A Review of Assumed and Reported Intracluster Correlations in Cluster Randomized Trials

CURRENT STATUS: POSTED

Xiaoran Han
Chinese University of Hong Kong

Jiaye Lin
Chinese University of Hong Kong

Jinjing Xu
Chinese University of Hong Kong

Maggie Wang
Chinese University of Hong Kong

Benny Zee
Chinese University of Hong Kong

Ka Chun Chong
Chinese University of Hong Kong

marc@cuhk.edu.hk Corresponding Author
ORCiD: https://orcid.org/0000-0001-5610-1298

DOI:
10.21203/rs.2.12500/v1

SUBJECT AREAS
Internal Medicine Integrative & Complementary Medicine

KEYWORDS
Cluster randomized trials, CRT, intraclass correlation, ICC, cluster effect
Abstract
Background Cluster randomized trials (CRTs) are widely adopted in health and primary care research. However, the cluster effect needs to be taken into account appropriately in the design and analysis of CRTs. The objectives of this study were (i) to review the reporting of intracluster correlations in CRTs; and (ii) to evaluate whether the assumed intracluster correlation measures in sample size planning are consistent with those obtained in the analysis.

Methods The Aggregate Analysis of ClinicalTrials.gov database was searched to identify CRTs registered between January 1, 2004 and March 27, 2016. The selected CRTs with accessible publications were screened according to eligibility criteria.

Results Of the 281 CRTs identified, the percentage of studies accounting for cluster effect increased annually. A total of 183 studies accounted for clustering in sample size estimation, among them 43% of CRTs adopted the intraclass correlation coefficient (ICC) but the exact estimated value of ICC was provided in only 26% of the included studies. In different intervention types, there were no statistically significant differences between the assumed and reported values of ICC (all p-values >0.05).

Conclusion Although the difference between the values of ICC assumed in sample size planning and that reported in the analysis was not statistically significant, deficiencies in CRTs are still common, such as low rates of considering cluster effect in sample size and reporting intracluster correlation estimates. We also suggest that researchers ought to be familiar with the properties of statistical approaches to improve the analysis of CRTs. Thus, more recommendations and guidelines such as the CONSORT statement for CRTs should be suggested to researchers.

Introduction
Cluster randomized trials (CRTs) or group randomized trials involve interventions that are randomly assigned to groups or clusters of participants instead of participants individually [1]. Allowing the study of interventions that cannot be implemented at the individual level and the control of experimental contamination across individuals, CRTs are increasingly adopted in research in education, lifestyle modification, and public health in recent years [2]. A principal difference between
CRTs and individually randomized trials is that participants within the same cluster tend to show more similar responses than do those in different clusters [3]. To maintain an equivalent statistical power to that in individually randomized trials, the number of participants in CRTs needs to be increased. One of the pivotal determinants of this increase is the intracluster correlation, which is the ratio of the between-cluster variance to the total variance of an outcome variable. Intracluster correlation quantifies how strongly individuals within the same cluster resemble each other and can be reported using the intracluster correlation coefficient (ICC), often denoted by $\rho$, and the coefficient of variation (CV) [4]. The Consolidated Standards of Reporting Trials (CONSORT) statement, published in 1996 and extended to CRTs in 2004, incorporates the reporting of intracluster correlation as an index to judge the quality and reliability of CRTs [5]. Meanwhile, several journals require the reporting of intracluster correlation as a precondition for publication to facilitate appropriate interpretation of the results between two levels of inferences [6].

Previous reviews of published CRTs have found that most studies do not take the cluster effect into account, although there has been an increasing trend of articles including adequate planning and analysis of CRTs [7-9]. A review of cancer CRTs between 2002 and 2006 found that only 18 (24%) of the 75 articles reviewed accounted for the cluster effect adequately in their sample size calculations [10]. Petros et al. [11] reviewed all CRTs conducted in sub-Saharan Africa and reported that only 10 (20%) of 51 articles accounted for clustering in sample size or power calculations, and only 19 (37%) took clustering into account in their analyses. Although a few CRTs reported their intracluster correlation outcomes, they didn’t compare the difference between assumed and reported intracluster correlation measures.

