Reporting non-adherence in cluster randomised trials: A systematic review

Schadrac C. Agbla and **Karla DiazOrdaz**
Department of Medical Statistics London School of Hygiene and Tropical Medicine, UK

Abstract

Background—Treatment non-adherence in randomised trials refers to situations where some participants do not receive their allocated treatment as intended. For cluster randomised trials, where the unit of randomisation is a group of participants, non-adherence may occur at the cluster or individual level. When non-adherence occurs, randomisation no longer guarantees that the relationship between treatment receipt and outcome is unconfounded and the power to detect the treatment effects in intention-to-treat analysis may be reduced. Thus, recording adherence and estimating the causal treatment effect adequately are of interest for clinical trials.

Objectives—To assess the extent of reporting of non-adherence issues in published cluster trials and to establish which methods are currently being used for addressing non-adherence, if any, and whether clustering is accounted for in these.

Methods—We systematically reviewed 132 cluster trials published in English in 2011 previously identified through a search in PubMed.

Results—One-hundred and twenty three cluster trials were included in this systematic review. Non-adherence was reported in 56 cluster trials. Amongst these, 19 reported a treatment efficacy estimate: per protocol in 15 and as-treated in 4. No study discussed the assumptions made by these methods, their plausibility or the sensitivity of the results to deviations from these assumptions.

Limitations—The year of publication of the cluster trials included in this review (2011) could be considered a limitation of the present study, however no new guidelines regarding the reporting and the handling of non-adherence for cluster trials have been published since. In addition, a single reviewer undertook the data extraction. To mitigate this, a second reviewer conducted a validation of the extraction process on 15 randomly selected reports. Agreement was satisfactory (93%).

Conclusions—Despite the recommendations of the CONSORT statement extension to cluster randomised trials, treatment adherence is under-reported. Amongst the trials providing adherence information, there was substantial variation in how adherence was defined, handled and reported. Researchers should discuss the assumptions required for the results to be interpreted causally, and whether these are scientifically plausible in their studies. Sensitivity analyses to study the robustness of the results to departures from these assumptions should be performed.

1   Introduction

Cluster randomised trials, where pre-existing groups of individuals are randomised, have become a common design to test public health and primary care interventions, as often the target of the intervention is a hospital or general practice, or their staff. Increased administrative convenience, reduction of contamination between experimental arms and improved adherence with allocated treatment, are often cited amongst the advantages of adopting this design [1, 2, 3]. Nevertheless, treatment adherence may be more challenging in cluster trials because of the hierarchical nature of the design and the delivery of the intervention, where at least two levels at which deviations from protocol called non-adherence, can occur, e.g. cluster or individual level [4]. The nature of the non-adherence patterns largely depends on the nature of the intervention. Some interventions are aimed exclusively at the clusters and thus all individuals within a cluster are exposed to the same treatment. Water fluoridation in a village would be one such example.

In other cluster trials, participants within the same cluster may individually stop adhering to the allocated treatment. An example of this is the study conducted by Sommer et al. [5] where villages were randomised to "vitamin A supplements" or not, to be offered to all infants. However, some children whose villages were randomised to "vitamin A supplements" did not receive the supplements. Finally, non-adherence at both levels is possible for interventions with components targeted at both levels. For example, the OPERA study (exercise for treating depression in care home residents) aimed to evaluate the impact of a 'whole home' exercise intervention on depressive symptoms in older adults living in care homes in England [6]. Clusters were randomly allocated to provide either a depression awareness training session for care home staff (control) or a twice-weekly physiotherapist-led exercise class (active intervention) for 12 months. Some of the exercise classes did not occur due to a shortage of staff (cluster-level non-adherence). In addition, even when the nursing home ran the exercise classes, some individuals recruited to attend these classes did not to do so (individual-level non-adherence). Nursing homes and individuals complied with their assigned treatment during some weeks but then deviated at a later time, introducing a time-varying non-adherence pattern.

The standard analysis of a trial with departures from allocated treatment is intention-to-treat, which compares outcomes between the groups as randomised, ignoring the actual treatments received. The intention-to-treat estimates the effect of being offered (or allocated to) the treatment, and cannot necessarily be interpreted as the causal effect of treatment received. An intention-to-treat analysis with poor adherence may dilute a true treatment effect; with negative outcomes such as side effects, adverse events or mortality, an intention-to-treat estimate which is closer to the null than the true causal effect may make a more toxic or aggressive treatment look less harmful. In addition, non-adherence leads to a loss of power in intention-to-treat analysis [7].

Where there is an interest in estimating treatment efficacy, as the causal effect of receiving the treatment according to the protocol is called, analytical approaches such as "per protocol" and "as treated" are often used [8]. Per protocol restricts the analysis only to participants who received their assigned treatment, whereas as treated analysis compares

participants according to their treatment receipt, regardless of their treatment assignment. Both per protocol and as treated may be subject to selection bias, and their validity as causal estimates depends on the assumption that the groups being compared are exchangeable, that is comparable in terms of their measured and unmeasured covariates. This is a very strong assumption. Since the original comparable groups achieved through randomisation are not preserved, any observed differences in outcomes are not necessarily due to the treatment effect, but potentially also due to differences in covariates [9]. Per protocol also leads to a reduction of statistical power [7]. However, some design features may increase the likelihood of per protocol analysis to be unbiased. One example is when the trial is double-blinded and the treatments have low rates of adverse events, because in this situation, both treatment switching and discontinuations are unlikely to be associated to outcomes, as described by White [7].

More recently, "modified" per protocol analyses, that adjust for potential confounders that may lead to selection bias, have been advocated [10, 11]. These modified per protocol analyses allow the investigators to adjust for baseline and post-randomisation variables believed to be sufficient to adjust for the confounding of the association between treatment received and outcome. The assumption of "no unobserved confounding", required for their validity is still strong.

There are statistical methods that do not assume "no unobserved confounding". Instead, they rely on randomised treatment being an instrumental variable [12, 13, 14], and have been proposed in the context of individually randomised trials [7, 15]. Extensions to cluster randomised trials exist [4, 16, 17, 18]. A brief summary of these is given in Box 1. However, methods which are applicable to cluster settings tend to be limited in their usefulness, requiring some programming or the use of specialised software. Previous systematic reviews investigating the reporting and statistical handling of non-adherence in individually randomised controlled trials have been published [8, 19]. These have found that adherence to treatment is often under-reported, and when reported, sufficient detail on how adherence was defined is often not included. They also found that the majority of studies used 'unadjusted' per protocol analyses to obtain treatment efficacy estimates [8]. Cluster randomised trials are more complex to run, analyse and report than individually randomised trials, and previous systematic reviews of cluster trials have found that despite the CONSORT extension mentioning explicitly the need to report numbers assigned, on treatment and analysed at the cluster and individual level, this information is often lacking [3, 20]. However, no previous study has focused on the conduct of statistical analyses aiming to estimate causal treatment effects in the presence of treatment non-adherence.

