Revised Cochrane risk of bias tool for randomized trials (RoB 2) Additional considerations for cluster-randomized trials (RoB 2 CRT)

Sandra Eldridge, Marion K Campbell, Michael J Campbell, Amy K Drahota, Bruno Giraudeau, Barnaby C Reeves, Nandi Siegfried, Julian PT Higgins

18 March 2021

Note: This document is a **supplement** to the main guidance document about the RoB 2 tool, which is a revised and improved version of the original Cochrane tool for assessing risk of bias in randomized trials (1). The format and structure of RoB 2 are aligned with the ROBINS-I tool for assessing risk of bias in non-randomized studies of interventions (2), using signalling questions within bias domains. The main RoB 2 tool is intended for trials in which participants are individually randomized to parallel groups. This document describes an adaptation for cluster-randomized trials, and should be read in conjunction with the guidance document for the main RoB 2 tool.

1 Cluster-randomized trials in the context of the Risk of Bias tool version 2.

In cluster-randomized trials, groups (or clusters) of individuals, rather than individuals themselves, are randomized to different interventions.

A key difference between cluster-randomized trials and individually-randomized trials is that in cluster-randomized trials the individuals of interest (those within the clusters) are not directly allocated to one intervention or another. This design feature raises the possibility of a specific type of bias because the individuals are sometimes recruited into the study (or otherwise selected for inclusion in the analysis) after the clusters have been allocated to different interventions. Recruitment or identification of individuals within clusters whose intervention assignment is known creates the potential for knowledge of the assigned intervention to influence whether individuals are recruited, or selected into the analysis. The bias that arises when this occurs is referred to in various ways: we prefer the term **identification/recruitment bias,** which distinguishes this bias from other types of bias and highlights its source**.** We have added an additional domain for cluster-randomized trials to the RoB 2 tool to address this bias.

Identifying the 'participants' is not always straightforward in cluster-randomized trials, because:

(i) in some trials there may be no formal recruitment of participants;
(ii) in some trials there may be two or more different groups of participants on whom different outcomes are measured; and
(iii) in some trials data are collected at two or more time points on different individuals (the design is then often referred to as a "repeated cross-sectional" design) (3).

For the purposes of the RoB 2 tool we **define participants in cluster-randomized trials as those on whom investigators seek to measure the outcome of interest.** In Box 1 we present examples of how this definition works for each of the three situations outlined above. The definition of 'participants' used here is from a data collection and analysis perspective, and not the same as in the Ottawa statement on the ethics of cluster-randomized trials (4). The definition in the Ottawa statement is wider (for example, including those who may be exposed to the intervention but from whom no data are collected). This provides the most thoughtful and robust exposition to date of who are the participants in a cluster-randomized trial from an ethical perspective.

Box 1. Examples of trials corresponding to the three reasons why identifying participtants in these trials is not straightforward

1. In the IRIS trial, clusters were general practices and the intervention aimed to increase rates of identification and referral for victims of domestic abuse (5). The trial focussed on women aged 16 and over, but women were not directly recruited into the trial. The researchers instead sought to collect data from routine practice records for all women aged 16 and over; these women are the participants in relation to the outcomes of identification and referral rates.
2. In a trial evaluating an open access urological investigation service, researchers measured several outcomes, including general practitioners' (GPs') compliance with referral guidelines and waiting time for patients' referral to initial outpatient appointment (6). Thus, both GPs and patients were participants.
3. In the WELL London 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 (7). The individuals in these surveys were not the same, and the baseline data were controlled for in the analysis of the follow-up population. To perform this analysis the researchers sought data on all those approached in both surveys. In trials using this method of data collection, all of these individuals would be counted as participants in relation to risk of bias. .

In recent years, a new type of design of cluster-randomized trial has become increasingly popular: **the stepped-wedge design**. The guidance presented here broadly applies to stepped-wedge cluster randomized trials. An additional source of bias in the analysis of stepped-wedge trials – analysis without adjustment for secular trends (8) can be addressed through signalling question 2.6, as discussed in Section 4.

Some aspects of the analysis of a cluster-randomized trial are important but not directly related to bias, so are not covered by the RoB 2 tool. Participants within a cluster often tend to respond in a similar manner, and thus their outcomes cannot be assumed to be statistically independent. It is important that the analysis of a cluster-randomized trial accounts for this issue: unfortunately many past studies were incorrectly analysed as though the unit of allocation were the individual participants. This is often referred to as a 'unit-of-analysis error'. If the clustering is ignored, confidence intervals will be too narrow and P values too small. In the context of a meta-analysis, studies in which clustering was ignored will receive more weight than is appropriate, and the confidence interval for the overall meta-analytic summary is likely to be too narrow. Often review authors can apply adjustments for clustering to overcome this problem, at least approximately. Unit-of-analysis errors are associated primarily with problems of precision rather than bias in the estimated effect of intervention.

Trials in which individuals are randomized and outcomes are measured on **different parts of the body** (e.g. on both eyes or multiple teeth) or **different time points** (e.g. multiple pregnancies) have properties very similar to those of cluster-randomized trials. Such designs are therefore also covered in principle by the tool below, although for simplicity of explanation, the wording of the signalling questions and the accompanying discussion do not explicitly refer to this situation. Where instructions in this document related to signalling questions state "As for individually randomised trials", reviewers should refer to the main guidance for the RoB-2 tool at www.riskofbias.info.

2 Bias arising from the randomization process

The main RoB 2 tool for individually-randomized trials includes signalling questions about generation of the allocation sequence, concealment of allocation and whether substantial imbalances between groups suggest a problem with the randomization process. These issues are also important in cluster-randomized trials. 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 RoB 2 guidance document. Minimization and biased coin (or biased urn) randomization are used more often in cluster-randomized trials than in individually-randomized trials, largely because substantial imbalances in cluster characteristics between intervention groups can occur by chance when the number of clusters is small. While its use is less frequent, biased coin randomisation may also be used for similar reasons (9). As discussed in the main guidance document for RoB 2, minimization is considered an adequate form of generating a randomization sequence, and the same applies in general to biased coin randomization.

The randomization process in cluster-randomized trials can involve randomizing clusters sequentially, in batches, or all at once. Allocation concealment may operate differently in trials with these different processes. When all clusters are randomized at once, concealment of the allocation sequence is not usually problematic.

In the IRIS trial, general practices were allocated using minimization(5). A researcher emailed practice details, including 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 of the allocation (deliberate tampering with the allocation so that clusters end up in a group to which they were not supposed to be randomized) by either the researcher or the individual undertaking the randomization.

In the Diabetes Manual trial, clusters were allocated by minimization. For logistical reasons, allocation was performed in batches (10). The individual carrying out the computerized 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 for each cluster ensured that it was impossible to subvert allocations in spite of the fact that the minimization was done in batches.

Bias arising from the randomization process may be rarer in cluster-randomized trials than it is in individually-randomized trials. This is because it is usually hard to predict how clusters will respond to an intervention (11). Also, in many cluster-randomized trials the main opportunity for subversion is by methodologists (who typically implement the randomization), who are unlikely to have motives or knowledge that predispose them to subvert the process.

In individually-randomized trials, imbalances in characteristics or numbers of participants may alert investigators to problems with randomization (see main RoB 2 guidance for examples). For cluster-randomized trials these judgements should be made in relation to the characteristics or numbers of clusters (the randomization units), particularly with reference to stratification or matching factors if these have been used. Relevant cluster characteristics may be of the clusters themselves (such as numbers or make up of staff, or geographical location), of the whole cluster population (such as the ethnic make-up of a general practice list or previous referral rates in a large hospital), or of trial participants (as in the Diabetes Manual trial described earlier).

