8-11-2021

Effect Sizes and Intra-Cluster Correlation Coefficients Measured from the Green Dot High School Study for Guiding Sample Size Calculations When Designing Future Violence Prevention Cluster Randomized Trials in School Settings

Md. Tofial Azam
University of Kentucky, tofial.azam@uky.edu

Heather M. Bush
University of Kentucky, heather.bush@uky.edu

Ann L. Coker
University of Kentucky, ann.coker@uky.edu

Philip M. Westgate
University of Kentucky, philip.westgate@uky.edu

Contemporary Clinical Trials Communications 23 (2021) 100831

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Effect sizes and intra-cluster correlation coefficients measured from the Green Dot High School study for guiding sample size calculations when designing future violence prevention cluster randomized trials in school settings

Md. Tofial Azam [a,*], Heather M. Bush [a], Ann L. Coker [b], Philip M. Westgate [a]

[a] University of Kentucky, Department of Biostatistics, College of Public Health, Lexington, KY, USA
[b] University of Kentucky, College of Medicine, 800 Rose St., Pavilion H, Room C361, Lexington, KY, 40536, USA

ARTICLE INFO

Keywords:
Cluster randomized controlled trial
Effect size
Intra-cluster correlation coefficient
Sample size calculation
Dating violence
Sexual violence

ABSTRACT

Purpose: Cluster randomized controlled trials (cRCTs) are popular in school-based research designs where schools are randomized to different trial arms. To help guide future study planning, we provide information on anticipated effect sizes and intra-cluster correlation coefficients (ICCs), as well as school sizes, for dating violence (DV) and interpersonal violence outcomes based on data from a cRCT which evaluated the bystander-based violence intervention *'Green Dot'*.
Methods: We utilized data from 25 schools from the Green Dot High School study. Effect size and ICC values corresponding to dating and interpersonal violence outcomes are obtained from linear mixed effect models. We also calculated the required number of schools needed for future studies utilizing available methods that do and do not consider variation in school size.
Results: Observed effect sizes for DV outcomes range from 0.06 to 0.11. Observed ICC values for DV outcomes range from 0.0006 to 0.0032. The upper limit of 95% CIs for the true ICCs range from 0.0023 to 0.0070.
Conclusion: School-based evaluations with violence outcomes are expected to have small effect sizes. Observed ICCs are less than 0.005 and upper limit of of 95% CIs for the true ICCs are less than 0.01. Designing school-based cRCTs should account for the ICC, even if its value is assumed to be negligible. Furthermore, variation in school sizes should also be accounted for to avoid having too few schools to achieve the desired power.

1. Introduction

Cluster randomized controlled trials (cRCTs) are trials in which entire clusters of subjects are randomized to different trial arms, and the intervention is applied to entire clusters [1,2]. Popular examples of possible clusters include, but are not limited to, health care organizations, communities, and in regard to our focus, schools. Although schools are randomized, outcomes are obtained at the student-level.

In violence prevention effectiveness evaluations, school-level cRCTs are regularly being used. There are several possible advantages to randomizing schools as opposed to individual students. In general, applying the intervention to the entire school is advantageous in that contamination is avoided [3–5] and school-wide implementation is logistically more convenient [6]. School-level cRCTs provide an approach to rigorously evaluate interventions that address outcomes that are unlikely to be disclosed without student anonymity (such as reports of violence victimization or perpetration) [7]. In many cases, interventions are better suited to school-wide delivery rather than to individuals [5,8].

Researchers conducting school randomized trials must explicitly account for between-school differences in outcomes that are not explained by covariates, i.e., clustering, both while designing the study,

Abbreviations: cRCT, Cluster randomized controlled trial; ICC, Intra-cluster correlation coefficient; DV, Dating violence; SV, Sexual violence; GDHS, Green dot high school.

* Corresponding author. Department of Biostatistics, College of Public Health 725 Rose Street, Multidisciplinary Science Building Suit 205, Lexington, KY, 40536, USA.

E-mail addresses: md.tofialazam@gmail.com (Md.T. Azam), heather.bush@uky.edu (H.M. Bush), ann.coker@uky.edu (A.L. Coker), philip.westgate@uky.edu (P.M. Westgate).

https://doi.org/10.1016/j.conctc.2021.100831
Received 18 September 2020; Received in revised form 9 July 2021; Accepted 4 August 2021
Available online 11 August 2021

which is our focus in this manuscript, as well as when analyzing data collected from the study. In practice, to calculate the number of schools needed to conduct a cRCT with a pre-specified significance level and statistical power to test for an intervention effect, multiple pieces of information are required. Besides the desired significance level, power, and effect size, numerical information is needed on school sizes (number of subjects from whom data are collected) and the intra-cluster correlation coefficient (ICC) [9], which quantifies the degree of clustering. Therefore, as emphasized in other settings (e.g. Refs. [3,10,11,14,15], there is a significant need for studies to publish not only their information on effect sizes but ICC values as well.

