcome Reporting Bias In Trials) study. These classifications were used to inform the assessment of the risk of bias for selective outcome reporting for each trial.

Results

Forty six Cochrane cystic fibrosis systematic reviews were included. The median number of primary outcomes, number of trials, and participants per trial in the reviews were 3 (IQR 2, 3), 4 (IQR 2,8) and 21 (IQR 14,41) respectively. Eighteen reviews (39%, 18/46) had a discrepancy in outcomes between protocol and full review. Thirty seven reviews were eligible to be included in the ORB assessment. When considering review primary outcomes and all review outcomes, ORB was suspected in at least one trial in 86% and 100% respectively.

Conclusion

Assessment of ORB within a systematic review, of a single primary, outcome underestimates the risk of ORB in comparison to the assessment of multiple primary and secondary outcomes. ORB in trials is highly prevalent within systematic reviews of cystic fibrosis when assessed across all outcomes. This could be reduced by the development of a core outcome set for trials and systematic reviews in cystic fibrosis.

Article summary

Article focus

- Assessment of discrepancies in outcome selection between systematic review protocols and full reviews.
- Assessment of outcome reporting bias at the outcome level across all efficacy systematic review outcomes.
- Assessment of the risk of bias of a trial from selective outcome reporting within a systematic review.

Key messages

- Assessment of ORB within a systematic review, of a single primary, outcome underestimates the risk of ORB in comparison to the assessment of multiple primary and secondary outcomes.Clearer guidance is needed on how to assess the risk of bias as a result of selective outcome reporting for each included trial within a systematic review, when considering multiple outcomes.
- The development of a core outcome set in Cystic Fibrosis would help reduce the problem of ORB.

Strengths and limitations

This is the first study to consider the assessment of outcome reporting bias in all efficacy review outcomes. However, this is limited to reviews of cystic fibrosis.

Background

The value of systematic reviews in establishing an evidence base is widely acknowledged with well conducted systematic reviews of randomised controlled trials being placed at the top of the hierarchy of evidence (Green and Byar, 1984). It is essential, when conducting systematic reviews, to consider the potential for bias and its impact on the review conclusions. Bias may be induced through the decisions and actions of the authors of the included clinical trials or systematic review authors.

Bias in a systematic review is frequently considered in relation to limitations of the search strategy. However, bias may also occur, for example, when outcomes are added, omitted or changed after a systematic review protocol is published if the decision to deviate from the protocol is based on the significance of the results. A study of an unselected cohort of Cochrane reviews revealed that over a fifth (64/288) of protocol/review pairings showed some discrepancy in at least one outcome measure with just 6% (4/64) describing the reason for the change in the review (Kirkham, 2010a). Results also indicated that outcomes promoted from primary to secondary between the protocol and the review were more likely to report statistically significant meta-analysis results in comparison to reviews where there was no discrepancy in outcome specification with the review protocol (relative risk 1.66 95% confidence interval (1.10, 2.49), p = 0.02).

Systematic reviews are only as valid as the trials they contain (Juni et al 2001), consequently much effort is given to assessing the risk of bias within the trials identified by assessing their methodological quality. However, it is also important to consider the content of trial reports in an assessment of bias. Outcome reporting bias (ORB) within a randomised controlled trial (RCT) is defined as the result-based selection of a subset of the original outcomes for

publication (Williamson and Hutton 2000). In a systematic empirical assessment of Cochrane reviews within which a single review primary outcome could be identified (Kirkham et al 2010b), ORB was suspected in at least one randomised controlled trial in more than a third of the systematic reviews that were examined (35%). This study may have underestimated this problem as review primary outcomes are chosen due to their clinical importance and are more likely to have been measured and reported in trials. Therefore, there is concern regarding the prevalence and impact of outcome reporting bias in reviews where multiple primary outcomes are specified, or in secondary outcomes.

Systematic reviews in cystic fibrosis are characterised by inclusion of small randomised trials specifying multiple primary outcomes. Reporting standards for trials of cystic fibrosis have also been shown to be low when comparing trial reports to the Consolidated Standards of Reporting Trails (CONSORT) statement (von Mosch and Dwan 2011). The aims of this current study were to

1. Examine the potential for bias created by review authors by identifying inconsistencies between outcomes published in review protocols and in the associated published reviews
2. Determine the prevalence of ORB in trials in systematic reviews of CF, extending previous work by considering all review efficacy outcomes (multiple primary and secondary).
3. Assess the risk of bias of trials from selective outcome reporting when considering review primary outcomes only in comparison to all review outcomes.

Methods

A cohort of systematic reviews published by the Cochrane Cystic Fibrosis and Genetic Disorders (CFGD) group on the Cochrane Library before 2010 were identified (The Cochrane Library, 2009). Reviews were eligible for inclusion if they compared interventions for cystic fibrosis and identified one or more eligible RCTs. RCTs that had been excluded (in the "characteristics of excluded studies" section) were also checked for any suggestion of outcome reporting bias. For example, if a review had excluded trials as a result of 'no relevant outcome data (NROD)', then these trials were also scrutinised for the presence of ORB and included in the assessment.

Changes in outcomes between systematic review protocol and full review – review level

The numbers of primary and secondary outcomes per review were compared to the recommendations for the number of outcomes (no more than three primary outcomes and a limited number of secondary outcomes) to include in a review in the Cochrane Handbook (Higgins and Green 2011). If a review did not distinguish between primary and secondary outcomes, the first three outcomes listed were taken to be the primary outcomes and the rest were considered as secondary outcomes. Protocols of the systematic reviews were accessed and outcomes stated in the protocol were compared to those stated in the full review. Changes in outcomes were identified and categorised by one author (KD) as: primary outcome downgraded to secondary (*downgrade*); secondary outcome upgraded to primary (*upgrade*); a new outcome not stated in the protocol was added to the full review (*addition*) or an outcome stated in the protocol was omitted from the full review (*omission*). If there had been a change in outcomes, the section 'changes between protocol and review' was examined for a declaration and explanation of the changes.

Assessing trial reports for full ORB – outcome level

For each eligible systematic review, all reports relating to included studies and studies excluded due to no relevant outcome data were obtained. Reviews were checked to see whether review authors had contacted trialists for further information or data for outcomes. Where this was not clear in the review, review authors were asked to clarify.

A nine-point classification system (Table 1) developed for missing or incomplete outcome reporting in randomised trials (Kirkham et al, 2010b) was used to make an assessment of the risk of bias. Table 1 also provides examples of outcomes that were not assessed because they had poor outcome definitions. An outcome matrix (Table 2) was created for each review using the ORBIT matrix generator (http://ctrc.liv.ac.uk/orbit/), with studies listed in the rows and review primary and secondary outcomes listed in the columns with the ORBIT classifications (Table 1) given for each review outcome that was not fully reported (e.g. not reported or partially reported e.g. $p>0.05$).

The outcomes listed or detailed in the method section and the outcomes reported in the results section were compared for all trial publications to determine whether each outcome of the systematic review was measured and analysed. In some instances it may be obvious that an outcome was measured given the other outcomes reported. For example, if cause-specific mortality is reported then overall mortality must have been measured, even if not reported. In other situations it may be that a battery of tests or measurements are usually undertaken together, for example FEV1 (forced expiratory volume in 1 second), FVC (forced vital capacity) and FEF_{25-75} (average expired flow over the middle half of the FVC manoeuvre). FEV1 is the outcome most often considered for lung function due to its validity, repeatability and it is the outcome most understood by clinicians. However, the device used to measure FEV1 also measures the majority of other lung function outcomes. Therefore if FEV1 was reported in a trial, it was assumed that other lung function outcomes were also measured but

not necessarily analysed (classification F) unless they were specifically stated as an outcome in the trial report. However, if FEV1 was not reported but other lung function outcomes were then an E classification was given to FEV1 as suspicion would be raised that the latter may have been selectively not reported. This was decided after discussion with clinical experts.

However, it is often difficult to assess whether an outcome was measured, and clinical judgment is required. The clinical lead for each review was contacted by email and asked for their input into the assessment of selective outcome reporting within the trials included in their review. An assessment of whether the review outcomes had been measured and reported within each trial using the classification system was completed. The clinical lead for the review and KD independently assessed the trials in the review and any disagreements were resolved through discussion and then checked with a third person (JJK or PRW).

Assessment of risk of bias for selective outcome reporting – trial level

If one or more of the outcomes for a trial was given a high risk classification according to Table 1, the trial was deemed at high risk of bias from selective reporting.

Analysis

Descriptive results are presented. The median and interquartile range for the number of review primary and secondary outcomes was calculated.

Data are tabulated and excerpts found in the trial reports relating to review outcomes are used to support decisions made regarding ORBIT classifications and the assessment of risk of bias.

Results

The CFGD group had 46 cystic fibrosis systematic reviews published as of 2010.

1. Changes in outcomes between systematic review protocol and full review – review level

Protocols were available for all 46 systematic reviews. Nine protocols (20%) did not distinguish between primary and secondary outcomes. Table 3 shows the median number of primary and secondary outcomes for the 46 reviews and the changes in outcomes between protocol and full review.

Eighteen reviews (39%, 18/46) had a discrepancy in outcomes between protocol and full review. Between review protocol and full review, five (28%) listed all changes, two (11%) listed some changes and 11 reviews (61%) did not mention any change in outcomes. Of the seven reviews describing the changes between protocol and full review, three provided no reason for the changes, two stated that the changes in recommendations in the Cochrane Handbook to have a maximum of three primary outcomes were the reason for downgrading outcomes and two reviews stated that they added clinically relevant outcomes that were discovered during the review process.

2. Assessing trial reports for full ORB – outcome level

Of the 46 published reviews, 38 were eligible to be assessed for outcome reporting bias (Figure 1).

