CLINICAL TRIALS WORKSHOP ARTICLE Clinical Trials 2005; 2: 108–118

Prior distributions for the intracluster correlation coefficient, based on multiple previous estimates, and their application in cluster randomized trials

Rebecca M Turner, Simon G Thompson and David J Spiegelhalter

Numerous estimates for the intracluster correlation coefficient (ICC) are available in research databases and publications. When planning a cluster randomized trial, an anticipated value for the ICC is required; currently, researchers base their choice informally on the magnitude of previous ICC estimates. In this paper, we make use of the wealth of ICC information by formally constructing informative prior distributions, while acknowledging the varying relevance and precision of the estimates available. Typically, for a planned trial in a given clinical setting, multiple relevant ICC estimates are available from each of several completed studies. Our preferred model allows for the imprecision in each ICC estimate around its underlying true value and, separately, allows for the similarity of ICC values from the same study. The relevance of each previous estimate to the planned clinical setting is considered, and estimates corresponding to less relevant outcomes or population types are given less influence. We find that such downweighting can increase the precision of the anticipated ICC. In trial design, the prior distribution constructed allows uncertainty about the ICC to be acknowledged, and we describe how to choose a design that provides adequate power across the range of likely ICC values. Prior information on the ICC can also be incorporated in analysis of the trial data, when taking a Bayesian approach. The methods proposed enable available ICC information to be summarised appropriately by an informative prior distribution, which is of direct practical use in cluster randomized trials. *Clinical Trials* 2005; **2**: 108–118. www.SCTjournal.com

Introduction

In cluster randomized trials, groups of individuals are randomized together to the trial arms [1]. The clusters used as randomization units in health care research are, for example, defined by health care centres, geographical districts, or health professional patient lists, and consequently patients within clusters are similar. The correlation of patient outcome measurements within clusters must be taken into account in both design and analysis of cluster trials. In design, the conventional approach is to multiply the sample size needed under individual randomization by a function of the intracluster correlation coefficient (ICC), to ensure that the cluster randomized design will provide adequate power [2].

When designing a planned cluster trial, it is common to calculate sample size using a single anticipated ICC value based on relevant ICC estimates available from previous studies. However, use of a single ICC value could well be dangerous because the required sample size is very sensitive to small differences in the ICC: if the value picked is too low, the trial will be underpowered, while a value intended to be conservative potentially leads to wasted resources. A Bayesian approach is advantageous here. In design, the uncertainty about the ICC can be acknowledged through specifying a prior distribution, and researchers can choose a design that provides adequate power across the range of likely ICC values [3]. In analysis, a Bayesian approach allows appropriately for the imprecision in model parameters and

MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK
Author for correspondence: Rebecca Turner, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge, CB2 2SR, UK. E-mail: rebecca.turner@mrc-bsu.cam.ac.uk

 10.1191/1740774505cn072oa

enables incorporation of prior information on the ICC [4,5].

A substantial amount of published ICC information is available, since authors have been encouraged to publish ICC estimates routinely in trial reports [1,6] and several research groups are compiling databases of ICC estimates (for example [7,8]). For a given clinical setting, researchers have access to multiple ICC estimates of varying relevance. Typically, observed ICC values representing the same outcome measured within the same cluster type on the same patient population are rare, but numerous estimates from similar outcomes, cluster types and populations are available. The estimates vary also in precision; while confidence intervals for ICCs are rarely presented, it is usual for basic information on the size of the source study to be published alongside the estimate.

In this paper we construct informative prior distributions for the ICC, in order to make use of the wealth of available ICC information in design and analysis of cluster randomized trials. Previous work has discussed constructing such priors, either on the basis of informal subjective beliefs [4], on the basis of a single previous ICC estimate [3], or by pooling a set of ICC estimates while ignoring differences in relevance [5]. Here, we propose and exemplify methods for formal construction of prior distributions from multiple ICC estimates, while allowing for differences in both relevance and precision.

Example

The cluster randomized design is frequently used in primary care and community settings, in particular when evaluating interventions which aim to modify the behaviour of health professionals or change patient management policies. As a typical setting, we consider secondary prevention of coronary heart disease (CHD) in primary care. Interventions designed to improve secondary prevention of CHD are often delivered to general (family) practices, while outcomes to evaluate their performance are collected from patients. Such trials commonly measure total cholesterol, and we will consider constructing a prior distribution for the ICC of total cholesterol measurements within general practices, in patients with existing CHD. Cholesterol, general practice and patients with existing CHD will be referred to as the target outcome, cluster type, and population type, respectively.

When designing a cluster trial, researchers carry out an informal search for relevant ICC estimates to inform their sample size calculation. Sources might include full data sets from previous studies to which they have access, published reports of studies, and ICC databases. We have followed this approach to identify a set of ICC estimates relevant to our setting, making use of three data sets available to us [9–11] and three ICC databases [7,8,12]. Each of cluster type, outcome and population type are likely to influence the magnitude of the ICC [12]. From our sources, we have extracted estimates representing clustering within general practices of total cholesterol or other outcomes related to CHD; these are estimates for the target cluster type and target or similar outcomes, obtained from any population. A total of 18 ICC estimates were identified (Table 1) which varied in magnitude from 0 to 0.06. The range of sample sizes (around 200 to 32 000 patients, 20 to 100 practices) led to very considerable variation in their precisions (Figure 1).

For convenience, ICC estimates for the target outcome within the target cluster type will be referred to as "type I" estimates, and those for similar outcomes within the target cluster type as "type II" estimates. Clearly, the type I estimates are most relevant to the target setting, and it is desirable to have as many of these available as possible. In our example, two of the five type I ICC estimates collected (Table 1) were unpublished and obtained through personal communication with the study authors [13,14]. ICC estimates were classified as type II estimates if they corresponded to outcomes representing other CHD risk factors or morbidity.

