Appendix 1: Formula and notation

We make large sample size approximations and assume normality throughout. We also assume equal cluster sizes. Count outcomes are expressed as rates per unit of person-time, this means each cluster has the same number of individuals and the same person time of follow-up. Of note, ICCs for binary (or rate outcomes) should be calculated on the proportions (or rate) scale and not on the logit or log scale.

Notation

n_I	Sample size per arm under individual randomisation
m	Cluster size
ρ	Intra-cluster correlation (ICC)
k	No. clusters per arm required
$2mk$	Total sample size
α	Significance level
$1-\beta$	Power

Formula

Continuous outcomes

δ	Target difference
σ^2	Variance of outcome
$n_I = 2\sigma^2((z_{\alpha/2}+z_\beta)^2/\delta^2)$	Sample size under individual randomisation
$(4\sigma^2(1+(m-1)\rho)/2mk$	Variance of treatment effect
$2mk/(4\sigma^2(1+(m-1)\rho)$	Total precision of treatment effect
$m/(2\sigma^2(1+(m-1)\rho)$	Precision per cluster of treatment effect

Binary outcomes

p_0	Proportion in control arm
p_1	Proportion in intervention arm
$\delta = p_1 - p_0$	Target difference
$\sigma^2 = (p_0(1-p_0)+p_1(1-p_1))$	Variance of outcome
$n_I = 2\sigma^2((z_{\alpha/2}+z_\beta)^2/\delta^2)$	Sample size under individual randomisation
$(4\sigma^2(1+(m-1)\rho)/2mk$	Variance of treatment effect
$2mk/(4\sigma^2(1+(m-1)\rho)$	Total precision of treatment effect
$m/(2\sigma^2(1+(m-1)\rho)$	Precision per cluster of treatment effect

Count outcomes

r_0	Rate in control arm (per person per unit time period)
r_1	Rate in intervention arm (per person per unit time period)
m	Cluster size = number of persons * person follow-up time
$\delta = r_1 - r_0$	Target difference
$\sigma^2 = r_0 + r_1$	Variance of outcome
$n_I = 2\sigma^2((z_{\alpha/2}+z_\beta)^2/\delta^2)$	Sample size under individual randomisation
$(4\sigma^2(1+(m-1)\rho)/2mk$	Variance of treatment effect
$2mk/(4\sigma^2(1+(m-1)\rho)$	Total precision of treatment effect
$m/(2\sigma^2(1+(m-1)\rho)$	Precision per cluster of treatment effect

Where z_β represents the area to the left of the value β on a standard normal distribution curve.

Appendix 2: Illustration of how to construct power and precision curves

Power and precision curves can be easily programmed in a spreadsheet for a visual inspection of the point of diminishing returns. The code below will construct power and precision curves in Stata for continuous, binary and rate outcomes. A similar Excel file is available from the authors which will construct the same curves in Excel.

The power curve is a plot of the cluster size against the corresponding power achievable at that cluster size. The power will depend, in addition to the cluster size, on the effect size, number of clusters, ICC and significance level. The precision curve is a plot of the cluster size against the corresponding precision achievable at that cluster size. For continuous outcomes, precision curves can thus be useful when the exact magnitude of the target difference is unknown, and they allow demonstration of the point at which confidence intervals around resulting effect sizes would no longer reduce. For binary outcomes, precision curves do depend on the target difference.

Stata code to plot power and precision curves – continuous variables

```
*Inputs (values in red have to be inputted by user)
local k = 3 // number of clusters per arm
local delta = 0.55 // minimally important clinical difference
local a = 0.05 // alpha level
local ro = 0.03 // intra-cluster correlation coefficient
local sigma = 1 // standard deviation
local  es = `delta'/`sigma' // effect size
local x1 = 1 // lowest cluster size considered
local x2 = 400 // highest cluster size considered

*Graph of Power vs Cluster size & Precision vs Cluster size
*Precision per cluster
twoway function power = normal(sqrt(x*`k'*(`es'^2)/(2*(1+(x-1)*`ro')))-invnormal(1-
`a'/2)),range(`r1' `r2') ylabel(#10) xtitle("cluster size") ytitle("Power") || ///
          function precision = x/(2*`sigma'^2*(1+(x-1)*`ro')), yaxis(2) range(`x1' `x2')  xtitle("cluster
size")   ytitle("Precision per cluster", axis(2)) ylabel(#6, axis(2)) ///
          title("effect size = `es', ICC = `ro', `k' clusters per arm")
*Total precision
twoway function power = normal(sqrt(x*`k'*(`es'^2)/(2*(1+(x-1)*`ro')))-invnormal(1-
`a'/2)),range(`r1' `r2') ylabel(#10) xtitle("cluster size") ytitle("Power") || ///
          function precision = x*`k'/(2*sigma^2*(1+(x-1)*`ro')), yaxis(2) range(`x1' `x2')  xtitle("cluster
size")   ytitle("Precision", axis(2)) ylabel(#6, axis(2)) ///
          title("effect size = `es', ICC = `ro', `k' clusters per arm")
```

