Supplementary Appendix for *Completeness of reporting and risks of overstating impact in cluster randomised trials: a systematic review*

Elizabeth L. Turner, Alyssa C. Platt, John A. Gallis, Kaitlin Tetreault, Christina Easter, Joanne E. McKenzie, Stephen Nash, Andrew B. Forbes, Karla Hemming on behalf of the *CRT Binary Outcome Reporting Group*

Appendix A: Systematic Review: Summary, Methods, Results and Discussion

Appendix B: Supplementary Tables

Appendix C: Supplementary Figures

Appendix D: Other Supplementary Materials

Appendix A Systematic Review Summary, Methods, Results and Discussion

Summary

Background To avoid scale-up of interventions with smaller than perceived impact, complete and accurate reporting of expected impact is needed. This is of great importance in global health research to protect precious resources. In this context, the cluster randomised trial (CRT) design is used to evaluate complex, multicomponent interventions. For binary outcomes, this means reporting both relative and absolute measures of effect. Otherwise, intervention impact may be overstated. This can arise when only a relative measure is provided for a rare outcome (risk ≤ 10%) or when the odds ratio is reported for a common outcome (risk > 10%) but is interpreted as a risk ratio. We assessed reporting practices and potential to overstate impact in contemporary CRTs with primary binary outcome.

Methods Systematic review of all reports of parallel-arm CRTs with primary binary outcome indexed in Cochrane CENTRAL and published in 2017. Data abstraction performed in duplicate.

Main outcome measures: whether relative and absolute effects were reported; type of relative effects reported; potential for overstating impact.

Findings Of 711 abstracts screened, 73 had a primary binary outcome and met inclusion criteria. Most (95·9%) reported risks by arm, some (8·2%) provided no effect measure, while few (17·8%) reported both relative and absolute effects. Instead, most (63·0%) reported a relative measure only. Of the 59 reporting a relative measure, most (64·4%) reported an odds ratio. Of 64 CRTs reporting an effect measure and risks by arm, most (62·5%, n=40) had the potential to overstate intervention impact. Of these, 12 (30%) with rare outcome and only a relative measure; 28 (70.0%) with common outcome and odds ratio.

Interpretation Given that reporting of CRTs with binary outcomes is often incomplete and that many have the potential to overstate impact, interventions with smaller than perceived impact may be adopted.

Funding: Partial funding from funding agencies in Australia (NHMRC), UK (DfiD, MRC, NIHR) and USA (NIH).

Key words: binary outcomes, cluster randomised trial, reporting, treatment effect measures, risk ratio, odds ratio, risk difference, rare outcome

Introduction

See details in the main manuscript text.

Methods

As described in the main manuscript text, we undertook a systematic review of design, analysis and reporting of CRTs with primary binary outcome. A detailed protocol and statistical analysis plan are included in **Supplementary Material 1.** The main manuscript includes a brief description of the methods and results of the review. To complement that text, here we provide a complete description of the methods, results and strengths and limitations of the review. In order that the text below can "stand alone" from the main text, we include all information that is provided in the main manuscript and have not removed that information. Note too that two figures are included in the manuscript. In the results below, they are referred to as **Main Text Figures 1 and 2**.

Search strategy and selection criteria

The Cochrane Central Register of Controlled Trials (CENTRAL) was searched to identify eligible reports of published (either online or "in print") CRTs (see **Table S1** for the search strategy). The search was restricted to 1st January 2017 to 31st December 2017, and was last conducted on October 29, 2018. We derived a search strategy based on identifying terms used to describe cluster randomised trials (title, abstract or key words) and binary outcomes (abstract), informed by previously developed search strategies to identify cluster randomised trials in PubMed.[1]

Selection process

Abstracts identified using the search strategy were exported to Rayyan software[2] where duplicates and references that only existed in trial registration sites (e.g. clinicaltrials.gov) were excluded. The remaining abstracts were randomly assigned for screening in duplicate by pairs of the core study team (all listed authors except KT and JEM) who independently screened their assigned abstracts within the Rayyan platform. Full text was retrieved where abstracts appeared to meet the inclusion criteria and the final assessment of eligibility was made based on the full text. When there was disagreement between the pair, the final decision was made by a third reviewer (either ELT or KH, who were intentionally not paired together as independent reviewers).

Data abstraction and management

A data extraction form was developed (**Supplementary Material 2**), with some response options based on a previously published systematic review of cluster randomised crossover studies.[1] An accompanying REDCap database was developed for data entry.[3, 4] Data abstractors were the core study team (except JM) plus individuals who agreed to extract data from two CRT articles and to participate at one of three data reconciliation workshops held by the: (1) conference on *Current Development of Methods in Cluster Randomised Trials*, Queen Mary University London, London, UK (11/18); (2) *Biostatistics, Epidemiology, and Research Design* (BERD) Core, Duke University, USA (02/19); and, (3) *The Clinical Trials Unit,* Birmingham, UK (04/19).

Procedures were the same at all three locations except for some minor differences noted below. For each workshop, all individuals who had agreed to participate as data abstractors were randomly paired together and then randomly matched to two of the identified articles. The Duke workshop of statisticians included some participants with limited CRT experience (MSc and PhD students), therefore participants at this location were randomly matched to contain an "experienced" and "novice" member. After pairing, each data abstractor was sent an individualized link to the REDCap data abstraction form with links to each of their two assigned articles in order to access the article from their own institutional website. Abstractors were asked to affirm they would not save or distribute the PDF without explicit permission from the publisher and, given such affirmation, were then given access to a PDF downloadable from the REDCap link. Both members of a pair were asked to independently extract the same set of items except for a sub-set of questions about study characteristics (e.g., study size and setting) for which one member of a pair was randomly assigned to extract (**Supplementary Material 2** and schematic in **Figure S2**). In the case of multiple primary outcomes, abstractors were told to choose the binary outcome that was first mentioned in the abstract. Data abstraction was completed independently by each member of a pair, after which each abstractor received an automated email with a PDF attachment of their own abstracted data (but not that of their partner).

In-person data reconciliation workshops were held so that pairs of data abstractors could review answers and reach a final agreement on the double-abstracted data. At least one of the first and last authors (ELT and KH, respectively) was present at each of these workshops (acting as the final decision maker in the case of disagreement). One member of each pair entered the final version of the abstracted data in a new REDCap data record. From this, the final data set of reconciled data was created. After all three workshops had been completed, we determined that some additional information would be valuable, including whether authors reported the use of the CONSORT statement,[5, 6] as well as

the CONSORT reporting requirements of the publishing journal (**Supplementary Material 3**). To achieve this, two members of the Duke team (AP, KT) independently extracted such data with a final reconciled version determined by agreement (with ELT, as needed).

Data analysis

Characteristics of data abstractors were summarized descriptively. Reliability was quantified using percentage agreement for the data abstracted in duplicate before the in-person reconciliation workshops. Analyses of characteristics and outcomes of the included studies were descriptive, using summary statistics and data visualizations and were performed using Stata software version 16.[7] We made two assumptions about study characteristics. First, where the primary outcome was unclear or measured at multiple time points, data abstractors inferred the primary outcome and/or time point from that which was emphasized in the abstract or results section of such manuscripts. Second, if study authors stated that some form of "logistic regression" was used for analysis, it was assumed that the link was logit and that the outcome family was binomial when this was not stated explicitly.

In order to classify whether there was the potential to overstate intervention impact in each of the included CRT articles we first classified the primary outcome of each CRT as 'rare' if the reported outcome risk in either trial arm was ≤10%, so that 'non-rare', henceforth referred to as 'common', was defined as both trial arm outcome risks exceeding 10%. Next, we classified each CRT report as having the potential to overstate intervention impact if either of the following two conditions was satisfied: the outcome was rare and only a relative measure (odds ratio or risk ratio or other measure) was reported (i.e. no absolute measure was provided), or, the outcome was common and the odds ratio was selected as the relative measure (irrespective of whether an absolute measure was reported). The rationale for our definition is as follows. For the first condition in the rare outcome setting, if only a relative measure is reported (either odds ratio or risk ratio or other), it is possible to overstate intervention impact unless the estimated relative effect is linked to the absolute values of the risk and of their absolute difference (see first example in the introduction of the main manuscript). For the second condition in the common outcome setting, if the odds ratio is selected as the relative measure and is interpreted as a risk ratio, then it overstates the relative impact of the intervention, irrespective of whether an absolute measure is also presented. To further facilitate an understanding of the potential to overstate intervention impact using the odds ratio, we estimated the unadjusted risk ratio using the reported outcome risks for articles reporting an odds ratio, from which we calculated the ratio of the odds ratio relative to the risk ratio.

Role of the funding source

The funders of this research (see details at end) had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Results

Study selection

The search strategy yielded 939 abstracts, of which 228 were excluded due to being duplicates, triplicates or indexed on a trials registration website only (**Figure S3**). Of the remaining 711 abstracts, 89 were determined to be eligible from title and abstract screening. Of these 89, 16 were excluded at full-text screening (primarily because the article did not present the main analysis of the trial but was instead a secondary or sub-group analysis paper), leaving 73 included articles. Reporting of the current manuscript is in accordance with the PRISMA statement (**Supplementary Material 4**).[8]

Data abstractor characteristics and data abstraction reliability

Final data from the 73 articles in the review were abstracted by 82 individuals who participated at one of the three in-person workshops (**Figure S1**). Five additional individuals from the London meeting performed pre-meeting data abstraction, which was subsequently not included in the final data set; 3 due to not being able to attend the in-person meeting and 2 due to their pair of articles being inadvertently reassigned to a later pair of participants. The 82 individuals who contributed final data reconciliations for at least one of the 73 articles were mostly statisticians (84·1%), working in academic settings (85·4%), UK-based (62·2%) with experience of at least one CRT (65·9%) (**Table S2**). Pairwise agreement between data abstractors on variables from independent pre-workshop data abstraction was high, with 85·7% agreement on 95 potential variables across the 73 articles and with agreement of 85% on all 28 variables that were abstracted for all 73 articles (**Table S3**).

Study characteristics

Of the 73 included CRTs, the most common domains studied were infectious diseases (19, 26·0%) and women's health (16, 21·9%), most were conducted in Europe or Africa (22, 30·1% and 19, 26·0%, respectively), most randomised health facilities or providers (in total 41, 56·2%) or by geographic area (14, 19·2%); most (46, 64·8%) studied direct participant health promotion or educational interventions and most used a comparator with no active intervention (54, 74·0%) (**Table S4**). In terms of design, few (17, 23·3%) used simple randomization (most adopted some form of restricted randomization such as stratification) and most (51, 69·9%) reported accounting for clustering in the power calculation (**Table S5**). Most (49, 67·1%) enrolled a cohort of individuals that was followed over time **(Table S4)** and most (48, 66·7%) had a single post-randomisation follow-up time-point. Follow-up data were typically collected using a questionnaire or survey (34, 46·6%) or via electronic/medical records (22, 30·1%) **(Table S5)**. In other design features, most (52, 71·2%) CRTs enrolled fewer than 40 clusters and median (25^{th}, 75^{th} percentile) cluster size was 48 (20, 220) **(Table S4)**.

Reporting of primary binary outcomes in CRTs

Of 73 CRTs with a primary binary outcome, the outcome was not explicitly identified as "primary" in the manuscript in 11 (15·1%) reports (item 6a of the CONSORT extension for cluster trials[6]), with a greater number (21 CRTs, 28·8%) not explicit in the abstract. Relatedly, of the 24 CRTs with more than one post-randomisation follow-up time-point, only 9 (37·5%) were explicit about which time point was the primary assessment time (**Table S5**). Most (70, 95·9%) of the 73 CRTs reported the outcome by study arm (item 17a), again with a smaller proportion doing so in the abstract (50, 68·5%) (**Table S6**). Few (13, 17·8%) reported both a relative and absolute measure (**Table S6, Main Text Figure 1**) and therefore did not satisfy CONSORT recommendations (item 17b). Instead, in the main text, most (46, 63·0%) reported a relative measure only, 8 (11%) an absolute measure only and 6 (8·2%) reported no effect measure, with a larger number (15, 20·5%) reporting no effect measure in the abstract. Of the 5 CRTs not reporting a treatment effect anywhere in the manuscript, most (4/5) reported only proportions by arm with no statistical inference (i.e. no p-value or confidence interval for the difference). Of the 59 CRTs (80·8%) reporting a relative measure, most (38, 64·4%) reported an odds ratio, with fewer (19, 32·2%) reporting a risk ratio **(Table S6).** Of the 21 CRTs (28·8%) reporting an absolute effect, most (17, 81·0%) reported a risk difference, with 2 (9·5%) reporting a difference-in-difference of proportions **(Table S6)**. Overall, most CRTs provided an incomplete picture of evidence of intervention impact.

Potential for overstating intervention effects

Of the 64 CRTs reporting an effect measure with accompanying risks by arm, most (40, 62·5%) were classified as having the potential for the intervention impact to be overstated (**Table S6**). Potential overstatement was primarily (28/64, 43·8%) because the odds ratio was the chosen relative measure for a common outcome (>10% risk), with the remaining 12 (12/64, 18·8%) because only a relative measure (odds ratio or risk ratio or other) was reported for a rare

outcome (<10% risk). The magnitude of this potential for overstatement is considerable and is illustrated for the 59 studies that reported a relative measure (**Main Text Figure 2**). For the 28 CRTs in the common outcome setting that reported an odds ratio as the relative measure (shown in orange with reference risk > 10%), many of those odds ratios are of large magnitude, and most are of large magnitude relative to the risk ratio; the estimated ratio of the odds ratio to risk ratio averages 1.4 (standard deviation = 0.6) with a maximum of 3.2, indicating a large potential for misinterpretation (**Main Text Figure 2**, footnote). Similarly, for the CRTs in the rare outcome setting that reported only a relative effect (shown in orange with reference risk ≤ 10%), those relative effects are typically of a large magnitude. For example, one CRT has a risk ratio of almost 25 and a risk in the reference arm less than 5%. Putting these statistics in context, not only do most CRTs provide an incomplete picture of evidence of intervention impact, most have the potential to overstate intervention impact.

Analysis of primary binary outcomes in CRTs

The predominant software used for analysis was Stata (24, 32·9%), followed by SAS (16, 21·9%) and R (14, 19·2%) **(Table S5)**. Data from most (62, 84·9%) of the 73 CRTs were analyzed using individual-level analysis only, with an additional 3 (4·1%) that used both cluster- and individual-level analysis versus 8 (11·0%) CRTs which used cluster-level analysis only (**Table S7**). Overall, most (64, 87·7%) analyses accounted for the clustered design: all of the 11 CRTs with cluster-level analysis and 86·2% (56) of the 65 CRTs with individual-level analysis. A few CRTs (5/65, 7·7%) presented only a statistical test accounting for clustering rather than a model-based analysis thus not providing a confidence interval (**Table S7**). Of the 51 CRTs that implemented an individual-level regression approach accounting for clustering, most common (31, 60·8%) was mixed effects modeling (e.g. logistic regression with random effects), followed by generalized estimating equations (13, 26·0%). Overall, a non-trivial fraction (12·3%) of the articles have the potential for incorrect evidence of impact as a result of analysis that does not account for the CRT design.