Judging that baseline imbalances suggest a problem with randomization is more difficult in cluster- than individually-randomized trials because the number of clusters is typically small, so that substantial baseline imbalances between the randomized groups, in terms of numbers or characteristics of clusters or individual participants, can occur by chance. **Only imbalances judged to suggest problems with randomization should be highlighted in this domain.** Issues of identification/recruitment bias, which can also cause baseline imbalances, are covered in the domain "Bias arising from the identification or recruitment of participants into clusters" in Section 3.

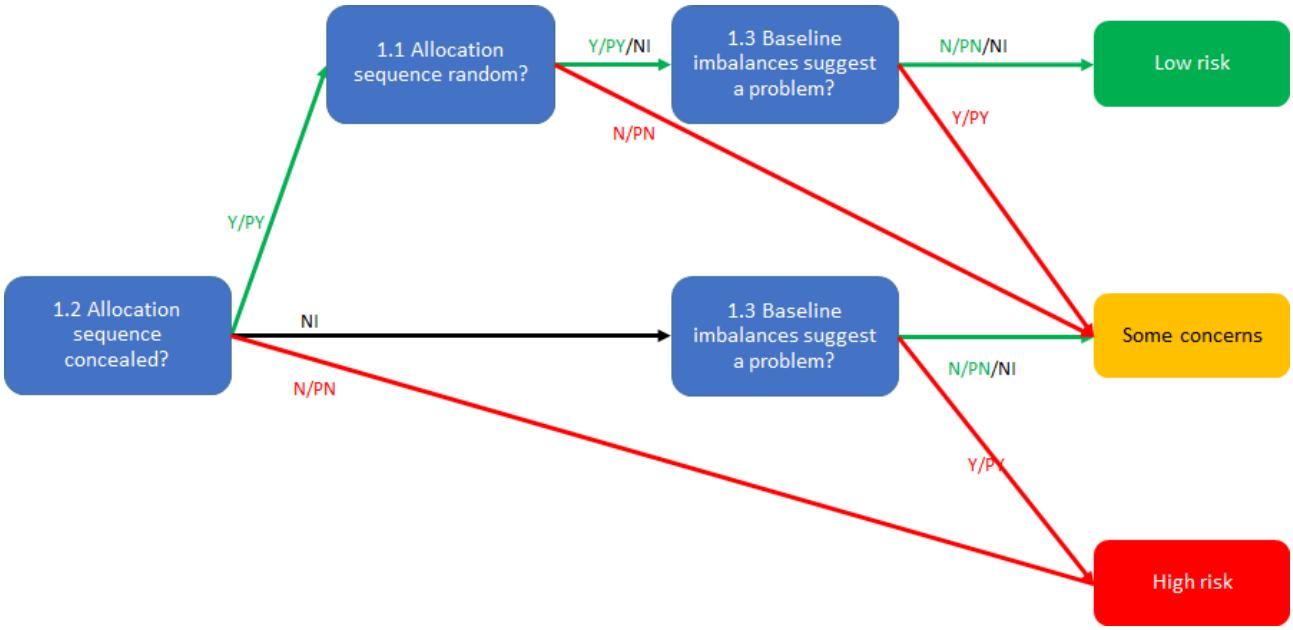
Chance baseline imbalances do not preclude inclusion in a meta-analysis, even if they are substantial. Providing that the analysis accounts for the clustered design chance imbalances will tend to be balanced across a large number of small cluster-randomized trials and the confidence interval around the effect from an individual trial will provide appropriate uncertainty around a single realisation of the randomisation. Nevertheless, the influence of chance baseline imbalances in a single trial can be reduced by using stratified or pair-matched randomization of clusters. However, pair-matching has disadvantages and must be specifically addressed in the analysis (12).

Signalling questions for this domain are provided in Box **2.** An algorithm for reaching risk of bias judgements is provided in Figure 1.

Box 2. Risk of bias arising from the randomization process in a cluster-randomized trial

Signalling questions	Elaboration
1a.1 Was the allocation sequence random?	Considerations are mostly the same as for individually randomized trials. Answer 'No' for non-random methods that might be seen in cluster-randomized trials, including those based on geography (e.g. clusters near the main research centre allocated to the intervention and those further away to the control).
1a.2 Was the allocation sequence concealed until clusters were enrolled and assigned to interventions?	As for individually randomized trials.
1a.3 Did baseline differences between intervention groups suggest a problem with the randomization process?	*Note that differences that are compatible with chance do not lead to a risk of bias.* Answer 'No' if any observed imbalances are compatible with chance or likely to be because of identification/recruitment bias, which are addressed in domain 1b (see section 3). Imbalances in numbers of clusters or in stratification/ matching/ minimization factors can provide evidence of problems with the randomization process, but such problems are likely to be unusual in cluster-randomized trials. Due to the small numbers of clusters randomized in most cluster-randomized trials, chance imbalances in either cluster or participant characteristics, which can be substantial, are more common than in individually-randomized trials. Answer 'Yes' if there are imbalances that indicate problems with the randomization process, including: (1) substantial differences between numbers of clusters between intervention arms, compared with the intended allocation ratio; or (2) a substantial excess in statistically significant differences in baseline cluster characteristics between intervention groups, beyond that expected by chance; or (3) imbalance in one or more baseline measures of outcome variables, that is very unlikely to be due to chance and for which the between-group difference is big enough to result in bias in the intervention effect estimate. Also answer 'Yes' if there are other reasons to suspect that the randomization process was problematic: (4) excessive similarity in baseline characteristics that is not compatible with chance. Answer 'No information' when there is no *useful* baseline information available (e.g. abstracts, or studies that reported only baseline characteristics of participants in the final analysis). 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 and when such information could reasonably be expected to be available/reported. 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 that are judged to be unlikely to be due to chance or identification/recruitment bias, 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. Algorithm for suggested judgement of risk of bias judgements for bias arising from the randomization process in a cluster-randomized trial



3 Bias arising from the identification or recruitment of participants into clusters

As highlighted in the introduction, a key difference between cluster-randomized trials and individually-randomized trials is that participants (those on whom researchers seek to collect outcome data) may be recruited to the trial or otherwise identified for inclusion in the analysis after the clusters have been randomized. This can lead to bias if knowledge of the intervention assigned to a cluster affects recruitment or identification of participants. For example, Farrin et al. identified differential participant recruitment after randomization in a trial of low back pain patients randomized by primary care practice: participants with less severe pain were more likely to be recruited to the practices in the 'active management' arm of the trial (13). We refer to this as **identification/recruitment bias**.

Table 1 shows the different potential orderings of (i) randomization of clusters, (ii) identification of individual participant and (iii) recruitment of individual participants in cluster-randomized trials. It includes scenarios in which individual participants are not directly recruited. In three scenarios, identification/recruitment bias is possible However, such bias is not inevitable and can be avoided by ensuring that recruiters are unaware of the intervention assigned to the cluster (14). The same principles apply in a stepped-wedge design.

It is possible for more than one of the scenarios in Table 1 to occur in a single trial. For example, the OPERA trial evaluated the effectiveness of a whole-home intervention to increase physical activity in nursing homes to reduce depression (15). Most participants were recruited before randomization (scenario 5), but residents who joined homes after randomization were also invited to participate (scenario 1). In Box 3 we present further examples of trials corresponding to the three scenarios in which there is potential for bias.