Methods

Initial model

To construct prior distributions for the ICC ρ in a target setting on the basis of published ICC information, the first step is to specify a model for the available ICC estimates. To allow for differences in precision of the estimates, the approach taken is to assume each estimate $\hat{\rho}_l (l = 1, \ldots, s)$ to be distributed around an underlying ICC value ρ_l, and to base inference on the set of underlying values ρ_l rather than on the estimates $\hat{\rho}_l$. To provide a form for the distribution of $\hat{\rho}_l$ given ρ_l, we follow Swiger's approach [15] which enables calculation of the variance of $\hat{\rho}_l$ using only the numbers N_l of individuals and k_l of clusters from whom the estimate was obtained:

$$\hat{\rho}_l \sim N(\rho_l, Var(\hat{\rho}_l)), Var(\hat{\rho}_l) = V(\rho_l, N_l, k_l)$$

$$= \frac{2(N_l - 1)(1 - \rho_l)^2\{1 + ((N_l/k_l) - 1)\rho_l\}^2}{(N_l/k_l)^2(N_l - k_l)(k_l - 1)}$$

The variance formula is referred to hereafter by the abbreviated expression $V(\rho, N, k)$. ICC estimates reported as equal to 0 are treated as censored observations and handled differently (see Appendix),

Table 1 Available ICC estimates relevant to the target setting, details of source data sets and subjective weights chosen

Outcome	ICC estimate	Source	Characteristics of patients recruited	Number of patients	Number of practices	Classification	w_{ml} (outcome weight)	w_m (study weight)
Total cholesterol	$\hat{\rho} = 0.025$	SHIP[10]	MI[a]/angina	473	67	Type I	1	1
Systolic blood pressure	$\hat{\rho}$ set to 0	SHIP[10]	MI[a]/angina	470	66	Type II	0.7	1
Distance walked in two minutes	$\hat{\rho} = 0.005$	SHIP[10]	MI[a]/angina	412	62	Type II	0.5	1
BMI[a]	$\hat{\rho} = 0.021$	SHIP[10]	MI[a]/angina	461	67	Type II	0.7	1
Total cholesterol	$\hat{\rho} = 0.016$	ASSIST[11]	Established CHD	1287	21	Type I	1	1
Systolic blood pressure	$\hat{\rho} = 0.019$	ASSIST[11]	Established CHD	1341	21	Type II	0.7	1
HDL cholesterol	$\hat{\rho} = 0.002$	ASSIST[11]	Established CHD	1265	21	Type II	0.7	1
Total cholesterol	$\hat{\rho} = 0.00003$	Diabetes Care[13]	Diabetes	231	40	Type I	1	0.5
Total cholesterol	$\hat{\rho} = 0.017$	FHS[9]	Ages 40–59	10 196	28	Type I	1	0.2
Total cholesterol	$\hat{\rho} = 0.06$	DiSC[14]	Diabetes	171	30	Type I	1	0.5
Smoked in last month	$\hat{\rho} = 0.0101$	Premaratne[31]	Ages 15–50	12 474	42	Type II	0.7	0.2
Smokes cigarettes	$\hat{\rho} = 0.00636$	MRC Elderly Trial[8]	$\geq$75, excl. nursing/ill	32 714	106	Type II	0.7	0.5
Definite angina	$\hat{\rho} = 0.01036$	MRC Elderly Trial[8]	$\geq$75, excl. nursing/ill	14 910	53	Type II	0.2	0.5
Pulse rate	$\hat{\rho} = 0.03207$	MRC Elderly Trial[8]	$\geq$75, excl. nursing/ill	14 868	53	Type II	0.5	0.5
BMI[a]	$\hat{\rho} = 0.02207$	MRC Elderly Trial[8]	$\geq$75, excl. nursing/ill	13 872	53	Type II	0.7	0.5
Controlled hypertension	$\hat{\rho} = 0.064$	Fahey and Peters[32]	Hypertension	882	49	Type II	0.2	0.7
SAQ[a] physical limitation	$\hat{\rho} = 0.05$	COGENT[33]	Angina	1752	60	Type II	0.2	1
SAQ[a] anginal stability	$\hat{\rho} = 0.02$	COGENT[33]	Angina	1752	60	Type II	0.2	1

[a]BMI = body mass index, SAQ = Seattle Angina Questionnaire, MI = myocardial infarction.

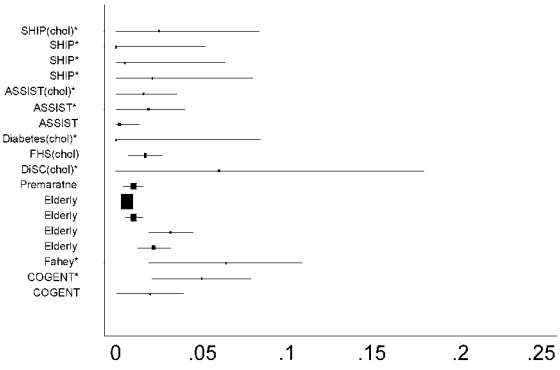


Figure 1 Available ICC estimates ordered as in Table 1, with 95% confidence intervals constructed using Swiger's approach [15]. Box sizes are inversely proportional to variances, except for outcomes marked by *, where sizes are set to a minimum level for visibility.

since these commonly represent negative ICC estimates which have been truncated upwards to 0.

When declaring how the underlying ICC values ρ_l ($l = 1, \ldots, s$) are related, one option would be to assume all ρ_l equal to a common value ρ_c. However, this would be unrealistic even if the ρ_l corresponded to identical cluster, outcome and population types. A more appropriate assumption is "exchangeability", meaning that the values ρ_l are similar rather than identical, and it is judged that there is no reason to believe any specific ICC is systematically different [16]. Under the assumption of exchangeability, the ρ_l may be regarded as independently drawn from a common random distribution. In order to constrain the ρ_l to the permissible range of values [0, 1], it is convenient to assume a Normal distribution on the logit-transformed scale in a random effects model.

$$\begin{aligned} \hat{\rho}_l &\sim N(\rho_l, V(\rho_l, N_l, k_l)) \\ \text{logit}(\rho_l) &\sim N(\mu, \sigma^2). \end{aligned} \tag{1}$$

The defining parameters μ and σ^2 are regarded as unknowns to be estimated from the available data within a Bayesian model. When using model (1) to construct a prior distribution for the ICC ρ in a target setting, ρ is assumed drawn from the same distribution as the ρ_l.