Although the information on effect sizes, school sizes, and the ICC is needed, accurate information may be difficult to have in hand when calculating the required number of clusters. Essentially, although results from trials include effect size estimates in some form, often design features such as variability in school sizes, and especially ICCs, are not fully reported. For instance Ref. [16], were able to obtain effect size estimates from 5 studies and perform a meta-analysis from which they reported effect sizes of −0.21 [95% CI: −0.41, −0.02] for dating violence victimization and −0.01 [95% CI: −0.04, 0.05] for dating violence perpetration. However, in a systematic review by Ref. [17], only one of the cRCTs reported ICC values [6] among 38 studies included in their review manuscript.

This lack of reporting in the literature can make designing a future cRCT challenging. The purpose of this manuscript is to offer guidance and best practice for the study design of future violence prevention cRCTs in school settings. Utilizing data from the Green Dot High School (GDHS) study, we provide information on observed school sizes, effect sizes, and ICC values specific to victimization and perpetration outcomes related to sexual and dating violence in teens. These student-level violence outcomes of focus are descriptively presented in Table 1. Specifically, sexual & dating violence in teens includes stalking (repeated pattern of following, harassing, or threat), sexual violence (SV) (unwanted sexual contact), physical and psychological dating violence, sexual harassment (unwanted derogatory sexual comments or gesture), and reproductive coercion with a dating relationship [12,13,18]. Finally, we present the required numbers of clusters for observed and generalized effect size and ICC values to facilitate expectations for future study planning.

Table 1
Individual student level outcome characteristics.

Variables	Mean (SD)	Median (Min - Max)	1st quartiles – 3rd quartiles
School/Cluster Size	553 (205)	576 (227–963)	
Victimization			
Unwanted sex	0.69 (3.12)	0 (0–30)	0–0
Physical DV	0.31 (1.43)	0 (0–10)	0–0
Psychological DV	2.90 (6.56)	0 (0–40)	0–2
Sexual Harassment	2.86 (5.44)	0 (0–30)	0–3
Stalking	1.49 (4.05)	0 (0–30)	0–1
Reproductive coercion	1.16 (3.45)	0 (0–30)	0–1
Perpetration			
Unwanted sex	0.41 (2.82)	0 (0–30)	0–0
Physical DV	0.19 (1.11)	0 (0–10)	0–0
Psychological DV	1.32 (4.52)	0 (0–40)	0–1
Sexual Harassment	0.77 (3.34)	0 (0–30)	0–0
Stalking	0.48 (2.86)	0 (0–30)	0–0
Total violence	11.42 (26.50)	3 (0–280)	0–12

Total number of individuals: n = 13816 (Intervention: 7046; Control: 6770).
DV = Dating Violence.

2. Method

2.1. Bystander-based violence prevention intervention evaluated

This analysis was based on previously collected data from the cRCT design which evaluated the bystander-based violence prevention intervention *'Green Dot'*. Details on the GDHS study can be found in previous publications [12,13,19,20]. The high school version of '*Green Dot*' was adapted by the developer based on the prior college version. Unique to the high school-based Green Dot is its use of Rape crisis center staff trained as intervention educators to provide this intervention to students in two phases. In phase 1, students received a 50-min long school-wide general speech. These cogent speeches provided students with training to identify the green and red dots and their potential role as active bystanders to prevent and reduce SV at schools [21]. In phase 2, students were selected by teachers and school administrators to participate in intensive popular opinion leader (POL) training. Students were chosen as leaders whom others respected, followed, or emulated, and they didn't need to have academic, athletic, or social leadership skills [12,13]. The training lasted for 5 h and included role-playing and more interactive skill-building for employing bystander behaviors. While POLs were identified and invited, the training sessions were kept open to other students when space permitted [12,13]. Both training phases focused on violence victimization, as well as perpetration and other prosocial behaviors to identify circumstances that may lead to violence. Both phases also trained students to act directly, to distract, or to delegate to others to prevent and reduce the risk of violence [12,13]. The intervention training focused not only on SV and DV, but also on sexual harassment, stalking, and other interpersonal violence. Using a diffusion approach, the speeches provide the foundational training, but the intensive POL education offers the opportunity for peers to diffuse the bystander behaviors as part of the school culture [22].