Thus, the aim of the present study is to assess the reporting and handling of non-adherence in cluster randomised trials, and in particular, to establish the prevalence of non-adherence and describe the methods used to obtain adherence-adjusted treatment effects. For this, we perform a secondary analysis of data originally extracted for a systematic review investigating the reporting and adjustment of missing data in cluster trials [21]. We also propose some guidelines for reporting adherence and for conducting adherence-adjusted analysis of cluster trials.

2 Methods

2.1 Search strategy and inclusion criteria

This review uses a database of 132 cluster trials previously identified using a published electronic search strategy [21]. The full electronic search strategy is reported in Box A in the Supplemental File. Reports were eligible for inclusion if they were full reports of cluster randomised controlled trials, published in English in 2011. They were excluded if they were quasi-experimental, self-identified as pilot, feasibility, or preliminary studies; only reported cost-effectiveness or where no data at the individual level were collected. We also excluded crossover trials, where deviations from randomised treatment may include failure to follow the randomised sequence of treatments, and studies reporting only sub-samples of previously published cluster trials data.

2.2 Piloting and validation

Two researchers independently piloted a data extraction form using five randomly selected reports. This helped to identify extra relevant information to extract and to improve the study protocol. After updating the study protocol and the data extraction form, a random sample of fifteen reports was used for validation of the extraction procedures. In case of discrepancy, a final decision was made by consensus and the appropriate information was recorded in the data extraction form. Once the team was satisfied with the extraction procedure, one researcher performed the data extraction in the whole sample. When there was doubt or ambiguity, this was reviewed by the second extractor and a consensus was reached.

2.3 Data extraction

Data were extracted on one primary outcome per report, defined as that specified by the authors or, if not specified, the outcome used in sample size calculations. If no primary outcome was specified and no sample size calculation was reported, the first outcome presented in the abstract or manuscript was considered as primary. Information was collected on the type of cluster, the type of primary outcome (binary, continuous, categorical), whether a harm outcome was investigated [22], and the type of intervention given in the control arm (placebo, standard care or active). Information on the level of adherence (cluster-level or individual-level) was also recorded. Non-adherence was considered to be at the cluster level if the treatment received was different from that assigned for all the participants within clusters, and it was considered to be at the individual level if the treatment received differed from the allocated treatment on an individual basis within the same cluster.

Additionally, data on total number of clusters and individuals randomised and analysed were extracted as well as numbers of clusters and individuals receiving the allocated treatment. We defined treatment non-adherence as discrepancy between the allocated course of treatment and the actual treatment received [8]. Descriptions of treatment adherence, including intra-cluster correlation coefficient for treatment adherence [17], were also recorded, when reported. We also recorded information on adherence-adjusted analyses and whether clustering was accounted for.

We adapted the definitions by Dodd et al. [8] and extracted data about the duration of the intervention. A "one-off" intervention is defined as that which is received at a single time point, e.g. a surgery. A "short-term" intervention is defined as an intervention implemented at different time points over a short period; for example, five training sessions on the importance of breastfeeding over one week. Any other recurrent intervention over an extended period of time was classified as a "long-term" intervention.

2.4 Analysis

Simple analyses were performed to describe the frequency of adherence-reporting and the reported methods used to adjust for non-adherence. The median (and the first and third quartiles) of the number of clusters and individuals randomised, on treatment and analysed are provided.

For the percentage of non-adherence, we used the author-reported non-adherence when this was reported numerically. If not, we calculated the percentage of non-adherence for each study, from the data provided in the manuscript (the ratio between "off allocated treatment" participants to the total number randomised).

3 Results

After excluding 7 reports that used only sub-samples of cluster trials data and 2 crossover trials, our final sample included 123 cluster trials. See the Flow Chart, Figure 1. During the validation phase, the two extractors had an initial agreement of 93%, ultimately achieving consensus by discussion.

3.1 Trial characteristics

Trial characteristics are shown in Table 1. Interventions were mainly concerned with changing healthcare practices (63 trials, 51%), educational practices (27 trials, 22%) or lifestyle (25 trials, 20%). In most trials, standard care was used as the control intervention (96 trials, 76%). The primary outcome was either continuous (65 trials, 53%) or binary (57 trials, 46%), with one exception (multi-category). Adverse events were investigated in 12 trials (10%).

The intervention was implemented exclusively at the cluster level in 65 trials (53%) and at the individual level in 58 trials (47%). Long-term interventions were the most common (83 trials, 68%), followed by short-term interventions (35 trials, 28%). The majority of the studies were two-arm trials (106 trials, 86%). The median (1st–3rd quartiles) number of clusters randomised in each trial arm was 12 (7–24) and the number of clusters per trial arm ranged from 2 to 199. The number of individuals per cluster had a median (1st–3rd quartiles) of 27 (10–65) with a range of 2 to 14350.

Intention-to-treat analysis was done as primary analysis in 119 trials (97%), with the remaining 4 trials (3.2%) using per protocol or as treated analysis. Only 6 trials (5%) used cluster-level analysis (primary outcome defined at the cluster level) while the remaining 117 trials use individual-level analysis. Among these, clustering was not accounted for in 12 trials (10%). See Table 2.

3.2 The reporting and handling of non-adherence

Sixty-one reports (50%) included information on adherence: full adherence was reported in 5 trials while the remaining 56 trials reported some form of non-adherence to the allocated treatment. Table 3 reports the adherence characteristics of these trials. The reporting of adherence was more frequent in interventions of short duration (57%) compared to those of long duration (47%). Forty-four trials (72%) used a binary treatment adherence definition, with only one report justifying the threshold used for this dichotomisation. Five trials (8%) recorded non-adherence as a continuous variable, while the remaining 12 trials (20%) gave no details on the definition of adherence used. Only 11 trials (9%) provided a flow chart with complete information on how many clusters and/or individuals received the assigned treatment. Nine trials reporting non-adherence performed adverse events analysis. Non-adherence at the cluster level was reported in 15 trials (24%), with a further 4 (6%) studies reporting full cluster adherence. Non-adherence at the individual level was reported in 41 trials, representing 71% of the 58 trials reporting treatment non-adherence at the individual level, while one trial (2%) reported full adherence. No study reported the use of an intra-cluster correlation coefficient for adherence.

3.2.1 Adherence by allocated groups

3.2.2 Active group—Five studies provided the percentage of cluster-level non-adherence, with a median (1st–3rd quartiles) of 44.8% (33%–50%), with a further 10 reporting cluster non-adherence without further details. At the individual level, 30 (73%) out of 41 studies reported this, with a corresponding median (1st–3rd quartiles) of 15% (9%–24%).