One review was excluded at this stage as the outcomes could not be assessed for ORB due to the different ways the outcome definitions could be measured and reported. The primary outcomes were psychosocial outcomes, which included any objective measure with adequate

psychometric properties and demonstrable reliability and validity quantifying psychological or social outcomes or both, including individual psychological adjustment, relational, social functioning and adaptation to life with cystic fibrosis.

Therefore 37 reviews were assessed for ORB, including 280 RCTs (278 included and 2 excluded due to no relevant outcome data but confirmed by review authors that they would have otherwise been included). The median number of trials per review was four (IQR 2, 8) and there was a median sample size of 21 (IQR 14, 41) per trial.

Review authors contacted trialists for missing outcome data in 33 reviews (89%), one stated that "trialists were not contacted but would be in updates of the review" and three reviews did not state if trialists were contacted for further data.

The lead authors of twelve reviews assessed the included trials and gave classifications for each outcome. For thirteen reviews, authors gave input on which outcomes they expected to be measured for trials in their review and which outcomes they expected to be measured in routine clinical practice but did not classify each outcome due to time restrictions. The authors of twelve reviews did not respond to our request.

For the twelve reviews where the authors assigned classifications, discussion was needed on all outcomes to come to an agreed classification. For the other 25 reviews it was difficult to assign a classification to all outcomes as some outcomes needed a large amount of clinical input in understanding the outcome and language used to describe the outcomes within the trial reports. Due to the number and complexity of outcomes and lack of reviewer input on the majority of reviews, it was decided that the assessment of all primary outcomes listed in

the full review that were well defined should take priority. Many outcomes were also split into sub outcomes or ill defined to maximise the ability of a trial to contribute data to the review.

The ORBIT classifications for the review primary outcomes for the 280 RCTs are shown in Table 1. For the 12 reviews where reviewer input was obtained, classifications for 64 included trials for review secondary outcomes are also shown in Table 1. Eligible trials within the reviews fully reported 383 (33.7%) review primary outcomes and 125 (18.7%) review secondary outcomes. In addition to the classifications in Table 1, a 'G, no events' classification (For example, mortality, where clinical judgement says it is likely to have been measured and it would have been reported had any deaths occurred. Therefore, it is assumed no deaths occurred during the trial.) was given to eligible trials within the reviews for 109 (9.5%) review primary outcomes and 22 (3.3%) review secondary outcomes. Due to limited reviewer input or the lack of a standard definition, we were unable to assess outcomes (including: adverse events, symptoms, complications, biochemical measures of glycaemic control, symptoms of sleep disordered breathing and measures of specific indices of strength, mass, effort and general fatigue) for 102 trials for review primary outcomes and 59 trials for review secondary outcomes.

Assessment of risk of bias from selective outcome reporting – trial level

Eighteen reviews (49%) had not yet assessed the risk of bias for selective outcome reporting as although the Cochrane guidance on the risk of bias was introduced in 2008 and the cut off for this study was the beginning of 2010, these reviews were still to be updated. Seventeen reviews (46%) had assessed the risk of bias for all included trials and two reviews (5%) assessed this for some of their included trials.

As we were unable to assess secondary outcomes for ORB for all reviews, the risk of bias assessments were made based on classifications of primary outcomes in order to be consistent across reviews. Only five (14%) of the 37 reviews had no trials at high risk of bias based on the review primary outcomes only. Table 4 shows the risk of bias for selective outcome reporting as defined in this study and also as assessed within the published reviews for the 280 trials assessed for ORB based on the consideration of review primary outcomes only. It was found that 69% of trials had either not been assessed for selective reporting or were assessed as an unclear risk.

Table 5 shows the risk of bias for selective outcome reporting based on the consideration of review primary and secondary outcomes separately for the 12 reviews (64 trials) where reviewers also provided classifications. This was to see if decisions regarding risk of bias would change if we considered all outcomes. Only four (6%) of the 64 RCTs had a low risk of bias when considering all outcomes.

Discrepancies in the risk of bias when considering all outcomes arose in 34 (53%) trials; 31 were at low risk when considering review primary outcomes only but high risk of bias (excluding G classifications: 13, G classification only: 18) when considering all outcomes; 3 were at high risk (G classifications only) when considering review primary outcomes only but high risk (excluding G classifications) when considering all outcomes. This often occurred in reviews were there was only one or two primary outcomes and a large number of secondary outcomes.

Based on all review outcomes, none of the 12 reviews had all included trials at low risk of bias.

Discussion

This is the first study to consider all review efficacy outcomes in an ORB assessment which has allowed us to make practical recommendations on assessing the risk of bias of selective reporting for systematic reviews at both the review and trial level. Over a third of Cochrane cystic fibrosis reviews (39%) examined had a discrepancy in outcomes between the review protocol and full review. This compares to 22% of reviews (64/288) that contained a discrepancy in at least one outcome measure in the main ORBIT study which looked at reviews covering all 50 Cochrane review groups (Kirkham et al 2010a). However, this is confounded by the different publication date ranges of the reviews (assessed as up to date between 2006 and 2009). Furthermore, for the cystic fibrosis reviews outcome reporting bias was suspected in at least one randomised controlled trial in 86% of reviews when considering all review primary outcomes. The prevalence of reviews containing at least one trial with high suspicion of outcome reporting bias from ORBIT, when only a single primary outcome was considered was substantially lower at 34% (96/283) (Kirkham et al 2010b). While this study is limited only to CF trials, it is clear that the problem of outcome reporting bias is much larger when considering more than just the single primary review outcome of importance that was used in the ORBIT study.

A study by von Mosch and Dwan (2011) that compared the reporting in trial reports of CF to the CONSORT statement found that from a maximum of 57 points available, the scores rose from a median of 17.5 (Inter quartile range (IQR) 15.5-24.5) in 1994 to a median of 32 (IQR

22.8-41.5) in 2008. Along with the current study, this also indicates that there is still room for an improvement in the reporting of outcomes.

Use of the ORBIT classification system offered a robust methodology for assessing the risk of bias for trials included within a systematic review. When considering the 64 trials in the 12 reviews where it was possible to assess both primary and secondary outcomes, when basing the risk of bias assessment on review primary outcomes, 45% of trials were at high risk of bias and when using all outcomes in the assessment, 94% were at high risk of bias. Using the current selective reporting item of the current Cochrane risk of bias tool, 69% of trials included in CF reviews were assessed by reviewers as 'unclear' risk of bias or not assessed at all, indicating the need for more informed guidance on assigning risk of bias in the systematic review process for all outcomes within a review.

The ORBIT classification system has already been validated as part of the original project. Sensitivity results for predicting that the outcome had been measured (G-classification) was 92% (23/25, 95% CI 81% to 100%), while the specificity for predicting that the outcome had not been measured (H-classification) was 77% (23/30, 95% CI 62% to 92%). With the additional requirement to assess all outcomes in this project, there were an increasing number of outcomes that were not mentioned in the trials reports and therefore clinical judgement was needed as to whether the outcome of interest was likely to have been measured in a particular trial. Many review authors did not respond to our request to provide classifications (68%), but for those with no response we did obtain clinical input for the primary outcome from within the CFGD group. Although we can not exclude the possibility of response bias it is likely the decision to respond was influenced by time commitments rather than review

characteristics. However, these assessments will be provided to the review authors when their review is due to be updated.

Reviewers should ensure that changes between protocol and reviews are listed and justifications provided to enhance the validity of these decisions. Eligible trials should not be excluded on the basis of "No relevant outcome data" because although an outcome was not reported it may have been measured and contact with the authors is advised. Reviewers should be encouraged to consider trials that have not reported an outcome of interest and to assess whether selective reporting has occurred for all review outcomes. They should consider the amount of missing data from their meta-analysis (i.e. the percentage of the sample sizes of the studies that were included compared to those that would have been eligible to be included in the meta-analysis but no outcome data reported) and this information should be included along with the pooled effect estimate. If appropriate, a sensitivity analysis should be applied to assess the robustness of the conclusions of the review, such as an imputation approach (Williamson and Gamble 2005), the Copas bound for maximum bias (Copas et al, 2004; Williamson and Gamble 2007, Dwan et al 2010) or a model based correction (Copas et al, 2013).

Individuals conducting systematic reviews need to address explicitly the issue of missing outcome data for their review to be considered a reliable source of evidence. Extra care is required during data extraction, reviewers should identify when a trial reports that an outcome was measured but no results were reported or events observed, and contact with trialists should be encouraged. Contacting authors is encouraged by the CRG and is standard practice within CFGD reviews which is reflected in our results as 89% of reviews stated that they contacted authors for extra information on outcomes.

It is recommended that review authors ensure that they limit the number of outcomes in the review and define them clearly as this will allow easier assessments of selective reporting, which can be done during data extraction of the included trials as long as a knowledgeable clinical person is involved. Lung function was specified as the first primary outcome in nineteen reviews (50%), as the second or third primary outcome in 11 reviews (29%), as a secondary outcome in six reviews (16%) and it was not included as an outcome in only one review (5%). However, as discussed earlier, lung function can be measured in different ways (FEV1, FVC, mid forced expiratory flow (FEF), peak expiratory flow rate (PEFR), residual volume (RV), total lung capacity (TLC), Lung clearance index (LCI) and maximum expiratory flow (MEF).) These outcomes can then be analysed and reported in different ways such as**:** percentage predicted, litres, litres/second and post treatment, absolute change from baseline, relative change from baseline or annual rate of change. Therefore there is a large scope for selective reporting. One solution is the development of a core outcome set for cystic fibrosis (Ramsey and Boat 1994 , Sinha et al 2008, Clarke 2007).

Unanswered questions and future research

Work is needed to consider what the best method is to assess the impact of ORB on the results of the meta-analysis when there are multiple outcomes. Multivariate meta-analysis has been suggested by Kirkham et al 2012 and a model based correction has been suggested by Copas et al (2013).