Furthermore, most literature reviews are based on published papers that were searched using the keywords “cluster randomized.” As a result, publications that do not mention this term in their text might be omitted. Since there is no specific item to distinguish CRTs from other trials in ClinicalTrials.gov, a major international registry for clinical trials, this systematic review was conducted using a relational database of ClinicalTrials.gov, the Aggregate Analysis of ClinicalTrials.gov (AACT), to avoid missing eligible trials because of keyword search. The AACT is an
aggregate database restructured by the Clinical Trials Transformation Initiative, containing information about all CRTs registered in ClinicalTrials.gov from February 2000 to March 2016 and including semi-annual updates.

In this study, we primarily evaluated whether the assumed ICC, a major intracluster correlation measure in sample size planning was consistent with that obtained in analysis. We hypothesized that the ICC in sample size planning should not be under- or over-estimated. Moreover, we determined the proportion of CRTs accounting for the intracluster correlation and profiled the intracluster correlation measures by different cluster and intervention types. The findings of this review may provide a valuable reference for researchers planning CRTs.

Method

Eligibility criteria

All completed CRTs registered in ClinicalTrials.gov between January 1, 2004 and March 27, 2016 were screened according to following criteria:

1. Quasi-randomized controlled trials and trials that further allocated participants within clusters were removed.

2. Studies without available publications were also excluded, and so were those published in a language other than English.

3. Articles were also excluded if they contained “cluster” with a different meaning (such as gene cluster, cluster of differentiation, cluster of symptoms, and cluster headache).

4. Secondary analyses of previous publications, case reports, protocols, and pilot studies were also excluded.

Search strategy

A total of 211,437 records were registered in ClinicalTrials.gov from January 1, 2004 to March 27, 2016. These were obtained through the AACT and downloaded in SAS CPORT format. CRTs were identified using the search term “cluster” in variables “brief_title”, “official_title”, “brief_summary”,
“detailed_description”, or “keywords”.

**Data extraction**

We reviewed each potentially eligible study to determine whether it satisfied the eligibility criteria. From each included study, we extracted the following information: author, journal, year of publication, title, country or countries of recruitment, sample size, trial phase, number of trial arms, type of intervention, number of clusters, and type of cluster. We also extracted whether 1) the study took into account the cluster effect (e.g., ICC) in sample size estimation; 2) the study reported the cluster effect; 3) the study accounted for the cluster effect in analysis; 4) the analysis was conducted at aggregate or individual level. One of the authors extracted data on all items. The other author assessed all questionable items. In case of disagreement, consensus was reached after discussion.

**Data synthesis**

Results were summarized using frequencies and percentages for categorical variables. We compared the difference between the assumed and reported ICCs using Wilcoxon signed-rank test depended on the sample size. The ICCs were summarized using medians and ranges. Statistical analyses were performed using SAS 9.4 (SAS Institute, Cary, NC, USA). All p-values reported are two-tailed.

**Results**

As summarized in Figure 1, a total of 211,437 records were registered in ClinicalTrials.gov from January 1, 2004 to March 27, 2016. Of these, 746 were obtained after computational screening with the search term “cluster” in either variables “brief_title,” “official_title,” “brief_summary,” “detailed_description,” or “keywords” of the AACT database. A total of 218 records were excluded after abstract screening and 147 records without publications were also removed. After full-text screening, 281 CRTs were included in the present review.

**Overview of CRT publications**

Table 1 shows the characteristics of included CRTs, such as cluster type, sample size, and number of arms and clusters. More than half of the selected CRTs were conducted in clinical settings, i.e., 59.1% (166 CRTs) of all trials. Of the CRTs in clinical settings, healthcare facility was the most frequent cluster type (96 CRTs, 34.2%). Of the CRTs in non-clinical settings, those conducted in schools
accounted for the largest number of trials (27 CRTs, 9.6%). The majority of trials had only two study arms. With the use of clusters, the median sample size was 1,262 (range: 45–170,932) and 1,193 (range: 69–118,588) in clinical and non-clinical settings, respectively. For clinical and non-clinical settings, the median number of clusters ranged from 20 to 50 and 18 to 99, respectively. Besides the common types of cluster listed in Table 1, several CRTs used other kinds of clusters, such as blood donation units, maternity huts, and churches, as randomization units.

