Intra-cluster correlation coefficients for pupil health outcomes estimated from school-based cluster randomised trials

Corresponding author:
Ms Kitty Parker[a]
Email: kp477@exeter.ac.uk

Co-authors:
Dr Michael Nunns[b]
Email: M.P.Nunns@exeter.ac.uk

Dr ZhiMin Xiao[c]
Email: zhimin.xiao@essex.ac.uk

Professor Tamsin Ford[d]
Email: tjf52@medschl.cam.ac.uk

Professor Obioha C Ukoumunne[a]
Email: O.C.Ukoumunne@exeter.ac.uk

Affiliations

[a] NIHR Applied Research Collaboration South West Peninsula, Department of Health and Community Sciences, Faculty of Health and Life Sciences, University of Exeter, Exeter, United Kingdom. Room 2.16, South Cloisters, St Luke's Campus, 79 Heavitree Rd, Exeter, EX1 2LU

[b] Faculty of Health and Life Sciences, University of Exeter, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU, United Kingdom

[c] School of Health and Social Care, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom.

[d] Department of Psychiatry, University of Cambridge, L5 Clifford Allbutt Building, Cambridge Biomedical Campus Box 58, Cambridge, CB2 0AH, United Kingdom

Abstract

Objective

To summarise intra-cluster correlation coefficient (ICC) estimates for pupil health outcomes from school-based cluster randomised trials (CRTs) across world regions and describe their relationship with study design characteristics and context.

Study Design and Setting

School-based CRTs reporting ICCs for pupil health outcomes were identified through a literature search of MEDLINE (via Ovid). ICC estimates were summarised both overall and for different categories of study characteristics.

Results

Two hundred and forty-six articles reporting ICC estimates were identified. The median (interquartile range) ICC was 0.031 (0.011 to 0.08) at the school level (N=210) and 0.063 (0.024 to 0.1) at the class level (N=46). The distribution of ICCs at the school level was well described by the beta and exponential distributions. Besides larger ICCs in definitive trials than feasibility studies, there were no clear associations between study characteristics and ICC estimates.

Conclusion

The distribution of school-level ICCs worldwide was similar to previous summaries from studies in the United States. The description of the distribution of ICCs will help to inform sample size calculations and assess their sensitivity when designing future school-based CRTs of health interventions.

Key words

children; cluster randomised trials; intra-cluster correlation coefficient; public health; randomised trials; schools

Running title: Intra-cluster correlation coefficients for pupil health outcomes estimated from school-based cluster randomised trials

Word count: 3485

What is new?

What is new?

- Few studies outside the United States (US) have summarised intra-cluster correlation coefficients (ICC) for pupil health outcomes and explored their size in relation to design characteristics in school-based cluster randomised trials (CRTs).
- This study collated 260 ICCs for school-related clusters from CRTs worldwide to inform sample size calculation for future trials.
- Two-thirds of school-level ICCs were no greater than 0.05 and three-quarters were under 0.08.
- The ICC distribution was similar to previous summaries from US-based studies and larger for definitive trials than feasibility studies.
- There was little evidence of relationships between ICC estimates and region, health outcome area and educational level.

1. Background

Cluster randomised trials (CRTs) are studies in which clusters (groups) of individuals are randomised to trial arms and outcomes are measured on individuals [1]. CRTs are increasingly undertaken in schools to evaluate public health interventions for improving outcomes of children and adolescents [2-5]. Schools provide a natural environment in which to recruit and study children and deliver interventions to improve their health due to the amount of time they spend there [3, 6, 7]. CRTs may be undertaken in schools because many of the interventions examined in such studies are designed to be delivered to entire schools or classrooms [4], interventions are theorised to affect change at those levels, and randomising clusters (e.g., schools, classes) helps to minimise contamination between trial arms that may otherwise occur if individuals are allocated [1, 6, 7].

CRTs require more participants than individually randomised trials because observations on individuals in the same cluster are usually more similar than those from different clusters [1]. Due to this lack of independence between individuals within clusters, if standard sample size formulae are used this may result in an underpowered study [1]. Correlation between pupils within clusters needs to be accounted for when designing and analysing data from CRTs. In the sample size calculation, this is done by inflating the number of participants required in an individually randomised trial by the design effect (DE):

$$\mathrm{DE} = 1 + (\bar{n} - 1)\rho$$

where $\bar{n}$ is the mean number of participants providing outcome data in each cluster (cluster size) and ρ is the intra-cluster correlation coefficient (ICC) of the outcome [1]. The ICC quantifies the similarity of observations on individuals within clusters. For continuous outcomes, it can be defined as the proportion of the total variability in the outcome that is between clusters as opposed to between individuals within clusters:

$$\mathrm{ICC} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$

where σ_b^2 is the between-cluster variance component and σ_w^2 is the within-cluster variance component [1]. Under this definition the ICC can take values between zero and one. The larger the ICC, the greater

the sample size required. Similarity between participants from the same cluster can also be quantified by the between-cluster coefficient of variation (CV) of the outcome (the ratio of the between-cluster standard deviation to the outcome mean [6]):

$$CV = \frac{\sigma_b}{\mu}$$

where σ_b is the between-cluster standard deviation and μ is the mean outcome across the clusters [6]. The CV can then be incorporated into a modified design effect formula.

In the context of school-based CRTs there are several reasons for the similarity of outcomes between pupils within schools. First, in some countries, pupils and their parents/guardians have some influence regarding the school they attend [8]. Schools are likely to attract pupils with similar characteristics and who are more likely to share similar behaviours [3]. Second, pupils interact in the school setting and may influence the behaviour of their peers in the same schools or classrooms [8]. Finally, the school itself can influence the behaviours of pupils through its physical environment, ethos and polices [9, 10].