Table 1. Possible orderings of the steps of randomizing clusters, identifying individual eligible participants and recruiting individual eligible participants in cluster-randomized trials

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

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

Although identification/recruitment bias is only possible under scenarios 1-3, and can still be avoided with careful trial design, evidence suggests that it is often present. Puffer et al. reviewed 36 cluster-randomized trials, and found possible identification/recruitment bias in 14 (39%) (16). Using slightly different methodology, Eldridge et al. (11), Froud et al. (17) and Diaz-Ordaz et al. (17) suggested fewer but still significant proportions open to such bias (21%, 22% and 7%, respectively). The last of these studies identified 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 (16). While an exploration of imbalance may be useful to identify major issues caused by awareness of cluster allocation before identification and/or recruitment, imbalances should be interpreted with caution as described in Section 2. There have been no reviews to date looking at the possibility of identification/recruitment bias in stepped wedge trials. However, when a repeated cross-sectional stepped wedge design is used, the potential for bias may be greater than in parallel cluster-randomized trials.

Some cluster-randomized trials end up with cluster(s) in which no participants are enrolled. This can happen only if individuals are identified/recruited after randomization. Thus, the presence of such 'empty' clusters (in which no participants were recruited) may indicate issues with the ordering of the randomization and recruitment processes and thus of potential bias.

Signalling questions for this domain are provided in Box 4. An algorithm for reaching risk of bias judgements is provided in Figure 2.

Box 3 Three scenarios in which there is potential for identification/recruitment bias

Scenario 1: Participants are identified and recruited after randomization (as a result of a visit to a cluster or an acute event) by someone who knows the cluster allocation, and/or the potential participant knows the cluster allocation before consenting.

In the Diabetes Care from Diagnosis trial the participants had incident Type II diabetes (18). The general practicitioners (GPs) in both groups diagnosed 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 recruit differently from control GPs. 142 patients were recruited in the intervention group and only 108 in the control group, despite similar numbers of clusters of similar sizes in the two groups. Although different numbers in intervention and control groups do not indicate bias per se, this substantial difference in numbers recruited may be indicative of differences in characteristics which could result in bias.

Potential participants who know about cluster allocation before being recruited can also introduce bias. In a trial evaluating the treatment of malnutrition in Burkina Faso, health centres were randomized into three groups. The local community was aware that health centres in the two control groups provided food supplements while the intervention health centres provided counselling (19). In an area of food scarcity, potential participants who were poorer and/or who had means of transport may have chosen to attend screening for entry into the trial in a health centre providing the food supplements, and bias could have ensued.

Scenario 2: Participants are not directly recruited, because to do so would compromise the science or logistics of the trial, but they are identified after randomization by someone whose knowledge of the cluster allocation can influence which individuals have their outcomes measured.

In a trial to assess feeding strategies for critically ill patients in intensive care unit (ICU) wards (20) staff in intervention wards developed relevant guidelines before identifying participants; control staff did not. This could have differentially affected identification in the two groups, although there is no evidence from the publication that this happened.

Scenario 3: Potential participants are identified before randomization 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.

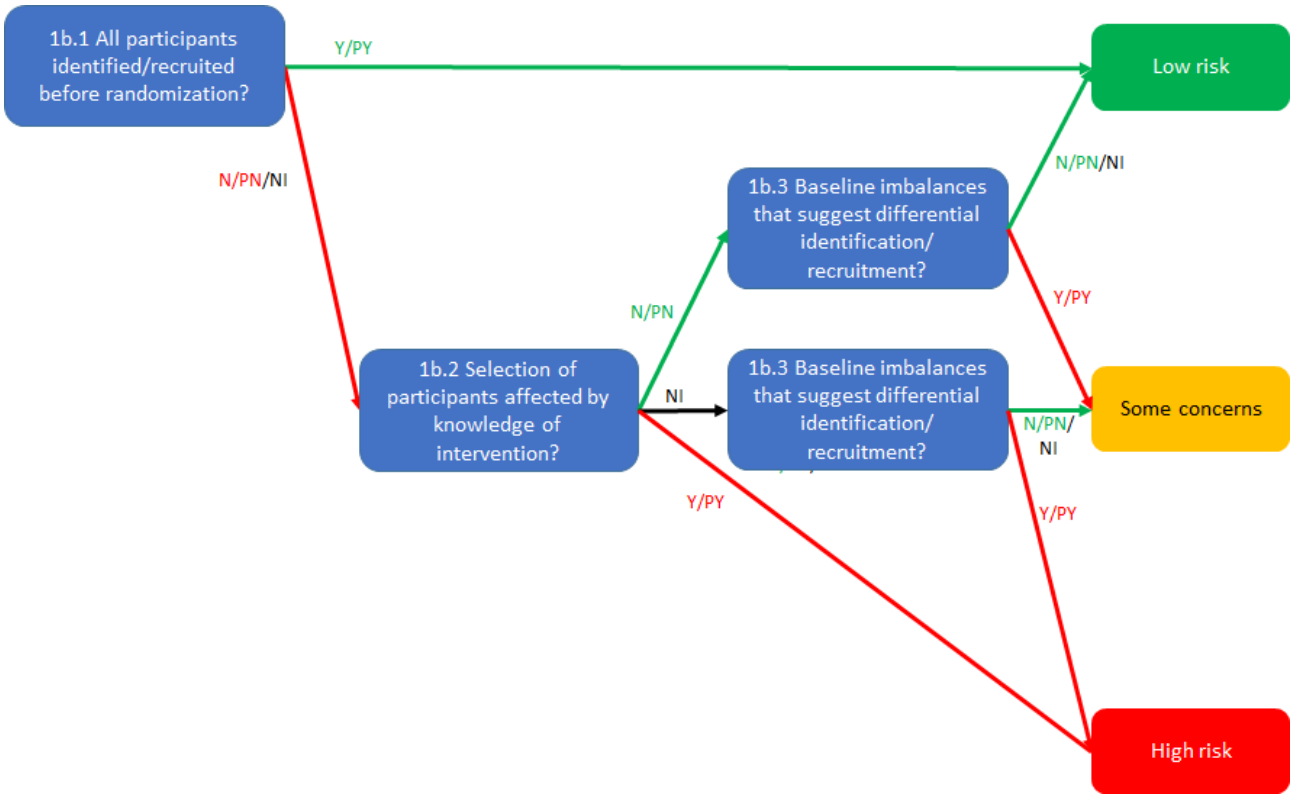
In a trial to evaluate hip protectors for preventing hip fractures the clusters were units for the care of the elderly within community-based health centres. Before randomization, all ambulatory men and women aged 70 or over and who had at least one easily identifiable risk factor for hip fracture were identified in each community health centre. After randomization these individuals were approached for consent to be recruited. 31 percent of the subjects in units assigned to the hip-protector group and 9 percent of the subjects in units assigned to the control group declined to participate, a substantial difference that may have led to bias.