Model assuming exchangeability between studies, equality within studies

Some source studies are likely to provide multiple relevant ICC estimates, as in Table 1, representing clustering of different outcome measurements. The exchangeability assumption underpinning model (1) is then inappropriate and the model must be

adapted to acknowledge the likely similarity of ICCs within studies. One approach would be to assume exchangeability between studies, while assuming a common ICC value for outcomes within the same study. Writing $\hat{\rho}_{ml}$ to represent the estimates for outcomes $l = 1, \ldots, s_m$ within studies $m = 1, \ldots, r$, model (2) below assumes each estimate $\hat{\rho}_{ml}$ within study m to be distributed around a common underlying value ρ_m, again following Swiger's distributional approach. The ρ_m are assumed exchangeable between studies and Normally distributed on the logit-transformed scale.

$$\begin{aligned} \hat{\rho}_{ml} &\sim N(\rho_m, V(\rho_m, N_{ml}, k_{ml})) \\ \text{logit}(\rho_m) &\sim N(\mu, \sigma^2). \end{aligned} \tag{2}$$

Model assuming exchangeability between and within studies

The assumption of a common ICC value across outcomes within each study is rather unrealistic. Model (3) below instead assumes exchangeability between and within studies: the estimates $\hat{\rho}_{ml}$ are distributed around an underlying value ρ_{ml}, where the ρ_{ml} are assumed exchangeable within studies and Normally distributed on the logit-transformed scale around a study-specific mean μ_m. Under exchangeability between studies, the μ_m are assumed independently drawn from a common Normal distribution.

$$\begin{aligned} \hat{\rho}_{ml} &\sim N(\rho_{ml}, V(\rho_{ml}, N_{ml}, k_{ml})) \\ \text{logit}(\rho_{ml}) &\sim N(\mu_m, \sigma_w^2) \\ \mu_m &\sim \text{N}(\mu, \sigma_b^2). \end{aligned} \tag{3}$$

When using model (3) to construct a prior distribution for the ICC ρ in a target setting, a new study mean μ_m is sampled and ρ is assumed centred around μ_m (on the logit scale) with variance σ_w^2.

Multiplicative weighting for relevance

Available ICC estimates differ in their relevance to the target setting. It is usually unrealistic to base the prior distribution for ρ only on estimates for the target outcome within the target cluster type on the target patient population type. Such estimates would be rare, or nonexistent when the outcome is constructed from a newly developed questionnaire. A prior distribution constructed from a few extremely relevant ICC estimates, ignoring the moderately relevant ICC estimates, will unnecessarily waste information. We will consider the relevance to the target setting of the outcome and population type, separately. The approach taken is to modify model (3) by increasing the variances assumed for less relevant ICC estimates, in order that these have less influence on the prior constructed.

We reconsider the interpretation of certain parameters in model (3). The parameter μ_m represents the study-specific mean of underlying ICC values ρ_{ml} for outcomes regarded as exchangeable with (and so fully relevant to) the target outcome. We assume that less relevant outcomes are distributed about the same mean μ_m with an increased variance. The increase in variance from σ_w^2 is specific to each outcome to reflect differences in relevance, and initially the variance is given a multiplicative form σ_w^2/w_{ml}, where the w_{ml} are fixed weights in the range $(0, 1]$ and $w_{ml} = 1$ for a fully relevant outcome. The parameter μ represents the mean of the μ_m across studies whose populations are regarded as exchangeable with (and so fully relevant to) the target population, and we assume that studies with less relevant populations are distributed about μ with an increased variance. The variance of each μ_m is given the form σ_b^2/w_m, where the w_m are again fixed weights in the range $(0, 1]$ and $w_m = 1$ for a fully relevant study population.

$$\begin{aligned} \hat{\rho}_{ml} &\sim N(\rho_{ml},\ V(\rho_{ml},\ N_{ml},\ k_{ml})) \\ \text{logit}(\rho_{ml}) &\sim N\left(\mu_m,\ \frac{\sigma_w^2}{w_{ml}}\right) \\ \mu_m &\sim N\left(\mu,\ \frac{\sigma_b^2}{w_m}\right). \end{aligned} \tag{4}$$

The weights w_{ml} and w_m are likely to be subjective in practice. Choice of weights is discussed later in the context of the example.

Additive weighting for relevance, incorporating uncertainty

It is undesirable in model (4) that the differential weighting for relevance disappears from the model as the variances σ_w^2 and σ_b^2 tend to zero, with the result that studies of limited relevance may still exert undue weight simply through being large. To avoid this, we assume that less relevant studies and endpoints introduce the possibility of bias [17], which leads us to apply relevance weights additively rather than multiplicatively. For study m we assume that $\mu_m \sim N(\mu + \delta_m, \sigma_b^2)$ where δ_m is an unknown bias parameter with mean 0 (since systematic bias is not anticipated). Assuming a $N(0, \tau_m^2)$ distribution for δ_m, the variance for μ_m in model (3) is increased through replacing σ_b^2 by $\sigma_b^2 + \tau_m^2$. Similarly, less relevant endpoints are downweighted through assuming bias with variance τ_{ml}^2, so that

$$\begin{aligned} \hat{\rho}_{ml} &\sim N(\rho_{ml}, V(\rho_{ml}, N_{ml}, k_{ml})) \\ \text{logit}(\rho_{ml}) &\sim N(\mu_m, \sigma_w^2 + \tau_{ml}^2) \\ \mu_m &\sim \text{N}(\mu, \sigma_b^2 + \tau_m^2). \end{aligned} \tag{5}$$

Options for supplying values for τ^2_{ml} and τ^2_m include: (a) selecting fixed values directly, (b) selecting fixed values based on the weights used in model (4), (c) incorporating uncertainty. The interpretation of τ^2_{ml} and τ^2_m does not facilitate direct choice of numerical values, and we therefore consider starting with option (b) and then extending to allow for uncertainty.

