Water Sensitive Cities proposal: Details of power calculations

17/8/2016

This document contains the details of the methods of power calculation from the main text of the proposal, which uses the approach of Hooper et al (2016) [1], together with details of the numerical simulation work to corroborate the power statements provided in the main text.

This document is structured as follows:

1. Transcription of the power calculation text from the main proposal document
2. Summary of Hooper et al (2016) [1] power formulation and its implementation
3. Use of Hooper et al for continuous outcomes in the proposal
4. Use of Hooper et al for binary outcomes in the proposal
5. Extensions required for use of Hooper et al for count outcomes in the proposal
6. Numerical simulation details, a comparison with Hooper et al detectable effects, and a brief sensitivity analysis to choice of ICCs

The overall conclusion from (6) is that numerical simulations using linear and generalised linear mixed models produce detectable effect sizes with 80% power that are at most 0.02 (in effect size units) larger than those calculated by the Hooper et al formulae.

1. Power calculation text from the proposal document:

The choice of 12 settlements per city, each consisting of ~50 dwellings with an average of 5–6 people per dwelling has been made as a balance between statistical power, construction logistics, and cost feasibility. Statistical power was calculated using formulae for cluster RCTs with repeated assessments of a cohort (Hooper R et al. 2016 [1]) (and confirmed by numerical simulation using generalised linear mixed models, as detailed in this document) for the three primary health outcomes assessed in children aged <5: prevalence of bacterial and parasitic gastrointestinal pathogen infection, concentration of intestinal inflammation markers , and number and abundance of antimicrobial resistance markers . For the primary environmental measure (0bjective 2)—the diversity and abundance of markers of drug resistance determined from soil samples—the power calculations assumed the same intra-cluster correlations as for the health outcomes (see below).

For the health outcomes and sub-sampling 30 children per settlement, there is 80% power to detect at 6 months post construction a 35% relative reduction in the prevalence of bacterial and parasitic gastrointestinal infections (assuming a baseline prevalence of 25%) and a 23% relative reduction in the average count of enteric

pathogens per child (assuming a baseline average count of 1.0; with a baseline of 2.0 the relative reduction is 17%), and at 12 months an absolute difference of 0.30 SDs (and 0.32 SDs for environmental assessments) in the average concentration of intestinal inflammation markers and average number (and average abundance) of antimicrobial resistance markers. The above calculations assume intra-cluster correlations as follows: between two persons with the same time period 0.10 [2]; between two persons in different time periods 0.067; and between repeated measurements of the same individual 0.10. The relative reduction of 35% is justified based on systematic reviews, which have found reductions of 30-50% due to water-quality interventions [3 4].

2. Summary of Hooper et al (2016) power formulation and its implementation

Hooper et al [1] describe a formulation for sample size determination for longitudinal cluster randomised trials. The required number of individuals for a cluster randomised trial to detect a certain effect size is determined from the number of individuals required for an individually randomised (non-cluster) trial, multiplied by two design effects, the first is for the loss of information from the clustering of individuals within clusters, and the second is for the potential regain of information from repeated measurements of clusters and individuals over time.

The underlying model in the paper is based on earlier work of Feldman and McKinlay [5] using a linear mixed effects model with random effects for cluster, cluster-time, and individual. We slightly reparameterise it as:

$$Y_{itk} = \alpha + \pi_t + A_{it}\beta + C_i + (CT)_{it} + S_{ik} + \epsilon_{itk}$$

where $i = 1, .., N$ clusters, $t = 1, \dots, T$ time points, and $k = 1, \dots, m$ subjects per cluster at each time point, $C_i \sim N(0, \sigma_C^2), (CT)_{it} \sim N(0, \sigma_{CT}^2), S_{ik} \sim N(0, \sigma_S^2), \epsilon_{itk} \sim N(0, \sigma_\epsilon^2)$., π_t are time fixed effects, and A_{it} are indicators of being in intervention ('treatment') or control condition for cluster i at time t . For control clusters $A_{it} = 0$ for all time points, whereas for intervention arm clusters $A_{it} = 0$ until the intervention is implemented, after which $A_{it} = 1$, and β is the treatment effect parameter of interest.

