



## Selective reporting of outcomes in randomised controlled trials of Cystic Fibrosis

Journal:	<i>BMJ Open</i>
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
<b>Primary Subject Heading</b>:	Evidence based practice
Secondary Subject Heading:	Genetics and genomics
Keywords:	STATISTICS & RESEARCH METHODS, MEDICAL EDUCATION & TRAINING, GENETICS

SCHOLARONE™  
Manuscripts

# Selective reporting of outcomes in randomised controlled trials of Cystic Fibrosis

Kerry Dwan<sup>1§</sup>, Jamie J Kirkham<sup>1</sup>, Paula R Williamson<sup>1</sup>, Carrol Gamble<sup>1</sup>

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

§Corresponding author

Email addresses:

KD: [kerry.dwan@liverpool.ac.uk](mailto:kerry.dwan@liverpool.ac.uk)

JJK: [jjk@liverpool.ac.uk](mailto:jjk@liverpool.ac.uk)

PRW: [prw@liverpool.ac.uk](mailto:prw@liverpool.ac.uk)

CG: [c.gamble@liverpool.ac.uk](mailto: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

1  
2  
3 This is the first study to consider the assessment of outcome reporting bias in all efficacy  
4  
5 review outcomes. However, this is limited to reviews of cystic fibrosis.  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## 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

1  
2  
3 Hutton 2000). In a systematic empirical assessment of Cochrane reviews within which a  
4 single review primary outcome could be identified (Kirkham et al 2010b), ORB was  
5 suspected in at least one randomised controlled trial in more than a third of the systematic  
6 reviews that were examined (35%). This study may have underestimated this problem as  
7 review primary outcomes are chosen due to their clinical importance so are more likely to  
8 have been measured and reported in trials increasing concern regarding the prevalence and  
9 impact of outcome reporting bias in reviews where multiple primary outcomes are specified,  
10 or in secondary outcomes.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

21  
22 Systematic reviews in cystic fibrosis are characterised by inclusion of small randomised trials  
23 specifying multiple primary outcomes. Reporting standards for trials of cystic fibrosis have  
24 also been shown to be low (von Mosch and Dwan 2011). The aims of this current study were  
25 to  
26  
27  
28  
29  
30

- 31 1. Examine the potential for bias created by review authors by identifying  
32 inconsistencies between outcomes published in review protocols and in the associated  
33 published reviews  
34  
35  
36  
37
- 38 2. Determine the prevalence of ORB in trials in systematic reviews of CF, extending  
39 previous work by considering all review efficacy outcomes (multiple primary and  
40 secondary).  
41  
42  
43  
44  
45  
46

## 47 **Methods**

48  
49 A cohort of systematic reviews published by the Cochrane Cystic Fibrosis and Genetic  
50 Disorders (CFGD) group on the Cochrane Library before 2010 were identified (The  
51 Cochrane Library, 2009). Reviews were eligible for inclusion if they compared interventions  
52 for cystic fibrosis and identified one or more eligible RCTs. RCTs that had been excluded (in  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 the “characteristics of excluded studies” section) were also checked for any suggestion of  
4  
5 outcome reporting bias. For example, if a review had excluded trials as a result of ‘no  
6  
7 relevant outcome data (NROD)’, then these trials were also scrutinised for the presence of  
8  
9 ORB and included in the assessment.  
10

### 11 12 13 14 **Changes in outcomes between systematic review protocol and full review**

15  
16 The numbers of primary and secondary outcomes per review were compared to the  
17  
18 recommendations for the number of outcomes (no more than three primary outcomes and a  
19  
20 limited number of secondary outcomes) to include in a review in the Cochrane Handbook  
21  
22 (Higgins and Green 2011). If a review did not distinguish between primary and secondary  
23  
24 outcomes, the first three outcomes listed were taken to be the primary outcomes and the rest  
25  
26 were considered as secondary outcomes. Protocols of the systematic reviews were accessed  
27  
28 and outcomes stated in the protocol were compared to those stated in the full review.  
29  
30 Changes in outcomes were identified and categorised by one author (KD) as: primary  
31  
32 outcome downgraded to secondary (*downgrade*); secondary outcome upgraded to primary  
33  
34 (*upgrade*); a new outcome not stated in the protocol was added to the full review (*addition*)  
35  
36 or an outcome stated in the protocol was omitted from the full review (*omission*). If there had  
37  
38 been a change in outcomes, the section ‘changes between protocol and review’ was examined  
39  
40 for a declaration and explanation of the changes.  
41  
42  
43  
44  
45

### 46 **Assessing trial reports for full ORB**

47  
48 For each eligible systematic review, all reports relating to included studies and studies  
49  
50 excluded due to no relevant outcome data were obtained. Reviews were checked to see  
51  
52 whether review authors had contacted trialists for further information or data for outcomes.  
53  
54  
55 Where this was not clear in the review, review authors were asked to clarify.  
56  
57  
58  
59  
60



1  
2  
3 A nine-point classification system (Table 1) developed for missing or incomplete outcome  
4 reporting in randomised trials (Kirkham et al, 2010b) was used to make an assessment of the  
5 risk of bias. Table 1 also provides examples of outcomes that were not assessed because they  
6 had poor outcome definitions. An outcome matrix (Table 2) was created for each review  
7 using the ORBIT matrix generator (<http://ctrc.liv.ac.uk/orbit/>), with studies listed in the rows  
8 and review primary and secondary outcomes listed in the columns with the ORBIT  
9 classifications (Table 1) given for each review outcome that was not fully reported (e.g. not  
10 reported or partially reported e.g.  $p>0.05$ ).  
11  
12  
13  
14  
15  
16  
17  
18  
19

20  
21  
22 The outcomes listed or detailed in the method section and the outcomes reported in the results  
23 section were compared for all trial publications to determine whether each outcome of the  
24 systematic review was measured and analysed. In some instances it may be obvious that an  
25 outcome was measured given the other outcomes reported. For example, if cause-specific  
26 mortality is reported then overall mortality must have been measured, even if not reported. In  
27 other situations it may be that a battery of tests or measurements are usually undertaken  
28 together, for example FEV1 (forced expiratory volume in 1 second) and FVC (forced vital  
29 capacity). If FVC is reported but FEV1 is not, suspicion should be raised that the latter may  
30 have been selectively not reported. However, it is often difficult to assess whether an  
31 outcome was measured, and clinical judgment is required. The clinical lead for each review  
32 was contacted by email and asked for their input into the assessment of selective outcome  
33 reporting within the trials included in their review. An assessment of whether the review  
34 outcomes had been measured and reported within each trial using the classification system  
35 was completed. The clinical lead for the review and KD independently assessed the trials in  
36 the review and any disagreements were resolved through discussion and then checked with a  
37 third person (JJK or PRW).  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 If one or more of the outcomes for a trial was given a high risk classification according to  
5  
6 Table 1, the trial was deemed at high risk of bias from selective reporting.  
7  
8

## 9 10 **Analysis**

11 Descriptive results are presented. The median and interquartile range for the number of  
12  
13 review primary and secondary outcomes was calculated.  
14  
15

16  
17  
18 Data are tabulated and excerpts found in the trial reports relating to review outcomes are used  
19  
20 to support decisions made regarding ORBIT classifications and the assessment of risk of bias.  
21  
22

## 23 24 25 **Results**

26  
27  
28 The CFGD group had 46 cystic fibrosis systematic reviews published as of 2010.  
29  
30

### 31 32 33 **1. Changes in outcomes between systematic review protocol and full review**

34  
35 Protocols were available for all 46 systematic reviews. Nine protocols (20%) did not  
36  
37 distinguish between primary and secondary outcomes. Table 3 shows the median number of  
38  
39 primary and secondary outcomes for the 46 reviews and the changes in outcomes between  
40  
41 protocol and full review.  
42  
43

44  
45  
46 Eighteen reviews (39%, 18/46) had a discrepancy in outcomes between protocol and full  
47  
48 review. Between review protocol and full review, five (28%) listed all changes, two (11%)  
49  
50 listed some changes and 11 reviews (61%) did not mention any change in outcomes. Of the  
51  
52 seven reviews describing the changes between protocol and full review, three provided no  
53  
54 reason for the changes, two stated that the changes in recommendations in the Cochrane  
55  
56 Handbook to have a maximum of three primary outcomes were the reason for downgrading  
57  
58  
59  
60

1  
2  
3 outcomes and two reviews stated that they added clinically relevant outcomes that were  
4  
5 discovered during the review process.  
6  
7

## 8 9 **2. ORBIT classifications**

10  
11 Of the 46 published reviews, 38 were eligible to be assessed for outcome reporting bias  
12  
13 (Figure 1).  
14  
15

16  
17 One review was excluded at this stage as the outcomes could not be assessed for ORB due to  
18  
19 the different ways the outcome definitions could be measured and reported. The primary  
20  
21 outcomes were psychosocial outcomes, which included any objective measure with adequate  
22  
23 psychometric properties and demonstrable reliability and validity quantifying psychological  
24  
25 or social outcomes or both, including individual psychological adjustment, relational, social  
26  
27 functioning and adaptation to life with cystic fibrosis.  
28  
29  
30

31  
32 Therefore 37 reviews were assessed for ORB, including 280 RCTs (278 included and 2  
33  
34 excluded due to no relevant outcome data but confirmed by review authors that they would  
35  
36 have otherwise been included). The median number of trials per review was four (IQR 2, 8)  
37  
38 and there was a median sample size of 21 (IQR 14, 41) per trial.  
39  
40  
41  
42

43  
44 Review authors contacted trialists for missing outcome data in 33 reviews (89%), one stated  
45  
46 that “trialists were not contacted but would be in updates of the review” and three reviews did  
47  
48 not state if trialists were contacted for further data.  
49  
50  
51

52  
53 Lead authors of each review were contacted. The lead authors of twelve reviews assessed the  
54  
55 included trials and gave classifications for each outcome. For thirteen reviews, authors gave  
56  
57  
58  
59  
60

1  
2  
3 input on which outcomes they expected to be measured for trials in their review and which  
4  
5 outcomes they expected to be measured in routine clinical practice but did not classify each  
6  
7 outcome due to time restrictions. The authors of twelve reviews did not respond to our  
8  
9 request.  
10

11  
12  
13  
14 For the twelve reviews where the authors assigned classifications, discussion was needed on  
15  
16 all outcomes to come to an agreed classification. For the other 25 reviews it was difficult to  
17  
18 assign a classification to all outcomes as some outcomes needed a large amount of clinical  
19  
20 input in understanding the outcome and language used to describe the outcomes within the  
21  
22 trial reports.  
23  
24

25  
26  
27 Due to the number and complexity of outcomes and lack of reviewer input on the majority of  
28  
29 reviews, it was decided that the assessment of all primary outcomes listed in the full review  
30  
31 that were well defined should take priority. Many outcomes were also split into sub  
32  
33 outcomes or ill defined to maximise the ability of a trial to contribute data to the review. For  
34  
35 example lung function was often split into FEV1 (Volume that has been exhaled at the end of  
36  
37 the first second of forced expiration), FVC (Forced vital capacity), PEF<sub>R</sub> Peak expiratory  
38  
39 flow rate), FEF<sub>25-75</sub> (average expired flow over the middle half of the FVC manoeuvre) and  
40  
41 these were assessed separately. FEV1 is the outcome most often considered for lung function  
42  
43 due to its validity, repeatability and it is the outcome most understood by clinicians.  
44  
45

46  
47 However, the device used to measure FEV1 also measures the majority of other lung function  
48  
49 outcomes. Therefore if FEV1 was reported in a trial, it was assumed that other lung function  
50  
51 outcomes were also measured but not necessarily analysed (classification F) unless they were  
52  
53 specifically stated as an outcome in the trial report. However, if FEV1 was not reported but  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 other lung function outcomes were then an E classification was given to FEV1. This was  
4  
5 decided after discussion with clinical experts.  
6  
7

8  
9  
10 The ORBIT classifications for the review primary outcomes for the 280 RCTs are shown in  
11  
12 Table 1. For the 12 reviews where reviewer input was obtained, classifications for 64  
13  
14 included trials for review secondary outcomes are also shown in Table 1. In addition to these  
15  
16 classifications, a 'G, no events' classification (For example, mortality, were clinical  
17  
18 judgement says it is likely to have been measured and it would have been reported had any  
19  
20 deaths occurred. Therefore, it is assumed no deaths occurred during the trial.) was given to  
21  
22 109 trials for review primary outcomes and 22 trials for review secondary outcomes. We  
23  
24 were unable to assess outcomes (including: adverse events, symptoms, complications,  
25  
26 biochemical measures of glycaemic control, symptoms of sleep disordered breathing and  
27  
28 measures of specific indices of strength, mass, effort and general fatigue) for 102 trials for  
29  
30 review primary outcomes and 59 trials for review secondary outcomes.  
31  
32  
33  
34  
35  
36