First, if we were to set $\tau^2_{ml} = \sigma^2_w(1 - w_{ml})/w_{ml}$ and $\tau^2_m = \sigma^2_b(1 - w_m)/w_m$, then we would exactly reproduce model (4): the relevance weights $w_{ml} = \sigma^2_w/(\sigma^2_w + \tau^2_{ml})$ and $w_m = \sigma^2_b/(\sigma^2_b + \tau^2_m)$ thus have the attractive interpretation as the proportion of the total variance that is *not* due to bias. We propose extending this by incorporating uncertainty in the extent to which less relevant studies and outcomes are downweighted relative to fully relevant studies and outcomes. We introduce unknown parameters λ_w and λ_b and set $\tau^2_{ml} = \lambda_w\sigma^2_w(1 - w_{ml})/w_{ml}$ and $\tau^2_m = \lambda_b\sigma^2_b(1 - w_m)/w_m$. The random-effects part of model (5) then can be expressed as

$$\text{logit}(\rho_{ml}) \sim N\left(\mu_m, \frac{\sigma^2_w}{w_{ml}}(w_{ml} + \lambda_w(1 - w_{ml}))\right)$$

$$\mu_m \sim N\left(\mu, \frac{\sigma^2_b}{w_m}(w_m + \lambda_b(1 - w_m))\right).$$

As mentioned above, assuming $\lambda_w = \lambda_b = 1$ gives exactly the multiplicative model (4). When $\lambda_w < 1$, model (5) downweights less relevant outcomes to a smaller degree than model (4), and the differential weighting disappears as λ_w tends to zero. When $\lambda_w > 1$, model (5) downweights less relevant outcomes to a greater degree, and the weighting becomes increasingly unequal as λ_w becomes large. Similarly, the extent of weighting of studies in model (5) varies as $\lambda_b < 1$ and $\lambda_b > 1$. Model (5) allows the data to influence the degree to which subjectively determined relevance is taken into account; if studies given low weight are not found to be outlying, a smaller estimate is obtained for λ_b and less downweighting is applied.

Application to example

Implementation

The methods in the previous section are used to construct prior distributions for the ICC ρ which represents clustering of total cholesterol measurements within general practices in patients with existing CHD. Initially, we construct prior distributions without weighting for relevance of outcomes and patient populations. First, we use the five type I estimates only; since these were obtained from five distinct studies, we assume exchangeability and fit model (1). The set of 18 estimates of types I and II were obtained from nine studies (Table 1, Figure 1), and we fit models (2) and (3).

To incorporate differential weighting according to relevance of outcomes and study populations, we use models (4) and (5) and begin by choosing weights w_{ml} and w_m (Table 1). Correlation within general practices is caused partly by geographical effects and partly by patient management strategies. The effects of practice on a commonly measured cardiovascular endpoint such as blood pressure may be expected to be similar to the effects on total cholesterol, so we expect the ICC values for these outcomes to be similar. To acknowledge that the ICC value for a blood pressure estimate is not however considered fully relevant to that for cholesterol, we choose a weight of $w_{ml} = 0.7$ to increase the assumed variance about the study mean. Amongst the other outcomes, indicators of pre-existing disease such as angina are allocated the lowest weights, $w_{ml} = 0.2$, since the effects of practice on these may be very different from the effects of practice on levels of cholesterol. Weights w_m to represent the relevance of the study population are chosen in a similar manner; here, the lowest weights are assigned to studies which recruited from the general population, since the effects of practice on patients who attend only rarely are likely to be very different from the effects on patients with CHD.

To fit models (1)–(5) within the Bayesian framework, vague prior distributions are declared for the unknown parameters. We declare $N(0, 10^4)$ distributions for the mean μ in each model, and Uniform[0, 5] distributions for the standard deviation σ in models (1) and (2), and for the between-study and within-study standard deviations σ_b and σ_w in models (3)–(5). In model (5) we declare priors for λ_w and λ_b, which represent the excess variability in the weights under model (5) compared to model (4). We declare N(0,1) prior distributions for $\log(\lambda_w)$ and $\log(\lambda_b)$; the implied priors for the ratios cover a fairly wide range of values around 1 (95% range 0.14–7.1). All models are fitted using Markov chain Monte Carlo (MCMC) methods within the WinBUGS software [18]. Using single chains, we discard an initial 5000 iterations as "burn-in" to ensure convergence [19], and base our results on the subsequent 100 000 iterations.

Application to prior distributions for the ICC

The prior distributions constructed for the ICC ρ are summarised in Table 2 as posterior medians with 95% intervals, and illustrated in Figure 2. Table 2 also presents the model standard deviations (σ, σ_w, σ_b) and Deviance Information Criterion (DIC), a

Table 2 Summaries of prior distributions constructed for the ICC ρ representing clustering of total cholesterol measurements within general practices, in patients with existing CHD

Set of ICC estimates and model used	Target ICC ρ	σ	σ_w	σ_b	p_D	*DIC*
Type I, model (1)	0.014 (<0.0001, 0.21)	0.69 (0.13, 4.23)	–	–	–	–[a]
Types I and II, model (2)	0.019 (0.003, 0.10)	0.66 (0.18, 1.78)	–	–	6.2	−68.3
Types I and II, model (3)	0.016 (0.002, 0.12)	–	0.69 (0.37, 1.35)	0.41 (0.02, 1.70)	11.0	−96.4
Types I and II, model (4)	0.015 (0.003, 0.056)	–	0.48 (0.27, 0.93)	0.18 (0.008, 0.98)	10.4	−97.5
Types I and II, model (5)	0.015 (0.003, 0.060)	–	0.48 (0.22, 0.99)	0.19 (0.006, 1.02)	10.5	−97.3

[a]DIC value not comparable with values from the other models and therefore not presented.

