

BMJ Open Quality of reporting of pilot and feasibility cluster randomised trials: a systematic review

Claire L Chan,¹ Clémence Leyrat,² Sandra M Eldridge¹

To cite: Chan CL, Leyrat C, Eldridge SM. Quality of reporting of pilot and feasibility cluster randomised trials: a systematic review. *BMJ Open* 2017;**7**:e016970. doi:10.1136/bmjopen-2017-016970

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-016970>).

Received 23 March 2017
Revised 14 July 2017
Accepted 17 July 2017

ABSTRACT

Objectives To systematically review the quality of reporting of pilot and feasibility of cluster randomised trials (CRTs). In particular, to assess (1) the number of pilot CRTs conducted between 1 January 2011 and 31 December 2014, (2) whether objectives and methods are appropriate and (3) reporting quality.

Methods We searched PubMed (2011–2014) for CRTs with ‘pilot’ or ‘feasibility’ in the title or abstract; that were assessing some element of feasibility and showing evidence the study was in preparation for a main effectiveness/efficacy trial. Quality assessment criteria were based on the Consolidated Standards of Reporting Trials (CONSORT) extensions for pilot trials and CRTs.

Results Eighteen pilot CRTs were identified. Forty-four per cent did not have feasibility as their primary objective, and many (50%) performed formal hypothesis testing for effectiveness/efficacy despite being underpowered. Most (83%) included ‘pilot’ or ‘feasibility’ in the title, and discussed implications for progression from the pilot to the future definitive trial (89%), but fewer reported reasons for the randomised pilot trial (39%), sample size rationale (44%) or progression criteria (17%). Most defined the cluster (100%), and number of clusters randomised (94%), but few reported how the cluster design affected sample size (17%), whether consent was sought from clusters (11%), or who enrolled clusters (17%).

Conclusions That only 18 pilot CRTs were identified necessitates increased awareness of the importance of conducting and publishing pilot CRTs and improved reporting. Pilot CRTs should primarily be assessing feasibility, avoiding formal hypothesis testing for effectiveness/efficacy and reporting reasons for the pilot, sample size rationale and progression criteria, as well as enrolment of clusters, and how the cluster design affects design aspects. We recommend adherence to the CONSORT extensions for pilot trials and CRTs.

BACKGROUND

In a cluster randomised trial (CRT) clusters, rather than individuals, are the units of randomisation. A cluster is a group (usually predefined) of one or more individuals. For example, clusters could be hospitals and the individuals, the patients within those hospitals. CRTs are often chosen for logistical reasons, prevention of contamination across individuals or because the intervention

Strengths and limitations of this study

- We used a robust search and data extraction procedure, including validation of the screening/sifting process and double data extraction.
- We may have missed some studies, since our criteria excluded studies not including ‘pilot’ or ‘feasibility’ in the title or abstract, and those not clearly in preparation for a main trial.

is targeted at the cluster level. CRTs are useful for evaluating complex interventions. However, they have added complexity in terms of design, implementation and analysis and so it is important to ensure that carrying out a CRT is feasible before conducting the future definitive trial.¹

A feasibility study conducted in advance of a future definitive trial is a study designed to answer the question about whether the study can be done and whether one should proceed with it. A pilot study answers the same question but in such a study part or all of the future trial is carried out on a smaller scale.² Thus, all pilot studies are also feasibility studies. Pilot studies can be randomised or non-randomised; for brevity we use the term pilot CRT throughout this paper to refer to a randomised study with a clustered design that is in preparation for a future definitive trial assessing effectiveness/efficacy.^{3 4} The focus of pilot trials is on investigating areas of uncertainty about the future definitive trial to see whether it is feasible to carry out, so the data, methods and analysis are different from an effectiveness/efficacy trial. In particular, more data might be collected on items such as recruitment and retention to assess feasibility, methods may include specifying criteria to judge whether to proceed with the future definitive trial, and analysis is likely to be based on descriptive statistics since the study is not powered for formal hypothesis testing for effectiveness/efficacy.



CrossMark

¹Centre for Primary Care and Public Health, Queen Mary University of London, London, UK

²Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

Correspondence to
Claire L Chan;
c.l.chan@qmul.ac.uk

Arnold *et al* highlight the importance of pilot studies being of high quality.⁵ Good reporting quality is essential to show how the pilot has informed the future definitive trial as well as to allow readers to use the results in preparing for similar future trials. The number of pilot and feasibility studies in the literature is increasing. However, Arain *et al* indicate that reporting of pilot studies is poor.⁶ There are no previous reviews of the reporting quality of pilot CRTs, despite the extra complications arising from the clustered structure. The aim of this review is to reveal the quality of reporting of pilot CRTs published between 1 January 2011 and 31 December 2014. We extracted information to describe the sample of pilot CRTs and to assess quality, with quality criteria based on the Consolidated Standards of Reporting Trials (CONSORT) extension for CRTs,⁷ and a CONSORT extension for pilot trials for which SE and CC were involved in the final stages of development during this review.^{3 4} We present recommendations for improving the conduct, analysis and reporting of these studies and expect this to improve the quality, usefulness and interpretation of pilot CRTs in the future. We know current reporting of CRTs is suboptimal,^{8–11} and thus we expected the reporting of pilot CRTs to be even poorer.

The questions addressed by this review are:

1. How many pilot CRTs have been conducted between 1 January 2011 and 31 December 2014?
2. Are pilot CRTs using appropriate objectives and methods?
3. To what extent is the quality of reporting of pilot CRTs sufficient?

METHODS

Inclusion and exclusion criteria

We included papers published in English with a publication date (print or electronic) between 1 January 2011 and 31 December 2014. We chose the start date to be after the updated CONSORT 2010 was published.¹² We estimated a search covering 4 years would give us a reasonable number of papers to perform our quality assessment, and that later papers would be similar in terms of quality of reporting since the CONSORT for pilot trials was not published until the end of 2016. The study had to be a CRT, have the word 'pilot' or 'feasibility' in the title or abstract, be assessing some element of feasibility and show evidence that the study was in preparation for a specific trial assessing effectiveness/efficacy that is planned to go ahead if the pilot trial suggests it is feasible (ie, not just a general assessment of feasibility issues to help researchers in general, although pilot trials may do this as an addition). Regardless of how authors described a study, we did not consider it to be a pilot trial if it was *only* looking at effectiveness/efficacy because we wanted to exclude those studies that claim to be a pilot/feasibility trial simply as justification for small sample size.¹³ The paper had to be reporting results (ie, not a protocol or statistical analysis plan) and had to be the first published paper reporting pilot outcomes (ie, not an extension/follow-up

study for a pilot study already reported, and not a second paper reporting further pilot outcomes). Interim analyses, analyses before the study was complete and internal pilots were excluded; the CONSORT extension for pilot trials on which we based the quality assessment does not apply to internal pilots.^{3 4} No studies were excluded on the basis of quality since the aim was to assess the quality of reporting.