The random effects define 3 intracluster correlations (ICCs). We redefine the ICCs of Hooper et al into 'conventional' correlations:

$$\rho_{CT} = Corr(\text{different individuals, same time period}) = \frac{\sigma_C^2 + \sigma_{CT}^2}{\sigma_C^2 + \sigma_{CT}^2 + \sigma_S^2 + \sigma_\epsilon^2}$$

$$\rho_C = Corr(\text{different individuals, different time periods}) = \frac{\sigma_C^2}{\sigma_C^2 + \sigma_{CT}^2 + \sigma_S^2 + \sigma_\epsilon^2}$$

$$\rho_S = Corr(\text{same individual, different time periods}) = \frac{\sigma_C^2 + \sigma_S^2}{\sigma_C^2 + \sigma_{CT}^2 + \sigma_S^2 + \sigma_\epsilon^2}$$

For cross-sectional samples of completely different individuals in each cluster at each time point one would take $\sigma_S^2 = 0$ and $\rho_S = \rho_C$. In the WSC trial we will have largely

overlapping samples of individuals at each time point so we allow a non-zero value of σ_S^2.

For any given study design consisting of multiple control and treatment periods and the corresponding treatment indicators $\{A_{it}\}$ together with estimates of the three ICCs, one can calculate the overall design effect/inflation factor relative to an individually randomised trial for a longitudinal cluster randomised trial to detect a specified effect with the required level of power.

In this trial's development we implemented the Hooper et al method as follows:
As described earlier, the final choice of number of settlements was made with reference to power, logistic and budgetary considerations. For this final choice of number of clusters, frequency of sampling and numbers sampled per occasion, the design effects of Hooper et al were used to determine the sample size of an individually randomised trial with the same total amount of information as the total sample size for the longitudinal cluster trial. The detectable intervention effect for the longitudinal cluster trial was then determined from standard power formulae for the equivalent individually randomised trial.

3. Use of Hooper et al for continuous outcomes in the proposal

Effect sizes for continuous outcomes are presented as standardised effect sizes, yielding differences that can be detected in terms of the (cross-sectional) standard deviation of the particular outcome measure. Values of the ICCs were assumed to be: $\rho_{CT} = 0.10, \rho_C = 0.067, \rho_S = 0.10$. The value of $\rho_{CT} = 0.10$ was chosen from Clasen et al [2] as the maximum within-time ICC across a range of continuous and binary outcomes from a cluster RCT of a rural sanitation program in India. The value $\rho_C = 0.067$ is equivalent to an assumption that the variability in outcome across occasions within a cluster, after removing systematic time trends, is one half of the variation between clusters at a single time point. The ratio ρ_c/ρ_{CT} is called a cluster autocorrelation in Hooper et al and takes the value 0.67 here. The value $\rho_S = 0.10$ is a conservative correlation between repeated measurements on the same individual over time. Although we would have chosen a larger value if the design was a pure closed cohort design, we follow the discussion of Feldman and McKinlay [5] and use the value of 0.10 as "a middle ground between independent samples (no overlap) and pure cohort design (complete overlap)".

As an example of the treatment indicators $\{A_{it}\}$ used in the formulae of Hooper et al, we consider the outcome environmental enteric dysfunction markers (measured by concentration of intestinal inflammation markers), which are measured 6-monthly but the effect of the WSC intervention is not expected to be yielded until 12 months post-construction. In the intervention arm there are therefore two baseline (pre-construction) measurements in Year 2 of the trial (i.e. quarters 2 and 4), and 3 measurements ≥ 12 months post-construction (quarter 4 of Year 4 and quarters 2 and

4 of Year 5). Hence in the formulation of Hooper et al, $\{A_{it}\} = \{0,0,1,1,1\}$ for the intervention arm and $\{A_{it}\} = \{0,0,0,0,0\}$ for the control arm.