### 37 **Assessment of risk of bias from selective reporting**

38  
39 Eighteen reviews (49%) had not yet assessed the risk of bias for selective outcome reporting  
40  
41 as they had not been updated since the new Cochrane guidance on the risk of bias had been  
42  
43 introduced and prior to this study. Seventeen reviews (46%) had assessed the risk of bias for  
44  
45 all included trials and two reviews (5%) assessed this for some of their included trials.  
46  
47  
48  
49

50  
51 As we were unable to assess secondary outcomes for ORB for all reviews, the risk of bias  
52  
53 assessments were made based on classifications of primary outcomes in order to be consistent  
54  
55 across reviews. Table 4 shows the risk of bias for selective outcome reporting as defined in  
56  
57  
58  
59  
60

1  
2  
3 this study and also as assessed within the published reviews for the 280 trials assessed for  
4  
5 ORB based on the consideration of review primary outcomes only.  
6  
7

8  
9  
10 Only five (14%) of the 37 reviews had no trials at high risk of bias based on the review  
11  
12 primary outcomes only.  
13

14  
15  
16 Table 5 shows the risk of bias for selective outcome reporting based on the consideration of  
17  
18 review primary and secondary outcomes separately for the 12 reviews (64 trials) were  
19  
20 reviewers also provided classifications. This was to see if decisions regarding risk of bias  
21  
22 would change if we considered all outcomes. Only four (6%) of the 64 RCTs had a low risk  
23  
24 of bias when considering all outcomes.  
25  
26

27  
28  
29 Discrepancies in the risk of bias when considering all outcomes arose in 34 (53%) trials; 31  
30  
31 were at low risk when considering review primary outcomes only but high risk of bias  
32  
33 (excluding G classifications: 13, G classification only: 18) when considering all outcomes; 3  
34  
35 were at high risk (G classifications only) when considering review primary outcomes only  
36  
37 but high risk (excluding G classifications) when considering all outcomes. This often  
38  
39 occurred in reviews where there was only one or two primary outcomes and a large number of  
40  
41 secondary outcomes.  
42  
43  
44

45  
46  
47 Based on all review outcomes, none of the 12 reviews had all included trials at low risk of  
48  
49 bias.  
50  
51

## 52 53 54 **Discussion** 55 56 57 58 59 60

1  
2  
3 This is the first study to consider all review efficacy outcomes in an ORB assessment which  
4 has allowed us to make practical recommendations on assessing the risk of bias of selective  
5 reporting for systematic reviews at both the review and trial level. Over a third of Cochrane  
6 cystic fibrosis reviews (39%) examined had a discrepancy in outcomes between the review  
7 protocol and full review. This compares to 22% of reviews (64/288) that contained a  
8 discrepancy in at least one outcome measure in the main ORBIT study which looked at  
9 reviews covering all 50 Cochrane review groups (Kirkham et al 2010a). However, this is  
10 confounded by the times in which the reviews were published. Furthermore, for the cystic  
11 fibrosis reviews outcome reporting bias was suspected in at least one randomised controlled  
12 trial in 86% of reviews when considering all review primary outcomes. The prevalence of  
13 reviews containing at least one trial with high suspicion of outcome reporting bias from  
14 ORBIT, when only a single primary outcome was considered was substantially lower at 34%  
15 (96/283) (Kirkham et al 2010b). While this study is limited only to CF trials, it is clear that  
16 the problem of outcome reporting bias is much larger when considering more than just the  
17 single primary review outcome of importance that was used in the ORBIT study.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37

38 Use of the ORBIT classification system offered a robust methodology for assessing the risk  
39 of bias for trials included within a systematic review. When considering the 64 trials in the  
40 12 reviews where it was possible to assess both primary and secondary outcomes, when basing  
41 the risk of bias assessment on review primary outcomes, 45% of trials were at high risk of  
42 bias and when using all outcomes in the assessment, 94% were at high risk of bias. Using the  
43 current selective reporting item of the current Cochrane risk of bias tool, 69% of trials  
44 included in CF reviews were assessed by reviewers as 'unclear' risk of bias indicating the  
45 need for more informed guidance on assigning risk of bias in the systematic review process  
46 for all outcomes within a review.  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5 The ORBIT classification system has already been validated as part of the original project.  
6  
7 Sensitivity results for predicting that the outcome had been measured (G-classification) was  
8  
9 92% (23/25, 95% CI 81% to 100%), while the specificity for predicting that the outcome had  
10  
11 not been measured (H-classification) was 77% (23/30, 95% CI 62% to 92%). With the  
12  
13 additional requirement to assess all outcomes in this project, there were an increasing number  
14  
15 of outcomes that were not mentioned in the trials reports and therefore clinical judgement  
16  
17 was needed as to whether the outcome of interest was likely to have been measured in a  
18  
19 particular trial. Many review authors did not respond to our request, therefore only primary  
20  
21 outcomes were assessed within the majority of reviews due to the clinical complexity of  
22  
23 many of the secondary outcomes.  
24  
25  
26  
27

28  
29 Reviewers should ensure that changes between protocol and reviews are listed and  
30  
31 justifications provided to enhance the validity of these decisions. Eligible trials should not be  
32  
33 excluded on the basis of “No relevant outcome data” because although an outcome was not  
34  
35 reported it may have been measured and contact with the authors is advised. Reviewers  
36  
37 should be encouraged to consider trials that have not reported an outcome of interest and to  
38  
39 assess whether selective reporting has occurred for all review outcomes. They should  
40  
41 consider the amount of missing data from their meta-analysis (i.e. the sample sizes of the  
42  
43 studies that would have been eligible to be included in the meta-analysis but no outcome data  
44  
45 reported) and this information should be included along with the pooled effect estimate. If  
46  
47 appropriate, a sensitivity analysis should be applied to assess the robustness of the  
48  
49 conclusions of the review (Dwan et al 2010).  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 Individuals conducting systematic reviews need to address explicitly the issue of missing  
4  
5 outcome data for their review to be considered a reliable source of evidence. Extra care is  
6  
7 required during data extraction, reviewers should identify when a trial reports that an  
8  
9 outcome was measured but no results were reported or events observed, and contact with  
10  
11 trialists should be encouraged. Contacting authors is encouraged by the CRG and is standard  
12  
13 practice within CFGD reviews which is reflected in our results as 89% of reviews stated that  
14  
15 they contacted authors for extra information on outcomes.  
16  
17

18  
19  
20 Reviewers also need to ensure that outcomes are well defined. Lung function was specified  
21  
22 as the first primary outcome in nineteen reviews (50%), as the second or third primary  
23  
24 outcome in 11 reviews (29%), as a secondary outcome in six reviews (16%) and it was not  
25  
26 included as an outcome in only one review (5%). However, it is often split into  
27  
28 ‘suboutcomes’, including FEV1, FVC, mid forced expiratory flow (FEF), residual volume  
29  
30 (RV), total lung capacity (TLC), Lung clearance index (LCI) and maximum expiratory flow  
31  
32 (MEF). These outcomes can then be analysed and reported in different ways such as: %  
33  
34 predicted, litres, litres/second and post treatment, absolute change from baseline, relative  
35  
36 change from baseline or annual rate of change. Therefore there is a large scope for selective  
37  
38 reporting. One solution is the development of a core outcome set for cystic fibrosis (Ramsey  
39  
40 and Boat 1994 , Sinha et al 2008, Clarke 2007). It is recommended that review authors  
41  
42 ensure that they limit the number of outcomes in the review and define them clearly. This  
43  
44 will allow easier assessments of selective reporting, which can be done during data extraction  
45  
46 of the included trials as long as a knowledgeable clinical person is involved.  
47  
48  
49  
50

#### 51 52 53 **Unanswered questions and future research** 54 55 56 57 58 59 60

1  
2  
3 Work is needed to consider what the best method is to assess the impact of ORB on the  
4 results of the meta-analysis when there are multiple outcomes. Multivariate meta-analysis  
5 has been suggested by Kirkham et al 2012 and a model based correction has been suggested  
6 by Copas et al (2012).  
7  
8  
9  
10

## 11 **Conclusion**

12 Systematic reviews need to clearly state the primary and secondary outcomes that they will  
13 consider and be consistent between review protocol and full review.  
14

15 Outcome reporting bias is a major problem for systematic reviews and more guidance needs  
16 to be included in the Cochrane handbook to allow assessment of this important item within  
17 the risk of bias tool. We recommend that an outcome matrix be completed during the  
18 production of a review to allow an ORB assessment for all review outcomes which can then  
19 inform the risk of bias assessment.  
20

21 A core set of outcomes should be agreed upon for cystic fibrosis which in turn will have a  
22 positive impact on systematic reviews.  
23  
24  
25  
26  
27  
28  
29  
30  
31

## 32 **Abbreviations**

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

40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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](http://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.

1  
2  
3 Kirkham JJ, Altman DG, Williamson PR. Bias Due to Changes in Specified Outcomes during  
4  
5 the Systematic Review Process. PLoS ONE 2010a; 5(3):  
6  
7 e9810.doi:10.1371/journal.pone.0009810  
8  
9

10  
11 Kirkham JJ, Dwan K, Dodd S, Altman DG, Smyth R, Jacoby A, Gamble C, Williamson PR.  
12  
13 The impact of outcome reporting bias in randomised controlled trials on a cohort of  
14  
15 systematic reviews. BMJ 2010b; 340: c365  
16  
17  
18  
19

20  
21 Kirkham JJ, Riley RD and Williamson PR. A multivariate meta-analysis approach for  
22  
23 reducing the impact of outcome reporting bias in systematic reviews. *Statistics in Medicine*  
24  
25 2012; 31:20, 2179-2195.  
26  
27  
28

29  
30 Ramsey BW, Boat TF (1994) Outcome measures for clinical trials in cystic fibrosis.  
31  
32 Summary of a Cystic Fibrosis Foundation consensus conference. *J Pediatr* 1994; 124: 177–  
33  
34 192.  
35  
36  
37

38  
39 Sinha I, Jones L, Smyth RL, Williamson PR. A Systematic Review of Studies That Aim to  
40  
41 Determine Which Outcomes to Measure in Clinical Trials in Children . *PLoS Med* 2008;  
42  
43 5(4): e96. doi:10.1371/journal.pmed.0050096  
44  
45  
46

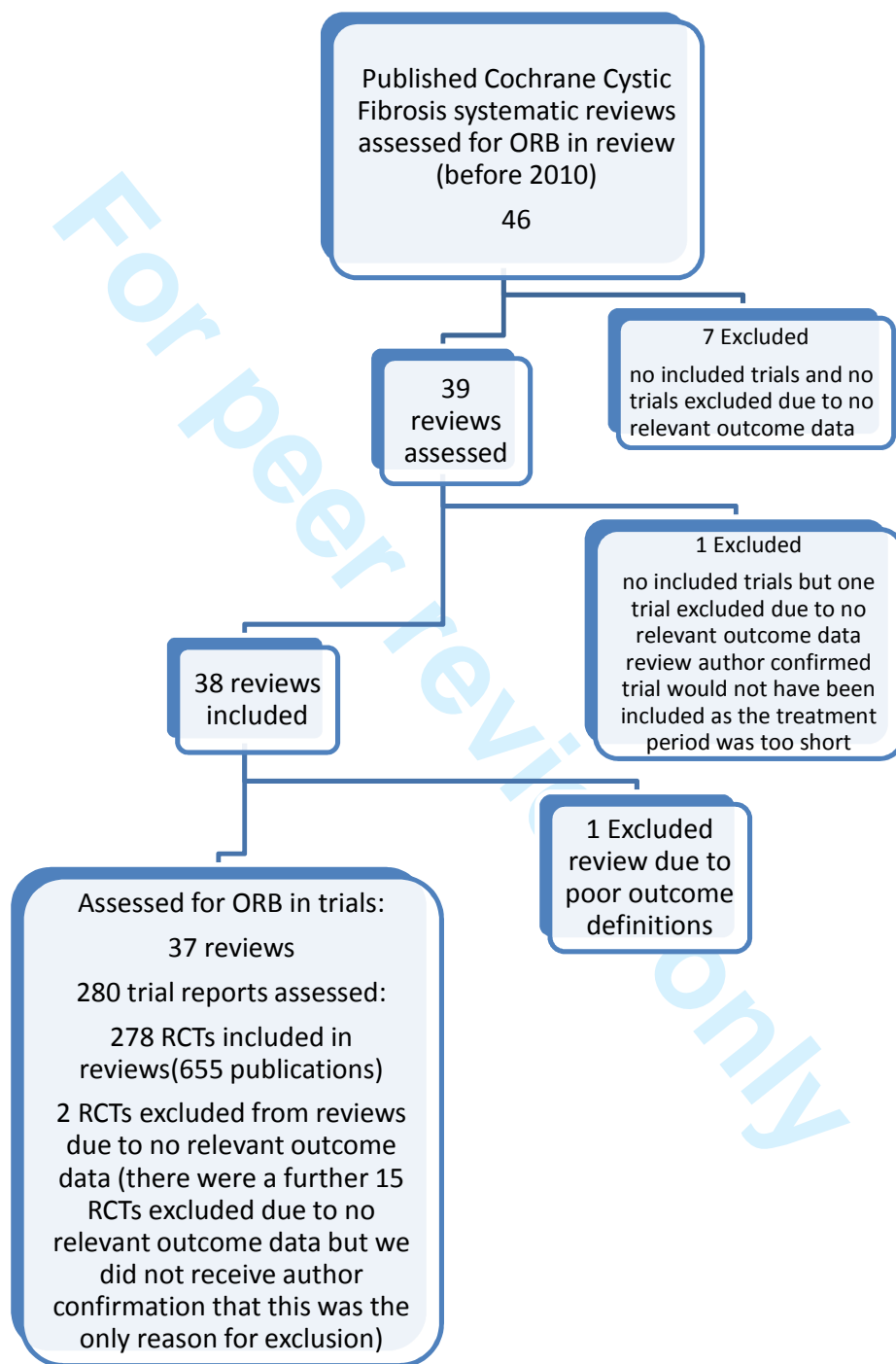
47  
48 Smyth AR, Walters S. Prophylactic anti-staphylococcal antibiotics for cystic fibrosis.  
49  
50 *Cochrane Database of Systematic Reviews* 2003, Issue 3. Art. No.: CD001912. DOI:  
51  
52 10.1002/14651858.CD001912.  
53  
54