2.2. Study design and data collection

As described in Ref. [21]; Green Dot was implemented in 26 schools randomized to the intervention (Green Dot) versus control (standard awareness education) conditions. Details for data collection methods are discussed elsewhere [21,23]. Briefly, school-based surveying was conducted such that all students, grades 9 to 12, were invited to complete an annual 99 item paper-and-pencil, anonymous questionnaire that took approximately 25–40 min to complete. Training began in fall 2010; surveys were conducted over a 5-year (2010–2014) period. Study personnel traveled to each school every year to identify 1 or 2 days in the spring semester on which the majority of the students would be present. These 26 schools were selected based on their statewide distribution from 13 administrative districts; two schools per district were randomized to intervention or control arm. In this manuscript, Year 4 (2013) data are used as this is the year when the study was fully implemented; i.e., the maximum proportion of POLs in GDHS study were selected at that year [12,13]. One school in the control arm dropped out at year 2 (2011) [21]; therefore, in this manuscript, data from 25 remaining schools (13 intervention, 12 control) are analyzed. To perform this secondary data analysis, we have obtained permission from the primary investigators of the study. The study protocol was approved by the University of Kentucky IRB (13-0680-F1V).

2.3. Calculation of effect size and ICC

The effect size of interest is based on *Cohen's d,* defined as the difference in means divided by the standard deviation [24]. The outcome variables in this manuscript are continuous variables calculated from instruments described in details in the appendix of [21]. In order to calculate the effect size corresponding to any given outcome, linear mixed effects models were utilized. For the *j*th student from the *i*th school in the GDHS, the linear mixed effects model [25] is formulated as

$$y_{ij} = \beta_0 + \beta_1 I_i(Green\ Dot) + \beta_2 X_i + b_i + \varepsilon_{ij};\ \ i = 1, 2, 3, \&, 25 j = 1, 2, 3, \&, m\ i.$$

In this model, y_{ij} is the value of the outcome from the *j*th student in the *i*th school, $I_i(Green\ Dot)$ is the intervention indicator for school, and X_i is the baseline school-level mean of the given outcome from the i^{th} school. Furthermore, b_i is the *i*th school's random effect/intercept which has a mean of *0* and a variance denoted by $\sigma^2_{between}$, which quantifies variability in school-specific means. Finally, ε_{ij} is a random error with mean *0* and a variance denoted by σ^2_{within}, which quantifies within-school variance. As a result, the variance of the outcome is $\sigma^2_{between} + \sigma^2_{within}$. Therefore, based on this statistical model, the effect size is calculated as the difference in means between the two trial arms, β_1, divided by the standard deviation of the outcome, given by $\sqrt{\sigma^2_{between} + \sigma^2_{within}}$. As a result,

$$Effect\ Size = \frac{\beta_1}{\sqrt{\sigma^2_{between} + \sigma^2_{within}}}$$

The ICC is inherently estimated with the linear mixed effects model via the use of the two variance components [9,25,26]. Specifically,

$$ICC = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}},$$

showing that the ICC is the proportion of the total variance in the outcome that is a result of between-school variation.

These models are fit using the GLIMMIX procedure of SAS 9.4 [27]. Confidence intervals (95%) are calculated for each ICC using the procedure proposed by Ref. [28], as implemented in SAS code from Ref. [10]. The upper bounds of these intervals are presented in our results to provide conservative values.

We note that the primary analysis of data from the GDHS study utilized school-level sums of student responses to account for clustering due to the longitudinal evaluation of the anonymous student survey. In this manuscript, we analogously analyze student-level data and account for clustering via the incorporation of a random effect.

2.4. Calculation of required number of clusters

In GDHS, students are clustered by schools, and schools were randomized. As violence prevention studies often examine student-reported outcomes for interventions delivered at the school level, study design must consider the number of clusters or schools. In order to calculate the required number of clusters to design a cRCT to assess the efficacy of the intervention versus control, information on desired level of significance (α), type II error rate ($\beta = 1 - power$), the standard deviation of the outcome variable, the cluster size ($\overline{m}$; the average high school size), the ICC, and the effect size are needed. It can be shown (see pages 54 and 78 of [3] that the required number of schools per arm, *k*, can be approximated by,

$$k = \frac{2\left[z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right]^2 \{1 + (\overline{m} - 1)ICC\}}{\overline{m}^* Effect\ Size^2}$$

This approximation only works well if *k* is "large" due to the use of critical values, z, from a standard normal distribution. Therefore, critical values from a t-distribution may be preferred. For instance, we utilize an R (R core [29] function provided by Ref. [3]. If the calculated number of clusters per arm is less than 20, their function replaces $z_{1-\alpha/2}$ and $z_{1-\beta}$ with $t\left(1 - \frac{\alpha}{2}, k - 2\right)$ and $t(1 - \beta,\ k - 2)$, respectively, which are the corresponding critical values from t-distribution with *k-2* degrees of freedom. We note that although an ICC value of 0 corresponds to the case of independence, the sample size approach utilizes degrees of freedom equaling to the number of clusters minus 2 which accounts for the possibility of clustering as is required in the data analysis. Due to the use of *k* within the degrees of freedom, their function utilizes an iterative process to provide a final value for *k*. To use this function, the initial value of *k*, calculated using the sample size formula above, has to be more than 2. Note also, the solution for *k* may not be an integer, and therefore the reported value for *k* should be the calculated value rounded up to the next integer to ensure the nominal power.