Conclusion

Systematic reviews need to clearly state the primary and secondary outcomes that they will consider and be consistent between review protocol and full review.

Outcome reporting bias is a major problem for systematic reviews and more guidance needs to be included in the Cochrane handbook to allow assessment of this important item within the risk of bias tool. We recommend that an outcome matrix be completed during the production of a review to allow an ORB assessment for all review outcomes which can then inform the risk of bias assessment.

A core set of outcomes should be agreed upon for cystic fibrosis which in turn will have a positive impact on systematic reviews as future trials are conducted they should specifically set out to measure and report these outcomes therefore reducing the prevalence of selective reporting.

Abbreviations

CFGD	Cystic Fibrosis and Genetic Disorders
CRG	Cochrane Review Group
COMET	Core Outcome Measures in Effectiveness Trials
CONSORT	Consolidated Standards of Reporting Trails
FEF_{25-75}	Average expired flow over the middle half of the FVC manoeuvre
FEV1	Forced expiratory volume in 1 second
FVC	Forced vital capacity
NROD	No Relevant Outcome Data
ORB	Outcome Reporting Bias
ORBIT	Outcome Reporting Bias In Trials
PEFR	Peak Expiratory Flow Rate
RCT	Randomised Controlled Trial

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing Interests

There are no financial or non-financial competing interests.

Authors' contributions

KD drafted the protocol, completed the ORB assessments and wrote the manuscript.

JJK completed the ORB assessments and commented on the manuscript.

CG commented on the manuscript.

PRW commented on the protocol and the manuscript and commented on the ORB assessments.

Acknowledgements

The authors would like to thank the Cochrane CFGD group reviewers for their assistance with clinical input for the ORB assessments of the included trials in their reviews.

References

Clarke, M. Standardising Outcomes for Clinical Trials and Systematic Reviews. *Trials; 2007,* 8**,** 39.

Copas J, Jackson D. A bound for publication bias based on the fraction 14 of unpublished studies. *Biometrics* 2004;60:146-53.

Copas J, Dwan K, Kirkham JJ, Williamson PR. A model based correction for outcome reporting bias in meta-analysis. (submitted 2013).

Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, et al. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. PLoS ONE 2008, 3(8): e3081.

Dwan K, Gamble C, Kolamunnage-Dona R, Mohammed S, Powell C, Williamson PR. Assessing the potential for outcome reporting bias in a review: A tutorial. *Trials 2010,* **11:**52 doi:10.1186/1745-6215-11-52.

Green, S. B. & Byar, D. P. Using observational data from registries to compare treatments: the fallacy of omnimetrics. *Stat Med, 1984;* 3**,** 361-73.

Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001; 323: 42-46.

Kirkham JJ, Altman DG, Williamson PR. Bias Due to Changes in Specified Outcomes during the Systematic Review Process. PLoS ONE 2010a; 5(3): e9810.doi:10.1371/journal.pone.0009810

Kirkham JJ, Dwan K, Dodd S, Altman DG, Smyth R, Jacoby A, Gamble C, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ 2010b; 340: c365

Kirkham JJ, Riley RD and Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. Statistics in Medicine 2012*; 31:20, 2179-2195.*

Ramsey BW, Boat TF (1994) Outcome measures for clinical trials in cystic fibrosis. Summary of a Cystic Fibrosis Foundation consensus conference. J Pediatr 1994; 124: 177–192.

Sinha I, Jones L, Smyth RL, Williamson PR. A Systematic Review of Studies That Aim to Determine Which Outcomes to Measure in Clinical Trials in Children . PLoS Med 2008; 5(4): e96. doi:10.1371/journal.pmed.0050096

Smyth AR,Walters S. Prophylactic anti-staphylococcal antibiotics for cystic fibrosis. *Cochrane Database of Systematic Reviews* 2003, Issue 3. Art. No.: CD001912. DOI: 10.1002/14651858.CD001912.

The Cochrane Library, Issue 4, 2009. Chichester: Wiley.

Williamson, P. R., Gamble, C., Altman, D. G. & Hutton, J. L. Outcome selection bias in meta-analysis. *Stat Methods Med Res, 2005;* 14**,** 515-24.

Williamson PR, Gamble C: Identification and impact of outcome selection bias in meta-analysis. Statistics in Medicine 2005, 24(10):1547-61.

Williamson PR, Gamble C: Application and investigation of a bound for outcome reporting bias. Trials 2007, 8:9.

von Mosch S, Dwan K. Reporting of harm and efficacy outcomes of randomised clinical trials in cystic fibrosis before and after the publication of the revised CONSORT (CONsolidated Standards Of Reporting Trials) statement in 2001 – a comparison of trial reports published from 1994 and from 2008. The University of Liverpool (thesis) 2011.

Figure 1: Study flow diagram

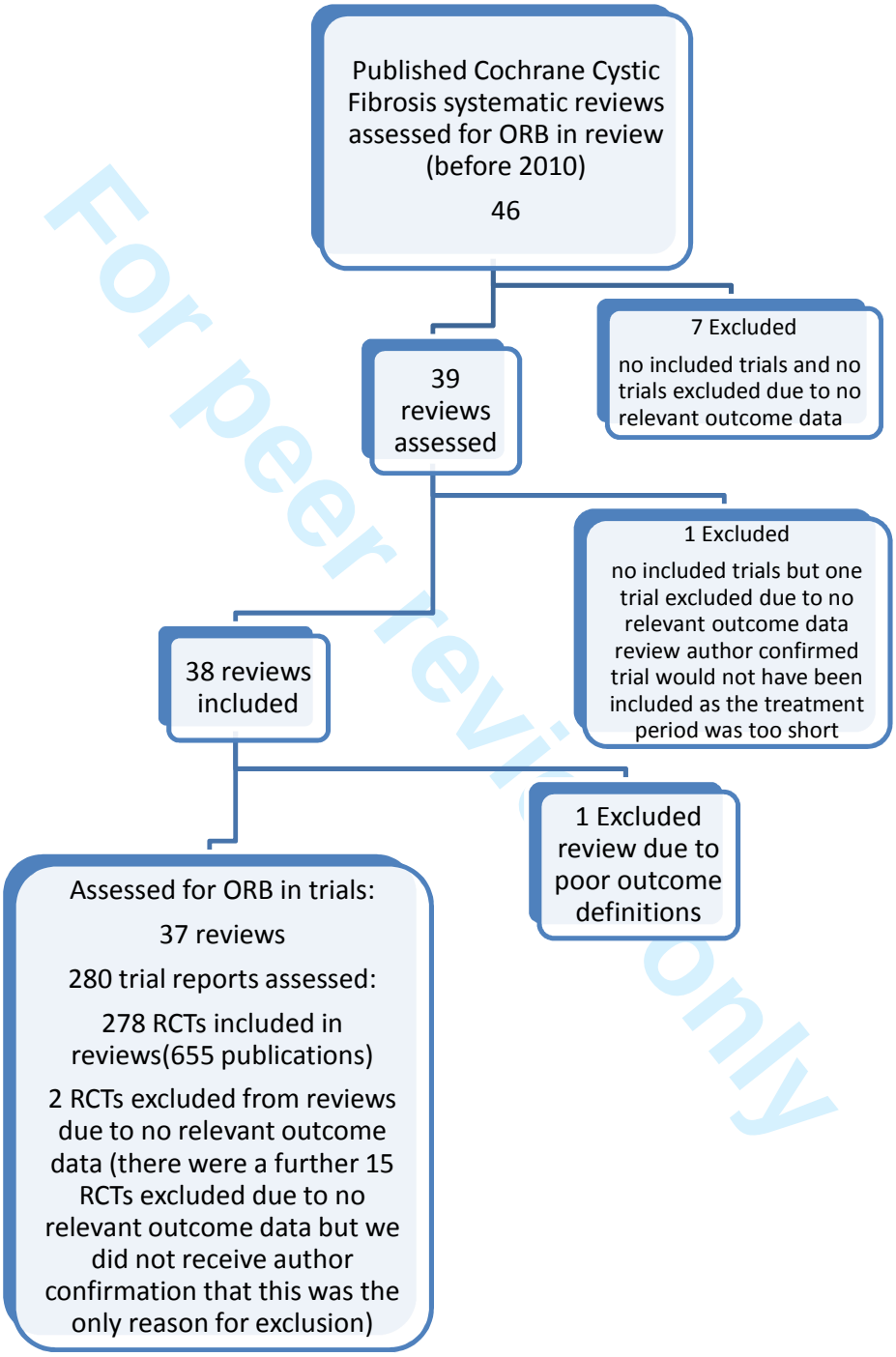


Table 1: ORBIT classifications

Classification	Description	Level of reporting	Level of suspicion of ORB	Primary outcome classifications	Secondary outcome classifications
				Number of trials (percentage overall)[1]	Number of trials (percentage overall)[2]
Clear that the outcome was measured and analysed					
A	States outcome analysed but only reported that result not significant (typically stating p-value >0.05).	Partial	High risk	75 (6.6%)	12 (1.8%)
B	States outcome analysed but only reported that result significant (typically stating p-value <0.05).	Partial	Low risk	13 (1.1%)	2 (0.3%)
C	States outcome analysed but insufficient data presented to be included in meta-analysis or to be considered to be fully tabulated.	Partial	Low risk	53 (4.7%)	15 (2.2%)
D	States outcome analysed but no results reported.	None	High risk	0 (0)	0 (0)
Clear that the outcome was measured					
E	Clear that outcome was measured but not necessarily analysed.	None	High risk	59 (5.2%)	26 (3.9%)
F	Clear that outcome was measured but not necessarily analysed.	None	Low risk	110 (9.7%)	15 (2.2%)
Unclear that the outcome was measured					
G	Not mentioned but clinical judgment says likely to have been measured and analysed.	None	High risk	195 (17.1%)	197 (29.4%)
H	Not mentioned but clinical judgment says unlikely to have been measured.	None	Low risk	141 (12.4%)	256 (38.2%)
Clear that the outcome was NOT measured					
I	Clear that outcome was not measured.	N/A	No risk	0 (0)	0 (0)

The ORBIT classifications for review primary outcomes for the 280 RCTs. For the 12 reviews where reviewer input was obtained, classifications for 64 included trials for review secondary outcomes are also shown.