55  
56 *The Cochrane Library*, Issue 4, 2009. Chichester: Wiley.  
57  
58  
59  
60

1  
2  
3 Williamson, P. R., Gamble, C., Altman, D. G. & Hutton, J. L. Outcome selection bias in  
4  
5 meta-analysis. *Stat Methods Med Res*, 2005; 14, 515-24.  
6  
7

8  
9  
10 von Mosch S, Dwan K. Reporting of harm and efficacy outcomes of randomised clinical  
11  
12 trials in cystic fibrosis before and after the publication of the revised CONSORT  
13  
14 (CONsolidated Standards Of Reporting Trials) statement in 2001 – a comparison of trial  
15  
16 reports published from 1994 and from 2008. The University of Liverpool (thesis) 2011.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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
<i>Clear that the outcome was measured and analysed</i>					
<b>A</b>	States outcome analysed but only reported that result not significant (typically stating p-value >0.05).	Partial	High risk	75	12
<b>B</b>	States outcome analysed but only reported that result significant (typically stating p-value <0.05).	Partial	Low risk	13	2
<b>C</b>	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
<b>D</b>	States outcome analysed but no results reported.	None	High risk	0	0
<i>Clear that the outcome was measured</i>					
<b>E</b>	Clear that outcome was measured but not necessarily analysed.	None	High risk	59	26
<b>F</b>	Clear that outcome was measured but not necessarily analysed.	None	Low risk	110	15
<i>Unclear that the outcome was measured</i>					
<b>G</b>	Not mentioned but clinical judgment says likely to have been measured and analysed.	None	High risk	195	197
<b>H</b>	Not mentioned but clinical judgment says unlikely to have been measured.	None	Low risk	141	256
<i>Clear that the outcome was NOT measured</i>					
<b>I</b>	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	○ (A) <sup>1</sup>	○ (A) <sup>1</sup>	✓	✓	✓	✗ (H) <sup>2</sup>	✓
Schlesinger 1984	✗ (H) <sup>3</sup>	✗ (H) <sup>3</sup>	✓	○ (C) <sup>4</sup>	✓	✗ (H) <sup>2</sup>	✓
Stutman 2002	✓	✓	✓	✓	✓	✗ (H) <sup>2</sup>	✗
Weaver 1994	✗ (H) <sup>3</sup>	✗ (H) <sup>3</sup>	✓	○ (C) <sup>4</sup>	✓	✗ (H) <sup>2</sup>	✗

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.

- |   |
|---|
| <ul style="list-style-type: none"> <li>✓ indicates full reporting of results for treatment comparison of interest</li> <li>✗ indicates no reporting</li> <li>○ indicates partial reporting</li> </ul> |
|---|

**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) <sup>1</sup>	14 (38%)	
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	4 (44%)	
Reviews which have <i>upgraded</i> 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) <sup>1</sup>	3 (8%) (3 outcomes)	
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	0	
Reviews which have <i>downgraded</i> 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) <sup>1</sup>	9 (24%) (16 outcomes; min 1, max 5)	
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	1 (11%) (2 outcomes)	
Reviews which have <i>added</i> 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) <sup>1</sup>	2 (5%) (3 outcomes)	2 (5%) (4 outcomes; min 1, max 3)
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	1(11%) (1 outcome)	2 (22%) (2 outcomes)
Reviews which have <i>excluded</i> 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) <sup>1</sup>	2 (5%) (10 outcomes; min 1, max 9)	3 (8%) (5 outcomes; min 1; max 2)
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	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.

<b>SELECTIVE OUTCOME REPORTING</b>	
<b>Are reports of the study free of suggestion of selective outcome reporting? [Short form: <i>Free of selective reporting?</i>]</b>	
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:	<i>BMJ Open</i>
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
<b>Primary Subject Heading</b>:	Evidence based practice
Secondary Subject Heading:	Genetics and genomics, Respiratory medicine, Research methods
Keywords:	STATISTICS & RESEARCH METHODS, MEDICAL EDUCATION & TRAINING, GENETICS

SCHOLARONE™  
Manuscripts

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

Kerry Dwan<sup>1§</sup>, Jamie J Kirkham<sup>1</sup>, Paula R Williamson<sup>1</sup>, Carrol Gamble<sup>1</sup>

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

<sup>§</sup>Corresponding author

Email addresses:

KD: [kerry.dwan@liverpool.ac.uk](mailto:kerry.dwan@liverpool.ac.uk)

JJK: [jjk@liverpool.ac.uk](mailto:jjk@liverpool.ac.uk)

PRW: [prw@liverpool.ac.uk](mailto:prw@liverpool.ac.uk)

CG: [c.gamble@liverpool.ac.uk](mailto: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

1  
2  
3 Assessment of ORB within a systematic review, of a single primary, outcome underestimates  
4 the risk of ORB in comparison to the assessment of multiple primary and secondary  
5 outcomes. ORB in trials is highly prevalent within systematic reviews of cystic fibrosis when  
6 assessed across all outcomes. This could be reduced by the development of a core outcome  
7 set for trials and systematic reviews in cystic fibrosis.  
8  
9  
10  
11  
12

## 13 14 15 16 17 **Article summary**

### 18 19 **Article focus**

- 20 • Assessment of discrepancies in outcome selection between systematic review  
21 protocols and full reviews.  
22
- 23 • Assessment of outcome reporting bias at the outcome level across all efficacy  
24 systematic review outcomes.  
25
- 26 • Assessment of the risk of bias of a trial from selective outcome reporting within a  
27 systematic review.  
28  
29  
30  
31  
32  
33  
34  
35

### 36 37 **Key messages**

- 38 • Assessment of ORB within a systematic review, of a single primary, outcome  
39 underestimates the risk of ORB in comparison to the assessment of multiple primary  
40 and secondary outcomes. Clearer guidance is needed on how to assess the risk of bias  
41 as a result of selective outcome reporting for each included trial within a systematic  
42 review, when considering multiple outcomes.  
43  
44  
45  
46  
47  
48
- 49 • The development of a core outcome set in Cystic Fibrosis would help reduce the  
50 problem of ORB.  
51  
52  
53  
54

### 55 56 **Strengths and limitations**

57  
58  
59  
60



1  
2  
3 This is the first study to consider the assessment of outcome reporting bias in all efficacy  
4  
5 review outcomes. However, this is limited to reviews of cystic fibrosis.  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## 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

1  
2  
3 publication (Williamson and Hutton 2000). In a systematic empirical assessment of  
4  
5 Cochrane reviews within which a single review primary outcome could be identified  
6  
7 (Kirkham et al 2010b), ORB was suspected in at least one randomised controlled trial in more  
8  
9 than a third of the systematic reviews that were examined (35%). This study may have  
10  
11 underestimated this problem as review primary outcomes are chosen due to their clinical  
12  
13 importance and are more likely to have been measured and reported in trials. Therefore,  
14  
15 there is concern regarding the prevalence and impact of outcome reporting bias in reviews  
16  
17 where multiple primary outcomes are specified, or in secondary outcomes.  
18  
19

20  
21  
22 Systematic reviews in cystic fibrosis are characterised by inclusion of small randomised trials  
23  
24 specifying multiple primary outcomes. Reporting standards for trials of cystic fibrosis have  
25  
26 also been shown to be low when comparing trial reports to the Consolidated Standards of  
27  
28 Reporting Trails (CONSORT) statement (von Mosch and Dwan 2011). The aims of this  
29  
30 current study were to  
31  
32

- 33  
34 1. Examine the potential for bias created by review authors by identifying  
35  
36 inconsistencies between outcomes published in review protocols and in the associated  
37  
38 published reviews  
39
- 40  
41 2. Determine the prevalence of ORB in trials in systematic reviews of CF, extending  
42  
43 previous work by considering all review efficacy outcomes (multiple primary and  
44  
45 secondary).  
46
- 47  
48 3. Assess the risk of bias of trials from selective outcome reporting when considering  
49  
50 review primary outcomes only in comparison to all review outcomes.  
51

## 52 53 54 **Methods**

55  
56  
57  
58  
59  
60

1  
2  
3 A cohort of systematic reviews published by the Cochrane Cystic Fibrosis and Genetic  
4 Disorders (CFGD) group on the Cochrane Library before 2010 were identified (The  
5 Cochrane Library, 2009). Reviews were eligible for inclusion if they compared interventions  
6 for cystic fibrosis and identified one or more eligible RCTs. RCTs that had been excluded (in  
7 the “characteristics of excluded studies” section) were also checked for any suggestion of  
8 outcome reporting bias. For example, if a review had excluded trials as a result of ‘no  
9 relevant outcome data (NROD)’, then these trials were also scrutinised for the presence of  
10 ORB and included in the assessment.  
11  
12  
13  
14  
15  
16  
17  
18  
19

### 20 21 22 23 **Changes in outcomes between systematic review protocol and full review – review level**

24 The numbers of primary and secondary outcomes per review were compared to the  
25 recommendations for the number of outcomes (no more than three primary outcomes and a  
26 limited number of secondary outcomes) to include in a review in the Cochrane Handbook  
27 (Higgins and Green 2011). If a review did not distinguish between primary and secondary  
28 outcomes, the first three outcomes listed were taken to be the primary outcomes and the rest  
29 were considered as secondary outcomes. Protocols of the systematic reviews were accessed  
30 and outcomes stated in the protocol were compared to those stated in the full review.  
31 Changes in outcomes were identified and categorised by one author (KD) as: primary  
32 outcome downgraded to secondary (*downgrade*); secondary outcome upgraded to primary  
33 (*upgrade*); a new outcome not stated in the protocol was added to the full review (*addition*)  
34 or an outcome stated in the protocol was omitted from the full review (*omission*). If there had  
35 been a change in outcomes, the section ‘changes between protocol and review’ was examined  
36 for a declaration and explanation of the changes.  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53

### 54 55 **Assessing trial reports for full ORB – outcome level**

1  
2  
3 For each eligible systematic review, all reports relating to included studies and studies  
4 excluded due to no relevant outcome data were obtained. Reviews were checked to see  
5 whether review authors had contacted trialists for further information or data for outcomes.  
6  
7 Where this was not clear in the review, review authors were asked to clarify.  
8  
9  
10

11  
12  
13 A nine-point classification system (Table 1) developed for missing or incomplete outcome  
14 reporting in randomised trials (Kirkham et al, 2010b) was used to make an assessment of the  
15 risk of bias. Table 1 also provides examples of outcomes that were not assessed because they  
16 had poor outcome definitions. An outcome matrix (Table 2) was created for each review  
17 using the ORBIT matrix generator (<http://ctrc.liv.ac.uk/orbit/>), with studies listed in the rows  
18 and review primary and secondary outcomes listed in the columns with the ORBIT  
19 classifications (Table 1) given for each review outcome that was not fully reported (e.g. not  
20 reported or partially reported e.g.  $p>0.05$ ).  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

32  
33 The outcomes listed or detailed in the method section and the outcomes reported in the results  
34 section were compared for all trial publications to determine whether each outcome of the  
35 systematic review was measured and analysed. In some instances it may be obvious that an  
36 outcome was measured given the other outcomes reported. For example, if cause-specific  
37 mortality is reported then overall mortality must have been measured, even if not reported. In  
38 other situations it may be that a battery of tests or measurements are usually undertaken  
39 together, for example FEV1 (forced expiratory volume in 1 second), FVC (forced vital  
40 capacity) and FEF<sub>25-75</sub> (average expired flow over the middle half of the FVC manoeuvre).  
41 FEV1 is the outcome most often considered for lung function due to its validity, repeatability  
42 and it is the outcome most understood by clinicians. However, the device used to measure  
43 FEV1 also measures the majority of other lung function outcomes. Therefore if FEV1 was  
44 reported in a trial, it was assumed that other lung function outcomes were also measured but  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 not necessarily analysed (classification F) unless they were specifically stated as an outcome  
4  
5 in the trial report. However, if FEV1 was not reported but other lung function outcomes were  
6  
7 then an E classification was given to FEV1 as suspicion would be raised that the latter may  
8  
9 have been selectively not reported. This was decided after discussion with clinical experts.  
10  
11

