Revised Cochrane risk of bias tool for randomized trials (RoB 2.0) Additional considerations for cluster-randomized trials

Sandra Eldridge, Marion Campbell, Michael Campbell, Amy Dahota, Bruno Giraudeau, Julian Higgins, Barney Reeves and Nandi Siegfried

21st October 2016

Note: This document is a **supplement** to the main guidance document about the RoB 2.0 tool.

1.1 Bias in cluster-randomized trials

In cluster-randomized trials, groups of individuals rather than individuals are randomized to different interventions. The groups may be, for example, schools, villages, medical practices or families. In some trials, individuals are allocated to interventions that are then applied to multiple parts of those individuals (for example, to both eyes or to several teeth), or repeated observations are made on a participant. If the analysis is by the individual units (for example, each tooth or each observation) then individual units are clustered within individuals in the same way that individuals are clustered within, for example, medical practices. However, the terminology used to describe the clusters (individuals) and cluster members (units) will often be different and trying to deal with both types of scenarios using one set of signalling questions is complex. In this tool we therefore focus on trials in which groups of individuals are randomized. This includes stepped-wedge trials in which randomization is by cluster, although there is a source of bias in these trials if they are analysed without adjustment for secular trends which is not directly covered in the tool (1).

One of the main consequences of a cluster design is that participants within any one cluster often tend to respond in a similar manner, and thus their data can no longer be assumed to be independent of one another. It is important that the analysis of a cluster-randomized trial takes into account this issue. Unfortunately, many studies have in the past been incorrectly analysed as though the unit of allocation had been the individual participants. This is often referred to as a "unit-of-analysis error" because the unit of analysis is different from the unit of allocation. If the clustering is ignored and cluster trials are analysed as if individuals had been randomized, resulting P values will be artificially small. This can result in false positive conclusions that the intervention had an effect. In the context of a meta-analysis, studies in which clustering has been ignored will have overly narrow confidence intervals and will receive more weight than is appropriate in the meta-analysis. Often review authors can apply adjustments for clustering to overcome this problem, at least approximately. Note, however, that although there are examples of analyses that result in biased results, unit of analysis errors are associated primarily with problems of precision rather than bias. Therefore the appropriateness of analyses in taking account of clustering is not addressed by the RoB 2.0 tool.

A key difference between cluster-randomized trials and individually-randomized trials is that the individuals of interest (those within the clusters) may not be directly allocated to one intervention or another. In particular, sometimes the individuals are recruited into the study (or otherwise selected for inclusion in the analysis) after the clusters have been allocated to different interventions, leaving the potential for knowledge of the cluster's intervention allocation to influence whether individuals are recuited or selected into the analysis. The selection bias that arises when knowledge of the cluster allocation leads to different types of individuals being included in clusters in different intervention arms is often called **recruitment bias** or **identification bias**. We have added an additional domain for cluster-randomized trials to address this bias.

A second key difference between cluster-randomized trials and individually-randomized trials is that identifying who the "participants" are is not always straightforward in cluster-randomized trials. There are two reasons for this. First, unlike in individually-randomized trials, there may be no formal recruitment of

participants. When this is the case, for the purposes of this Risk of Bias tool, participants are defined as those individuals on whom it has been decided to collect data for the outcome of interest. In the IRIS trial (2), for example, clusters were general practices and the intervention aimed to increase rates of identification and referral for victims of domestic abuse. The researchers sought to collect data from routine records for all women in a practice aged 16 and over; these women are then the participants in relation to these outcomes. Second, for some trials there may be two or more different groups of participants on whom different outcomes are measured. For example, patients in the clusters may have a range of clinical outcomes measured, while health professionals in the clusters may have their knowledge or competency measured; who the participants are then depends on the outcome being measured. A specific example of this is a trial evaluating an open access urological investigation service. Researchers measured a number of outcomes in this trial, including GPs' compliance with referral guidelines and waiting time for patients referral to initial out-patient appointment (3), thus both GPs and patients were participants. In any assessment of bias outcomes on different sets of individuals should be considered different outcomes. A third issue around identifying participants arises in some trials in which data are collected at a number of time points from different individuals. For example, in WELL London (4), a trial of community engagement activity to increase physical activity, healthy eating and mental health and wellbeing, data were collected via baseline and follow-up population surveys, with baseline data controlled for in the analysis. In trials with this method of data collection, participants are all those whose data we wish to include in the analysis of the outcome of interest. Thus for the purposes of the RoB 2.0 tool we generally **define participants as those on whom investigators seek to measure the outcome of interest**, and when data are collected from different individuals at different time points, all those whose data we wish to include in the analysis of the outcome of interest. This interpretation of "participants" serves the purposes of the RoB 2.0 tool. However, it is worth noting that it is not the same as the interpretation of "participants" in the Ottawa statement on the ethics of cluster-randomized trials (5) which provides the most thoughtful and robust exposition to date of who the participants are in a cluster-randomized trial, from the standpoint of ethical considerations. The interpretation of participants in this statement is wider.

1.1.1 Bias arising from the randomization process

See also the section about bias arising from the randomization process in the main guidance document.

Bias arising from the randomization process operates in the same way as for individually-randomized trials but at the level of the cluster. An adequate allocation sequence needs to be devised as described in the main guidance document. Minimization is used more often in cluster-randomized trials than in individually-randomized trials, largely because it achieves a better balance in cluster characteristics between intervention groups when the number of clusters in small. In terms of risk of bias, minimization is regarded as equivalent to randomization when it includes a random element.

The randomization process in cluster-randomized trials can involve randomizing clusters sequentially, randomizing clusters in batches or randomizing clusters all at once. Allocation concealment may operate differently in trials with these different processes. Here we give two examples of adequate allocation concealment pertaining to the first two processes. When all clusters are randomized at once, concealment of the allocation sequence is not usually an issue.

In the IRIS trial (2), general practices were allocated using minimization with a random element. A researcher emailed details of a practice, included minimization factors, to an individual who used a computerized minimization programme to allocate the practice, and then sent details of the practice allocation to the researcher who communicated this with the practice. Practices were randomized one at a time. It would have been almost impossible for there to be any subversion (deliberate tampering with the allocation so that clusters end up in a group they were not supposed to be randomized to) of the allocation by either the researcher or the individual undertaking the randomization.

In the Diabetes Manual trial (6), clusters were allocated by minimization with a random element. For logistic reasons, allocation was performed in batches. The individual carrying out the minimization was provided with several cluster characteristics, which formed the basis of the minimization factors. These characteristics were predominantly continuous variables based on aggregating measures over all participants in each cluster. The uniqueness of these characteristics ensured that it was impossible to subvert allocations in spite of the fact that the minimization was done in batches.