3.2.3 Control group—Adherence to the control protocol was less frequently reported; 5 trials stated full adherence, while a further 15 studies reported some form of non-adherence. Cluster-level non-adherence was reported in one trial, while full adherence was reported in a further 4 trials. At the individual level, 19 trials reported control-treatment non-adherence, with full adherence in one study.

3.2.4 Adherence-adjusted analyses—Fifteen trials performed per protocol analyses, with the remaining 4 studies carrying out as treated analyses either as primary or secondary analyses. No study reported the complier average causal effect estimate. Amongst the 9 studies with a safety outcome, 4 trials performed a per protocol analysis [23, 24, 25, 26], with a further trial using an as treated analysis [27]. Two studies did not account for clustering in their adherence-adjusted analyses [23, 28]. No study reported the assumptions necessary for their adherence-adjusted analyses to be unbiased causal treatment estimates. In any case, none of these studies was double-blinded. We summarise some of the characteristics of these adherence-adjusted analyses in Table 4.

4 Discussion

This is the first systematic review of reporting practices of non-adherence with randomised treatment in cluster randomised trials. Our findings show that about half of the studies include information on treatment adherence, but details on numbers of clusters and

individuals that adhered to the intended treatment are often incomplete. Schulz et al. [29, 30] found that trials reporting exclusions after treatment initiation (i.e. deviations from protocol) tend to be of higher methodological quality than those that did not report it. This is known as the "exclusion paradox". It is therefore possible that those studies that did not report on adherence also experienced protocol deviations. On this basis, we estimate that in this study the proportion of trials with non-adherence lies within the range 23% to 94% at the cluster level and 71% to 98% at the individual level. In addition, we found that studies tended to report more often adherence at the individual level. This potential under-reporting may be due to the complexity of defining adherence in cluster trials, and that as cluster trials are often pragmatic in nature, recording adherence to treatment protocol is not often a primary concern.

Amongst the studies reporting non-adherence, only one-third specified departures from protocol in the control group. This has to be interpreted in light of the fact that in our review, "usual care" was used as control in approximately three quarters of studies, and that defining and measuring adherence to "usual care" may be difficult or impractical. In general, the nature of the departures from protocol was very poorly reported, and it was not possible to ascertain whether alternative treatments to those in the trial, i.e. not originally included in the design of the study, were taken. Knowledge of the alternative regimes followed by those individuals who did not adhere to their allocated treatment is important if we want to judge the impact of such non-adherence on the causal interpretation of an intention-to-treat analysis. If no external treatments are available, then the intention-to-treat estimate will be diluted towards the null, when compared with the true treatment effect. Moreover, the reported non-adherence details (numbers initiating and completing the treatment protocol, period of discontinuation, etc.) were often inadequate for a meaningful interpretation of the study results.

All of the studies reporting adherence-adjusted estimates performed per protocol or as treated analyses, without discussing the plausibility of the necessary assumptions for the results to be interpretable as unbiased causal estimates. These "unadjusted" analyses rely on the assumption that the association between treatment received and outcome is completely unconfounded [7, 9]. Since this assumption is very strong, we would advise adjusting for measured confounders in a "modified" per protocol analysis (as suggested by Hern´an and Robins [11]), thus relying instead on the "no unobserved confounding" assumption. Moreover, in the context of randomised trials, randomised treatment is very plausibly a valid instrumental variable, and the monotonicity assumption may sometimes hold by design, for example where the experimental intervention is not available to the controls. These assumptions are sufficient to identify the CACE, allowing the analyst to obtain causal effects in the presence of unobserved confounding. In the absence of complex interactions between compliance classes at the cluster and individual level, statistical methods to estimate CACEs accounting for the clustering in the data could be used (See Box 1). However, no report included in this review performed a complier average causal effect.

4.1 Comparison with previous studies

A previous systematic review including 152 cluster trials published between 1997 and 2000 found that non-adherence was reported in about 24% of the studies [31]. For individual randomised trials, Dodd et al. [8], in a review of 100 trials published in 2008 in five leading medical journals, found this percentage to be 98%. In contrast, the review performed by Zhang et al. [19], which considered individual randomised drug trials published in 2010, found a prevalence of non-adherence reporting of 46%. Both of these results are thus in line with the lower and upper bounds found in our study. These two previous individually-randomised trials reviews noted a lack of justification in the threshold used in defining a binary measure of non-adherence [8, 19]. In the present review, only one justified this choice.

The median percentage of individual-level non-adherence reported by the cluster trials included here was 13%. Similar median percentages of non-adherence were found in previous reviews of adherence in individually randomised trials, 10–20% in Dodd [8], and 11.6% in Zhang [19]. While the latter reported finding a monotonic trend of adherence with regards to intervention duration [19], we did not find any such trend. This could be because adherence was not clearly reported in over 40% of both long and short-term interventions. In fact, in view of the "exclusion paradox", non-adherence with short-term interventions could be as high as the non-adherence reported in long-term interventions.

Only 3% of the studies included in the present review presented an adherence-adjusted analysis as primary, with the great majority reporting an intention-to-treat approach. Of those studies assessing treatment efficacy, per protocol analysis was the most used. Dodd et al. [8] also found that the majority of studies attempting to adjust for non-adherence in an analysis used per protocol.

Although the extended CONSORT statement for cluster randomised trials [32, 33] recommends reporting the numbers of clusters and individuals randomised and receiving their assigned treatment, we found that the reporting of these numbers in the flow chart was low (9%). This is in contrast to the results reported by Dodd et al. [8], who found that 58% stated the number of participants actually initiating their allocated treatment. A possible explanation may be the lower adherence to CONSORT guidelines among cluster trial reports [20] as well as the extra complexities of defining, measuring and recording adherence at both levels.

Strengths and limitations

The cluster trials database used for this review was identified using a rigorous electronic search procedure previously published [21]. This search strategy was calibrated with a previously published one [34], which had been validated with an ideal set of cluster randomised trials identified from manual examination of a large sample of health journals and was found to have high sensitivity (90.1%). Nevertheless, we may have missed some cluster randomised trials, as reports may fail to clearly identify the cluster randomisation design in either title or abstract.

Our inclusion criteria were broad, and thus our sample should be representative of the quality of conducting and reporting of cluster trials. The included reports were published in 2011, but we do not expect a change in practice for adherence reporting, as the updated CONSORT statement for cluster trials [33] was available in pre-print form since 2010 and did not contain any new guidelines with regards to adherence reporting or handling over and above those included in the 2004 version [32].

As with reviews of this nature, our assessments were based only on the information included in the trial reports. It is possible that non-adherence is more common but under-reported, the so-called exclusion paradox [29, 30]. We calculated ranges of non-adherence to reflect this possibility.