1. The denominator used is the total number of data points expected if all 280 eligible trials reported on all review primary outcomes in the 37 reviews (i.e. the number of review primary outcomes multiplied by the number of trials within the review for all reviews. This does not include the 102 trials where we were unable to assess primary outcomes).

2. The denominator used is the total number of data points expected if all 64 trials reported on all review secondary outcomes in the 12 reviews (i.e. the number of review secondary outcomes multiplied by the number of trials within the review for all reviews. This does not include the 59 trials where we were unable to assess secondary outcomes).

Table 2: Example of review outcome matrix for 6 of 17 outcomes in a review of Prophylactic anti-staphylococcal antibiotics for cystic fibrosis (Smyth and Walters, 2003).

Study ID (author, date of publication)	Review primary outcomes			Review secondary outcomes			Other study outcomes
	Lung function FEV1	Lung function FVC	Number of people with one or more isolates of S. aureus	Growth	Survival	Quality of life	Serum levels of IgG
Chatfield 1991	o (A)[1]	o (A)[1]	✓	✓	✓	× (H)[2]	✓
Schlesinger 1984	× (H)[3]	× (H)[3]	✓	o (C)[4]	✓	× (H)[2]	✓
Stutman 2002	✓	✓	✓	✓	✓	× (H)[2]	×
Weaver 1994	× (H)[3]	× (H)[3]	✓	o (C)[4]	✓	× (H)[2]	×

1. Reasons for A classifications: 'no significant difference' reported in the text.
2. Reason for H classifications for quality of life: clinical judgement says it is unlikely to have been measured in these trials.
3. Reason for H classifications for lung function tests: both trials involve young children and these tests are not usually carried out on young children.
4. Reason for C classifications for Growth: trial reports give means but no standard deviations and also present the data in a graph.

✓ indicates full reporting of results for treatment comparison of interest
× indicates no reporting
o indicates partial reporting

Table 3: Changes in outcomes between review protocol and publication

		Primary outcomes	Secondary outcomes
Total number of outcomes included in the review (Median, IQR, range)		3 (IQR 2, 3 and range 1,8)	7 (IQR 5, 9 and range 2,13)
Reviews with any discrepancy in outcomes between protocol and full review	Protocol distinguished outcomes (n=37)[1]	14 (38%)	
	Protocol did not distinguish outcomes (n=9)[2]	4 (44%)	
Reviews which have *upgraded* at least one outcome from secondary in the protocol to primary in the full review (number of outcomes; minimum per review; maximum per review)	Protocol distinguished outcomes (n=37)[1]	3 (8%) (3 outcomes)	
	Protocol did not distinguish outcomes (n=9)[2]	0	
Reviews which have *downgraded* at least one outcome from primary in the protocol to secondary in the full review (number of outcomes; minimum per review; maximum per review)	Protocol distinguished outcomes (n=37)[1]	9 (24%) (16 outcomes; min 1, max 5)	
	Protocol did not distinguish outcomes (n=9)[2]	1 (11%) (2 outcomes)	
Reviews which have *added* a new outcome in the full review which was not included in the protocol (number of outcomes; minimum per review; maximum per review)	Protocol distinguished outcomes (n=37)[1]	2 (5%) (3 outcomes)	2 (5%) (4 outcomes; min 1, max 3)
	Protocol did not distinguish outcomes (n=9)[2]	1(11%) (1 outcome)	2 (22%) (2 outcomes)
Reviews which have *excluded* an outcome from the full review which was included in the protocol (number of outcomes; minimum per review; maximum per review)	Protocol distinguished outcomes (n=37)[1]	2 (5%) (10 outcomes; min 1, max 9)	3 (8%) (5 outcomes; min 1; max 2)
	Protocol did not distinguish outcomes (n=9)[2]	0	0

1. Protocol distinguished primary from secondary outcomes
2. Protocol did not distinguish primary from secondary outcomes

Table 4: Risk of bias of RCTs based on review primary outcomes only

		As assessed in review			Total
		High risk	Low risk	Unclear risk/ Not assessed	
As assessed in this study on the primary outcomes of the review only	High risk excluding G	10	18	50	78 (28%)
	High risk (based on G classifications only)	3	17	64	84 (30%)
	Low risk	14	24	80	118 (42%)
Total		27 (10%)	59 (21%)	194 (69%)	280

Note that 'As assessed in this study on the primary outcomes of the review only' is split into three categories: high risk excluding G; high risk (based on G classifications only) and low risk. This is because G classifications, although high risk of bias, are subjective as they are given based on clinical judgment only when there are no details mentioned in the trial report. However, as shown in the original ORBIT study (Kirkham et al, 2010b) the sensitivity and specificity of assigning G and H classifications was high.

Table 5: Risk of bias of RCTs based on review primary and secondary outcomes

		Risk of bias based on review primary outcomes only			Total
		High risk excluding G	High risk (based on G classifications only)	Low risk	
Risk of bias based on review primary and secondary outcomes	High risk excluding G	13	3	13	29 (45%)
	High risk (based on G classifications only)	0	13	18	31 (49%)
	Low risk	0	0	4	4 (6%)
Total		13 (20%)	16 (25%)	35 (55%)	64

Table 6: Risk of bias table for selective outcome reporting.

SELECTIVE OUTCOME REPORTING **Are reports of the study free of suggestion of selective outcome reporting? [Short form: *Free of selective reporting*?]**	
Criteria for a judgement of 'YES' (i.e. low risk of bias).	Any of the following: The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way; The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon).
Criteria for the judgement of 'NO' (i.e. high risk of bias).	Any one of the following: Not all of the study's pre-specified primary outcomes have been reported; One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; The study report fails to include results for a key outcome that would be expected to have been reported for such a study.
Criteria for the judgement of 'UNCLEAR' (uncertain risk of bias).	Insufficient information to permit judgement of 'Yes' or 'No'. It is likely that the majority of studies will fall into this category.

Selective reporting of outcomes in randomised controlled trials in~~and~~ systematic reviews of Cystic Fibrosis

Kerry Dwan[1§], Jamie J Kirkham[1], Paula R Williamson[1], Carrol Gamble[1]

1. Department of Biostatistics, University of Liverpool, Liverpool, England.

[§]Corresponding author

Email addresses:

KD: kerry.dwan@liverpool.ac.uk

JJK: jjk@liverpool.ac.uk

PRW: prw@liverpool.ac.uk

CG: c.gamble@liverpool.ac.uk

Abstract

Background

Outcome reporting bias (ORB) in randomised trials ~~and systematic reviews~~ has been identified as a threat to the validity of systematic reviews. Previous work highlighting this problem is limited to considering a single primary review outcome. ~~Cochrane Cystic fibrosis systematic reviews are often characterised by inclusion of small randomised trials and specify multiple review primary outcomes increasing concern about ORB.~~ The aim of this study was to assess ORB across all efficacy outcomes in Cochrane systematic reviews of cystic fibrosis.

Methods

Systematic reviews of interventions for cystic fibrosis published on the Cochrane Library by the Cochrane Cystic Fibrosis and Genetic Disorders Group before 2010 were assessed for ~~differences~~ discrepancies in outcomes between review protocol and full review. ORB in eligible trials was also assessed for all efficacy review outcomes. Two authors independently classified each outcome using a nine point classification system developed by the ORBIT (Outcome Reporting Bias In Trials) study. These classifications were used to inform the assessment of the risk of bias for selective outcome reporting for each trial.

Results

Forty six Cochrane cystic fibrosis systematic reviews were included. The median number of primary outcomes, number of trials, and participants per trial in the reviews were 3 (IQR 2, 3), 4 (IQR 2,8) and 21 (IQR 14,41) respectively. Eighteen reviews (39%, 18/46) had a discrepancy in outcomes between protocol and full review. Thirty seven reviews were eligible to be included in the ORB assessment. When considering review primary outcomes and all review outcomes, ~~outcome reporting bias~~ORB was suspected in at least one trial in 86% and 100% respectively.

Conclusion

Assessment of ORB within a systematic review, of a single primary, outcome underestimates the risk of ORB in comparison to ~~When considering~~ the assessment of ~~Systematic reviews with~~ multiple primary and secondary ~~efficacy~~ review outcomes. ~~, systematic reviews are at greater risk of ORB than when considering a single review primary outcome.~~

ORB in trials is highly prevalent within systematic reviews of cystic fibrosis when assessed across all outcomes. This could be reduced by the development of a core outcome set for trials and systematic reviews in cystic fibrosis.

Article summary

Article focus

- Assessment of discrepancies in outcome selection between systematic review protocols and full reviews.
- Assessment of outcome reporting bias at the outcome level across all efficacy systematic review outcomes.
- Assessment of the ~~overall~~ risk of bias of a trial from selective outcome reporting ~~of outcomes of a trial~~ within a systematic review.

Key messages

- Assessment of ORB within a systematic review, of a single primary, outcome underestimates the risk of ORB in comparison to the assessment of multiple primary and secondary outcomes. ~~When considering multiple primary and secondary efficacy review outcomes, systematic reviews are at greater risk of ORB than when~~

~~considering a single review primary outcome.~~Systematic reviews with multiple primary and secondary efficacy outcomes are at greater risk of ORB.