Figure 2 describes the number of CRTs considering and reporting intracluster correlation measures between 2004 and 2016. The total number of published CRTs was found to increase with time, and the percentage of studies considering the cluster effect in sample size planning was more than 50%, while that for studies providing their estimates of intracluster correlation measures was less than 40%. Although the CONSORT statement has recommended stating the ICC assumption used in sample size estimation and providing cluster effect measures in CRTs since 2004, articles neither considering nor reporting ICC were more numerous than those following the CONSORT recommendation.

**Intracluster correlation measures**

Table 2 summarizes the types of intracluster correlation measures used in sample size calculations and reported in primary results of the included studies. Overall, ICC was the most widely adopted measure for accounting the sample size and reporting the cluster effect, which is in accord with the CONSORT recommendations. About 43% (121) of the included CRTs accounted for cluster effect by assuming a value of ICC in sample size estimation. However, ICC was reported in only in 25.3% (71) of the articles. Besides ICC, the CV was also used in considering the cluster effect in 26 trials (9.3%) and was reported in 5 (1.8%). A total of 36 trials (12.8%) mentioned that the cluster effect was took into account by using other intracluster correlation measures, such as the between-cluster standard deviation, and only 4 (1.4%) of them reported those measures in their results.

**ICC for different intervention types**

Table 3 compares the assumed and reported ICC values between various intervention types. The most common intervention type was behavioral interventions, which represented 65.5% of all
included CRTs. Behavioral CRTs also had the largest proportion of trials using ICC for their sample size calculations (44.0%, 81 trials) and reporting the ICC in their results (30.4%, 56 trials). CRTs of drug interventions had the smallest number of trials using ICC for sample size estimation (25.9%, 7 trials), and dietary intervention CRTs had the lowest rate of reporting ICC outcomes (12.5%, 2 trials). The median values of ICC assumed in sample size planning ranged from 0.03 to 0.05. Several behavior and drug intervention CRTs assumed the values greater than 0.3. The median estimated ICC values reported in the studies ranged from 0.03 to 0.08 among all intervention types. No statistically significant differences between assumed ICC in sample size planning and estimated ICC reported in the studies were found in different intervention types (all p-values >0.05).

**Statistical analysis of cluster effect**

Table 4 summarizes the statistical analysis methods adopted in the included CRTs. Most CRTs (229 trials, 81.5%) considered the cluster effect in their analysis, 78.6% (221 trials) accounted for it at the individual level and 27.1% (67 trials) at the aggregate level. For aggregate level analysis, 29 (10.3%) of the total 67 trials regarded clusters as the analysis unit, and 38 trials (13.5%) only summarized baseline characteristics at the cluster level. The most widely used method to address the cluster effect in the analysis was the mixed or random effect models, which accounted for 38.8% (109 trials) of all CRTs. Besides the mixed/random effect model, the generalized estimating equation (GEE) and hypothesis testing adjusted with inflated variations were two predominant analysis methods accounting for 17.1% (48 trials) and 13.5% (38 trials) of all CRTs, respectively. A small percentage of trials used other methods to address the cluster effect (9 trials, 3.2%), such as fixed effects model and bootstrapping methods.

**Discussion**

Accounting for clustering in the design of CRTs is an important step to prevent a reduction in study power. In this review, we profiled the measures of clustering used in 281 selected CRTs registered in ClinicalTrials.gov from 2004 to 2016. Primarily, we assessed the difference between measure of clustering assumed in sample size planning and that reported in the analysis. This difference was not statistically significant; however, only around 50% of CRTs accounted for the cluster effect in sample
size planning and around 30% to 40% of them reported their estimated intracluster correlation measures in the analysis. Therefore, CRTs with correct analytic methods accounting for the cluster effect following recommendations such as the CONSORT statement are still insufficient in the current literature.

Our review shows that that there was an increase in the number of CRTs accounting for cluster effects and reporting measures of clustering since 2004. Such a trend might be due to the development of specialized software for hierarchical data analysis. Several statistical packages designed as add-ons, such as SAS PROC MIXED, could account for the dependency in nested data when calculating standard errors. General statistical software, such as the advanced statistics module of SPSS and R, have also implemented multilevel modeling [12]. In the meantime, the revision of CRTs guidelines also suggests authors to report intracluster correlation measurers; consequently, the number of these articles grew dramatically after 2010.