Experience suggests that bias arising from the randomization process may be rarer in cluster-randomized trials than it is in individually-randomized trials. Reasons for this are that it is usually less easy to understand how clusters will react to an intervention (7), and also that in many cluster-randomized trials the main opportunity for subversion is by methodologists, who are usually less likely to have any motives or knowledge that may predispose them to do so.

On the other hand, if subversion is suspected, judging whether baseline imbalances suggest a problem with randomization processes is more difficult in these trials than in individually-randomized trials. In individually-randomized trials, imbalance in characteristics or numbers of participants may alert investigators to the likelihood of problems with randomization. For example, in the 1948 trial comparing anticoagulation medication to conventional treatment for myocardial infarction (8) described in the randomization methods section of the main guidance document for RoB 2.0, there were 589 participants in the intervention arm and 442 in the control arm, raising suspicion that investigators manipulated the allocation. For cluster-randomized trials these judgements should ideally be made in relation to the numbers or characteristics of clusters (the randomization units), particularly stratification or matching factors if these have been used. Stratification or matching factors, and other relevant cluster characteristics, are often characteristics such as numbers or make up of staff, or geographical location. They may also include characteristics of the whole cluster population, for example the ethnic make up of a general practice list or previous referral rates in a large hospital, and less often characteristics of actual participants (as in the Diabetes Manual trial described earlier). It is, however, usually not possible to make the sort of judgements about imbalance required to suggest problems with randomization because the number of clusters is too small. This leads to a greater possibility than in individually-randomized trials of chance baseline imbalance between the randomized groups, in terms of numbers or characteristics of clusters or individual participants. There is, additionally, another possible bias that might cause imbalance: identification/recruitment bias covered in the next domain and described below in Section 1.1.2. it is important that chance imbalances and imbalances likely to be because of identification/recruitment bias are not highlighted in this domain. The only imbalance that should be highlighted here is that judged to be due to problems with randomization.

The risk of chance baseline differences can be reduced by using stratified or pair-matched randomization of clusters, although pair-matching can bring its own issues. Reporting of the baseline comparability of clusters, or statistical adjustment for baseline characteristics, can help reduce concern about the effects of baseline imbalance. Furthermore, it is worth noting that extreme imbalances do not necessarily preclude inclusion in a meta-analysis so long as these imbalances arise by chance. For example, if only two clusters are present in a trial, the intervention effect is completely confounded with all differences between the clusters. However, if there are a large number of small cluster-randomized trials and confounding of clusters was occurring at random, then a meta-analysis would not necessarily be biased.

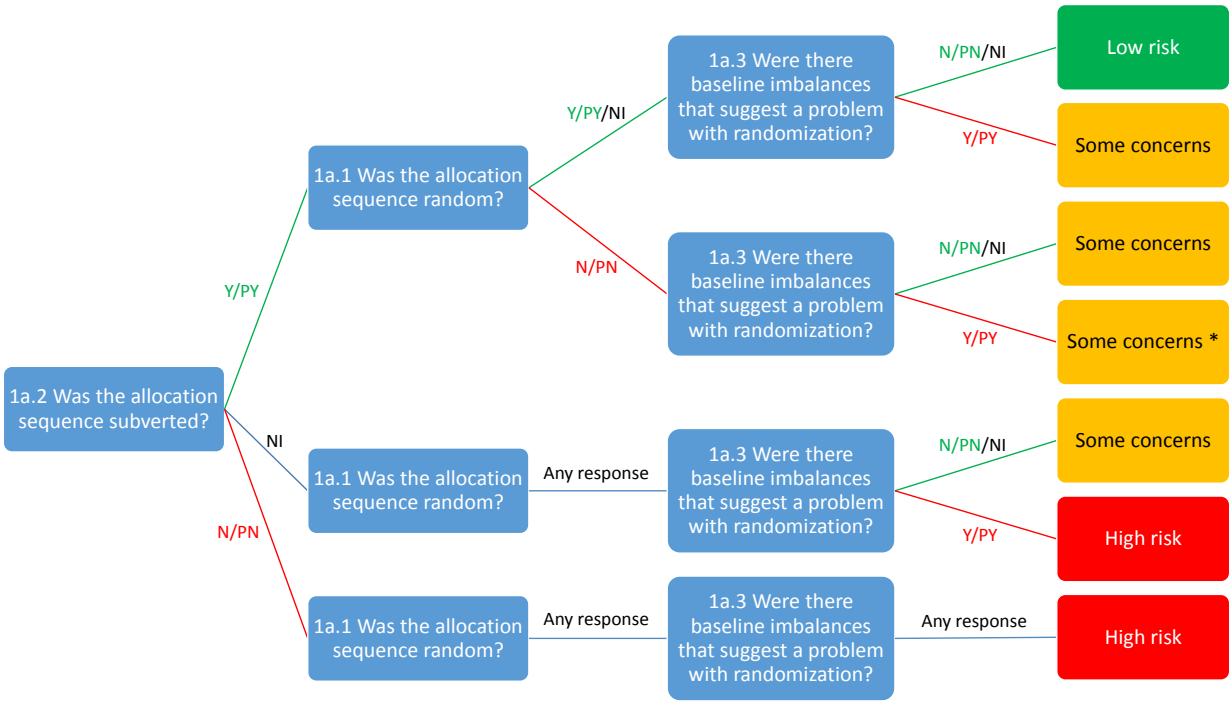
Signalling questions for this domain are provided in Box 1. An algorithm for reaching risk of bias judgements is provided in Figure 1.

Box 1. Risk of bias arising from the randomization process in a cluster-randomized trial

Signalling questions	Elaboration
1a.1 Was the allocation sequence random?	"Yes" if a random component was used in the sequence generation process such as using a computer generated random numbers, referring to a random number table, minimization, coin tossing; shuffling cards or envelopes; throwing dice; or drawing of lots. Minimization may be implemented without a random element, and this is considered to be equivalent to being random. "No" if the sequence is non-random, such that it is either likely to introduce confounding, or is predictable or difficult to conceal, e.g. alternation, methods based on dates (of birth or admission) or patient record numbers, allocation decision made by clinicians or participants, based on the availability of the intervention, or any other systematic or haphazard method. If the only information about randomization methods is to state that the study is randomized, then this signalling question should generally be answered as "No information". There may be situations in which a judgement is made to answer "Probably No" or "Probably yes". For example, if the study was large, conducted by an independent trials unit or carried out for regulatory purposes, then it may be reasonable to assume that the sequence was random. Alternatively, if other (contemporary) trials by the same investigator team have clearly used non-random sequences, it might be reasonable to assume that the current study was done using similar methods. Similarly, if participants and personnel are all unaware of intervention assignments throughout/during the trial (blinding or masking), this may be an indicator that the allocation process was also concealed, but this will not necessarily always be the case. If the allocation sequence was clearly concealed but there is no information about how the sequence was generated, it will often be reasonable to assume that the sequence was random (although this will not necessarily always be the case).
1a.2 Is it likely that the allocation sequence was subverted?	Processes of randomizing clusters vary. It is important first to consider carefully whether there are any ways in which the allocation could potentially have been subverted (deliberately tampered with so that clusters end up in a group they were not supposed to be randomized to if the randomization was conducted properly). This will usually include a consideration of whether any individuals were aware of any potential allocations prior to those allocations being made. However, although subversion may be possible, it is often the case that in cluster randomized trials those who could subvert the randomization have less motivation and/or knowledge to do so (see text for further explanation), so a judgement must be made as to whether this is likely.

