Why Cluster-Randomised Community Health Trials are Underpowered: Design Effects, Intracluster Correlation and a Set of Reference Tables

Abdul Mukit

Assistant Professor, Nemcare Group of Institutions, Guwahati, Assam, India.

Corresponding author: abdulmukit13@gmail.com

Received: 13 February 2026; Revised: 24 April 2026; Accepted: 01 May 2026; Published: 18 July 2026

Abstract

Background: Community health interventions are usually allocated to groups — villages, clinics, schools, wards — rather than to individuals, and the resulting correlation between outcomes within a group reduces the information each additional participant contributes. Trials that ignore or underestimate this correlation are systematically underpowered, and a literature of inconclusive community trials is the predictable result.

Objective: To set out, for practitioners designing community health trials, the quantitative consequences of clustering, and to provide reference tables from which design parameters can be read directly.

Methods: Standard sample-size relationships for cluster-randomised designs with a continuous outcome were evaluated across the parameter ranges encountered in community health research. All tabulated and plotted values were computed analytically under two-sided testing at the 5 % level with 80 % power. Published empirical estimates of the intracluster correlation coefficient were collated from methodological studies to establish plausible input ranges.

Results: The design effect is linear in cluster size and in the intracluster correlation, so the penalty for clustering grows without bound as clusters get larger: at a correlation of 0.02 — near the middle of the published empirical range — clusters of 30 inflate the required sample by 58 %, and clusters of 200 inflate it fivefold. Enlarging clusters yields sharply diminishing returns: raising cluster size from 30 to 200 increases the effective information per cluster from 19 to 40 participants while trebling the number recruited. The number of clusters, not the number of participants, governs power, and at a correlation of 0.05 a standardised effect of 0.20 requires 33 clusters per arm. Unequal cluster sizes impose a further penalty reaching 38 to 67 % when the coefficient of variation of cluster sizes approaches unity. Published intracluster correlations span three orders of magnitude and are consistently larger for process outcomes than for clinical outcomes.

Conclusions: Most of the power problem in community trials is fixed at the design stage by the choice of number of clusters, and no analytic technique recovers it afterwards. Twelve reporting items are proposed, each verifiable by a reviewer from the manuscript alone.

Keywords: Cluster Randomised Trial, Design Effect, Intracluster Correlation, Sample Size, Statistical Power, Research Methodology.

1. INTRODUCTION

Most interventions in community health cannot be allocated to individuals. A water-supply improvement, a change in clinic staffing, a school-based curriculum, a village health-committee model — each is delivered to a group, and the group is therefore the unit of allocation whether or not the analyst treats it as such. Randomising groups rather than individuals is often the only defensible design and frequently the only feasible one.

The statistical cost is that outcomes within a group are not independent. People in the same village share a water source, a market, a set of health workers, a local disease ecology and a set of social norms; patients

in the same clinic share a clinician's habits. Their outcomes are therefore more similar to one another than to outcomes drawn at random from the population, and each additional person recruited from a group already in the study adds less new information than the first person did. Because the conventional sample-size formula assumes independent observations, applying it unmodified to a clustered design produces a sample that is too small — sometimes by a factor of ten — and a trial that cannot answer its own question.
This is textbook material and is nevertheless a persistent source of failure in practice. Three reasons recur. The required inputs are unfamiliar: an investigator can usually state a plausible effect size but rarely a plausible intracluster correlation, and in the absence of a value the temptation is to assume it negligible. The arithmetic is counterintuitive, in that correlations of 0.02 — which sound negligible — produce substantial inflation once cluster sizes reach the tens or hundreds. And the intuitive response to a power problem, recruiting more people, is largely ineffective here, because power in a clustered design is governed by the number of clusters rather than the number of participants.
This review sets out those relationships quantitatively and supplies reference tables from which design parameters can be read directly. It is aimed at investigators and reviewers in community health rather than at statisticians, and it is deliberately generic: the relationships are properties of the design, not of any disease, setting or health system.