At the time of sample size calculation the ICC is usually unknown and specification of a suitable value for the outcome and type of cluster should be informed by the empirical literature [1]. Researchers have reported ICCs for pupil health outcomes to be generally smaller than those for educational outcomes in schools [11-13]. This might be expected given that the main purpose of schools is to provide education [8]. Although ICCs for health outcomes in health care settings are well established, particularly in primary care [1, 14, 15], there is a relative lack of reported estimates in the school setting. Several studies have provided estimates of ICCs from school-based CRTs or surveys for outcomes related to substance use [8, 16-24], nutrition [25-27], physical activity [24, 27-29], and mental health and behaviour [12, 24, 30], but the vast majority of these were undertaken in the United States. It is not known whether these estimates are transferable to other regions and education systems, and outcome areas such as infectious diseases and dental health are not well represented. Furthermore, although patterns in the size of the ICC have been investigated [14, 15, 31-34], little is known about the extent to which ICCs from school-based CRTs differ by study characteristics.

A summary of ICCs for a range of health outcomes in different settings would aid the design of future school-based CRTs by providing plausible values that can be used in sample size calculations. Estimates from CRTs specifically, rather than surveys, are potentially especially relevant as they may better reflect the level of variation in outcomes across the types of schools that tend to participate in health-related trials [1] (p177).

1.1 Objectives

This paper collates and summarises ICC estimates for health outcomes from school-based CRTs and examines the relationship between the size of the ICC and study characteristics.

2. Methods

2.1 Data sources and search methods

A systematic searching approach was used to identify papers reporting ICC estimates from school-based CRTs. MEDLINE (Ovid) was exclusively searched for published peer-reviewed articles reporting school-based CRTs from inception to 18th October 2021. The search strategy was developed based on a strategy by Taljaard and colleagues [35] used to identify CRTs, combined with school-related terms (Table 1).

Search strategy
Terms for Randomised Controlled trials:
1. random:.mp.
2. trial.ab, kw, ti.
Cluster design-related terms:
3. "cluster*".ab, kw, ti.
4. "communit*".ab, kw, ti.
5. group*adj2 random*.ab, kw, ti.
6. 3 OR 4 OR 5
School terms:
7. exp Schools/
8. School*.ab, kw, ti.
9. 7 OR 8
Final search stages:
10. 1 AND 2 AND 6 AND 9
11. 10 limited to English language

Table 1: Search strategy using MEDLINE (through Ovid)

2.2 Inclusion and exclusion criteria

Eligible articles reported school-based studies with a CRT design, including articles reporting baseline data, follow-up outcomes, or secondary data analyses that used the data to address additional questions that were unrelated to the main trial objectives. To be eligible, the article had to report the estimate of an ICC/CV for at least one health outcome measured on pupils. The eligible study population was pupils attending pre-primary, primary, lower secondary and higher secondary educational settings according to the United Nations Educational, Scientific and Cultural Organisation (UNESCO) International Standard Classification of Education (ISCED) system [36]. Eligible clusters were any school-related unit (e.g., schools, classes/classrooms, year groups, teachers). Any intervention(s) were considered. Articles were excluded if they randomised after-school clubs, school-based health centres or childcare centres. Articles that only reported protocol/design information, process evaluations, economic evaluations/cost-effectiveness analyses, statistical analysis plans, commentaries and mediation/mechanism analyses were also excluded.

If more than one publication of the same eligible study was identified, the key study report (index paper) for data extraction was determined by identifying the article that first published the outcomes.

2.3 Sifting and validation

Titles and abstracts were screened by two independent researchers (KP & OU) for eligibility against the inclusion criteria. Any studies for which the reviewers were uncertain of inclusion status were progressed to full text screening. Two independent researchers (KP & OU) examined the full text of each article against the inclusion criteria. Any disagreements over inclusion were resolved through discussion with a third researcher (MN).

2.4 Data extraction

One researcher (KP) extracted data from all included articles, while a second (OU) independently validated the process. Any uncertainty regarding the data extraction was resolved through discussion, or consultation with a third researcher (MN). The information extracted is specified in Table 2.

Aspect	**Information extracted**
Publication details	Author surname, year of publication, title of article, type of study (i.e., definitive or feasibility study).
Setting information	Country in which the study took place (e.g., France), stage of education (e.g., primary, secondary), gender of pupils, age(s) of pupils at baseline.
Study design	Type of cluster unit allocated, cluster unit of ICC/CV estimate.
Sample size information	ICC/CV assumed in the sample size calculation, number of clusters and pupils that provided outcome data, number of classes per school.
Health outcome information	Health area of outcome (e.g., physical activity), outcome description (e.g. amount of moderate-to-vigorous physical activity), outcome type (e.g., continuous, binary), timing (months post-randomisation) at which outcome was measured.
ICC information	ICC/CV of the outcome (and 95% CIs where provided), analytical method used to calculate ICC/CV (e.g., multilevel model [37], marginal model using Generalised Estimating Equations [38]), whether the ICC/CV estimate was pooled across trial arms, whether the ICC/CV estimate was unadjusted or adjusted for prognostic factors, whether the ICC/CV estimate was adjusted for the baseline value of the outcome, whether the ICC/CV was estimated from an analysis of change scores between baseline

	and follow-up, whether a repeated measures analysis was used to estimate the ICC.

Table 2: Data extracted

The ICC/CV estimate(s) of one pupil health outcome was extracted from each article, as estimates for multiple outcomes from the same study would likely be correlated and contribute relatively little additional information to the analyses in this paper which are focussed on comparing the ICC/CV across different study scenarios. Where estimates were reported for the chosen outcome at multiple levels (e.g., school and class) these were all extracted. The criteria used to select the ICC/CV when multiple estimates were reported for a given paper are presented in Table 3. Where studies reported both unadjusted and adjusted ICCs, the former was extracted on the basis that this would be of more general use to future researchers who may want to adjust their estimate of the intervention effect for a specific set of prognostic factors. Where the ICC for a given outcome was reported for multiple time points the ICC for the earliest wave was extracted, as the ICC estimate would be less likely to be impacted by the intervention. For a similar reason, where the ICC was reported separately for the control and intervention arms the former was chosen.

Aspect	**Criteria**
Outcome measure	In the first instance, the ICC/CV for the primary health outcome was selected. If there was more than one primary health outcome, the ICC/CV for the first primary outcome presented in the Results section of the paper was selected. If no primary health outcome was declared, the ICC/CV for the health outcome on which the sample size calculation was based was selected. If no primary health outcome was declared and the sample size was not based on a health outcome, the ICC/CV for the first health outcome reported in the Results section of the paper was selected.