1a.3 Were there baseline imbalances that suggest a problem with the randomization process?	Imbalances in numbers of clusters or stratification factors or other cluster characteristics are usually the best evidence of problems with the randomization process, but such problems are relatively unusual as explained in 1a.2. On the other hand, due to the small numbers of clusters randomized in most cluster randomized trials, chance imbalances in either cluster or participant characteristics are more common than in individually-randomized trials and can sometimes appear substantial. As for the tool for individually-randomized trials, chance imbalances should not be highlighted here, and neither should imbalances that are due to identification/recruitment bias (which are assessed in Domain 1b). Answer "No" if no imbalances are apparent or if any observed imbalances are compatible with chance Answer "Yes" only if there is clear evidence of imbalances that appear to be due to problems with randomization. In some circumstances, it may be reasonable to answer "Yes/Probably yes" (rather than "No information") when there is a surprising lack of information on baseline characteristics when such information could reasonably be expected to be available/reported. If there is no information about cluster characteristics record "No information". The answer to this question should not be used to influence answers to questions 1a.1 or 1a.2. For example, if the trial has large baseline imbalances, but authors report adequate randomization methods, questions 1a.1 and 1a.2 should still be answered on the basis of the reported adequate methods, and any concerns about the imbalance should be raised in the answer to the question 1a.3 and reflected in the domain-level risk of bias judgement).

Figure 1. Suggested algorithm for reaching risk of bias judgements for bias arising from the randomization process in a cluster-randomized trial



1.1.2 Bias arising from the timing of identification and recruitment of participants in relation to timing of randomization

Bias can occur when participants (those on whom we seek to collect outcome data) are identified and/or recruited to the trial after the clusters have been randomized, because the knowledge of whether each cluster is an "intervention" or "control"' cluster could affect the types of participants recruited or identified. For example, Farrin et al. showed differential participant recruitment (which took place after randomization) in a trial of low back pain patients randomized by primary care practice; a greater number of less severe participants were recruited to the 'active management' practices (9). To distinguish this bias arising from the timing of the identification and/or recruitment of individual participants from other biases, we refer to it as identification/recruitment bias. Table 1 shows the different potential orderings of randomization of clusters, individual participant identification and individual participant recruitment in cluster-randomized trials, including scenarios in which individual participants are not recruited. In three scenarios identification/recruitment bias is possible although not inevitable; there are methods of protecting against this bias through trial design (10). Here we present some examples of where bias did exist corresponding to these three scenarios.

Scenario 1: Individuals are identified and recruited after randomization as a result of a visit to a cluster or an acute event (e.g asthma exacerbation or, as in the example above, episode of back pain) by someone who knows the cluster allocation, and/or the potential participant knows the cluster allocation before consenting.

If the individuals recruiting participants know the cluster allocation, they can then consciously or subconsciously influence the numbers and type of individuals recruited in that cluster, or may share knowledge of the cluster allocation with those being recruited, influencing the likelihood of them agreeing to participate. In a similar example to the back pain example given above, in the Diabetes Care from Diagnosis trial (11) the participants were incident cases of Type II diabetes. The general practicitions (GPs) in both arms had to diagnose the patients before they could become trial participants. Intervention GPs were trained in new ways of treating people with Type II diabetes. One might therefore expect them to behave differently from control GPs. In the end, 142 patients were recruited in the intervention arm and only 108 in the control arm, despite similar numbers of clusters in the two arms.

Potential participants who know about cluster allocation before being recruited can also induce bias. For example, in a trial evaluating the treatment of malnutrition in Burkino Faso in which clusters were randomized into three groups, the local community were aware that clusters (health centres) in the two control arms provided food supplements while the intervention clusters provided counselling. In an area of food scarcity, potential participants may have chosen to attend the screening for entry into the trial in the catchment area of a health centre providing the intervention of their choice, most likely the food supplements.

Table 1. Possible orderings of randomization of clusters, individual participant identification and individual participant recruitment in cluster-randomized trials

Scenario 1	Scenario 2	Scenario 3	Scenario 4 (identical to 6)	Scenario 5	Scenario 6 (identical to 4)
Randomization	Randomization	Identification of *potential* individual participants	Identification of individual participants	Identification of *potential* individual participants	Identification of individual participants
Identification of *potential* individual participants	Identification of individual participants	Randomization	Randomization	Recruitment of individual participants	Participants not directly recruited
Recruitment of individual participants	Participants not directly recruited	Recruitment of individual participants	Participants not directly recruited	Randomization	Randomization
Potential for identification/recruitment bias although this could be avoided through trial design			**No potential for identification/recruitment bias because randomization happens after**		

Note: In scenarios 2, 4 and 6 individual participants are not recruited as indicated. This also means that when individual participants are identified they become the *actual* participants in the study rather than being identified as potential participants.

Scenario 2: Individuals are not recruited at all, perhaps because outcomes will be measured on routine data, but they are identified after randomization by someone whose knowledge of the cluster allocation can influence which individuals are chosen to have their outcomes measured.

> This happened, for example, in a trial to assess feeding strategies for critically ill patients in intensive care unit (ICU) wards (12). Staff in intervention wards developed guidelines prior to identifying participants; control staff did not. This could have differentially affected identification in the two arms, although there is no evidence of that from the publication.

Scenario 3: Potential individual participants are identified prior to randomization, for example from a clinic list, but actual participants are recruited after randomization, at which stage knowledge of the cluster allocation by those recruiting or by the potential participants themselves can influence the number and types of individuals recruited in that cluster.

> This was the case in a trial to evaluate hip protectors for preventing hip fractures. The clusters were units for the care of the elderly within community based health centres. Prior to randomization of the units, All ambulatory men and women who were 70 years old or older and who had at least one easily identifiable risk factor for hip fracture were identified in each community health centre. After randomization these individuals were approached to be recruited. After the study had been explained to them, 204 of the subjects in units assigned to the hip-protector group (31 percent) and 94 of the subjects in units assigned to the control group (9 percent) declined to participate, a not insubstantial difference that may have caused bias if those left in the intervention group had different characteristics from those in the control.