4. Use of Hooper et al for binary outcomes in the proposal

The formulation in Hooper et al is for linear mixed models with a Gaussian outcome. We describe here our adaptation for binary outcomes, and in a later section describe our validation of this approach by numerical simulation using generalised linear mixed models.

For binary outcomes we use as input the three ICCs, each being correlations on the 'original' or 'raw binary' data scale. We follow the same process as for continuous outcomes using Hooper et al's design effects to calculate the sample size of an individually randomised trial with the same total amount of information as the cluster trial. From this equivalent sample size we then use standard power formulae (e.g. the `power twoproportions` function in Stata v14) to find the detectable difference in prevalence of the outcome, which is then converted to a prevalence ratio.

5. Use of Hooper et al for count outcomes in the proposal

For count outcomes we commence as for continuous and binary outcomes using Hooper et al's design effects to calculate the sample size of an individually randomised trial with the same total amount of information as the cluster trial. However, here the subsequent approach is more complicated because we cannot assume the marginal distribution of the count outcome is Poisson, that is, we cannot assume the variance equals the mean, even for an individually randomised trial. We therefore allow for overdispersion in which the marginal variance is a multiple of the marginal mean, and allow it to differ between treatment and control outcomes.[1] Details are provided in the Appendix at the conclusion of this document.

6. Numerical simulation details and comparison with Hooper et al detectable effects

(a) Simulation details

In order to validate the detectable effect sizes from Hooper et al and its extensions described above, we performed numerical simulations using the linear mixed models of Hooper et al or the corresponding generalised linear mixed models for binary and count outcomes.

The common format of the mixed models were $E[Y_{itk}|\mu_{itk}] = \mu_{itk}$, where

$$h(\mu_{itk}) = \alpha + \gamma I(city_i = 1) + \pi_t + A_{it}\beta + C_i + (CT)_{it} + S_{ik}$$

[1] Note that this complication was not present for the binary case because the marginal variance does not depend on the assumed ICCs, since the marginal variance is equal to prevalence*(1-prevalence) and hence is fixed for a given prevalence.

where h() is the identity function for continuous outcomes and the logarithm function for both binary or count outcomes, and $C_i \sim N(0, \sigma_C^2), (CT)_{it} \sim N(0, \sigma_{CT}^2), S_{ik} \sim N(0, \sigma_S^2)$, as described previously. For continuous outcomes, $Y_{itk}|\mu_{itk} \sim N(\mu_{itk}, \sigma_\epsilon^2)$, for binary outcomes $Y_{itk}|\mu_{itk} \sim Binomial(\mu_{itk})$, and for count outcomes $Y_{itk}|\mu_{itk} \sim Poisson(\mu_{itk})$.

The parameter γ enables a different baseline mean per city (Suva, Makassar), and the time effects π_t were parametrised as (a) year of trial, regarded as continuous, and (b) quarter of year, in order to define wet and dry seasons (wet seasons in quarters 1 and 4).

Because the effect of the WSC intervention may not be evident until 6 or 12 months post-construction, the treatment covariate for the intervention arm was partitioned (as in [7]) into 4 time components: during-construction-period, 3 months post-construction, 6-9 months post-construction and ≥ 12 months post-construction. Overall the fixed effects components of the above model were (with $G_i = 1$ if cluster i is in the intervention arm, and $G_i = 0$ for control arm):

$$\alpha + \gamma I(city = 1) + \delta * year + \sum_{q=2}^{4} \eta_q I(quarter = q) + \beta_1 * G_i * I(construction\ period)$$
$$+\beta_2 * G_i * I(3\ months\ post - construction)$$
$$+\beta_3 * G_i * I(6 - 9\ months\ post - construction)$$
$$+\beta_4 * G_i * I(\geq 12\ months\ post - construction).$$

Input parameters into each data generation model were values of the fixed effects of time and treatment effect lags as above, the three ICCs (on the 'raw' data scale), the baseline mean or prevalence or mean count, the number of settlements/clusters (24), the number of dwellings per settlement (drawn as $N(30, SD = 5)$ with an assumption of one individual per dwelling being sampled), and the frequency of measurement (annual, six-monthly or quarterly).