Data sources and search methods

We searched PubMed for relevant papers in September 2015. We searched for the words 'pilot' or 'feasibility' in the title or abstract, a search strategy similar to that used by Lancaster *et al*.¹⁴ We combined this with a search strategy to identify CRTs; this was similar to the strategy used by Diaz-Ordaz *et al*.⁸ The full electronic search strategy is given in online supplementary appendix 1.

Sifting and validation

The titles and abstracts of all papers identified by the electronic search were screened by CC for possible inclusion. Full texts were obtained for those papers identified as definitely or possibly satisfying the inclusion criteria and sifted by CC for inclusion. As validation, CL carried out the same screening and sifting process independently on a 10% random sample of electronically identified papers. For full texts where there was uncertainty whether the paper should be included, it was referred to SE for a final decision.

Refining the inclusion process

We refined the screening and sifting process following piloting. In particular, we rejected a more restrictive PubMed search that required 'pilot' or 'feasibility' in the title rather than allowing these words to occur in the title *or* abstract because this missed relevant papers; we altered the order of the exclusion criteria to make the process more streamline; and we relaxed one inclusion criteria, requiring evidence that the pilot trial was in preparation for a future definitive trial rather than an explicit statement that authors were planning a future definitive trial. The protocol was updated, and is available from the corresponding author.

Data extraction

CC and CL independently extracted data from all papers selected for inclusion in the review, and followed rules on what to extract (see 'Further information' column of online supplementary appendix 2). Extracted data were recorded in an Excel spreadsheet. Discrepancies were resolved by discussion between CC and CL, and where agreement could not be reached a final decision was made by SE.

For each pilot CRT included in the review, we extracted information to describe the trials, including publication date (print date unless there was an earlier electronic date), country in which the trial was set, number of clusters randomised, method of cluster randomisation and following the CONSORT extension for pilot trials'

recommendation to focus on objectives rather than outcomes, the primary objective. We defined the primary objective using method similar to that used by Diaz-Ordaz *et al*⁸ for primary outcomes that is, as that specified by the author, else the objective used in the sample size justification, or else the first objective mentioned in the abstract or else main text.

To assess whether the pilot trials were using appropriate objectives and methods, we collected information on whether the primary objective was about feasibility, the method used to address the main feasibility objective, the rationale for numbers in the pilot trial and whether there was formal hypothesis testing for, or statements about, effectiveness/efficacy without a caveat about the small sample size.

To assess reporting quality, we created a list of quality assessment items based on the CONSORT extension for pilot trials.^{3,4} We also looked at the CONSORT extension for CRTs,⁷ and incorporated any cluster-specific items into our quality assessment items. Where a CRT item became less relevant in the context of a pilot trial, we did not extract it (eg, whether variation in cluster sizes was formally considered in the sample size calculation). In addition, where there was a substantial difference between the item for the CONSORT extension for CRTs and that for the pilot trial extension and the items were not compatible, we used the latter item (eg, focusing on objectives rather than outcomes). We recognised the need to balance comprehensiveness and feasibility.¹¹ Therefore, where items referred to objectives or methods, we extracted this for the primary objective only. We also did not extract on whether papers reported a structured summary of trial design, methods, results and conclusions. The final version of the full list of data extracted, and further information on each item extracted, is included in online supplementary appendix 2.

Refining data extraction

Initially, CC extracted data on a random 10% sample of papers. However, some of the items were difficult to extract in a clear, standardised way, as similarly noted by Ivers *et al*,¹¹ so these items were removed. In particular, whether the objectives, intervention or allocation concealment were at the individual level, cluster level or both; and other analyses performed or other unintended consequences (difficult to decipher from papers whether it classified as an 'other'). Furthermore, some items were deemed easier to extract if split into two items, for example, 'reported why the pilot trial ended/stopped' which we subsequently split into 'reported the pilot trial ended/stopped' and 'if so, what was the reason'.

Analysis

Data were analysed using Excel V.2013. We describe the characteristics of the pilot CRTs using descriptive statistics. Where we extracted text, we established categories during analysis by grouping similar data, for example, grouping the different primary objectives. To assess adherence to

the CONSORT checklists, we present the number and percentage reporting each item. This report adheres, where appropriate, to the Preferred Reporting Items for Systematic reviews and Meta-Analyses statement.¹⁵

Patient involvement

No patients were involved in the development of the research question, design or conduct of the study, interpretation or reporting. No patients were recruited for this study. There are no plans to disseminate results of the research to study participants.

RESULTS

The electronic PubMed search identified 257 published papers. We rejected 108 during screening (29 not reporting results; 32 not about a single randomised trial; 46 not cluster randomised; 1 interim analysis). The remaining 149 full-text articles were assessed for eligibility, and 131 more papers were rejected (1 not reporting results; 13 not about a single randomised trial; 25 not cluster randomised; 8 analyses before study complete/internal pilot; 32 not assessing feasibility; 50 not in preparation for a future definitive effectiveness/efficacy trial; 2 not the first published paper reporting pilot outcomes). This left 18 studies to be included in the analysis. [A1-A18]. The full list of studies is included in table 1, with citations in online Supplementary file 2. Figure 1 shows the flow diagram of the identification process for the sample of 18 pilot CRTs.

There was 96% agreement between CC and CL for the 10% random sample used for the screening and sifting validation (based on 26 papers), with a kappa coefficient of 0.84.

Trial characteristics

In general, the more recent the publication date, the more pilot CRTs were identified, but with the most identified in 2013 (table 2). Of the 18 included studies, the majority (56%) were set in the UK. All other countries were represented only once except for Canada (three trials) and the USA (two trials). Of those reporting the method of randomisation, the majority (69%) used stratified with blocked randomisation. The median number of clusters randomised was 8 (IQR: 4–16) with a range from 2 to 50.