Discussion

Strengths of this systematic review include the rigorous and comprehensive methods used to screen and abstract data, with duplication at key stages of the review. This review used an innovative methodology whereby data abstraction leveraged the expertise of a large group of individuals.[9] This is a methodology that allows reviews of this nature to be conducted to scale, in a timely way, and capitalizes on the knowledge of experts. In other strengths, rather than focus exclusively on the frequency of reporting of both absolute and relative measures, we also presented visual information as to magnitudes of reported effects (**Main Text Figure 2**) and, in order to more directly translating our findings in to a more interpretable form, we evaluated potential for overstatement of intervention impact. A weakness of this evaluation is that we did not assess whether authors of each CRT had misinterpreted the reported results. In particular, it is possible that authors correctly interpreted an odds ratio for a report with common outcome but our concern is that readers may overstate impact if they misinterpret the odds ratio as a risk ratio. Relatedly, whilst our review identified incomplete reporting for most trials, we are uncertain as to the reason why, particularly given that most (79·5%) of the 73 CRTs appeared in journals (48 in total) that endorsed the 2010 CONSORT statement on reporting of RCTs[5] (**Tables S8-S9**). Other weaknesses include the potential for inconsistent quality in data abstraction due to the large number of individuals involved in this process. Nevertheless, quality control measures were implemented by the core study team and reliability assessments of the independent data abstractions shows that most data were abstracted without any disagreements, although this does not necessarily mean that the abstracted data were a reliable assessment of the "truth". Whilst we searched only one source for CRTs - Cochrane CENTRAL - this is the most comprehensive source of randomised trials available, including records from multiple bibliographic databases and trial registers (PubMed, Embase, CINAHL, ClinicalTrials,gov, WHO's International Clinical Trials Registry Platform) as well as records contributed from other sources (e.g. handsearching). Finally, we assessed reporting of primary evidence from

randomised controlled trials and not evidence in systematic reviews, which might be argued to lead more directly to changes in health care and in policy.

References

1. Arnup SJ, Forbes AB, Kahan BC, Morgan KE, McKenzie JE. Appropriate statistical methods were infrequently used in cluster-randomized crossover trials. J Clin Epidemiol. 2016;74:40-50.
2. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Systematic reviews. 2016;5(1):210.
3. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics. 2009;42(2):377-81.
4. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. Journal of biomedical informatics. 2019;95:103208.
5. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.
6. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012;345:e5661.
7. StataCorp L. Stata statistical software: release 16. College Station, TX. 2019.
8. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264-9.
9. Hemming K, Carroll K, Thompson J, Forbes A, Taljaard M, Dutton SJ, et al. Quality of stepped-wedge trial reporting can be reliably assessed using an updated CONSORT: crowd-sourcing systematic review. J Clin Epidemiol. 2019;107:77-88.

Appendix B: Supplementary Tables

Table S1. Inclusion and exclusion criteria and search terms used in systematic review

Inclusion criteria were all of the following:
Two-arm parallel CRT
At least one binary primary outcome
Published in the peer-reviewed literature
Full-scale CRT (i.e., not a pilot or feasibility CRT)
Main analysis (i.e., not secondary or subgroup)
Report published in 2017 (either online first or "in print")
Exclusion criteria were at least one of the following:
Any trial design other than a two-arm parallel CRT (i.e. including but not limited to stepped wedge, crossover and factorial designs)
Pilot or feasibility CRT
Secondary or subgroup analysis of a CRT
Protocol or study design paper
Methodological paper
Report appearing in conference proceedings
CRT that appears only on trial registration website (e.g. clinicaltrials.gov).
Rationale for choice of inclusion and exclusion criteria
We restricted attention to fully-powered two-arm parallel CRTs because different methodological and reporting issues may arise in complex designs (e.g. crossover and stepped wedge) and in secondary or subgroup analyses, and because statistical inference should not the focus of a pilot or feasibility CRT.[1] The year 2017 was selected as we wished to review from a 12-month period of publication and selecting the most recent calendar year was the easiest way to ensure a simple search strategy and to ensure that all reports had appeared within the index at the time of the search (October 2018).
Search terms used
The Cochrane CENTRAL search engine was used to search for any eligible report published (either online or "in print") in the year 2017 in any journal and in any language using the following search criteria: (1) Title/Abstract/Keyword: ((unit? Or school? Or hospital? Or cluster* or region? Or ward* or practice* or communit* or population* or facility or faciliites or practitioner or group) next random*) (2) Title/Abstract/Keyword: (odds* or "odds ratio" or risk or "risk ratio" or "risk difference" or "prevalence ratio" or "prevalence difference" or "relative risk" or nnt or "number needed to treat" or binary or dichot* or proportion or fraction or "absolute difference" or event or probability) (3) Title: (protocol or pilot or feasibility) (4) ((#1) and (#2)) not (#3) The search was last conducted on October 29, 2018.

Reference:
1. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ. 2016;355:i5239.

Table S2. Characteristics of N=82 data abstractors[a]

	Total (N = 82)
Highest career level	
Student (PhD/MSc)	35 (42.7%)
Post Doctoral/Masters level researcher	25 (30.5%)
Lecturer/Assistant Professor	7 (8.5%)
Associate Professor, Professor, Senior Lecturer	14 (17.1%)
Other	1 (1.2%)
Main role	
Methodologist (statistician)	69 (84.1%)
Trialist	5 (6.1%)
Other	8 (9.8%)
Type of work setting[b]	
Healthcare	15 (18.3%)
University	70 (85.4%)
Other	1 (1.2%)
Country of work	
Australia	3 (3.7%)
Canada	2 (2.4%)
United Kingdom	51 (62.2%)
United States	19 (23.2%)
Other	7 (8.5%)
Previous CRT experience	
Yes; one trial	14 (17.1%)
Yes; two trials	13 (15.9%)
Yes, three or more trials	27 (32.9%)
No experience	28 (34.1%)

[a]Reviewer defined as having contributed to *at least one* collaborative data extraction via in-person meeting. Participants who performed only individual extractions or refused participation after randomization were excluded. N=5 reviewers performed multiple reviews.

[b]Categories not mutually exclusive

Table S3. Percent agreement between two independent data abstractors prior to data reconciliation workshop across the 73 papers included in the systematic review

	All Variables (N=95)				Common Variables (N=28)			
Workshop	Variables	Comparisons	Discrepancies	Agreement (%)	Variables	Comparisons	Discrepancies	Agreement (%)
London, UK	95	3990	625	84.3	28	1176	193	83.6
Durham, NC	95	1615	191	88.2	28	476	59	87.6
Birmingham, UK	95	1330	178	86.6	28	392	54	86.2
All Workshops	95	6935	994	85.7	28	2044	306	85.0

"Common variables" are those that were abstracted for every one of the 73 CRTs; "All variables" includes all those that were abstracted from at least one of the 73 articles, including some that were not abstracted for every CRT as they might not have been relevant (e.g. "type of absolute measure" would be relevant to only the 21 CRTs that reported an absolute measure – see **Table S6**).

Table S4. Characteristics of N=73 CRTs in systematic review [a]

	N(%) (N=73)
Disease or domain under study[b]	
Bodily systems	10 (13.7%)
Cancer	6 (8.2%)
General health	9 (12.3%)
Infectious diseases	19 (26.0%)
Mental health and behavioural conditions	11 (15.1%)
Nutritional and metabolic	5 (6.8%)
Respiratory disease	6 (8.2%)
Women's health	16 (21.9%)
Other[c]	23 (31.5%)
Geographic region[b]	
Africa	19 (26.0%)
Asia	14 (19.2%)
Europe	22 (30.1%)
North America	12 (16.4%)
Central America/South America/Caribbean[d]	3 (4.1%)
Oceania	5 (6.8%)
Low- or Middle Income Country (LMIC)[1]	35 (47.9%)
Type of experimental intervention[b]	
Targeted at health care professionals	32 (44.4%)
Targeted at the organisation of health care or health delivery service	24 (33.8%)
Participant health promotion or educational intervention	46 (64.8%)
Direct participant therapeutic intervention	12 (16.9%)
Other	7 (9.7%)
Type of control intervention	
Not reported	1 (1.4%)
Placebo, no active intervention	54 (74.0%)
Minimal application of experimental intervention	8 (11.0%)
Other	10 (13.7%)
Unit of randomisation	
Health facility	30 (41.1%)
Health care provider	11 (15.1%)
School, School district	10 (13.7%)
Geographic areas (e.g. village or county)	14 (19.2%)
Workplace	2 (2.7%)
Household/family	1 (1.4%)
Other	5 (6.8%)
Total number of clusters randomised	
Median (Q1, Q3)	29.0 (20.0, 44.0)
<6	3 (4.1%)
6-10	5 (6.8%)
11-20	17 (23.3%)
21-40	27 (37.0%)
>40	21 (28.8%)
Size of analyzed clusters	
Median (Q1, Q3)	48.0 (20.0, 219.7)
<20	21 (28.8%)
20-49	18 (24.7%)
50-99	10 (13.7%)
100-199	5 (6.8%)
>200	19 (26.0%)
Study design[e]	
Cohort	49 (67.1%)
Cross-sectional	23 (31.5%)
Mix of cohort and cross-sectional	1 (1.4%)

Abbreviations: SD – Standard Deviation; Q1 – Quartile 1 (i.e. 25th percentile); Q3 – Quartile 3 (i.e. 75th percentile); CRT – Cluster Randomised Trial

[a] Reported as n(%), unless otherwise stated, and based on data abstraction from one data abstractor

[b] Categories not mutually exclusive

[c] Including 5 categories (with number , % out of total of 73 studies) of: accidents and injuries (2, 2.7%), genetic disorders (1, 1.4%), mouth and dental (0, 0%), pathological conditions (0, 0%) and symptoms and signs (1, 1.4%)

[d] Of which, all 3 in South America.

[e] Cohort design: such that individuals enrolled at the baseline time point are measured for outcomes at the primary follow-up time point; Cross-sectional design: whereby a sample of individuals is taken at the primary follow-up time-point and these are not individuals who were enrolled at the baseline time point.

[1] LMIC is determined as defined by the Organisation for Economic Co-operation and Development (OECD). Organisation for Economic Co-operation and Development. DAC List of ODA Recipients 2020 [Available from: http://www.oecd.org/dac/financing-sustainable-development/development-finance-standards/daclist.htm.] See also: https://wellcome.org/grant-funding/guidance/low-and-middle-income-countries. Website last accessed March 18, 2021.

Table S5: Additional design and analysis characteristics of N=73 CRTs in systematic review [a]

	N(%)
Design characteristics	
Nature of power/sample size calculation	(N=73)
Accounted for CRT design	51 (69.9%)
Did not account for CRT design	8 (11.0%)
Did not report power/sample size calculation	14 (19.2%)
Type of randomization	(N=73)
Not restricted (e.g., simple randomization, block randomization, etc.)	17 (23.3%)
Restricted randomization*[b]	
Stratification	38 (52.1%)
Pair-matching	11 (15.1%)
Constrained randomization	8 (11.0%)
Unclear/Not reported	4 (5.5%)
Number of follow-up time points	(N=73)
1	49 (67.1%)
2	9 (12.3%)
3+	9 (12.3%)
Ambiguous	6 (8.2%)
Single primary time point explicitly identified[c]	(N=49)
Unclear/Not Reported	5 (20.8%)
Yes, a single time point is explicitly identified as primary.	9 (37.5%)
No, multiple time points are specified as being of equal importance	2 (8.3%)
No, authors did not explicitly identify any specific time point as primary	8 (33.3%)
Method of data collection of primary binary outcome*	(N=73)
Questionnarie or survey	34 (46.6%)
Administrative data	12 (16.4%)
Laboratory data	5 (6.8%)
Electronic health/medical record	22 (30.1%)
Other	12 (16.4%)
Unclear	5 (6.8%)
Analysis characteristics	
P-values reported in manuscript "Table 1" of baseline characteristics	(N=73)
No	39 (53.4%)
No Table 1	3 (4.1%)
Yes	31 (42.5%)
Software used for analysis*	(N=73)
SAS	16 (21.9%)
Stata	24 (32.9%)
R	14 (19.2%)
SPSS	12 (16.4%)
MPlus	2 (2.7%)
Other	4 (5.5%)
Unclear/Not Reported	10 (13.7%)

Abbreviations: SD – Standard Deviation; Q1 – Quartile 1 (i.e. 25th percentile); Q3 – Quartile 3 (i.e. 75th percentile); CRT – Cluster Randomised Trial

[a] Reported as n(%), unless otherwise stated, and based on data abstraction from one data abstractor

[b] Of all those either explicitly identified or of which it is implied by reporting

[c] Of the 24 papers with more than one follow-up time point (or "ambiguous" on number of follow-up time points)

*Categories are not mutually exclusive

Table S6. Reporting of primary binary outcome for N=73 CRTs in systematic review

	Abstract N (%) (N = 73)	Main text N (%) (N = 73)
Binary outcome explicitly identified, n(%)	52 (71.2%)	62 (84.9%)
Outcome reported[a], n(%)		
By study arm	50 (68.5%)	70 (95.9%)
Overall	9 (12.3%)	9 (12.3%)
None reported	13 (17.8%)	1 (1.4%)[b]
Other	4 (5.5%)[c]	1 (1.4%)[d]
Treatment effect measure		
Absolute only	8 (11.0%)	8 (11.0%)
Relative only	46 (63.0%)	46 (63.0%)
Absolute and Relative	4 (5.5%)	13 (17.8%)
No treatment effect (i.e. neither absolute nor relative)	15 (20.5%)	6 (8.2%)
Details[e]	**(N=15)**	**(N=6)**
Reports proportions per arm with p-value	0 (0.0%)	1 (20.0%)
Reports only proportions per arm (with no p-val or stat. sig).	0 (0.0%)	4 (80.0%)
Reports only p-value or stat sig (with no proportions per arm).	10 (66.7%)	1 (20.0%)
Reports neither proportions per arm nor p-value or stat. sig.	5 (33.3%)	0 (0.0%)
Type of absolute measure reported[a, f]	**(N=12)**	**(N=21)**
Difference in prevalence, risk, or proportions (e.g. risk difference)[g]	11 (91.7%)	19 (90.5%)
Number needed to treat (NNT)	2 (16.7%)	3 (14.3%)
Type of relative measure reported[a,h]	**(N=50)**	**(N=59)**
Odds ratio[i]	34 (47.8%)	40 (67.8%)
Risk ratio/Relative risk/Prevalence ratio	17 (34.0%)	19 (32.2%)
Magnitude of binary outcome[j]		**(N=64)**
Rare (risk ≤ 10%[k])		16 (25%)
Common (risk > 10%[k])		48 (75%)
Potential for overstating intervention effect[j]		40 (62.5%)
Rare outcome (risk ≤ 10%[k]) and only relative measure reported		12 (18.8%)
Common outcome (risk > 10%[k]) and odds ratio reported as relative measure		28 (43.8%)

[a]Categories not mutually exclusive

[b]1 article only reports outcome percentages at baseline. Follow-up results are only reported as treatment effects.