Box 4. Risk of bias arising from the timing of identification or recruitment of participants in a cluster-randomized trial

Signalling questions	Elaboration
1b.1 Were all the individual participants identified and recruited (if appropriate) before randomization of clusters?	Answer 'Yes' if: (1) all participants were identified and recruited before the clusters were randomized; or (2) individual participants were not recruited at all but all were identified before randomization. In these cases identification/recruitment bias is not possible. Answer 'No' if: (1) some or all participants were identified or recruited after randomization; or

	(2) there are any clusters in which no participants were recruited (empty clusters).
1b.2 <u>If N/PN/NI to 1b.1</u>: Is it likely that selection of individual participants was affected by knowledge of the intervention assigned to the cluster?	Answer 'Yes' if: (1) those recruiting individuals were aware of cluster allocation before recruitment and this is likely, consciously or subconsciously, to have affected recruitment differentially between the intervention groups; (2) some participants were aware of cluster allocation before their recruitment and this is likely to have affected recruitment differentially between the intervention groups; or (3) those identifying potential participants (when recruitment is to take place subsequently) are aware of cluster allocation and are likely, consciously or subconsciously, to have differentially included potential individual participants in different trial groups or those identifying actual participants (when there is no subsequent recruitment) are aware of cluster allocation and are likely consciously or subconsciously, to have differentially included potential individual participants in different trial groups. 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.
1b.3 Were there baseline imbalances that suggest differential identification or recruitment of individual participants between intervention groups?	As for signalling question 1a.3, imbalances that are compatible with chance should not be interpreted as suggesting differential identification or recruitment of participants. Such imbalances are more common in cluster-randomized trials than imbalances due to problems with randomization. They can be in the numbers of participants recruited into each group or in the characteristics of such individuals.

Figure 2. Algorithm for suggested judgement of risk of bias judgements for bias arising from the timing of identification and recruitment of participants in a cluster-randomized trial



4 Bias due to deviations from intended intervention

As described in the guidance for the main RoB 2 tool, this domain relates to biases that arise when there are deviations from the intended interventions, which could be the administration of additional interventions that are inconsistent with the trial protocol, failure to implement the protocol interventions as intended, or non-adherence by trial participants to their assigned intervention. Deviations from intended interventions only lead to bias if they affect the outcome.

The interventions that were intended should ideally be fully specified in the trial protocol. However, this information is often not available in published reports. It may be necessary for users of the RoB 2 tool to document whether they consider any changes to interventions to be consistent with the intended intervention. Understanding the intended interventions can be particularly challenging for cluster-randomized trials, because interventions are often aimed at multiple levels. For example, interventions may:

- be aimed at individual participants (e.g. a leaflet, a blood test, a self-management course);
- be aimed at health professionals (e.g. feedback on some aspects of care, educational sessions, a computerized tool) or clusters (e.g. new flooring in hospital wards, posters or videos in a waiting room). In one of the largest reviews of cluster-randomized trials to date, 90% of 157 trials included interventions aimed at either professionals or clusters or both (21).;
- be the explicit provision of additional staff who are not normally based in the clusters who may see patients and support staff (e.g. liaison nurse or health advocate in general practices); or
- comprise a combination of the above.

In addition, protocols often do not detail how interventions aimed at individual participants should change or evolve in response to their health, or events they experience. These limitations make it difficult to determine whether deviations from the intended interventions have occurred, and users of the tool may have to make judgements about what intended courses of action are likely to have been.

4.1 The role of the effect of interest

4.1.1 Assessing the effect of assignment to intervention

As for individually randomized trials, we expect that the effect of interest will most often be that of **assignment to intervention**. Indeed, the effect of assignment to intervention is often the only effect of interest in cluster-randomized trials. The question of interest is whether the intervention *as implemented in the trial* is effective, regardless of what was intended in the intervention protocol. When assessing the effect of assignment to intervention, the analysis should follow the principles of intention-to-treat. In Box 5 we discuss identification of who should be included in an intention-to-treat analysis for cluster-randomized trials with different approaches to selecting participants.

The risk of bias assessment for the effect of assignment to intervention should focus on whether deviations from the intended interventions arose because of the trial context: that is, because of either recruitment and engagement activities that may influence trial participants' adherence to their assigned intervention, or because of unconscious or conscious undermining of trial comparisons by trial personnel. Contamination is one consequence of recruitment and engagement activities and would not be expected to happen outside of the trial context. Cluster-randomization is often implemented to reduce the possibility of contamination (application of one of the interventions in participants intended to receive the other) between individual participants that might occur in an individually-randomized trial (see Box 6), thus risk of bias due to contamination may arise less frequently in cluster randomized trials. Furthermore, recruitment and engagement may be minimal when participants are not directly recuited into the trial. For example, in the IRIS trial discussed in Box 5, in which outcomes were collected cross-sectionally from routine data at the end of the trial, there was a possibility that not all women included in the analysis had been exposed to the intervention (5). However, since the women were not directly recruited, it is unlikely that such non-exposure to the intervention would have been related to the trial context, so the risk of bias in the effect of assignment to intervention, due to deviations from intended intervention, is low.

4.1.2 Assessing the effect of adhering to intervention

It is less common for researchers (or review authors) focusing on cluster randomized trials to be concerned with the effect of **adhering to the intervention**. Because of the complex, multifaceted nature of many interventions in cluster-randomized trials, and the fact that in some trials interventions are tailored according to the cluster, it is often not obvious how to define adherence to the intended intervention, nor to identify which participants adhered. For example, in the OPERA trial, all those who consented to data collection were considered the population of interest and were included in the analysis. To identify which participants adhered would have been challenging because the intervention had multiple components aimed at both residents and staff: 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.

When the intervention is simpler, the effect of **adhering to the intervention** may be of interest, and appropriate analyses may be based on clear definitions of adherence. However, actually *identifying* the individuals who do not adhere to the protocol (and the nature of these deviations, including co-interventions) may be difficult in the context of a cluster-randomized trial. To illustrate this difficulty, consider 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. It would be possible to estimate the effect of assignment to intervention without bias, because the change in behaviour is a natural consequence of the intervention in usual practice. However, 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 from the intended intervention. From a practical point of view, identifying health professionals whose behaviour was or was not affected by the flooring would be difficult.

Occasionally, the effect of adhering to the intervention is relevant and possible to estimate in a cluster-randomized trial. This usually occurs when the intervention is a drug or therapy, or an interventional event such as attending screening. If this effect is of interest in a review, reviewers should carefully consider how deviations, including co-intervntions, will be assessed.

Box 5. Intention-to-treat analyses of cluster-randomized trials

In an intention-to-treat analysis of a cluster-randomized trial, it is important that clusters are analysed according to the intervention group to which they were randomized, regardless of intervention received. The selection of which participants should be included is not always immediately obvious. We discuss three situations in which identification of participants within clusters differs, which we refer to as a cohort-design, a simple cross-sectional design and a repeated cross-sectional design.

A **cohort design** refers to the situation in which individual participants are recruited within clusters and each is followed up. An intention-to treat analysis requires that these participants should be analysed in the intervention group to which their clusters were randomized.

A **simple cross-sectional design** is one in which participants are identified only at the end of the trial. Some cluster-randomized trials do not recruit participants directly, but collect outcome data from routine data sources at the end of the trial. It is usually assumed that including all eligible participants in the clusters from which their data arose is sufficient for an intention-to-treat analysis. In the IRIS trial, for example, outcomes were measured using routine data on all women over 16 years of age who were registered at participating general practices (the clusters) at the end of the trial period (5). It was assumed that the women on whom data were collected had the potential to be exposed to the intervention. 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 all the women included in the analysis had, strictly, been analysed in the group to which their original clusters were randomized. When the movement between clusters is more fluid, or when it is expected that the intervention effect requires participants to be exposed to the intervention over a very specific period of time (for example antenatal care), it may be necessary to take a different approach to identify the appropriate population for an intention-to-treat analysis.