Pilot trial objectives and methods

Ten (56%) of the 18 included pilot trials had feasibility as their primary objective, for example, assessing feasibility of implementing the intervention (6 trials), of recruitment and retention (3 trials) and of the cluster design (1 trial) (table 3). All 10 trials reported a corresponding measure to assess the feasibility objective; most (90%) used descriptive statistics and/or qualitative methods to address the objective. In one trial, a statistical test was used to address their primary feasibility objective without

Table 1 Pilot cluster randomised trials included in this review

Author	Year*	Journal	Title	Cluster
Begh [A1]	2011	<i>Trials</i>	Promoting smoking cessation in Pakistani and Bangladeshi men in the UK: pilot cluster randomised controlled trial of trained community outreach workers.	Census lower layer super output areas
Jones [A2]	2011	<i>Paediatric exercise science</i>	Promoting fundamental movement skill development and physical activity in early childhood settings: a cluster randomised controlled trial.	Childcare centres
Légaré [A3]	2010	<i>Health expectations</i>	Training family physicians in shared decision making for the use of antibiotics for acute respiratory infections: a pilot clustered randomised controlled trial.	Family medicine groups
Hopkins [A4]	2012	<i>Health education research</i>	Implementing organisational physical activity and healthy eating strategies on paid time: process evaluation of the UCLA WORKING pilot study.	Worksites—health and human service organisations
Jago [A5]	2012	<i>International Journal of Behavioral Nutrition and Physical Activity</i>	Bristol girls dance project feasibility trial: outcome and process evaluation results.	Secondary schools
Taylor [A6]	2011	<i>Clinical rehabilitation</i>	A pilot cluster randomised controlled trial of structured goal-setting following stroke.	Rehabilitation services
Drahota [A7]	2013	<i>Age and ageing</i>	Pilot cluster randomised controlled trial of flooring to reduce injuries from falls in wards for older people.	Study areas—bays within hospitals
Frenn [A8]	2013	<i>Journal for Specialists in Pediatric Nursing</i>	Authoritative feeding behaviours to reduce child BMI through online interventions.	Classrooms
Gifford [A9]	2012	<i>World views on evidence-based nursing</i>	Developing leadership capacity for guideline use: a pilot cluster randomised control trial.	Service delivery centres with nursing care for diabetic foot ulcers
Jones [A10]	2013	<i>Journal of Medical Internet Research</i>	Recruitment to online therapies for depression: pilot cluster randomised controlled trial.	Postcode areas
Moore [A11]	2013	<i>Substance abuse treatment, prevention, and policy</i>	An exploratory cluster randomised trial of a university halls of residence-based social norms marketing campaign to reduce alcohol consumption among first year students.	Residence halls
Pai [A12]	2013	<i>Implementation science</i>	Strategies to enhance venous thromboprophylaxis in hospitalised medical patients (SENTRY): a pilot cluster randomised trial.	Hospitals
Reeves [A13]	2013	<i>BMC health services research</i>	Facilitated patient experience feedback can improve nursing care: a pilot study for a phase III cluster randomised controlled trial.	Wards
Teut [A14]	2013	<i>Clinical interventions in Ageing</i>	Effects and feasibility of an Integrative Medicine programme for geriatric patients: a cluster randomised pilot study.	Shared apartments
Jago [A15]	2014	<i>International journal of Behavioural nutrition and physical activity</i>	Randomised feasibility trial of a teaching assistant-led extracurricular physical activity intervention for those aged 9–11 years: action 3:30.	Primary schools

Continued

Table 1 Continued

Author	Year*	Journal	Title	Cluster
Michie [A16]	2014	<i>Contraception</i>	Pharmacy-based interventions for initiating effective contraception following the use of emergency contraception: a pilot study.	Pharmacies
Mytton [A17]	2014	Health technology assessment	The feasibility of using a parenting programme for the prevention of unintentional home injuries in the under-fives: a cluster randomised controlled trial.	Children's centres
Thomas [A18]	2014	<i>Trials</i>	Identifying continence options after stroke (ICONS): a cluster randomised controlled feasibility trial.	Stroke services

*We extracted the earlier of the print and electronic publication year.

the authors designing the study to be adequately powered to do so.

The remaining eight trials had an effectiveness/efficacy primary objective, and used statistical tests to address this. Nevertheless, these eight trials all had feasibility as one of their other objectives (this was an inclusion criterion). The feasibility objectives were similar to those where the feasibility was primary, but expressed more generally in two trials, for example, looking at the feasibility of the future definitive trial,[A16] and looking at whether the future definitive trial could answer the effectiveness question and which study design would enable this.[A10] In only three trials was a measure to assess the feasibility objective reported, using either quantitative or qualitative measures.

Eight trials reported a rationale for the numbers in the pilot trial, with all of these following best practice in not basing the rationale on a formal sample size calculation for effectiveness/efficacy. Nine (50%) trials performed any formal hypothesis testing for effectiveness/efficacy, regardless of whether this was for the primary or a secondary objective. Of these nine trials, four of the conclusions about effectiveness/efficacy were made without any caveats about the imprecision of estimates or possible lack of representativeness because of the small samples.

Quality of reporting—by items

The pilot CRTs in our review are published after the CONSORT 2010 for RCTs but before the CONSORT extension for pilot trials. Therefore, to present data on quality of reporting, we looked at our list of quality assessment items based on the CONSORT extension for pilot trials, and grouped reporting items into three categories (table 4): (1) items in the CONSORT extension for pilot trials that are new compared with CONSORT 2010 for RCTs, (2) items in the CONSORT extension for pilot trials that are substantially adapted from CONSORT 2010 for RCTs and (3) items in the CONSORT extension for pilot trials that are the same as or have only minor

differences from CONSORT 2010 for RCTs, plus items in the CONSORT extension for CRTs.^{3 4 7 12}

In the tables, denominators for proportions are based on papers for which the item is relevant. Not all items are relevant for all trials, due to their design, so we highlight where this applies in the table footnotes. The footnote of table 4 also explains where the quality assessment items come from, with different font differentiating items based on the CONSORT extension for pilot trials and the CONSORT extension for CRTs, and a key to highlight which of the three categories above the item falls under.

New items

Five new items were added to the CONSORT extension for pilot trials on the identification and consent process, progression criteria, other unintended consequences, implications for progression and ethical approval.^{3 4} See items with [N] in column 2 of table 4. In our review, how participants were identified and consented was reported by 50% and 76% of the pilot CRTs, respectively, but how clusters were identified and consented was reported by just 33% and 11%, respectively. Only three trials (17%) reported criteria used to judge whether or how to proceed with the future definitive trial, with two giving numbers that must be exceeded such as recruitment, retention, attendance and data collection percentages,[A17, A2] and one giving categories of 'definitely feasible', 'possibly feasible' and 'not feasible'. [A12] The item on other unintended consequences was reported by none of the pilot CRTs, although it is unclear whether this is due to poor reporting or because no unintended consequences occurred. Implications for progression from pilot to future definitive trial was reported by 16 trials (89%), with 9 reporting to proceed/proceed with changes, 5 reporting further research or piloting is needed first and 2 reporting to not go ahead with the future definitive trial. Ninety-four per cent reported ethical approval/research review committee approval, but only 47% of them also reported the corresponding reference number.