[c]1 abstract reported only individual elements of the composite outcome by study arm, 1 abstract reported the primary outcome by study arm but only for sub-groups, 1 abstract reported within-group differences in outcome from baseline to follow-up rather than of the main outcome at follow-up, and 1 abstract reported results only in qualitative terms

[d]1 article only presents outcome stratified by gender in the main text

[e]Of those reporting no treatment effect i.e. neither absolute nor relative

[f]Of those reporting an absolute measure

[g]2 articles reported "difference-in-differences" in the abstract and main text as the between-arm (i.e. intervention vs. control) difference in the within-arm change in proportion from baseline to endline

[h]Of those reporting a relative measure

[i]2 articles reported a ratio of odds ratios (ROR) in the abstract and main text. More specifically, 1 ROR was a comparison between intervention and control arms of the within-arm odds ratio for baseline to endline change, and 1 ROR was the ratio of the between-arm odds ratio (i,e, intervention effect) based on two levels of a post-randomization covariate.

[j] For 64 articles that report both an intervention effect as well as outcome proportions by arm. Note that, of the 73 articles, 3 articles do not report outcome proportions by arm and an additional 6 report no intervention effect.

[k] Rare outcome defined as: risk of the primary binary outcome is ≤ 10% in either the intervention arm or the control arm; Common outcome defined as: risk of the primary binary outcome is > 10% in both the intervention arm and the control arm.

Table S7. Analysis of primary binary outcome for N=73 CRTs in systematic review

	N (%) (N = 73)
Unit of analysis	
Cluster-level only	8 (11.0%)
Individual-level only	62 (84.9%)
Cluster- and individual-level	3 (4.1%)
Accounted for clustering in the analysis[a]	64 (87.7%)
Cluster-level analysis	**(N = 11)**
Main cluster-level summary statistic analyzed[b]	
Proportions	9 (81.8%)
Mean residuals	1 (9.1%)
Other[c]	2 (18.2%)
Method to compare cluster-level summary statistic[b]	
T-test	5 (45.5%)
Z-test	0 (0%)
Wilcoxon Rank Sum test	1 (9.1%)
Permutation test	1 (9.1%)
Other[d]	4 (36.4%)
Individual-level analysis	**(N = 65)**
Main method of analysis[b]	
Regression model accounting for clustering	51 (78.5%)
Statistical test accounting for clustering	5 (7.7%)
Regression model not accounting for clustering[e]	3 (4.6%)
Statistical test not accounting for clustering	6 (9.2%)
Regression method taking clustering into account[f,g]	**(N=51)**
Mixed Effects	31 (60.8%)
Generalized Estimating Equations (GEE)[h]	13 (26.0%)
Cluster Robust Standard Errors	4 (8.0%)
Other[i]	3 (6.0%)

[a] Of the 65 with individual-level analysis, 56 accounted for clustering and all 8 with cluster-level analysis (which implicitly accounts for clustering) accounted for clustering.
[b]Categories not mutually exclusive; [c]Two papers report log cluster-level proportions
[d]All 4 of these used some form of regression analysis: 2 used regression of cluster proportions, 1 used regression of log-cluster proportions and 1 classified each cluster into a binary category based on the level of the cluster proportion and then analyzed that dichotomous variable using logistic regression.
[e] 1 article tested for the presence of clustering, determined it wasn't present, and then did an analysis that ignored the clustering, specifically doing a chi-square test for the primary outcome
[f] Categories are not mutually exclusive; [g] Of those reporting regression model accounting for clustering
[h]Of which 2 (15.4%) reported using exchangeable working correlation matrix, 1 (7.7%) reported using independence and 10 (76.9%) were unclear or did not report the working correlation matrix
[i] 1 article states that GLM was used to correct for clustering, though does not state how. 1 article states that G-side GLIMMIX modeling was used but does not state how clustering was accounted for. 1 article states logistic regression accounted for clustering but does not state how.

Table S8. Journal policy regarding use of CONSORT statement for reporting, for N=73 CRTs in systematic review[a]

Journal Name	CONSORT 2010 Statement		CONSORT 2010 Extension to Cluster Trials		
	Recommended	Required	Recommended	Required	Count
Total Papers					73
Total Journals					48
Plos one	_	√	_	√*	8
The Lancet	_	√	_	√	5
The Lancet Global Health	_	√	_	√	4
Journal of Adolescent Health	_	_	_	_	3
Cancer	_	√	_	√*	3
JAIDS Journal of Acquired Immune Deficiency Syndromes	√	_	_	_	3
BMC public health	√*	_	√*	_	2
Bulletin of the world health organization	√	_	_	_	2
BMC medicine	√	_	√*	_	2
BMC health services research	_	√	_	√*	2
Plos medicine	√	_	√*	_	1
AIDS (london, england)	_	_	_	_	1
British journal of sports medicine	√	_	√*	_	1
Psycho-oncology	√	_	√*	_	1
Medical journal of Australia	_	√	_	√	1
Health expectations	√	_	_	_	1
Family practice	_	√	_	_	1
Thorax	√	_	√*	_	1
Pharmacoepidemiology and drug safety	√	_	√*	_	1
Journal of the International AIDS Society	√	_	√*	_	1
Anaesthesia	_	√	_	√	1
JAMA surgery	√*	_	_	_	1
Journal of Consulting and Clinical Psychology	_	_	_	_	1
Medical care	_	_	_	_	1
Journal of substance abuse treatment	√	_	√*	_	1
New England journal of medicine	√	_	√*	_	1
Translational behavioral medicine	_	_	_	_	1
The American journal of tropical medicine and hygiene	√	_	_	_	1
AJPH	√	_	√	_	1
Global health, science and practice	_	_	_	_	1
Globalization and health	√	_	_	_	1
Annals of internal medicine	√	_	√	_	1
BMJ open	√	_	_	_	1
European journal of public health	_	√	_	√*	1
BJOG	√	_	√*	_	1
Stroke	_	√	_	√	1
JAMA pediatrics	_	_	_	_	1
Journal of youth and adolescence	√	_	√*	_	1
Reproductive health	_	_	_	_	1
International Journal for Quality in Health Care	√	_	√*	_	1
Osteoarthritis and cartilage	_	√	_	√	1
JAMA internal medicine	_	√	_	√	1
The lancet. HIV	_	_	_	_	1
Trials	_	√	_	√	1
Journal of community health	_	_	_	_	1
Archives of women's mental health	_	√	_	√	1
JAMA	_	√*	_	√*	1
Intensive care medicine					

Note: * indicates that recommendation or requirement is implicit because it references either the EQUATOR Network or another repository of reporting guidelines; [a] Classification of journals' policies was conducted post-hoc (i.e. not as part of initial data abstraction). Classification was based on information contained in journals' "Instruction to Authors". Classification for each journal was coded as one of the following five options: (1) Explicitly required (2) Explicitly recommended (3) Implicitly required (4) Implicitly recommended (5) Not

mentioned; where an "implicit" indicates that journal referred to guidelines on the EQUATOR network website or to some4 other respository of reporting guidelines.

Methods and Results for Table S8: In an effort to better understand reporting omissions identified through analysis of data from the 73 CRTs in the review, we performed a post-hoc analysis in which we investigated whether CONSORT was endorsed by the 48 journals in which those CRTs were published. To do so, we (KT and AP) reviewed the instructions to authors of each of the 48 journals to determine whether there was mention of the 2010 CONSORT statement on reporting of RCTs,[1] of the 2010 CONSORT extension to CRTs,[2] or instruction to use a relevant reporting checklist from the EQUATOR network (which itself lists the 2010 CRT extension).[3] This investigation showed that most (58, 79.5%) of the 73 CRTs appeared in journals that endorsed (either required or recommended) the 2010 CONSORT statement on reporting of RCTs, which itself recommends the joint use of relative and absolute measures when reporting results for binary outcomes. Some journals including The Lancet Global Health, are quite explicit in their preferences and specifically prescribe the use of both relative and absolute measures and state "absolute differences are more useful than relative ones". This sort of explicit direction to authors might be of benefit.

References:
1. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.
2. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012;345:e5661.
3. EQUATOR Network. Enhancing the QUAlity and Transparency Of health Research [Available from: https://www.equator-network.org/.

Table S9. Cross-tabulation of journal policy and evidence of author use of CONSORT statement for reporting, for N=73 CRTs in systematic review

		Journal policy regarding use of CONSORT statement for reporting[a,b]		
Evidence of author use of CONSORT statement for reporting [a,c]	Overall (N=73)	Requires CONSORT (N = 32, 43.8%[d])	Recommends CONSORT (N = 26, 35.6%[d])	Does not mention CONSORT (N = 15, 20.5%[d])
No evidence	3 (4.1%)	1 (3.1%)	0 (0.0%)	2 (13.3%)
Some evidence	70 (95.9%)	31 (96.9%)	26 (100%)	13 (86.7%)
Nature of evidence				
Author(s) included CONSORT flow diagram only	43 (58.9%)	20 (62.5%)	18 (69.2%)	5 (33.3%)
Author(s) reported using CONSORT RCT reporting guidelines alone[e1]	9 (12.3%)	7 (21.9%)	1 (3.8%)	1 (6.7%)
Author(s) reported using CONSORT extension for CRTs[2]	18 (24.7%)	4 (12.5%)	7 (26.9%)	7 (46.7%)

Abbreviations: CRT – Cluster Randomised Trial; RCT – Randomised Controlled Trial; CONSORT - Consolidated Standards of Reporting Trials

[a] Reported as n(%) and based on data abstraction from Duke team after initial data set compiled; [b] Either the CONSORT statement for reporting of standard parallel-arm randomised controlled trials or the CONSORT extension for reporting of CRTs. These data were determined using each journal's "instructions to authors" & relevant sections of the journal website (see details by journal in Table S5); [c] Based on data reported only in the manuscript text, tables, figures and and supplementary materials and not using information from journal "instructions to authors". Therefore, for journals that "require CONSORT" but for which there was no explicit mention within the manuscript of the use of one of the CONSORT statements for reporting of trials, such an article was classified here as "no evidence" for use of CONSORT; [d] Row percentages out of N=73; [e] i.e. did not report using the CONSORT extension for CRTs.[1]

Methods and Results for Table S9: In further post-hoc investigation , we revisited each of the 73 CRT articles to investigate whether there was reference to some form of the CONSORT statement within the manuscript itself (irrespective of whether it was included in the journal's "Instructions to Authors"). Almost all (95.9%) indicated use of the statement in some form, primarily through reference to the "CONSORT flow-chart" alone with no explicit mention of any CONSORT statement on reporting, noting however, that authors may not mention CONSORT if its use is recommended or required by the journal. Nevertheless, nearly all of the 15 manuscripts appearing in journals for which CONSORT was not mentioned by the journal showed evidence of using some form of the CONSORT guidelines**.** Thus, overall, even though it is evident that authors and journals are familiar with CONSORT, this does not lead to authors reporting both absolute and relative measures of effect for binary outcomes. Again, explicit direction to authors might be of benefit.

References:

1. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.
2. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012;345:e5661.

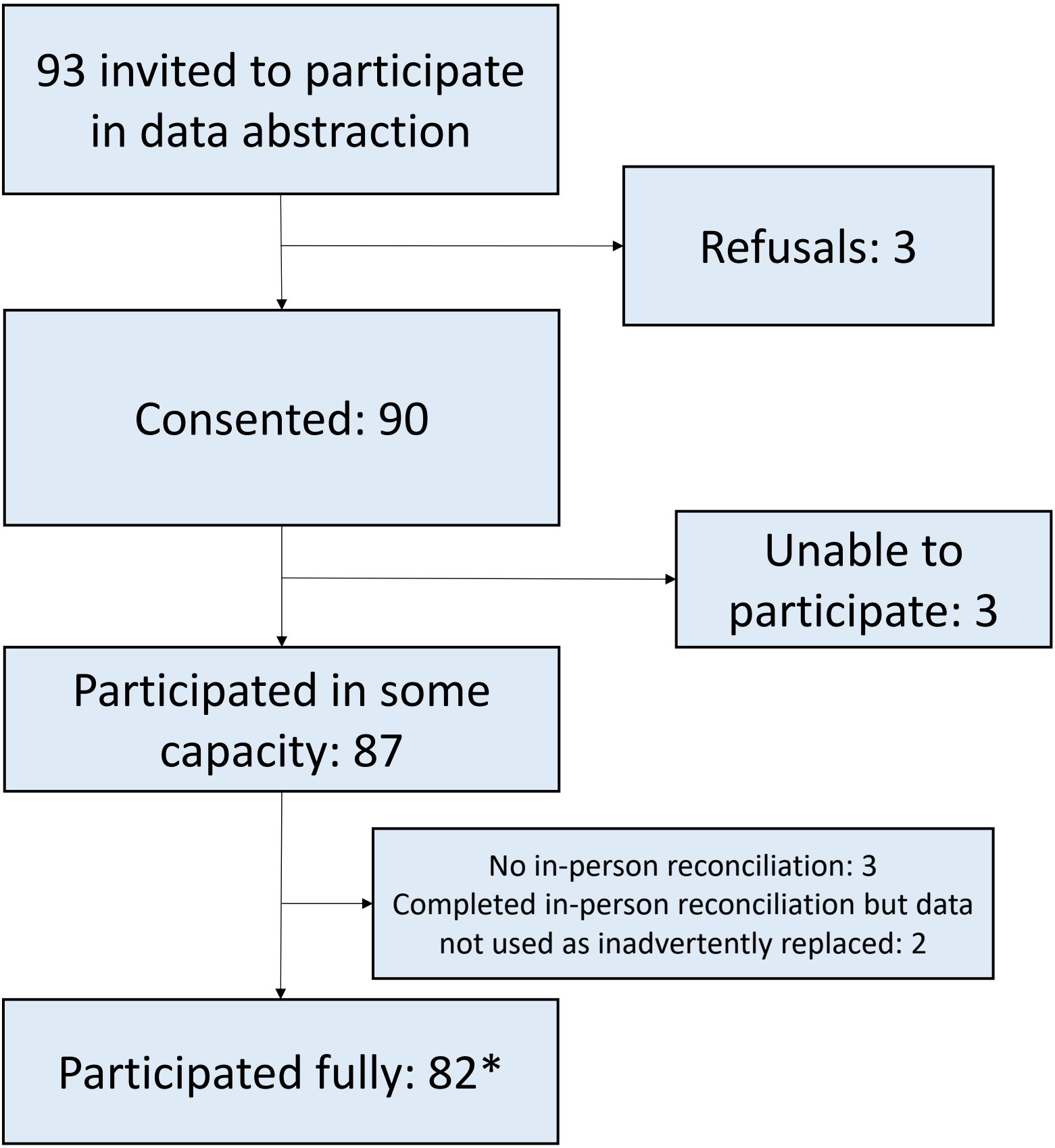
Table S10. List of LMICs*

Afghanistan
Albania
Algeria
Angola
Antigua and Barbuda
Argentina
Armenia
Azerbaijan
Bangladesh
Belarus
Belize
Benin
Bhutan
Bolivia
Bosnia and Herzegovina
Botswana
Brazil
Burkina Faso
Burundi
Cabo Verde
Cambodia
Cameroon
Central African Republic
Chad
China (People's Republic of)
Colombia
Comoros
Democratic Republic of Congo
Congo
Costa Rica
Côte d'Ivoire
Cuba
Djibouti
Dominica
Dominican Republic
Ecuador
Egypt
El Salvador
Equatorial Guinea
Eritrea
Eswatini
Ethiopia
Fiji
Gabon
Gambia
Georgia
Ghana
Grenada
Guatemala
Guinea
Guinea-Bissau
Guyana
Haiti
Honduras
India
Indonesia
Iran
Iraq
Jamaica
Jordan
Kazakhstan
Kenya
Kiribati
Democratic People's Republic of Korea
Kosovo
Kyrgyzstan
Lao People's Democratic Republic
Lebanon
Lesotho
Liberia
Libya
North Macedonia
Madagascar
Malawi
Malaysia
Maldives
Mali
Marshall Islands
Mauritania
Mauritius
Mexico
Micronesia
Moldova
Mongolia
Montenegro
Montserrat
Morocco
Mozambique
Myanmar
Namibia
Nauru
Nepal
Nicaragua
Niger
Nigeria
Niue
Pakistan
Palau
Panama
Papua New Guinea
Paraguay
Peru
Philippines
Rwanda
Saint Helena
Samoa
São Tomé and Príncipe
Senegal
Serbia
Sierra Leone
Solomon Islands
Somalia
South Africa
South Sudan
Sri Lanka
Saint Lucia
Saint Vincent and the Grenadines
Sudan
Suriname
Syrian Arab Republic
Tajikistan
Tanzania
Thailand
Timor-Leste
Togo
Tokelau
Tonga
Tunisia
Turkey
Turkmenistan
Tuvalu
Uganda
Ukraine
Uzbekistan
Vanuatu
Venezuela
Vietnam
Wallis and Futuna
West Bank and Gaza Strip
Yemen
Zambia
Zimbabwe