2. METHODS

The quantities reported here are analytic, not empirical. Standard sample-size relationships for parallel-group cluster-randomised trials with a continuous outcome were evaluated across the parameter ranges encountered in community health research. All computations assume two-sided testing at the 5 % significance level with 80 % power, equal allocation between two arms, complete follow-up, and analysis at the individual level with appropriate adjustment for clustering. Effects are expressed as standardised mean differences so that results are scale-free.
Plausible ranges for the intracluster correlation were established by collating published empirical estimates from methodological studies reporting correlations from completed trials and routine datasets. These sources are described in Section 3.5 and tabulated in Table 6; they are used to bound the parameter space, not to estimate any pooled value.
Table 1 defines the notation used throughout.

Table 1. Notation

Symbol	Quantity	Typical value in community health
ρ	Intracluster correlation coefficient	0.005 to 0.10; higher for process outcomes
m	Average number of participants per cluster	10 to 200
k	Number of clusters per arm	4 to 60
DE	Design effect, or variance inflation factor	1.1 to 20
d	Standardised effect size (mean difference divided by SD)	0.20 to 0.50
CV	Coefficient of variation of cluster sizes	0 to 1.0
α	Two-sided significance level	0.05
1 − β	Power	0.80
n_eff	Effective sample size per cluster, m / DE	always less than m

3. RESULTS

3.1 The design effect

The variance of a treatment-effect estimate from a clustered design exceeds that from an individually randomised design of the same size by the design effect,

$$DE = 1 + (m - 1)\rho \qquad (1)$$

where m is the average cluster size and ρ the intracluster correlation. Two properties of (1) drive everything that follows. It is linear in cluster size, so the penalty has no ceiling: doubling cluster size approximately doubles the excess variance. And it depends on the product of cluster size and correlation,

so a correlation small enough to seem negligible becomes consequential once clusters are large. Figure 1 and Table 2 give the design effect across the relevant parameter space.

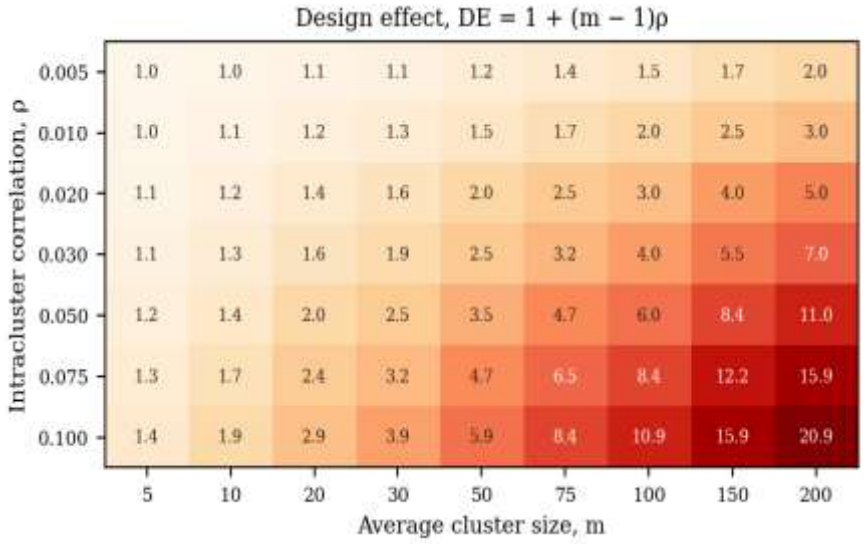

Figure 1. Design effect as a function of cluster size and intracluster correlation. Shading is on a logarithmic scale

Table 2. Design effect, DE = 1 + (m − 1)ρ

ρ	m = 10	m = 20	m = 30	m = 50	m = 100	m = 200
0.005	1.04	1.09	1.15	1.25	1.50	2.00
0.010	1.09	1.19	1.29	1.49	1.99	2.99
0.020	1.18	1.38	1.58	1.98	2.98	4.98
0.050	1.45	1.95	2.45	3.45	5.95	10.95
0.100	1.90	2.90	3.90	5.90	10.90	20.90

Note. Read the design effect as the multiplier applied to the sample size an individually randomised trial would require. A design effect of 3.90 means the clustered trial needs 3.9 times as many participants for the same power.

The entries in the lower right of Table 2 are the ones that surprise investigators. A correlation of 0.10 with clusters of 200 — a plausible combination when randomising large clinics on a process-of-care outcome — requires twenty-one times the sample size of the corresponding individually randomised trial. A trial designed without that multiplier will have power in the region of 10 to 15 % where it believed it had 80 %.