Another possible limitation is the use of a single reviewer for data extraction. However, single-reviewer extraction was only carried out after a validation phase, where a second reviewer conducted extraction. Agreement between the two extractors was high during validation. Additionally, during full-extraction, whenever there was ambiguity, the second reviewer's opinion was sought and disagreements were resolved by consensus.

5 Conclusion

Non-adherence with allocated treatment is common in cluster randomised trials but it is not sufficiently well reported. Our study suggests that cluster-level non-adherence is less common than individual-level non-adherence. However, after taking into consideration possible under-reporting, the overall level of non-adherence in cluster randomised trials may well be comparable with that previously observed in individually randomised trials [8], weakening the claim that a cluster randomisation design improves treatment-adherence [1].

A greater effort should be made to improve the quality of reporting of adherence data and analyses. When undertaking causal analyses as part of a cluster randomised trial with non-adherence, researchers should consider carefully the assumptions necessary for their analyses to result in valid inferences, and discuss their plausibility in the context of their trial. Sensitivity analyses should be undertaken when departures from these assumptions are suspected. It is also important to remember that in cluster randomised trials, the validity of the results also relies on obtaining an appropriate estimate for the standard errors, for which it is crucial to use a method that correctly models the dependence structure of the data.

Methodologists should make existing causal methods that accommodate the clustering more widely available and easy to implement in commonly used software. To promote their use in practical applications, methodologists should also publish more tutorial papers describing clearly the assumptions needed and detailing the challenges of performing such adherence-adjusted analyses in the context of a good empirical example.

We conclude by making some recommendations for trialists conducting cluster randomised trials. See Box 2. Previous recommendations for conducting and reporting adherence analyses for individually randomised trials are still relevant [8]. A new framework for the conduct and interpretation of randomised trials in the presence of treatment non-adherence

has been recently published, and we encourage the readers to follow these guidelines as much as possible [35].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

[1]. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. American Journal of Public Health. 2004; 94(3):416–422. [PubMed: 14998805]

[2]. Glynn RJ, Brookhart MA, Stedman M, Avorn J, Solomon DH. Design of cluster-randomized trials of quality improvement interventions aimed at medical care providers. Medical Care. 2007; 45(10):S38–S43. [PubMed: 17909381]

[3]. Wright N, Ivers N, Eldridge S, Taljaard M, Bremner S. A review of the use of covariates in cluster randomized trials uncovers marked discrepancies between guidance and practice. Journal of Clinical Epidemiology. 2015; 68(6):603–609. [PubMed: 25648791]

[4]. Schochet PZ, Chiang HS. Estimation and identification of the complier average causal effect parameter in education RCTs. Journal of Educational and Behavioral Statistics. 2011; 36(3):307–345.

[5]. Sommer A, Djunaedi E, Loeden A, Tarwotjo I, West K, Tilden R, et al. Impact of Vitamin A supplementation on childhood mortality. A randomised controlled community trial. The Lancet. 1986 May; 327(8491):1169–1173.

[6]. Underwood M, Lamb SE, Eldridge S, Sheehan B, Slowther AM, Spencer A, et al. Exercise for depression in elderly residents of care homes: a cluster-randomised controlled trial. The Lancet. 2013; 382(9886):41–49.

[7]. White IR. Uses and limitations of randomization-based efficacy estimators. Statistical Methods in Medical Research. 2005; 14(4):327–347. [PubMed: 16178136]

[8]. Dodd S, White IR, Williamson P. Nonadherence to treatment protocol in published randomised controlled trials: a review. Trials. 2012; 13(1):84. [PubMed: 22709676]

[9]. Ten Have TR, Normand SLT, Marcus SM, Brown CH, Lavori P, Duan N. Intent-to-Treat vs. Non-Intent-to-Treat Analyses under Treatment Non-Adherence in Mental Health Randomized Trials. Psychiatric Annals. 2008 Dec; 38(12):772–783. [PubMed: 20717484]

[10]. Murray EJ, Hernn MA. Adherence adjustment in the Coronary Drug Project: A call for better per-protocol effect estimates in randomized trials. Clinical Trials. 2016; 13(4):372–378. DOI: 10.1177/1740774516634335 [PubMed: 26951361]

[11]. Hernán MA, Robins JM. per-protocol Analyses of Pragmatic Trials. The New England Journal of Medicine. 2017; 377(14):1391–1398. [PubMed: 28976864]

[12]. Imbens GW, Angrist JD. Identification and Estimation of Local Average Treatment Effects. Econometrica. 1994; 62(2):467–475.

[13]. Angrist, J, Imbens, G. Identification and estimation of local average treatment effects. National Bureau of Economic Research Cambridge; Mass, USA: 1995.

[14]. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics. 2002; 58(1):21–29. [PubMed: 11890317]

[15]. Bellamy SL, Lin JY, Ten Have TR. An introduction to causal modeling in clinical trials. Clinical Trials. 2007; 4(1):58–73. [PubMed: 17327246]

[16]. Frangakis CE, Rubin DB, Zhou XH. Clustered encouragement designs with individual noncompliance: Bayesian inference with randomization, and application to advance directive forms. Biostatistics. 2002; 3(2):147–164. [PubMed: 12933609]

[17]. Jo B, Asparouhov T, Muthén BO, Ialongo NS, Brown CH. Cluster randomized trials with treatment noncompliance. Psychological Methods. 2008; 13(1):1. [PubMed: 18331150]

[18]. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Statistics in Medicine. 2014; 33(13):2297–2340. [PubMed: 24599889]

[19]. Zhang Z, Peluso MJ, Gross CP, Viscoli CM, Kernan WN. Adherence reporting in randomized controlled trials. Clinical Trials. 2014; 11(2):195–204. [PubMed: 24357665]

[20]. Ivers N, Taljaard M, Dixon S, Bennett C, McRae A, Taleban J, et al. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8. British Medical Journal. 2011; 343:d5886. [PubMed: 21948873]

[21]. Díaz-Ordaz K, Kenward MG, Cohen A, Coleman CL, Eldridge S. Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines. Clinical Trials. 2014

[22]. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC Medicine. 2010; 8(1):1. [PubMed: 20051100]

[23]. Neuzil KM, Thiem VD, Janmohamed A, Huong VM, Tang Y, Diep NTN, et al. Immunogenicity and reactogenicity of alternative schedules of HPV vaccine in Vietnam: a cluster randomized noninferiority trial. The Journal of the American Medical Association. 2011; 305(14):1424–1431. [PubMed: 21486975]

[24]. Boorsma M, Frijters DH, Knol DL, Ribbe ME, Nijpels G, van Hout HP. Effects of multidisciplinary integrated care on quality of care in residential care facilities for elderly people: a cluster randomized trial. Canadian Medical Association Journal. 2011; 183(11):E724–E732. [PubMed: 21708967]