- Clearer guidance is needed on how to assess the ~~'overall'~~ risk of bias as a result of ~~ORB~~ selective outcome reporting for each included trial within a systematic review, when considering multiple outcomes.
- The development of a core outcome set in Cystic Fibrosis would help reduce the problem of ORB.

Strengths and limitations

This is the first study to consider the assessment of outcome reporting bias in all efficacy review outcomes. However, this is limited to reviews of cystic fibrosis.

Background

The value of systematic reviews in establishing an evidence base is widely acknowledged with well conducted systematic reviews of randomised controlled trials being placed at the top of the hierarchy of evidence (Green and Byar, 1984). It is essential, when conducting systematic reviews, to consider the potential for bias and its impact on the review conclusions. Bias may be induced through the decisions and actions of the authors of the included clinical trials or systematic review authors.

Bias in a systematic review is frequently considered in relation to limitations of the search strategy. However, bias may also occur, for example, when outcomes are added, omitted or changed after a systematic review protocol is published if the decision to deviate from the protocol is based on the significance of the results. A study of an unselected cohort of Cochrane reviews revealed that over a fifth (64/288) of protocol/review pairings showed some discrepancy in at least one outcome measure with just 6% (4/64) describing the reason for the change in the review (Kirkham, 2010a). Results also indicated that outcomes promoted from primary to secondary between the protocol and the review were more likely to report statistically significant meta-analysis results in comparison to reviews where there was no discrepancy in outcome specification with the review protocol (relative risk 1.66 95% confidence interval (1.10, 2.49), p = 0.02).

Systematic reviews are only as valid as the trials they contain (Juni et al 2001), consequently much effort is given to assessing the risk of bias within the trials identified by assessing their methodological quality. However, it is also important to consider the content of trial reports in an assessment of bias. Outcome reporting bias (ORB) within a randomised controlled trial (RCT) is defined as the result-based selection of a subset of the original outcomes for

publication (Williamson and Hutton 2000). In a systematic empirical assessment of Cochrane reviews within which a single review primary outcome could be identified (Kirkham et al 2010b), ORB was suspected in at least one randomised controlled trial in more than a third of the systematic reviews that were examined (35%). This study may have underestimated this problem as review primary outcomes are chosen due to their clinical importance ~~so~~ and are more likely to have been measured and reported in trials. Therefore, there is ~~increasing~~ concern regarding the prevalence and impact of outcome reporting bias in reviews where multiple primary outcomes are specified, or in secondary outcomes.

Systematic reviews in cystic fibrosis are characterised by inclusion of small randomised trials specifying multiple primary outcomes. Reporting standards for trials of cystic fibrosis have also been shown to be low when comparing trial reports to the Consolidated Standards of Reporting Trails (CONSORT) statement (von Mosch and Dwan 2011). The aims of this current study were to

1. Examine the potential for bias created by review authors by identifying inconsistencies between outcomes published in review protocols and in the associated published reviews
2. Determine the prevalence of ORB in trials in systematic reviews of CF, extending previous work by considering all review efficacy outcomes (multiple primary and secondary).
~~2.~~3. Assess the risk of bias of trials from selective outcome reporting when considering review primary outcomes only in comparison to all review outcomes.

Methods

A cohort of systematic reviews published by the Cochrane Cystic Fibrosis and Genetic Disorders (CFGD) group on the Cochrane Library before 2010 were identified (The Cochrane Library, 2009). Reviews were eligible for inclusion if they compared interventions for cystic fibrosis and identified one or more eligible RCTs. RCTs that had been excluded (in the "characteristics of excluded studies" section) were also checked for any suggestion of outcome reporting bias. For example, if a review had excluded trials as a result of 'no relevant outcome data (NROD)', then these trials were also scrutinised for the presence of ORB and included in the assessment.

Changes in outcomes between systematic review protocol and full review – review level

The numbers of primary and secondary outcomes per review were compared to the recommendations for the number of outcomes (no more than three primary outcomes and a limited number of secondary outcomes) to include in a review in the Cochrane Handbook (Higgins and Green 2011). If a review did not distinguish between primary and secondary outcomes, the first three outcomes listed were taken to be the primary outcomes and the rest were considered as secondary outcomes. Protocols of the systematic reviews were accessed and outcomes stated in the protocol were compared to those stated in the full review. Changes in outcomes were identified and categorised by one author (KD) as: primary outcome downgraded to secondary (*downgrade*); secondary outcome upgraded to primary (*upgrade*); a new outcome not stated in the protocol was added to the full review (*addition*) or an outcome stated in the protocol was omitted from the full review (*omission*). If there had been a change in outcomes, the section 'changes between protocol and review' was examined for a declaration and explanation of the changes.

Assessing trial reports for full ORB – outcome level

For each eligible systematic review, all reports relating to included studies and studies excluded due to no relevant outcome data were obtained. Reviews were checked to see whether review authors had contacted trialists for further information or data for outcomes. Where this was not clear in the review, review authors were asked to clarify.

A nine-point classification system (Table 1) developed for missing or incomplete outcome reporting in randomised trials (Kirkham et al, 2010b) was used to make an assessment of the risk of bias. Table 1 also provides examples of outcomes that were not assessed because they had poor outcome definitions. An outcome matrix (Table 2) was created for each review using the ORBIT matrix generator (http://ctrc.liv.ac.uk/orbit/), with studies listed in the rows and review primary and secondary outcomes listed in the columns with the ORBIT classifications (Table 1) given for each review outcome that was not fully reported (e.g. not reported or partially reported e.g. $p>0.05$).

The outcomes listed or detailed in the method section and the outcomes reported in the results section were compared for all trial publications to determine whether each outcome of the systematic review was measured and analysed. In some instances it may be obvious that an outcome was measured given the other outcomes reported. For example, if cause-specific mortality is reported then overall mortality must have been measured, even if not reported. In other situations it may be that a battery of tests or measurements are usually undertaken together, for example FEV1 (forced expiratory volume in 1 second) ~~and,~~ FVC (forced vital capacity) and FEF_{25-75} (average expired flow over the middle half of the FVC manoeuvre). ~~If FVC is reported but FEV1 is not, suspicion should be raised that the latter may have been selectively not reported.~~ FEV1 is the outcome most often considered for lung function due to its validity, repeatability and it is the outcome most understood by clinicians. However, the device used to measure FEV1 also measures the majority of other lung function outcomes.

Therefore if FEV1 was reported in a trial, it was assumed that other lung function outcomes were also measured but not necessarily analysed (classification F) unless they were specifically stated as an outcome in the trial report. However, if FEV1 was not reported but other lung function outcomes were then an E classification was given to FEV1 as suspicion would be raised that the latter may have been selectively not reported. This was decided after discussion with clinical experts.

However, it is often difficult to assess whether an outcome was measured, and clinical judgment is required. The clinical lead for each review was contacted by email and asked for their input into the assessment of selective outcome reporting within the trials included in their review. An assessment of whether the review outcomes had been measured and reported within each trial using the classification system was completed. The clinical lead for the review and KD independently assessed the trials in the review and any disagreements were resolved through discussion and then checked with a third person (JJK or PRW).

Assessment of risk of bias for selective outcome reporting – trial level

If one or more of the outcomes for a trial was given a high risk classification according to Table 1, the trial was deemed at high risk of bias from selective reporting.

Analysis

Descriptive results are presented. The median and interquartile range for the number of review primary and secondary outcomes was calculated.

Data are tabulated and excerpts found in the trial reports relating to review outcomes are used to support decisions made regarding ORBIT classifications and the assessment of risk of bias.

Results

The CFGD group had 46 cystic fibrosis systematic reviews published as of 2010.

1. Changes in outcomes between systematic review protocol and full review – review level

Protocols were available for all 46 systematic reviews. Nine protocols (20%) did not distinguish between primary and secondary outcomes. Table 3 shows the median number of primary and secondary outcomes for the 46 reviews and the changes in outcomes between protocol and full review.

Eighteen reviews (39%, 18/46) had a discrepancy in outcomes between protocol and full review. Between review protocol and full review, five (28%) listed all changes, two (11%) listed some changes and 11 reviews (61%) did not mention any change in outcomes. Of the seven reviews describing the changes between protocol and full review, three provided no reason for the changes, two stated that the changes in recommendations in the Cochrane Handbook to have a maximum of three primary outcomes were the reason for downgrading outcomes and two reviews stated that they added clinically relevant outcomes that were discovered during the review process.

2. Assessing trial reports for full ORB – outcome level~~ORBIT classifications~~

Of the 46 published reviews, 38 were eligible to be assessed for outcome reporting bias (Figure 1).

One review was excluded at this stage as the outcomes could not be assessed for ORB due to the different ways the outcome definitions could be measured and reported. The primary outcomes were psychosocial outcomes, which included any objective measure with adequate psychometric properties and demonstrable reliability and validity quantifying psychological or social outcomes or both, including individual psychological adjustment, relational, social functioning and adaptation to life with cystic fibrosis.

Therefore 37 reviews were assessed for ORB, including 280 RCTs (278 included and 2 excluded due to no relevant outcome data but confirmed by review authors that they would have otherwise been included). The median number of trials per review was four (IQR 2, 8) and there was a median sample size of 21 (IQR 14, 41) per trial.

Review authors contacted trialists for missing outcome data in 33 reviews (89%), one stated that "trialists were not contacted but would be in updates of the review" and three reviews did not state if trialists were contacted for further data.

~~Lead authors of each review were contacted.~~ The lead authors of twelve reviews assessed the included trials and gave classifications for each outcome. For thirteen reviews, authors gave input on which outcomes they expected to be measured for trials in their review and which outcomes they expected to be measured in routine clinical practice but did not classify each outcome due to time restrictions. The authors of twelve reviews did not respond to our request.