12  
13  
14  
15  
16 However, it is often difficult to assess whether an outcome was measured, and clinical  
17  
18 judgment is required. The clinical lead for each review was contacted by email and asked for  
19  
20 their input into the assessment of selective outcome reporting within the trials included in  
21  
22 their review. An assessment of whether the review outcomes had been measured and  
23  
24 reported within each trial using the classification system was completed. The clinical lead for  
25  
26 the review and KD independently assessed the trials in the review and any disagreements  
27  
28 were resolved through discussion and then checked with a third person (JJK or PRW).  
29  
30  
31

### 32 **Assessment of risk of bias for selective outcome reporting – trial level**

33  
34  
35 If one or more of the outcomes for a trial was given a high risk classification according to  
36  
37 Table 1, the trial was deemed at high risk of bias from selective reporting.  
38  
39

### 40 **Analysis**

41  
42 Descriptive results are presented. The median and interquartile range for the number of  
43  
44 review primary and secondary outcomes was calculated.  
45  
46  
47

48  
49 Data are tabulated and excerpts found in the trial reports relating to review outcomes are used  
50  
51 to support decisions made regarding ORBIT classifications and the assessment of risk of bias.  
52  
53  
54

### 55 **Results**

1  
2  
3 The CFGD group had 46 cystic fibrosis systematic reviews published as of 2010.  
4  
5  
6

7  
8 **1. Changes in outcomes between systematic review protocol and full review – review**  
9  
10 **level**

11 Protocols were available for all 46 systematic reviews. Nine protocols (20%) did not  
12 distinguish between primary and secondary outcomes. Table 3 shows the median number of  
13 primary and secondary outcomes for the 46 reviews and the changes in outcomes between  
14 protocol and full review.  
15  
16  
17  
18  
19

20  
21  
22 Eighteen reviews (39%, 18/46) had a discrepancy in outcomes between protocol and full  
23 review. Between review protocol and full review, five (28%) listed all changes, two (11%)  
24 listed some changes and 11 reviews (61%) did not mention any change in outcomes. Of the  
25 seven reviews describing the changes between protocol and full review, three provided no  
26 reason for the changes, two stated that the changes in recommendations in the Cochrane  
27 Handbook to have a maximum of three primary outcomes were the reason for downgrading  
28 outcomes and two reviews stated that they added clinically relevant outcomes that were  
29 discovered during the review process.  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

43 **2. Assessing trial reports for full ORB – outcome level**  
44

45 Of the 46 published reviews, 38 were eligible to be assessed for outcome reporting bias  
46 (Figure 1).  
47  
48  
49  
50

51 One review was excluded at this stage as the outcomes could not be assessed for ORB due to  
52 the different ways the outcome definitions could be measured and reported. The primary  
53 outcomes were psychosocial outcomes, which included any objective measure with adequate  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 psychometric properties and demonstrable reliability and validity quantifying psychological  
4 or social outcomes or both, including individual psychological adjustment, relational, social  
5 functioning and adaptation to life with cystic fibrosis.  
6  
7  
8

9  
10  
11 Therefore 37 reviews were assessed for ORB, including 280 RCTs (278 included and 2  
12 excluded due to no relevant outcome data but confirmed by review authors that they would  
13 have otherwise been included). The median number of trials per review was four (IQR 2, 8)  
14 and there was a median sample size of 21 (IQR 14, 41) per trial.  
15  
16  
17  
18  
19  
20

21  
22  
23 Review authors contacted trialists for missing outcome data in 33 reviews (89%), one stated  
24 that “trialists were not contacted but would be in updates of the review” and three reviews did  
25 not state if trialists were contacted for further data.  
26  
27  
28

29  
30  
31 The lead authors of twelve reviews assessed the included trials and gave classifications for  
32 each outcome. For thirteen reviews, authors gave input on which outcomes they expected to  
33 be measured for trials in their review and which outcomes they expected to be measured in  
34 routine clinical practice but did not classify each outcome due to time restrictions. The  
35 authors of twelve reviews did not respond to our request.  
36  
37  
38  
39  
40  
41  
42  
43  
44

45 For the twelve reviews where the authors assigned classifications, discussion was needed on  
46 all outcomes to come to an agreed classification. For the other 25 reviews it was difficult to  
47 assign a classification to all outcomes as some outcomes needed a large amount of clinical  
48 input in understanding the outcome and language used to describe the outcomes within the  
49 trial reports. Due to the number and complexity of outcomes and lack of reviewer input on  
50 the majority of reviews, it was decided that the assessment of all primary outcomes listed in  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 the full review that were well defined should take priority. Many outcomes were also split  
4  
5 into sub outcomes or ill defined to maximise the ability of a trial to contribute data to the  
6  
7 review.  
8

9  
10 The ORBIT classifications for the review primary outcomes for the 280 RCTs are shown in  
11  
12 Table 1. For the 12 reviews where reviewer input was obtained, classifications for 64  
13  
14 included trials for review secondary outcomes are also shown in Table 1. Eligible trials  
15  
16 within the reviews fully reported 383 (33.7%) review primary outcomes and 125 (18.7%)  
17  
18 review secondary outcomes. In addition to the classifications in Table 1, a 'G, no events'  
19  
20 classification (For example, mortality, where clinical judgement says it is likely to have been  
21  
22 measured and it would have been reported had any deaths occurred. Therefore, it is assumed  
23  
24 no deaths occurred during the trial.) was given to eligible trials within the reviews for 109  
25  
26 (9.5%) review primary outcomes and 22 (3.3%) review secondary outcomes. Due to  
27  
28 limited reviewer input or the lack of a standard definition, we were unable to assess outcomes  
29  
30 (including: adverse events, symptoms, complications, biochemical measures of glycaemic  
31  
32 control, symptoms of sleep disordered breathing and measures of specific indices of strength,  
33  
34 mass, effort and general fatigue) for 102 trials for review primary outcomes and 59 trials for  
35  
36 review secondary outcomes.  
37  
38  
39  
40  
41  
42  
43

#### 44 **Assessment of risk of bias from selective outcome reporting – trial level**

45  
46 Eighteen reviews (49%) had not yet assessed the risk of bias for selective outcome reporting  
47  
48 as although the Cochrane guidance on the risk of bias was introduced in 2008 and the cut off  
49  
50 for this study was the beginning of 2010, these reviews were still to be updated. Seventeen  
51  
52 reviews (46%) had assessed the risk of bias for all included trials and two reviews (5%)  
53  
54 assessed this for some of their included trials.  
55  
56  
57  
58  
59  
60

1  
2  
3 As we were unable to assess secondary outcomes for ORB for all reviews, the risk of bias  
4 assessments were made based on classifications of primary outcomes in order to be consistent  
5 across reviews. Only five (14%) of the 37 reviews had no trials at high risk of bias based on  
6 the review primary outcomes only. Table 4 shows the risk of bias for selective outcome  
7 reporting as defined in this study and also as assessed within the published reviews for the  
8 280 trials assessed for ORB based on the consideration of review primary outcomes only. It  
9 was found that 69% of trials had either not been assessed for selective reporting or were  
10 assessed as an unclear risk.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

25 Table 5 shows the risk of bias for selective outcome reporting based on the consideration of  
26 review primary and secondary outcomes separately for the 12 reviews (64 trials) where  
27 reviewers also provided classifications. This was to see if decisions regarding risk of bias  
28 would change if we considered all outcomes. Only four (6%) of the 64 RCTs had a low risk  
29 of bias when considering all outcomes.  
30  
31  
32  
33  
34  
35  
36  
37  
38

39 Discrepancies in the risk of bias when considering all outcomes arose in 34 (53%) trials; 31  
40 were at low risk when considering review primary outcomes only but high risk of bias  
41 (excluding G classifications: 13, G classification only: 18) when considering all outcomes; 3  
42 were at high risk (G classifications only) when considering review primary outcomes only  
43 but high risk (excluding G classifications) when considering all outcomes. This often  
44 occurred in reviews where there was only one or two primary outcomes and a large number of  
45 secondary outcomes.  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Based on all review outcomes, none of the 12 reviews had all included trials at low risk of  
4  
5 bias.  
6  
7

## 8 9 **Discussion**

10  
11  
12 This is the first study to consider all review efficacy outcomes in an ORB assessment which  
13  
14 has allowed us to make practical recommendations on assessing the risk of bias of selective  
15  
16 reporting for systematic reviews at both the review and trial level. Over a third of Cochrane  
17  
18 cystic fibrosis reviews (39%) examined had a discrepancy in outcomes between the review  
19  
20 protocol and full review. This compares to 22% of reviews (64/288) that contained a  
21  
22 discrepancy in at least one outcome measure in the main ORBIT study which looked at  
23  
24 reviews covering all 50 Cochrane review groups (Kirkham et al 2010a). However, this is  
25  
26 confounded by the different publication date ranges of the reviews (assessed as up to date  
27  
28 between 2006 and 2009). Furthermore, for the cystic fibrosis reviews outcome reporting bias  
29  
30 was suspected in at least one randomised controlled trial in 86% of reviews when considering  
31  
32 all review primary outcomes. The prevalence of reviews containing at least one trial with  
33  
34 high suspicion of outcome reporting bias from ORBIT, when only a single primary outcome  
35  
36 was considered was substantially lower at 34% (96/283) (Kirkham et al 2010b). While this  
37  
38 study is limited only to CF trials, it is clear that the problem of outcome reporting bias is  
39  
40 much larger when considering more than just the single primary review outcome of  
41  
42 importance that was used in the ORBIT study.  
43  
44  
45  
46  
47  
48  
49  
50

51 A study by von Mosch and Dwan (2011) that compared the reporting in trial reports of CF to  
52  
53 the CONSORT statement found that from a maximum of 57 points available, the scores rose  
54  
55 from a median of 17.5 (Inter quartile range (IQR) 15.5-24.5) in 1994 to a median of 32 (IQR  
56  
57  
58  
59  
60

1  
2  
3 22.8-41.5) in 2008. Along with the current study, this also indicates that there is still room  
4  
5 for an improvement in the reporting of outcomes.  
6  
7

8  
9  
10 Use of the ORBIT classification system offered a robust methodology for assessing the risk  
11 of bias for trials included within a systematic review. When considering the 64 trials in the  
12  
13 12 reviews where it was possible to assess both primary and secondary outcomes, when  
14 basing the risk of bias assessment on review primary outcomes, 45% of trials were at high  
15 risk of bias and when using all outcomes in the assessment, 94% were at high risk of bias.  
16  
17 Using the current selective reporting item of the current Cochrane risk of bias tool, 69% of  
18 trials included in CF reviews were assessed by reviewers as 'unclear' risk of bias or not  
19 assessed at all, indicating the need for more informed guidance on assigning risk of bias in  
20 the systematic review process for all outcomes within a review.  
21  
22  
23  
24  
25  
26  
27  
28  
29

30  
31  
32 The ORBIT classification system has already been validated as part of the original project.  
33  
34 Sensitivity results for predicting that the outcome had been measured (G-classification) was  
35 92% (23/25, 95% CI 81% to 100%), while the specificity for predicting that the outcome had  
36 not been measured (H-classification) was 77% (23/30, 95% CI 62% to 92%). With the  
37 additional requirement to assess all outcomes in this project, there were an increasing number  
38 of outcomes that were not mentioned in the trials reports and therefore clinical judgement  
39 was needed as to whether the outcome of interest was likely to have been measured in a  
40 particular trial. Many review authors did not respond to our request to provide classifications  
41 (68%), but for those with no response we did obtain clinical input for the primary outcome  
42 from within the CFGD group. Although we can not exclude the possibility of response bias it  
43 is likely the decision to respond was influenced by time commitments rather than review  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 characteristics. However, these assessments will be provided to the review authors when their  
4  
5 review is due to be updated.  
6  
7

8  
9 Reviewers should ensure that changes between protocol and reviews are listed and  
10  
11 justifications provided to enhance the validity of these decisions. Eligible trials should not be  
12  
13 excluded on the basis of “No relevant outcome data” because although an outcome was not  
14  
15 reported it may have been measured and contact with the authors is advised. Reviewers  
16  
17 should be encouraged to consider trials that have not reported an outcome of interest and to  
18  
19 assess whether selective reporting has occurred for all review outcomes. They should  
20  
21 consider the amount of missing data from their meta-analysis (i.e. the percentage of the  
22  
23 sample sizes of the studies that were included compared to those that would have been  
24  
25 eligible to be included in the meta-analysis but no outcome data reported) and this  
26  
27 information should be included along with the pooled effect estimate. If appropriate, a  
28  
29 sensitivity analysis should be applied to assess the robustness of the conclusions of the  
30  
31 review, such as an imputation approach (Williamson and Gamble 2005), the Copas bound for  
32  
33 maximum bias (Copas et al, 2004; Williamson and Gamble 2007, Dwan et al 2010) or a  
34  
35 model based correction (Copas et al, 2013).  
36  
37  
38  
39  
40  
41  
42