Abbreviation: LMIC – Low- and middle-income countries. *Defined by the Organisation for Economic Co-operation and Development (OECD).
List downloaded from: https://wellcome.org/grant-funding/guidance/low-and-middle-income-countries on 18 March 2021

Appendix C: Supplementary Figures

Figure S1: Flow-chart of participation of data abstractors



*Includes 1 paper for which 1 person of a pair at the data reconciliation meeting had not completed the pre-meeting data abstraction.

Figure S2: Schematic of process of data abstraction for each member of pair

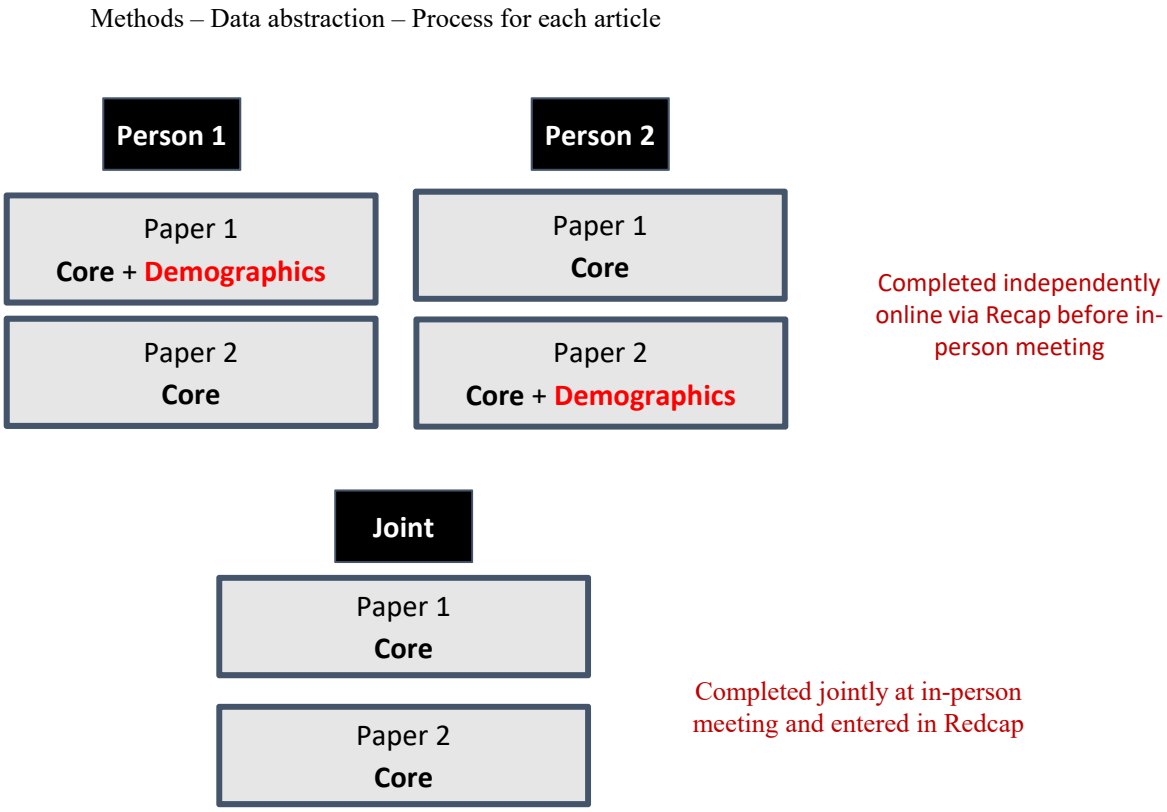
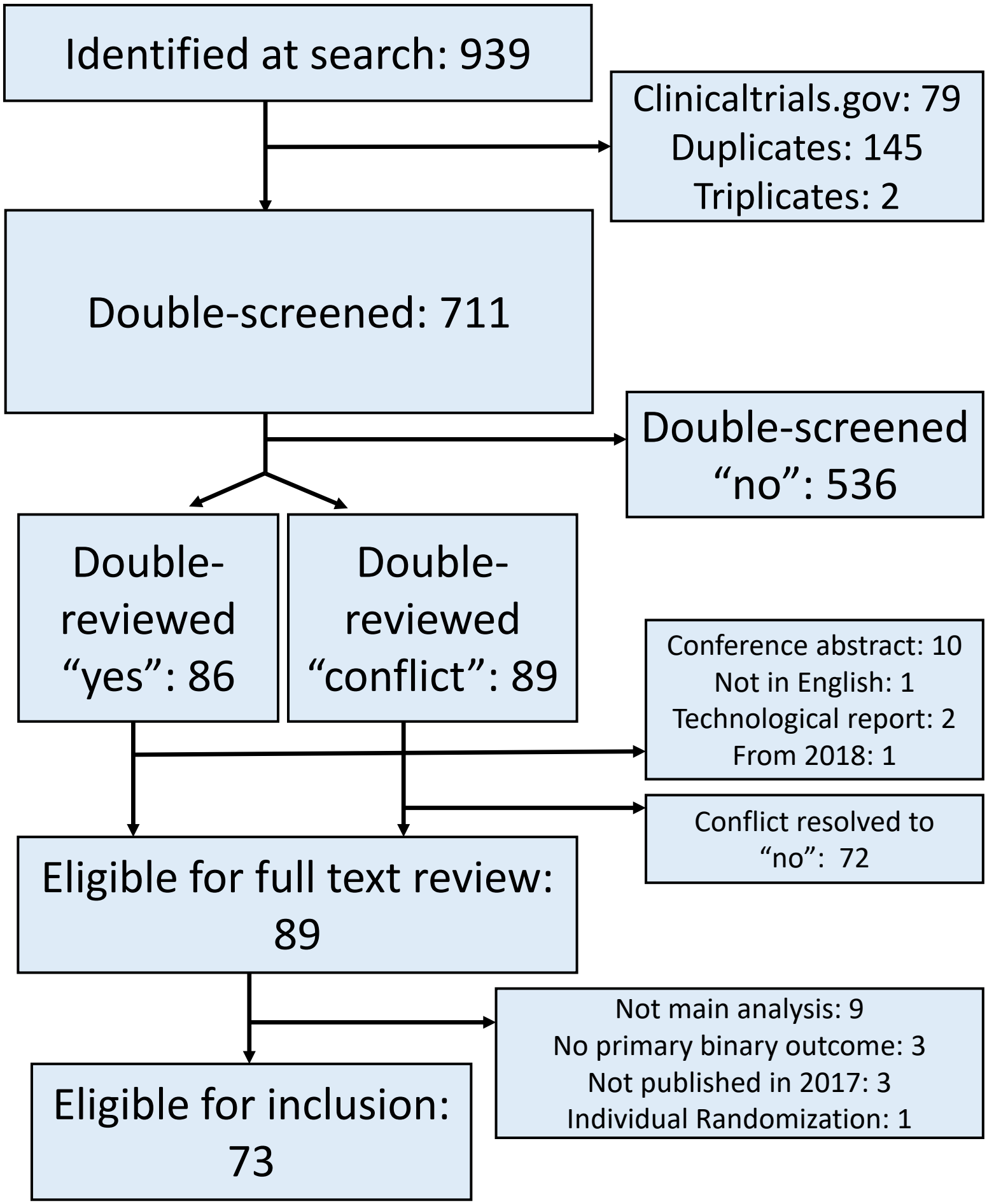


Figure S3: Flow-chart of CRTs included in systematic review (N=73)



Appendix D: Other Supplementary Materials

Supplementary Material 1. Protocol for systematic review.

Title	**Rapid Review of Reporting of CRTs with Binary Outcomes**
CRU/Department/Division/Center	Duke Global Health Institute
IRB Number	Exempt
Investigators:	Elizabeth Turner
Lead Investigator	Elizabeth Turner
Mentors	
Co-authors	Liz Turner (corresponding), Alyssa Platt, John Gallis, Kaitlin Tetreault, Jo McKenzie, Stephen Nash, Andrew Forbes, Karla Hemming
Biostatistician(s)	Liz Turner, Alyssa Platt, John Gallis
Supervising Biostatistician	Liz Turner
Original Creation Date	March 8, 2019
Project Folder Location	~\PROJECT_DGHI_RDAC\Core Consultations\Faculty\Liz Turner\cRCT Binary Outcome Rapid Review
Project Goal(s)	Manuscript
Submission Deadline(s)	July 30, 2019
Investigator Agreement	☒ All statistical analyses included in an abstract or manuscript should reflect the work of the biostatistician(s) listed on this SAP. No changes or additional analyses should be made to the results or findings without discussing with the project biostatistician(s). ☒ All biostatisticians on this SAP should be given sufficient time to review the full presentation, abstract, manuscript, or grant and be included as co-authors on any abstract or manuscript resulting from the analyses. ☒I have reviewed the SAP and understand that any changes must be documented.
Activity Log	1. Updated March 8, 2019 2. Updated April 4, 2019 3. April 10, 2019 – Adding information on reliability piece; adding descriptions under the eligibility and study design headings. 4. May 28, 2019 – Adding information about data extraction process (Alyssa) 5. August 7, 2019 – Adding proposed analysis for agreement calculation 6. August 8, 2019 – Adding description of additional variables extracted from articles (CONSORT, trials registration) + info on CONSORT from journal instructions to authors 7. October 4, 2019 – Adding information to the flow chart as well as shell tables in the addendum

1 Study Overview

Background/Introduction: The CONSORT extension for reporting of results of cluster-randomized trials (CRTs) recommends that binary outcomes be reported as both a relative and absolute effect and indeed some leading journals indicate a preference for absolute effects (e.g. Lancet instructions to authors states: “absolute differences are more useful than relative ones”). Nevertheless, informal reviews indicate that most CRT results for binary outcome measures report only a relative effect. Moreover, the relative effect is typically an odds ratio, which has the potential to be misinterpreted as a risk ratio. This is problematic when the outcome of interest is common.

This study aims to make an assessment of recent reporting practices for CRTs using an innovative rapid review method to quickly and accurate collect needed information for qualifying CRTs with a primary binary outcome.

Study Aims

1. Describe reporting and interpretation practices for published studies of binary outcomes in CRTs.
2. Describe the extent of agreement between data extraction by reviewers of the same paper(s)

1.1 Study Hypotheses

There are no formal hypotheses for this study, it is purely descriptive.

2 Study Description

2.1 CRT Inclusion Criteria

• Two-arm parallel CRT
• At least one binary primary outcome
• Full trial (i.e., not pilot/feasibility)
• Main analysis (i.e., not secondary or subgroup 2017

2.2 CRT Exclusion Criteria

• Pilot/feasibility trial
• Secondary/subgroup analysis paper
• Protocol/study design paper
• Methods paper
• A report in conference proceedings
• Any trial design other than a two-arm parallel CRT (i.e. stepped wedge, crossover and factorial designs)

2.3 Data Acquisition

Abstract Identification

Cochrane CENTRAL search engine was used to search for any publication in the year 2017 in any journal using the following search criteria:

(5) Title/Abstract/Keyword: ((unit? Or school? Or hospital? Or cluster* or region? Or ward* or practice* or communit* or population* or facility or faciliites or practitioner or group) next random*)
(6) Title/Abstract/Keyword: (odds* or “odds ratio” or risk or “risk ratio” or “risk difference” or “prevalence ratio” or “prevalence difference” or “relative risk” or nnt or “number needed to treat” or binary or dichot* or proportion or fraction or “absolute difference” or event or probability)
(7) Title: (protocol or pilot or feasibility)
(8) ((#1) and (#2)) not (#3)

From these search terms **939** abstracts were identified and exported to Rayyan software where duplicates and references that only existed in ClinicalTrials.gov were excluded. All 711 remaining abstracts were double reviewed by named authors.

Manuscript Assignment

The final 89 abstracts were organized in a list and each assigned a random number between 0 and 1 generated using the RAND function in MS Excel. The list of studies was then and sorted in ascending order by the random generated number to determine their order for review. Each participating reviewer was assigned to review 2 papers in duplicate with a randomly assigned partner. For each workshop, a final count of reviewers was first established and this same number of papers was selected from the sorted list (starting with the lowest number) to be used in the workshop. The sequential first half of the selected workshop papers were chosen to be Paper #1 for review and the second half were selected as Paper #2 for review. Paper #2 was matched up with Paper #1 in the order in which they were listed (i.e. the first paper in the first list was matched with the first paper in the second list). The list of pairs of papers was then duplicated and lined up with the list of participants. Finally, participants were assigned a random number in the same fashion as was used to assign random numbers to papers and sorted in that order to line up with the listed paper pairs.

The Duke workshop included participants with varying levels of statistical experience and expertise with cluster randomized trials and therefore a stratum was used for paper/partner assignment. Each participant was labelled as “experienced” or “novice” and random numbers were assigned within strata. The person assigned lowest number in the “experienced” group was assigned to the person assigned the lowest number in the “novice” group to establish partnerships.

Electronic data abstraction forms were administered using REDCap survey software. Each participant was sent an individualized link to the survey form revealing their assigned papers and survey questions. Participants were to abstract data independently for their two assigned papers before meeting in person to reconcile their responses to each question.
Questions describing the demographics of each study were not double reviewed. We randomly selected one partner in a pair to extract demographic characteristics for the first paper while the other partner extracted demographic characteristics for the second paper.