For school-based studies, school sizes are often not fixed. Unfortunately, variation in cluster sizes is a source of imprecision in the above sample size calculation [30–32]. To account for this potential imprecision, the above formula can be adjusted using the coefficient of variation ($cv = \frac{sd_{clustersize}}{\overline{m}}$; here, $sd_{clustersize}$ is the standard deviation of school size) of school sizes. As given by Ref. [32], the adjusted number of clusters per arm should be, $k_{adj} = \frac{k}{1 - cv^2\theta(1-\theta)}$, where $\theta = \frac{\overline{m}ICC}{\overline{m}ICC + (1 - ICC)}$. We note that this adjustment should be made prior to rounding, as described in the previous paragraph, or the required number of clusters per arm may be one more cluster than is actually needed. To carry out calculations that account for cluster size variation, we incorporate the adjustment of [32] within the R (R core [29] function provided by Ref. [3]. In the Appendix, we demonstrated a way to modify the R code provided by Ref. [3] to calculate the required number of schools to design a cRCT.

3. Results

The GDHS included 25 schools having an average size of 553 (SD = 205) (Table 1), with sizes ranging from 227 to 963 (Table 1). Table 2 presents calculated effect sizes, ICCs, and upper limits of corresponding 95% confidence intervals based on analyzing violence outcomes from GDHS, adjusted for baseline characteristics. For various cluster sizes and variations in cluster sizes, Table 3, 4 and 5 provide the minimum number of clusters required per trial arm to have at least 80% or 90% power for two-sided tests at the 5% significance level.

The estimated effect sizes for different violence outcomes range from −0.06 (physical DV) to −0.10 (sexual harassment) for victimization and −0.06 (unwanted sex) to −0.08 (sexual harassment) for perpetration. The effect size for all violence is −0.11 (victimization or perpetration or both). We note that the negative effect size indicates that the intervention reduces the incidence of violence in the intervention arm compared to the control arm.

The estimated ICC values for violence outcomes range from 0.0006 to 0.0032. The ICC value for all violence (victimization or perpetration or both) is 0.0018. The upper limit of 95% CIs for the true ICCs range

Table 2
Observed effect sizes, ICCs and upper limits of corresponding 95% confidence intervals, and minimum numbers of clusters per arm for fixed and varying cluster sizes for given ICC and ICC UL values for both 80% and 90% power at a two-sided significance level of 0.05.

Variable	Observed Effect Size[a]	ICC	ICC UL	Number of clusters required per trial arm							
				Power = 80%				Power = 90%			
				For ICC		For ICC UL		For ICC		For ICC UL	
				Fixed cluster size[b]	Varying cluster size[c]	Fixed cluster size[b]	Varying cluster size[c]	Fixed cluster size[b]	Varying cluster size[c]	Fixed cluster size[b]	Varying cluster size[c]
Victimization											
Unwanted sex	−0.08	0.0006	0.0026	9	9	13	14	11	11	17	18
Physical DV	−0.06	0.0008	0.0032	14	14	22	23	18	18	30	31
Psychological DV	−0.07	0.0032	0.0070	19	19	29	29	22	23	38	39
Sexual Harassment	−0.10	0.0017	0.0044	8	8	12	13	10	11	16	16
Stalking	−0.08	0.0003	0.0024	8	9	13	13	10	10	16	17
Reproductive coercion	−0.06	0.0022	0.0055	20	21	32	33	24	25	43	44
Perpetration											
Unwanted sex	−0.06	0.0019	0.0050	19	19	30	31	22	23	40	41
Physical DV	−0.07	0.0008	0.0032	11	12	19	19	14	14	22	23
Psychological DV	−0.07	0.0003	0.0023	10	10	16	16	12	12	20	21
Sexual Harassment	−0.08	0.0025	0.0060	13	14	22	22	17	17	26	27
Stalking	−0.07	0.0017	0.0049	14	14	22	23	18	18	29	30
Total violence	−0.11	0.0018	0.0048	7	8	11	12	9	9	14	14

ICC =Intra-cluster correlation coefficient.
ICC UL = upper limit of the 95% confidence interval of ICC, represents a potential upper bound on ICC.
DV = Dating Violence.
All regression equations are adjusted for a specific baseline school-level covariate.