3.2 Sample Size and the Number of Clusters

For a continuous outcome, the number of clusters required per arm is

$$k = 2(z_{1-\alpha/2} + z_{1-\beta})^2 \times DE / (m\, d^2) \qquad (2)$$

and the corresponding minimum detectable effect for a given design is

$$d_{min} = (z_{1-\alpha/2} + z_{1-\beta}) \sqrt{(2\, DE / (k\, m))} \qquad (3)$$

Table 3 evaluates (2) for clusters of 30, a common size when randomising villages, schools or clinics.

Table 3. Clusters required per arm, cluster size m = 30, 80 % power, α = 0.05

ρ	d = 0.20	d = 0.30	d = 0.40	d = 0.50
0.005	15	7	4	3
0.010	17	8	5	3
0.020	21	10	6	4

0.050	33	15	9	6
0.100	52	23	13	9

Note. Values are rounded up to the next whole cluster. Designs requiring fewer than about eight clusters per arm should be treated with caution: the normal approximation underlying (2) performs poorly with few clusters, and a t-based correction or an exact method should be used.

Table 3 contains the practical message of this review. Detecting a small effect of 0.20 requires 33 clusters per arm at a correlation of 0.05, which is 66 clusters in total and roughly 2,000 participants. Many published community trials randomise eight or twelve clusters, which Table 3 shows is adequate only for effects in the region of 0.40 to 0.50 — effects large enough that a trial is rarely needed to detect them.

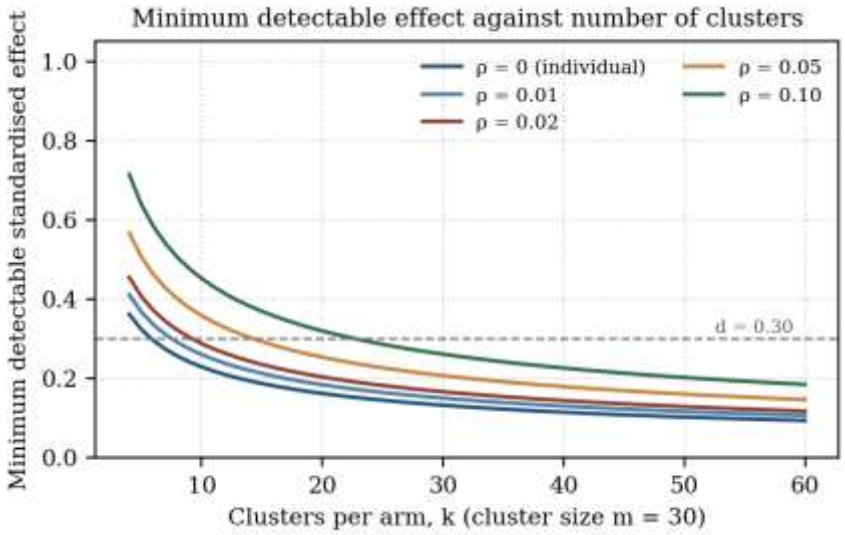

Figure 2. Minimum detectable standardised effect against number of clusters per arm, for cluster size 30. The dashed line marks a standardised effect of 0.30

Figure 2 shows why adding a few clusters to an already small trial rarely rescues it. The curves are steep at low cluster counts and flat thereafter, so the marginal cluster is worth a great deal when there are eight of them and very little when there are forty. The corresponding power curves appear in Figure 3.