43 Individuals conducting systematic reviews need to address explicitly the issue of missing  
44  
45 outcome data for their review to be considered a reliable source of evidence. Extra care is  
46  
47 required during data extraction, reviewers should identify when a trial reports that an  
48  
49 outcome was measured but no results were reported or events observed, and contact with  
50  
51 trialists should be encouraged. Contacting authors is encouraged by the CRG and is standard  
52  
53 practice within CFGD reviews which is reflected in our results as 89% of reviews stated that  
54  
55 they contacted authors for extra information on outcomes.  
56  
57  
58  
59  
60

1  
2  
3  
4  
5 It is recommended that review authors ensure that they limit the number of outcomes in the  
6 review and define them clearly as this will allow easier assessments of selective reporting,  
7 which can be done during data extraction of the included trials as long as a knowledgeable  
8 clinical person is involved. Lung function was specified as the first primary outcome in  
9 nineteen reviews (50%), as the second or third primary outcome in 11 reviews (29%), as a  
10 secondary outcome in six reviews (16%) and it was not included as an outcome in only one  
11 review (5%). However, as discussed earlier, lung function can be measured in different ways  
12 (FEV1, FVC, mid forced expiratory flow (FEF), peak expiratory flow rate (PEFR), residual  
13 volume (RV), total lung capacity (TLC), Lung clearance index (LCI) and maximum  
14 expiratory flow (MEF).) These outcomes can then be analysed and reported in different ways  
15 such as: percentage predicted, litres, litres/second and post treatment, absolute change from  
16 baseline, relative change from baseline or annual rate of change. Therefore there is a large  
17 scope for selective reporting. One solution is the development of a core outcome set for  
18 cystic fibrosis (Ramsey and Boat 1994 , Sinha et al 2008, Clarke 2007).  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37

### 38 **Unanswered questions and future research**

39  
40 Work is needed to consider what the best method is to assess the impact of ORB on the  
41 results of the meta-analysis when there are multiple outcomes. Multivariate meta-analysis  
42 has been suggested by Kirkham et al 2012 and a model based correction has been suggested  
43 by Copas et al (2013).  
44  
45  
46  
47  
48  
49  
50

### 51 **Conclusion**

52  
53 Systematic reviews need to clearly state the primary and secondary outcomes that they will  
54 consider and be consistent between review protocol and full review.  
55  
56  
57  
58  
59  
60

1  
2  
3 Outcome reporting bias is a major problem for systematic reviews and more guidance needs  
4  
5 to be included in the Cochrane handbook to allow assessment of this important item within  
6  
7 the risk of bias tool. We recommend that an outcome matrix be completed during the  
8  
9 production of a review to allow an ORB assessment for all review outcomes which can then  
10  
11 inform the risk of bias assessment.  
12  
13

14 A core set of outcomes should be agreed upon for cystic fibrosis which in turn will have a  
15  
16 positive impact on systematic reviews as future trials are conducted they should specifically  
17  
18 set out to measure and report these outcomes therefore reducing the prevalence of selective  
19  
20 reporting.  
21  
22

## 23 24 25 **Abbreviations** 26

27		
28	CFGD	Cystic Fibrosis and Genetic Disorders
29		
30	CRG	Cochrane Review Group
31		
32	COMET	Core Outcome Measures in Effectiveness Trials
33		
34	CONSORT	Consolidated Standards of Reporting Trails
35		
36		
37	FEF <sub>25-75</sub>	Average expired flow over the middle half of the FVC manoeuvre
38		
39	FEV1	Forced expiratory volume in 1 second
40		
41	FVC	Forced vital capacity
42		
43		
44	NROD	No Relevant Outcome Data
45		
46	ORB	Outcome Reporting Bias
47		
48	ORBIT	Outcome Reporting Bias In Trials
49		
50	PEFR	Peak Expiratory Flow Rate
51		
52		
53	RCT	Randomised Controlled Trial
54		
55		
56		
57		
58		
59		
60		

## 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.



1  
2  
3  
4  
5 Copas J, Dwan K, Kirkham JJ, Williamson PR. A model based correction for outcome  
6 reporting bias in meta-analysis. (submitted 2013).  
7  
8

9  
10  
11 Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, et al. Systematic Review of the  
12 Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. PLoS ONE  
13 2008, 3(8): e3081.  
14  
15  
16  
17

18  
19  
20 Dwan K, Gamble C, Kolamunnage-Dona R, Mohammed S, Powell C, Williamson PR.  
21 Assessing the potential for outcome reporting bias in a review: A tutorial. *Trials* 2010, **11**:52  
22 doi:10.1186/1745-6215-11-52.  
23  
24  
25  
26

27  
28  
29 Green, S. B. & Byar, D. P. Using observational data from registries to compare treatments:  
30 the fallacy of omnimetrics. *Stat Med*, 1984; 3, 361-73.  
31  
32

33  
34  
35  
36 Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions*  
37 Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from  
38 www.cochrane-handbook.org.  
39  
40  
41  
42

43  
44  
45 Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of  
46 controlled clinical trials. *BMJ* 2001; 323: 42-46.  
47  
48

49  
50  
51 Kirkham JJ, Altman DG, Williamson PR. Bias Due to Changes in Specified Outcomes during  
52 the Systematic Review Process. PLoS ONE 2010a; 5(3):  
53 e9810.doi:10.1371/journal.pone.0009810  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5 Kirkham JJ, Dwan K, Dodd S, Altman DG, Smyth R, Jacoby A, Gamble C, Williamson PR.

6  
7 The impact of outcome reporting bias in randomised controlled trials on a cohort of  
8  
9 systematic reviews. *BMJ* 2010b; 340: c365  
10

11  
12  
13  
14 Kirkham JJ, Riley RD and Williamson PR. A multivariate meta-analysis approach for  
15  
16 reducing the impact of outcome reporting bias in systematic reviews. *Statistics in Medicine*  
17  
18 2012; 31:20, 2179-2195.  
19

20  
21  
22  
23 Ramsey BW, Boat TF (1994) Outcome measures for clinical trials in cystic fibrosis.  
24  
25 Summary of a Cystic Fibrosis Foundation consensus conference. *J Pediatr* 1994; 124: 177–  
26  
27 192.  
28

29  
30  
31  
32 Sinha I, Jones L, Smyth RL, Williamson PR. A Systematic Review of Studies That Aim to  
33  
34 Determine Which Outcomes to Measure in Clinical Trials in Children . *PLoS Med* 2008;  
35  
36 5(4): e96. doi:10.1371/journal.pmed.0050096  
37

38  
39  
40 Smyth AR, Walters S. Prophylactic anti-staphylococcal antibiotics for cystic fibrosis.  
41  
42 *Cochrane Database of Systematic Reviews* 2003, Issue 3. Art. No.: CD001912. DOI:  
43  
44 10.1002/14651858.CD001912.  
45

46  
47  
48  
49 *The Cochrane Library*, Issue 4, 2009. Chichester: Wiley.  
50

51  
52  
53  
54 Williamson, P. R., Gamble, C., Altman, D. G. & Hutton, J. L. Outcome selection bias in  
55  
56 meta-analysis. *Stat Methods Med Res*, 2005; 14, 515-24.  
57

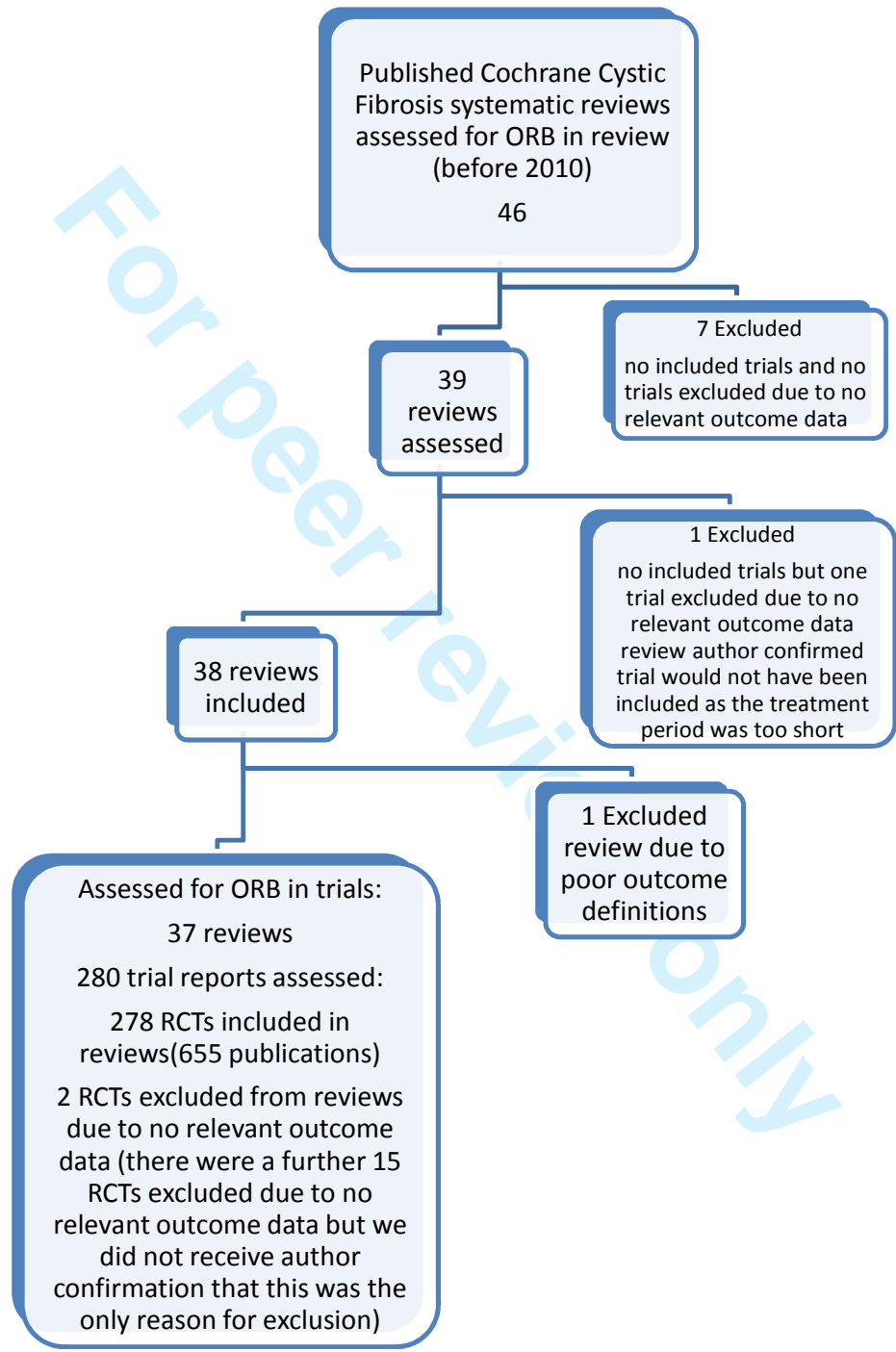
1  
2  
3 Williamson PR, Gamble C: Identification and impact of outcome selection bias in meta-  
4  
5 analysis. *Statistics in Medicine* 2005, 24(10):1547-61.  
6  
7

8  
9  
10 Williamson PR, Gamble C: Application and investigation of a bound for outcome reporting  
11  
12 bias. *Trials* 2007, 8:9.  
13

14  
15  
16 von Mosch S, Dwan K. Reporting of harm and efficacy outcomes of randomised clinical  
17  
18 trials in cystic fibrosis before and after the publication of the revised CONSORT  
19  
20 (CONsolidated Standards Of Reporting Trials) statement in 2001 – a comparison of trial  
21  
22 reports published from 1994 and from 2008. The University of Liverpool (thesis) 2011.  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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) <sup>1</sup>	Number of trials (percentage overall) <sup>2</sup>
<i>Clear that the outcome was measured and analysed</i>					
<b>A</b>	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>B</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%)
<b>C</b>	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%)
<b>D</b>	States outcome analysed but no results reported.	None	High risk	0 (0)	0 (0)
<i>Clear that the outcome was measured</i>					
<b>E</b>	Clear that outcome was measured but not necessarily analysed.	None	High risk	59 (5.2%)	26 (3.9%)
<b>F</b>	Clear that outcome was measured but not necessarily analysed.	None	Low risk	110 (9.7%)	15 (2.2%)
<i>Unclear that the outcome was measured</i>					
<b>G</b>	Not mentioned but clinical judgment says likely to have been measured and analysed.	None	High risk	195 (17.1%)	197 (29.4%)
<b>H</b>	Not mentioned but clinical judgment says unlikely to have been measured.	None	Low risk	141 (12.4%)	256 (38.2%)
<i>Clear that the outcome was NOT measured</i>					
<b>I</b>	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	○ (A) <sup>1</sup>	○ (A) <sup>1</sup>	✓	✓	✓	✗ (H) <sup>2</sup>	✓
Schlesinger 1984	✗ (H) <sup>3</sup>	✗ (H) <sup>3</sup>	✓	○ (C) <sup>4</sup>	✓	✗ (H) <sup>2</sup>	✓
Stutman 2002	✓	✓	✓	✓	✓	✗ (H) <sup>2</sup>	✗
Weaver 1994	✗ (H) <sup>3</sup>	✗ (H) <sup>3</sup>	✓	○ (C) <sup>4</sup>	✓	✗ (H) <sup>2</sup>	✗

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.