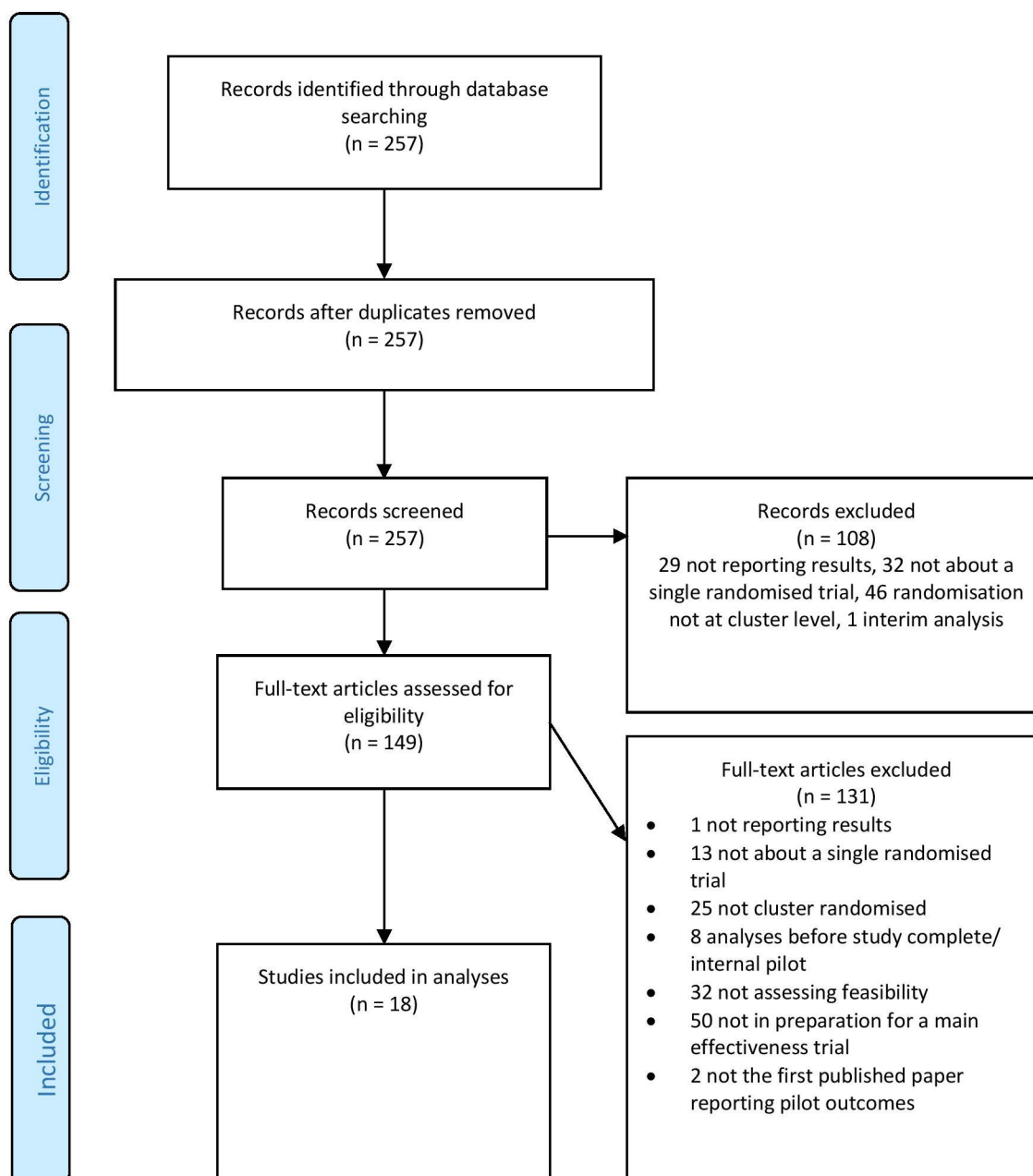


Figure 1 Flow diagram of the identification process for the sample of 18 pilot cluster randomised trials included in this review.

Substantially adapted items

Six items in the CONSORT extension for pilot trials were substantially adapted from CONSORT 2010 for RCTs, regarding reasons for the randomised pilot trial, sample size rationale for the pilot trial, numbers approached and/or assessed for eligibility, remaining uncertainty about feasibility, generalisability of pilot trial methods and findings and where the pilot trial protocol can be accessed.^{3 4} See items with [S] in column 2 of [table 4](#). Reasons for the randomised pilot trial were reported by 39% of the pilot CRTs. Eight trials (44%) gave a rationale for the sample size of the pilot trial. Pilot trials should always report a rationale for their sample size; this can be qualitative or quantitative, but should not be based on a formal sample size calculation for effectiveness/efficacy.

In this review, the rationales were based on logistics,[A15] resources,[A14] time,[A16] a balance of practicalities and need for reasonable precision,[A18] a general statement that it was considered sufficient to address the objectives of the pilot trial,[A17] formal [A6] and non-formal [A7] calculation to enable estimation of parameters in the future definitive trial, and a formal calculation based on the primary feasibility outcome.[A12] Of these rationales, good examples include ‘The decision to include eight apartment-sharing communities was based on practical feasibility that seemed appropriate according to funding and the personal resources available’,[A14] as well as ‘The sample size was chosen in order to have two clusters per randomised treatment and the number of participants per cluster was based on the number of

Table 2 Characteristics of pilot cluster randomised trials included in this review

Characteristic	Number of trials (%)
Publication year (earlier of the print and electronic publication date)	
2010*	1 (6)
2011	3 (17)
2012	3 (17)
2013	7 (39)
2014	4 (22)
Country	
UK	10 (56)
Canada	3 (17)
USA	2 (11)
Germany	1 (6)
New Zealand	1 (6)
Australia	1 (6)
Method of cluster randomisation†	
Simple	1 (8)
Stratified with blocks	9 (69)
Blocked only	2 (15)
Bias coin method	1 (8)
Number of clusters randomised‡	
Median (IQR)	8 (4 to 16)
Range	2 to 50
Average cluster size§	
Median (IQR)	32 (14 to 82)
Range	7 to 588

*One paper has an extracted publication year outside of the 2011–2014 range. This is because the print publication date for this paper was 2011 but the online publication date was 2010, so the paper satisfies the inclusion criteria which states that the publication date, print or electronic, must be between 2011 and 2014, but we extract the earlier of the print and electronic dates.

†13 of the 18 trials reported their method of randomisation. Percentages are given as a percentage of these 13 trials.

‡Not reported for one trial.