Although identification/recruitment bias is only possible under these three scenarios, and even then can be avoided in many cases with careful trial design, evidence suggests that it is not uncommon. Puffer et al. (13) reviewed 36 cluster-randomized trials, and found possible identification/recruitment bias in 14 (39%). Using slightly different methodology, Eldridge et al. (7), Froud et al. (14) and Diaz-Ordaz et al. (15) suggested slightly fewer but still not insignificant proportions open to such bias (21%, 22% and 7%, respectively), with the last of these studies finding a larger proportion of trials in which it was not possible to judge bias because of lack of reported information. Puffer et al judged the potential for identification/recruitment bias by looking for imbalance in baseline characteristics of individual participants (13). While an exploration of imbalance may be useful to identify major suspected issues caused by awareness of cluster allocation prior to randomization as described above, it should be used with caution: as described in the previous domain (Section 1.1.1), it may be

difficult to disentangle the different possible reasons for imbalance in individual participant characteristics: chance, problems with randomization, and identification/recruitment bias.

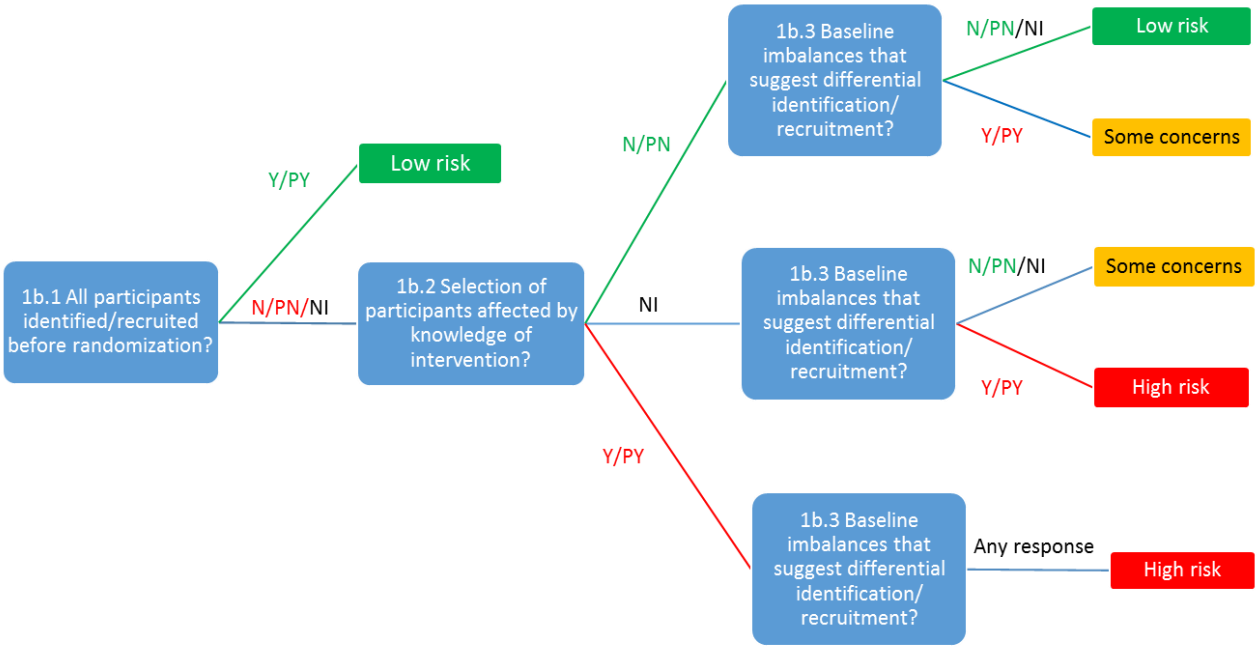
Some cluster-randomized trials end up with no participants in a cluster. This can happen only if individuals are identified/recruited after randomization, and poses a particularly awkward type of recruitment bias. If clusters are completely empty of individual participants, any sort of imputation becomes difficult. It is important that authors report the existence of any of these empty clusters.

Signalling questions for this domain are provided in Box 2. An algorithm for reaching risk of bias judgements is provided in Figure 2.

Box 2. Risk of bias arising from the timing of identification and recruitment of participants in a cluster-randomized trial

Signalling questions	Elaboration
1b.1 Were all the individual participants identified before randomization of clusters (and if the trial specifically recruited patients were they all recruited before randomization of clusters)?	Answer "Yes" if participants were identified and recruited prior to the clusters being randomized or if individual participants were not recruited at all but were identified prior to randomization. In these cases identification/recruitment bias is not possible. Answer "No" if either identification or recruitment of participants (or both) takes place after randomization. Also answer "No" if some participants are identified and/or recruited before and some after randomization as the potential for bias still exists in these trials.
1b.2 If N/PN/NI to 1b.1: Is it likely that selection of individual participants was affected by knowledge of the intervention?	Answer "Yes" if those recruiting individuals are aware of cluster allocation prior to recruitment and are likely to consciously or subconsciously have differentially recruited in the trial arms; if some of those being recruited are aware of cluster allocation prior to their own recruitment and this is likely to have differentially affected recruitment in the trial arms; if those identifying potential participants (when recruitment is to take place subsequently) or those identifying actual participants (when there is no subsequent recruitment) are aware of cluster allocation and are likely to have consciously or subconsciously differentially include potential individual participants in different trial arms. Answer "No" if all of the following (as relevant depending on the trial) are unaware of cluster allocation at recruitment: (1) those identifying actual participants, (2) those identifying potential participants, (3) those recruiting and (4) potential participants themselves.
1b.3 Were there baseline imbalances that suggest differential identification or recruitment of individual participants between arms?	As for signalling question 1a.3, imbalances that are compatible with chance should not be highlighted here. Imbalances due to differential identification or recruitment of participants are more common in cluster randomized trials than imbalances due to problems with randomization. Such imbalances are usually in the numbers of participants recruited into each arm or, less commonly, in the characteristics of such individuals. If there is a noticeable imbalance and imbalance due to the randomization process and due to identification/recruitment of individuals are both possible a judgement will need to be made about which is the most likely cause of any imbalance or whether they are both likely.

Figure 2. Suggested algorithm for reaching risk of bias judgements for bias arising from the timing of identification and recruitment of participants in a cluster-randomized trial



1.1.3 Bias due to deviations from intended intervention

Interventions in cluster-randomized trials are commonly multifaceted and very often do not involve drugs. We can usefully think of "the intervention" in such a trial as comprising a number of different interventions (sometimes also referred to as components) which may be aimed at individual participants (e.g. a leaflet, a blood test, a self-management course), or at health professionals (e.g. feedback on some aspects of care, educational sessions, a computerized tool) or at clusters (e.g. new flooring in hospital wards, posters or videos in a waiting room), or may be the addition of staff (e.g. liaison nurse, health advocate). Most cluster-randomized trials include interventions aimed at professionals and/or whole clusters; in one of the largest reviews of cluster-randomized trials to date, 90% of 157 trials included one or both of these types of interventions (unpublished data, Eldridge PhD thesis). This multifaceted complex nature of interventions in cluster-randomized trials has implications for interpreting bias due to deviations from the intended intervention, Before considering this source of bias it is important to be clear exactly what the intended intervention is.