For the twelve reviews where the authors assigned classifications, discussion was needed on all outcomes to come to an agreed classification. For the other 25 reviews it was difficult to

assign a classification to all outcomes as some outcomes needed a large amount of clinical input in understanding the outcome and language used to describe the outcomes within the trial reports. Due to the number and complexity of outcomes and lack of reviewer input on the majority of reviews, it was decided that the assessment of all primary outcomes listed in the full review that were well defined should take priority. Many outcomes were also split into sub outcomes or ill defined to maximise the ability of a trial to contribute data to the review. ~~For example lung function was often split into FEV1 (Volume that has been exhaled at the end of the first second of forced expiration), FVC (Forced vital capacity), PEFR Peak expiratory flow rate), FEF_{25-75} (average expired flow over the middle half of the FVC manoeuvre) and these were assessed separately. FEV1 is the outcome most often considered for lung function due to its validity, repeatability and it is the outcome most understood by clinicians. However, the device used to measure FEV1 also measures the majority of other lung function outcomes. Therefore if FEV1 was reported in a trial, it was assumed that other lung function outcomes were also measured but not necessarily analysed (classification F) unless they were specifically stated as an outcome in the trial report. However, if FEV1 was not reported but other lung function outcomes were then an E classification was given to FEV1. This was decided after discussion with clinical experts.~~

The ORBIT classifications for the review primary outcomes for the 280 RCTs are shown in Table 1. For the 12 reviews where reviewer input was obtained, classifications for 64 included trials for review secondary outcomes are also shown in Table 1. Eligible trials within the reviews fully reported 383 (33.7%) review primary outcomes and 125 (18.7%) review secondary outcomes. In addition to the~~se~~ classifications in Table 1, a 'G, no events' classification (For example, mortality, where clinical judgement says it is likely to have been measured and it would have been reported had any deaths occurred. Therefore, it is assumed

no deaths occurred during the trial.) was given to eligible trials within the reviews for 109 ~~trials~~ (9.5%) ~~for~~ review primary outcomes and 22 ~~trials~~ (3.3%) ~~for~~ review secondary outcomes. Due to limited reviewer input or the lack of a standard definition ~~for some outcomes~~, ~~W~~we were unable to assess outcomes (including: adverse events, symptoms, complications, biochemical measures of glycaemic control, symptoms of sleep disordered breathing and measures of specific indices of strength, mass, effort and general fatigue) for 102 trials for review primary outcomes and 59 trials for review secondary outcomes.

Assessment of risk of bias from selective outcome reporting – trial level

Eighteen reviews (49%) had not yet assessed the risk of bias for selective outcome reporting as although ~~they had not been updated since the new~~ the Cochrane guidance on the risk of bias was introduced in 2008 and the cut off for this study was the beginning of 2010, these reviews were still to be updated~~had been introduced and prior to this study~~. Seventeen reviews (46%) had assessed the risk of bias for all included trials and two reviews (5%) assessed this for some of their included trials.

As we were unable to assess secondary outcomes for ORB for all reviews, the risk of bias assessments were made based on classifications of primary outcomes in order to be consistent across reviews. Only five (14%) of the 37 reviews had no trials at high risk of bias based on the review primary outcomes only.

Table 4 shows the risk of bias for selective outcome reporting as defined in this study and also as assessed within the published reviews for the 280 trials assessed for ORB based on the consideration of review primary outcomes only. It was found that 69% of trials had either not been assessed for selective reporting or were assessed as an unclear risk.

~~Only five (14%) of the 37 reviews had no trials at high risk of bias based on the review primary outcomes only.~~

Table 5 shows the risk of bias for selective outcome reporting based on the consideration of review primary and secondary outcomes separately for the 12 reviews (64 trials) where reviewers also provided classifications. This was to see if decisions regarding risk of bias would change if we considered all outcomes. Only four (6%) of the 64 RCTs had a low risk of bias when considering all outcomes.

Discrepancies in the risk of bias when considering all outcomes arose in 34 (53%) trials; 31 were at low risk when considering review primary outcomes only but high risk of bias (excluding G classifications: 13, G classification only: 18) when considering all outcomes; 3 were at high risk (G classifications only) when considering review primary outcomes only but high risk (excluding G classifications) when considering all outcomes. This often occurred in reviews were there was only one or two primary outcomes and a large number of secondary outcomes.

Based on all review outcomes, none of the 12 reviews had all included trials at low risk of bias.

Discussion

This is the first study to consider all review efficacy outcomes in an ORB assessment which has allowed us to make practical recommendations on assessing the risk of bias of selective reporting for systematic reviews at both the review and trial level. Over a third of Cochrane cystic fibrosis reviews (39%) examined had a discrepancy in outcomes between the review

protocol and full review. This compares to 22% of reviews (64/288) that contained a discrepancy in at least one outcome measure in the main ORBIT study which looked at reviews covering all 50 Cochrane review groups (Kirkham et al 2010a). However, this is confounded by the different publication date ranges of~~times in which~~ the reviews (assessed as up to date between 2006 and 2009) ~~were published~~. Furthermore, for the cystic fibrosis reviews outcome reporting bias was suspected in at least one randomised controlled trial in 86% of reviews when considering all review primary outcomes. The prevalence of reviews containing at least one trial with high suspicion of outcome reporting bias from ORBIT, when only a single primary outcome was considered was substantially lower at 34% (96/283) (Kirkham et al 2010b). While this study is limited only to CF trials, it is clear that the problem of outcome reporting bias is much larger when considering more than just the single primary review outcome of importance that was used in the ORBIT study.

A study by von Mosch and Dwan (2011) that compared the reporting in trial reports of CF to the CONSORT statement found that from a maximum of 57 points available, the scores rose from a median of 17.5 (Inter quartile range (IQR) 15.5-24.5) in 1994 to a median of 32 (IQR 22.8-41.5) in 2008. Along with the current study, this also indicates that there is still room for an improvement in the reporting of outcomes.

Use of the ORBIT classification system offered a robust methodology for assessing the risk of bias for trials included within a systematic review. When considering the 64 trials in the 12 reviews where it was possible to assess both primary and secondary outcomes, when basing the risk of bias assessment on review primary outcomes, 45% of trials were at high risk of bias and when using all outcomes in the assessment, 94% were at high risk of bias. Using the current selective reporting item of the current Cochrane risk of bias tool, 69% of

trials included in CF reviews were assessed by reviewers as 'unclear' risk of bias or not assessed at all, indicating the need for more informed guidance on assigning risk of bias in the systematic review process for all outcomes within a review.

The ORBIT classification system has already been validated as part of the original project. Sensitivity results for predicting that the outcome had been measured (G-classification) was 92% (23/25, 95% CI 81% to 100%), while the specificity for predicting that the outcome had not been measured (H-classification) was 77% (23/30, 95% CI 62% to 92%). With the additional requirement to assess all outcomes in this project, there were an increasing number of outcomes that were not mentioned in the trials reports and therefore clinical judgement was needed as to whether the outcome of interest was likely to have been measured in a particular trial. Many review authors did not respond to our request to provide classifications (68%), but for those with no response we did obtain clinical input ~~from within the CFGD group~~for the primary outcome from within the CFGD group. Although we can not exclude the possibility of response bias it is likely the decision to respond was influenced by time commitments rather than review characteristics. ~~therefore only primary outcomes were assessed within the majority of reviews due to the clinical complexity of many of the secondary outcomes.~~However, these assessments will be provided to the review authors when their review is due to be updated.

Reviewers should ensure that changes between protocol and reviews are listed and justifications provided to enhance the validity of these decisions. Eligible trials should not be excluded on the basis of "No relevant outcome data" because although an outcome was not reported it may have been measured and contact with the authors is advised. Reviewers should be encouraged to consider trials that have not reported an outcome of interest and to

assess whether selective reporting has occurred for all review outcomes. They should consider the amount of missing data from their meta-analysis (i.e. the percentage of the sample sizes of the studies that were included compared to those that would have been eligible to be included in the meta-analysis but no outcome data reported) and this information should be included along with the pooled effect estimate. If appropriate, a sensitivity analysis should be applied to assess the robustness of the conclusions of the review, such as an imputation approach (Williamson and Gamble 2005), the Copas bound for maximum bias (Copas et al, 2004; Williamson and Gamble 2007, Dwan et al 2010) or a model based correction (Copas et al, 2013).

Individuals conducting systematic reviews need to address explicitly the issue of missing outcome data for their review to be considered a reliable source of evidence. Extra care is required during data extraction, reviewers should identify when a trial reports that an outcome was measured but no results were reported or events observed, and contact with trialists should be encouraged. Contacting authors is encouraged by the CRG and is standard practice within CFGD reviews which is reflected in our results as 89% of reviews stated that they contacted authors for extra information on outcomes.

It is recommended that review authors ensure that they limit the number of outcomes in the review and define them clearly. ~~Reviewers also need to ensure that outcomes are well defined~~ as t~~T~~his will allow easier assessments of selective reporting, which can be done during data extraction of the included trials as long as a knowledgeable clinical person is involved~~.~~. Lung function was specified as the first primary outcome in nineteen reviews (50%), as the second or third primary outcome in 11 reviews (29%), as a secondary outcome in six reviews (16%) and it was not included as an outcome in only one review (5%).

However, as discussed earlier, lung function can be measured in different ways ~~it is often split into 'suboutcomes', including~~(-FEV1, FVC, mid forced expiratory flow (FEF), peak expiratory flow rate (PEFR), residual volume (RV), total lung capacity (TLC), Lung clearance index (LCI) and maximum expiratory flow (MEF).) These outcomes can then be analysed and reported in different ways such as**:** percentage~~%~~ predicted, litres, litres/second and post treatment, absolute change from baseline, relative change from baseline or annual rate of change. Therefore there is a large scope for selective reporting. One solution is the development of a core outcome set for cystic fibrosis (Ramsey and Boat 1994 , Sinha et al 2008, Clarke 2007). ~~It is recommended that review authors ensure that they limit the number of outcomes in the review and define them clearly. This will allow easier assessments of selective reporting, which can be done during data extraction of the included trials as long as a knowledgeable clinical person is involved.~~

Unanswered questions and future research

Work is needed to consider what the best method is to assess the impact of ORB on the results of the meta-analysis when there are multiple outcomes. Multivariate meta-analysis has been suggested by Kirkham et al 2012 and a model based correction has been suggested by Copas et al (2013~~2~~).