[a] .Effect size $= \frac{\textit{Adjusted mean difference of outcome in intervention arm and the control arm}}{\textit{pooled standard deviation}}$

[b] Fixed cluster size is the school size average of 553 (Range: 227–963).

[c] Allow school size to vary utilizing the standard deviation of school size of 205.

Table 3
Minimum number of clusters required per trial arm for a range of effect sizes, ICCs and 80% and 90% power for average school size of 400 for fixed and varying cluster size at a two-sided significance level of 0.05.

Effect Size[a]	ICC	Number of clusters required per trial arm							
		Power = 80%				Power = 90%			
		Fixed Cluster Size[b]	Varying Cluster Size[c]			Fixed Cluster Size[b]	Varying Cluster Size[c]		
			SD = 100	SD = 200	SD = 300		SD = 100	SD = 200	SD = 300
0.05	0	18	18	18	18	22	22	22	22
	0.0001	19	19	19	19	22	22	23	23
	0.001	22	23	24	25	30	30	31	34
	0.005	48	48	50	54	63	64	67	72
	0.01	79	80	82	87	105	106	110	116
0.10	0	7	7	7	7	8	8	8	8
	0.0001	8	8	8	8	8	8	8	8
	0.001	8	8	9	9	10	10	11	12
	0.005	14	15	15	16	18	19	19	21
	0.01	22	22	23	24	27	27	28	29
0.15	0	3	3	3	3	4	4	4	4
	0.0001	3	3	3	3	4	4	4	4
	0.001	4	4	4	4	5	5	5	5
	0.005	8	9	9	10	10	10	11	11
	0.01	12	12	12	13	14	14	15	16

SD = Standard deviation.
ICC= Intra-cluster correlation coefficient.

[a] Effect size $= \frac{\textit{Mean of outcome in intervention arm} - \textit{Mean of outcome in control arm}}{\textit{pooled standard deviation}}$

[b] Average school size is 400.

[c] Allows school (cluster) size to vary utilizing standard deviation of cluster size.

Table 4
Minimum number of clusters required per trial arm for a range of effect sizes, ICCs and 80% and 90% power for average school size of 500 for fixed and varying cluster size at a two-sided significance level of 0.05.

Effect Size[a]	ICC	Number of clusters required per trial arm							
		Power = 80%				Power = 90%			
		Fixed Cluster Size[b]	Varying Cluster Size[c]			Fixed Cluster Size[b]	Varying Cluster Size[c]		
			SD = 100	SD = 200	SD = 300		SD = 100	SD = 200	SD = 300
0.05	0	15	15	15	15	19	19	19	19
	0.0001	16	16	16	16	20	20	20	21
	0.001	21	22	22	23	26	26	27	28
	0.005	44	45	46	48	59	60	61	64
	0.01	76	76	77	80	101	102	103	106
0.10	0	5	5	5	5	7	7	7	7
	0.0001	5	5	5	5	8	8	8	8
	0.001	7	7	8	8	10	10	10	10
	0.005	14	14	14	15	17	17	18	19
	0.01	21	21	22	22	26	26	26	27
0.15	0	3	3	3	3	3	3	3	3
	0.0001	3	3	3	3	3	3	3	4
	0.001	4	4	4	4	4	4	4	5
	0.005	8	8	8	9	10	10	10	10
	0.01	11	11	11	12	14	14	14	14

SD = Standard deviation.
ICC= Intra-cluster correlation coefficient.

[a] Effect size $= \frac{\textit{Mean of outcome in intervention arm} - \textit{Mean of outcome in control arm}}{\textit{pooled standard deviation}}$

[b] Average school size is 500.

[c] Allows school (cluster) size to vary utilizing standard deviation of cluster size.

Table 5
Minimum number of clusters required per trial arm for a range of effect sizes, ICCs and 80% and 90% power for average school size of 600 for fixed and varying cluster size at a two-sided significance level of 0.05.

Effect Size[a]	ICC	Number of clusters required per trial arm							
		Power = 80%				Power = 90%			
		Fixed Cluster Size[b]	Varying Cluster Size[c]			Fixed Cluster Size[b]	Varying Cluster Size[c]		
			SD = 100	SD = 200	SD = 300		SD = 100	SD = 200	SD = 300
0.05	0	13	13	13	13	17	17	17	17
	0.0001	14	14	14	14	17	17	18	18
	0.001	19	19	20	20	23	23	24	24
	0.005	42	43	43	44	56	57	58	59
	0.01	74	74	75	76	98	99	100	101
0.10	0	4	4	4	4	7	7	7	7
	0.0001	4	4	4	4	6	6	6	6
	0.001	8	8	8	8	8	8	8	9
	0.005	13	13	13	14	17	17	17	17
	0.01	21	21	21	22	25	25	25	26
0.15	0	3	3	3	3	3	3	3	3
	0.0001	3	3	3	3	3	3	3	3
	0.001	3	3	3	4	4	4	4	4
	0.005	7	7	7	8	9	9	9	9
	0.01	11	11	11	11	13	14	14	14

SD = Standard deviation.
ICC= Intra-cluster correlation coefficient.