Time point at which outcome was measured	In the first instance, the ICC/CV from the baseline time point was selected. If this was not reported, the ICC/CV from the earliest time point of measurement was selected.
Unadjusted versus adjusted ICC/CV	If the study presented both unadjusted ICCs/CVs estimates and estimates that are adjusted for prognostic factors, the unadjusted ICC/CV was extracted.
Control verses intervention arm	If the ICC/CV was reported separately for the intervention and control arms, the ICC/CV from the control arm was selected.

Table 3: Criteria used to select which ICC/CV to extract

2.5 Data analysis

Study characteristics were summarised using medians, interquartile ranges (IQR) and ranges for continuous variables, and numbers and percentages for categorical variables. Mann-Whitney and Kruskal-Wallis tests were used to compare the ICC estimates across subgroups. Analyses were undertaken using Stata 17 [39].

3. Results

3.1 Search results

3632 articles were identified through searching MEDLINE. 1590 articles were included in the full text screening stage and 246 articles were identified as eligible for inclusion in the review. One paper reported an estimate of the between-cluster coefficient of variation of the outcome, but this was negative and therefore the paper was not included. The PRISMA flow diagram is presented in Figure 1.

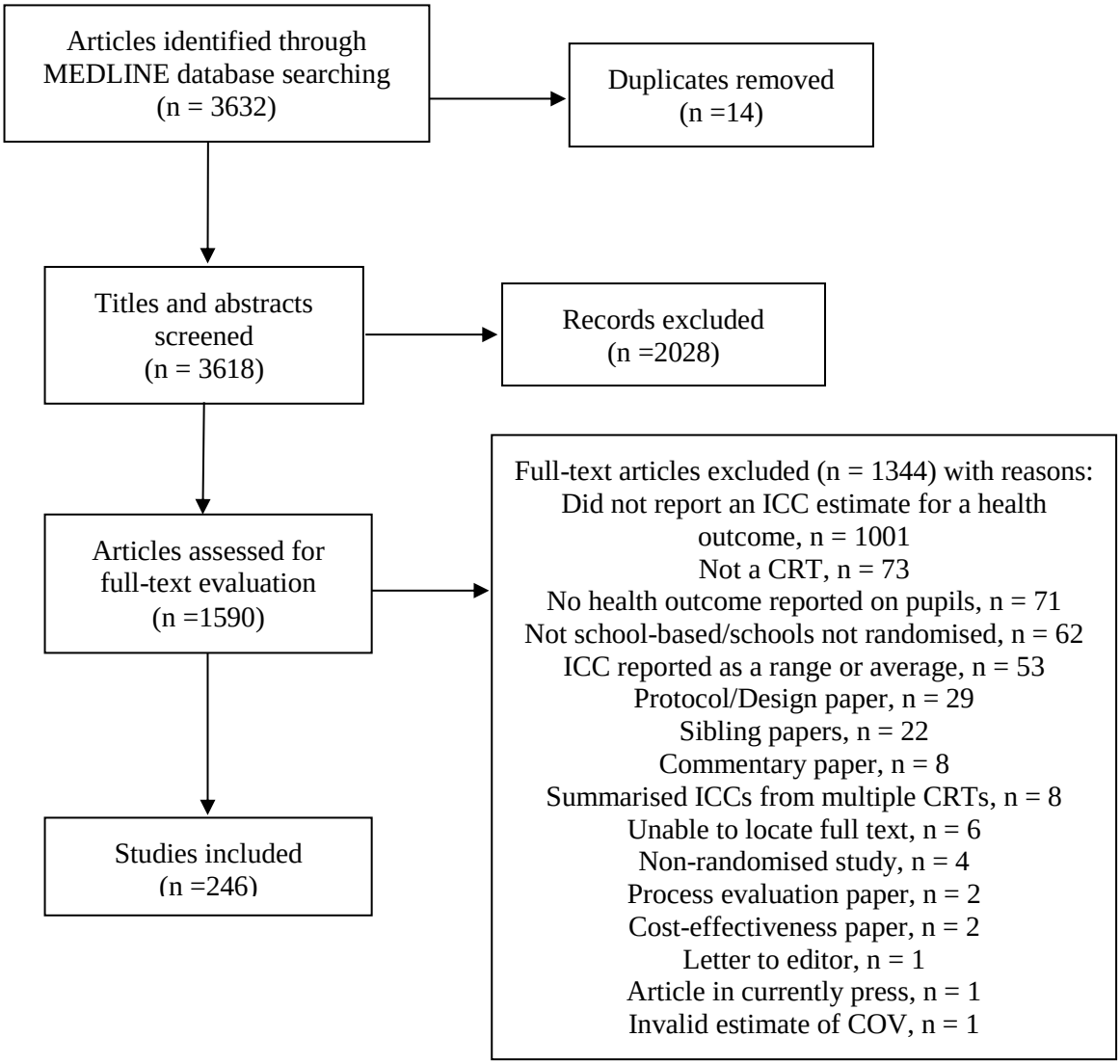


Figure 1: PRISMA flowchart summarising the results of the literature search and screening for eligibility

3.2 Publication characteristics

Worldwide, the rate of publication of articles reporting ICC estimates from school-based CRTs that evaluate interventions for improving pupil health outcomes has increased since the first publication in 1999; 44 articles were published between 1999 and 2010, compared to 25 in 2021 alone. Of the 246 included studies, 226 (91.9%) were definitive trials and 20 (8.1%) were feasibility studies. The settings of included studies spanned all regions of the world and different stages of education. The majority of studies (n=227; 92.3%) included males and females. In most of the studies schools were the units of randomisation (n=220; 89.4%); classes were randomised in 23 (9.3%) studies; and school buildings

[40], student groups [41] and year groups [42] were randomised in one study each. The studies spanned a range of different health outcome areas, the most common being socioemotional functioning and its influences (n=53; 21.5%), physical activity (n=34; 13.8%), adiposity (n=28; 11.4%) and smoking (n=21; 8.5%).