[25]. Dangour AD, Albala C, Allen E, Grundy E, Walker DG, Aedo C, et al. Effect of a nutrition supplement and physical activity program on pneumonia and walking capacity in Chilean older people: a factorial cluster randomized trial. PLoS Medicine. 2011; 8(4):e1001023. [PubMed: 21526229]

[26]. Luoto R, Kinnunen TI, Aittasalo M, Kolu P, Raitanen J, Ojala K, et al. Primary prevention of gestational diabetes mellitus and large-for-gestational-age newborns by lifestyle counseling: a cluster-randomized controlled trial. PLoS Medicine. 2011; 8(5):e1001036. [PubMed: 21610860]

[27]. Zamorano J, Erdine S, Pavia A, Kim JH, Al-Khadra A, Westergaard M, et al. Proactive multiple cardiovascular risk factor management compared with usual care in patients with hypertension and additional risk factors: the CRUCIAL trial. Current Medical Research and Opinion. 2011; 27(4):821–833. [PubMed: 21306285]

[28]. Tagbor H, Cairns M, Nakwa E, Browne E, Sarkodie B, Counihan H, et al. The clinical impact of combining intermittent preventive treatment with home management of malaria in children aged below 5 years: cluster randomised trial. Tropical Medicine & International Health. 2011; 16(3): 280–289. [PubMed: 21159081]

[29]. Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. British Medical Journal. 1996; 312(7033):742–744. [PubMed: 8605459]

[30]. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. The Lancet. 2002; 359(9308):781–785.

[31]. Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. Clinical Trials. 2004; 1(1):80–90. [PubMed: 16281464]

[32]. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. British Medical Journal. 2004; 328(7441):702–708. [PubMed: 15031246]

[33]. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. British Medical Journal. 2012

[34]. Taljaard M, McGowan J, Grimshaw JM, Brehaut JC, McRae A, Eccles MP, et al. Electronic search strategies to identify reports of cluster randomized trials in MEDLINE: low precision will

improve with adherence to reporting standards. BMC Medical Research Methodology. 2010; 10(1):1. [PubMed: 20053272]

[35]. Dodd S, White IR, Williamson P. A framework for the design, conduct and interpretation of randomised controlled trials in the presence of treatment changes. Trials. 2017 Oct.18(1):498.doi: 10.1186/s13063-017-2240-9 [PubMed: 29070048]

[36]. Bhutta ZA, Soofi S, Cousens S, Mohammad S, Memon ZA, Ali I, et al. Improvement of perinatal and newborn care in rural Pakistan through community-based strategies: a cluster-randomised effectiveness trial. The Lancet. 2011; 377(9763):403–412.

[37]. Acolet D, Allen E, Houston R, Wilkinson AR, Costeloe K, Elbourne D. Improvement in neonatal intensive care unit care: a cluster randomised controlled trial of active dissemination of information. Archives of Disease in Childhood-Fetal and Neonatal Edition. 2011; 96(6):F434–F439. [PubMed: 21393310]

[38]. Auger N, Daniel M, Knäuper B, Raynault MF, Pless B. Children and youth perceive smoking messages in an unbranded advertisement from a NIKE marketing campaign: a cluster randomised controlled trial. BMC Pediatrics. 2011; 11(1):1. [PubMed: 21214908]

[39]. Beer C, Horner B, Flicker L, Scherer S, Lautenschlager NT, Bretland N, et al. A cluster-randomised trial of staff education to improve the quality of life of people with dementia living in residential care: the DIRECT study. PloS One. 2011; 6(11):e28155. [PubMed: 22140531]

[40]. Bickman L, Kelley SD, Breda C, de Andrade AR, Riemer M. Effects of routine feedback to clinicians on mental health outcomes of youths: results of a randomized trial. Psychiatric Services. 2011

[41]. Cooke LJ, Chambers LC, Añez EV, Croker HA, Boniface D, Yeomans MR, et al. Eating for pleasure or profit the effect of incentives on childrens enjoyment of vegetables. Psychological Science. 2011; 22(2):190–196. [PubMed: 21191095]

[42]. Cutrer WB, Castro D, Roy KM, Turner TL. Use of an expert concept map as an advance organizer to improve understanding of respiratory failure. Medical Teacher. 2011; 33(12):1018–1026. [PubMed: 22225439]

[43]. Estrada CA, Safford MM, Salanitro AH, Houston TK, Curry W, Williams JH, et al. A web-based diabetes intervention for physician: a cluster-randomized effectiveness trial. International Journal for Quality in Health Care. 2011; 23(6):682–689. [PubMed: 21831967]

[44]. Smith SM, Paul G, Kelly A, Whitford DL, OShea E, ODowd T. Peer support for patients with type 2 diabetes: cluster randomised controlled trial. British Medical Journal. 2011; 342:d715. [PubMed: 21324992]

[45]. Taveras EM, Gortmaker SL, Hohman KH, Horan CM, Kleinman KP, Mitchell K, et al. Randomized controlled trial to improve primary care to prevent and manage childhood obesity: the High Five for Kids study. Archives of Pediatrics & Adolescent Medicine. 2011; 165(8):714–722. [PubMed: 21464376]

[46]. Zurovac D, Sudoi RK, Akhwale WS, Ndiritu M, Hamer DH, Rowe AK, et al. The effect of mobile phone text-message reminders on Kenyan health workers' adherence to malaria treatment guidelines: a cluster randomised trial. The Lancet. 2011; 378(9793):795–803.

[47]. Stiell IG, Nichol G, Leroux BG, Rea TD, Ornato JP, Powell J, et al. Early versus later rhythm analysis in patients with out-of-hospital cardiac arrest. New England Journal of Medicine. 2011; 365(9):787–797. [PubMed: 21879896]

[48]. LaBella CR, Huxford MR, Grissom J, Kim KY, Peng J, Christoffel KK. Effect of neuromuscular warm-up on injuries in female soccer and basketball athletes in urban public high schools: cluster randomized controlled trial. Archives of Pediatrics & Adolescent Medicine. 2011; 165(11):1033–1040. [PubMed: 22065184]

[49]. Levine DA, Funkhouser EM, Houston TK, Gerald JK, Johnson-Roe N, Allison JJ, et al. Improving Care After Myocardial Infarction Using a 2-Year Internet-Delivered Intervention: The Department of Veterans Affairs Myocardial Infarction–Plus Cluster-Randomized Trial. Archives of Internal Medicine. 2011; 171(21):1910–1917. [PubMed: 22123798]

[50]. Stuart EA, Perry DF, Le HN, Ialongo NS. Estimating intervention effects of prevention programs: Accounting for noncompliance. Prevention Science. 2008; 9(4):288–298. [PubMed: 18843535]

[51]. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. Journal of the American Statistical Association. 1996; 91(434):444–455.