Our results indicated an insufficient proportion of trials considering the cluster effect in sample size estimation and there thus remains considerable room for improvement. Although accounting for clustering in CRTs to maintain statistical power is recommended in the CONSORT statement, several authors just simply mentioned that the cluster effect was considered without providing an exact value of intracluster correlation measure or only provided an unclear variation estimates. For example, some studies claimed that “sample size calculation took into account the complexity of the study design (i.e., cluster-randomized)” or “estimation included the accounting for cluster effects” without reporting concrete methods [13, 14]. Moreover, the present study found that the majority of selected publications did not comply with the CONSORT statement and only one third reported their intracluster correlation measures in the results. CRTs ought to report the magnitude of correlation within clusters for their primary outcomes to enable readers to assess whether the original sample size estimation as well as study power is adequate.

In our study, we found that the difference between ICCs for considered in sample size calculation and that reported in the results was not statistically significant. This indicated that most CRTs did not under- or over-assume the cluster effect estimates in their specific trial setting. However, intracluster
correlation measures from previous studies are not always suitable for different trials, since they are highly variable and depend on cluster levels, cluster size, and outcomes. In previous study, it was suggested that the upper confidence interval limit of the ICC estimates be used in sample size determination. However, this approach may on the other hand induce an overestimation of the required sample size, and the redundant sample would require extra resources and enhance the risk of loss to follow-up particularly when involving trial settings across a wide geographical area [15]. Furthermore, to ensure that appropriate sample size estimations are undertook, a sensitivity analysis has been recommended to help determining the required sample size for a range of plausible values of ICC instead of a single value [16, 17].

Our results indicate that a large proportion of authors (>80%) applied statistical methods to consider cluster effects in their CRTs. Although the improvement of existing methods and development of novel methods provide authors with more options, unawareness of the limitations of various approaches may lead to biased results [18]. For example, it is well documented that the application of GEE for CRTs with few clusters or small cluster size would underestimate the standard error of intervention effect [19]. Among the selected studies, a CRT evaluating the performance of a computerized malnutrition-screening system adopted GEE for analysis despite having only six clusters [20]. Applying such a statistical method increases the likelihood of obtaining a significant intervention effect but may inflate the type I error rate [21]. Previous comparison studies found that for both binary and continuous outcomes, mix-effect models are more efficient in terms of attaining a larger coverage of power, even if the number of clusters and cluster size are highly variable [22]. Conversely, mix-effect models are less accurate and precise than GEE when considering the presence of missing values [23]. Therefore, researchers ought to be familiar with valid approaches to improve the analysis of their CRTs.

This study has several major limitations. Previous studies found that about 30–50% of all completed clinical trials are not published in academic journals [24]. As only CRTs with publications were extracted from the AACT for our analysis, a proportion of unpublished trials is likely to be missing. In addition to this publication bias, selective reporting of outcomes in articles is another potential
limitation to our study. Among trials evaluated by Chan et al. [25], 92% had at least one incompletely reported efficacy outcome and 81% at least one harm outcome. Most of these were motivated by the absence of statistical significance or clinical importance, and journal space limitations. Any missing primary outcomes would thus result in bias in our study.

Conclusions
In conclusion, we present a systematic review profiling intracluster correlation measures in 281 selected CRTs from the AACT database of ClinicalTrials.gov. Our study showed that the assumed ICC, a common intracluster correlation measure in sample size planning, was generally consistent with that obtained in the analysis. However, deficiencies are still common, such as low rates of accounting cluster effect in sample size calculation and reporting intracluster correlation measure estimates. More recommendations and guidelines such as the CONSORT statement for CRTs should thus be suggested to researchers.

List Of Abbreviations
CRT Cluster randomized trial
ICC Intraclass correlation coefficient
AACT Aggregate Analysis of ClinicalTrials.gov
GEE Generalized estimating equation

Declarations
Ethics approval and consent to participate
No formal ethical approval and consent to participate are required due to no collection of primary data.