- |   |
|---|
| <ul style="list-style-type: none"> <li>✓ indicates full reporting of results for treatment comparison of interest</li> <li>✗ indicates no reporting</li> <li>○ indicates partial reporting</li> </ul> |
|---|

**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) <sup>1</sup>	14 (38%)	
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	4 (44%)	
Reviews which have <i>upgraded</i> 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) <sup>1</sup>	3 (8%) (3 outcomes)	
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	0	
Reviews which have <i>downgraded</i> 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) <sup>1</sup>	9 (24%) (16 outcomes; min 1, max 5)	
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	1 (11%) (2 outcomes)	
Reviews which have <i>added</i> 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) <sup>1</sup>	2 (5%) (3 outcomes)	2 (5%) (4 outcomes; min 1, max 3)
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	1(11%) (1 outcome)	2 (22%) (2 outcomes)
Reviews which have <i>excluded</i> 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) <sup>1</sup>	2 (5%) (10 outcomes; min 1, max 9)	3 (8%) (5 outcomes; min 1; max 2)
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	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.

<b>SELECTIVE OUTCOME REPORTING</b>	
<b>Are reports of the study free of suggestion of selective outcome reporting? [Short form: <i>Free of selective reporting?</i>]</b>	
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<sup>1§</sup>, Jamie J Kirkham<sup>1</sup>, Paula R Williamson<sup>1</sup>, Carrol Gamble<sup>1</sup>

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

<sup>§</sup>Corresponding author

Email addresses:

KD: [kerry.dwan@liverpool.ac.uk](mailto:kerry.dwan@liverpool.ac.uk)

JJK: [jjk@liverpool.ac.uk](mailto:jjk@liverpool.ac.uk)

PRW: [prw@liverpool.ac.uk](mailto:prw@liverpool.ac.uk)

CG: [c.gamble@liverpool.ac.uk](mailto: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~~

1  
2  
3 considering a single review primary outcome. Systematic reviews with multiple  
4  
5 primary and secondary efficacy outcomes are at greater risk of ORB.  
6

- 7 • Clearer guidance is needed on how to assess the ~~‘overall’~~ risk of bias as a result of  
8 ORB-selective outcome reporting for each included trial within a systematic review,  
9  
10 when considering multiple outcomes.  
11
- 12 • The development of a core outcome set in Cystic Fibrosis would help reduce the  
13  
14 problem of ORB.  
15  
16  
17

### 18 19 **Strengths and limitations**

20  
21 This is the first study to consider the assessment of outcome reporting bias in all efficacy  
22  
23 review outcomes. However, this is limited to reviews of cystic fibrosis.  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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

1  
2  
3 publication (Williamson and Hutton 2000). In a systematic empirical assessment of  
4  
5 Cochrane reviews within which a single review primary outcome could be identified  
6  
7 (Kirkham et al 2010b), ORB was suspected in at least one randomised controlled trial in more  
8  
9 than a third of the systematic reviews that were examined (35%). This study may have  
10  
11 underestimated this problem as review primary outcomes are chosen due to their clinical  
12  
13 importance ~~so and~~ are more likely to have been measured and reported in trials. Therefore,  
14  
15 there is increasing concern regarding the prevalence and impact of outcome reporting bias in  
16  
17 reviews where multiple primary outcomes are specified, or in secondary outcomes.  
18  
19

20  
21  
22 Systematic reviews in cystic fibrosis are characterised by inclusion of small randomised trials  
23  
24 specifying multiple primary outcomes. Reporting standards for trials of cystic fibrosis have  
25  
26 also been shown to be low when comparing trial reports to the Consolidated Standards of  
27  
28 Reporting Trails (CONSORT) statement (von Mosch and Dwan 2011). The aims of this  
29  
30 current study were to  
31  
32

- 33  
34 1. Examine the potential for bias created by review authors by identifying  
35  
36 inconsistencies between outcomes published in review protocols and in the associated  
37  
38 published reviews  
39
- 40  
41 2. Determine the prevalence of ORB in trials in systematic reviews of CF, extending  
42  
43 previous work by considering all review efficacy outcomes (multiple primary and  
44  
45 secondary).  
46
- 47  
48 2-3. Assess the risk of bias of trials from selective outcome reporting when considering  
49  
50 review primary outcomes only in comparison to all review outcomes.  
51  
52

## 53 54 **Methods**

55  
56  
57  
58  
59  
60

1  
2  
3 A cohort of systematic reviews published by the Cochrane Cystic Fibrosis and Genetic  
4 Disorders (CFGD) group on the Cochrane Library before 2010 were identified (The  
5 Cochrane Library, 2009). Reviews were eligible for inclusion if they compared interventions  
6 for cystic fibrosis and identified one or more eligible RCTs. RCTs that had been excluded (in  
7 the “characteristics of excluded studies” section) were also checked for any suggestion of  
8 outcome reporting bias. For example, if a review had excluded trials as a result of ‘no  
9 relevant outcome data (NROD)’, then these trials were also scrutinised for the presence of  
10 ORB and included in the assessment.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22

### 23 **Changes in outcomes between systematic review protocol and full review – review level**

24 The numbers of primary and secondary outcomes per review were compared to the  
25 recommendations for the number of outcomes (no more than three primary outcomes and a  
26 limited number of secondary outcomes) to include in a review in the Cochrane Handbook  
27 (Higgins and Green 2011). If a review did not distinguish between primary and secondary  
28 outcomes, the first three outcomes listed were taken to be the primary outcomes and the rest  
29 were considered as secondary outcomes. Protocols of the systematic reviews were accessed  
30 and outcomes stated in the protocol were compared to those stated in the full review.  
31 Changes in outcomes were identified and categorised by one author (KD) as: primary  
32 outcome downgraded to secondary (*downgrade*); secondary outcome upgraded to primary  
33 (*upgrade*); a new outcome not stated in the protocol was added to the full review (*addition*)  
34 or an outcome stated in the protocol was omitted from the full review (*omission*). If there had  
35 been a change in outcomes, the section ‘changes between protocol and review’ was examined  
36 for a declaration and explanation of the changes.  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53

### 54 **Assessing trial reports for full ORB – outcome level**



1  
2  
3 For each eligible systematic review, all reports relating to included studies and studies  
4 excluded due to no relevant outcome data were obtained. Reviews were checked to see  
5 whether review authors had contacted trialists for further information or data for outcomes.  
6  
7 Where this was not clear in the review, review authors were asked to clarify.  
8  
9  
10

11  
12  
13 A nine-point classification system (Table 1) developed for missing or incomplete outcome  
14 reporting in randomised trials (Kirkham et al, 2010b) was used to make an assessment of the  
15 risk of bias. Table 1 also provides examples of outcomes that were not assessed because they  
16 had poor outcome definitions. An outcome matrix (Table 2) was created for each review  
17 using the ORBIT matrix generator (<http://ctrc.liv.ac.uk/orbit/>), with studies listed in the rows  
18 and review primary and secondary outcomes listed in the columns with the ORBIT  
19 classifications (Table 1) given for each review outcome that was not fully reported (e.g. not  
20 reported or partially reported e.g.  $p>0.05$ ).  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

32  
33 The outcomes listed or detailed in the method section and the outcomes reported in the results  
34 section were compared for all trial publications to determine whether each outcome of the  
35 systematic review was measured and analysed. In some instances it may be obvious that an  
36 outcome was measured given the other outcomes reported. For example, if cause-specific  
37 mortality is reported then overall mortality must have been measured, even if not reported. In  
38 other situations it may be that a battery of tests or measurements are usually undertaken  
39 together, for example FEV1 (forced expiratory volume in 1 second) ~~and~~ FVC (forced vital  
40 capacity) and FEF<sub>25-75</sub> (average expired flow over the middle half of the FVC manoeuvre). ~~If~~  
41 FVC is reported but FEV1 is not, suspicion should be raised that the latter may have been  
42 selectively not reported. FEV1 is the outcome most often considered for lung function due to  
43 its validity, repeatability and it is the outcome most understood by clinicians. However, the  
44 device used to measure FEV1 also measures the majority of other lung function outcomes.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Therefore if FEV1 was reported in a trial, it was assumed that other lung function outcomes  
4 were also measured but not necessarily analysed (classification F) unless they were  
5 specifically stated as an outcome in the trial report. However, if FEV1 was not reported but  
6 other lung function outcomes were then an E classification was given to FEV1 as suspicion  
7 would be raised that the latter may have been selectively not reported. This was decided after  
8 discussion with clinical experts.  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19

20  
21 However, it is often difficult to assess whether an outcome was measured, and clinical  
22 judgment is required. The clinical lead for each review was contacted by email and asked for  
23 their input into the assessment of selective outcome reporting within the trials included in  
24 their review. An assessment of whether the review outcomes had been measured and  
25 reported within each trial using the classification system was completed. The clinical lead for  
26 the review and KD independently assessed the trials in the review and any disagreements  
27 were resolved through discussion and then checked with a third person (JJK or PRW).  
28  
29  
30  
31  
32  
33  
34  
35  
36

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

38  
39  
40 If one or more of the outcomes for a trial was given a high risk classification according to  
41 Table 1, the trial was deemed at high risk of bias from selective reporting.  
42  
43  
44

### 45 **Analysis**

46  
47 Descriptive results are presented. The median and interquartile range for the number of  
48 review primary and secondary outcomes was calculated.  
49  
50  
51

52  
53  
54 Data are tabulated and excerpts found in the trial reports relating to review outcomes are used  
55 to support decisions made regarding ORBIT classifications and the assessment of risk of bias.  
56  
57  
58  
59  
60

## 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).

1  
2  
3 One review was excluded at this stage as the outcomes could not be assessed for ORB due to  
4 the different ways the outcome definitions could be measured and reported. The primary  
5 outcomes were psychosocial outcomes, which included any objective measure with adequate  
6 psychometric properties and demonstrable reliability and validity quantifying psychological  
7 or social outcomes or both, including individual psychological adjustment, relational, social  
8 functioning and adaptation to life with cystic fibrosis.  
9

10  
11  
12 Therefore 37 reviews were assessed for ORB, including 280 RCTs (278 included and 2  
13 excluded due to no relevant outcome data but confirmed by review authors that they would  
14 have otherwise been included). The median number of trials per review was four (IQR 2, 8)  
15 and there was a median sample size of 21 (IQR 14, 41) per trial.  
16  
17

18  
19  
20 Review authors contacted trialists for missing outcome data in 33 reviews (89%), one stated  
21 that “trialists were not contacted but would be in updates of the review” and three reviews did  
22 not state if trialists were contacted for further data.  
23  
24  
25

26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39 ~~Lead authors of each review were contacted.~~—The lead authors of twelve reviews assessed the  
40 included trials and gave classifications for each outcome. For thirteen reviews, authors gave  
41 input on which outcomes they expected to be measured for trials in their review and which  
42 outcomes they expected to be measured in routine clinical practice but did not classify each  
43 outcome due to time restrictions. The authors of twelve reviews did not respond to our  
44 request.  
45  
46  
47  
48  
49  
50

51  
52  
53  
54 For the twelve reviews where the authors assigned classifications, discussion was needed on  
55 all outcomes to come to an agreed classification. For the other 25 reviews it was difficult to  
56  
57  
58  
59  
60

1  
2  
3 assign a classification to all outcomes as some outcomes needed a large amount of clinical  
4 input in understanding the outcome and language used to describe the outcomes within the  
5  
6  
7 trial reports. Due to the number and complexity of outcomes and lack of reviewer input on  
8  
9  
10 the majority of reviews, it was decided that the assessment of all primary outcomes listed in  
11  
12 the full review that were well defined should take priority. Many outcomes were also split  
13  
14 into sub outcomes or ill defined to maximise the ability of a trial to contribute data to the  
15  
16 review. ~~For example lung function was often split into FEV1 (Volume that has been exhaled~~  
17 ~~at the end of the first second of forced expiration), FVC (Forced vital capacity), PEFR (Peak~~  
18 ~~expiratory flow rate), FEF<sub>25-75</sub> (average expired flow over the middle half of the FVC~~  
19 ~~manoeuvre) and these were assessed separately. FEV1 is the outcome most often considered~~  
20 ~~for lung function due to its validity, repeatability and it is the outcome most understood by~~  
21 ~~clinicians. However, the device used to measure FEV1 also measures the majority of other~~  
22 ~~lung function outcomes. Therefore if FEV1 was reported in a trial, it was assumed that other~~  
23 ~~lung function outcomes were also measured but not necessarily analysed (classification F)~~  
24 ~~unless they were specifically stated as an outcome in the trial report. However, if FEV1 was~~  
25 ~~not reported but other lung function outcomes were then an E classification was given to~~  
26 ~~FEV1. This was decided after discussion with clinical experts.~~