1.1.3.1 The role of the target comparison

As well as involving multifaceted interventions, the majority of cluster-randomized trials are also pragmatic, designed to answer the question of whether an intervention works in real life conditions. Indeed, it is the pragmatic nature of the research question being asked that often leads to a clustered design. One example is the OPERA trial (16). Investigators wanted to see if a whole-home intervention to increase physical activity in nursing homes would reduce depression. Interventions were depression awareness training for nursing home staff, physiotherapist-led exercise classes, physiotherapist assessment and feedback on activity for individual residents, and physiotherapist-led interventions to increase activity in the homes in general.

The appropriate analysis for addressing pragmatic research questions is an intention-to-treat (ITT) analysis measuring the effect of assignment to the intervention arm at baseline. In an ITT analysis it is important that both clusters and participants are analysed according to the intervention group to which they were randomized, regardless of intervention received. For clusters, it is usually relatively straightforward to ascertain

whether this has been the case or not. However, for participants this can be more difficult, particularly if they are not recruited. The IRIS trial is an example: in this trial outcomes were measured using routine data on all women over 16 on general practices (the clusters) lists at the end of the trial period (2). The data were not interrogated to find out whether any women had moved from one trial practice to another or moved from a practice outside the trial during the trial period. It was therefore not possible to say whether the analysis was strictly an ITT analysis. Nevertheless, it was not expected that movement of women between practices would have had anything to do with the trial, thus bias was thought to be negligible.

In fact, the effect of assignment to intervention is usually the only effect of interest in cluster-randomized trials; consideration of per-protocol effects measuring the effect of starting and adhering to "the intervention", is uncommon. Indeed, because of the complex multifaceted nature of many of the interventions in these trials, it is often not obvious how to define a strict per-protocol population. For example, in the OPERA trial, all those who consented to data collection were considered the intention-to-treat population. However, it is difficult to think how to define a strict per-protocol population or per-protocol effect for the whole intervention when this had multiple components aimed at both residents and staff, and this effect was not of interest in this trial. When the intervention in the trial is simpler, per-protocol effects may be of interest and per-protocol populations may be *defined*, but actually *identifying* the per-protocol population and/or protocol deviations may be difficult. One example of this comes from a trial evaluating the installation of flooring in hospital wards to reduce the incidence of injurious falls. In this trial the clusters were wards. Suppose that the new flooring caused health professionals to relax (i.e. become lax in their patient management) leading to more opportunities for patients to fall. If the intended intervention is the new flooring in addition to usual practice then the relaxed attitude of the health professionals would constitute a deviation beyond that expected in usual practice. In the intervention group, the per protocol population could then be defined as those individual patients who were treated by health professionals as they would have been if the flooring were not present, but from a practical point of view this population would be hard to identify. Less commonly, interventions in cluster-randomized trials are placebo controlled drugs. This is usually when the reason for conducting a trial is the risk of contamination between individual patients in an individually-randomized trial. In these trials, per protocol effects are often relevant and possible. We provide two examples in Box 3. We have not included signalling questions for the per-protocol effect in this tool because of the rarity of these analyses.

Box 3. Examples of trials which used a cluster-randomized design because of the possibility of contamination

Example 1: The intervention was a drug administered to those with difficult to treat head lice

In this trial, clusters were households; a cluster-randomized trial was chosen because of the possibility of contamination within households. The active treatment was tablets and the control treatment was lotion. To ensure that individual participants, clusters in general, and those delivering the intervention remained unaware of the intervention, investigators used a double dummy design in which both intervention groups administered tablets and lotion but the intervention group administered placebo lotion and the control group administered placebo tablets. Thus, in relation to the effect of assignment to intervention, the trial was protected against bias due to deviations from intended intervention beyond usual practice. In terms of the effect of starting and adhering to intervention, however, questions about the success of the implementation, adherence and co-interventions are relevant. In this trial, both intention to treat and per protocol analyses were conducted (17).

Example 2: The intervention was vitamin D administered to residents and carers within sheltered accommodation to prevent acute respiratory infection in residents

The clusters in this trial were sheltered accommodation schemes; a cluster-randomized design was chosen because of the possibility of contamination within the sheltered schemes. Masking was achieved through a placebo. Intention to treat and per-protocol analyses were conducted.

1.1.3.2 The role of masking or blinding

In some cluster-randomized trials participants may not be aware that they are in a trial. This can happen when participants are not directly recruited, for example in the IRIS trial (2). It can also happen when participants

are told at recruitment that they are in a study but not that they are in a trial; this is sometimes done if participants are recruited after randomization in order to lessen the chance of identification/recruitment bias. If participants do not know that they are in a trial then they cannot deliberately switch to an alternative treatment or cause any deviations from intended interventions beyond what would occur in usual practice, even if they are aware of their intervention.

There may be more personnel, or more levels of personnel, involved in a cluster-randomized trial than an individually-randomized trial. When some components of a multifaceted intervention are aimed at health professionals and/or clusters it is usually not possible, or desirable, for those receiving these interventions to be unaware of the fact that they are receiving an intervention. In fact, in most cluster-randomized trials there will be at least one intervention for which at least some of those involved in the trial (individual participants, professionals within clusters, other trial personnel) will be aware of the intervention(s) being administered. In the OPERA trial, for example, nursing home staff, and trial personnel who visited nursing homes, were aware of the allocation of homes. Residents would have been aware of some of the interventions, for example the exercise classes (16).

1.1.3.3 Co-interventions

It is possible that whole clusters and/or specific individuals within clusters do not receive the intervention intended.