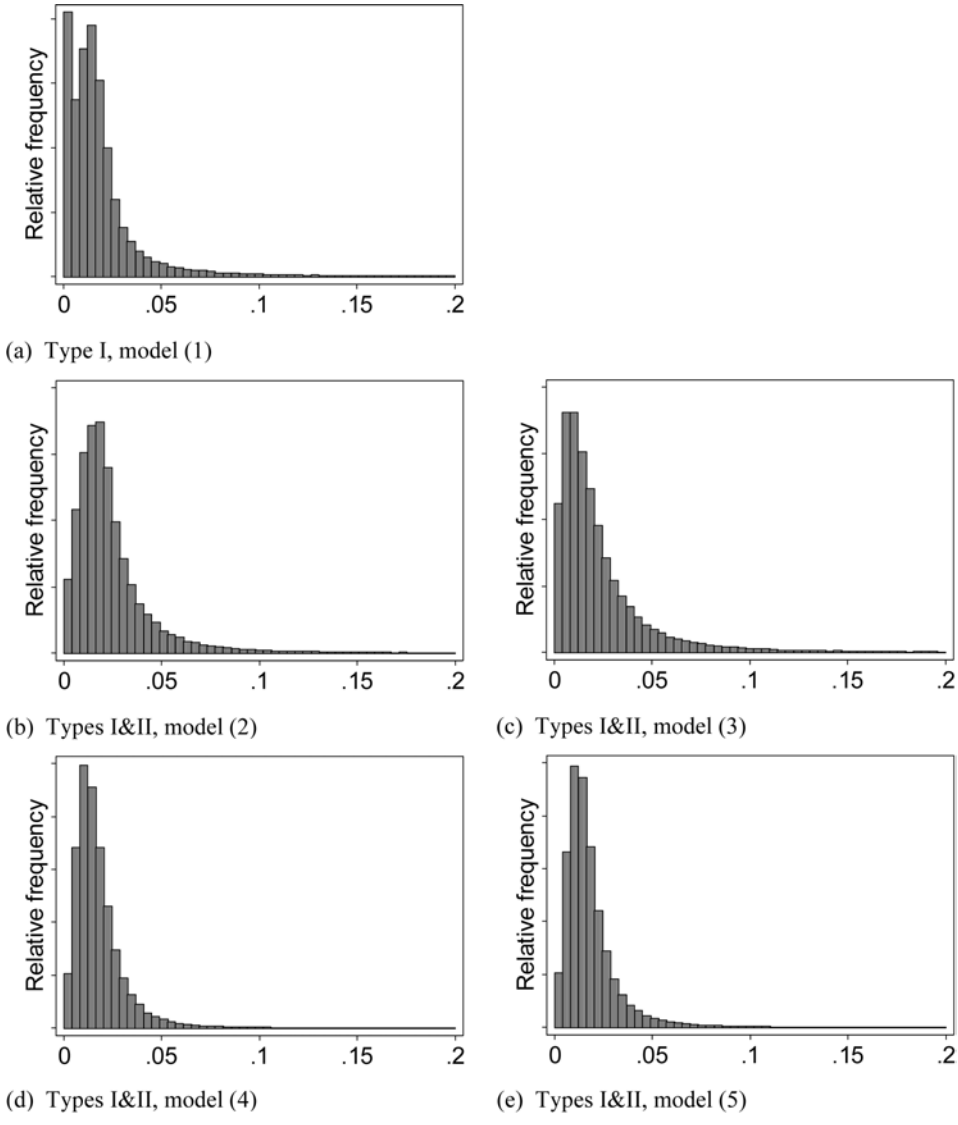


Figure 2 Prior distributions constructed for the ICC ρ by fitting several different models to the available ICC estimates.

measure of model adequacy [20], to aid comparison between the models.

The prior distribution for ρ is considerably narrower when based on ICC estimates of types I and II rather than type I alone, due to increased precision in estimating the variation between underlying ICC values (Table 2). The results demonstrate the benefit of making use of additional

ICC information, rather than relying on the few most relevant ICC estimates. To compare the adequacy of models (2) and (3), we examine the number p_D of effective parameters, and the corresponding DIC values [20]. The fit is considerably worse under the simpler model (2), showing clearly that the assumption of equality within studies is not supported by the data.

Weighting for the varying relevance of outcomes and populations to the target setting using models (4) or (5) results in narrower prior distributions for ρ (Table 2, Figure 2). This is caused by the differences in estimates for the within-study and between-study standard deviations σ_b and σ_w, which have smaller central estimates and narrower intervals when weights are applied. The results show that the ICC estimates which were considered less relevant lie further out in the distribution fitted under model (3), so that downweighting these reduces the estimated magnitude of variation for ICC values which are fully relevant to the target ICC. We note that application of weights need not always result in narrower prior distributions; weighting reduces the amount of information extracted from the less relevant estimates, and may therefore cause the distribution for ρ to widen in other examples.

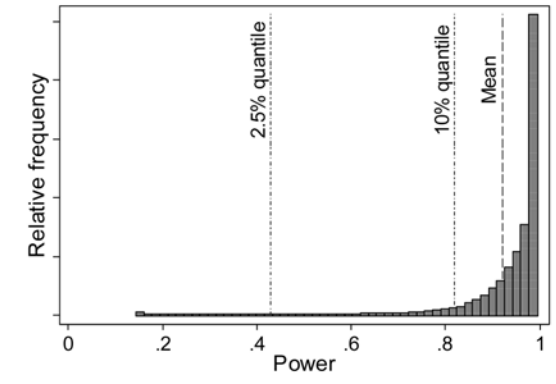
The differences between results from models (4) and (5) depend on estimation of λ_w and λ_b in model (5), which represent the extra variability in the weights under model (5) compared to model (4). The central estimates (95% intervals) for λ_w and λ_b are, respectively, 0.87 (0.13, 5.09) and 0.76 (0.12, 4.88). The prior 95% intervals for λ_w and λ_b are (0.14, 7.1) and we note that very little information is provided by these example data. However, the distributions favour values below 1, meaning that on average model (5) downweights less relevant outcomes and studies to a smaller degree than model (4). The effective number of parameters is approximately 11 in each of models (3)–(5); the DIC values are almost identical and no model shows reduced adequacy in this example (Table 2). In practice, we would recommend model (5), partly because an additive form of weighting is preferred and partly in order to avoid relying entirely on the subjective weights chosen.

Application in trial design

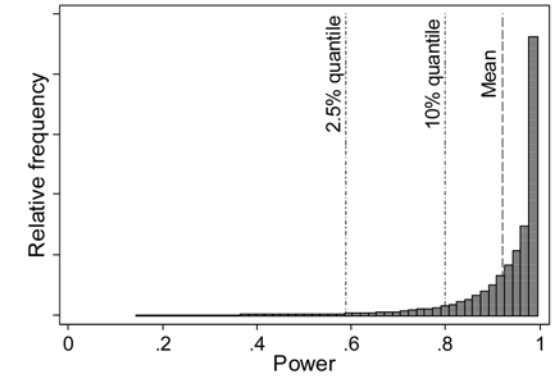
Turner *et al.* [3] described how to allow for the imprecision in one ICC estimate within a Bayesian approach to design, and as an extension discussed use of multiple equally relevant ICC estimates. The methods presented in this paper enable cluster trial design to be based on a set of available ICC estimates of varying relevance. A prior distribution which expresses our uncertainty about the target ICC ρ may be used to construct a corresponding distribution for the power of each design under consideration (see Appendix) [3]. For any given design including a particular number of clusters and patients, the mean of the power distribution provides a point estimate for the power of the trial, allowing for ICC uncertainty. The lower quantiles of the distribution describe the probabilities of obtaining unacceptably low levels of power [16]. It is desirable to protect against the eventuality of the trial having inadequate power due to a higher than expected ICC, but it seems insufficient to simply provide a series of "what-if" scenarios conditioning on various ICC values: a measure of their respective plausibility appears necessary.