[52]. Ding P, Lu J. Principal stratification analysis using principal scores. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2017; 79(3):757–777. DOI: 10.1111/rssb. 12191

[53]. Joffe MM, Ten Have TR, Brensinger C. The compliance score as a regressor in randomized trials. Biostatistics. 2003; 4(3):327–340. [PubMed: 12925501]

[54]. Jo B, Stuart EA. On the use of propensity scores in principal causal effect estimation. Statistics in Medicine. 2009; 28(23):2857–2875. [PubMed: 19610131]

[55]. Hirano K, Imbens GW, Rubin DB, Zhou XH. Assessing the effect of an influenza vaccine in an encouragement design. Biostatistics. 2000; 1(1):69–88. [PubMed: 12933526]

[56]. Stuart EA, Jo B. Assessing the sensitivity of methods for estimating principal causal effects. Statistical Methods in Medical Research. 2015; 24(6):657–674. [PubMed: 21971481]

Box 1

Causal methods for obtaining adherence-adjusted treatment estimates

There exists many statistical methods estimating causal treatment effects in randomised trials. They target different estimands, i.e. quantities of interest in a defined population, and they also differ in the assumptions required to guarantee identification and unbiased estimation. Interested readers are directed to introductory materials by Bellamy et al. [15], Baiocchi et al. [18] and Stuart et al. [50].

A key idea is that of potential outcomes, i.e. the outcome that would have been observed had the randomised allocation been different. Likewise, the potential treatment received is the treatment that individuals/clusters would have received had their randomised allocation been different. Assuming all-or-nothing compliance, the most common assumptions are:

(i)   **Stable Unit Treatment Value Assumption**: the potential outcomes of the i-th individual are unrelated to the treatment status of all other individuals (known as *no interference*). In addition, we assume *consistency*: for those who actually received treatment level z, the observed outcome is the potential outcome corresponding to that level of treatment. In cluster trials, Stable Unit Treatment Value Assumption as above is unlikely to hold. Instead, we may assume that *no interference* holds at the cluster level, i.e. the potential outcome of an individual is unrelated to the treatment status of individuals in different clusters, but may depend on those within the same cluster [4, 17].

(ii)   **Ignorability of the treatment assignment**: Randomised allocation is independent of unmeasured confounders (conditional on measured covariates) and the potential outcomes.

(iii) **Instrument relevance**: The random allocation predicts treatment received.

(iv)   **Exclusion restriction**: The random allocation cannot affect the outcomes directly.

(v)   **Monotonicity**: There are no *defiers*, i.e. individuals who receive treatment if and only if they are not randomised to it. Generalisations of these assumptions to cluster trials settings can be found in Schochet [4].

Here we concentrate on methods that target the complier average causal effect.

•   **Principal stratification** [14]: Under assumptions (i)-(v), each individual may be grouped into a compliance *principal stratum*, which is a latent class, and can be thought of as a baseline covariate.

   a   Never-takers receive no active treatment, regardless of their randomised treatment;

   b   Compliers receive the active treatment if and only if they are randomised to it;

c   Always-takers, who receive the active treatment, regardless of their randomised treatment.

The complier average causal effect can then be identified from the observed data. Estimation for the principal stratification is based on a mixture of distributions across compliance strata. Extensions to cluster trials are possible, by using multilevel mixture models, in either a Bayesian [16] or likelihood approaches [17].

- **Instrumental variables**: Under assumptions (i)-(v) Angrist et al. [51] showed that the instrumental variable estimand is the intention-to-treat effect in compliers. This is then usually estimated using two-stage least squares. To account for the clustering, it has been recommended to use two-stage least squares using Huber-White variance estimator [18].

- **Principal scores**: While this method is based on principal stratification, it differs from the previous version, because it does not assume exclusion restriction, but instead assumes **principal ignorability**[52]: the observed covariates are sufficient for identifying principal stratum membership [53, 54]. The compliance or principal score is a function solely of pre-randomisation covariates. Similar to the use of propensity scores, this method uses baseline covariates to model principal stratum membership. Once principal scores are obtained, the complier average causal effect is estimated using either matching or weighting, as it is usually done in propensity scores literature. Because it does not assume exclusion restriction, this method is attractive when this assumption is believed to not hold, but the principal ignorability assumption is more plausible.

The choice of causal estimation methods depends on the estimand that investigators are interested in and whether the trial setting supports the plausibility of the underlying causal assumptions. Many of these assumptions are untestable, and often their plausibility is questionable. There are several options to study the sensitivity to departures from these assumptions. For example, when exclusion restriction does not hold, researchers could use the principal scores methods. Alternatively, a Bayesian parametric model can use priors on the non-zero direct effect of randomisation on the outcome for identification on the mixture models used in the principal stratification approach [55]. In the frequentist instrumental variables framework, such modelling is also possible, see Baiocchi et al. [18] for a tutorial on how to perform sensitivity analyses to departures from exclusion restriction and monotonicity assumptions. See also [56] for a comparison of the sensitivity to departures from assumptions of principal stratification under exclusion restriction and principal scores under principal ignorability.

Box 2

Guidelines for analysing and reporting cluster randomised trials with non-adherence to treatment

1. Report how adherence to treatment is defined and measured. Describe adherence at the cluster and individual level. If dichotomised, justify the choice of threshold made. These choices should be pre-specified in the protocol [8].

2. Where there is interest in the causal treatment effect, this should be stated clearly in the trial protocol, prior to data collection.

3. Adherence measures should be collected alongside other trial data.

4. Report the number of clusters and individuals that received the intended treatment in each trial arm [33].

5. Details of the planned causal analyses should be included in the statistical analysis plan, in advance of receiving the data.

6. Efforts should be made in the statistical analysis to reduce any bias introduced by the fact that treatment received may be associated with other variables affecting the outcome.

7. Choose a statistical method that relies on a set of plausible assumptions for the trial at hand and interpret non-adherence adjusted analyses as explanatory.

8. Discuss the assumptions necessary for the chosen analysis method to result in unbiased causal treatment effect estimates and their plausibility in the context of the cluster trial being analysed and reported.

9. In particular, the use of per protocol analysis must be supported by an explanation of why it is reasonable to assume that the group of participants and clusters who did and did not deviate from their allocated treatment are equivalent. If a set of baseline or post-randomisation variables available is believed to be sufficient to adjust for the confounding, a "modified" per protocol analysis may be valid [10, 11].

10. If clusters or individuals are excluded from analyses, describe if the fraction excluded is similar between arms, and that the included groups were comparable at baseline [7].

11. Use a method that accounts for clustering adequately. Principal stratification can be used to estimate the complier average causal effect while accounting for clustering ; alternatives include multilevel mixture models [17] and Bayesian hierarchical models [14]. Alternatively, instrumental variable methods can use sandwich variance estimation, which is robust to clustering [18].