[a] Effect size $= \frac{\textit{Mean of outcome in intervention arm} - \textit{Mean of outcome in control arm}}{\textit{pooled standard deviation}}$

[b] Average school size is 600.

[c] Allows school (cluster) size to vary utilizing standard deviation of cluster size.

from 0.0023 to 0.0070. The upper limit of CIs of the true ICC of all violence (victimization or perpetration or both) is 0.0048. Overall, the estimated ICCs from the GDHS study data is less than 0.005. The upper limits of corresponding ICC values are less than 0.01.

Results presented in Table 3, 4 and 5 focus on general effect size and ICC ranges based on values observed in Table 2. More tables are included in the Appendix. These tables demonstrate multiple important findings. First, as expected, magnitudes of effect sizes, ICC values, and desired power influence the calculation of the required number of schools. As calculated by the sample size formula, effect size has an inverse relationship with the required number of schools. The required number of clusters or schools drastically decreases as the effect size increases by only 0.05 (Table 3, 4 and 5). In contrast to effect sizes, the sample size formula dictates that the required number of schools increases with the ICC value. Similarly, the greater the desired power, the greater the required number of schools. Second, it is evident from the results in Table 3, 4 and 5 that the presence of even apparently negligible ICC values can have a considerable impact on the required number of clusters. For instance, a realistic ICC value of 0.005, which is seemingly negligible, more than doubled or tripled the required number of clusters across the settings. Third, accounting for variation in cluster sizes could increase the required number of schools by more than 1 per arm in order to avoid sub-nominal power. The higher the cluster size variation, the greater the impact on the required number of clusters.

4. Discussion

The effect sizes observed from this study indicate small, but notable, reductions in student-level dating violence victimization and perpetration in the intervention schools compared to the control schools. Student-level effect sizes for victimization and perpetration of physical DV in our study fall within the range of effect sizes calculated by Ref. [16] from other school-based dating violence studies in their review paper. However, the magnitude of effect size for the perpetration of physical DV in our study (−0.07) is larger compared to the effect size computed in their meta-analysis (−0.01; 95% CI: −0.04, 0.05). Furthermore, for the victimization of physical DV, the effect size in our study (−0.06) is smaller compared to their findings (−0.21; 95% CI: −0.41, −0.02) as their effect sizes were based on different dating violence programs. Overall, these results demonstrate that when designing future cRCTs for testing an intervention, it may be reasonable to assume the magnitude of effect size between 0.05 and 0.10 for SV and DV outcomes in school settings.

The ICCs observed in this study are small (less than 0.005), as is often the case when cluster sizes are large [30,33–35]. This is also consistent with previous findings that ICC and prevalence are associated [15]. Specifically, SV and DV at the student-level are not relatively frequent, and therefore small ICCs for corresponding violence outcomes are expected. Researchers may utilize these reported effect sizes and ICCs as guideline values when calculating the number of schools required for their future trials.

To calculate effect sizes for each outcome, we utilized standardized mean differences proposed by Ref. [24]. These effect size measures can be generalized despite the unit of measurements of these outcomes in different studies. Although these results are specific to sexual and dating violence in school-based settings, these results can further be generalized to other outcomes in school settings.

The current study provides one of the broadest collections of effect sizes and ICCs for variables relating to dating violence. Future research is needed with other study populations and violence outcomes to determine if similar effect sizes and ICC values are observed. In general, we encourage the publication of effect size, ICC, and cluster size information from cRCTs.

When there is a presence of natural clustering among subjects, accounting for it is a must. When cluster sizes are large, which is the case for school-based cluster designs, even very small ICC values have a notable impact on the required number of clusters through design effect, as design effect is a function of ICC values and mean cluster size, hence the total number of subjects. For instance, the design effect is $1 + (\overline{m} - 1)ICC$. For an average school size of 500 and an ICC value of 0.001, the design effect is 1.50. This means that the number of subjects needed per trial arm needs to be inflated by 1.5 times compared to when the ICC value was ignored (see page 56 of [3]. Therefore, even though the ICC values we observed appear small in magnitude, their impact can be quite notable, and hence even what may seem to be a negligible ICC should not be ignored when designing a cRCT.

Our sample size calculations used the average school size as well as variation in school sizes based on the GDHS study. At times, ignoring the school size variation can result in too few clusters, and hence sub-nominal power. As such, we strongly recommend the use of an appropriate sample size calculation, such as the use of the [3] function along with the [32] adjustment, as demonstrated in the Appendix.