The impact of the uncertainty expressed in the prior distributions constructed for our target setting is illustrated through application to a possible trial design. We suppose that a difference of 0.4 standard deviations in total cholesterol is to be detected at a (two-sided) 5% significance level, and consider a design including 20 clusters of size 25. Distributions for the power of this design are obtained from the prior distributions constructed by fitting model (1) to the 5 type I ICC estimates, and by fitting models (3) or (5) to all 18 ICC estimates (Figure 3). The point estimates (means) of power are similar, but the lower quantiles of the distribution are substantially higher when making use of all ICC estimates, in particular when also weighting for varying relevance using model (5). Thus resources can be saved in terms of numbers of patients and clusters recruited. After using all available information to allow formally for the uncertainty about the ICC, it would seem very reasonable to choose a design including 20 clusters of size 25 since we believe there to be only a 2.5% chance that the power of this design will fall below 75%. If relying on the type I ICC information alone, in this example we would feel less confident about choosing this design, although of course this might not always be the case.

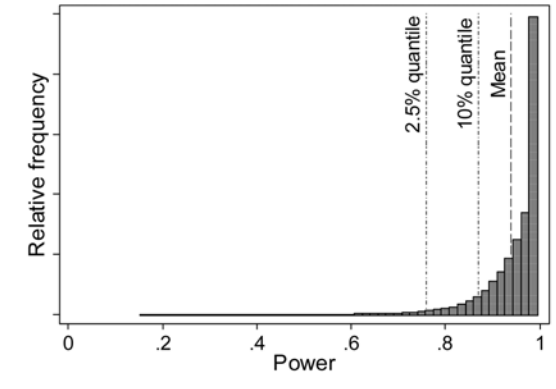
A practical approach to using a prior distribution for ρ to aid choice between several possible study designs is demonstrated in Figure 4, using the distribution resulting from model (5). Because resources are limited, it is very common for some aspect of a cluster trial design such as number of clusters, cluster sizes or total number of patients to be fixed in advance. Supposing that the maximum number of patients for the trial in our target setting is fixed at 300, Figure 4 plots summary characteristics of the distributions for power against different possible numbers of clusters. The point estimate (mean) of power reduces as the number of clusters decreases, but may be thought reasonably adequate under all four designs. However, the spread of the



(a) Type I, model (1)



(b) Types I&II, model (3)



(c) Types I&II, model (5)

Figure 3 Distributions for the power of a trial in the target setting including 20 clusters of size 25, constructed using prior distributions for the underlying ICC ρ.

probability distributions over lower levels of power show that the designs including 20 or 30 clusters are more likely to provide acceptable levels of power, given our uncertainty about the magnitude of the ICC. If possible, for our example, it would therefore be preferable to divide the 300 patients between 20 or 30 clusters.

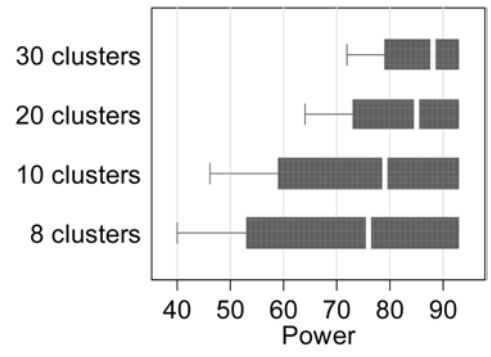


Figure 4 Power of designs including 300 patients divided between 30, 20, 10 or eight clusters, derived from model (5). Plots show mean power (central markers) with 10 and 90% quantiles (box), and 2.5 and 97.5% quantiles (whiskers). (The 97.5 and 90% quantiles are equal in these examples.)

Application in data analysis

The ICC is often imprecisely estimated in a single cluster trial because the number of clusters randomised tends to be small. Several conventional approaches to analysis such as classically estimated hierarchical models and marginal models based on generalized estimating equations assume the ICC to be known when estimating the intervention effect, which may lead to inappropriately narrow intervals. A Bayesian approach to analysis allows for the imprecision in model parameters, while also allowing incorporation of external evidence in the form of prior distributions [5]. To demonstrate the potential advantages of including prior information, we analyse a small data set in our target setting. The POST trial [21] provides measurements of total cholesterol for 123 patients from 43 general practices. We fit the simple hierarchical linear model below [4], while declaring $N(0, 10^4)$ prior distributions for α and β, and a Uniform[0,10] prior for the within-cluster standard deviation σ_e.

$$\mathrm{E}(y_{ij}) = \mu_{ij} = \alpha + \beta x_i + u_i$$
$$y_{ij} \sim N(\mu_{ij}, \sigma_e^2), u_i \sim N(0, \sigma_u^2).$$

Results obtained when specifying three of the informative prior distributions constructed earlier for the ICC $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$ are compared against those from specifying a Uniform[0, 10] distribution for σ_u as a vague prior for the extent of clustering (Table 3). When incorporating prior information on the ICC, the 95% intervals for ρ and σ_u are considerably narrower and the central estimates are smaller. Consequently, the 95% intervals for the intervention effect are narrower. In this example, the conclusions are not sensitive to choice of informative prior distribution: the results from the prior constructed by fitting model (1) to the type I

Table 3 Results from Bayesian analysis of the POST trial cholesterol measurements, under a vague prior distribution for the between-cluster standard deviation σ_u or informative prior distributions constructed for ρ using the available ICC estimates

Prior for extent of clustering	Intervention effect β	σ_u	ρ
Uniform[0,10] prior for σ_u	0.23 (−0.20, 0.68)	0.24 (0.01, 0.59)	0.049 (0.0001, 0.27)
Prior constructed for ρ: Type I, model (1)	0.21 (−0.19, 0.62)	0.13 (0.008, 0.28)	0.015 (0.0001, 0.068)
Prior constructed for ρ: Types I and II, model (3)	0.22 (−0.19, 0.62)	0.14 (0.038, 0.32)	0.016 (0.004, 0.088)
Prior constructed for ρ: Types I and II, model (5)	0.21 (−0.19, 0.62)	0.13 (0.061, 0.25)	0.015 (0.003, 0.056)

estimates, and from the priors constructed by fitting models (3) or (5) to all available ICC estimates are numerically similar. In other examples, the conclusions may be affected by the choice of informative prior.