After the workshops had taken place, the study team identified some important key variables that had not been included in the original data abstraction form. Specifically, given the role of the CONSORT statement in supporting researchers in reporting of findings from CRTs, it was important to know if the article indicated (either explicitly, or by the fact that information was included in a supplement) that the CONSORT extension statement on the reporting of CRTs was used in developing the article. Therefore, the Duke team re-reviewed all included articles to determine this information. A single person (KT) performed the full data abstraction, which was reviewed by LT. Additional variables extracted at this time included: whether p-values were used in "Table 1", was the trial registered on an open access registry and was there a protocol paper published? Similarly, for all journals in which articles were published, the instructions to authors were reviewed (by KT, and verified by AP) to determine whether journal policy stated that CONSORT (the extension statement or the 2010 standard RCT statement) should be used in reporting the results of the CRT. These details are important for understanding the context of the reporting of the research.

Description of Workshops
Three rapid review sessions were held to facilitate the reconciliation of the data extracted from the assigned manuscripts:

(1) *Current Developments in Cluster Randomised Trials and Stepped Wedge Designs Meeting,* Royal London Hospital, London UK, November 2018
(2) Duke University School of Medicine BERD Core Workshop, Durham, NC USA, February 2019
(3) University of Birmingham – Birmingham Clinical Trials Unit (BCTU), Birmingham, UK April 2019

Of the final **89** papers **7** were found to be ineligible for the following reasons:

- 1 protocol paper
- 1 individual randomization
- were secondary/sub-analyses
- 1 was interim/observational
- 1 had no binary outcome
- 1 paper was a duplicate not identified previously

Questions about the REDCap form and data collection process can be directed to Alyssa Platt (alyssa.platt@duke.edu).
Final data were downloaded on May 24, 2019
Raw and derived datasets are stored in the following location:
~\Box\PROJECT_DGHI_RDAC\Core Consultations\Faculty\Liz Turner\cRCT Binary Outcome Rapid Review\4 - Analysis\Data

2.4 Outcomes, Exposures, and Additional Variables of InterestPrimary Outcome(s)

Description	Variables and Source	Specifications
ABSTRACT		
Is binary outcome explicitly identified in the abstract?	**ab_binout_primary**	**1, Yes** 0, No -99, Unclear/Don't Know
How does the abstract report the prevalence, risk or proportion of the binary outcome at a follow-up time point?	**ab_binout_prev** • ab_binout_prev___1 • ab_binout_prev___2 • ab_binout_prev___99	1, Overall 2, By study arm 99, Other (Please Specify)
Does the abstract report a treatment effect for the results of analysis for the binary outcome as:	**ab_result_type** • ab_result_type___1 • ab_result_type___2 • ab_result_type___99 • ab_result_type____99	1, Absolute measure such as difference in proportions, prevalence difference or risk difference 2, Relative measure such as odds ratios or risk ratios 99, Other (Please Specify) -99, Does not quantify results for binary outcome or it is unclear from the abstract what the form of the primary outcome of analysis is.
What type of ABSOLUTE measure is reported?	**ab_binout_abtype** • ab_binout_abtype___1 • ab_binout_abtype___2 • ab_binout_abtype___99	1, Difference in prevalence, risk, or proportions (e.g. risk difference) 2, Number needed to treat (NNT) 99, Other (Specify below)
What type of RELATIVE measure is reported?	**ab_binout_reltype** • ab_binout_reltype___1 • ab_binout_reltype___2 • ab_binout_reltype___99	1, Odds ratio 2, Risk ratio/Relative risk/Prevalence ratio 99, Other (Specify below)
MAIN TEXT		
Does the manuscript identify the binary outcome as primary?	**mn_binout_primary**	**1, Yes** 0, No
Does the results section (including tables or figures) report the prevalence, risk or proportion of the binary outcome at follow-up:	**mn_rep_prev** • mn_rep_prev___1 • mn_rep_prev___2 • mn_rep_prev____99	1, Overall 2, By study arm -99, Unclear/Not Reported
Does the results section report a treatment effect for the results of analysis for the binary outcome as:	**mn_result_type** • mn_result_type___1 • mn_result_type___2 • mn_result_type___99 • mn_result_type____99	1, Absolute measure such as difference in proportions, prevalence difference or risk difference 2, Relative measure such as odds ratios or risk ratios 99, Other (Please Specify) -99, Does not quantify results for binary outcome or it is unclear from the results section what the form of the primary outcome of analysis is.
What type of ABSOLUTE measure is reported?	**mn_binout_abtype** • mn_binout_abtype___1 • mn_binout_abtype___2 • mn_binout_abtype___99	1, Difference in prevalence, risk, or proportions (e.g. risk difference) 2, Number needed to treat (NNT) 99, Other measure
What type of RELATIVE measure is reported?	**mn_binout_reltype**	1, Odds ratio 2, Risk ratio/Relative risk/Prevalence ratio

	• mn_binout_reltype___1 • mn_binout_reltype___2 • mn_binout_reltype___99	99, Other measure

2.5 Additional Variables of Interest

Description	**Variables and Source**	**Specifications**
DESCRIPTIVE STATISTICS		
Disease or domain under study:	mn_disease_domn 1, Accidents and wounds 2, Blood and immune system 3, Cancer 4, Cardiovascular 5, Central nervous system/musculoskeletal 6, Digestive/endocrine 7, Ear and nose 8, Eye 9, General health 10, Genetic disorders 11, Gynaecology 12, Infectious diseases 13, Injuries 14, Mental health and behavioural conditions 15, Mouth and dental 16, Nutritional and metabolic 17, Pathological conditions 18, Pregnancy and birth 19, Respiratory disease 20, Skin 21, Symptoms and signs 22, Urogenital 99, Other (Please Specify)	1,13 Accidents and injuries 11, 18 Women's health 2, 4, 5, 6, 7, 8, 20, 22 Bodily systems 3, Cancer 9, General health 10, Genetic disorders 12, Infectious diseases 14, Mental health and behavioural conditions 15, Mouth and dental 16, Nutritional and metabolic 17, Pathological conditions 19, Respiratory disease 21, Symptoms and signs 99, Other (Please Specify)

Geographic Region	• mn_country1 • mn_country2 • mn_country3 • mn_country4 • mn_country5 • mn_country6 • mn_country7 • mn_country8 • mn_country9 • mn_country10 • mn_country11 • mn_country12 • mn_country13 • mn_country14 • mn_country15	Africa Asia Central America Eastern Europe European Union Middle East North America Oceania South America Caribbean
There are many types of experimental interventions that can evaluated in a CRT. These may include but are not limited to the following: • Educational/quality improvement interventions targeted at health care professionals • Quality improvement interventions targeted at the organisation of health care or health delivery service • Participant health promotion or educational intervention • Direct participant therapeutic intervention • Other, specify	• mn_int_qi_hcp • mn_int_qi_org • mn_int_hlth_promo • mn_int_thrp • mn_int_other	**1, Yes** 2, No 99, Unclear (Code as missing)
Please select type of CONTROL intervention	mn_cont_type 1, Not reported 2, No active intervention, i.e. usual care 3, Minimal application for experimental intervention 4, Placebo intervention 5, Other active intervention 99, Other (Please Specify)	1, Not reported 2, 4, No active intervention, i.e. usual care or placebo 3, Minimal application for experimental intervention 5, Other active intervention 99, Other (Please Specify)
What is the cluster (i.e. unit of randomization)?	mn_clstr_type 1, Hospital 2, Other health facility 3, Individual School 4, School district 5, Geographic areas (e.g. village or county) 6, General Practitioner/Primary Care Provider/Health specialist 7, Nursing home/aged care facility 8, Workplace 9, Household/family 99, Other (Please Specify)	1,2,7, Health Facilities 3,4, Schools and school districts 5, Geographic areas (e.g. village or county) 6, General Practitioner/Primary Care Provider/Health specialist 8, Workplace 9, Household/family 99, Other (Please Specify)
Was some form of restricted randomization procedure used? Please indicate all that apply:	mn_restrand	1, Stratification 2, Pair-matching 3, Constrained randomization 99, Other (Please Specify) If none are selected, code as "simple randomization"
Total number of clusters randomized:	mn_num_clstrs	Use as continuous (Mean, SD) Code as categories:

		• < 6 • 6-10 • 11-20 • 21-40 • >40
What is the average size of analyzed clusters?	mn_clstr_m	
In order to evaluate the effect of the intervention on the binary outcome, did the CRT use data from a cohort of the same individuals who were followed-up over time, take a cross-sectional sample of individuals at each follow-up time point or a mix of both?	mn_coll_type	1, Cohort 2, Cross-sectional 3, A mix of cohort and cross-sectional -99, Unclear/Not Reported
Was power or sample size calculation reported that accounted for the CRT design?	mn_power_rptd	1, Yes 2, No, sample size calculation was reported but it did not account for CRT design 3, No, sample size calculation was NOT reported -99, Unclear/Not Reported
METHODS USED TO ANALYZE PRIMARY BINARY OUTCOME		
What was the unit of analysis of the primary binary outcome?	IF mn_result_type___1 == 1 mn_analysis_unit	1, Cluster level 2, Individual level analysis
What was the unit of analysis of the primary binary outcome?	IF mn_result_type___2 == 1 mn_analysis_unit	1, Cluster level 2, Individual level analysis
Journal name	journal_name	
P-values are used to compare groups in baseline table	pvaluesareusedtocomparegro	Yes No
Is the trial registered?	Isthetrialregistered	Yes No
Is there a published, peer-reviewed protocol?	Isthereapublishedpeerrevie	Yes No
Is the protocol accessible in a non-peered-reviewed format?	Istheprotocolaccessibleina	Yes No
The study was reportedly approved by an ethics committee	Thestudywasreportedlyapprove	Yes No
Are the reporting guidelines in the author guidelines?	Arethereportingguidelinesin	Yes No
Journal statement regarding CRT reporting using CONSORT recommendations	consort_3categ	Requires CONSORT Recommends CONSORT No mention of CONSORT
Reporting Practices	consort_use_3categ	No apparent use of CONSORT Use CONSORT only for flow diagram Evidence of using additional CONSORT reporting guidelines Evidence of using CONSORT extension for cluster trials

3 Statistical Analysis Plan

Aim 1. Analyses will be purely descriptive. Categorical variables will be expressed with counts and percentages, noting when categories are not mutually exclusive (as is the case with several primary variables). Continuous variables will be summarized with means and standard deviations when distributions are relatively normal and will be presented with medians and interquartile ranges (IQR) in the case of skewed distributions. Where meaningful, continuous variables may also be expressed in categorical ranges to better illustrate important cut-points in the continuous scale.

For categorical variables with greater than 5 levels, we will consider collapsing into broader categories.

Data will also be displayed visually, such as with stacked bar charts and other methods that give the reader a better sense of the relative sizes of various categories across multiple strata.

Aim 2..For each of the double-entered variables, the percent agreement will be calculated using the Stata function cfout[1]. Prior to the final run of percent concordance, minimal cleaning will be conducted to ensure that only meaningful differences will show up as such (e.g. differences in decimal places will be rounded, text answers that are qualitatively the same will be corrected as such).

We will compute an overall agreement percentage followed by agreement percentages stratified by workshop (to examine any substantial differences between agreement between workshop formats – which had varying levels of experience and expertise with CRTs). We will consider any difference greater than 5% in total agreement to be substantial.

We will also compute agreement statistics on a subset of questions that would have been answered by ALL reviewers, regardless of study assignment.

4 Appendices

4.1 Shell Tables

4.2 Table 1. Characteristics of Included Studies[a]

	N(%) (N=XX) Mean(SD)
Disease or domain under study[b], n(%)	
Accidents and injuries	
Bodily systems	
Cancer	
General health	
Genetic disorders	
Infectious diseases	
Mental health and behavioural conditions	
Mouth and dental	
Nutritional and metabolic	
Pathological conditions	
Respiratory disease	
Symptoms and signs	
Women's health	
Other	

[1] Authors: Ryan Knight, Matthew White
For questions or suggestions, submit a GitHub issue or e-mail researchsupport@poverty-action.org.

Geographic Region[b], n(%)
- Africa
- Asia
- Eastern Europe
- European Union
- Middle East
- North America
- Central America
- South America
- Caribbean
- Oceania

Type of Experimental Intervention[b], n(%)
- Educational/quality improvement interventions targeted at health care professionals
  - Yes
  - No
  - Unclear
- Quality improvement interventions targeted at the organisation of health care or health delivery service
  - Yes
  - No
  - Unclear
- Participant health promotion or educational intervention
  - Yes
  - No
  - Unclear
- Direct participant therapeutic intervention
  - Yes
  - No
  - Unclear
- Other
  - Yes
  - No
  - Unclear

Type of Control Intervention, n(%)
- Not reported
- Placebo, no active intervention
- Minimal application for experimental intervention
- Other

Unit of Randomization, n(%)
- Health facility
- School, School district
- Geographic areas (e.g. village or county)
- Health care provider
- Workplace
- Household/family
- Other

Type of Randomization[b]
- Simple
- Restricted randomization
  - Stratification
  - Pair-matching
  - Constrained randomization
- Other
- Unclear/Not reported

Total Number of Clusters Randomized, n(%)
- Mean (SD)

<6
6-10
11-20
21-40
>40

Average Size of Analyzed Clusters, Mean(SD)

Study Design, n(%)

Cohort
Cross-sectional
A mix of cohort, cross-sectional
Unclear, not reported

Power/Sample Size Accounting for CRT Design, n(%)

Yes
No, sample size calculation was reported but it did not account for CRT design
No, sample size calculation was NOT reported
Unclear/Not Reported

Journal statement regarding CRT reporting using CONSORT recommendations, n(%)

Requires CONSORT
Recommends CONSORT
Does not mention CONSORT

[a]Based on data abstraction from one reviewer

[b]Categories not mutually exclusive

4.3 Table 2. Reporting of Binary Outcomes

	Abstract N (%) (N = XX)	Main text N (%) (N = XX)
Binary outcome explicitly identified, n(%)		
Outcome reported[a], n(%)		
By study arm		
Overall		
Unclear, not reported		
Treatment effect measure		
Absolute (e.g. difference in proportions, prevalence difference or risk difference)		
Absolute WITHOUT relative measure		
Relative: measure such as odds ratios or risk ratios		
Relative WITHOUT absolute measure		
Absolute and Relative		
Other		
Does not quantify or is unclear		
What type of ABSOLUTE measure is reported[a, b]?		
Difference in prevalence, risk, or proportions (e.g. risk difference)		
Number needed to treat (NNT)		
Other measure		
What type of RELATIVE measure is reported[a,c]?		
Odds ratio		
Risk ratio/Relative risk/Prevalence ratio		
Other measure		

[a]Categories not mutually exclusive

[b]Of those reporting an absolute measure

[c] Of those reporting a relative measure

4.4 Table 3a: Reporting of Unit of Analysis for Primary Outcomes

	N (%) (N = XX)
Unit of analysis	
Cluster-level only	
Individual-level only	
Cluster- and Individual-level	
Neither	