A **repeated cross-sectional design** is one in which data are collected on different participants at baseline and at end of the trial (and possibly at intermediate time points). The assumption is usually made that those on whom data are collected at the end of the trial have been affected in some way by the intervention and those at baseline have not, and for an intention-to-treat analysis, individuals are again normally included in the cluster in which they were measured. The same caveats apply to this situation as to the simple cross-sectional design.

Box 6. Examples of trials which used a cluster-randomized design because of the possibility of contamination between individuals

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 experimental treatment was tablets and the comparator treatment was lotion. 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, individual participants, clusters in general, and those delivering the intervention remained unaware of the intervention and in relation to the effect of assignment to intervention, the trial was protected against bias due to deviations from intended intervention that arose because of the trial context. The effects of both assignment to intervention, and adhering to the intervention were estimated (22).

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 interaction of residents and staff within the sheltered schemes, which meant there was a possibility of the effect of the intervention affecting control participants if there were both intervention and control participants in the same cluster. The experimental treatment was vitamin D and the control treatment was placebo. Thus, as for the example above, the trial was protected against bias due to deviations from intended intervention that arose because of the trial context. The effects of both assignment to intervention and adhering to the intervention were estimated (23).

4.2 The role of masking or blinding

As for individually-randomized trials, blinding can, if successful, prevent knowledge of the intervention assignment from influencing contamination, switches to non-protocol interventions or non-adherence by trial participants. Blinding is rare in cluster randomized trials, but two examples of its use are given in Box 6.

Blinding can occur in cluster-randomized trials without the use of placebos. Participants may not be aware that they are in a trial if they are not directly recruited, such as in the IRIS trial described in Box 1 (5). Researchers may secure a waiver that allows them to tell participants at recruitment that they are in a study, but not that they are in a comparative trial (4). This is sometimes done to lessen the chance that the recruitment process affects participants' subsequent behaviour. This may reduce the likelihood that participants switch to the comparator intervention because of recruitment or engagement effects.

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. This is because there are usually more levels of personnel involved in a cluster-randomized trial than an individually-randomized trial. In particular, 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 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 (15). Assessment of the likely effect of this lack of blinding usually means considering how an awareness of the assigned intervention could have effected the extent to which participants, and people delivering the intervention, deviated from the protocol-specified interventions.

Signalling questions for this domain are provided in Box 7 for the effect of assignment to intervention. An algorithm for reaching risk of bias judgements is provided in Figure 3. Our expectation is that most reviewers will be interested in the effect of assignment to intervention. We also include signalling questions and an

algorithm for assessing risk of bias in the effect of adherence to intervention in Box 8 and Figure 4, respectively, which we expect to be relevant to only a small number of cluster-randomized trials.

4.3 Analysis issues for stepped-wedge trials

Although most of the issues arising in stepped-wedge trials are covered by this RoB 2 version for cluster-randomized trials without the need for additional consideration of how they operate in stepped wedge trials, there is an additional source of bias in the analysis of a stepped-wedge trial. In a cluster randomizedrandomized stepped-wedge trial, rather than assign a predefined proportion of the clusters to the experimental intervention and the rest to a comparator intervention, a stepped-wedge design starts with all clusters allocated to the comparator intervention and sequentially randomizes individual clusters (or groups of clusters) to switch to the experimental intervention. By the end of the trial, all clusters are implementing the experimental intervention (8). Stepped-wedge trials are increasingly used to evaluate health service and policy interventions, and are often attractive to policy makers because all clusters can expect to receive (or implement) the experimental intervention.

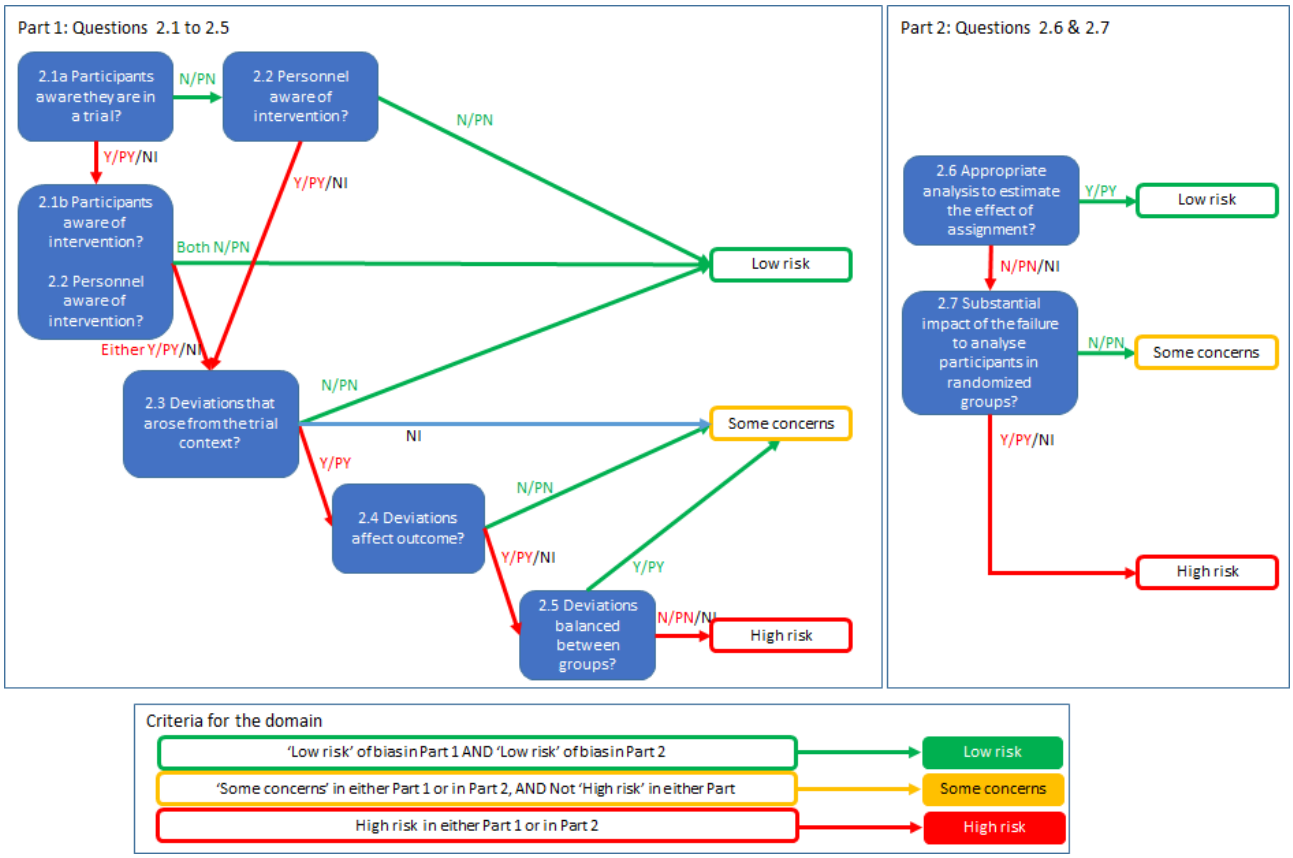
The analysis of a stepped-wedge trial must take into account the possibility of time trends. A naïve comparison of experimental intervention periods with comparator intervention periods will be confounded by any variables that change over time, since more clusters are receiving the experimental intervention during the later stages of the trial. When assessing risk of bias in a stepped-wedge trial, we suggest that users of the RoB 2 tool consider this potential source of bias when answering question 2.6 ("Was an appropriate analysis used to estimate the effect of assignment to intervention?" for the effect of assignment to intervention, or "Was an appropriate analysis used to estimate the effect of adhering to the intervention?" for the effect of adhering to intervention).