12. Sensitivity analyses should be considered when the assumptions necessary for the primary causal analysis are likely to be violated [18, 16].

13. A discussion of potential bias introduced by assumptions' violations in any of the causal analyses should be included in the published report.
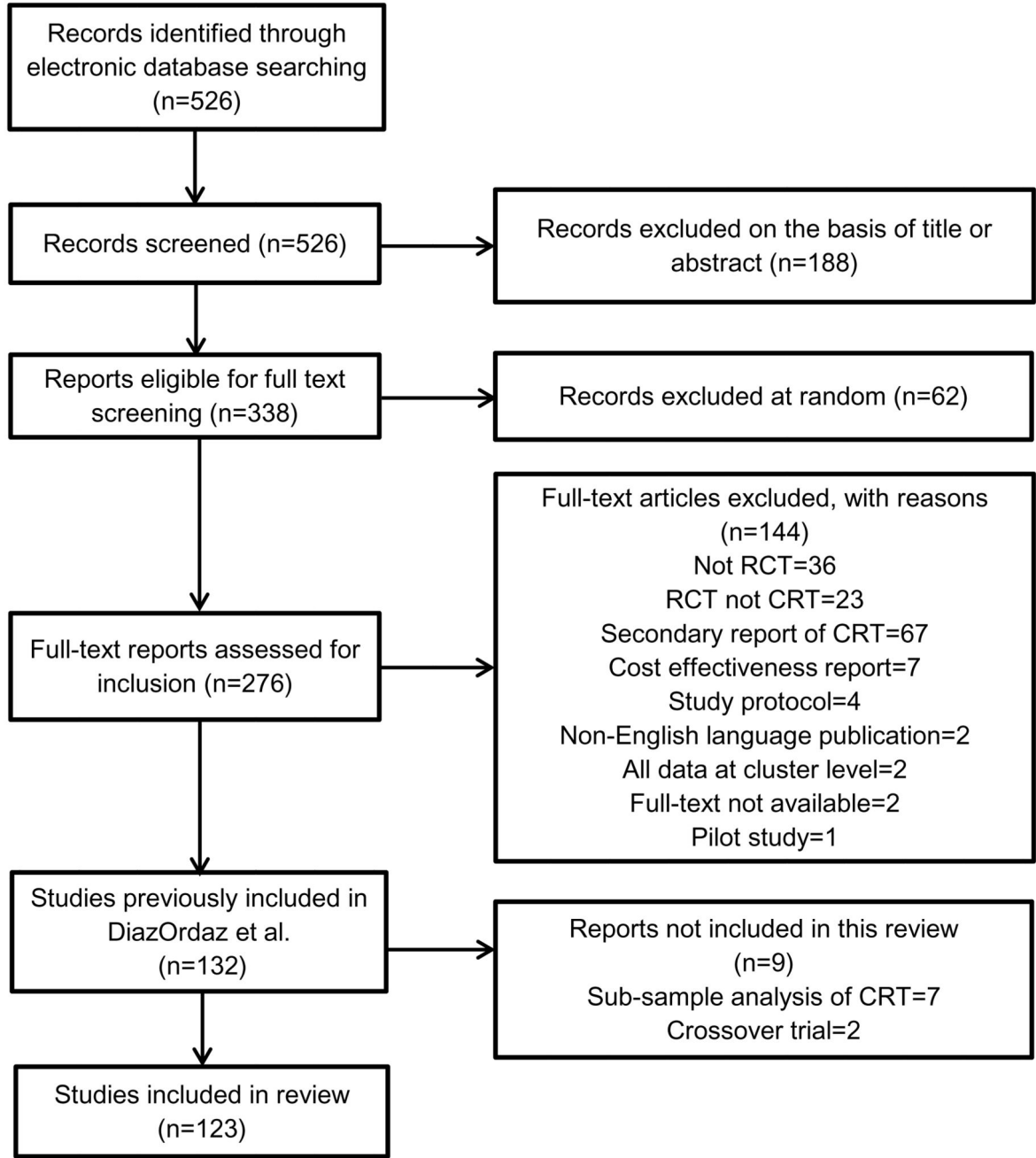
Figure 1.
Flow diagram of the identification process for the sample of 123 cluster trials included in this review

Table 1

Characteristics of the cluster trials included in this review

Characteristics	Included trials (123 cluster trials)
Type of intervention, n (%)	
Healthcare practice	63 (51.2)
Lifestyle changes	25 (20.3)
Educational	27 (22.0)
New drug	5 (4.1)
Vaccination/screening	3 (2.4)
Type of control intervention, n (%)	
Standard practice	94 (76.4)
Active control	27 (22.0)
Placebo	2 (1.6)
Primary outcome, n (%)	
Continuous	65 (52.8)
Binary	57 (46.4)
Categorical	1 (0.8)
Investigation of adverse events, n (%)	12 (9.8)
Number of trial arms, n (%)	
2	106 (86.2)
3-4	17 (13.8)
Level of intervention, n (%)	
Cluster level	65 (52.8)
Individual level	58 (47.2)
Unit of analysis, n (%)	
Clusters	6 (4.9)
Individuals	117 (95.1)
Length of intervention, n (%)	
One-off	5 (4.1)
Short term	35 (28.4)
Long term	83 (67.5)
Clusters randomised per arm, Median (1st-3rd quartiles)	12 (7-24)
Range	2-199
Cluster size, Median (1st-3rd quartiles)[a]	27 (10-65)
Range[a]	2-14350
Primary analysis population, n (%)	
Intention-to-treat	119 (96.8)
Per protocol/as treated	4 (3.2)

[a]Based on the average number of individuals per cluster reported in each trials.

Table 2

Analysis methods stratified by unit of analysis

	Cluster-level analysis	Individual-level analysis
Methods of analysis	6 (100)	117 (100)
Generalized estimating equations	-	27 (23.1)
Mixed effects models	-	55 (47.0)
Repeated measures analysis of variance	-	5 (4.3)
Generalized linear model with sandwich variance	-	16 (13.7)
Chi square accounting for clustering	-	1 (0.8)
Survival analysis accounting for clustering	-	1 (0.8)
Other methods ignoring clustering [a]	-	12 (10.3)
Weighted regression [b]	1 (16.7)	-
Other methods without weighting [a]	5 (83.3)	-
Methods of analysis when non-adherence was addressed	1 (100)	18 (100)
Generalized estimating equations	-	4 (22.2)
Mixed effects models	-	9 (50.0)
Generalized linear model with sandwich variance	-	4 (22.2)
Poisson regression ignoring clustering [c]	-	1 (5.6)
Unweighted t-test [d]	1 (100)	-

The numbers in brackets are the column percentages.

[a] Generalized linear model, analysis of variance, analysis of covariance, T-test, Mann-Whitney U test, Chi square test.