Conclusion

Systematic reviews need to clearly state the primary and secondary outcomes that they will consider and be consistent between review protocol and full review.

Outcome reporting bias is a major problem for systematic reviews and more guidance needs to be included in the Cochrane handbook to allow assessment of this important item within the risk of bias tool. We recommend that an outcome matrix be completed during the

production of a review to allow an ORB assessment for all review outcomes which can then inform the risk of bias assessment.

A core set of outcomes should be agreed upon for cystic fibrosis which in turn will have a positive impact on systematic reviews as future trials are conducted they should specifically set out to measure and report these outcomes therefore reducing the prevalence of selective reporting. -

Abbreviations

CFGD	Cystic Fibrosis and Genetic Disorders
CRG	Cochrane Review Group
COMET	Core Outcome Measures in Effectiveness Trials
CONSORT	Consolidated Standards of Reporting Trails
FEF_{25-75}	Average expired flow over the middle half of the FVC manoeuvre
FEV1	Forced expiratory volume in 1 second
FVC	Forced vital capacity
NROD	No Relevant Outcome Data
ORB	Outcome Reporting Bias
ORBIT	Outcome Reporting Bias In Trials
PEFR	Peak Expiratory Flow Rate
RCT	Randomised Controlled Trial

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing Interests

There are no financial or non-financial competing interests.

Authors' contributions

KD drafted the protocol, completed the ORB assessments and wrote the manuscript.

JJK completed the ORB assessments and commented on the manuscript.

CG commented on the manuscript.

PRW commented on the protocol and the manuscript and commented on the ORB assessments.

Acknowledgements

The authors would like to thank the Cochrane CFGD group reviewers for their assistance with clinical input for the ORB assessments of the included trials in their reviews.

References

Clarke, M. Standardising Outcomes for Clinical Trials and Systematic Reviews. *Trials; 2007,* 8**,** 39.

Copas J, Jackson D. A bound for publication bias based on the fraction 14 of unpublished studies. *Biometrics* 2004;60:146-53.

Copas J, Dwan K, Kirkham JJ, Williamson PR. A model based correction for outcome reporting bias in meta-analysis. (submitted 2013).

Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, et al. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. PLoS ONE 2008, 3(8): e3081.

Dwan K, Gamble C, Kolamunnage-Dona R, Mohammed S, Powell C, Williamson PR. Assessing the potential for outcome reporting bias in a review: A tutorial. *Trials 2010,* **11:**52 doi:10.1186/1745-6215-11-52.

Green, S. B. & Byar, D. P. Using observational data from registries to compare treatments: the fallacy of omnimetrics. *Stat Med, 1984;* 3**,** 361-73.

Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001; 323: 42-46.

Kirkham JJ, Altman DG, Williamson PR. Bias Due to Changes in Specified Outcomes during the Systematic Review Process. PLoS ONE 2010a; 5(3): e9810.doi:10.1371/journal.pone.0009810

Kirkham JJ, Dwan K, Dodd S, Altman DG, Smyth R, Jacoby A, Gamble C, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ 2010b; 340: c365

Kirkham JJ, Riley RD and Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. Statistics in Medicine 2012*; 31:20, 2179-2195.*

Ramsey BW, Boat TF (1994) Outcome measures for clinical trials in cystic fibrosis. Summary of a Cystic Fibrosis Foundation consensus conference. J Pediatr 1994; 124: 177–192.

Sinha I, Jones L, Smyth RL, Williamson PR. A Systematic Review of Studies That Aim to Determine Which Outcomes to Measure in Clinical Trials in Children . PLoS Med 2008; 5(4): e96. doi:10.1371/journal.pmed.0050096

Smyth AR,Walters S. Prophylactic anti-staphylococcal antibiotics for cystic fibrosis. *Cochrane Database of Systematic Reviews* 2003, Issue 3. Art. No.: CD001912. DOI: 10.1002/14651858.CD001912.

The Cochrane Library, Issue 4, 2009. Chichester: Wiley.

Williamson, P. R., Gamble, C., Altman, D. G. & Hutton, J. L. Outcome selection bias in meta-analysis. *Stat Methods Med Res, 2005;* 14**,** 515-24.

Williamson PR, Gamble C: Identification and impact of outcome selection bias in meta-analysis. Statistics in Medicine 2005, 24(10):1547-61.

Williamson PR, Gamble C: Application and investigation of a bound for outcome reporting bias. Trials 2007, 8:9.

von Mosch S, Dwan K. Reporting of harm and efficacy outcomes of randomised clinical trials in cystic fibrosis before and after the publication of the revised CONSORT (CONsolidated Standards Of Reporting Trials) statement in 2001 – a comparison of trial reports published from 1994 and from 2008. The University of Liverpool (thesis) 2011.

Figure 1: Study flow diagram

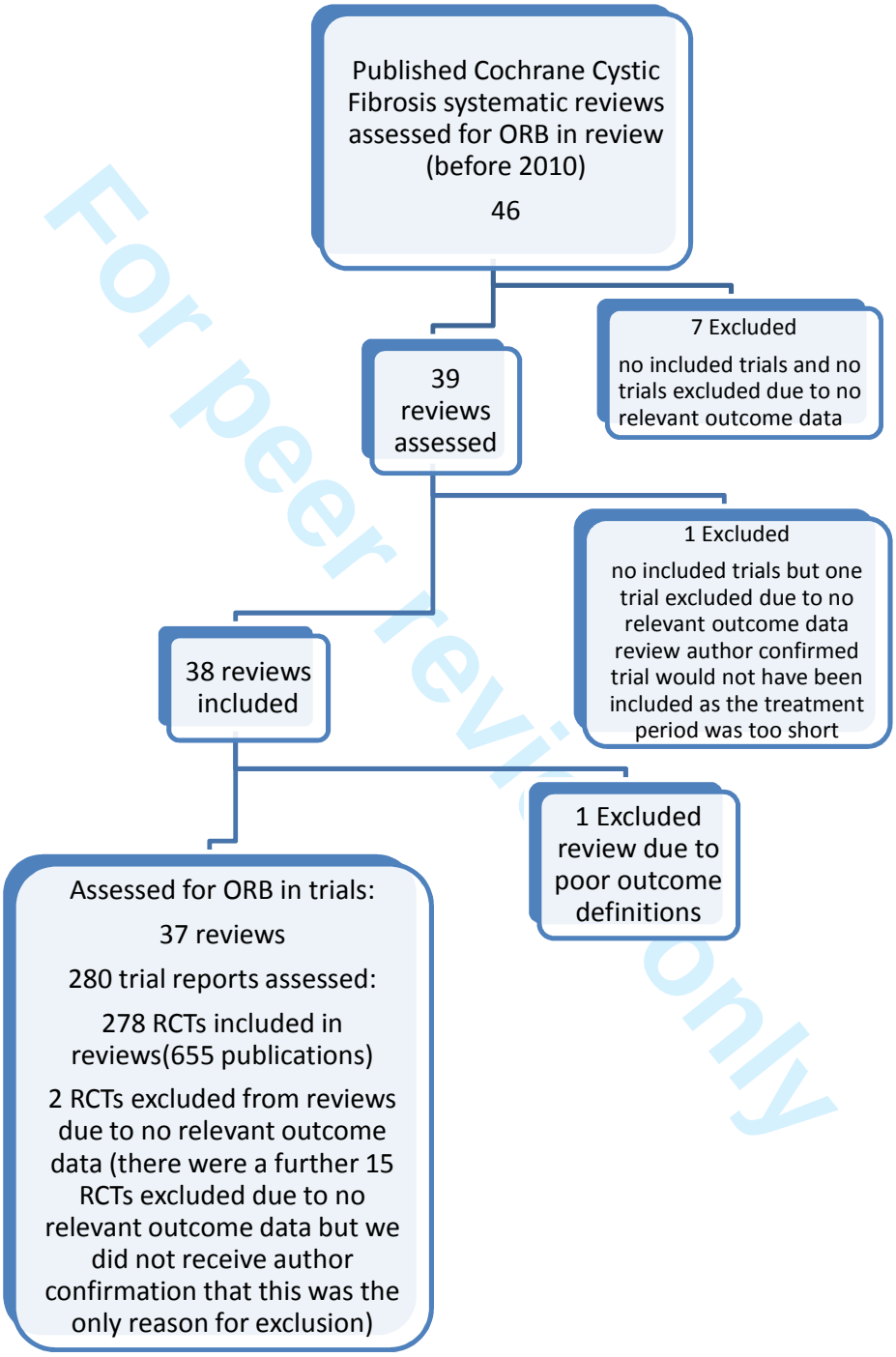


Table 1: ORBIT classifications

Classification	Description	Level of reporting	Level of suspicion of ORB	Primary outcome classifications	Secondary outcome classifications
				Number of trials (percentage overall)[1]	Number of trials (percentage overall)[2]
Clear that the outcome was measured and analysed					
A	States outcome analysed but only reported that result not significant (typically stating p-value >0.05).	Partial	High risk	75 (6.6%)	12 (1.8%)
B	States outcome analysed but only reported that result significant (typically stating p-value <0.05).	Partial	Low risk	13 (1.1%)	2 (0.3%)
C	States outcome analysed but insufficient data presented to be included in meta-analysis or to be considered to be fully tabulated.	Partial	Low risk	53 (4.7%)	15 (2.2%)
D	States outcome analysed but no results reported.	None	High risk	0 (0)	0 (0)
Clear that the outcome was measured					
E	Clear that outcome was measured but not necessarily analysed.	None	High risk	59 (5.2%)	26 (3.9%)
F	Clear that outcome was measured but not necessarily analysed.	None	Low risk	110 (9.7%)	15 (2.2%)
Unclear that the outcome was measured					
G	Not mentioned but clinical judgment says likely to have been measured and analysed.	None	High risk	195 (17.1%)	197 (29.4%)
H	Not mentioned but clinical judgment says unlikely to have been measured.	None	Low risk	141 (12.4%)	256 (38.2%)
Clear that the outcome was NOT measured					
I	Clear that outcome was not measured.	N/A	No risk	0 (0)	0 (0)

The ORBIT classifications for review primary outcomes for the 280 RCTs. For the 12 reviews where reviewer input was obtained, classifications for 64 included trials for review secondary outcomes are also shown.