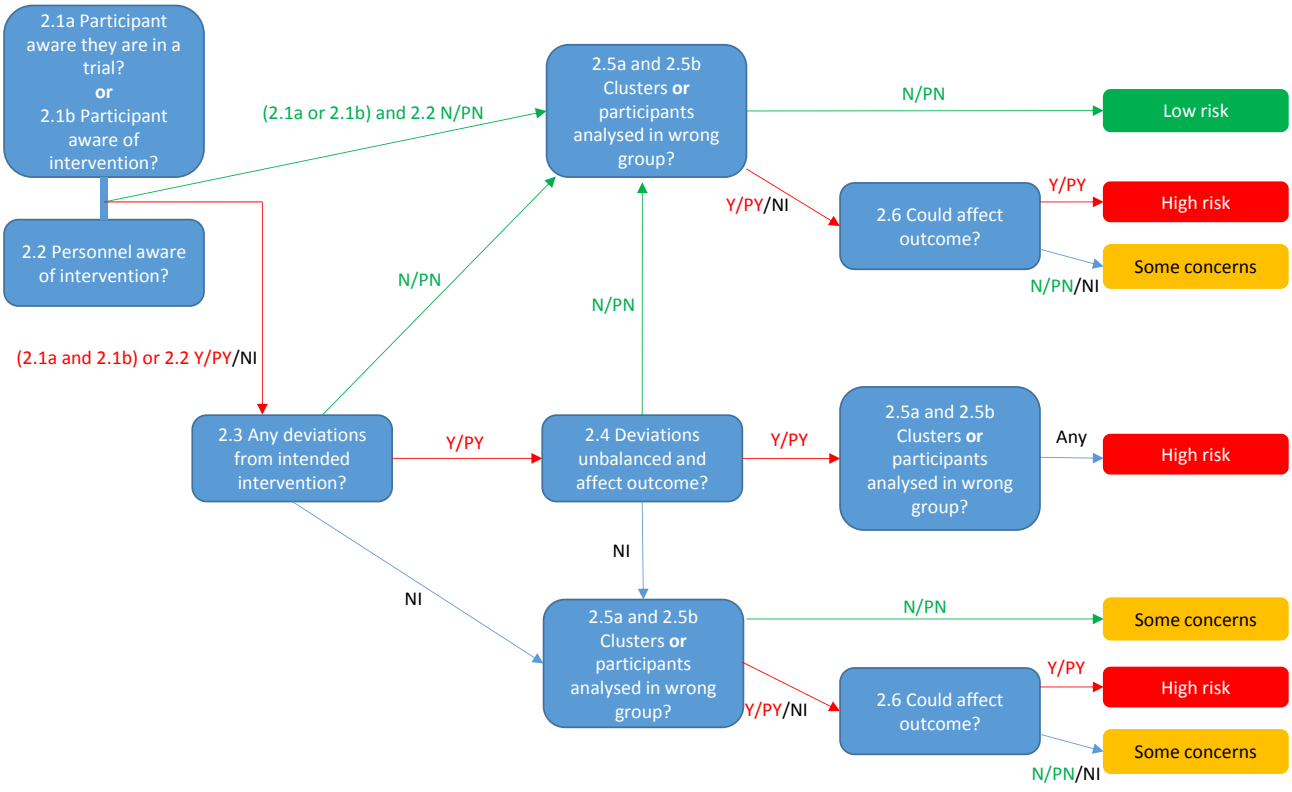
Signalling questions for this domain are provided in Box 4. An algorithm for reaching risk of bias judgements is provided in Figure 3.

Box 4. Risk of bias due to deviations from intended intervention in a cluster-randomized trial (*effect of assignment to intervention*)

Signalling questions	Elaboration
2.1a Were participants aware that they were in a trial?	In cluster randomized trials it is possible for participants to know they are receiving an intervention or that they are in a study but not that they are in a trial. Thus they may not know that other evaluations are being evaluated or what these interventions are. This makes it impossible for them to cause deviations from the intended interventions beyond what would be expected in usual practice.
2.1b If Y/PY/NI to 2.1a: Were participants aware of their assigned intervention during the trial?	Cluster randomized trials frequently involve multifaceted interventions. Answer "Yes" if participants were aware of any part of the allocated intervention during the trial.
2.2. Were carers and trial personnel aware of participants' assigned intervention during the trial?	If those involved in caring for participants or making decisions about their health care are aware of the assigned intervention, then implementation of the intended intervention, or administration of additional co-interventions, may differ between the assigned intervention groups. Masking carers and trial personnel, which is most commonly achieved through use of a placebo, may prevent such differences.
2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention beyond what would be expected in usual practice?	When interest focusses on the effect of assignment to intervention, it is important to distinguish between: (a) deviations that happen in usual practice following the intervention and so are part of the intended intervention (for example, cessation of an exercise programme for health related issues); and (b) deviations from intended intervention that arise due to expectations of a difference between intervention and comparator (for example because participants feel 'unlucky' to have been assigned to the comparator group and therefore seek the active intervention, or components of it, or other interventions). We use the term "usual practice" to refer to the usual course of events in a non-trial context. Because deviations that arise due to expectations of a difference between intervention and comparator are not part of usual practice, they may lead to biased effect estimates that do not reflect what would happen to participants assigned to the interventions in practice. Deviations from the intended intervention that arise due to expectations of a difference between intervention and comparator are rarely reported in cluster randomized trials and may, in fact, occur rarely. This is likely to be partly because it is very often the case in these trials that those who might have the opportunity to introduce deviations will not have any inclination to deliberately affect the results of the trial by doing so. In addition the more complex the intervention, the more difficult it might be to practically identify such deviations. The answer "No information" will therefore be appropriate in many cases, but "Probably yes" should be used if it seems likely that such deviations occurred.
2.4. If Y/PY to 2.3: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	As for individually-randomized trials.

2.5a Were any clusters analysed in a group different from the one to which they were assigned?	As for 2.5 for individually-randomized trials.
2.5b Were any participants analysed in a group different from the one to which their original cluster was randomized?	In some cluster randomized trials it may not be possible to ascertain the original cluster that individuals were in. This could happen, for example, when clusters split or merge or participants are not recruited and outcomes are collected from routine data. In this case a judgement will need to be made about whether the answer to this question is "PY" or "NI".
2.6 <u>If Y/PY/NI to 2.5:</u> Was there potential for a substantial impact (on the estimated effect of intervention) of analysing participants in the wrong group?	Risk of bias will be high in a randomized trial in which sufficiently many clusters or participants were analysed in the wrong intervention group that there could have been a substantial impact on the results. There is potential for a substantial impact if more than 5% of participants were analysed in the wrong group, but for rare events there could be an impact for a smaller proportion.

Figure 3. Suggested algorithm for reaching risk of bias judgements for bias due to deviations from intended interventions in a cluster-randomized trial (*effect of assignment to intervention*)