Stata code to plot power and precision curves – binary variables

```
*Inputs (values in red have to be inputted by user)
local k = 15 // number of clusters per arm
local p0 = 0.23 // proportion in control arm
local p1 = 0.44 // proportion in intervention arm
local a = 0.05 // alpha level
local ro = 0.3 // intra-cluster correlation coefficient
local delta=`p1'-`p0'
```

```
local sigma0= `p0'*(1-`p0')
local sigma1= `p1'*(1-`p1')
local sigma2=`sigma0'+`sigma1'
local x1 = 1 // lowest cluster size considered
local x2 = 1400 // highest cluster size considered


*Precision per cluster
twoway function power = normal(`delta'*sqrt((x*`k')/(`sigma2'*(1+((x-1)*`ro'))))-invnormal(1-
`a'/2)),range(`r1' `r2') ylabel(0 0.2 0.4 0.6 0.8 1) xtitle("cluster size") ytitle("Power") || ///
            function precision = x/(2*`sigma2'*(1+(x-1)*`ro')), yaxis(2) range(`x1' `x2')  xtitle("cluster
size")   ytitle("Precision per cluster", axis(2)) ylabel(#6, axis(2)) ///
            title("effect size = `p0' vs `p1', ICC = `ro', `k' clusters per arm")
*Total precision
twoway function power = normal(`delta'*sqrt((x*`k')/(`sigma2'*(1+((x-1)*`ro'))))-invnormal(1-
`a'/2)),range(`r1' `r2') ylabel(0 0.2 0.4 0.6 0.8 1) xtitle("cluster size") ytitle("Power") || ///
            function precision = (x*`k')/(2*`sigma2'*(1+(x-1)*`ro')), yaxis(2) range(`x1' `x2')
xtitle("cluster size")   ytitle("Precision", axis(2)) ylabel(#6, axis(2)) ///
            title("effect size = `p0' vs `p1', ICC = `ro', `k' clusters per arm")
```

Stata code to plot power and precision curves – count variables

```
*Inputs (values in red have to be inputted by user)
local k = 15 // number of clusters per arm
local r0 = 0.01 // rate in control arm
local r1 = 0.062 // rate in intervention arm
local a = 0.05 // alpha level
local ro = 0.03 // intra-cluster correlation coefficient
local delta=`r1'-`r0'
local sigma2=`r0'+`r1'
local x1 = 1 // lowest cluster size considered
local x2 = 1800 // highest cluster size considered

*Precision per cluster
twoway function power = normal(`delta'*sqrt((x*`k')/(`sigma2'*(1+((x-1)*`ro'))))-invnormal(1-
`a'/2)),range(`r1' `r2') ylabel(0 0.2 0.4 0.6 0.8 1) xtitle("cluster size") ytitle("Power") || ///
            function precision = x/(2*`sigma2'*(1+(x-1)*`ro')), yaxis(2) range(`x1' `x2')  xtitle("cluster
size")   ytitle("Precision per cluster", axis(2)) ylabel(#6, axis(2)) ///
            title("effect size = `p0' vs `p1', ICC = `ro', `k' clusters per arm")
*Total precision
twoway function power = normal(`delta'*sqrt((x*`k')/(`sigma2'*(1+((x-1)*`ro'))))-invnormal(1-
`a'/2)),range(`r1' `r2') ylabel(0 0.2 0.4 0.6 0.8 1) xtitle("cluster size") ytitle("Power") || ///
            function precision = (x*`k')/(2*`sigma2'*(1+(x-1)*`ro')), yaxis(2) range(`x1' `x2')
xtitle("cluster size")   ytitle("Precision", axis(2)) ylabel(#6, axis(2)) ///
            title("effect size = `p0' vs `p1', ICC = `ro', `k' clusters per arm")
```

Appendix 3: Derivation of "A simple rule to determine if a minimal increase in the number of clusters can lead to a significant reduction in cluster size"

In this section we derive the result presented in the paper to determine

Notation

n_I	Sample size per arm under individual randomisation
m	Cluster size
ρ	Intra-cluster correlation (ICC)
k	No. clusters per arm required
$\lceil x \rceil$	Value x rounded up to nearest integer value

Lower bound for number of clusters per arm

The lower bound for the number of clusters per arm required to detect a difference for a given power and (two-sided) significance level is$\lceil n_I \rho \rceil$. This is an accepted and known result (see references cited in main paper). Technically this minimum number of clusters requires the cluster size to be infinite, but in practice the cluster size needed to achieve this minimum will be a non-infinite number.

Cluster size when number of clusters is C more than minimum number of clusters

For equally sized clusters the number of observations in k equally sized clusters required to detect a difference for a given power and significance level is [Hemming et al 2011]:

$$m = \left\lceil \frac{(1-\rho)n_I}{k - n_I \rho} \right\rceil$$

If $k = \lceil n_I \rho \rceil + C$, that is if the number of clusters is C more than the lower bound for the number of clusters required, then the following inequality holds:

$$m = \left\lceil \frac{(1-\rho)n_I}{\lceil n_I \rho \rceil + C - n_I \rho} \right\rceil \leq \left\lceil \frac{(1-\rho)n_I}{n_I \rho + C - n_I \rho} \right\rceil = \left\lceil \frac{(1-\rho)n_I}{C} \right\rceil \leq \left\lceil \frac{n_I}{C} \right\rceil$$

This therefore means that if we have C more clusters than the minimum number of clusters required per arm then the required cluster size will be less than $\frac{n_I}{C}$.

Number of clusters required when $m = n_I/C$

In general the number of clusters per arm required to detect a difference for a given power and significance level per arm is [Hemming et al 2011]:

$$k = \left\lceil \frac{n_I(1 + (m-1)\,\rho)}{m} \right\rceil$$

If the cluster size is equal to $m = n_I/C$ then the number of clusters (k) needed is such that:

$$k = C\left(1 + \left(\frac{n_I}{C} - 1\right)\rho\right) = (C + (n_I - C)\rho) \leq C + \lceil \rho n_I \rceil$$

Therefore the number of clusters required is at most C more than the minimum number of clusters needed (assuming an infinite cluster size). From this it therefore follows that when the cluster size is equal to the sample size per arm under individual randomisation, the number of clusters needed is at most one more than the minimum number of clusters required.