§Defined as number of individuals randomised divided by number of clusters randomised, based on 12 trials that reported information on both.

degrees of freedom (df) needed within each cluster to have reasonable precision to estimate a variance'. [A6] The number of individuals approached and/or assessed for eligibility was reported by 47%, and the number of clusters by 56%. Remaining uncertainty was reported by 56% of the pilot CRTs. 89% reported generalisability of pilot trial methods/findings to the future definitive trial or other studies, but clarity of reporting was lacking as it was difficult to distinguish between references to the future definitive trial versus other future studies due to ambiguous phrases such as 'in a future trial'. Only 39% reported where the pilot trial protocol could be accessed.

Table 3 Pilot trial objectives and methods

Characteristic	Number of trials (%)
Primary objective is feasibility*	10 (56)
Main feasibility objective given	
Where feasibility is primary objective	
Implementing intervention	6/10 (60)
Recruitment and retention	3/10 (30)
Feasibility of cluster design	1/10 (10)
Where feasibility is not primary objective†	
Implementing intervention	3/8 (38)
Recruitment	2/8 (25)
Cluster design	1/8 (13)
Feasibility of trial being able to answer the effectiveness question (and what study design would enable this)	1/8 (13)
Feasibility of larger study	1/8 (13)
Method used to address main feasibility objective given	
Where feasibility is primary objective	
Descriptive statistics and/or qualitative	9/10 (90)
Statistical test	1/10 (10)
Where feasibility is not primary objective	
Descriptive statistics/qualitative	3/8 (38)
None given/reported elsewhere	5/8 (63)
Rationale for numbers in pilot trial based on formal power calculation for effectiveness/efficacy‡	0/8 (0)
Performing any formal hypothesis testing for effectiveness/efficacy	9/18 (50)
Making any statements about effectiveness/efficacy without a caveat	4/18 (22)

*Where the primary objective was not feasibility, the primary objective was effectiveness/potential effectiveness and was addressed using statistical tests.

†One of the inclusion criteria was that studies were assessing feasibility, but it did not have to be the primary objective.

‡Based on eight trials that reported a rationale for the sample size of the pilot trial.

Items essentially taken from CONSORT 2010 for RCTs or the CONSORT extension for CRTs

For the remaining items, reporting quality was variable. Some were reported by fewer than 20% of the pilot CRTs, for example, considering the cluster design in the sample size rationale for the pilot trial (17%) (item 7a), reporting whether consent was sought from clusters (11%) and who enrolled them (17%) (items 10c and 10a), how people were blinded (7% of applicable trials) (item 11a), number of excluded individuals (6% of applicable trials) and clusters (18% of applicable trials) after randomisation (item 13b) and a table showing baseline cluster characteristics (11%) (item 15). Those reported most well, by >80% of the pilot CRTs, included reporting 'pilot' or 'feasibility' in the title (83%) (item 1a), scientific background and

Table 4 Number (%) of reports adhering to pilot cluster randomised trial quality criteria

	Item	Criterion	n (%)
Title and abstract	1a	Term 'pilot' or 'feasibility' included in the title Identification as a pilot or feasibility randomised trial in the title	15 (83) 12 (67)
	1a	Term 'cluster' included in the title Identification as a cluster randomised trial in the title	12 (67) 12 (67)
Introduction	2a [S]	Scientific background and explanation of rationale for future definitive trial reported Reasons for randomised pilot trial reported	18 (100) 7 (39)
	2a	Rationale given for using cluster design	6 (33)
Methods—trial design	3a	Description of pilot trial design	18 (100)
	3a	Definition of cluster	18 (100)
	3b	Reported any changes to methods after pilot trial commencement If yes, reported reasons	5 (28) 5/5 (100)
Methods—participants	4a	Reported eligibility criteria for participants	13 (72)
	4a	Reported eligibility criteria for clusters	9 (50)
	4b	Reported settings and locations where the data were collected	18 (100)
	4c	Reported how participants were identified	9 (50)
	[N]	Reported how clusters were identified Reported how participants were consented * Reported how clusters were consented	6 (33) 13/17 (76) 2 (11)
	5	Described the interventions for each group	13 (72)
Methods—interventions	6b	Reported any changes to pilot trial assessments or measurements after pilot trial commencement If yes, reported reasons	1 (6) 1/1 (100)
	6c [N]	Reported criteria used to judge whether, or how, to proceed with the future definitive trial	3 (17)
Methods—sample size	7a [S]	Reported a rationale for the sample size of the pilot trial	8 (44)
	7a	Cluster design considered during the description of the rationale for numbers in the pilot trial	3 (17)
	7b	Reported stopping guidelines	0 (0)

Continued

Table 4 Continued

Item	Criterion	n (%)
Methods — randomisation		
8a	Reported method used to generate the random allocation sequence	9 (50)
8b	Reported randomisation method	13 (72)
9	Reported mechanism used to implement the random allocation sequence	4 (22)
	Reported allocation concealment	7 (39)
10/	Reported who:	8 (44)
10a	Generated the random allocation sequence	3 (17)
	Enrolled clusters	4 (22)
	Assigned clusters to interventions	
10c	Reported from whom consent was sought	2 (11)
	<i>Reported whether consent was sought from participants</i>	17 (94)
	Reported whether consent was sought from clusters	2 (11)
	Reported whether participant consent was sought before or after randomisation*	8/17 (47)
Methods — blinding		
11a	Reported on whether there was blinding	10 (56)
	Reported who was blinded†	6/14 (43)
	Reported how they were blinded†	1/14 (7)
Methods — analytical methods		
12a	Reports clustering accounted for in any of the methods used to address pilot trial objectives/research questions‡	13/17 (76)
Results — participant flow		
13§	Reports a diagram with flow of individuals through the trial	12 (67)
13§	Reports a diagram with flow of clusters through the trial	10 (56)
13a/	Reported number of:	8/17 (47); 10/18 (56)
13a	Individuals (clusters) approached and/or assessed for eligibility¶	13/17 (76); 17/18 (94)
[S]	Individuals (clusters) randomly assigned¶	8/17 (47); 5/17 (29)
	Individuals (clusters) that received intended treatment¶; ¶	16/17 (94); 14/17 (82)
	Individuals (clusters) that were assessed for primary objective¶; ¶	
13b/	Reported number of:	11/16 (69); 6/17 (35)
13b	Losses for individuals (clusters) after randomisation*; ¶	1/17 (6); 3/17 (18)
	Exclusions for individuals (clusters) after randomisation¶; ¶	
14a	Reported on dates defining the periods of recruitment	8 (44)
	Reported on dates defining the periods of follow-up	11 (61)
14b	Reported the pilot trial ended/stopped	0 (0)
Results — baseline data		
15	Reported a table showing baseline characteristics for the individual level if yes, by group	12 (67)
		11/12 (92)
15	Reported a table showing baseline characteristics for the cluster level if yes, by group	2 (11)
		2/2 (100)
Results — outcomes and estimation		
17a	Reported results for main feasibility objective (quantitative or qualitative)††	13/17 (76)