Box 7. 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 even to know that they are in a study) but not to know that they are in a trial. They therefore may not know that another interventions is being compared with theirs or what this other intervention is. This makes it impossible for them to cause deviations from the intended interventions that arise because of the trial context. Answer 'No' if participants are not aware that they are in a study or aware that they are in a study but not that they are in trial.
2.1b If Y/PY/NI to 2.1a: Were participants aware of their assigned intervention during the trial?	Answer 'Yes' if participants were aware of any part of the assigned intervention during the trial. It is important to consider all parts of the assigned intervention. Note that, for the purposes of the risk of bias tool, participants are defined as those on whom investigators seek to measure the outcome under consideration, and may be patients, the public, health professionals or other cluster staff.
2.2. Were carers and people delivering the interventions 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 non-protocol interventions, may differ between the intervention groups. Blinding carers and trial personnel, which is most commonly achieved through use of a placebo, may prevent such differences, but this is rare in cluster randomized trials.
2.3. If Y/PY/NI to 2.1b or 2.2: Were there deviations from the intended intervention that arose because of the trial context?	The guidance mostly applies as for individually-randomized trials. Deviations from the intended intervention that arise due to the trial context are rarely reported in cluster-randomized trials and may, in fact, occur rarely. This is likely to be partly because in these trials interventions

	are often aimed at clusters and cluster staff. These staff may not have the authority to introduce deviations, and if they do, may have less motivation to do so than caregivers or participants in individually randomized trials who are more directly aware of the intervention. In addition, the more complex the intervention, the more difficult it might be practically to 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 likely to have affected the outcome?	As for individually-randomized trials.
2.5. If Y/PY/NI to 2.4: Were these deviations from intended intervention balanced between groups?	As for individually-randomized trials.
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	Answer 'Yes' if all clusters and individuals were analysed according to the groups to which they were assigned. Note that there are various reasons why, in some cluster-randomized trials, it is not possible to identify with certainty the groups to which individuals in the trial were assigned, or whether some individuals change clusters part-way through the trial. If the number of such individuals can reasonably be expected to be very small and unrelated to the individual's assigned group, an analysis that analyses all individuals in the groups to which they were assigned *as far as possible* should be considered appropriate. When analyses exclude only participants with missing outcome data, these should be considered appropriate with regard to this signalling question: missing outcome data are addressed in a separate domain. Answer 'No' if trial participants were analysed according to the intervention they received, rather than according to the intervention to which they were assigned, or if analyses exclude trial participants or clusters not receiving their assigned intervention., or a stepped wedge trial does not take into account the time trend. Analyses excluding eligible trial participants after randomization should be considered inappropriate, but exclusions of ineligible participants after randomization (when eligibility was not confirmed until after randomization, and could not have been influenced by intervention group assignment) can be considered appropriate.
2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomized ?	As for individually randomized trials but bearing in mind that reviewers need to look out for entire clusters analysed in the wrong intervention group as well as individual participants.

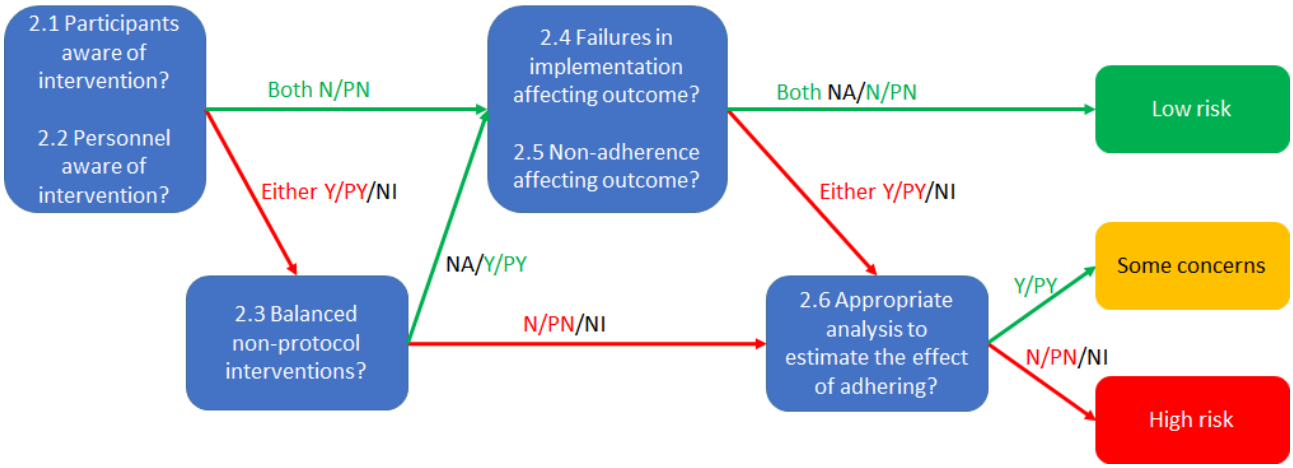
Figure 3. Algorithm for suggested judgement of risk of bias judgements for bias due to deviations from intended interventions in a cluster-randomized trial (*effect of assignment to intervention*)



Box 8. Risk of bias due to deviations from the intended interventions (*effect of adhering to ~~the assigned~~ intervention*)

Signalling questions	Elaboration
2.1 Were participants aware of their assigned intervention during the trial?	Answer 'Yes' if participants were aware of any part of the assigned intervention during the trial. It is important to consider all parts of the assigned intervention. Note that, for the purposes of the risk of bias tool, participants are defined as those on whom investigators seek to measure the outcome under consideration, and may be patients, the public, health professionals or other cluster staff.
2.2. Were carers and people delivering the interventions 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 non-protocol interventions, may differ between the intervention groups. Blinding carers and trial personnel, which is most commonly achieved through use of a placebo, may prevent such differences, but this is rare in cluster randomized trials.
2.3. [If applicable:] <u>If Y/PY/NI to 2.1 or 2.2</u>: Were important co-interventions balanced across intervention groups?	Mostly as for individually-randomized trials. It is important to consider co-interventions at both the individual and cluster level.
2.4. [If applicable:] Were there failures in implementing the intervention that could have affected the outcome?	Mostly as for individually-randomized trials. When interventions are multifacteted, it is important to consider all interventions for which implementation failures could have affected the outcome. These include interventions aimed at whole clusters and professionals in clusters, as well as those aimed at individual patients and members of the public.
2.5. [If applicable:] Was there non-adherence to the assigned intervention regimen that could have affected participants' outcomes?	As for individually-randomized trials.
2.6. <u>If N/PN/NI to 2.3 or 2.5 or Y/PY/NI to 2.4</u>: Was an appropriate analysis used to estimate the effect of adhering to the intervention?	As for individually-randomized trials.

Figure 4. Algorithm for suggested judgement of risk of bias judgements for bias due to deviations from intended interventions in a cluster-randomized trial (*effect of adhering to intervention*)



5 Bias due to missing outcome data

Broadly the same considerations apply here as for individually-randomized trials. 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. Empirical research has shown that most cluster-randomized trials have missing data but that this missingness is poorly reported and inadequately handled in analyses (24).

