Author's response to reviews

Title: Planning a cluster randomized trial with unequal cluster sizes: practical issues involving continuous outcomes

Authors:

Lydia Guittet (lguittet@infonie.fr)
Philippe Ravaud (philippe.ravaud@bch.ap-hop-paris.fr)
Bruno Giraudieu (giraudieu@med.univ-tours.fr)

Version: 2 Date: 7 March 2006

Author's response to reviews: see over
Point by point answer to A. Donner’s review.

« 1. The recommended sample size approach based on "minimum variance weights" may be more sensitive to the guessed value of the ICC than the other methods, since the ICC appears in four separate terms. In this case, for situations where there is little prior knowledge available concerning the value of the ICC, one or more of the other approaches considered may be preferable in practice. Therefore it would be interesting to compare the robustness of the different approaches to misspecification of this parameter.”

We assessed the sensitivity of the corrections to the guessed value of the ICC by calculating the expected power assuming a different a posteriori ICC than the a priori postulated one. We considered two a priori postulated ICCs (0.005 and 0.020). Pareto repartition of cluster sizes was assumed. Sample sizes were calculated by use of the different modified variance inflation factors. Expected power was calculated with the variance inflation factor modified by use of minimum variance weights. Indeed, this latter correction has been shown to lead to a calculated power that differs from the real power (estimated through simulations) by no more than 3.8%. Results are presented in a new section (section 5) and are illustrated by a figure (figure 1).

The paper was corrected as follows:

- A new section (section 5) and a new figure (see figure 1) were added:

“5. ROBUSTNESS OF SAMPLE SIZE ADJUSTMENT FOR UNBALANCED TRIALS WITH MISSPECIFICATION OF THE ICC

5.1 Method

We assessed the robustness of the different sample size adjustments for Pareto-like unbalanced trials with misspecification of the ICC. We considered an effect size of 0.25, a priori postulated ICCs of 0.005 and 0.020 and the combinations of number of clusters and cluster sizes previously used (see sample sizes in table IIa). Then, for each weighting method, (i.e., for each total number of subjects of each arm Nw1, Nw2, Nw3) we plotted the expected power calculated for a pre-specified ICC as a function of the real ICC (which will be a posteriori assessed). This power was calculated by use of the variance inflation factor VIFw3 derived from minimum variance weights, because it allows for calculating an expected power that does not differ from the empirical one by more than 3.8% in the situations explored in table IIa (data not shown). For reference, we also plotted the expected power (calculated with the usual VIF) as a function of the real ICC in cases of no imbalance in cluster size.

5.2 Results

Results are displayed in figure 1(a) and 1(b) for an effect size of 0.25 and a priori postulated ICC values of 0.005 and 0.020, respectively.

-- Insert Figure 1 about here --

As expected [20], in any situation, the power decreases as the ICC increases, and this result is all the more important when the number of clusters is low. In the planning situations explored, minimum variance weights and cluster size weights curves are very close, except when 20 clusters per intervention arm are randomized and the ICC is a priori fixed at 0.020, but this latter situation is extreme, as discussed in section 4.3. Otherwise, the power
associated with equal weights remains greater than that associated with minimum variance weights in any situation. However, this finding probably just reflects that the use of this weighting system leads to higher required sample sizes than the use of a minimum variance weights system (cf Table II) and therefore higher power. In any case, imbalance in cluster size is associated with a higher sensitivity to the a priori-specified ICC than constant cluster size. For example, let us consider the case of 20 clusters per intervention arm: if the ICC is a priori postulated at 0.005, but in reality equals 0.015, the power associated with constant cluster size decreases from 0.80 to 0.75 only, whereas the power associated with Pareto repartition decreases from 0.80 to 0.68 (with the minimum variance weighting system). However, all weighting systems show great sensitivity to the actual value of the ICC. Consider the former example (ES=0.25, g=20 and Pareto repartition, increase in ICC from 0.005 to 0.015), the power associated with equal weights will decrease from 0.98 to 0.90, and the power associated with cluster size weights from 0.80 to 0.68. Thus, if little prior knowledge is available concerning the value of the ICC, the sensitivity analysis involving several values of ICC is of major importance, particularly when imbalance in cluster size is expected.”

- The first paragraph of the discussion was modified by adding the following before the last sentence:

“The higher sensitivity of severely unbalanced trials to the a priori-postulated value of the ICC compared to that of balanced trials emphasized the necessity of a sensitivity analysis on this parameter.”