Two hundred and sixty ICC estimates were extracted: 210 at school level, 46 at classroom level, and 1 each at the levels of school building [40], student group [41], year group [42] and sports-team [43]. Forty-five (17.3%) ICCs were estimated using the baseline measurement of the outcome. ICCs were extracted for 172 continuous outcomes, 78 binary outcomes, 6 count/rate outcomes and 2 ordinal outcomes; for 2 extracted ICCs the outcome type was unclear. Of the studies that reported school-level ICCs, the median (IQR) number of clusters and pupils were 22 (12 to 40) and 1110 (441 to 2443), respectively. Of the studies that reported class-level ICCs, the median (IQR) number of clusters and pupils were 47 (25 to 88) and 647.5 (288 to 1477), respectively.

3.3 Summary of ICC estimates

The median (IQR; range) school-level ICC estimate was 0.031 (0.011 to 0.08; 0 to 0.47); 51 (24.3%) of the school-level ICCs were less than or equal to 0.01 and 135 (64.3%) were less than or equal to 0.05. The mean (SD) school-level ICC was 0.060 (0.076). Figure 2 summarises the distribution of school-level ICCs. Both the beta distribution (with shape parameters 0.77 and 11.0) and the exponential distribution provided a good fit to the school-level ICC estimates. The median (IQR; range) of the class-level ICC estimates was 0.063 (0.024 to 0.1; -0.009 to 0.262); the only negative reported ICC was at this level. All ICC estimates are reported in Appendix 1. School- and class-level ICCs are reported side-by-side for 14 studies that reported at both those levels in Appendix 2.

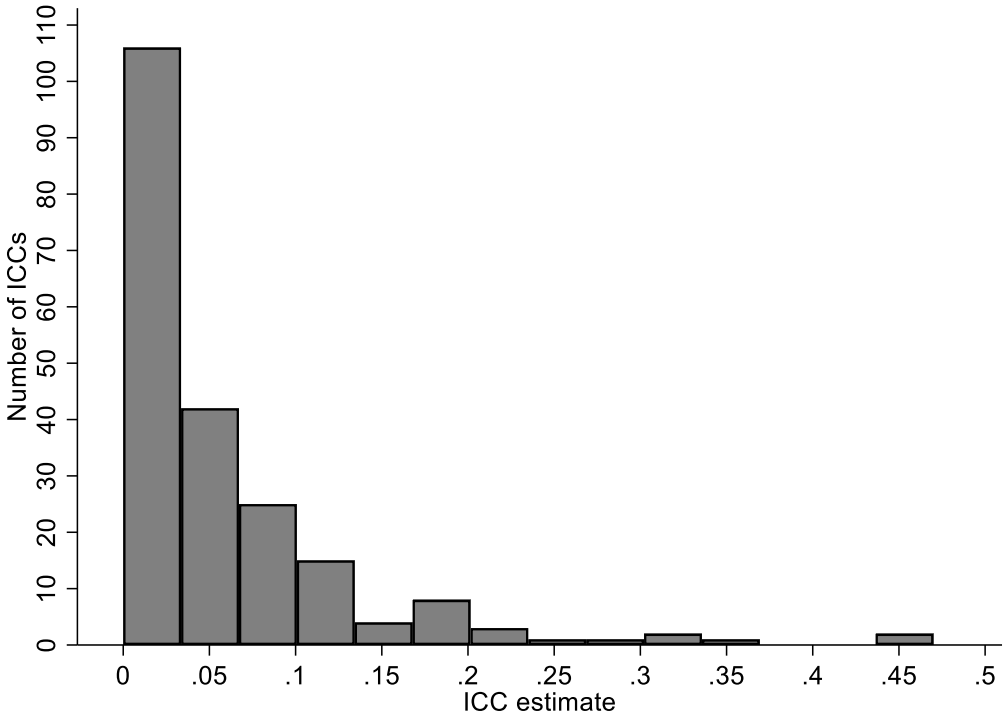


Figure 2. The distribution of school-level ICCs in school-based CRTs (N=210)

Table 4 reports the median school-level ICC by region, health outcome area (for the 10 most common areas – there were at least 8 ICC estimates for each area) and education stage. Figure 3 uses dot plots to describe the distributions. Tests of significance indicated little evidence of differences across these subgroups and there was generally a fair amount of overlap in ICC distributions. The distribution of ICCs for the USA/Canada region (median 0.033 and 75% of estimates being lower than 0.073) is in keeping with summaries of USA-based estimates that have previously been reported [12, 16-22, 25-30]. There was reasonable overlap with the distributions in the other regions with the exception of Australia/New Zealand, for which the median and upper quartile were notably lower. The school-level ICC distributions for adiposity, physical activity and general health were lower than for other outcome areas. For two specific outcomes there were more than 10 estimates of the school-level ICC. For the 17 articles that reported the school-level ICC for body mass index (BMI) the median (IQR) was 0.021 (0.015 to 0.04) and for the 11 articles that reported amount of moderate-to-vigorous physical activity (MVPA) the median (IQR) school-level ICC was 0.018 (0.01 to 0.057). Although the median was

higher, the overall distribution of ICCs was located at lower values for the pre-primary stage compared to the later stages of education.

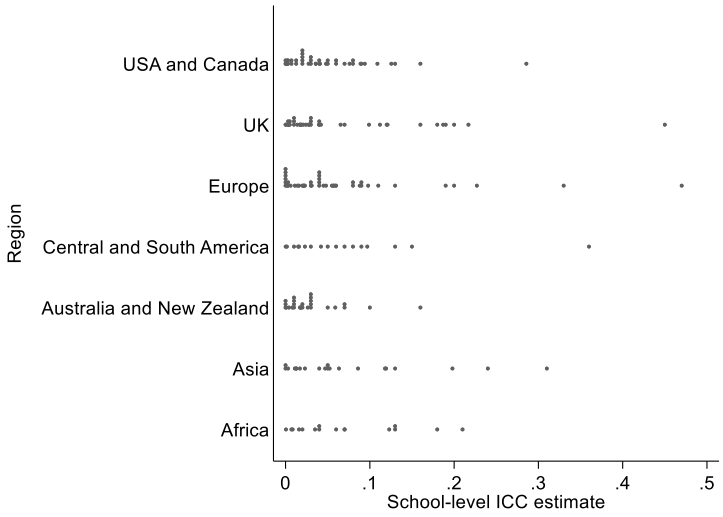
Characteristic	N	Median ICC (IQR; range)	p value
Region			0.26
Europe[1]	45	0.04 (0.014 to 0.08; 0 to 0.47)	
USA and Canada	44	0.033 (0.010 to 0.073; 0 to 0.286)	
UK[2]	40	0.029 (0.01 to 0.106; 0 to 0.45)	
Australia and New Zealand	27	0.02 (0.01 to 0.03; 0 to 0.16)	
Asia[3]	21	0.05 (0.013 to 0.118; 0 to 0.31)	
Central and South America[4]	17	0.05 (0.016 to 0.09; 0.0001 to 0.36)	
Africa[5]	16	0.05 (0.018 to 0.127; 0.0005 to 0.21)	
Health outcome area			0.76
Socioemotional functioning and its influences[6]	39	0.05 (0.02 to 0.097; 0 to 0.217)	
Physical activity	30	0.035 (0.013 to 0.059; 0 to 0.19)	