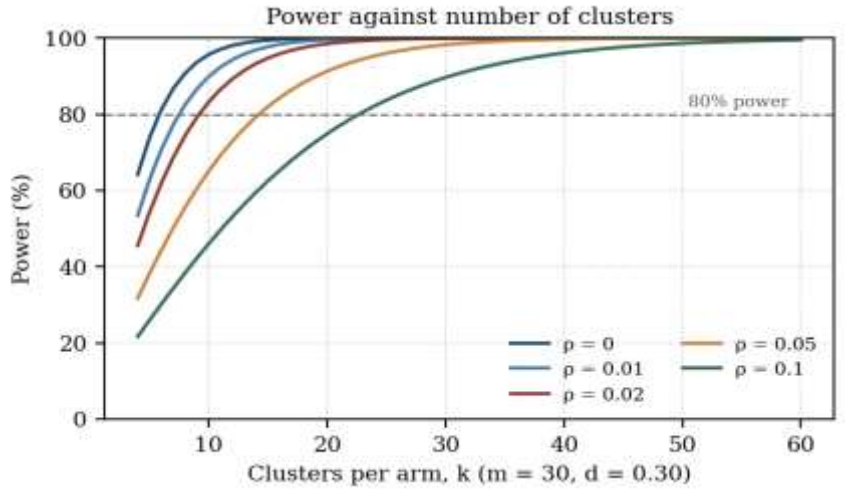

Figure 3. Power against number of clusters per arm for a standardised effect of 0.30 and cluster size 30.

The separation between the curves in Figure 3 is the cost of the correlation. At twenty clusters per arm a design with no clustering would have essentially complete power to detect an effect of 0.30, while the same design with a correlation of 0.10 has roughly half that. The difference is created entirely at the design stage and is not recoverable by any choice of analysis.

3.3 Why Recruiting More People per Cluster Does Not Help

When a cluster trial is found to be underpowered, the instinctive remedy is to recruit more participants from the clusters already enrolled. Equation (1) explains why this works poorly. Adding participants raises m, which raises the design effect proportionally, so the additional participants are progressively discounted. The effective sample size per cluster,

$n_{eff} = m / DE = m / (1 + (m - 1)\rho)$ (4)

approaches an asymptote of $1/\rho$ as cluster size grows without limit. At a correlation of 0.05 that ceiling is twenty participants: a cluster of a thousand people carries no more information about the treatment effect than a cluster of about twenty, because everyone beyond the first twenty is telling the analyst substantially what the first twenty already said. Table 4 evaluates this for a fixed design target.

Table 4. Effect of cluster size on total sample requirement (d = 0.30, ρ = 0.02, 80 % power)

Cluster size m	Clusters per arm k	Participants per arm	Design effect	Effective n per cluster
10	21	210	1.18	8.5
20	13	260	1.38	14.5
30	10	300	1.58	19.0
50	7	350	1.98	25.3
100	6	600	2.98	33.6
200	5	1000	4.98	40.2

Note. Moving from clusters of 30 to clusters of 200 more than trebles the number of participants recruited while reducing the number of clusters only from ten to five, and the effective information per cluster rises from 19 to 40. The trial becomes far more expensive and only slightly more informative.

Table 4 shows the trade-off in its starkest form. If clusters are cheap to enrol and participants expensive, many small clusters is the efficient design. If the reverse holds — which is common when clusters are whole facilities requiring negotiated access — the temptation is to enrol few large clusters, and the table shows what that costs. Figure 4 plots the sample-size relationship across the full range of cluster sizes, and Figure 5 shows the underlying information loss.

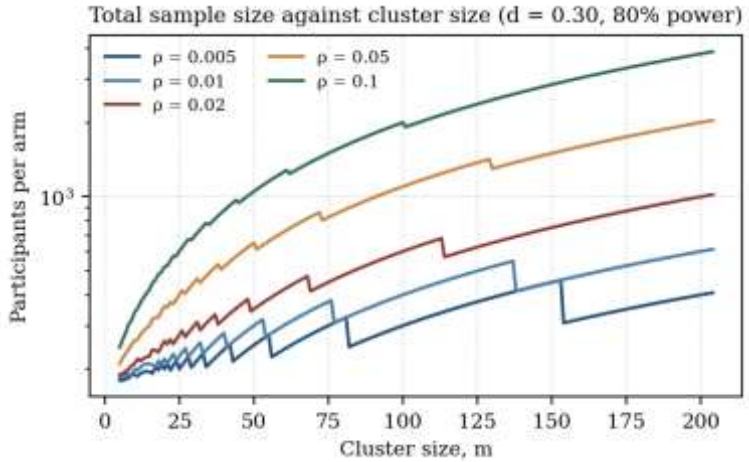

Figure 4. Participants required per arm against cluster size for a standardised effect of 0.30, on a logarithmic vertical scale