[b] Number of events (cluster size) used as weights (Buttha et al [36]). The use of weights is applicable when cluster-level summaries analysis is performed while accounting for clustering may be required for individual-level analysis.

[c] t-test with multiple testing adjustment but ignoring clustering was applied to perform a per protocol analysis at individual-level (Neuzil et al. [23]).

[d] Per protocol analysis with unweighted t-test comparing rates at cluster level (Tagbor et al. [28]).

Table 3

Reporting of non-adherence by length of intervention, randomised arm and level of adherence.

	One-off	Short term	Long term	Total
Reporting of any non-adherence, n (%)	**5 (100)**	**35 (100)**	**83 (100)**	**123 (100)**
Non-adherence reported in both active and control groups	-	4 (11.4)	16 (19.3)	20 (16.2)
Non-adherence reported in active group only	2 (40.0)	15 (42.9)	19 (22.9)	36 (29.3)
Non-adherence reported in control group only	-	-	-	-
Full adherence reported	-	1 (2.9)	4 (4.8)	5 (4.1)
Not reported	3 (60.0)	11 (31.4)	36 (43.4)	50 (40.6)
Unclear	-	4 (11.4)	8 (9.6)	12 (9.8)
Trials with adherence at cluster level	**2 (100)**	**21 (100)**	**42 (100)**	**65 (100)**
Non-adherence reported in both active and control groups	-	-	1 (2.4)	1 (1.5)
Non-adherence reported in active group only	-	9 (42.9)	5 (11.9)	14 (21.5)
Non-adherence reported in control group only	-	-	-	-
Full adherence reported	-	1 (4.7)	3 (7.1)	4 (6.2)
Not reported	2 (100)	9 (42.9)	29 (69.1)	40 (61.6)
Unclear	-	2 (9.5)	4 (9.5)	6 (9.2)
Trials with adherence at individual level	**3 (100)**	**14 (100)**	**41 (100)**	**58 (100)**
Non-adherence reported in both active and control groups	-	4 (28.6)	15 (36.6)	19 (32.8)
Non-adherence reported in active group only	2 (66.7)	6 (42.9)	14 (34.1)	22 (38.0)
Non-adherence reported in control group only	-	-	-	-
Full adherence reported	-	-	1 (2.4)	1 (1.7)
Not reported	1 (33.3)	2 (14.3)	7 (17.1)	10 (17.2)
Unclear	-	2 (14.3)	4 (9.8)	6 (10.3)
Percentage of non-adherence at cluster level[a]				
Total number of trials reporting non-adherence, n (%)	**-**	**9 (100)**	**6 (100)**	**15 (100)**
Trials reporting % of non-adherence in active group, n (%)	-	2 (22.2)	3 (50.0)	5 (33.3)
Median % of non-adherence in active group[b]	-	37.4 (30–44.8)	50 (33–80)	44.8 (33–50)
Percentage of non-adherence at individual level				
Total number of trials reporting non-adherence, n (%)	**2 (100)**	**10 (100)**	**29 (100)**	**41 (100)**
Trials reporting % of non-adherence in active group, n (%)	2 (100)	7 (70.0)	21 (72.4)	30 (73.2)
Median % of non-adherence in active group[b]	16.5 (0.5–32.4)	13.7 (5.3–25)	15 (10–20)	15 (9–24)
Total number of trials reporting non-adherence, n (%)	**2 (100)**	**10 (100)**	**29 (100)**	**41 (100)**
Trials reporting % of non-adherence in control group, n (%)	-	3 (30.0)	11 (37.9)	14 (34.1)
Median % of non-adherence in control group[b]	-	8.1 (1.7–32)	8.2 (3.4–20)	8.2 (3.4–20)
Total number of trials, n (%)	**5 (100)**	**35 (100)**	**83 (100)**	**123 (100)**
Flow chart with adherence information	1 (20.0)	4 (11.4)	6 (7.2)	11 (8.9)
Flow chart without adherence information	1 (20.0)	19 (54.3)	65 (78.3)	85 (69.1)

	One-off	Short term	Long term	Total
No flow char	3 (60.0)	12 (34.3)	12 (14.5)	27 (22.0)
Adherence type, n (%)[c]	**2 (100)**	**20 (100)**	**39 (100)**	**61 (100)**
Binary adherence	2 (100)	14 (70.0)	28 (71.8)	44 (72.1)
Continuous adherence	-	2 (10.0)	3 (7.7)	5 (8.2)
Unclear	-	4 (20.0)	8 (20.5)	12 (19.7)
Trials using adherence-adjusted methods, n (%)	**1 (100)**	**4 (100)**	**14 (100)**	**19 (100)**
Per protocol	1 (100)	4 (100)	10 (71.4)	15 (78.9)
As treated	-	-	4 (28.6)	4 (21.1)

[a]No report provided non-adherence % at cluster level in the control group.

[b]Numbers in brackets are the 1st and 3rd quartiles.

[c]Total number of trials reporting non-adherence or full adherence.

Table 4

Details of the adherence-adjusted analyses performed.

Study	Reason	Type	Differences in inference
Per protocol			
Acolet et al.[37]	Exploratory	Binary	PP[*] not shown, stated similar to ITT[*]
Auger et al. [38]	Unclear	Binary	ITT not done
Beer et al. [39]	Unclear	Binary	Evidence of effect with PP, but not with ITT
Bickman et al.[40]	Unclear	Binary	No change
Boorsma et al.[24][c]	Unclear	Binary	Evidence of effect with PP, but not with ITT
Cooke et al. [41]	Unclear	Binary[a]	ITT not done
Cutrer et al. [42]	Unclear	Binary	ITT not done
Dangour et al. [25][c]	Exploratory	Binary	No change
Estrada et al.[43]	Unclear	Binary	No change
Luoto et al.[26][c]	Unclear	Binary	No change
Neuzil et al.[23][c,d]	Safety	Binary	No change
Smith et al. [44]	Additional analyses	Binary	No change
Tagbor et al. [28][d]	Unclear	Binary[b,d]	Evidence of effect with PP, but not with ITT
Taveras et al. [45]	Unclear	Binary	No change
Zurovac et al.[46]	Unclear	Binary	No change
As-treated			
Stiell et al. [47]	Additional analyses	Binary	No change
Zamorano et al. [27][c]	Efficacy	Binary	ITT not done
LaBella et al. [48]	Unclear	Continuous	Evidence of effect with ITT, but not with AT
Levine et al. [49]	Unclear	Continuous	AT not shown

[*] PP: Per protocol analysis, ITT: Intention to treat analysis

[a] The threshold chosen to define the binary non-adherence was based on a previous study.

[b] All possible definitions of binary adherence explored (> 1 dose,> 2 doses and full exposure)

[c] Carried out a safety outcome analysis.

[d] Failed to adjust for clustering in the analysis.