Although the continuous outcomes of focus are not normally distributed, linear mixed effects models are robust and provide the correct formula for the ICC for continuous outcomes [25,26,36]. Furthermore, our models did not adjust for any covariates except for the baseline school-level mean of the given outcome variable to mimic a primary data analysis. However, secondary data analyses adjusting for other meaningful covariates may increase power via the potential reduction of the ICC value and/or the within-school variance of the given outcome. In short, adjustment for school-level covariates may reduce the between-school variance, hence lowering the ICC value and increasing power. Similarly, adjusting for student-level covariates may reduce within-school variance, thus increasing power.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical approval

The study protocol was approved by the University of Kentucky IRB (13-0680-F1V).

Conflicts of interest

The authors have no relevant financial or non-financial interests to disclose.

Informed consent

This study was based on secondary data analysis and as such informed consent was not required.

Contributions of each author in the preparation of this manuscript

Tofial Azam, Ph.D.: Methodology, Writing – Original Draft, Writing-Review & Editing, Data Curation, Resources, Formal analysis, Software, Heather M. Bush, Ph.D.: Writing- Review & Editing, Ann L. Coker, Ph.D.: Writing- Review & Editing, Philip M Westgate, Ph.D.: Conceptualization, Methodology, Supervision, Writing - Review & Editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.conctc.2021.100831.

References

[1] G. Adams, M.C. Gulliford, O.C. Ukoumunne, S. Eldridge, S. Chinn, M.J. Campbell, Patterns of intra-cluster correlation from primary care research to inform study design and analysis, J. Clin. Epidemiol. 57 (8) (2004) 785–794.

[2] D. Barnhart, E. Hertzmark, E. Liu, E. Mungure, A.N. Muya, D. Sando, D. Spiegelman, Intra-cluster correlation estimates for HIV-related outcomes from care and treatment clinics in Dar Es Salaam, Tanzania, Contemporary clinical trials communications 4 (2016) 161–169.

[3] J.T. Bertrand, Diffusion of innovations and HIV/AIDS, J. Health Commun. 9 (S1) (2004) 113–121.

[4] N. Beyler, M. Bleeker, S. James-Burdumy, J. Fortson, M. Benjamin, The impact of Playworks on students' physical activity during recess: findings from a randomized controlled trial, Prev. Med. 69 (2014) S20–S26.

[5] M. Breiding, K.C. Basile, S.G. Smith, M.C. Black, R.R. Mahendra, Intimate Partner Violence Surveillance: Uniform Definitions and Recommended Data Elements, Version 2.0, 2015.

[6] H.M. Bush, A.L. Coker, S. DeGue, E.R. Clear, C.J. Brancato, B.S. Fisher, Do violence acceptance and bystander actions explain the effects of Green Dot on reducing violence perpetration in high schools? J. Interpers Violence (2019), 0886260519888206.

[7] M.J. Campbell, S.J. Walters, How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research, John Wiley & Sons, 2014.

[8] J. Cohen, in: Statistical Power Analysis for the Behavioral Sciences, second ed., Lawrence Earlbaum Associates, Hillsdale, NJ, 1988.

[9] A.L. Coker, H.M. Bush, C.J. Brancato, E.R. Clear, E.A. Recktenwald, Bystander program effectiveness to reduce violence acceptance: RCT in high schools, J. Fam. Violence 34 (3) (2019) 153–164.

[10] A.L. Coker, H.M. Bush, P.G. Cook-Craig, S.A. DeGue, E.R. Clear, C.J. Brancato, E. A. Recktenwald, RCT testing bystander effectiveness to reduce violence, Am. J. Prev. Med. 52 (5) (2017) 566–578.

[11] A.L. Coker, P.G. Cook-Craig, C.M. Williams, B.S. Fisher, E.R. Clear, L.S. Garcia, L. M. Hegge, Evaluation of Green Dot: an active bystander intervention to reduce sexual violence on college campuses, Violence Against Women 17 (6) (2011) 777–796.

[12] P.G. Cook-Craig, A.L. Coker, E.R. Clear, L.S. Garcia, H.M. Bush, C.J. Brancato, B. S. Fisher, Challenge and opportunity in evaluating a diffusion-based active bystanding prevention program: green dot in high schools, Violence Against Women 20 (10) (2014) 1179–1202.

[13] P.G. Cook-Craig, P.H. Millspaugh, E.A. Recktenwald, N.C. Kelly, L.M. Hegge, A. L. Coker, T.S. Pletcher, From empower to Green Dot: successful strategies and lessons learned in developing comprehensive sexual violence primary prevention programming, Violence Against Women 20 (10) (2014) 1162–1178.