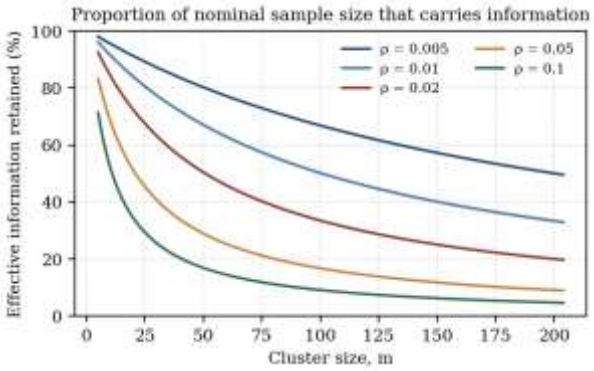

Figure 5. Proportion of the nominal sample size that carries independent information, as a function of cluster size

Figure 5 states the problem in its most compact form. At a correlation of 0.05 and clusters of 100, only about 17 % of the nominal sample size is contributing independent information; the remaining 83 % of recruitment cost buys very little. This is the quantity an investigator should compute before committing to a recruitment target, and it is rarely reported.

3.4 Unequal Cluster Sizes

Clusters in community health are rarely of equal size: villages differ in population, clinics in caseload, schools in enrolment. Variation in cluster size inflates the design effect further, because the effective sample size is dominated by the smaller clusters while the larger contribute less than proportionally. A widely used approximation replaces (1) with

$$DE = 1 + ((CV^2 + 1)\bar{m} - 1)\rho \qquad (5)$$

where CV is the coefficient of variation of cluster sizes and $\bar{m}$ the mean cluster size. Table 5 evaluates the penalty.

Table 5. Additional inflation from unequal cluster sizes

CV of cluster sizes	m = 30, ρ = 0.02	m = 30, ρ = 0.05	m = 100, ρ = 0.02
0.0 (equal)	1.58 (reference)	2.45 (reference)	2.98 (reference)
0.2	1.60 (+2 %)	2.51 (+2 %)	3.06 (+3 %)
0.4	1.68 (+6 %)	2.69 (+10 %)	3.30 (+11 %)
0.6	1.80 (+14 %)	2.99 (+22 %)	3.70 (+24 %)
0.8	1.96 (+24 %)	3.41 (+39 %)	4.26 (+43 %)
1.0	2.18 (+38 %)	3.95 (+61 %)	4.98 (+67 %)

Note. Percentages are the additional sample size required relative to the equal-size design at the same mean cluster size and correlation.

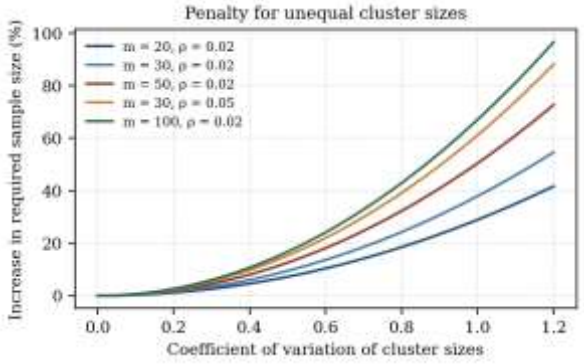

Figure 6. Additional sample size required as a function of the coefficient of variation of cluster sizes

A coefficient of variation below about 0.2 can reasonably be ignored, adding only two or three per cent. Above about 0.6 the penalty becomes material, and in community settings such variation is ordinary rather than exceptional: a set of villages ranging from 200 to 2,000 inhabitants will comfortably exceed it. The practical response is either to stratify or restrict on cluster size at the design stage, or to build the inflation from (5) into the sample-size calculation explicitly.

3.5 What Values of the Correlation Should Be Assumed

Equations (1) to (5) require a value for the intracluster correlation, and the investigator will usually not have one for the specific outcome and setting. Published empirical work provides bounds. Table 6 collates estimates from methodological studies, and Figure 7 displays them.