27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43 The ORBIT classifications for the review primary outcomes for the 280 RCTs are shown in  
44  
45 Table 1. For the 12 reviews where reviewer input was obtained, classifications for 64  
46  
47 included trials for review secondary outcomes are also shown in Table 1. Eligible trials  
48 within the reviews fully reported 383 (33.7%) review primary outcomes and 125 (18.7%)  
49 review secondary outcomes. In addition to these classifications in Table 1, a 'G, no events'  
50  
51 classification (For example, mortality, where clinical judgement says it is likely to have been  
52  
53 measured and it would have been reported had any deaths occurred. Therefore, it is assumed  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 no deaths occurred during the trial.) was given to eligible trials within the reviews for 109  
4 trials (9.5%) for review primary outcomes and 22 trials (3.3%) for review secondary  
5 outcomes. Due to limited reviewer input or the lack of a standard definition for some  
6 outcomes. We were unable to assess outcomes (including: adverse events, symptoms,  
7 complications, biochemical measures of glycaemic control, symptoms of sleep disordered  
8 breathing and measures of specific indices of strength, mass, effort and general fatigue) for  
9 102 trials for review primary outcomes and 59 trials for review secondary outcomes.  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

### 21 **Assessment of risk of bias from selective outcome reporting – trial level**

22  
23 Eighteen reviews (49%) had not yet assessed the risk of bias for selective outcome reporting  
24 as although they had not been updated since the new the Cochrane guidance on the risk of  
25 bias was introduced in 2008 and the cut off for this study was the beginning of 2010, these  
26 reviews were still to be updated ~~had been introduced and prior to this study.~~ Seventeen  
27 reviews (46%) had assessed the risk of bias for all included trials and two reviews (5%)  
28 assessed this for some of their included trials.  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

40 As we were unable to assess secondary outcomes for ORB for all reviews, the risk of bias  
41 assessments were made based on classifications of primary outcomes in order to be consistent  
42 across reviews. Only five (14%) of the 37 reviews had no trials at high risk of bias based on  
43 the review primary outcomes only.  
44  
45  
46  
47

48 Table 4 shows the risk of bias for selective outcome reporting as defined in this study and  
49 also as assessed within the published reviews for the 280 trials assessed for ORB based on the  
50 consideration of review primary outcomes only. It was found that 69% of trials had either  
51 not been assessed for selective reporting or were assessed as an unclear risk.  
52  
53  
54  
55  
56  
57  
58  
59  
60

~~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 where 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

1  
2  
3 protocol and full review. This compares to 22% of reviews (64/288) that contained a  
4  
5 discrepancy in at least one outcome measure in the main ORBIT study which looked at  
6  
7 reviews covering all 50 Cochrane review groups (Kirkham et al 2010a). However, this is  
8  
9 confounded by the different publication date ranges of times in which the reviews (assessed as  
10  
11 up to date between 2006 and 2009) were published. Furthermore, for the cystic fibrosis  
12  
13 reviews outcome reporting bias was suspected in at least one randomised controlled trial in  
14  
15 86% of reviews when considering all review primary outcomes. The prevalence of reviews  
16  
17 containing at least one trial with high suspicion of outcome reporting bias from ORBIT, when  
18  
19 only a single primary outcome was considered was substantially lower at 34% (96/283)  
20  
21 (Kirkham et al 2010b). While this study is limited only to CF trials, it is clear that the  
22  
23 problem of outcome reporting bias is much larger when considering more than just the single  
24  
25 primary review outcome of importance that was used in the ORBIT study.  
26  
27  
28  
29  
30  
31

32 A study by von Mosch and Dwan (2011) that compared the reporting in trial reports of CF to  
33 the CONSORT statement found that from a maximum of 57 points available, the scores rose  
34 from a median of 17.5 (Inter quartile range (IQR) 15.5-24.5) in 1994 to a median of 32 (IQR  
35 22.8-41.5) in 2008. Along with the current study, this also indicates that there is still room  
36 for an improvement in the reporting of outcomes.  
37  
38  
39  
40  
41  
42  
43  
44

45 Use of the ORBIT classification system offered a robust methodology for assessing the risk  
46  
47 of bias for trials included within a systematic review. When considering the 64 trials in the  
48  
49 12 reviews where it was possible to assess both primary and secondary outcomes, when  
50  
51 basing the risk of bias assessment on review primary outcomes, 45% of trials were at high  
52  
53 risk of bias and when using all outcomes in the assessment, 94% were at high risk of bias.  
54  
55 Using the current selective reporting item of the current Cochrane risk of bias tool, 69% of  
56  
57  
58  
59  
60



1  
2  
3 trials included in CF reviews were assessed by reviewers as ‘unclear’ risk of bias or not  
4  
5 assessed at all, indicating the need for more informed guidance on assigning risk of bias in  
6  
7 the systematic review process for all outcomes within a review.  
8  
9

10  
11 The ORBIT classification system has already been validated as part of the original project.  
12  
13 Sensitivity results for predicting that the outcome had been measured (G-classification) was  
14  
15 92% (23/25, 95% CI 81% to 100%), while the specificity for predicting that the outcome had  
16  
17 not been measured (H-classification) was 77% (23/30, 95% CI 62% to 92%). With the  
18  
19 additional requirement to assess all outcomes in this project, there were an increasing number  
20  
21 of outcomes that were not mentioned in the trials reports and therefore clinical judgement  
22  
23 was needed as to whether the outcome of interest was likely to have been measured in a  
24  
25 particular trial. Many review authors did not respond to our request to provide classifications  
26  
27 (68%), but for those with no response we did obtain clinical input from within the CFGD  
28  
29 group for the primary outcome from within the CFGD group. Although we can not exclude  
30  
31 the possibility of response bias it is likely the decision to respond was influenced by time  
32  
33 commitments rather than review characteristics. therefore only primary outcomes were  
34  
35 assessed within the majority of reviews due to the clinical complexity of many of the  
36  
37 secondary outcomes. However, these assessments will be provided to the review authors when  
38  
39 their review is due to be updated.  
40  
41  
42  
43  
44  
45

46  
47 Reviewers should ensure that changes between protocol and reviews are listed and  
48  
49 justifications provided to enhance the validity of these decisions. Eligible trials should not be  
50  
51 excluded on the basis of “No relevant outcome data” because although an outcome was not  
52  
53 reported it may have been measured and contact with the authors is advised. Reviewers  
54  
55 should be encouraged to consider trials that have not reported an outcome of interest and to  
56  
57  
58  
59  
60

1  
2  
3 assess whether selective reporting has occurred for all review outcomes. They should  
4  
5 consider the amount of missing data from their meta-analysis (i.e. the percentage of the  
6  
7 sample sizes of the studies that were included compared to those that would have been  
8  
9 eligible to be included in the meta-analysis but no outcome data reported) and this  
10  
11 information should be included along with the pooled effect estimate. If appropriate, a  
12  
13 sensitivity analysis should be applied to assess the robustness of the conclusions of the  
14  
15 review, such as an imputation approach (Williamson and Gamble 2005), the Copas bound for  
16  
17 maximum bias (Copas et al, 2004; Williamson and Gamble 2007, Dwan et al 2010) or a  
18  
19 model based correction (Copas et al, 2013).  
20  
21  
22  
23  
24

25 Individuals conducting systematic reviews need to address explicitly the issue of missing  
26  
27 outcome data for their review to be considered a reliable source of evidence. Extra care is  
28  
29 required during data extraction, reviewers should identify when a trial reports that an  
30  
31 outcome was measured but no results were reported or events observed, and contact with  
32  
33 trialists should be encouraged. Contacting authors is encouraged by the CRG and is standard  
34  
35 practice within CFGD reviews which is reflected in our results as 89% of reviews stated that  
36  
37 they contacted authors for extra information on outcomes.  
38  
39  
40  
41  
42

43 It is recommended that review authors ensure that they limit the number of outcomes in the  
44  
45 review and define them clearly. ~~Reviewers also need to ensure that outcomes are well~~  
46  
47 ~~defined as t~~ This will allow easier assessments of selective reporting, which can be done  
48  
49 during data extraction of the included trials as long as a knowledgeable clinical person is  
50  
51 involved. Lung function was specified as the first primary outcome in nineteen reviews  
52  
53 (50%), as the second or third primary outcome in 11 reviews (29%), as a secondary outcome  
54  
55 in six reviews (16%) and it was not included as an outcome in only one review (5%).  
56  
57  
58  
59  
60

1  
2  
3 However, as discussed earlier, lung function can be measured in different ways ~~it is often~~  
4 ~~split into ‘suboutcomes’, including~~ (-FEV1, FVC, mid forced expiratory flow (FEF), peak  
5 expiratory flow rate (PEFR), residual volume (RV), total lung capacity (TLC), Lung  
6  
7 clearance index (LCI) and maximum expiratory flow (MEF).) These outcomes can then be  
8  
9 analysed and reported in different ways such as: percentage% predicted, litres, litres/second  
10  
11 and post treatment, absolute change from baseline, relative change from baseline or annual  
12  
13 rate of change. Therefore there is a large scope for selective reporting. One solution is the  
14  
15 development of a core outcome set for cystic fibrosis (Ramsey and Boat 1994 , Sinha et al  
16  
17 2008, Clarke 2007). ~~It is recommended that review authors ensure that they limit the number~~  
18  
19 ~~of outcomes in the review and define them clearly. This will allow easier assessments of~~  
20  
21 ~~selective reporting, which can be done during data extraction of the included trials as long as~~  
22  
23 ~~a knowledgeable clinical person is involved.~~  
24  
25  
26  
27  
28  
29  
30  
31

### 32 **Unanswered questions and future research**

33  
34 Work is needed to consider what the best method is to assess the impact of ORB on the  
35  
36 results of the meta-analysis when there are multiple outcomes. Multivariate meta-analysis  
37  
38 has been suggested by Kirkham et al 2012 and a model based correction has been suggested  
39  
40 by Copas et al (2013<sup>2</sup>).  
41  
42  
43  
44

### 45 **Conclusion**

46  
47 Systematic reviews need to clearly state the primary and secondary outcomes that they will  
48  
49 consider and be consistent between review protocol and full review.  
50  
51

52  
53 Outcome reporting bias is a major problem for systematic reviews and more guidance needs  
54  
55 to be included in the Cochrane handbook to allow assessment of this important item within  
56  
57 the risk of bias tool. We recommend that an outcome matrix be completed during the  
58  
59  
60

1  
2  
3 production of a review to allow an ORB assessment for all review outcomes which can then  
4  
5 inform the risk of bias assessment.  
6

7 A core set of outcomes should be agreed upon for cystic fibrosis which in turn will have a  
8  
9 positive impact on systematic reviews as future trials are conducted they should specifically  
10 set out to measure and report these outcomes therefore reducing the prevalence of selective  
11 reporting.  
12  
13  
14  
15  
16  
17  
18

## 19 Abbreviations

20  
21 CFGD Cystic Fibrosis and Genetic Disorders

22  
23 CRG Cochrane Review Group

24  
25 COMET Core Outcome Measures in Effectiveness Trials

26  
27 CONSORT Consolidated Standards of Reporting Trails

28  
29 FEF<sub>25-75</sub> Average expired flow over the middle half of the FVC manoeuvre

30  
31 FEV1 Forced expiratory volume in 1 second

32  
33 FVC Forced vital capacity

34  
35  
36 NROD No Relevant Outcome Data

37  
38 ORB Outcome Reporting Bias

39  
40 ORBIT Outcome Reporting Bias In Trials

41  
42 PEFR Peak Expiratory Flow Rate

43  
44  
45 RCT Randomised Controlled Trial  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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.](#)

1  
2  
3  
4  
5 Copas J, Dwan K, Kirkham JJ, Williamson PR. A model based correction for outcome  
6 reporting bias in meta-analysis. (submitted [2013](#)).  
7  
8  
9

10  
11 Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, et al. Systematic Review of the  
12 Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. PLoS ONE  
13 2008, 3(8): e3081.  
14  
15  
16  
17  
18

19  
20 Dwan K, Gamble C, Kolamunnage-Dona R, Mohammed S, Powell C, Williamson PR.  
21 Assessing the potential for outcome reporting bias in a review: A tutorial. *Trials* 2010, **11**:52  
22 doi:10.1186/1745-6215-11-52.  
23  
24  
25  
26  
27

28  
29 Green, S. B. & Byar, D. P. Using observational data from registries to compare treatments:  
30 the fallacy of omnimetrics. *Stat Med*, 1984; 3, 361-73.  
31  
32  
33

34  
35 Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions*  
36 Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from  
37 www.cochrane-handbook.org.  
38  
39  
40  
41  
42

43  
44 Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of  
45 controlled clinical trials. *BMJ* 2001; 323: 42-46.  
46  
47  
48  
49

50  
51 Kirkham JJ, Altman DG, Williamson PR. Bias Due to Changes in Specified Outcomes during  
52 the Systematic Review Process. PLoS ONE 2010a; 5(3):  
53 e9810.doi:10.1371/journal.pone.0009810  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5 Kirkham JJ, Dwan K, Dodd S, Altman DG, Smyth R, Jacoby A, Gamble C, Williamson PR.