From the values of the baseline mean, the effect size, and the three ICCs, the variance components were calculated. To obtain these for binary and count data, we approximated a constant mean outcome as $\bar{\theta} = \frac{\theta_0(1+R_{intervention})(1+R_{city})}{4}$. For binary outcomes we then solved $\rho = \frac{\left(e^{\sigma_g^2}-1\right)\bar{\theta}}{1-\bar{\theta}}$ [8], and for count outcomes $\rho = \frac{\bar{\theta}\left(e^{\sigma_g^2}-1\right)}{1+\bar{\theta}\left(e^{\sigma_T^2}-1\right)}$ [6] (see Appendix for details). The intercept parameter α in the GLMM was obtained from $\log(marginal\ baseline\ mean) - \frac{1}{2}\sigma_T^2$. This then fully defined the parameters of the GLMM and the data were then simulated, with 2000 replications per parameter combination.

For binary and count outcomes data were generated from models with:

(i) no city or time effects, but with the lagged intervention effects as follows: a 20% increased adverse outcome during construction, no intervention effect at 3 months post-construction, and either 25% or 100% intervention effect at 6 months depending on the outcome measure, and full intervention effect at 12 months and thereafter.

(ii) As above, plus inclusion of city effects in which the baseline mean outcome in one city was twice that of the other, and 'wet season' fixed effects in which the mean outcome was 50% larger during quarters 1 and 4.

For continuous outcomes models were simulated with the full fixed effect specification, but the detectable effects did not vary with specification of city or time effects.

Analysis of the simulated data was performed by first aggregating the individual-level data to cluster-time means for continuous outcomes, or to cluster-time totals for binary and count outcomes, and then estimating intervention effects at the various time periods (during construction and post-construction) with generalised estimating equations with identity link for continuous outcomes, and logarithmic link for binary and count outcomes, with an offset of the logarithm of the total number of individuals sampled per time period in each cluster. Models with fixed effects for city, year and quarter were used in all analyses regardless of whether or not the true parameters were set to zero in the data generation model (e.g. model generation (i) above). An exchangeable working correlation was used, together with robust sandwich standard errors clustered at the settlement level.

It is well known that the robust sandwich estimator can be anti-conservative when the number of clusters is not large. We observed sub-nominal 95% confidence interval coverage in early simulation testing, at which stage we were using simpler fixed effect models with 6 parameters (intercept, city, 3 quarters, and treatment). We applied a simple standard error correction of $\sqrt{\frac{N}{N-6}}$ which provided approximately correct 95% CI coverage, and because this CI coverage was maintained across all later models, no matter how detailed the fixed effect component specification, we retained this correction factor in all simulations.

(b) Simulation results and comparison with Hooper et al detectable effects

In brief, the table below indicates that detectable effects with 80% power by numerical simulation differed from Hooper et al's formulation by at most 0.02 in any effect size (e.g. detect ES=0.31 rather than ES=0.29).

Each configuration was simulated with 2000 replications. Monte Carlo SE for 80% power is 0.9%; for 95% CI coverage it is 0.5%. Observed power between 78.2% and 81.8% is within simulation error of the value of 80%.

For continuous outcomes, the effect size is the difference between intervention and control in SD units; for binary outcomes it is the Prevalence Ratio, and for count outcomes it is the ratio of mean counts.