Table 6. Published empirical estimates of the intracluster correlation

Source and context	Number of estimates	Central value	Range or IQR
Reanalysis of primary-care datasets	1,039 variables from 31 studies	0.010 (median)	IQR 0 to 0.032; range 0 to 0.840
Implementation research datasets	220 ICCs from 21 datasets	—	0 to 0.415
Primary care, clinical measures	Multiple outcomes, one trial	—	0.020 to 0.047 (DE 2.6 to 4.1)
Primary care, height and diastolic BP	Two outcomes, one trial	—	0.05 to 0.12 (DE 4.4 to 9.4)
Diabetes cohort, HbA1c continuous	Routine data, 430 practices	0.032	95 % CI 0.026 to 0.038
Diabetes trials, values assumed at design	Systematic search of published trials	0.047 (median)	IQR 0.047 to 0.050
Chronic disease clinics, clinical outcomes	Multiple outcomes, 40 clinics	—	0.01 to 0.09
Chronic disease clinics, processes of care	Multiple outcomes, 40 clinics	—	0.01 to 0.48
Hospital care pathways, baseline characteristics	Multiple variables	0.043 (median)	IQR 0.026 to 0.052

Note. Values are as reported by the original investigators. Where a source reported only a range, no central value is given. DE = design effect as reported in that source.

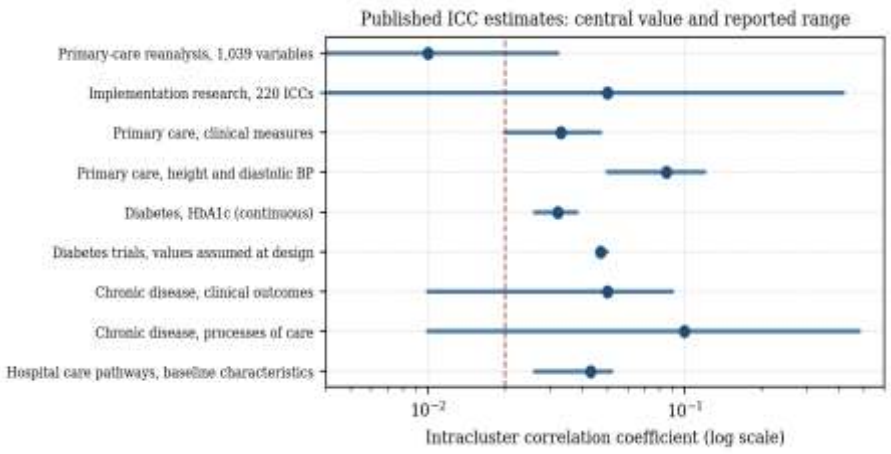

Figure 7. Published intracluster correlation estimates on a logarithmic scale. Markers show the reported central value where one was given and the midpoint of the reported range otherwise; bars show the reported range or interquartile range

Four regularities emerge from Table 6 and are worth carrying into any design. Correlations are usually small in absolute terms, with medians around 0.01 to 0.05, which is why they are so often dismissed — and Table 2 shows why dismissing them is a mistake. Their range is enormous, spanning zero to above 0.4, so a point estimate borrowed from an unrelated study is a weak basis for a design. Process outcomes carry systematically higher correlations than clinical outcomes, consistent with process being determined largely by the practice of the provider, who is shared by everyone in the cluster; one source reports 0.01 to 0.48 for processes against 0.01 to 0.09 for clinical outcomes in the same clinics. And investigators designing trials tend to assume more conservative values than the data subsequently support, which is the correct direction in which to err.

The practical recommendation is to compute the sample size across a range of plausible correlations rather than at a single value, to report the range, and to select the design from the pessimistic end. Where a correlation must be assumed without local data, a value of 0.02 for clinical outcomes and 0.05 to 0.10 for process outcomes is consistent with the published distribution, and should be stated as an assumption rather than presented as a parameter.

3.6 Reporting

Table 7 sets out twelve items, each of which a reviewer can verify from the manuscript alone and none of which requires access to data.

Table 7. Reporting items for sample size in cluster-randomised community trials

#	Item	Why it matters
1	Unit of randomisation stated explicitly	Determines whether clustering applies at all
2	Number of clusters per arm	The quantity that governs power
3	Average cluster size and its expected variation	Required for (1) and (5)
4	Assumed intracluster correlation, with its source	Without it the calculation cannot be checked
5	Design effect used, stated numerically	Makes the inflation visible rather than implicit
6	Target effect size and its justification	Distinguishes a detectable effect from an important one
7	Power and significance level assumed	Standard, but frequently omitted
8	Sample size under a range of correlations	Shows sensitivity to the least certain input
9	Whether analysis accounts for clustering, and how	Individual-level analysis ignoring clusters inflates type I error
10	Handling of unequal cluster sizes	Table 5 shows the penalty can exceed 50 %
11	Allowance for cluster-level attrition	Losing a cluster costs far more than losing participants
12	Observed intracluster correlation reported at analysis	Supplies the empirical values future trials need

Note. Item 12 is the one most often omitted and the one with the greatest external value: the scarcity documented in Table 6 exists because trials do not routinely report the correlations they observe.