Discussion

This paper addresses the issue of how to make use of multiple ICC estimates of varying relevance and precision in design and analysis of cluster randomized trials. We have presented models to incorporate estimates representing clustering of the target outcome or similar outcomes within the target cluster type, with allowance for the relevance of the outcome and patient population. In our example, downweighting less relevant estimates increased the precision of the anticipated ICC.

Our Bayesian approach has a number of advantages. It takes into account the full distribution of the ICC, representing the uncertainty given prior information. In sample size calculations, this is preferable to simply choosing an unnecessarily high ICC value in order to provide a conservative conventional power calculation. Our approach yields a design that provides adequate power across the likely range of ICC values, and may save resources in terms of numbers of patients and clusters recruited. Uncertainty in other parameters, such as the target treatment effect, may be as important as uncertainty in the ICC. The Bayesian approach readily extends to encompass uncertainty in such multiple parameters. Also, in analysis, greater precision in the treatment effect estimate may result from using the informative priors that we have developed for the ICC.

Researchers who wish to adopt these methods need to consider the following steps:

1) Identify relevant ICC estimates from data-bases or through systematic review, and arrange them as in Table 1.
2) Assess relevance weights for outcomes and studies: broadly speaking this is the proportion of the total (nonsampling) variance that is not due to bias.
3) Explore a range of models and check information in data about important parameters.
4) Examine the sensitivity of main conclusions to reasonably supported models.
5) Report "robust" result if conclusions broadly unchanged (as in the illustrated example). Otherwise acknowledge sensitivity of conclusions.

The models could be extended to incorporate ICC estimates from clusters other than the target cluster type. When appropriate to assume the ICC values to be distributed around the same mean as the target ICC, estimates for other cluster types could be included and weighted for relevance using exactly the methods presented. In other cases, systematic bias in one direction will be expected; for example, correlation within households is expected to be higher than within general practices. The magnitude of the bias could be assumed known or given a prior distribution and estimated [16] (although estimation may be difficult in a small data set). Similarly, it is possible in our example to include a bias term which distinguishes estimates corresponding to total cholesterol from estimates corresponding to other outcomes, rather than acknowledging relevance through the weights alone. Since the total number of ICC estimates is small, the bias is imprecisely estimated, which leads to a wider distribution for ρ. In principle, given enough data, one could fit a model which includes separate fixed effects for each of the outcome types represented, in addition to weighting for relevance.

To allow for the imprecision in each ICC estimate, the models made use of Swiger's distributional approach to constructing confidence intervals for the ICC [15]. Alternative methods which similarly require only minimal information about the source data set are Fisher's transformation approach [22] and Searle's approach [23]. When allowing for the imprecision in a single ICC estimate in the context of trial design, Turner *et al.* [3] found the practical conclusions drawn from these three approaches to be similar. Different assumptions for the distribution of the underlying ICC values could be used [3], but employment of a Normal distribution on the logit scale offers greater elegance than truncation to the range $[0, 1]$ on the natural scale.

A general framework for constructing prior distributions through modelling observed estimates

from a number of earlier studies is provided by Spiegelhalter *et al*. [16]. The approach of weighting for relevance through multiplicatively increasing the variance assumed for less relevant observations has been commonly used in other contexts [24–27]. However, the method has been criticized on the basis that there is no means for determining appropriate values for the fixed weights [28]. It is advantageous instead to increase the variance additively in order that weighting does not disappear from the model when variances are small [17], and to introduce uncertainty. We note that the subjective weights w_{ml} and w_m could be obtained through formal procedures for elicitation of prior beliefs (for example see [29]). Tritchler [30] took a different approach and based weighting on subjective values expressing the probability of relevance; a likelihood-based method was proposed for estimating a parameter of interest while incorporating the varying probabilities of relevance for each of the source studies.

The additive approach to weighting presented in this paper is novel in making use of the ordering of subjective weights, while allowing for uncertainty in the extent to which less relevant data should be downweighted. This approach could be applied directly in other areas such as meta-analysis or long-term cost-effectiveness modelling, in which studies of varying quality, with different populations and differently measured outcomes, are combined for estimation of a target parameter.

Acknowledgements

We would like to thank the POST trial authors for permission to use their data.