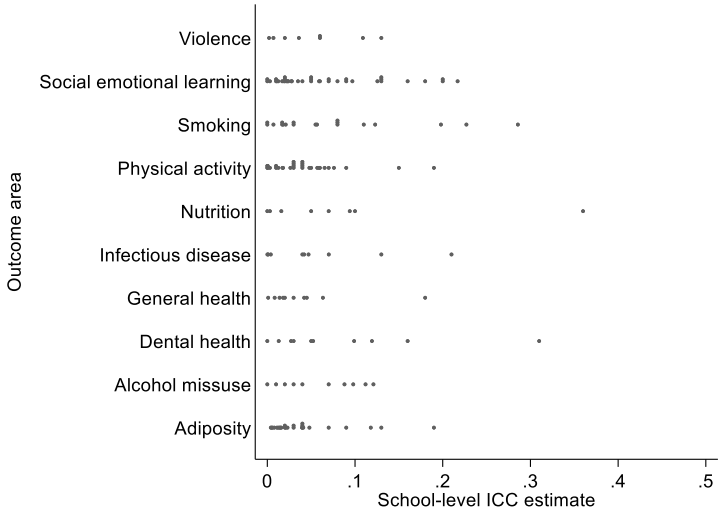
[1] Included countries stated as: Finland, The Netherlands, Denmark, Belgium, Norway, Germany, Estonia, Poland, Spain, Switzerland, Cyprus, Italy, Greece, Hungary, Sweden, Austria, Majorca, France, Ireland, Romania, Slovenia.
[2] Included countries stated as: England, Northern Ireland, Scotland, Wales.
[3] Included countries stated as: Israel, China, Iran, India, Japan, Bangladesh, Nepal, Taiwan, Peru, Pakistan, Thailand, Indonesia, Hong Kong.
[4] Included countries stated as: Jamaica, Brazil, Ecuador, Chile, Haiti, Belize.
[5] Included countries stated as: Uganda, South Africa, Kenya, Tanzania, Burundi.
[6] Includes mental health, behaviour, neurodiversity, wellbeing, quality of life, bullying, social and emotional learning, body image and self-esteem.

Adiposity	26	0.027 (0.014 to 0.041; 0.004 to 0.19)	
Smoking	19	0.055 (0.017 to 0.11; 0 to 0.286)	
Alcohol use	10	0.055 (0.02 to 0.098; 0 to 0.121)	
Dental/oral health	10	0.051 (0.027 to 0.119; 0 to 0.31)	
General health	10	0.025 (0.014 to 0.045; 0.001 to 0.18)	
Infectious disease	9	0.042 (0.004 to 0.070; 0.0001 to 0.21)	
Nutrition	8	0.06 (0.010 to 0.097; 0 to 0.36)	
Violence	8	0.048 (0.014 to 0.085; 0.002 to 0.13)	
Education stage			0.40
Pre-primary education only[7]	13	0.048 (0.03 to 0.063; 0 to 0.097)	
Primary education only[8]	81	0.04 (0.013 to 0.094; 0 to 0.47)	
Secondary education only[9]	81	0.03 (0.01 to 0.07; 0 to 0.31)	

Table 4: Median (IQR; range) school-level ICC by region, outcome area and education stage

[7] Includes preschools, kindergartens, educational childcare centres and head-start schools
[8] Includes elementary schools, middle schools (Grade 6)
[9] Includes secondary schools, middle schools (>= Grade 7), high schools, junior high schools, lower secondary schools, higher/upper secondary schools, vocational schools, intermediate vocational schools, secondary-level vocational schools and continuation schools.





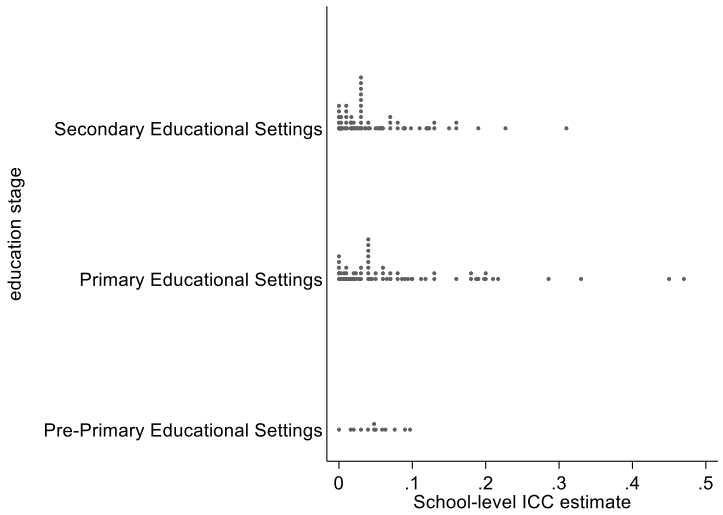


Figure 3. Dot plots of school-level ICCs by region, outcome area and education stage

The median (IQR) school-level ICC was higher for definitive studies (N=192) than feasibility studies (N=18) (0.038 (0.016 to 0.08) versus 0.01 (0.0005 to 0.04); p=0.005). The median (IQR) school-level ICC was larger for continuous outcomes (N=135) than binary outcomes (N=68) although there was little evidence of a true difference in the distributions (0.04 (0.014 to 0.08) versus 0.025 (0.008 to 0.08); p=0.21). Summaries of the school-level ICCs are reported separately for continuous and binary outcomes in Appendix 3.