Continued

Table 4 Continued

Item	Criterion	n (%)
Results— harms		
19	Reported on harms or unintended effects	4 (22)
19a [N]	Reported other unintended consequences	0 (0)
Discussion		
20 [S]	Reported limitations of pilot trial	17 (94)
	Reported sources of potential bias	10 (56)
	Reported remaining uncertainty	10 (56)
21 [S]	Reported generalisability of pilot trial methods/findings to future definitive trial or other studies	16 (89)
22	Interpretation of feasibility consistent with main feasibility objectives and findings††	12/17 (71)
22A [N]	Reported implications for progression from the pilot to the future definitive trial	16 (89)
Other information		
23	Reported registration number for pilot trial	11 (61)
	Reported name of registry for pilot trial	11 (61)
24 [S]	Reported where the pilot trial protocol can be accessed	7 (39)
25	Reported source of funding	18 (100)
26 [N]	Reported ethical approval/research review committee approval If yes, reported reference number	17 (94) 8/17 (47)

Item numbers in normal font refer to the item in the CONSORT extension for pilot trials that the quality assessment item is based on.

Item numbers in bold italics refer to the item in the CONSORT extension for CRTs that the quality assessment item is based on.

[N] represents new items in the CONSORT extension for pilot trials compared with the CONSORT 2010 for RCTs.

[S] represents items in the CONSORT extension for pilot trials that are substantially adapted from the CONSORT 2010 for RCTs.

*Item not relevant for one trial [A12] because they said that the Ethics Board determined it could be conducted without informed consent from patients or surrogates.

†Item not relevant for four trials [A7, A10, A12, A18] because they reported that blinding was not used.

‡Item not relevant for one trial because no CIs/p values were given, [A17] so clustering did not need to be accounted for in any of their methods because effect estimates are not biased by cluster randomisation, only CIs/p values.

§The CONSORT statements do not include an item 13 but there is a participant flow subheading which strongly recommends a diagram. We therefore reference this subheading as 'item 13' here.

¶Not relevant for one trial due to the design of the study.[A10] (This paper was different from the others such that it was not relevant to extract these items. The clusters were postcode areas and they were assessing two online recruitment interventions and comparing the success of the recruitment interventions. As such, participants were those who completed the online questions, and each arm of the study had a 'total population ranging from 1.6 to 2 million people clustered in four postcode areas'.)

**Not relevant for two trials due to the design of these studies.[A10, A12] (See reason above for A10. For A12, data were collected from medical patient charts so these items were not relevant to extract.)

††One paper reports the feasibility results in a separate paper so is not included.[A3]

Table 5 Number (%) of quality assessment criteria reported by each pilot cluster randomised trial in this review

Study	Overall n (%)*	Title and abstract and introduction n (%)	Methods n (%)	Results n (%)	Discussion and other information n (%)
Drahota [A7]	50 (70)	6 (86)	17 (59)	18 (78)	9 (75)
Pai [A12]	48 (69)	5 (71)	17 (61)	18 (78)	8 (67)
Mytton [A17]	50 (68)	4 (57)	21 (66)	13 (57)	12 (100)
Thomas [A18]	46(67)	5 (71)	17 (59)	15 (65)	9 (90)
Teut [A14]	49 (66)	6 (86)	20 (63)	14 (61)	9 (75)
Taylor [A6]	47 (64)	7 (100)	16 (52)	13 (57)	11 (92)
Légaré [A3]	42 (58)	3 (43)	18 (56)	14 (61)	7 (64)
Begh [A1]	41 (56)	5 (71)	16 (52)	11 (48)	9 (75)
Jago [A15]	39 (55)	4 (57)	11 (38)	13 (57)	11 (92)
Jones [A10]	32 (52)	7 (100)	10 (33)	6 (50)	9 (75)
Moore [A11]	37 (52)	5 (71)	13 (45)	8 (35)	11 (92)
Michie [A16]	36 (51)	3 (43)	15 (52)	8 (36)	10 (83)
Jones [A2]	37 (51)	3 (43)	15 (48)	10 (45)	9 (75)
Jago [A5]	33 (46)	4 (57)	13 (45)	10 (43)	6 (50)
Gifford [A9]	33 (45)	6 (86)	12 (39)	8 (35)	7 (58)
Reeves [A13]	29 (41)	6 (86)	11 (38)	7 (32)	5 (42)
Frenn [A8]	18 (26)	1 (14)	5 (17)	7 (32)	5 (42)
Hopkins [A4]	16 (23)	2 (29)	4 (14)	4 (18)	6 (50)

*This is the overall number (percentage) of the quality assessment items in [table 4](#) that are reported by each study. The other columns look at this within categories. Note that the denominator varies between studies because not all quality assessment items are relevant for all studies (see footnote of [table 4](#)) and not applicable for some items if a related item is not reported (see items 3b, 6b, 15, 26 in [table 4](#)).

explanation of rationale for future definitive trial (100%) (item 2a), pilot trial design (100%) (item 3a), nature of the cluster (100%) (item 3a), settings and locations where the data were collected (100%) (item 4b), whether consent was sought from participants (94%) (item 10c), number of clusters randomised (94%) and assessed for primary objective (82% of applicable trials) (item 13a), number of individuals assessed for primary objective (94% of applicable trials) (item 13a), limitations of pilot trial (94%) (item 20) and source of funding (100%) (item 25).

Quality of reporting—by study

Finally, in [table 5](#) we present the number (percentage) of quality assessment items reported by each study. We provide an overall score, as well as a score by categories of CONSORT. The quality of reporting varies across studies, with five of the pilot CRTs reporting over 65% of the quality assessment items and two of the pilot CRTs reporting under 30%. There does not appear to be a trend of reporting quality with time. Five of the studies report 90% or more of the quality assessment items in the ‘discussion and other information’ category, and only two

studies report <50%. Two of the studies report 100% of the items in the 'title and abstract and introduction' category, and five studies report <50%. The highest percentage of items reported by a study in the 'methods' category is 66% and the lowest is 14%. Similarly, the highest percentage of items reported by a study in the 'results' category is 78% and the lowest is 18%. Within studies, the category that is best reported tends to be the 'discussion and other information' category (had the highest percentage for 10 of the 18 pilot CRTs).

DISCUSSION

Main findings

This is the first study to assess the reporting quality of pilot CRTs using the recently developed CONSORT checklist for pilot trials.^{3,4} Our search strategy and inclusion criteria identified 18 pilot CRTs published between 2011 and 2014. Most studies were published in the UK, perhaps driven by the availability of funding or the large number of CRTs and interest in complex interventions in the UK.