Consent for publication
Not Applicable

Availability of data and material
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Funding
The work is supported by National Natural Science Foundation of China (Grant number: 31871340).
Authors’ contributions
KCC contributed in study concept construction, interpretation, and manuscript writing. XH, JL, JX contributed in data analysis and interpretation. XH, JL, JX, and BCY contributed in review of manuscript. All authors read and approved the final manuscript.

Acknowledgement
We thank Dr. Joey Yang providing helps in the review process.

Reference
1. Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. Int J Epidemiol. 2006;35(5):1292–300.
2. Wu S, Wong WK, Crespi CM. Maximin optimal designs for cluster randomized trials. Biometrics. 2017;73(3):916–26.
3. Simpson JM, Klar N, Donnor A. Accounting for cluster randomization: a review of primary prevention trials, 1990 through 1993. Am J Public Health. 1995;85(10):1378–83.
4. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. Am J Public Health. 2004;94(3):416–22.
5. Campbell MK, Elbourne DR, Altman DG, group C. CONSORT statement: extension to cluster randomised trials. BMJ. 2004;328(7441):702–8.
6. Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group C. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012;345:e5661.
7. Varnell SP, Murray DM, Janega JB, Blitstein JL. Design and analysis of group-randomized trials: a review of recent practices. Am J Public Health. 2004;94(3):393–9.
8. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. BMJ. 2003;327(7418):785–9.
9. Bowater Rj, Abdelmalik SM, Lilford RJ. The methodological quality of cluster randomised controlled trials for managing tropical parasitic disease: a review of trials published from 1998 to 2007. Trans R Soc Trop Med Hyg. 2009;103(5):429–36.
10. Murray DM, Pals SL, Blitstein JL, Alfano CM, Lehman J. Design and analysis of group-randomized trials in cancer: a review of current practices. J Natl Cancer Inst. 2008;100(7):483–91.
11. Isaakidis P, Ioannidis JP. Evaluation of cluster randomized controlled trials in sub-Saharan Africa. Am J Epidemiol. 2003;158(9):921–6.

12. Korendijk EJH HJ, editor A SAMPLING OF ANALYSES AND SOFTWARE USE FOR CLUSTER RANDOMIZED TRIALS OVER THE LAST DECADE. Data and context in statistics education: Towards an evidence-based society Proceedings of the 8th International Conference on Teaching Statistics; 2010 Jul; Ljubljana, Slovenia.

13. Stiell IG, Clement CM, Grimshaw JM, Brison RJ, Rowe BH, Lee JS, et al. A prospective cluster-randomized trial to implement the Canadian CT Head Rule in emergency departments. CMAJ. 2010;182(14):1527–32.

14. Hanson LC, Zimmerman S, Song MK, Lin FC, Rosemond C, Carey TS, et al. Effect of the Goals of Care Intervention for Advanced Dementia: A Randomized Clinical Trial. JAMA Intern Med. 2017;177(1):24–31.

15. Donner A KN. Design and analysis of cluster randomization trials in health research. London: Arnold; 2000.

16. Campbell MK, Mollison J, Grimshaw JM. Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. Stat Med. 2001;20(3):391–9.

17. Feng Z, Grizzle JE. Correlated binomial variates: properties of estimator of intraclass correlation and its effect on sample size calculation. Stat Med. 1992;11(12):1607–14.

18. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. Am J Public Health. 2004;94(3):423–32.

19. Yotebieng M, Labbok M, Soeters HM, Chalachala JL, Lapika B, Vitta BS, et al. Ten Steps to Successful Breastfeeding programme to promote early initiation and exclusive breastfeeding in DR Congo: a cluster-randomised controlled trial. Lancet Glob Health. 2015;3(9):e546–55.

20. Duclos A, Touzet S, Restier L, Occelli P, Cour-Andlauer F, Denis A, et al. Implementation of a computerized system in pediatric wards to improve nutritional care: a cluster randomized trial. Eur J Clin Nutr. 2015;69(7):769–75.

21. Barker D, D’Este C, Campbell MJ, McElduff P. Minimum number of clusters and comparison of
analysis methods for cross sectional stepped wedge cluster randomised trials with binary outcomes: A simulation study. Trials. 2017;18(1):119.

22. Chu R, Thabane L, Ma J, Holbrook A, Pullenayegum E, Devereaux PJ. Comparing methods to estimate treatment effects on a continuous outcome in multicentre randomized controlled trials: a simulation study. BMC Med Res Methodol. 2011;11(1):21.