4. DISCUSSION

The relationships set out above are not new, and none of this review's arithmetic is contestable. What is worth restating is the asymmetry between when the problem is created and when it is discovered. Power in a cluster trial is fixed almost entirely by the number of clusters, which is determined at the design stage, usually under budgetary and political constraints that have nothing to do with statistical considerations. By the time results are analysed, no technique recovers the lost information. Mixed models, generalised estimating equations and permutation tests all handle clustering correctly; correct handling of a small number of clusters yields a correctly computed wide confidence interval.

This has a consequence for how null results from community trials should be read. A trial of twelve clusters per arm reporting no effect on a clinical outcome has, by Table 3, adequate power only for effects of around 0.40. If the intervention's plausible effect is 0.20, the trial was never capable of distinguishing that effect from nothing, and its null result is uninformative rather than negative. Since such trials are common and their abstracts rarely say this, the community health literature accumulates apparent failures that are in fact non-tests. The minimum detectable effect, computed from (3) with the observed design, should be reported alongside every null finding; it takes one line and changes the interpretation.
Table 8 lists the errors that recur, with their consequences.

Table 8. Common errors in cluster trial design and their consequences

Error	Consequence	Remedy
Sample size computed as if individually randomised	Power overstated by the factor in Table 2; trial cannot detect its target effect	Apply the design effect; state it numerically
Intracluster correlation assumed to be zero	Same as above; usually implicit rather than stated	Assume 0.02 for clinical and 0.05 to 0.10 for process outcomes
Power sought by enlarging clusters	Cost rises steeply, information barely rises (Table 4)	Add clusters, not participants per cluster
Very few clusters per arm	Normal approximation fails; chance imbalance in cluster characteristics	Use t-based or exact methods; consider stratified allocation
Unequal cluster sizes ignored	Requirement understated by up to 67 % (Table 5)	Use equation (5); restrict or stratify on size
Analysis ignores clustering	Standard errors too small; type I error inflated	Mixed model, GEE, or cluster-level summary analysis
Attrition modelled at participant level only	Loss of a whole cluster is not anticipated	Inflate the number of clusters, not only participants
Observed correlation not reported	Future trials in the same field have no design input	Report the observed ICC with the primary result

4.1 Limitations

Four limitations bound the scope of these results. All computations assume a continuous outcome analysed as a mean difference; binary outcomes follow the same structure but require the variance to be expressed through the expected proportions, and rare-event outcomes behave differently in ways not addressed here. All computations assume a parallel two-arm design, whereas stepped-wedge and other longitudinal cluster designs involve within-period and between-period correlations that decay over time, which equation (1) does not represent. The approximation in (5) for unequal cluster sizes is one of several in use and performs less well at extreme variation. And the collation in Table 6 is illustrative rather than systematic: sources were selected to span the range of contexts encountered in community health, not identified by protocol, so Table 6 should be used to bound plausible values and not as a meta-analysis.
A broader limitation is that this review addresses only statistical power. A trial can be adequately powered and still uninformative if the intervention is poorly specified, if contamination between clusters dilutes the contrast, or if the outcome measured is not the outcome that matters. Power is necessary and not sufficient.

5. CONCLUSION

Clustering is not a nuisance to be corrected at the analysis stage; it is a design parameter that determines whether a trial can answer its question. The arithmetic is unforgiving in a specific direction: correlations that sound negligible produce large inflations once clusters are of realistic size, additional participants within existing clusters buy progressively less information, and the effective sample size per cluster is bounded above by the reciprocal of the correlation regardless of how many people are recruited.

Three consequences follow for practice. Design should proceed from the number of clusters available, not from the number of participants obtainable. The intracluster correlation should be treated as a range to be explored rather than a constant to be assumed, with the design taken from the pessimistic end. And every trial should report the correlation it observed, so that the next investigator is not designing from the same absence of information. The reference tables supplied here are intended to make the first two of those steps a matter of reading a value rather than deriving one.