With respect to the pilot CRT objectives and methods, a considerable proportion of papers did not have feasibility as their primary objective. Of the trials reporting a sample size rationale for the pilot, all followed best practice in not carrying out a formal sample size calculation for effectiveness/efficacy, yet a substantial proportion performed formal hypothesis testing for effectiveness/efficacy. This could indicate an inappropriate attachment to hypothesis testing, although many did explain it was an indication of *potential* effectiveness or that the study was underpowered. Investigators wanting to assess effectiveness/efficacy and use statistical tests to do so should be performing a properly powered definitive trial, otherwise there is the potential for misleading conclusions affecting clinical decisions as well as misinformed decisions about the future definitive trial.¹⁶ One may however look at *potential* effectiveness, for example, using an interim or surrogate outcome, with a caveat about the lack of power.^{3,4} Moreover, one may include a progression criterion based on potential effect. If so, Eldridge and Kerry recommend any interpretation of potential effect is done by looking at the limits of the CI,¹³ and one should also pay attention to features of the pilot which might have biased the result (eg, convenience sampling of clusters). A positive effect finding excluding the null value would still justify the future definitive trial to estimate the effect with greater certainty, but a negative effect finding excluding the null value (ie, strongly suggesting harm), or even a finding where the clinically important difference is excluded, might suggest not proceeding. It is good practice to prestate such progression criteria. Finally, one may use estimates from outcome data, for example, as inputs for the sample size calculation for the future definitive trial. In particular, for pilot CRTs we may be interested in estimating the intraclass correlation coefficient (ICC), although we note that the ICC estimate from a pilot CRT should not be the only

source for the future definitive trial sample size, because of the large amount of imprecision in a pilot trial.¹⁷

Reporting quality of pilot CRTs was variable. Items reported well included reporting the term 'pilot' or 'feasibility' in the title, generalisability of pilot trial methods/findings to the future definitive trial or other studies and implications for progression from the pilot to the future definitive trial, although clarity could be improved when referring to the future definitive trial rather than other future studies in general. Items least well reported included reasons for the randomised pilot trial, sample size rationale for the pilot trial, criteria used to judge whether or how to proceed with the future definitive trial and where the pilot trial protocol can be accessed. These items are important so that readers can understand whether the uncertainty they are facing about their future trial has already been addressed in a pilot, researchers can make sure they have enough patients to achieve the pilot trial objectives, readers can understand the criteria for progression and to prevent against selective reporting.

For items related to the cluster aspect of pilot CRTs, most pilot CRTs reported the nature of the cluster, and the number of clusters randomised and assessed for the primary objective. The items reported least well included considering the cluster design during the sample size rationale for the pilot trial, reporting who enrolled clusters and how they were consented, number of exclusions for clusters after randomisation and a table showing baseline cluster characteristics. Although the number of clusters in a pilot trial is usually small, it is still important to, for example, describe the cluster-level characteristics using a baseline table as it may give helpful information important for planning the future definitive trial. Moreover, while nearly all trial reports described whether consent was sought from individuals or not, seeking agreement from clusters was only described in a small minority. The items on agreement from and enrolment of clusters, baseline cluster characteristics and number of excluded clusters are particularly important to report, since they may affect assessment of feasibility.

If we consider why some items may have been well adhered to and others not, it is interesting to observe that new items added to the CONSORT extension for pilot trials and items substantially adapted from CONSORT 2010 for RCTs were in general not well adhered to. This could perhaps be because of somewhat newer ideas that may not have been considered during design such as specifying progression criteria and considering a rationale for numbers in the pilot. Alternatively, perhaps there were aspects sometimes done but not reported due to lack of reporting guidance to remind authors; for example, the new items on how clusters were identified and consented, other unintended consequences and ethical approval/research review committee approval reference number, and the substantially adapted items on reporting reasons for the pilot trial, number of individuals approached and/or assessed for eligibility and where the pilot trial protocol can be accessed. With the item on unintended

consequences, we recognise that investigators are free to choose what they interpret and report as an unintended consequence. We recommend careful thought that all unintended consequences that may affect the future definitive trial are reported. It is also interesting to observe that many of the most poorly reported items concerned methods/design (progression criteria; enrolment and consent of clusters), and in particular, justification of design aspects (reasons for randomised pilot trial; sample size rationale for pilot trial including consideration of cluster design). Within studies, the category that is worst reported is the methods, despite being crucial to allow the reader to judge the quality of the trial.

Comparison with other studies

There has not been a previous review of pilot trials using the new CONSORT extension for pilot trials.^{3,4} However, the review by Arain *et al* looking at pilot and feasibility studies reported that 81% were performing hypothesis testing with sample sizes known to be insufficient,⁶ compared with 50% of pilot CRTs in our review. Arain *et al* also reported 36% of studies performing sample size calculations for the pilot. In our review, 17% performed calculations (all based on feasibility objectives), but if we include those that also correctly reported a rationale for the numbers in the pilot but without any calculation then this was 44%.

The general message that reporting of CRTs is suboptimal still holds.⁸⁻¹¹ The review by Diaz-Ordaz *et al*⁸ of definitive trial CRTs reported that 37% presented a table showing baseline cluster characteristics, compared with 11% of pilot CRTs in our review. Diaz-Ordaz *et al*⁸ also reported that 27% accounted for clustering in sample size calculations,⁸ and a recent review by Fiero *et al* reported 53%.¹⁰ However, just 17% of pilot CRTs in our review considered the cluster design in the sample size rationale for the pilot trial. Both these CRT reviews reviewed effectiveness/efficacy CRTs, for which the need to take account of clustering in sample sizes is generally well understood compared with pilot trials. In pilot trials, the rationale for considering the clustered design in deciding on numbers in the pilot may be different, for example, considering the number of df needed within each cluster to estimate a variance. [A6] In pilot trials, including a number of clusters with different characteristics may also be important to get an idea about the implementation of an intervention across different clusters.