References

1. **Donner A, Klar N.** *Design and analysis of cluster randomization trials in health research.* New York: Arnold 2000.
2. **Campbell MJ.** Cluster randomized trials in general (family) practice research. *Statistical Methods in Medical Research* 2000; **9**: 81–94.
3. **Turner RM, Prevost AT, Thompson SG.** Allowing for imprecision of the intracluster correlation coefficient in the design of cluster randomized trials. *Statistics in Medicine* 2004; **23**: 1195–214.
4. **Spiegelhalter DJ.** Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine* 2001; **20**: 435–52.
5. **Turner RM, Omar RZ, Thompson SG.** Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine* 2001; **20**: 453–72.
6. **Campbell MK, Grimshaw JM.** Cluster randomised trials: time for improvement. *British Medical Journal* 1998; **317**: 1171–72.
7. **Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ.** Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technology Assessment* 1999; **3**: iii–92.
8. **Smeeth L, Siu-Woon Ng E.** Intraclass correlation coefficients for cluster randomized trials in primary care: data from the MRC trial of the Assessment and Management of Older People in the Community. *Controlled Clinical Trials* 2002; **23**: 409–21.
9. **Family Heart Study Group.** Randomised controlled trial evaluating cardiovascular screening and intervention in general practice: principal results of British family heart study. *British Medical Journal* 1994; **308**: 313–20.
10. **Jolly K, Bradley F, Sharp S** *et al*. Randomised controlled trial of follow up care in general practice of patients with myocardial infarction and angina: final results of Southampton heart integrated care project (SHIP). *British Medical Journal* 1999; **318**: 706–11.
11. **Moher M, Yudkin P, Wright L** *et al*. Cluster randomised controlled trial to compare three methods of promoting secondary prevention of coronary heart disease in primary care. *British Medical Journal* 2001; **322**: 1338–42.
12. **Campbell M, Grimshaw J, Steen N.** Sample size calculations for cluster randomised trials. *Journal of Health Services Research and Policy* 2000; **5**: 12–16.
13. **Kinmonth A-L, Woodcock A, Griffin S, Spiegal N, Campbell MJ.** Randomised controlled trial of patient centred care of diabetes in general practice: impact on current wellbeing and future disease risk. *British Medical Journal* 1998; **317**: 1202–208.
14. **Smith SM.** The North Dublin Diabetes Shared Care (DiSC) Project: a randomised controlled tiral with qualitative and economic analyses. MD Thesis, University College Dublin, 2002.
15. **Donner A, Wells G.** A comparison of confidence interval methods for the intraclass correlation coefficient. *Biometrics* 1986; **42**: 401–12.
16. **Spiegelhalter DJ, Abrams KR, Myles JP.** *Bayesian approaches to clinical trials and health-care evaluation.* Chichester: Wiley 2004.
17. **Spiegelhalter DJ, Best NG.** Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statistics in Medicine* 2003; **22**: 3687–709.
18. **Spiegelhalter DJ, Thomas A, Best NG.** *WinBUGS Version 1.2 User Manual.* Cambridge: MRC Biostatistics Unit 1999.
19. **Brooks SP.** Markov chain Monte Carlo method and its application. *The Statistician* 1998; **47**: 69–100.
20. **Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A.** Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 2002; **64**: 583–640.
21. **Feder G, Griffiths C, Eldridge S, Spence M.** Effect of postal prompts to patients and general practitioners on the quality of primary care after a coronary event (POST): randomised controlled trial. *British Medical Journal* 1999; **318**: 1522–26.
22. **Ukoumunne OC.** A comparison of confidence interval methods for the intraclass correlation coefficient in cluster randomized trials. *Statistics in Medicine* 2002; **21**: 3757–74.
23. **Searle SR.** *Linear models*. New York: Wiley 1971.
24. **Greenhouse JB, Wasserman L.** Robust Bayesian methods for monitoring clinical trials. *Statistics in Medicine* 1995; **14**: 1379–91.
25. **Prevost TC, Abrams KR, Jones DR.** Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. *Statistics in Medicine* 2000; **19**: 3359–76.

26. **Ibrahim JG, Chen MH.** Power prior distributions for regression models. *Statistical Science* 2000; **15**: 46–60.
27. **Tan SB, Dear KB, Bruzzi P, Machin D.** Strategy for randomised clinical trials in rare cancers. *British Medical Journal* 2003; **327**: 47–49.
28. **Eddy DM, Hasselblad V, Schachter R.** *Meta-analysis by the confidence profile method: the statistical synthesis of evidence*. San Diego, CA: Academic Press 1992.
29. **Chaloner K.** Elicitation of prior distributions. In Berry DA, Stangl DK eds. *Bayesian biostatistics*. New York: Marcel Dekker 1996; 141–156.
30. **Tritchler D.** Modelling study quality in meta-analysis. *Statistics in Medicine* 1999; **18**: 2135–45.
31. **Premaratne UN, Sterne JAC, Marks GB, Webb JR, Azima H, Burney PGJ.** Clustered randomised trial of an intervention to improve the management of asthma: Greenwich asthma study. *British Medical Journal* 1999; **318**: 1251–55.
32. **Fahey TP, Peters TJ.** What constitutes controlled hypertension? Patient based comparison of hypertension guidelines. *British Medical Journal* 1996; **313**: 93–96.
33. **Eccles M, Grimshaw J, Steen N** *et al.* The design and analysis of a randomized controlled trial to evaluate computerized decision support in primary care: the COGENT study. *Family Practice* 2000; **17**: 180–86.

Appendix

Estimates reported as equal to zero

Since ICC estimates reported as equal to 0 commonly represent negative ICC estimates which have been truncated upwards, it is inappropriate to base the distribution for the underlying ICC ρ_l on the value $\hat{\rho}_l = 0$. Instead, we use the information that $\hat{\rho}_l \leq 0$; the probability that $\hat{\rho}_l \leq 0$ is $\Phi(-\rho_l/\sqrt{V(\rho_l, N_l, k_l)})$ under Swiger's assumption, where Φ represents the cumulative distribution function for the Normal $(0, 1)$ distribution. By assuming a Bernoulli distribution with probability $\Phi(-\rho_l/\sqrt{V(\rho_l, N_l, k_l)})$ for the binary variable indicating $\hat{\rho}_l \leq 0$, we obtain a distribution for ρ_l.

Constructing distributions for power

Suppose that the trial will include N individuals, divided between k clusters. The ICC value affecting the power is equal to the future observed ICC $\hat{\rho}$, if analysis will be based on a method such as marginal models using generalized estimating equations or classical hierarchical modelling, in which $\hat{\rho}$ is regarded as known [3]. Using Swiger's assumption [15], we calculate a distribution for $\hat{\rho}$ from the informative prior distribution for the underlying ICC ρ:

$$\hat{\rho} \sim N(\rho, V(\rho, N, k)).$$

Negative values obtained for $\hat{\rho}$ under this expression are set to 0, since this is standard practice for analysis of cluster trial data, and we generate a mixed distribution for $\hat{\rho}$ [3].

The power to detect a difference of δ standard deviations at a two-sided significance level of α is calculated from $\hat{\rho}$ using the formula below, where Φ is the cumulative Normal function. This supplies a distribution for the power of a design including N individuals and k clusters.

$$\pi(\hat{\rho}) = \Phi\left[\delta\sqrt{\frac{N}{4\{1 + (N/k - 1)\hat{\rho}\}}} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right].$$