6  
7 The impact of outcome reporting bias in randomised controlled trials on a cohort of  
8  
9 systematic reviews. *BMJ* 2010b; 340: c365  
10

11  
12  
13  
14 Kirkham JJ, Riley RD and Williamson PR. A multivariate meta-analysis approach for  
15  
16 reducing the impact of outcome reporting bias in systematic reviews. *Statistics in Medicine*  
17  
18 2012; 31:20, 2179-2195.  
19

20  
21  
22  
23 Ramsey BW, Boat TF (1994) Outcome measures for clinical trials in cystic fibrosis.  
24  
25 Summary of a Cystic Fibrosis Foundation consensus conference. *J Pediatr* 1994; 124: 177–  
26  
27 192.  
28

29  
30  
31  
32 Sinha I, Jones L, Smyth RL, Williamson PR. A Systematic Review of Studies That Aim to  
33  
34 Determine Which Outcomes to Measure in Clinical Trials in Children . *PLoS Med* 2008;  
35  
36 5(4): e96. doi:10.1371/journal.pmed.0050096  
37

38  
39  
40 Smyth AR, Walters S. Prophylactic anti-staphylococcal antibiotics for cystic fibrosis.  
41  
42 *Cochrane Database of Systematic Reviews* 2003, Issue 3. Art. No.: CD001912. DOI:  
43  
44 10.1002/14651858.CD001912.  
45

46  
47  
48  
49 *The Cochrane Library*, Issue 4, 2009. Chichester: Wiley.  
50

51  
52  
53  
54 Williamson, P. R., Gamble, C., Altman, D. G. & Hutton, J. L. Outcome selection bias in  
55  
56 meta-analysis. *Stat Methods Med Res*, 2005; 14, 515-24.  
57  
58  
59  
60

1  
2  
3 [Williamson PR, Gamble C: Identification and impact of outcome selection bias in meta-](#)  
4 [analysis. Statistics in Medicine 2005, 24\(10\):1547-61.](#)  
5  
6  
7

8  
9 [Williamson PR, Gamble C: Application and investigation of a bound for outcome reporting](#)  
10 [bias. Trials 2007, 8:9.](#)  
11  
12  
13

14  
15  
16 von Mosch S, Dwan K. Reporting of harm and efficacy outcomes of randomised clinical  
17 trials in cystic fibrosis before and after the publication of the revised CONSORT  
18 (CONsolidated Standards Of Reporting Trials) statement in 2001 – a comparison of trial  
19 reports published from 1994 and from 2008. The University of Liverpool (thesis) 2011.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



Figure 1: Study flow diagram

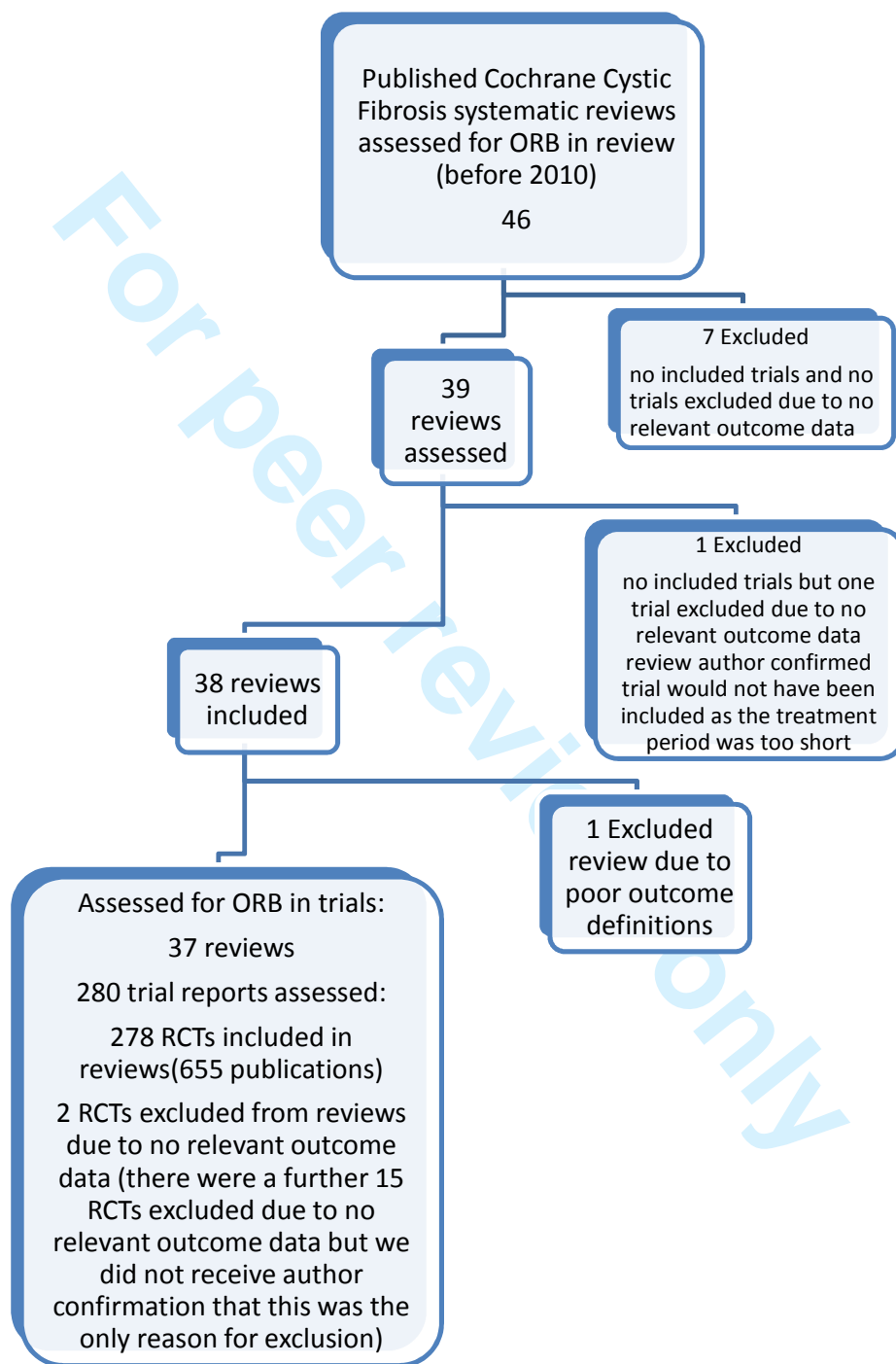


Table 1: ORBIT classifications

Classification	Description	Level of reporting	Level of suspicion of ORB	Primary outcome classifications	Secondary outcome classifications
				Number of trials (percentage overall) <sup>1</sup>	Number of trials (percentage overall) <sup>2</sup>
<i>Clear that the outcome was measured and analysed</i>					
<b>A</b>	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>B</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%)
<b>C</b>	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%)
<b>D</b>	States outcome analysed but no results reported.	None	High risk	0 (0)	0 (0)
<i>Clear that the outcome was measured</i>					
<b>E</b>	Clear that outcome was measured but not necessarily analysed.	None	High risk	59 (5.2%)	26 (3.9%)
<b>F</b>	Clear that outcome was measured but not necessarily analysed.	None	Low risk	110 (9.7%)	15 (2.2%)
<i>Unclear that the outcome was measured</i>					
<b>G</b>	Not mentioned but clinical judgment says likely to have been measured and analysed.	None	High risk	195 (17.1%)	197 (29.4%)
<b>H</b>	Not mentioned but clinical judgment says unlikely to have been measured.	None	Low risk	141 (12.4%)	256 (38.2%)
<i>Clear that the outcome was NOT measured</i>					
<b>I</b>	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	○ (A) <sup>1</sup>	○ (A) <sup>1</sup>	✓	✓	✓	✗ (H) <sup>2</sup>	✓
Schlesinger 1984	✗ (H) <sup>3</sup>	✗ (H) <sup>3</sup>	✓	○ (C) <sup>4</sup>	✓	✗ (H) <sup>2</sup>	✓
Stutman 2002	✓	✓	✓	✓	✓	✗ (H) <sup>2</sup>	✗
Weaver 1994	✗ (H) <sup>3</sup>	✗ (H) <sup>3</sup>	✓	○ (C) <sup>4</sup>	✓	✗ (H) <sup>2</sup>	✗

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.

- |   |
|---|
| <ul style="list-style-type: none"> <li>✓ indicates full reporting of results for treatment comparison of interest</li> <li>✗ indicates no reporting</li> <li>○ indicates partial reporting</li> </ul> |
|---|

**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) <sup>1</sup>	14 (38%)	
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	4 (44%)	
Reviews which have <i>upgraded</i> 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) <sup>1</sup>	3 (8%) (3 outcomes)	
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	0	
Reviews which have <i>downgraded</i> 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) <sup>1</sup>	9 (24%) (16 outcomes; min 1, max 5)	
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	1 (11%) (2 outcomes)	
Reviews which have <i>added</i> 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) <sup>1</sup>	2 (5%) (3 outcomes)	2 (5%) (4 outcomes; min 1, max 3)
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	1(11%) (1 outcome)	2 (22%) (2 outcomes)
Reviews which have <i>excluded</i> 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) <sup>1</sup>	2 (5%) (10 outcomes; min 1, max 9)	3 (8%) (5 outcomes; min 1; max 2)
	Protocol did not distinguish outcomes (n=9) <sup>2</sup>	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.

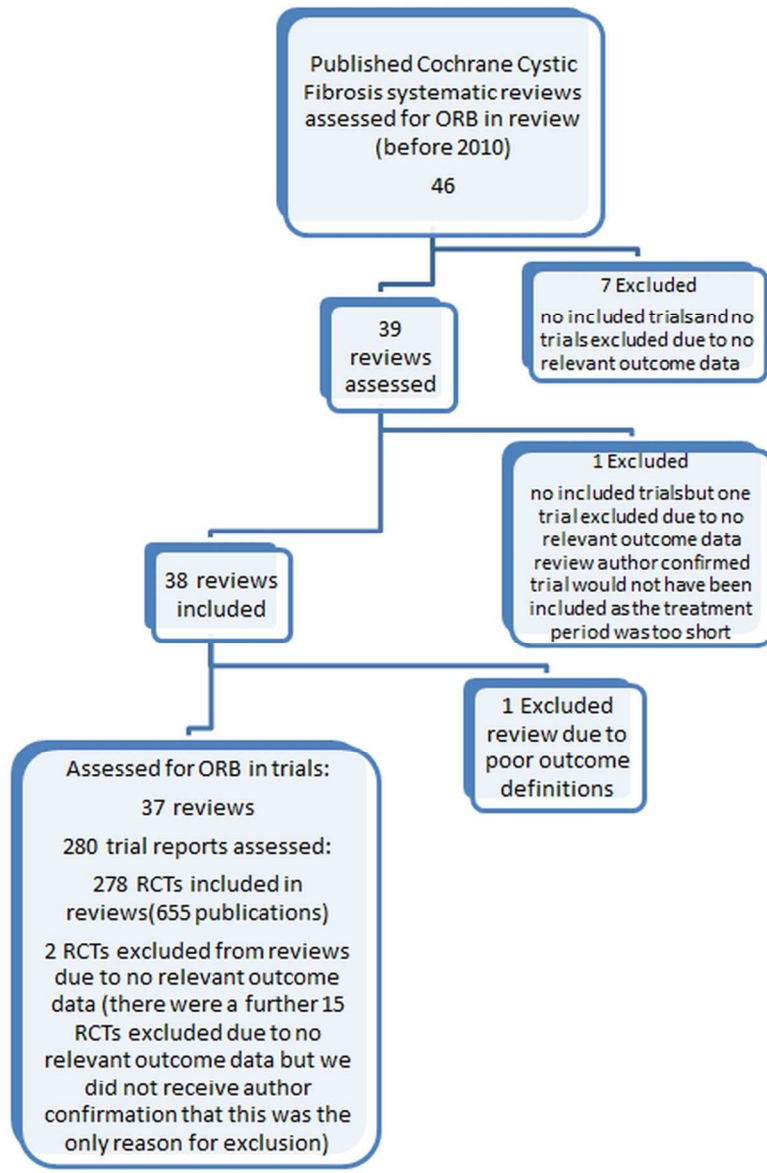
**Table 5: Risk of bias of RCTs based on review primary and secondary outcomes**

		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.

<b>SELECTIVE OUTCOME REPORTING</b>	
<b>Are reports of the study free of suggestion of selective outcome reporting? [Short form: <i>Free of selective reporting?</i>]</b>	
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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



90x118mm (300 x 300 DPI)