4.5 Table 3b: Reporting Cluster-Level Analysis

	N (%) (N = xx)	N (%) (N = XX)
Main method of analysis[a]		
Comparison of Proportions		
Comparison of Mean residuals		
Comparison of other summary measure		
Other		
Method to compare summary statistics[a]		
T-test		
Z-test		
Wilcoxon Rank Sum test		
Permutation test		
Other		
Accounted for method of restricted randomization used		
Yes		
No		
Unclear		

[a]Categories not mutually exclusive

4.6 Table 3c: Individual-level Analysis

	N (%) (N = xx)
Main method of analysis[a]	
Regression model accounting for clustering	
Statistical test accounting for clustering	
Regression model WITHOUT accounting for clustering	
Statistical test WITHOUT accounting for clustering	
Other	
Method taking clustering into account[a,b]	
Mixed Effects	
Random slopes were used to account for repeated measurements on clusters or individuals over time[e]	

Yes

No

Method section explicitly states random intercepts were included for cluster[c]

Yes

No

Generalized Estimating Equations (GEE)

Type of working correlation matrix[d]

Exchangeable

Independent

Other

Unclear/Not reported

Restricted randomization used

Accounted for method of restricted randomization used

Yes

No

Unclear/Not reported

[a]Categories are not mutually exclusive
[b]Of those reporting model accounting for clustering
[c]Of those reporting mixed effects modeling
[d]Of those reporting GEE modeling

4.7 Table 3d: Outcome distribution and link function

	Absolute	Relative
	N (%) (N = xx)	N (%) (N = xx)
Outcome distribution and link function used to estimate the intervention effect		
Binomial Identity		
Binomial Logit		
Binomial Other		
Gaussian/Normal Identity		
Poisson Log		
Unclear/Not reported Log		
Unclear/Not reported Logit		
Unclear/Not reported Other		
Unclear/Not reported Unclear/Not reported		
Effects were obtained via transformation		
Yes		
No		
Unclear / Not reported		

Table 4. Cross-tabulation of reporting practices versus Journal Requirements of CONSORT guidelines

	Requires CONSORT	Recommends CONSORT	No Mention
No apparent use of CONSORT	N(%)	N(%)	N(%)
Use CONSORT only for flow diagram	N(%)	N(%)	N(%)
Evidence of using additional CONSORT reporting guidelines[1]	N(%)	N(%)	N(%)
Evidence of use of CONSORT extension for cluster trials	N(%)	N(%)	N(%)

[1]n=xx include RCT or cRCT checklist, n=XX state explicitly

Table 5. Characteristics of Reviewers

	N (%) (N=XX)
Highest Career Level	
Student (PhD/MSc)	
Post-Doctoral/Masters level researcher	
Lecturer/Assistant Professor	
Associate Professor, Professor, Senior Lecturer	
Other	
Main Role	
Methodologist (statistician)	
Trialist	
Other	
Type of Work Setting[a]	
Healthcare	
University	
Other (Please Specify)	
Country of work	
Australia	
Canada	
Ireland	
United Kingdom	
United States	
Other	
Previous CRT Experience	
1 trial	
2 trials	
3+ trials	
No experience	
Review Location	
London, UK	
Birmingham, UK	
Durham, NC USA	

[a]Categories not mutually exclusive

4.8 Table 5. Percent Agreement between reviewers pre-collaborative meeting

	Percent Agreement	
Workshop	All questions	Common Questions
London, UK		
Durham, NC USA		
Birmingham, UK		
Overall		

4.9 Survey questions with indicator for double vs single review

See Supplementary Material 2.

4.10 Additional survey questions extracted by Duke team in duplicate

See Supplementary Material 3

Supplementary Material 2. Data extraction form used by two independent data abstractors before data reconciliation workshop and by pair of data abstractors during data reconciliation workshop, with indication as to whether data abstraction was by one or both members of the pair.

Field Label	Choices, Calculations, OR Slider Labels	Double Abstracted	Single Abstracted	Asked of All Reviewers
What is the primary binary outcome identified in the abstract?		X		X
Is the binary outcome explicitly identified as primary in the abstract?	1, Yes 0, No -99, Unclear/Don't Know	X		X
How does the abstract report the prevalence, risk or proportion of the binary outcome at a follow-up time point?	1, Overall 2, By study arm 99, Other (Please Specify)	X		X
Does the abstract report a treatment effect for the results of analysis for the binary outcome as: (CHECK ALL THAT APPLY. If none apply then leave unchecked):	1, Absolute measure such as difference in proportions, prevalence difference or risk difference 2, Relative measure such as odds ratios or risk ratios 99, Other (Please Specify) -99, Does not quantify results for binary outcome or it is unclear from the abstract what the form of the primary outcome of analysis is.	X		X
What type of ABSOLUTE measure is reported? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, Difference in prevalence, risk, or proportions (e.g. risk difference) 2, Number needed to treat (NNT) 99, Other (Specify below)	X		
What type of RELATIVE measure is reported? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, Odds ratio 2, Risk ratio/Relative risk/Prevalence ratio 99, Other (Specify below)	X		
Does the manuscript identify the binary outcome as primary?	1, Yes 0, No	X		X

Disease or domain under study:	1, Accidents and wounds 2, Blood and immune system 3, Cancer 4, Cardiovascular 5, Central nervous system/musculoskeletal 6, Digestive/endocrine 7, Ear and nose 8, Eye 9, General health 10, Genetic disorders 11, Gynaecology 12, Infectious diseases 13, Injuries 14, Mental health and behavioural conditions 15, Mouth and dental 16, Nutritional and metabolic 17, Pathological conditions 18, Pregnancy and birth 19, Respiratory disease 20, Skin 21, Symptoms and signs 22, Urogenital 99, Other (Please Specify)	X
In how many different countries was the study conducted? (Once you enter an integer number here, one or more drop-down boxes will appear below in order for you to enter the names of individual countries):		X
Country 1…15:	1, Afghanistan 2, Albania 3, Algeria .... 195, Zimbabwe	X
There are many types of experimental interventions that can evaluated in a CRT. These may include but are not limited to the following:		
• Educational/quality improvement interventions targeted at health care professionals	1, Yes 2, No 99, Unclear	X
• Quality improvement interventions targeted at the organisation of health care or health delivery service	1, Yes 2, No 99, Unclear	X
• Participant health promotion or educational intervention	1, Yes 2, No 99, Unclear	X

• Direct participant therapeutic intervention	1, Yes 2, No 99, Unclear		X	
• Other, specify	1, Yes 2, No 99, Unclear		X	
Please briefly describe details of the experimental intervention:			X	
Please select type of CONTROL intervention:	1, Not reported 2, No active intervention, i.e. usual care 3, Minimal application for experimental intervention 4, Placebo intervention 5, Other active intervention 99, Other (Please Specify)		X	
What is the cluster (i.e. unit of randomization)?	1, Hospital 2, Other health facility 3, Individual School 4, School district 5, Geographic areas (e.g. village or county) 6, General Practitioner/Primary Care Provider/Health specialist 7, Nursing home/aged care facility 8, Workplace 9, Household/family 99, Other (Please Specify)		X	
Was some form of restricted randomization procedure used? Please indicate all that apply: CHECK ALL THAT APPLY. If simple, unrestricted randomization was used (i.e. if none apply), please leave all boxes unchecked	1, Stratification 2, Pair-matching 3, Constrained randomization 99, Other (Please Specify) -99, Unclear/Not Reported		X	
Total number of clusters randomised:			X	
Of those **X** clusters randomised at the beginning of the trial, were there an equal number of clusters per arm?			X	
Number of clusters in control arm:			X	
Number of clusters in intervention arm:			X	
Was power or sample size calculation reported that accounted for the CRT design?	1, Yes 2, No, sample size calculation was reported but it did not account for CRT design 3, No, sample size calculation was NOT reported -99, Unclear/Not Reported		X	
In order to evaluate the effect of the intervention on the binary outcome, did the CRT use data from a cohort of the same individuals who were followed-up over time, take a cross-sectional sample of individuals at each follow-up time point or a mix of both?	1, Cohort 2, Cross-sectional 3, A mix of cohort and cross-sectional -99, Unclear/Not Reported	X		X

Was the binary outcome measured at baseline?	1, Yes (e.g. high blood pressure) 2, No (e.g. death) -99, Unclear/Not Reported	X		X
At how many follow-up time points was the binary outcome collected? *If follow-up data is collected via administrative data or electronic health record, please list the number of separate aggregations of data* *If the number of follow-up time points is ambiguous due to the nature of the binary outcome (e.g. death) please enter '999' and describe the nature of the ambiguity in the subsequently provided comment box:*		X		X
Was there a single primary follow-up time point of interest for the binary outcome?	1, Yes, a single time point is explicitly identified as primary. 2, No, multiple time points are specified as being of equal importance 3, No, authors did not explicitly identify any specific time point as primary -99, Unclear/Not Reported	X		
What method of data collection was used to measure the binary outcome? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, Questionnaire or survey 2, Administrative data 3, Laboratory data (e.g. test for malaria or HIV) 4, Electronic Health/Medical Record 99, Other (Please Specify) -99, Unclear from text what the data source was		X	
What OTHER method of data collection was used to measure the binary outcome?			X	
What is the average size of analyzed clusters? If the values are not reported directly, but can be calculated from the supplied data, then perform the calculation and enter the value e.g. 600 participants from 10 clusters gives an average cluster size of 60.			X	
Does the results section (including tables or figures) report the prevalence, risk or proportion of the binary outcome at follow-up: (CHECK ALL THAT APPLY).	1, Overall 2, By study arm -99, Unclear/Not Reported	X		X
Does the results section report a treatment effect for the results of analysis for the binary outcome as: (CHECK ALL THAT APPLY):	1, Absolute measure such as difference in proportions, prevalence difference or risk difference 2, Relative measure such as odds ratios or risk ratios 99, Other (Please Specify) -99, Does not quantify results for binary outcome or it is unclear from the results section what the form of the primary outcome of analysis is.	X		X
What type of ABSOLUTE measure is reported? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, Difference in prevalence, risk, or proportions (e.g. risk difference) 2, Number needed to treat (NNT) 99, Other (Specify below)	X		

If you've selected multiple types of ABSOLUTE treatment effect measures. For subsequent questions, you will be asked about estimation and magnitude of a single ABSOLUTE measure. Please select the PRIMARY reported measure type below and answer subsequent questions with respect to this measure (if primary is uncertain, then select the first mentioned):	1, Difference in prevalence, risk, or proportions (e.g. risk difference) 2, Number needed to treat (NNT) 99, Other measure	X	
What type of RELATIVE measure is reported? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, Odds ratio 2, Risk ratio/Relative risk/Prevalence ratio 99, Other (Specify below)	X	
You've selected multiple types of RELATIVE treatment effect measures. For subsequent questions, you will be asked about estimation and magnitude of a single RELATIVE measure. Please select the PRIMARY reported measure type below and answer subsequent questions with respect to this measure (if primary is uncertain, then select the first mentioned):	1, Odds ratio 2, Risk ratio/Relative risk/Prevalence ratio 99, Other measure	X	
What was the unit of analysis of the primary binary outcome? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, Cluster level 2, Individual level analysis	X	X
What was the main method of analysis at the CLUSTER level? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, Comparison of cluster level proportions 2, Comparison of mean of cluster level residuals (e.g. when the data are analyzed in two-stages with the first stage a logistic regression model with adjustment for individual-level covariates and not adjustment for clustering from which individual-level residuals are obtained and a cluster-level mean obtained for analysis in the second stage) 3, Comparison of some other cluster-level summary measure (Please Specify) 99, Other (Please Specify)	X	
What method was used to compare cluster-level summary statistics (i.e. proportions, cluster-level mean residuals or some other cluster-level summary statistic)? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, T-test 2, Z-test 3, Wilcoxon Rank Sum test 4, Permutation test 99, Other (Please Specify)	X	
What was the main method of analysis at the INDIVIDUAL level? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, Regression model WITHOUT accounting for clustering 2, Regression model accounting for clustering 99, Other (Please Specify)	X	
What method was used to account for clustering for the INDIVIDUAL level regression analysis? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, Mixed effects model (also referred to as a random effects model or a mixed model or, sometimes, as a multilevel or hierarchical model) 2, Generalized Estimating Equations (GEE) 3, Regular regression with cluster robust standard errors 99, Other (Please Specify)	X	
For the mixed effects model, it is expected that random intercepts were used. Does the methods section explicitly state that random intercept terms were included for cluster (i.e. for the unit of randomization)?	1, Yes 0, No -99, Unclear/Not Reported	X	

For the mixed effects model, were random slopes used to account for repeated measurements on clusters or individuals over time?	1, Yes 0, No -99, Unclear/Not Reported	X
For the GEE analysis, what type of working correlation matrix was chosen?	1, Exchangeable 2, Independent 99, Other (Please Specify) -99, Unclear/Not Reported	X
What was the OUTCOME DISTRIBUTION used to estimate the ABSOLUTE intervention effect?	1, Binomial 2, Gaussian/Normal 3, Poisson 99, Other (Please Specify) -99, Unclear/Not Reported	X
What was the LINK FUNCTION used to estimate the ABSOLUTE intervention effect? Note that a difference measure can be directly estimated from a regression model with an identity link (e.g. a linear regression-type approach) or could be obtained by transformation (e.g. by using, say, a logistic regression from which the outcome proportions are estimated by arm and then the absolute measure estimated using those).	1, Identity 2, Log 3, Logit 4, Probit 99, Other (Please Specify) -99, Unclear/Not Reported	X
What was the OUTCOME DISTRIBUTION used to estimate the RELATIVE intervention effect	1, Binomial 2, Gaussian/Normal 3, Poisson 99, Other (Please Specify) -99, Unclear/Not Reported	X
What was the LINK FUNCTION used to estimate the RELATIVE intervention effect	1, Identity 2, Log 3, Logit 4, Probit 99, Other (Please Specify) -99, Unclear/Not Reported	X
Were ABSOLUTE effects obtained via transformation?	1, Yes 0, No -99, Unclear/Not Reported	X
Were RELATIVE effects obtained via transformation?	1, Yes 0, No -99, Unclear/Not Reported	X
Was small sample bias acknowledged?	1, Yes 0, No	X
Was a small sample correction applied?	1, Yes 0, No -99, Unclear/Not Reported	X
Did the CLUSTER-level analysis account for the method of restricted randomization that was used?	1, Yes (Please Specify) 0, No -99, Unclear/Not Reported	X