Declarations

Ethical approval: Not applicable.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of interes: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author contributions:

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Abdul Mukit	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

REFERENCES

1. A. Donner and N. Klar, Design and Analysis of Cluster Randomization Trials in Health Research. London: Arnold, 2000. doi.org/10.1191/096228000669355658
2. M. K. Campbell, G. Piaggio, D. R. Elbourne, and D. C. Altman, 'CONSORT 2010 statement: extension to cluster randomised trials', BMJ, vol. 345, 2012. doi.org/10.1136/bmj.e5661
3. S. M. Eldridge, D. Ashby, and S. Kerry, 'Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method', Int J Epidemiol, vol. 35, no. 5, pp. 1292-1300, 2006. doi.org/10.1093/ije/dyl129
4. G. Adams, M. C. Gulliford, O. C. Ukoumunne, S. Eldridge, S. Chinn, and M. J. Campbell, 'Patterns of intra-cluster correlation from primary care research to inform study design and analysis', J Clin Epidemiol, vol. 57, no. 8, pp. 785-794, 2004. doi.org/10.1016/j.jclinepi.2003.12.013
5. M. K. Campbell, P. M. Fayers, and J. M. Grimshaw, 'Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research', Clin Trials, vol. 2, no. 2, pp. 99-107, 2005. doi.org/10.1191/1740774505cn071oa
6. D. R. Parker, E. Evangelou, and C. B. Eaton, 'Intraclass correlation coefficients for cluster randomized trials in primary care: the cholesterol education and research trial (CEART)', Contemp Clin Trials, vol. 26, no. 2, pp. 260-267, 2005. doi.org/10.1016/j.cct.2005.01.002
7. R. Littleford and I. Nazareth, 'Intra-cluster and inter-period correlation coefficients for cross-sectional cluster randomised controlled trials for type-2 diabetes in UK primary care', Trials, vol. 17, 2016. doi.org/10.1186/s13063-016-1532-9
8. R. Van Zelm, 'Intraclass correlation coefficients for cluster randomized trials in care pathways and usual care: hospital treatment for heart failure', BMC Health Serv Res, vol. 14, 2014. doi.org/10.1186/1472-6963-14-84
9. O. C. Ukoumunne, M. C. Gulliford, S. Chinn, J. Sterne, and P. Burney, 'Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review', Health Technol Assess, vol. 3, no. 5, 1999. doi.org/10.3310/hta3050

10. S. Killip, Z. Mahfoud, and K. Pearce, 'What is an intracluster correlation coefficient? Crucial concepts for primary care researchers', Ann Fam Med, vol. 2, no. 3, pp. 204-208, 2004. doi.org/10.1370/afm.141
11. C. Rutterford, A. Copas, and S. Eldridge, 'Methods for sample size determination in cluster randomized trials', Int J Epidemiol, vol. 44, no. 3, pp. 1051-1067, 2015. doi.org/10.1093/ije/dyv113
12. S. M. Kerry and J. M. Bland, 'The intracluster correlation coefficient in cluster randomisation', BMJ, vol. 316, no. 7142, 1998. doi.org/10.1136/bmj.316.7142.1455
13. L. Guittet, P. Ravaud, and B. Giraudeau, 'Planning a cluster randomized trial with unequal cluster sizes: practical issues involving continuous outcomes', BMC Med Res Methodol, vol. 6, 2006. doi.org/10.1186/1471-2288-6-17
14. K. Hemming, S. Eldridge, G. Forbes, C. Weijer, and M. Taljaard, 'How to design efficient cluster randomised trials', BMJ, vol. 358, 2017. doi.org/10.1136/bmj.j3064
15. S. M. Eldridge, O. C. Ukoumunne, and J. B. Carlin, 'The intra-cluster correlation coefficient in cluster randomized trials: a review of definitions', Int Stat Rev, vol. 77, no. 3, pp. 378-394, 2009. doi.org/10.1111/j.1751-5823.2009.00092.x
16. M. Taljaard, A. Donner, and N. Klar, 'Accounting for expected attrition in the planning of community intervention trials', Stat Med, vol. 26, no. 13, pp. 2615-2628, 2007. doi.org/10.1002/sim.2733