For continuous outcomes, the median (IQR) school-level ICC was higher for studies that adjusted for the baseline of the outcome at the pupil level (N=35) compared with those that did not (N=95), but there was little evidence of a real difference (0.045 (0.013 to 0.09) versus 0.040 (0.016 to 0.07); p=0.50). Also, for continuous outcomes, the median school-level ICC was identical for studies that did (N=11) and did not (N=124) analyse change scores (0.04; p=0.37). The median (IQR) school-level ICC was lower for studies that estimated the ICC from a repeated measures analysis (N=37) compared with those that did not (N=173) (0.027 (0.01 to 0.057) versus 0.036 (0.013 to 0.088)), but with little evidence of a systematic difference (p=0.15). Finally, for binary outcomes, there was weak evidence that the median (IQR) ICC was higher for studies that use multilevel logistic regression to estimate this parameter on the logistic scale (N=42) than those that use other methods to estimate it on the proportions (natural) scale (N=14) (0.049 (0.014 to 0.109) versus 0.014 (0.007 to 0.023); p=0.08). The direction of this difference is consistent with the fact that the ICC on the logistic scale is generally larger than on the proportions scale [44]. Appendix 4 summarises the relationship between ICC estimates and the prevalence for binary outcomes.

4. Discussion

To our knowledge, this is the first paper to report the distribution of ICCs for pupil health outcomes from school-based CRTs worldwide. 260 ICC estimates from 246 school-based CRTs were extracted for outcomes spanning a range of health areas. There were few clear patterns regarding the relationship of the ICC with aspects of the design and analysis. Indeed, comparison of the ICC across categories of the study features examined was characterised by overlap in the distributions, although the differences in medians would be large in terms of the impact they would have on the sample size requirement for a

CRT. Imprecision in the ICC estimates may have reduced the power to detect differences between subgroups defined by design and analysis characteristics.

The large number of different outcomes represented (Appendix 1) partly accounts for the variation in the estimates, although there was even a marked variation in ICC estimates across studies for the same outcome (i.e., amount of MVPA and BMI). Sampling variability, the methodological context of the trials and the models specified to estimate the parameter will also contribute to variability in the ICC estimates. Given the clinical and methodological heterogeneity across CRTs, an individual ICC estimate for a given outcome from a single study may have poor generalisability [28], and it has been recommended that researchers use the distribution of ICCs from many studies to model the sensitivity of sample size calculations [1, 14, 34]. Distributions of ICC estimates for health outcomes in primary care-based clusters have been found to be well described by the beta distribution [14, 45]. The beta distribution was a good fit to the school-level ICCs reported in this paper as was the exponential distribution. The distribution parameters of these ICC estimates are of value for constructing informative priors when using a Bayesian framework to incorporate uncertainty about the ICC in sample size calculations for school-based CRTs [46, 47].

There was little difference between ICC estimates that were adjusted for the baseline outcome measurement and those that were not. This may be due to differences in aspects of the design and setting across the studies and the fact that adjustment for individual-level prognostic factors may increase or decrease the ICC depending on the extent to which the between- and within-cluster components of variance are reduced following adjustment [48].

The ICC from a repeated measures analysis using outcome data from across all study waves does not necessarily estimate the same parameter as an ICC for the outcome at a specific study wave. The correlation between observations from the same cluster from different waves may be smaller than the correlation between observations from the same cluster at the same study wave [34, 49]. In this study, however, there was little evidence that the school-level ICC is lower for studies that estimate the ICC from a repeated measures analysis than those that do not, although the median was lower for the former set of studies.

Previously reported summaries of school-based ICCs for pupil health outcomes have largely used data from trials and surveys in the United States [12, 16-22, 25-30]. The distribution of school-level ICCs worldwide in the current paper was broadly similar to those previous summaries, with most estimates less than 0.05 and few greater than 0.1. Only the distribution for the Australia/New Zealand region was notably different (smaller).

The median ICC for pupil health outcomes was 0.031 at the school level and 0.063 at the class level. The difference is intuitive given the greater opportunity for interaction within classes as opposed to between classes within the same school and that the ICC has been reported to be larger when the natural cluster size is smaller [20, 50]. The median ICC was markedly smaller for feasibility studies than in definitive trials. This may reflect that schools recruited in feasibility studies are a more restricted and less representative subset of the wider types of schools that are recruited in larger definitive studies [1] (p180/181). There was little evidence of a relationship between the ICC for pupil health outcomes and stage of education. Previously, it has been reported that there is a tendency for ICCs for educational outcomes to be larger for lower education grades [48].

4.1 Strengths and limitations

This is the first study to collate and summarise ICCs for pupil outcomes across different health areas from school-based CRTs worldwide. The study used a systematic searching approach with dual screening and data validation. The sample of 246 CRTs was not sufficiently large to describe the ICC within different combinations of categories of the study design parameters (e.g., only one combination of region and health outcome area provided at least 10 school-level ICC estimates). Partly for this reason, when investigating geographic variation in the ICC, we grouped countries into regions which will have obscured differences between individual countries. Based on empirical evidence from a European-based survey, it has been suggested that the ICCs assumed in the sample size calculation for school-based trials should be country-specific and outcome-specific [8]. As more school-based CRTs are undertaken the pool of reported ICCs will increase, enabling a more detailed examination with greater power to detect ICC patterns in relation to key study characteristics.

A potential limitation was the decision to use only the MEDLINE database. Although findings from a previous systematic review of similar studies indicated that few additional studies would have been found by searching other databases (specifically, EMBASE, DARE, PsychINFO and ERIC) [4], we acknowledge that further articles may have been found by searching the grey literature. Additionally, some older articles may have been missed because the titles and abstracts did not refer to using a cluster design.

It was decided to extract the ICC estimate for only one outcome from each study even when multiple ones were reported. We anticipated that ICCs would be more similar within studies and wanted to avoid a scenario where a small number of studies that reported many ICCs had a disproportionate impact on the observed distribution of ICCs.