Strengths and limitations

We used a robust search and data extraction procedure, including validation of the screening/sifting process and double data extraction. However, the use of only one database, PubMed, which is comprehensive but not exhaustive, may have missed eligible papers, and the use of conditions #3, #5 and #6 (see online supplementary appendix 1) may have been restrictive. Our aim was to get a general idea of reporting issues in the area, rather than doing a completely comprehensive search. Our inclusion criteria stipulated that

papers must have the word 'pilot' or 'feasibility' in the title or abstract, so we may have missed some pilot CRTs and thus may have overestimated the percentage reporting 'pilot' or 'feasibility' in the title. This strategy may also have resulted in a skewed sample of papers with a greater tendency to adhere to CONSORT guidelines. However, our review suggests reporting of pilot CRTs need improving, so our conclusion would remain the same. We required authors to report that the trial was in preparation for a future definitive trial, so we expect that items related to the future definitive trial (eg, progression criteria, generalisability, implications) may be better reported than they would for all publications of pilot CRTs, which might include papers that did not report that they were in preparation for a future definitive trial clearly enough to be included. During sifting, we identified 32 trials that had 'pilot' or 'feasibility' in the title/abstract, but were not assessing feasibility. A third of these were identified because they referred to 'pilot' or 'feasibility' at some point in the abstract but it was not in reference to the current trial (eg, stating feasibility has already been shown), but the other two-thirds were labelled as a pilot or feasibility trial yet showed no evidence of assessing feasibility and were only assessing effectiveness. This is an important point as our review may appear to overestimate reporting quality by not including these studies. That there are underpowered main trials being published as pilot or feasibility studies is something that the academic community should look to prevent. During sifting, we also identified 50 trials that were assessing feasibility but did not show evidence of being in preparation for a future definitive trial. Most were assessing the feasibility of implementing an intervention targeted at members of the public, or discussing feasibility of the intervention with the aim of providing information to help researchers wanting to implement a similar intervention in similar settings or to raise questions for future research, rather than being in preparation for a trial assessing effectiveness/efficacy. Some of these 50 trials also appeared to be small effectiveness studies labelled as a pilot, usually only mentioning feasibility once or twice throughout the paper, with one trial explicitly stating that "Because of organisational changes... we had to stop the inclusion after 46 participants, and the study is consequently defined as a pilot study".¹⁸ For the few trials that were potentially pilot CRTs not reported clearly enough, the authors only spoke of future studies in general rather than clearly specifying the study was in preparation for a specific future definitive trial. Related to this, it is of interest to know the proportion of our 18 pilot CRTs that are actually followed by a future definitive trial, and we plan to investigate this in future.

CONCLUSION

We may have overestimated the reporting quality of pilot CRTs; nevertheless, our review demonstrates that reporting of pilot CRTs need improving. The identification of just 18 pilot CRTs between 2011 and 2014, mainly from the UK, highlights the need for increased awareness

of the importance of carrying out and publishing pilot CRTs and good reporting so that these studies can be identified. Pilot CRTs should primarily be assessing feasibility, and avoiding formal hypothesis testing for effectiveness/efficacy. Improvement is needed in reporting reasons for the pilot, rationale for the pilot trial sample size and progression criteria, as well as the enrolment stage of clusters and how the cluster design affects aspects of design such as numbers of participants. We recommend adherence to the new CONSORT extension for pilot trials, in conjunction with the CONSORT extension for CRTs.^{3 4 7} We encourage journals to endorse the CONSORT statement, including extensions.

Exclusive license The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in BMJ Open and any other BMJPG products to exploit all subsidiary rights, as set out in the licence <http://journals.bmj.com/site/authors/editorial-policies.xhtml#copyright> and the Corresponding Author accepts and understands that any supply made under these terms is made by BMJPG to the Corresponding Author. All articles published in BMJ Open will be made available on an Open Access basis (with authors being asked to pay an open access fee - see <http://bmjopen.bmj.com/site/about/resources.xhtml>) Access shall be governed by a Creative Commons licence – details as to which Creative Commons licence will apply to the article are set out in the licence referred to above.

Transparency declaration This manuscript is an honest, accurate, and transparent account of the study being reported. No important aspects of the study have been omitted, and any discrepancies from the study as planned have been explained.

Contributors SE conceived the study and advised on the design and protocol. CLC developed the design of the study, wrote the protocol and designed the screening/sifting and data extraction sheet. CLC performed screening and sifting on all papers identified by the electronic search, and CL carried out validation of the screening/sifting process. CLC and CL performed independent data extraction on all papers included in the review. CLC conducted the analyses of the data and took primary responsibility for writing the manuscript. All authors provided feedback on all versions of the paper. All authors read and approved the final manuscript. CLC is the study guarantor.

Funding CLC (nee Coleman) was funded by a National Institute for Health Research (NIHR) Research Methods Fellowship. This article presents independent research funded by the NIHR.

Disclaimer The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests SME and CLC are authors on the new CONSORT extension for pilot trials.

Ethics approval No ethics approval was necessary because this is a review of published literature. No patient data were used in this manuscript.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Extraction data are available from the corresponding author.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which

permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Eldridge SM, Ashby D, Feder GS, *et al.* Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials* 2004;1:80–90.
- Eldridge SM, Lancaster GA, Campbell MJ, *et al.* Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. *PLoS One* 2016;11:e0150205.
- Eldridge SM, Chan CL, Campbell MJ, *et al.* CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *Pilot Feasibility Stud* 2016;2:64.
- Eldridge SM, Chan CL, Campbell MJ, *et al.* CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ* 2016;355:i5239.
- Arnold DM, Burns KE, Adhikari NK, *et al.* The design and interpretation of pilot trials in clinical research in critical care. *Crit Care Med* 2009;37:S69–74.
- Araim M, Campbell MJ, Cooper CL, *et al.* What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Med Res Methodol* 2010;10:67.
- Campbell MK, Piaggio G, Elbourne DR, *et al.* Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012;345:e5661.
- Diaz-Ordaz K, Froud R, Sheehan B, *et al.* 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.
- Diaz-Ordaz K, Kenward MG, Cohen A, *et al.* Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines. *Clin Trials* 2014;11:590–600.
- Fiero MH, Huang S, Oren E, *et al.* Statistical analysis and handling of missing data in cluster randomized trials: a systematic review. *Trials* 2016;17:72.
- Ivers NM, Taljaard M, Dixon S, *et al.* Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000–8. *BMJ* 2011;343:d5886.
- Moher D, Hopewell S, Schulz KF, *et al.* CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
- Eldridge S, Kerry S. *A practical guide to cluster randomised trials in health services research*. Hoboken, New Jersey: Wiley, 2012.
- Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. *J Eval Clin Pract* 2004;10:307–12.
- Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264–9.
- Kistin C, Silverstein M. Pilot studies: a critical but potentially misused component of interventional research. *JAMA* 2015;314:1561–2.
- Eldridge SM, Costelloe CE, Kahan BC, *et al.* How big should the pilot study for my cluster randomised trial be? *Stat Methods Med Res* 2016;25:1039–56.
- Mikkelsen LR, Mikkelsen SS, Christensen FB. Early, intensified home-based exercise after total hip replacement—a pilot study. *Physiother Res Int* 2012;17:214–26.