[14] L. De La Rue, J.R. Polanin, D.L. Espelage, T.D. Pigott, A meta-analysis of school-based interventions aimed to prevent or reduce violence in teen dating relationships, Rev. Educ. Res. 87 (1) (2017) 7–34.

[15] N. Demetrashvili, E.C. Wit, E.R. van den Heuvel, Confidence intervals for intraclass correlation coefficients in variance components models, Stat. Methods Med. Res. 25 (5) (2016) 2359–2376.

[16] A. Donner, N. Klar, Design and Analysis of Cluster Randomization Trials in Health Research, 2010 (New York).

[17] A. Donner, N. Birkett, C. Buck, Randomization by cluster: sample size requirements and analysis, Am. J. Epidemiol. 114 (6) (1981) 906–914.

[18] S.M. Eldridge, D. Ashby, S. Kerry, Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method, Int. J. Epidemiol. 35 (5) (2006) 1292–1300.

[19] S.M. Eldridge, O.C. Ukoumunne, J.B. Carlin, The intra-cluster correlation coefficient in cluster randomized trials: a review of definitions, Int. Stat. Rev. 77 (3) (2009) 378–394.

[20] G.L. Fellmeth, C. Heffernan, J. Nurse, S. Habibula, D. Sethi, Educational and skills-based interventions for preventing relationship and dating violence in adolescents and young adults: a systematic review, Campbell Systematic Reviews 9 (1) (2013) i–124.

[21] G.M. Fitzmaurice, N.M. Laird, J.H. Ware, Applied Longitudinal Analysis, vol. 998, John Wiley & Sons, 2012.

[22] B. Goesling, A practical guide to cluster randomized trials in school health research, J. Sch. Health 89 (11) (2019) 916–925.

[23] M.C. Gulliford, G. Adams, O.C. Ukoumunne, R. Latinovic, S. Chinn, M.J. Campbell, Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data, J. Clin. Epidemiol. 58 (3) (2005) 246–251.

[24] E.M. Hade, D.M. Murray, M.L. Pennell, D. Rhoda, E.D. Paskett, V.L. Champion, J. J. Fenton, Intraclass correlation estimates for cancer screening outcomes: estimates and applications in the design of group-randomized cancer screening studies, Journal of the National Cancer Institute Monographs, *2010* (40) (2010) 97–103.

[25] K.Y. Liang, S.L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1) (1986) 13–22.

[26] D.M. Murray, Design and Analysis of Group-Randomized Trials, vol. 29, Oxford University Press, USA, 1998.

[27] SAS Institute, Base SAS 9.4 Procedures Guide: Statistical Procedures, SAS Institute, 2017.

[28] M. Taljaard, A. Donner, J. Villar, D. Wojdyla, A. Velazco, V. Bataglia, G. Carroli, Intracluster correlation coefficients from the 2005 WHO global survey on maternal and perinatal health: implications for implementation research, Paediatr. Perinat. Epidemiol. 22 (2) (2008) 117–125.

[29] R.C. Team, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2019, 2012. URL, https://www.R-project.org.

[30] D.M. Thompson, D.H. Fernald, J.W. Mold, Intraclass correlation coefficients typical of cluster-randomized studies: estimates from the Robert Wood Johnson Prescription for Health projects, Ann. Fam. Med. 10 (3) (2012) 235–240.

[31] D.J. Torgerson, Contamination in trials: is cluster randomisation the answer? Bmj 322 (7282) (2001) 355–357.

[32] E.L. Turner, F. Li, J.A. Gallis, M. Prague, D.M. Murray, Review of recent methodological developments in group-randomized trials: part 1—design, Am. J. Publ. Health 107 (6) (2017) 907–915. Alteristic. (2019, January 8). Green Dot. Retrieved from, https://alteristic.org/services/green-dot/.

[33] G.J. Van Breukelen, M.J. Candel, Efficiency loss because of varying cluster size in cluster randomized trials is smaller than literature suggests, Stat. Med. 31 (4) (2012) 397–400.

[34] C. Weijer, J.M. Grimshaw, M.P. Eccles, A.D. McRae, A. White, J.C. Brehaut, M. Taljaard, The Ottawa statement on the ethical design and conduct of cluster randomized trials, PLoS Med. 9 (11) (2012), e1001346.

[35] P.M. Westgate, A readily available improvement over method of moments for intra-cluster correlation estimation in the context of cluster randomized trials and fitting a GEE–type marginal model for binary outcomes, Clin. Trials 16 (1) (2019) 41–51.

[36] D.A. Wolfe, C. Crooks, P. Jaffe, D. Chiodo, R. Hughes, W. Ellis, A. Donner, A school-based program to prevent adolescent dating violence: a cluster randomized trial, Arch. Pediatr. Adolesc. Med. 163 (8) (2009) 692–699.