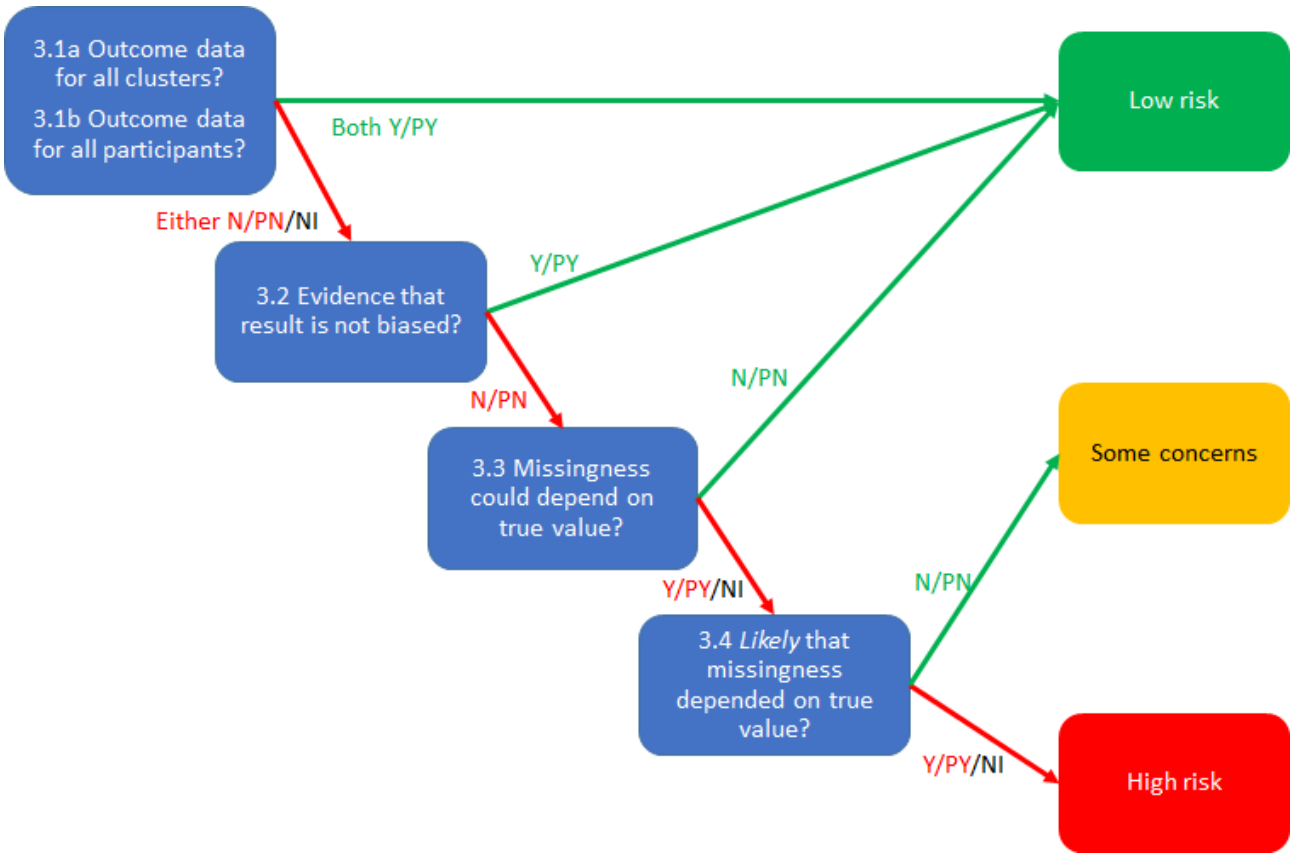
As discussed in the main RoB2 guidance document, bias arises if missingness in the outcome depends on both the intervention group and the true value of the outcome. Considerations will generally be the same as for individually randomized trials. A key consideration is whether there is differential missingness between intervention arms or, in stepped-wedge trials, between experimental and comparator intervention periods. Differential missingness may be more acute in some stepped-wedge trials because of the longitudinal nature of these designs: clusters, participants or people delivering the interventions may drop out over time, resulting in more missing data among experimental intervention periods (which are concentrated in the later stages of the trial) than among comparator intervention periods (which are concentrated in the early stages of the trial). Another consideration is the effect of whole clusters being missing. If outcomes are variable between clusters and the number of clusters is small then, almost by definition, missingness will be related to the true value of the outcome, and likely to cause bias.

Signalling questions for this domain are provided in Box 9. An algorithm for reaching risk of bias judgements is provided in Figure 5.

Box 9. Risk of bias due to missing data in a cluster-randomized trial

Signalling questions	Elaboration
3.1a Were data for this outcome available for all clusters that recruited participants?	Note that in some cluster randomized trials there may be some clusters in which no participants are recruited. This can happen only when participants are recruited following randomization and is dealt with in domain 1b. Given that there are usually a relatively small number of clusters in a cluster randomized trial, there is potential for bias in some trials even if only one cluster has no analysable participants.
3.1b Were data for this outcome available for all, or nearly all, participants within clusters?	The issues here are broadly as for individually-randomized trials. In cluster-randomized trials there may be particular complexities when clusters merge, split, or disappear.
3.2 If N/PN/NI to 3.1a or 3.1b: Is there evidence that the result was not biased by missing data?	As for individually-randomized trials.
3.3 If N/PN to 3.2 Could missingness in the outcome depend on its true value?	As for individually-randomized trials.
3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?	As for individually-randomized trials.

Figure 5. Algorithm for suggested judgement of risk of bias judgements for bias due to missing data in a cluster-randomized trial