5. Conclusions

The 260 reported ICC estimates from studies spanning all world regions and different health outcome areas, and the summaries of their distribution are a valuable resource to researchers for calculating sample size for future school-based CRTs. The ICCs had a similar distribution to published summaries of the parameter from studies based in the United States. Better reporting of the ICC in CRTs, in keeping with CONSORT guidance [51], will provide a larger pool of data that can be used to explore the distribution of ICC values and the factors that determine them in greater detail.

List of abbreviations

BMI – body mass index

CI – confidence interval

CONSORT – consolidated standards of reporting trials

CV – between coefficient of variation of the outcome

CRT – cluster randomised trial

ICC – intra-cluster correlation coefficient

IQR – interquartile range

MEDLINE – Medical Literature Analysis and Retrieval System Online

MeSH – Medical Subject Headings

MVPA – moderate-to-vigorous physical activity

PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses

SD – standard deviation

UK – United Kingdom

USA – United States of America

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available because they are also being used for a wider ongoing programme of research but are available from the corresponding author on reasonable request.

Competing interests

Not applicable

Funding

This research was funded by the National Institute for Health and Care Research Applied Research Collaboration South West Peninsula.

Authors' contributions

KP, MN, ZMX, TF and OU conceived the study. MN, ZMX and TF advised on the design of the study and contributed to the protocol. KP and OU contributed to the design of the study, wrote the protocol and designed the data extraction form. KP and OU undertook data extraction. KP conducted the analyses of the data. All authors had full access to all the data. KP took primary responsibility for writing the manuscript. All authors provided feedback on all versions of the paper. All authors read and approved the final manuscript.

Acknowledgements

Kitty Parker and Obioha Ukoumunne were supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration South West Peninsula. The NIHR had no role in the study design, collection, management, analysis or interpretation of data, writing of the report, or the decision to submit the report for publication. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

References

1. Eldridge S, Kerry S. A practical guide to cluster randomised trials in health services research. Chichester: John Wiley & Sons; 2012.
2. Walleser S, Hill SR, Bero LA. Characteristics and quality of reporting of cluster randomized trials in children: reporting needs improvement. Journal of Clinical Epidemiology. 2011;64(12):1331-40.
3. Goesling B. A Practical Guide to Cluster Randomized Trials in School Health Research. Journal of School Health. 2019;89(11):916-25.
4. Parker K, Nunns M, Xiao Z, Ford T, Ukoumunne OC. Characteristics and practices of school-based cluster randomised controlled trials for improving health outcomes in pupils in the United Kingdom: a methodological systematic review. BMC Medical Research Methodology. 2021;21(1):152.
5. Parker K, Eddy S, Nunns M, Xiao Z, Ford T, Eldridge S, et al. Systematic review of the characteristics of school-based feasibility cluster randomised trials of interventions for improving the health of pupils in the UK. Pilot Feasibility Studies. 2022;8(1):132.
6. Hayes R, Moulton L. Cluster Randomised Trials. Florida: CRC Press; 2009.
7. Murray DM. Design and Anaylsis of Group-Randomized Trials. New York: Oxford University Press; 1998.
8. Shackleton N, Hale D, Bonell C, Viner RM. Intraclass correlation values for adolescent health outcomes in secondary schools in 21 European countries. SSM - Population Health. 2016;2:217-25.
9. Bonell C, Parry W, Wells H, Jamal F, Fletcher A, Harden A, et al. The effects of the school environment on student health: A systematic review of multi-level studies. Health & Place. 2013;21:180-91.
10. Langford R, Bonell CP, Jones HE, Pouliou T, Murphy SM, Waters E, et al. The WHO Health Promoting School framework for improving the health and well-being of students and their academic achievement. Cochrane Database of Systematic Reviews. 2014(4).
11. Bonell C, Jamal F, Harden A, Wells H, Parry W, Fletcher A. Systematic review of the effects of schools and school environment interventions on health: evidence mapping and synthesis. Public Health Research. 2013;1(1).
12. Dong N, Reinke WM, Herman KC, Bradshaw CP, Murray DW. Meaningful Effect Sizes, Intraclass Correlations, and Proportions of Variance Explained by Covariates for Planning Two- and Three-Level Cluster Randomized Trials of Social and Behavioral Outcomes. Evaluation Review. 2016;40(4):334-77.
13. Hale DR, Fitzgerald-Yau N, Viner RM. A systematic review of effective interventions for reducing multiple health risk behaviors in adolescence. American Journal of Public Health. 2014;104(5):e19-41.
14. Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. Journal of Clinical Epidemiology. 2004;57(8):785-94.
15. Stuart B, Becque T, Moore M, Little P. Clustering of continuous and binary outcomes at the general practice level in individually randomised studies in primary care - a review of 10 years of primary care trials. BMC Medical Research Methodology. 2020;20(1):83.
16. Murray DM, Hannan PJ. Planning for the appropriate analysis in school-based drug-use prevention studies. Journal of Consulting and Clinical Psychology. 1990;58(4):458-68.
17. Murray DM, Rooney BL, Hannan PJ, Peterson AV, Ary DV, Biglan A, et al. Intraclass correlation among common measures of adolescent smoking: estimates, correlates, and applications in smoking prevention studies. American Journal of Epidemiology. 1994;140(11):1038-50.
18. Murray DM, Short B. Intraclass correlation among measures related to alcohol use by young adults: estimates, correlates and applications in intervention studies. Journal of Studies on Alcohol. 1995;56(6):681-94.
19. Murray DM, Short BJ. Intraclass correlation among measures related to tobacco use by adolescents: estimates, correlates, and applications in intervention studies. Addictive Behaviors. 1997;22(1):1-12.