1.1.4 Bias due to missing outcome data

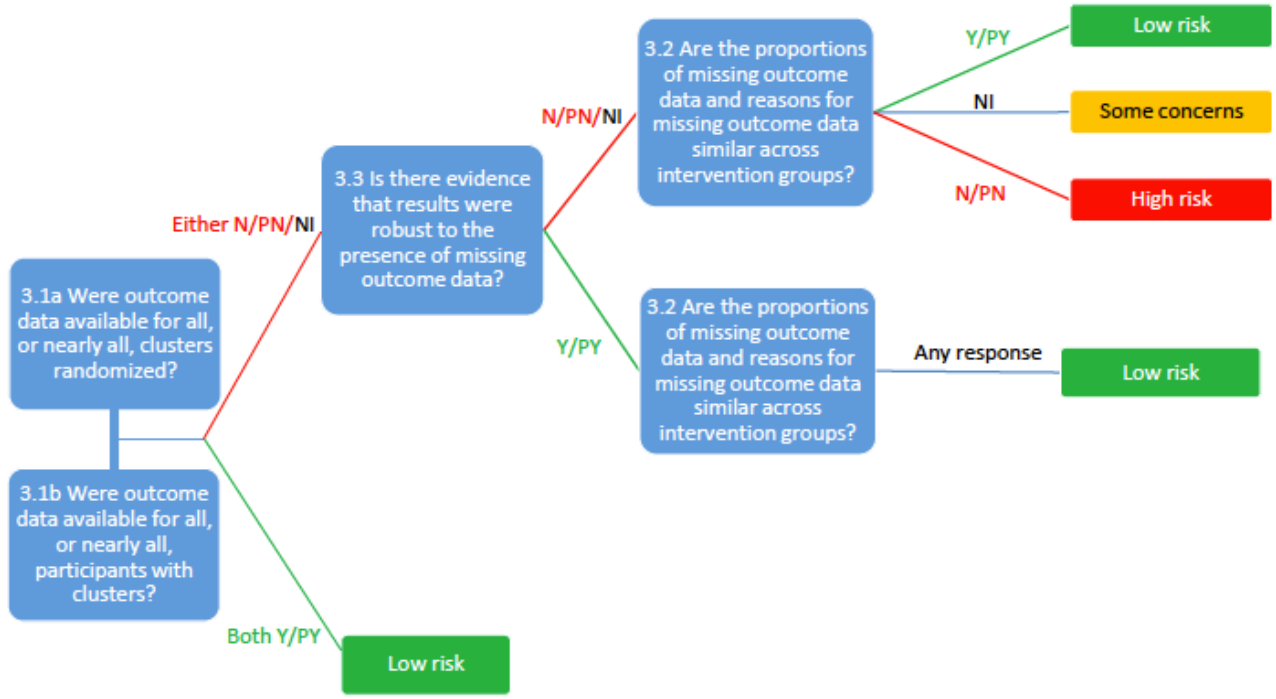
Missing outcome data should be considered at both the level of the cluster and the level of the individual. In most cases some participants will be missing outcome data, but occasionally complete clusters are lost from a trial. Broadly the same considerations apply here as for individually-randomized trials. Empirical research has shown that most cluster-randomized trials have missing data but that this poorly reported and inadequately handled in analyses (18). In considering the possibility of bias, attention should be paid to the amount of missing data, the reasons for missing data and the way the missingness has been dealt with in analyses.

Signalling questions for this domain are provided in Box 5. An algorithm for reaching risk of bias judgements is provided in Figure 4.

Box 5. Risk of bias due to missing data in a cluster-randomized trial

Signalling questions	Elaboration
3.1a Were outcome data available for all, or nearly all, clusters randomized?	As for individually-randomized trials.
3.1b Were outcome data available for all, or nearly all, participants within clusters?	The issues here are broadly as for question 3.1a. In cluster-randomized trials there may be particular complexities when clusters merge, split, or disappear.
3.2 If N/PN/NI to 3.1: Are the proportions of missing outcome data and reasons for missing outcome data similar across intervention groups?	As for individually-randomized trials.
3.3 If N/PN/NI to 3.1: Is there evidence that results were robust to the presence of missing outcome data?	As for individually-randomized trials.

Figure 4. Suggested algorithm for reaching risk of bias judgements for bias due to missing data in a cluster-randomized trial



1.1.5 Bias in measurement of the outcome

Issues in measurement of outcomes are broadly similar for cluster-randomized trials and individually-randomized trials. The key issues are identifying who is the outcome assessor and whether the assessment of the outcome is likely to be influenced by knowledge of intervention received. In individually-randomized

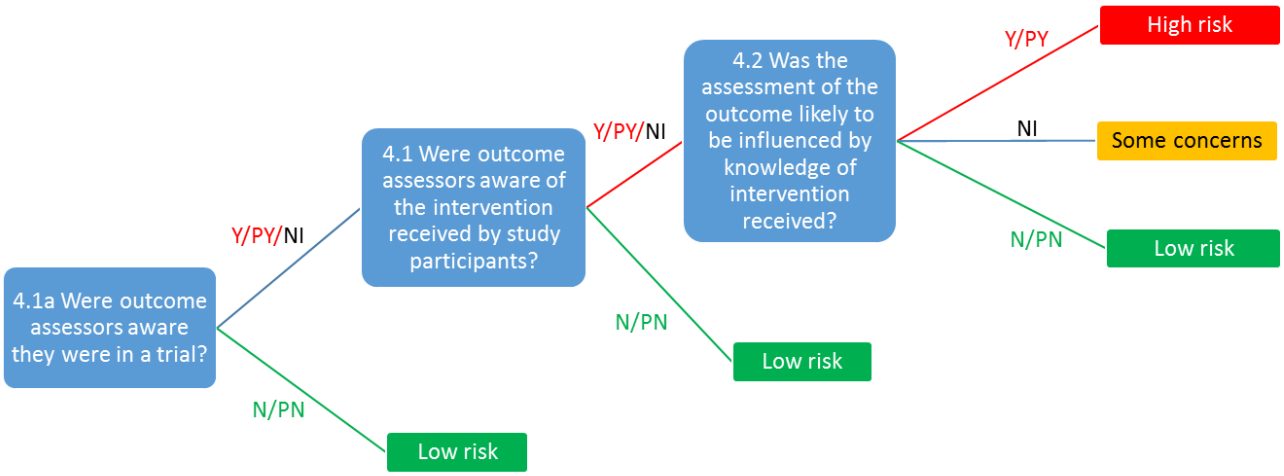
trials, more objective measures such death or cure are expected to be less subject to bias than measures such as participant-reported outcomes, because participants may be influenced by the knowledge of which intervention they received. This will also be true in many cluster-randomized trials. However, in cluster-randomized trials in which participants do not know they are part of a trial, participant reported outcomes may not be subject to bias in the same way, even if participants are aware of the intervention they receive because they will not be aware of other possible interventions that they could have received.

Signalling questions for this domain are provided in Box 6. An algorithm for reaching risk of bias judgements is provided in Figure 5.

Box 6. Risk of bias in measurement of the outcome in a cluster-randomized trial

Signalling questions	**Elaboration**
4.1a Were outcome assessors aware that a trial was taking place?	This question largely applies to studies in which participants report their outcomes themselves, for example in a questionnaire. The participant is then the outcome assessor. In individually randomized trials self-assessment may be influenced by assignment if participants are aware of their assignment. In cluster randomized trials, if participants are not aware that they are in a trial then their self-assessment cannot be affected by assignment regardless of whether they are aware of the intervention they receive or not.
4.1b If Y/PY/NI to 4.1: Were outcome assessors aware of the intervention received by study participants?	"No" if outcome assessors were blinded to intervention status. In studies where participants report their outcomes themselves (i.e., participant-reported outcome), the outcome assessor is the study participant. In cases where outcomes are collected using routine data, the outcome assessor is the individual responsible for extracting the data.
4.2 If Y/PY/NI to 4.1: Was the assessment of the outcome likely to be influenced by knowledge of intervention received?	As for individually-randomized trials.