Did the INDIVIDUAL-level analysis account for the method of restricted randomization that was used?	1, Yes (Please Specify) 0, No -99, Unclear/Not Reported	X	
What software was used for analysis ? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, SAS 2, Stata 3, R 4, SPSS 5, MPLUS 99, Other (Specify) -99, Unclear/Not Reported	X	X
What procedures were used for SAS software		X	
What procedures were used for Stata software		X	
What procedures were used for R software		X	
What procedures were used for SPSS software		X	
What procedures were used for MPLUS software		X	
What procedures were used for Other software		X	
What is the OVERALL prevalence, risk or proportion of the binary outcome at the primary follow-up time point of interest. Please report as a number without units and specify the units below		X	
What is the prevalence, risk or proportion of the binary outcome in the CONTROL ARM at the primary follow-up time point of interest: Please report as a number without units and specify the units below		X	
What is the prevalence, risk or proportion of the binary outcome in the INTERVENTION ARM at the primary follow-up time point of interest: Please report as a number without units and specify the units below		X	
Please specify UNITS of prevalence, risk or proportions listed above (e.g. percentage points, proportions)	1, Percentage points (i.e. %) 2, Fraction/proportion (i.e. on scale from 0 to 1) 99, Other (Please Specify)	X	
Please list the estimate for the ABSOULTE intervention effect at the primary time-point of interest		X	
Please specify units for the ABSOLUTE treatment effect at the primary time-point of interest (e.g. percentage points, proportion, etc....)	1, Percentage points (i.e. %) 2, Fraction/proportion (i.e. on scale from 0 to 1) 99, Other (Please Specify)	X	
Please list the LOWER confidence limit for the ABSOLUTE intervention effect at the primary time-point of interest		X	

Please list the UPPER confidence limit for the ABSOLUTE intervention effect at the primary time-point of interest		X		
Was a p-value reported for the ABSOLUTE intervention effect at the primary time point of interest?		X		
Please list the p-value for the ABSOLUTE intervention effect at the primary time-point of interest		X		
Please list the estimate for the RELATIVE intervention effect at the primary time-point of interest		X		
Please list the LOWER confidence limit for the RELATIVE intervention effect at the primary time-point of interest		X		
Please list the UPPER confidence limit for the RELATIVE intervention effect at the primary time-point of interest		X		
Was a p-value reported for the RELATIVE intervention effect at the primary time point of interest?		X		
Please list the p-value for the RELATIVE intervention effect at the primary time-point of interest		X		
Is a measure of the degree of clustering provided for the binary outcome (e.g. Intracluster correlation, coefficient of variation)?	1, Yes 0, No -99, Unclear/Not Reported	X		X
What type(s) of clustering measure was reported? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, Intracluster correlation (ICC) 2, Coefficient of variation 99, Other (Please Specify)	X		
How was the clustering measure(s) reported? (CHECK ALL THAT APPLY. If none apply then leave unchecked)	1, From the primary outcome analysis model 2, From another model 99, Another method (Please Specify) -99, Source of clustering measure unclear	X		
If the ICC was calculated, what scale is it reported on?	1, Logistic scale 2, Log scale 3, Linear scale 99, Other (Please Specify) -99, Unclear/Not Reported	X		

Supplementary Material 3. Data extraction form used for additional data abstraction by Duke team after all three data reconciliation workshops had taken place.

Field Label	Choices, Calculations, OR Slider Labels
use CONSORT flowchart	Explicit Implicit No mention
use CONSORT RCT checklist	Explicit Implicit No mention
use CONSORT RCT checklist with cluster extension	Explicit Implicit No mention
Flowchart of sample included	Yes – Main text Yes – Supplemental No
Checklist included	Yes – Main text Yes – Supplemental No
p-values are used to compare groups in baseline table	Yes No
Is the trial registered?	Yes No
If so, registry used	
If so, the registration number	
Is there a published, peer-reviewed protocol?	Yes No
If so, what is the reference/citation?	
Is the protocol accessible in a non-peered-reviewed format?	Yes No
Link, if available	
The study was reportedly approved by an ethics committee	Yes No
If so, what board or committee?	
Title of Guidelines	
Link to instructions	
Are the reporting guidelines in the author guidelines?	Yes No
Title of Reporting Guidelines	
Link to Reporting Guidelines	
CONSORT 2010	Explicitly Required Explicitly Recommended Implicitly Required – EQUATOR Implicitly Required – Other Implicitly Recommended – EQUATOR Implicitly Recommended – Other No mention

CONSORT (CRTs)	Explicitly Required
	Explicitly Recommended
	Implicitly Required – EQUATOR
	Implicitly Required – Other
	Implicitly Recommended – EQUATOR
	Implicitly Recommended – Other
	No mention
CONSORT (other)	Explicitly Required
	Explicitly Recommended
	Implicitly Required – EQUATOR
	Implicitly Required – Other
	Implicitly Recommended – EQUATOR
	Implicitly Recommended – Other
	No mention
Non-CONSORT guidelines	Explicitly Required
	Explicitly Recommended
	Implicitly Required – EQUATOR
	Implicitly Required – Other
	Implicitly Recommended – EQUATOR
	Implicitly Recommended – Other
	No mention

Supplementary Material 4. PRISMA checklist for reporting of systematic reviews

Section/topic	#	Checklist item	Where reported
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Title Page
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Abstract; Supplementary Material pg. 2 (S2)
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Introduction
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Methods; S2-3; Table S1
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Methods; S2; Supplementary Material 1
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Methods; S3; Table S1
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	S3; Table S1
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Table S1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	S3; Table S1
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	S3; Supplementary Material 2 and 3
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	S3; Supplementary Material 2 & 3

Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	N/A
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	S4
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	S4
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	S4
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	S4
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	S4; Figure S3
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	S5; Tables S4-S5
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	S5-6
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figure 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	S5-6
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	S5-6
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	S6; Discussion
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	S6; Discussion
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Discussion
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Acknowledgements

Supplementary Material 5. Reference list of N=73 papers included in the systematic review

1. Alagiyawanna A, Rajapaksa-Hewageegana N, Gunawardena N. The impact of multiple interventions to reduce household exposure to second-hand tobacco smoke among women: a cluster randomized controlled trial in Kalutara district, Sri Lanka. BMC Public Health. 2017;17(1):810.
2. Andersson SH, Bahr R, Clarsen B, Myklebust G. Preventing overuse shoulder injuries among throwing athletes: a cluster-randomised controlled trial in 660 elite handball players. Br J Sports Med. 2017;51(14):1073-80.
3. Arts DL, Abu-Hanna A, Medlock SK, Van Weert H. Effectiveness and usage of a decision support system to improve stroke prevention in general practice: a cluster randomized controlled trial. PLoS One. 2017;12(2).
4. Bloos F, Rüddel H, Thomas-Rüddel D, Schwarzkopf D, Pausch C, Harbarth S, et al. Effect of a multifaceted educational intervention for anti-infectious measures on sepsis mortality: a cluster randomized trial. Intensive Care Med. 2017;43(11):1602-12.
5. Boone P, Eble A, Elbourne D, Frost C, Jayanty C, Lakshminarayana R, et al. Community health promotion and medical provision for neonatal health-CHAMPION cluster randomised trial in Nagarkurnool district, Telangana (formerly Andhra Pradesh), India. PLoS Med. 2017;14(7):e1002324.
6. Chandler CI, Webb EL, Maiteki-Sebuguzi C, Nayiga S, Nabirye C, DiLiberto DD, et al. The impact of an intervention to introduce malaria rapid diagnostic tests on fever case management in a high transmission setting in Uganda: A mixed-methods cluster-randomized trial (PRIME). PLoS One. 2017;12(3):e0170998.
7. Chen CC, Li HC, Liang JT, Lai IR, Purnomo JDT, Yang YT, et al. Effect of a Modified Hospital Elder Life Program on Delirium and Length of Hospital Stay in Patients Undergoing Abdominal Surgery: a Cluster Randomized Clinical Trial. JAMA surgery. 2017;152(9):827-34.
8. Den Ouden H, Vos RC, Rutten G. Effectiveness of shared goal setting and decision making to achieve treatment targets in type 2 diabetes patients: a cluster-randomized trial (OPTIMAL). Health Expect. 2017;20(5):1172-80.
9. Devries K, Kuper H, Knight L, Allen E, Kyegombe N, Banks LM, et al. Reducing Physical Violence Toward Primary School Students With Disabilities. J Adolesc Health. 2017.
10. Downs JA, Mwakisole AH, Chandika AB, Lugoba S, Kassim R, Laizer E, et al. Educating religious leaders to promote uptake of male circumcision in Tanzania: a cluster randomised trial. The Lancet. 2017;389(10074):1124-32.
11. Ehlers LH, Simonsen KB, Jensen MB, Rasmussen GS, Olesen AV. Unannounced versus announced hospital surveys: a nationwide cluster-randomized controlled trial. Int J Qual Health Care. 2017;29(3):406-11.
12. Erismann S, Diagbouga S, Schindler C, Odermatt P, Knoblauch AM, Gerold J, et al. School Children's Intestinal Parasite and Nutritional Status One Year after Complementary School Garden, Nutrition, Water, Sanitation, and Hygiene Interventions in Burkina Faso. The American journal of tropical medicine and hygiene. 2017;97(3):904-13.
13. Fang CY, Ma GX, Handorf EA, Feng Z, Tan Y, Rhee J, et al. Addressing multilevel barriers to cervical cancer screening in Korean American women: A randomized trial of a community-based intervention. Cancer. 2017;123(6):1018-26.
14. Federman AD, Kil N, Kannry J, Andreopolous E, Toribio W, Lyons J, et al. An Electronic Health Record-based Intervention to Promote Hepatitis C Virus Testing Among Adults Born Between 1945 and 1965: a Cluster-randomized Trial. Med Care. 2017;55(6):590-7.
15. Fitzpatrick SL, Dickins K, Avery E, Ventrelle J, Shultz A, Kishen E, et al. Effect of an obesity best practice alert on physician documentation and referral practices. Transl Behav Med. 2017;7(4):881-90.
16. Foster G, Orne-Gliemann J, Font H, Kangwende A, Magezi V, Sengai T, et al. Impact of facility-based mother support groups on retention in care and PMTCT outcomes in rural Zimbabwe: the EPAZ cluster-randomized controlled trial. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2017;75:S207-S15.
17. Foxcroft DR, Callen H, Davies EL, Okulicz-Kozaryn K. Effectiveness of the strengthening families programme 10-14 in Poland: cluster randomized controlled trial. Eur J Public Health. 2017;27(3):494-500.
18. Fransen AF, van de Ven J, Schuit E, van Tetering AAC, Mol BW, Oei SG. Simulation-based team training for multi-professional obstetric care teams to improve patient outcome: a multicentre, cluster randomised controlled trial. BJOG. 2017;124(4):641-50.
19. Green J, Oman RF, Lu M, Clements-Nolle KD. Long-Term Improvements in Knowledge and Psychosocial Factors of a Teen Pregnancy Prevention Intervention Implemented in Group Homes. J Adolesc Health. 2017;60(6):698-705.

20. Guidet B, Leblanc G, Simon T, Woimant M, Quenot JP, Ganansia O, et al. Effect of Systematic Intensive Care Unit Triage on Long-term Mortality Among Critically Ill Elderly Patients in France: a Randomized Clinical Trial. JAMA. 2017;318(15):1450-9.
21. Gullo S, Galavotti C, Kuhlmann AS, Msiska T, Hastings P, Marti CN. Effects of a social accountability approach, CARE's Community Score Card, on reproductive health-related outcomes in Malawi: a cluster-randomized controlled evaluation. PLoS One. 2017;12(2).
22. Gupta J, Falb KL, Ponta O, Xuan Z, Campos PA, Gomez AA, et al. A nurse-delivered, clinic-based intervention to address intimate partner violence among low-income women in Mexico City: findings from a cluster randomized controlled trial. BMC Med. 2017;15(1):128.
23. Hailegebriel TD, Mulligan B, Cousens S, Mathewos B, Wall S, Bekele A, et al. Effect on Neonatal Mortality of Newborn Infection Management at Health Posts When Referral Is Not Possible: a Cluster-Randomized Trial in Rural Ethiopia. Global health, science and practice. 2017;5(2):202-16.
24. Hannon TS, Dugan TM, Saha CK, McKee SJ, Downs SM, Carroll AE. Effectiveness of Computer Automation for the Diagnosis and Management of Childhood Type 2 Diabetes: a Randomized Clinical Trial. JAMA pediatrics. 2017;171(4):327-34.
25. Harris MF, Parker SM, Litt J, van Driel M, Russell G, Mazza D, et al. An Australian general practice based strategy to improve chronic disease prevention, and its impact on patient reported outcomes: evaluation of the preventive evidence into practice cluster randomised controlled trial. BMC Health Serv Res. 2017;17(1):637.
26. Haug S, Paz Castro R, Kowatsch T, Filler A, Dey M, Schaub MP. Efficacy of a web-and text messaging-based intervention to reduce problem drinking in adolescents: Results of a cluster-randomized controlled trial. J Consult Clin Psychol. 2017;85(2):147.
27. Haug S, Paz Castro R, Kowatsch T, Filler A, Schaub MP. Efficacy of a technology-based, integrated smoking cessation and alcohol intervention for smoking cessation in adolescents: Results of a cluster-randomised controlled trial. J Subst Abuse Treat. 2017;82:55-66.
28. Hodder RK, Freund M, Bowman J, Wolfenden L, Campbell E, Dray J, et al. Effectiveness of a pragmatic school-based universal resilience intervention in reducing tobacco, alcohol and illicit substance use in a population of adolescents: cluster-randomised controlled trial. BMJ open. 2017;7(8):e016060.
29. Holt C, Milgrom J, Gemmill AW. Improving help-seeking for postnatal depression and anxiety: a cluster randomised controlled trial of motivational interviewing. Arch Womens Ment Health. 2017.
30. Holt TA, Dalton A, Marshall T, Fay M, Qureshi N, Kirkpatrick S, et al. Automated software system to promote anticoagulation and reduce stroke risk: cluster-randomized controlled trial. Stroke. 2017;48(3):787-90.
31. Jordan K, Edwards J, Porcheret M, Healey E, Jinks C, Bedson J, et al. Effect of a model consultation informed by guidelines on recorded quality of care of osteoarthritis (MOSAICS): a cluster randomised controlled trial in primary care. Osteoarthritis Cartilage. 2017;25(10):1588-97.
32. Joseph J, Gotora T, Erlwanger AS, Mushavi A, Zizhou S, Masuka N, et al. Impact of point-of-care CD4 testing on retention in care among HIV-positive pregnant and breastfeeding women in the context of Option B+ in Zimbabwe: a cluster randomized controlled trial. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2017;75:S190-S7.
33. Kesztyüs D, Lauer R, Kesztyüs T, Kilian R, Steinacker JM. Costs and effects of a state-wide health promotion program in primary schools in Germany - the Baden-Württemberg Study: a cluster-randomized, controlled trial. PLoS One. 2017;12(2):e0172332.
34. Kimani-Murage EW, Griffiths PL, Wekesah FM, Wanjohi M, Muhia N, Muriuki P, et al. Effectiveness of home-based nutritional counselling and support on exclusive breastfeeding in urban poor settings in Nairobi: a cluster randomized controlled trial. Globalization and health. 2017;13(1):90.
35. Lau BD, Shaffer DL, Hobson DB, Yenokyan G, Wang J, Sugar EA, et al. Effectiveness of two distinct web-based education tools for bedside nurses on medication administration practice for venous thromboembolism prevention: A randomized clinical trial. PLoS One. 2017;12(8):e0181664.
36. Lazzerini M, Shukurova V, Davletbaeva M, Monolbaev K, Kulichenko T, Akoev Y, et al. Improving the quality of hospital care for children by supportive supervision: a cluster randomized trial, Kyrgyzstan. Bull World Health Organ. 2017;95(6):397-407.
37. Lean ME, Leslie WS, Barnes AC, Brosnahan N, Thom G, McCombie L, et al. Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. Lancet. 2017.
38. Lemani C, Tang JH, Kopp D, Phiri B, Kumvula C, Chikosi L, et al. Contraceptive uptake after training community health workers in couples counseling: a cluster randomized trial. PLoS One. 2017;12(4):e0175879.
39. Leslie T, R, M, Mikhail A, Cundill B, Willey B, et al. Use of malaria rapid diagnostic tests by community health workers in Afghanistan: cluster randomised trial. BMC Med. 2017;15(1):124.