The simulations use intra-cluster correlations as follows: between two persons within the same time period $\rho_{CT} = 0.10$ [2]; between two persons in different time periods $\rho_C = 0.067$; and between repeated measurements of the same individual $\rho_s = 0.10$.

Outcome measure: description and scale	**Frequency of measurement; delay until impact**	**Sample size per occasion**	**Baseline control mean**	**Effect size Hooper**	**Simulated power at Hooper effect size (95% CI coverage)**	**Detectable effect size with 80% power from simulation**
Environmental enteric dysfunction markers (Continuous)	6 monthly, Effect at 12 months	30	-	0.24 SDs	75% (95.7%)	0.25 SDs (95.1%)
Antimicrobial resistance markers (Continuous)	6 monthly, Effect at 12 months	15 [Human]	-	0.28 SDs	77% (96.1%)	0.29 SDs (95.6%)
		10 [Envnm't]		0.31 SDs	75% (96.0%)	0.32 SDs (95.7%)
Prevalence of gastrointestinal bacterial and parasitic pathogens (Binary)	6-monthly, Effect at 6 months	30	25%	0.65	77% (95.2%)	0.64 (95.7%)
As above, but with city and temporal trends			17%, 33%*	0.65	79% (94.6%)	0.65 (94.6%)

Outcome measure: description and scale	Frequency of measurement; delay until impact	Sample size per occasion	Baseline control mean	Effect size Hooper	Simulated power at Hooper effect size (95% CI coverage)	Detectable effect size with 80% power from simulation
Number of gastrointestinal bacterial and parasitic pathogens in stools [Human] (Count)	6 monthly, Effect at 6 months	30	1.0	0.78	75% (96.2%)	0.77 (95.4%)
As above, but with city and temporal trends			0.67, 1.33		75% (94.7%)	0.77 (95.6%)
		30	2.0	0.84	77% (96.2%)	0.83 (95.9%)
As above, but with city and temporal trends			1.0, 3.0		75% (95.0%)	0.83 (95.6%)
In environmental samples	6 monthly, Effect at 6 months	20	1.0	0.76	76% (95.1%)	0.75 (96.4%)
		20	2.0	0.83	76% (95.8%)	0.82 (95.5%)
Diarrhoea (Binary)**	Quarterly, Effect at 6 months	100	7.5%	0.72	79% (95.0%)	0.72 (95.0%)
As above, but with city and temporal trends			5%, 10%		82% (95.7%)	0.72 (95.7%)

*Prevalence of 17% in one city and 33% in the other city.

**The 7-day diarrhoea period prevalence will be assessed each quarter using approximately 100 children aged <5 per settlement and assumes a within-cluster-within-time correlation of 0.02 [2].

(c) Sensitivity analysis to larger values of ICCs

Here we present effect sizes detectable with 80% power for larger ICCs. In particular, we use larger values of the within-cluster ICCs ρ_{CT} and ρ_C, and a between-subject variance $\sigma_S^2 = 0$ corresponding to non-overlapping repeated cross-sectional samples (noting that $\sigma_S^2 = 0$ leads to $\rho_S = \rho_C$).

The ICCs in the 'base' configuration in the table below are $\rho_{CT} = 0.10, \rho_C = 0.067, \rho_S = 0.10$. We consider two alternative configurations of ICCs:

(1) $\rho_{CT} = 0.134, \rho_C = 0.067, \rho_S = 0.067$ and (2) $\rho_{CT} = 0.20, \rho_C = 0.134, \rho_S = 0.134$

These correspond to (1) the variation between times within clusters being equal to the variation between clusters [as opposed to 50%], and (2) the variation between clusters and between times within clusters is double that previously assumed.

The table overleaf presents detectable effect sizes with 80% power using the formulation of Hooper et al. Although it is evident as expected that detectable effect sizes increase (relative to the null) for configurations (1) and (2) , effect sizes with these far more conservative ICCs for continuous outcomes are in the order of 0.30 SDs, for count outcomes 20%-30% reductions in means, and a prevalence reduction of 45%.