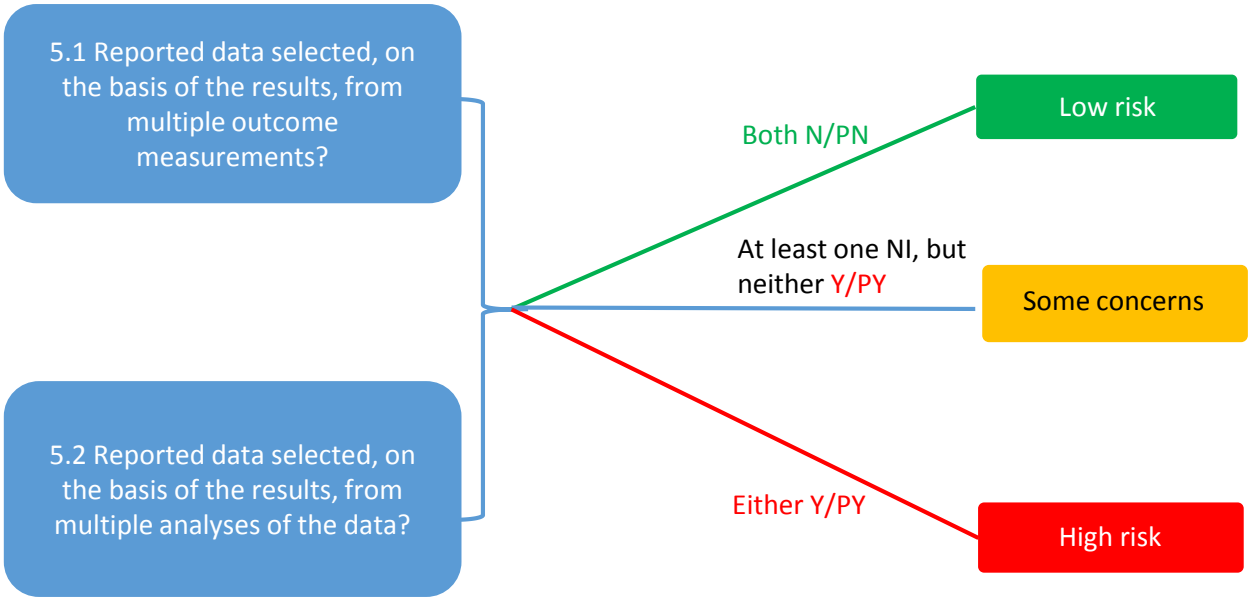
Figure 5. Suggested algorithm for reaching risk of bias judgements for bias in measurement of the outcome in a cluster-randomized trial (*effect of assignment to intervention*)



1.1.6 Bias in selection of the reported result

Issue of selective reporting are generally the same for cluster-randomized trials as for individually-randomized trials.The algorithm for reaching risk of bias judgements is provided in Figure 6.

Figure 6. Suggested algorithm for reaching risk of bias judgements for bias due to selection of the reported result in a cluster-randomized trial



1.2 References

1. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ. 2015;350:h391.
2. Feder G, Davies RA, Baird K, Dunne D, Eldridge S, Griffiths C, et al. Identification and Referral to Improve Safety (IRIS) of women experiencing domestic violence with a primary care training and support programme: a cluster randomised controlled trial. Lancet. 2011;378(9805):1788-95.
3. Thomas RE, Grimshaw JM, Mollison J, McClinton S, McIntosh E, Deans H, et al. Cluster randomized trial of a guideline-based open access urological investigation service. Fam Pract. 2003;20(6):646-54.
4. Phillips G, Bottomley C, Schmidt E, Tobi P, Lais S, Yu G, et al. Measures of exposure to the Well London Phase-1 intervention and their association with health well-being and social outcomes. J Epidemiol Community Health. 2014;68(7):597-605.
5. Taljaard M, Weijer C, Grimshaw JM, Eccles MP. The Ottawa Statement on the ethical design and conduct of cluster randomised trials: precis for researchers and research ethics committees. BMJ. 2013;346:f2838.
6. Sturt JA, Whitlock S, Fox C, Hearnshaw H, Farmer AJ, Wakelin M, et al. Effects of the Diabetes Manual 1:1 structured education in primary care. Diabet Med. 2008;25(6):722-31.
7. Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: systematic review of recent trials. BMJ. 2008;336(7649):876-80.
8. Wright IS, Marple CD, Beck DF. Report of the Committee for the Evaluation of Anticoagulants in the Treatment of Coronary Thrombosis with Myocardial Infarction; a progress report on the statistical analysis of the first 800 cases studied by this committee. Am Heart J. 1948;36(6):801-15.
9. Farrin A, Russell I, Torgerson D, Underwood M, Team UBT. Differential recruitment in a cluster randomized trial in primary care: the experience of the UK back pain, exercise, active management and manipulation (UK BEAM) feasibility study. Clin Trials. 2005;2(2):119-24.
10. Eldridge S, Kerry S, Torgerson DJ. Bias in identifying and recruiting participants in cluster randomised trials: what can be done? BMJ. 2009;339:b4006.
11. Kinmonth AL, Woodcock A, Griffin S, Spiegal N, Campbell MJ. Randomised controlled trial of patient centred care of diabetes in general practice: impact on current wellbeing and future disease risk. The Diabetes Care From Diagnosis Research Team. BMJ. 1998;317(7167):1202-8.
12. Doig GS, Simpson F, Finfer S, Delaney A, Davies AR, Mitchell I, et al. Effect of evidence-based feeding guidelines on mortality of critically ill adults: a cluster randomized controlled trial. JAMA. 2008;300(23):2731-41.
13. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. BMJ. 2003;327(7418):785-9.
14. Froud R, Eldridge S, Diaz Ordaz K, Marinho VC, Donner A. Quality of cluster randomized controlled trials in oral health: a systematic review of reports published between 2005 and 2009. Community Dent Oral Epidemiol. 2012;40 Suppl 1:3-14.
15. Diaz-Ordaz K, Froud R, Sheehan B, Eldridge S. A systematic review of cluster randomised trials in residential facilities for older people suggests how to improve quality. BMC Med Res Methodol. 2013;13:127.
16. Underwood M, Lamb SE, Eldridge S, Sheehan B, Slowther A, Spencer A, et al. Exercise for depression in care home residents: a randomised controlled trial with cost-effectiveness analysis (OPERA). Health Technol Assess. 2013;17(18):1-281.
17. Chosidow O, Giraudeau B, Cottrell J, Izri A, Hofmann R, Mann SG, et al. Oral ivermectin versus malathion lotion for difficult-to-treat head lice. N Engl J Med. 2010;362(10):896-905.
18. Diaz-Ordaz K, Kenward MG, Cohen A, Coleman CL, Eldridge S. Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines. Clin Trials. 2014;11(5):590-600.