40. Liebschutz JM, Xuan Z, Shanahan CW, LaRochelle M, Keosaian J, Beers D, et al. Improving Adherence to Long-term Opioid Therapy Guidelines to Reduce Opioid Misuse in Primary Care: a Cluster-Randomized Clinical Trial. JAMA internal medicine. 2017;177(9):1265-72.
41. Ma GX, Fang CY, Seals B, Feng Z, Tan Y, Siu P, et al. A Community-Based Randomized Trial of Hepatitis B Screening Among High-Risk Vietnamese Americans. Am J Public Health. 2017;107(3):433-40.
42. Ma GX, Lee MM, Tan Y, Hanlon AL, Feng Z, Shireman TI, et al. Efficacy of a community-based participatory and multilevel intervention to enhance hepatitis B virus screening and vaccination in underserved Korean Americans. Cancer. 2017.
43. Manfrin A, Tinelli M, Thomas T, Krska J. A cluster randomised control trial to evaluate the effectiveness and cost-effectiveness of the Italian medicines use review (I-MUR) for asthma patients. BMC Health Serv Res. 2017;17(1):300.
44. McNairy ML, Lamb MR, Gachuhi AB, Nuwagaba-Biribonwoha H, Burke S, Mazibuko S, et al. Effectiveness of a combination strategy for linkage and retention in adult HIV care in Swaziland: The Link4Health cluster randomized trial. PLoS Med. 2017;14(11):e1002420.
45. Mmbaga EJ, Kajula L, Aarø LE, Kilonzo M, Wubs AG, Eggers SM, et al. Effect of the PREPARE intervention on sexual initiation and condom use among adolescents aged 12-14: a cluster randomised controlled trial in Dar es Salaam, Tanzania. BMC Public Health. 2017;17(1):322.
46. More NS, Das S, Bapat U, Alcock G, Manjrekar S, Kamble V, et al. Community resource centres to improve the health of women and children in informal settlements in Mumbai: a cluster-randomised, controlled trial. The Lancet Global Health. 2017;5(3):e335-e49.
47. Norskov AK, Wetterslev J, Rosenstock CV, Afshari A, Astrup G, Jakobsen JC, et al. Prediction of difficult mask ventilation using a systematic assessment of risk factors vs. existing practice - a cluster randomised clinical trial in 94,006 patients. Anaesthesia. 2017;72(3):296-308.
48. Nsangi A, Semakula D, Oxman AD, Austvoll-Dahlgren A, Oxman M, Rosenbaum S, et al. Effects of the Informed Health Choices primary school intervention on the ability of children in Uganda to assess the reliability of claims about treatment effects: a cluster-randomised controlled trial. Lancet (london, england). 2017;390(10092):374-88.
49. Obsuth I, S, A, Cope A, Pilbeam L, Murray AL, et al. London Education and Inclusion Project (LEIP): results from a Cluster-Randomized Controlled Trial of an Intervention to Reduce School Exclusion and Antisocial Behavior. J Youth Adolesc. 2017;46(3):538-57.
50. Oliveira-Ciabati L, Vieira CS, Franzon ACA, Alves D, Zaratini FS, Braga GC, et al. PRENACEL–a mHealth messaging system to complement antenatal care: a cluster randomized trial. Reproductive health. 2017;14(1):146.
51. Oyeledun B, Phillips A, Oronsaye F, Alo OD, Shaffer N, Osibo B, et al. The effect of a continuous quality improvement intervention on retention-in-care at 6 months postpartum in a PMTCT program in northern Nigeria: results of a cluster randomized controlled study. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2017;75:S156-S64.
52. Rona RJ, Burdett H, Khondoker M, Chesnokov M, Green K, Pernet D, et al. Post-deployment screening for mental disorders and tailored advice about help-seeking in the UK military: a cluster randomised controlled trial. Lancet (london, england). 2017;389(10077):1410-23.
53. Ruzagira E, Grosskurth H, Kamali A, Baisley K. Brief counselling after home-based HIV counselling and testing strongly increases linkage to care: a cluster-randomized trial in Uganda. J Int AIDS Soc. 2017;20(2):e25014.
54. Sarkar BK, West R, Arora M, Ahluwalia JS, Reddy KS, Shahab L. Effectiveness of a brief community outreach tobacco cessation intervention in India: a cluster-randomised controlled trial (the BABEX Trial). Thorax. 2017;72(2):167-73.
55. Schmidt-Mende K, Andersen M, Wettermark B, Hasselstrom J. Educational intervention on medication reviews aiming to reduce acute healthcare consumption in elderly patients with potentially inappropriate medicines-A pragmatic open-label cluster-randomized controlled trial in primary care. Pharmacoepidemiol Drug Saf. 2017.
56. Semrau KEA, Hirschhorn LR, Marx Delaney M, Singh VP, Saurastri R, Sharma N, et al. Outcomes of a Coaching-Based WHO Safe Childbirth Checklist Program in India. N Engl J Med. 2017;377(24):2313-24.
57. Sibanda EL, Tumushime M, Mufuka J, Mavedzenge SN, Gudukeya S, Bautista-Arredondo S, et al. Effect of non-monetary incentives on uptake of couples' counselling and testing among clients attending mobile HIV services in rural Zimbabwe: a cluster-randomised trial. The Lancet Global Health. 2017;5(9):e907-e15.
58. Singer S, Danker H, Roick J, Einenkel J, Briest S, Spieker H, et al. Effects of stepped psychooncological care on referral to psychosocial services and emotional well-being in cancer patients: A cluster-randomized phase III trial. Psychooncology. 2017;26(10):1675-83.

59. Soofi S, Cousens S, Turab A, Wasan Y, Mohammed S, Ariff S, et al. Effect of provision of home-based curative health services by public sector health-care providers on neonatal survival: a community-based cluster-randomised trial in rural Pakistan. The Lancet Global health. 2017;5(8):e796-e806.
60. Szilagyi PG, Schaffer S, Rand CM, Goldstein NPN, Vincelli P, Hightower AD, et al. School-located Influenza Vaccinations for Adolescents: A Randomized Controlled Trial. J Adolesc Health. 2017.
61. Tawfik MY. The Impact of Health Education Intervention for Prevention and Early Detection of Type 2 Diabetes in Women with Gestational Diabetes. J Community Health. 2017;42(3):500-10.
62. Tong EK, Nguyen TT, Lo P, Stewart SL, Gildengorin GL, Tsoh JY, et al. Lay health educators increase colorectal cancer screening among Hmong Americans: a cluster randomized controlled trial. Cancer. 2017;123(1):98-106.
63. Tong EY, Roman CP, Mitra B, Yip GS, Gibbs H, Newnham HH, et al. Reducing medication errors in hospital discharge summaries: a randomised controlled trial. Med J Aust. 2017;206(1):36-9.
64. Tran VD, Lee AH, Jancey J, James AP, Howat P, Mai LT. Physical activity and nutrition behaviour outcomes of a cluster-randomized controlled trial for adults with metabolic syndrome in Vietnam. Trials. 2017;18(1):18.
65. Vinereanu D, Lopes RD, Bahit MC, Xavier D, Jiang J, Al-Khalidi HR, et al. A multifaceted intervention to improve treatment with oral anticoagulants in atrial fibrillation (IMPACT-AF): an international, cluster-randomised trial. Lancet. 2017.
66. Wambura M, Mahler H, Grund JM, Larke N, Mshana G, Kuringe E, et al. Increasing voluntary medical male circumcision uptake among adult men in Tanzania. AIDS (london, england). 2017;31(7):1025-34.
67. Wei X, Zhang Z, Walley JD, Hicks JP, Zeng J, Deng S, et al. Effect of a training and educational intervention for physicians and caregivers on antibiotic prescribing for upper respiratory tract infections in children at primary care facilities in rural China: a cluster-randomised controlled trial. The Lancet Global Health. 2017;5(12):e1258-e67.
68. Wells S, Rafter N, Kenealy T, Herd G, Eggleton K, Lightfoot R, et al. The impact of a point-of-care testing device on CVD risk assessment completion in New Zealand primary-care practice: a cluster randomised controlled trial and qualitative investigation. PLoS One. 2017;12(4):e0174504.
69. West SK, Munoz B, Mkocha H, Dize L, Gaydos CA, Swenor B, et al. Treating village newcomers and travelers for trachoma: results from ASANTE cluster randomized trial. PLoS One. 2017;12(6):e0178595.
70. Wingfield T, Tovar MA, Huff D, Boccia D, Montoya R, Ramos E, et al. A randomized controlled study of socioeconomic support to enhance tuberculosis prevention and treatment, Peru. Bull World Health Organ. 2017;95(4):270-80.
71. Wong LY, Chua SS, Husin AR, Arshad H. A pharmacy management service for adults with asthma: a cluster randomised controlled trial. Fam Pract. 2017;34(5):564-73.
72. Wouters H, Scheper J, Koning H, Brouwer C, Twisk JW, van der Meer H, et al. Discontinuing Inappropriate Medication Use in Nursing Home Residents: A Cluster Randomized Controlled Trial. Ann Intern Med. 2017.
73. Wu Z, Tang Z, Mao Y, Van Veldhuisen P, Ling W, Liu D, et al. Testing and linkage to HIV care in China: a cluster-randomised trial. The Lancet HIV. 2017.

Supplementary Material 6. Group Author Names and Location and Affiliation

Name	Country	Name of institution
London, UK Workshop		
Christine Adrion	Germany	Ludwig Maximilians University, Germany
Naseerah Akooji	United Kingdom	University of Central Lancashire
Aneel Bhangu	United Kingdom	University of Birmingham
Bernadeta Bridgwood	United Kingdom	University of Leicester
Eric Budgell	United Kingdom	Nuffield Department of Medicine, University of Oxford
Agnès Caille	France	University of Tours and University of Nantes, INSERM, SPHERE 1246, France
Michael Campbell	United Kingdom	University of Sheffield
Claire Louise Chan	United Kingdom	Queen Mary, University of London
Michelle Collinson	United Kingdom	Clinical Trials Research Unit, University of Leeds
Andrew Copas	United Kingdom	MRC Clinical Trials Unit, University College London
Stephanie N Dixon	Canada	Institute for Clinical Evaluative Sciences, Canada
Sandra Eldridge	United Kingdom	Queen Mary, University of London
Alice S Forster	United Kingdom	University College London
Bruno Giraudeau	France	INSERM U1246, Tours, France
Alan Girling	United Kingdom	Institute of Applied Health Research, University of Birmingham
James Glasbey	United Kingdom	Department of Surgery, Insitution of Translational Medicine, University of Birmingham
Beatriz Goulao	United Kingdom	Health Services Research Unit, University of Aberdeen
Kelsey L Grantham	Australia	Monash University, Australia
Simon Hackett	United Kingdom	Newcastle University
Thomas Hamborg	United Kingdom	Queen Mary, University of London
Richard Hooper	United Kingdom	Queen Mary University of London
Kirsty James	United Kingdom	Institute of Psychiatry, Psychology and Neurosciences, Psychology and Neuroscience King's College London
Christopher I Jarvis	United Kingdom	London School of Hygiene and Tropical Medicine
Ben Jones	United Kingdom	University of Plymouth
Brennan C Kahan	United Kingdom	MRC Clinical Trials Unit, University College London
Mona Kanaan	United Kingdom	University of York
Jessica Kasza	Australia	Monash University, Australia
Lindsay Kendall	United Kingdom	GlaxoSmithKline
Caroline Kristunas	United Kingdom	University of Birmingham
Clémence Leyrat	United Kingdom	London School of Hygiene and Tropical Medicine
S J Macneill	United Kingdom	University of Bristol
Vichithranie W Madurasinghe	United Kingdom	Nuffield Department of Population Health, University of Oxford
James Martin	United Kingdom	University of Birmingham
Ariane Murielle Mbekwe Yepnang	France	University of Tours, France
Mirjam Moerbeek	The Netherlands	Utrecht University, Netherlands
Lazaro Mwakesi Mwandigha	United Kingdom	Imperial College London
Lee Aymar Ndounga Diakou	France	UMR INSERM U1246 SPHERE, University of Tours, France
Dmitri Nepogodiev	United Kingdom	University of Birmingham
Omar Omar	United Kingdom	University of Birmingham
Laura A Pankhurst	United Kingdom	NHS Blood and Transplant
Hayley Perry	United Kingdom	University of Reading
Ines Rombach	United Kingdom	University of Oxford
Beth Stuart	United Kingdom	University of Southamptom
Monica Taljaard	Canada	Ottawa Hospital Research Institute, Ottawa, Canada
Elsa Tavernier	France	INSERM 1415, Tours, France
Jennifer A Thompson	United Kingdom	London School of Hygiene and Tropical Medicine
Adam P Wagner	United Kingdom	University of East Anglia

Nina Wilson	United Kingdom	Newcastle University
Durham, NC, USA Workshop		
Shiwei Cao	United States	Duke University
Monica Harding	United States	Duke University
Kristie Kusibab	United States	Duke University
Hui-Jie Lee	United States	Duke University
Kara McCormack	United States	Duke University
Kelly Moran	United States	Duke University
Alice Parish	United States	Duke University
Ryan Simmons	United States	Duke University
Tracy Truong	United States	Duke University
Joao Ricardo Vissoci	United States	Duke University
Tongrong Wang	United States	Duke University
Xueqi Wang	United States	Duke University
Jeremy Weber	United States	Duke University
Jonathan Wilson	United States	Duke University
Siyun Yang	United States	Duke University
Zidanyue Yang	United States	Duke University
Birmingham, UK Workshop		
Hannah Bensoussane	United Kingdom	University of Birmingham
Jon Bishop	United Kingdom	University of Birmingham
Versha Cheed	United Kingdom	University of Birmingham
Alicia Gill	United Kingdom	University of Birmingham
Kelly Handley	United Kingdom	University of Birmingham
Pollyanna Hardy	United Kingdom	University of Birmingham
Catherine A Hewitt	United Kingdom	University of Birmingham
Natalie Ives	United Kingdom	University of Birmingham
Samir Mehta	United Kingdom	University of Birmingham
Smitaa Patel	United Kingdom	University of Birmingham
Yongzhong Sun	United Kingdom	University of Birmingham
Rebecca Woolley	United Kingdom	University of Birmingham