- The end of the introduction was modified as follows:

“Section 2 describes the random effects model used to simulate clustered data; section 3, the simulation design used to evaluate the loss of power due to imbalance in cluster size and the findings; section 4, corrections of the variance inflation factor to allow for cluster size inequality evaluated by simulation; section 5, robustness of these corrections to misspecification of the ICC; section 6, practical guidelines for the planning stage of cluster randomized trials; and section 7, perspectives for future research.”

2. On a related point, would the recommended approach also be expected to show optimality for non-normal distributions in which the response variance depends on the mean? In particular it is important to know whether the authors’ recommendation would hold for binary outcome measures, since they are so common in practice.

We agree that our work is limited to a continuous outcome with a normal distribution. We consider that the extension of our results to non-normal distributions (especially binary ones) would be a better subject of a complementary but different work; otherwise, the present manuscript would become even more longer and harder to read. Thus, we specified in the discussion that this type of investigation could be an extension of our work for future research, and, to make things perfectly clear, we specified in our text that our work is restricted to continuous normal data.

The paper was modified as follows:

- Title: “Planning a cluster randomized trial with unequal cluster sizes: practical issues involving continuous outcomes”.
Another limitation is that our work focused on normally distributed continuous outcomes. More work is needed to extend our results to non-normal distributions, especially with binary variables.

3. The authors assume in their development that the underlying design involves unrestricted randomization. However for studies in which the cluster sizes are expected to be highly imbalanced, matching or stratification on cluster size is often built into the design. How does this affect the authors’ recommendations?

In our simulations, data were generated according to a Pareto distribution of cluster size. We therefore considered two strata: one of “big” clusters and one of “small” clusters. We acknowledged that this stratification may appear to be rough, but a more precise stratification on cluster size may be difficult to build into the design.

Regarding the analysis of the simulated trials ignoring the stratification, matched or stratified analyses were demonstrated to be attractive only if the number of clusters is high and the matching or stratifying correlation is high\(^1\).\(^4\).

We thus added the following in section 3.3 to clarify the use of unrestricted analysis:

“Data analysis involved no stratification on cluster size.”

4. The differential ability of physicians to recruit subjects within each cluster may not only result in imbalance in cluster size, as stated on page 3, but could also seriously threaten validity. As discussed by Puffer et al (2003), the characteristics of subjects in the two groups may then be systematically different.

We fully agree with this remark. In page 3 of our manuscript, our aim was to list reasons why imbalance in cluster size may occur, and a differential ability of physicians to recruit subjects could be one of these. In this section, the Introduction, we did not mention the potential risk of bias discussed by Puffer et al because our work focuses on consequences of an imbalance in cluster size. However, we acknowledged this in the Discussion and we rephrased the end of this section to clarify that the imbalance in cluster size is not the only limitation of cluster randomization trial results.

The end of the second paragraph of the Discussion section and the last paragraph were therefore modified as follows:

“Finally, we restricted our work to cases of no differential recruitment between arms, thus considering that imbalance is the same in the two arms. Such a hypothesis may be questionable in cluster randomized trials: since inclusion is posterior to randomization, this may indeed induce differential recruitment and imbalance in patient characteristics, which may lead to questioning the results of the study [23].

In conclusion, our study demonstrates that severely imbalanced trials with continuous outcomes may be highly underpowered. If such imbalance in cluster size can be anticipated at the design stage, minimum variance weights correction should be used to inflate the required sample size. A priori estimation of the expectable imbalance would be facilitated if more
details on cluster sizes were given in published cluster randomized trials, as was recently advised in the extension of the CONSORT statement for cluster randomized trials [24]. Moreover, such publication of cluster sizes would be of particular interest to assess the real power of the trial conducted.

5. The median number of 34 trials referred to on page 5, does not come from a review of cluster randomized trials in general, but rather from a review only of trials in primary care settings.

We agree with this remark. This median number of clusters, and even smaller ones, were also described in more general settings in recent reviews of cluster randomized trials [5-7].

We modified our manuscript as follows:

“The ICC values were chosen according to previously published estimates [5-15], and the number of clusters is in agreement with that from a recent review of cluster randomized trials in primary care settings in which the median number of randomized clusters was estimated at 34 [13].”

6. With respect to the discussion in Section 2.2, it should be pointed out that in the balanced case considered, an analysis at the individual level using ANOVA is algebraically equivalent to an analysis of the cluster means using a two-sample t-test.

We suppose that this remark concerns the second paragraph of the Background section rather than section 2.2. We agree that cluster-level analysis and ANOVA individual-level analysis are strictly applicable only to balanced designs, and the resulting test statistics are then identical \(^4\). In the case of unbalanced trials, the SAS procedure Proc Mixed we used is preferable.
Point by point answer to Torgerson’s review.

The authors are indebted to David J Torgerson for his kind comments.
References