1. The denominator used is the total number of data points expected if all 280 eligible trials reported on all review primary outcomes in the 37 reviews (i.e. the number of review primary outcomes multiplied by the number of trials within the review for all reviews. This does not include the 102 trials where we were unable to assess primary outcomes).
2. The denominator used is the total number of data points expected if all 64 trials reported on all review secondary outcomes in the 12 reviews (i.e. the number of review secondary outcomes multiplied by the number of trials within the review for all reviews. This does not include the 59 trials where we were unable to assess secondary outcomes).

Table 2: Example of review outcome matrix for 6 of 17 outcomes in a review of Prophylactic anti-staphylococcal antibiotics for cystic fibrosis (Smyth and Walters, 2003).

Study ID (author, date of publication)	Review primary outcomes			Review secondary outcomes			Other study outcomes
	Lung function FEV1	Lung function FVC	Number of people with one or more isolates of S. aureus	Growth	Survival	Quality of life	Serum levels of IgG
Chatfield 1991	o (A)[1]	o (A)[1]	✓	✓	✓	× (H)[2]	✓
Schlesinger 1984	× (H)[3]	× (H)[3]	✓	o (C)[4]	✓	× (H)[2]	✓
Stutman 2002	✓	✓	✓	✓	✓	× (H)[2]	×
Weaver 1994	× (H)[3]	× (H)[3]	✓	o (C)[4]	✓	× (H)[2]	×

1. Reasons for A classifications: 'no significant difference' reported in the text.
2. Reason for H classifications for quality of life: clinical judgement says it is unlikely to have been measured in these trials.
3. Reason for H classifications for lung function tests: both trials involve young children and these tests are not usually carried out on young children.
4. Reason for C classifications for Growth: trial reports give means but no standard deviations and also present the data in a graph.

✓ indicates full reporting of results for treatment comparison of interest
× indicates no reporting
o indicates partial reporting

Table 3: Changes in outcomes between review protocol and publication

		Primary outcomes	Secondary outcomes
Total number of outcomes included in the review (Median, IQR, range)		3 (IQR 2, 3 and range 1,8)	7 (IQR 5, 9 and range 2,13)
Reviews with any discrepancy in outcomes between protocol and full review	Protocol distinguished outcomes (n=37)[1]	14 (38%)	
	Protocol did not distinguish outcomes (n=9)[2]	4 (44%)	
Reviews which have *upgraded* at least one outcome from secondary in the protocol to primary in the full review (number of outcomes; minimum per review; maximum per review)	Protocol distinguished outcomes (n=37)[1]	3 (8%) (3 outcomes)	
	Protocol did not distinguish outcomes (n=9)[2]	0	
Reviews which have *downgraded* at least one outcome from primary in the protocol to secondary in the full review (number of outcomes; minimum per review; maximum per review)	Protocol distinguished outcomes (n=37)[1]	9 (24%) (16 outcomes; min 1, max 5)	
	Protocol did not distinguish outcomes (n=9)[2]	1 (11%) (2 outcomes)	
Reviews which have *added* a new outcome in the full review which was not included in the protocol (number of outcomes; minimum per review; maximum per review)	Protocol distinguished outcomes (n=37)[1]	2 (5%) (3 outcomes)	2 (5%) (4 outcomes; min 1, max 3)
	Protocol did not distinguish outcomes (n=9)[2]	1(11%) (1 outcome)	2 (22%) (2 outcomes)
Reviews which have *excluded* an outcome from the full review which was included in the protocol (number of outcomes; minimum per review; maximum per review)	Protocol distinguished outcomes (n=37)[1]	2 (5%) (10 outcomes; min 1, max 9)	3 (8%) (5 outcomes; min 1; max 2)
	Protocol did not distinguish outcomes (n=9)[2]	0	0

1. Protocol distinguished primary from secondary outcomes
2. Protocol did not distinguish primary from secondary outcomes

Table 4: Risk of bias of RCTs based on review primary outcomes only

		As assessed in review			Total
		High risk	Low risk	Unclear risk/ Not assessed	
As assessed in this study on the primary outcomes of the review only	High risk excluding G	10	18	50	78 (28%)
	High risk (based on G classifications only)	3	17	64	84 (30%)
	Low risk	14	24	80	118 (42%)
Total		27 (10%)	59 (21%)	194 (69%)	280

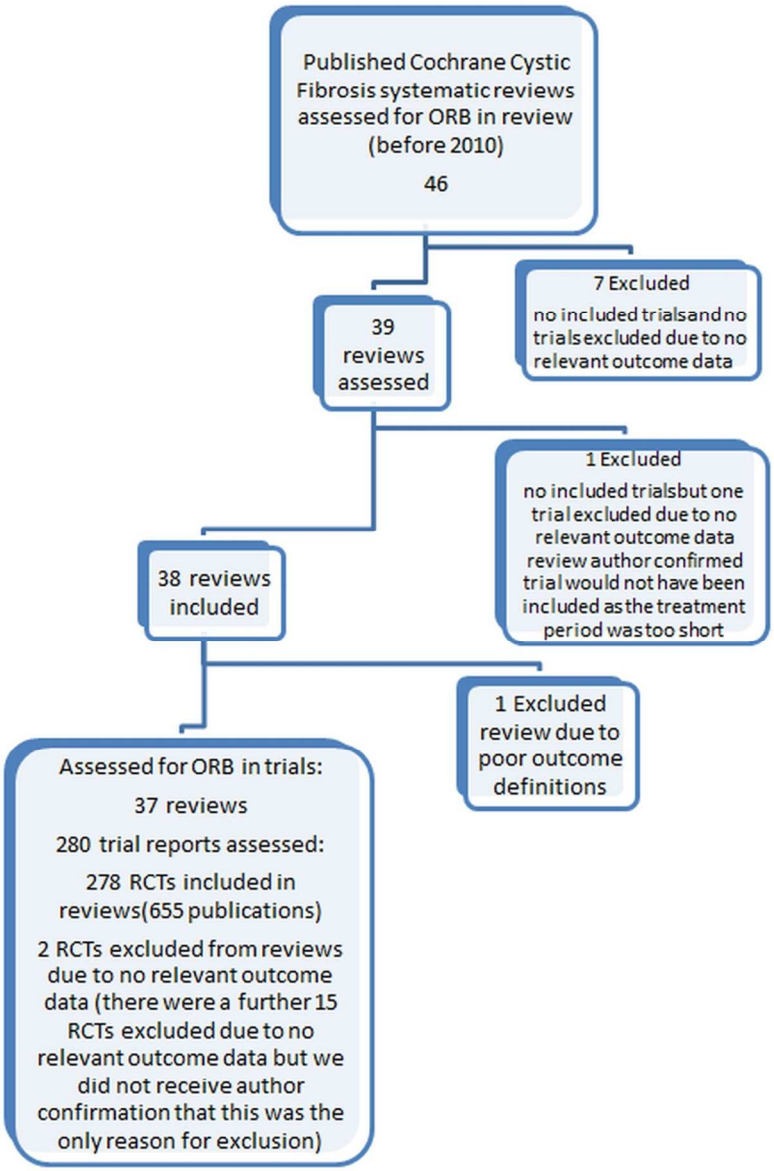
Note that 'As assessed in this study on the primary outcomes of the review only' is split into three categories: high risk excluding G; high risk (based on G classifications only) and low risk. This is because G classifications, although high risk of bias, are subjective as they are given based on clinical judgment only when there are no details mentioned in the trial report. However, as shown in the original ORBIT study (Kirkham et al, 2010b) the sensitivity and specificity of assigning G and H classifications was high.

Table 5: Risk of bias of RCTs based on review primary and secondary outcomes

		Risk of bias based on review primary outcomes only			Total
		High risk excluding G	High risk (based on G classifications only)	Low risk	
Risk of bias based on review primary and secondary outcomes	High risk excluding G	13	3	13	29 (45%)
	High risk (based on G classifications only)	0	13	18	31 (49%)
	Low risk	0	0	4	4 (6%)
Total		13 (20%)	16 (25%)	35 (55%)	64

Table 6: Risk of bias table for selective outcome reporting.

SELECTIVE OUTCOME REPORTING **Are reports of the study free of suggestion of selective outcome reporting? [Short form: *Free of selective reporting*?]**	
Criteria for a judgement of 'YES' (i.e. low risk of bias).	Any of the following: The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way; The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon).
Criteria for the judgement of 'NO' (i.e. high risk of bias).	Any one of the following: Not all of the study's pre-specified primary outcomes have been reported; One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; The study report fails to include results for a key outcome that would be expected to have been reported for such a study.
Criteria for the judgement of 'UNCLEAR' (uncertain risk of bias).	Insufficient information to permit judgement of 'Yes' or 'No'. It is likely that the majority of studies will fall into this category.



90x118mm (300 x 300 DPI)