20. Siddiqui O, Hedeker D, Flay BR, Hu FB. Intraclass correlation estimates in a school-based smoking prevention study: outcome and mediating variables, by sex and ethnicity. American Journal of Epidemiology. 1996;144(4):425-33.
21. Murray DM, Clark M, Wagenaar AC. Intraclass correlations from a community-based alcohol prevention study: the effect of repeat observations on the same communities. Journal of studies on alcohol. 2000;61(6):881-90.
22. Ennett ST, Flewelling RL, Lindrooth RC, Norton EC. School and neighborhood characteristics associated with school rates of alcohol, cigarette, and marijuana use. Journal of Health and Social Behavior. 1997;38(1):55-71.
23. Resnicow K, Zhang N, Vaughan RD, Reddy SP, James S, Murray DM. When intraclass correlation coefficients go awry: a case study from a school-based smoking prevention study in South Africa. American Journal of Public Health. 2010;100(9):1714-8.
24. Sellström E, Bremberg S. Is there a "school effect" on pupil outcomes? A review of multilevel studies. Journal of Epidemiology and Community Health. 2006;60(2):149-55.
25. Murray DM, Phillips GA, Birnbaum AS, Lytle LA. Intraclass Correlation for Measures from a Middle School Nutrition Intervention Study: Estimates, Correlates, and Applications. Health Education & Behavior. 2001;28(6):666-79.
26. Juras R. Estimates of Intraclass Correlation Coefficients and Other Design Parameters for Studies of School-Based Nutritional Interventions. Evaluation Review. 2016;40(4):314-33.
27. Gray HL, Burgermaster M, Tipton E, Contento IR, Koch PA, Di Noia J. Intraclass Correlation Coefficients for Obesity Indicators and Energy Balance-Related Behaviors Among New York City Public Elementary Schools. Health, Education and Behavior. 2016;43(2):172-81.
28. Murray DM, Catellier DJ, Hannan PJ, Treuth MS, Stevens J, Schmitz KH, et al. School-level intraclass correlation for physical activity in adolescent girls. Medicine & Science in Sports & Exercise. 2004;36(5):876-82.
29. Murray DM, Stevens J, Hannan PJ, Catellier DJ, Schmitz KH, Dowda M, et al. School-level intraclass correlation for physical activity in sixth grade girls. Medicine & Science in Sports & Exercise. 2006;38(5):926-36.
30. Hedberg EC. Academic and Behavioral Design Parameters for Cluster Randomized Trials in Kindergarten: An Analysis of the Early Childhood Longitudinal Study 2011 Kindergarten Cohort (ECLS-K 2011). Evaluation Review. 2016;40(4):279-313.
31. Murray DM, Blitstein JL. Methods To Reduce The Impact Of Intraclass Correlation In Group-Randomized Trials. Evaluation Review. 2003;27(1):79-103.
32. Gulliford MC, Adams G, Ukoumunne OC, Latinovic R, Chinn S, Campbell MJ. Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. Journal of Clinical Epidemiology. 2005;58(3):246-51.
33. Campbell MK, Fayers PM, Grimshaw JM. Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research. Clinical Trials. 2005;2(2):99-107.
34. Korevaar E, Kasza J, Taljaard M, Hemming K, Haines T, Turner EL, et al. Intra-cluster correlations from the CLustered OUtcome Dataset bank to inform the design of longitudinal cluster trials. Clinical Trials. 2021;18(5):529-40.
35. 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):15.
36. UNESCO. International Standard Classification of Education: ISCED 2011. Montreal: UIS; 2012. p. 85.
37. Goldstein H, Browne W, Rasbash J. Multilevel modelling of medical data. Statistics in Medicine. 2002;21(21):3291-315.
38. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986;73(1):13-22.

39. StataCorp. Stata. Release 17. College Station, TX: StataCorp LLC; 2021.
40. Lopata C, Thomeer ML, Rodgers JD, Donnelly JP, McDonald CA, Volker MA, et al. Cluster Randomized Trial of a School Intervention for Children with Autism Spectrum Disorder. Journal of Clinical Child and Adolescent Psychology. 2019;48(6):922-33.
41. Nykänen M, Sund R, Vuori J. Enhancing safety competencies of young adults: A randomized field trial (RCT). Journal of Safety Research. 2018;67:45-56.
42. Stallard P, Sayal K, Phillips R, Taylor JA, Spears M, Anderson R, et al. Classroom based cognitive behavioural therapy in reducing symptoms of depression in high risk adolescents: pragmatic cluster randomised controlled trial. British Medical Journal. 2012;345:e6058.
43. Emery CA, Rose MS, McAllister JR, Meeuwisse WH. A prevention strategy to reduce the incidence of injury in high school basketball: a cluster randomized controlled trial. Clinical Journal of Sport Medicine. 2007;17(1):17-24.
44. Eldridge SM, Ukoumunne OC, Carlin JB. The Intra-Cluster Correlation Coefficient in Cluster Randomized Trials: A Review of Definitions. International Statistical Review. 2009;77(3):378-94.
45. Turner RM, Omar RZ, Thompson SG. Bayesian methods of analysis for cluster randomized trials with binary outcome data. Statistics in Medicine. 2001;20(3):453-72.
46. Turner RM, Thompson SG, Spiegelhalter DJ. Prior distributions for the intracluster correlation coefficient, based on multiple previous estimates, and their application in cluster randomized trials. Clinical Trials. 2005;2(2):108-18.
47. Jones BG, Streeter AJ, Baker A, Moyeed R, Creanor S. Bayesian statistics in the design and analysis of cluster randomised controlled trials and their reporting quality: a methodological systematic review. Systematic Reviews. 2021;10(1):91.
48. Hedges LV, Hedberg EC. Intraclass Correlation Values for Planning Group-Randomized Trials in Education. Educational Evaluation and Policy Analysis. 2007;29(1):60-87.
49. Li F, Hughes JP, Hemming K, Taljaard M, Melnick ER, Heagerty PJ. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. Statistical Methods in Medical Research. 2021;30(2):612-39.
50. Gulliford MC, Ukoumunne OC, Chinn S. Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: data from the Health Survey for England 1994. American Journal of Epidemiology. 1999;149(9):876-83.
51. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. British Medical Journal. 2012;345:e5661.

Table and Figure legends

Table 1: Search strategy for MEDLINE (through Ovid)

Table 2: Data extracted

Table 3: Criteria used when selecting which ICC/CV to extract

Table 4: Median (IQR; range) school-level ICC by region, outcome area and education stage

Figure 1: PRISMA flowchart summarising the results of the literature search and screening for eligibility

Figure 2: The distribution of school-level ICCs in school-based CRTs (N=210)

Figure 3. Dot plots of school-level ICCs by region, outcome area and education stage