Outcome measure: description and scale	Baseline control mean	Effect size with base configuration	Effect size configuration (1)	Effect size configuration (2)
Environmental enteric dysfunction markers **6 monthly**	-	0.24 SDs	0.28 SDs	0.30 SDs
Antimicrobial resistance markers **6 monthly, 50% of samples**	-	0.28 SDs	0.31 SDs	0.33 SDs
Prevalence of gastrointestinal bacterial and parasitic pathogens **6 monthly**	25%	0.65	0.59	0.56
Number of gastrointestinal bacterial and parasitic pathogens in stools **6 monthly**	1.0	0.78	0.74	0.71
	2.0	0.84	0.81	0.79

REFERENCES:

1. Hooper R, Teerenstra S, de Hoop E, et al. Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. Statistics in medicine 2016:in press. Published online June 28. DOI 10.1002/sim.7028
2. Clasen T, Boisson S, Routray P, et al. Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. The Lancet Global health 2014;**2**(11):e645-53
3. Clasen TF, Alexander KT, Sinclair D, et al. Interventions to improve water quality for preventing diarrhoea. The Cochrane database of systematic reviews 2015(10):CD004794
4. Ziegelbauer K, Speich B, Mausezahl D, et al. Effect of sanitation on soil-transmitted helminth infection: systematic review and meta-analysis. PLoS medicine 2012;**9**(1):e1001162
5. Feldman HA, McKinlay SM. Cohort versus cross-sectional design in large field trials: Precision, sample size, and a unifying model. Statistics in Medicine 1994;**13**(1):61-78 doi: 10.1002/sim.4780130108[published Online First: Epub Date]|.
6. Vangeneugden T, Molenberghs G, Verbeke G, et al. Marginal correlation from an extended random-effects model for repeated and overdispersed counts. Journal of Applied Statistics 2011;**38**(2):215-32 doi: 10.1080/02664760903406405[published Online First: Epub Date]|.
7. Hughes JP, Granston TS, Heagerty PJ. Current issues in the design and analysis of stepped wedge trials. Contemporary Clinical Trials 2015;**45**:55-60
8. Yelland LN, Salter AB, Ryan P, et al. Adjusted intraclass correlation coefficients for binary data: methods and estimates from a cluster-randomized trial in primary care. Clinical Trials 2011;**8**(1):48-58

APPENDIX: Use of Hooper et al for count outcomes in the proposal

For count outcomes we commence as for continuous and binary outcomes using Hooper et al's design effects to calculate the sample size of an individually randomised trial with the same total amount of information as the cluster trial. However, here the subsequent approach is more complicated because we cannot assume the marginal distribution of the count outcome is Poisson, that is, we cannot assume the variance equals the mean, even for an individually randomised trial. We therefore allow for overdispersion in which the marginal variance is a multiple of the marginal mean, and allow it to differ between treatment and control outcomes.

In order to present a unified description of the three ICCs and the resulting overdispersion, we assume a generalised linear mixed model for the count outcomes, analogous to that for continuous outcomes of Hooper et al.

Generalised linear mixed model (GLMM) for count outcomes

We adopt the GLMM for count outcomes in which the logarithm of the mean outcome, $\log(\mu_{itk})$, is a linear function of the covariates and random effects:

$Y_{itk}|\mu_{itk} \sim Poisson(\mu_{itk})$, where

$$\log(\mu_{itk}) = \alpha + A_{it}\beta + C_i + (CT)_{it} + S_{ik}$$

and where the random effects are normally distributed as previously, and where for simplicity we have assumed no time effects π_t.