6 Bias in measurement of the outcome

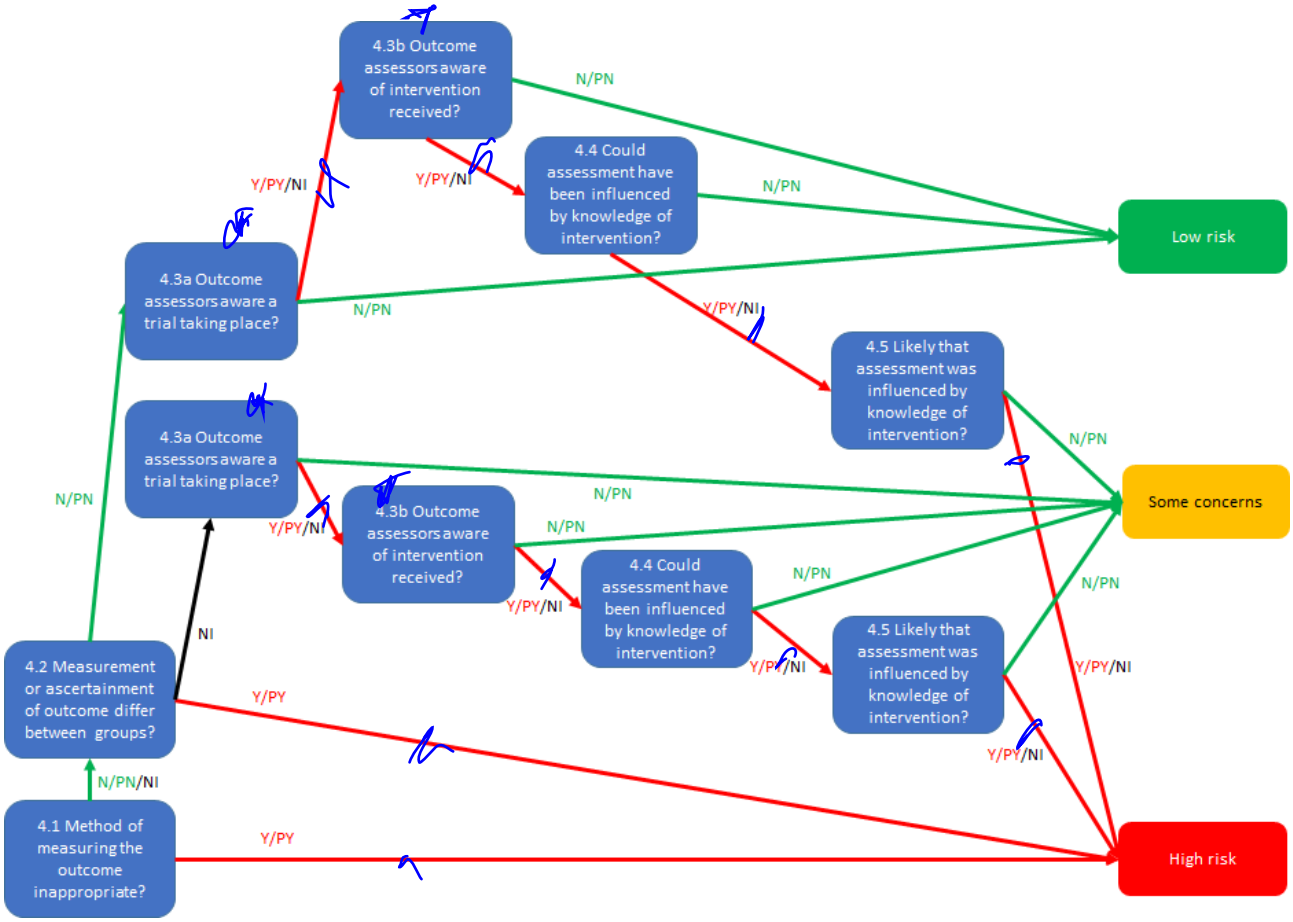
Issues in measurement of outcomes are broadly similar for cluster-randomized trials as individually-randomized trials. The key issues are identifying who is the outcome assessor and whether the assessment of the outcome is likely to be influenced by knowledge of intervention received. In individually-randomized trials, outcome measures that are (reasonably) objective such as death or cure are expected to be less subject to bias than measures such as participant-reported outcomes, because participants' assessments may be influenced by their knowledge of which intervention they received. This may also be true in cluster-randomized trials. However, in cluster-randomized trials in which participants do not know they are part of a trial, it may be unlikely that knowledge of the intervention received affects participants'self-reportedoutcome assessments. Furthermore, in such cluster-randomized trials, outcomes derived from electronic health reports, or from administrative databases, are unlikely to be affected by knowledge of intervention received at the individual level particularly when data are aggregated to cluster level by those with no knowledge of the trial purposes before being shared with researchers.

Signalling questions for this domain are provided in Box 10. An algorithm for reaching risk of bias judgements is provided in Figure 6.

Box 10. Risk of bias in measurement of the outcome in a cluster-randomized trial

Signalling questions	Elaboration
4.1 Was the method of measuring the outcome inappropriate?	As for individually randomized trials.
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	As for individually randomized trials.
4.3a If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware that a trial was taking place?	This question applies to cluster-randomized trials in which participants report their outcomes themselves, for example in a questionnaire. If they are not aware that they are in a trial then their self-assessment cannot be affected by assignment even if they are aware of the intervention they received.
4.3b If Y/PY/NI to 4.3a: Were outcome assessors aware of the intervention received by study participants?	Answer '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 the individual who provides the data (usually patients or clinicians)and the individual responsible for extracting the data can be considered as oucome assessors.
4.4 If Y/PY/NI to 4.3b: Could assessment of the outcome have been influenced by knowledge of intervention received?	As for individually-randomized trials.
4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	As for individually-randomized trials.

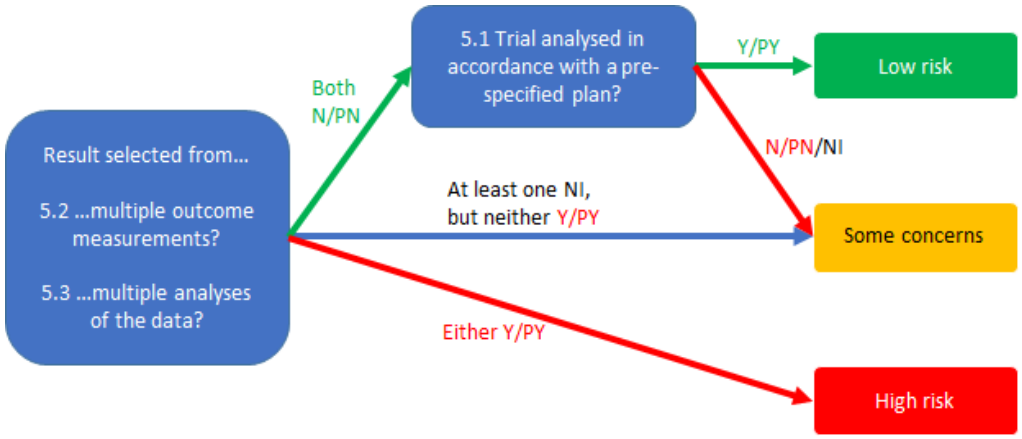
Figure 6. Algorithm for suggested judgement of risk of bias judgements for bias in measurement of the outcome in a cluster-randomized trial (*effect of assignment to intervention*)



7 Bias in selection of the reported result

Issues of selective reporting are the same for cluster-randomized trials as for individually-randomized trials. The algorithm for reaching risk of bias judgements is provided in Figure 7.

Figure 7. Algorithm for suggested judgement of risk of bias judgements for bias due to selection of the reported result in a cluster-randomized trial



8 References

1. Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
2. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. BMJ. 2016;355:i4919.
3. Feldman HA, McKinlay SM. Cohort versus cross-sectional design in large field trials: precision, sample size, and a unifying model. Stat Med. 1994;13(1):61-78.
4. Taljaard M, Weijer C, Grimshaw JM, Eccles MP, Ottawa Ethics of Cluster Randomised Trials Consensus G. The Ottawa Statement on the ethical design and conduct of cluster randomised trials: precis for researchers and research ethics committees. BMJ. 2013;346:f2838.
5. 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.
6. 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.
7. 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.
8. 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.
9. Efron B. Forcing a sequential experiment to be balanced. Biometrika. 1971;58(3):403-17.
10. 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.
11. 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.
12. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. Am J Public Health. 2004;94(3):416-22.
13. 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.
14. Eldridge S, Kerry S, Torgerson DJ. Bias in identifying and recruiting participants in cluster randomised trials: what can be done? BMJ. 2009;339:b4006.
15. 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.
16. 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.
17. 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.
18. 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.
19. Nikiema L, Huybregts L, Kolsteren P, Lanou H, Tiendrebeogo S, Bouckaert K, et al. Treating moderate acute malnutrition in first-line health services: an effectiveness cluster-randomized trial in Burkina Faso. Am J Clin Nutr. 2014;100(1):241-9.
20. 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.
21. Eldridge SM. Assessing, understanding and improving the efficiency of cluster randomised trials in primary care (PhD thesis): University of London; 2005.
22. 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.
23. Martineau AR, Hanifa Y, Witt KD, Barnes NC, Hooper RL, Patel M, et al. Double-blind randomised controlled trial of vitamin D3 supplementation for the prevention of acute respiratory infection in older adults and their carers (ViDiFlu). Thorax. 2015;70(10):953-60.
24. 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.