23. Ma J, Raina P, Beyene J, Thabane L. Comparison of population-averaged and cluster-specific models for the analysis of cluster randomized trials with missing binary outcomes: a simulation study. BMC Med Res Methodol. 2013;13(1):9.

24. Lampert A, Hoffmann G F, Ries M. Ten years after the International Committee of Medical Journal Editors’ Clinical Trial Registration Initiative, one quarter of phase 3 pediatric epilepsy clinical trials still remain unpublished: a cross sectional analysis. PloS one. 2016;11(1): e0144973.

25. Chan A W, Hróbjartsson A, Haahr M T, Gøtzsche P C, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. Jama. 2004;291(20): 2457–65.

Tables

Table 1. Characteristics of cluster randomized trials by cluster types (N=281)
Table 2. Assumed and reported intracluster correlation measures in CRTs (N=281)

| Type of ICC measures                  | Number of CRTs (%) | Accounted in sample size | Reported in analysis |
|---------------------------------------|--------------------|--------------------------|----------------------|
| ICC                                   | 121 (43.0)         | 71 (25.3)                |                      |
| CV                                    | 26 (9.3)           | 5 (1.8)                  |                      |
| Other intracluster correlation measures *| 36 (12.8)         | 4 (1.4)                  |                      |

ICC: intracluster correlation coefficient; CV: coefficient of variation;

*Other intracluster correlation measures: design effect, between cluster standard deviation...

Table 3. The planned and reported ICCs by different intervention types (N=281)
| Intervention type | Number of publications (%) | Number of pairs | Number of ICC accounted in sample size (%) | ICC accounted in sample size, median (range) | Number of ICC reported (%) | ICC reported, median (range) | p-value* |
|-------------------|-----------------------------|----------------|-------------------------------------------|---------------------------------------------|-----------------------------|----------------------------|----------|
| Behavioral        | 184 (65.5)                  | 41             | 81 (44.0)                                 | 0.05 (0.001-0.5)                            | 56 (30.4)                   | 0.04 (0.01-0.41)            | 0.71     |
| Device and procedure | 54 (19.2)                  | 6              | 17 (31.5)                                 | 0.05 (0.003-0.1)                            | 8 (14.8)                    | 0.03 (0.007-0.27)           | 0.07     |
| Dietary Supplement | 16 (5.7)                    | 2              | 7 (43.8)                                  | 0.03 (0.006-0.05)                           | 2 (12.5)                    | 0.08 (0-0.16)               | /        |
| Drug              | 27 (9.6)                    | 5              | 7 (25.9)                                  | 0.05 (0.001-0.3)                            | 7 (25.9)                    | 0.07 (0.003-0.29)           | 0.89     |
| Total publications | 281                        | 54             | 112 (39.9)                                | 0.05 (0.001-0.5)                            | 73 (26.0)                   | 0.04 (0-0.41)               | 0.53     |

ICC: intraclass correlation coefficient; SD: standard deviation.

*p-values were obtained by Wilcoxon signed-rank test.

Table 4. Comparison of different statistical analysis methods for clustering (N=281)

| Method of accounting cluster effect | N (%)       |
|------------------------------------|-------------|
| Total CRTs accounted cluster effect in analysis | 229 (81.5) |
| Individual level analysis          | 221 (78.6)  |
| Aggregate level analysis           | 29 (10.3)   |
| Aggregate level analysis only performed at baseline | 38 (13.5) |
| Method of accounting cluster effect |             |
| Mixed or random effect model       | 109 (38.8)  |
| Generalized estimating equation    | 48 (17.1)   |
| Hypothesis testing adjusted with inflated variations | 38 (13.5) |
| Other methods*                     | 9 (3.2)     |
| Unclassified**                     | 25 (8.9)    |

* Other methods: Fixed effects model, bootstrapping methods, adjusted with geometric means.

**Unclassified: Cannot be distinguished into categories due to limited information e.g. only mention cluster effect was took into account in regression models.

Figures
Figure 1
Flow chart of CRTs extraction
Figure 2

Number of publications with different reporting behavior of intracluster correlation by years