This formulation defines the μ_{itk} as being log-normally distributed, and we follow a similar approach to Vangeneugden et al. [6]. Using the properties of the log-normal distribution,

$$E[Y_{itk}] = E[\mu_{itk}] = e^{\alpha + A_{it}\beta + \frac{1}{2}\sigma_T^2}$$

where $\sigma_T^2 = \sigma_C^2 + \sigma_{CT}^2 + \sigma_S^2$. The treatment effect parameter of interest is the ratio of mean counts e^{β}, which has equivalent interpretation in marginal and cluster-specific models.

For convenience we will write the marginal means of Y under control and treatment as $\theta_0 = e^{\alpha + \frac{1}{2}\sigma_T^2}$ and $\theta_1 = e^{\alpha + \beta + \frac{1}{2}\sigma_T^2}$. Then the marginal variance of Y can be written as

$$Var(Y_{itk}) = \theta_j + \theta_j^2\left(e^{\sigma_T^2} - 1\right) = \theta_j\left[1 + \theta_j\left(e^{\sigma_T^2} - 1\right)\right], \quad j = 0,1$$

and hence the overdispersion factors are $\Gamma_j = 1 + \theta_j\left(e^{\sigma_T^2} - 1\right), j = 0,1.$

Using the standard correlation expression $Corr(Z_1, Z_2) = Cov(Z_1, Z_2)/\sqrt{Var(Z_1)\ ar(Z_2)}$ we can then determine the three within cluster correlations between observations on the 'raw count' data scale in a common format (for observations in the same treatment arm) as

$$\rho = \frac{\theta_j\left(e^{\sigma_g^2} - 1\right)}{1 + \theta_j\left(e^{\sigma_T^2} - 1\right)}, j = 0{,}1\ .$$

For ρ_{CT} we use $\sigma_g^2 = \sigma_C^2 + \sigma_{CT}^2$; for ρ_C we use $\sigma_g^2 = \sigma_C^2$; and for ρ_S we use $\sigma_g^2 = \sigma_C^2 + \sigma_S^2$.

For two observations across different treatments arms it is, for ρ_C and ρ_S

$$\rho = \frac{\theta_0\theta_1\left(e^{\sigma_g^2} - 1\right)}{\sqrt{\theta_0 + \theta_0^2\left(e^{\sigma_T^2} - 1\right)}\sqrt{\theta_1 + \theta_1^2\left(e^{\sigma_T^2} - 1\right)}}$$

To simplify matters we make an approximation to the ICCs by replacing θ_0 and θ_1 by the average outcome across control and treated observations, $\bar{\theta} = \frac{\theta_0 + \theta_1}{2}$, yielding the generic form for all ICCs, regardless of treatment condition, of

$$\rho = \frac{\bar{\theta}\left(e^{\sigma_g^2} - 1\right)}{1 + \bar{\theta}\left(e^{\sigma_T^2} - 1\right)}\ ,$$

with σ_g^2 determined as above for ρ_{CT}, ρ_C and ρ_S.

Progressing now to an expression for power, we write $\theta_1 = R\theta_0$, where $R = e^{\beta} = \frac{\theta_1}{\theta_0}$ is the ratio of mean counts of treated and control outcomes. The above expression for ρ is then a function of R via $\bar{\theta} = \frac{\theta_0(1+R)}{2}$. The power to detect a ratio of means R for an individually randomised trial in which there are n individuals per arm and in which the marginal mean and variance are θ_j and $\Gamma_j\theta_j, j = 0{,}1$, respectively, is:

$$Power = \Phi\left(-Z_{\frac{\alpha}{2}} + \frac{\sqrt{n}\ |\log(R)|}{\sqrt{\frac{1}{\theta_0}\left(\frac{\Gamma_1}{R} + \Gamma_0\right)}}\right).$$

In order to calculate the detectable value of R at 80% power for a given baseline mean count θ_0, an effective sample size n and the 3 ICCs $\rho_{CT}, \rho_C, \rho_S$ we need to solve a system of (nonlinear) equations. We performed this using the `optimize` function in the programming language Mata within Stata.