Reviewer’s report

**Title:** Randomised trial investigating the relationship of response rate for blood sample donation to site of biospecimen collection, fasting status and reminder letter: the 45 and Up Study

**Version:** 1  **Date:** 14 June 2012

**Reviewer:** Bruno Giraudeau

**Reviewer’s report:**

**General comment**

The present manuscript reports the results of a pragmatic trial aimed at increasing response rate for blood sample donation. The research question is pragmatic and of real relevance because very few is known on the subject and because collecting data and biospecimens always raises difficulties in large cohort studies.

However, I think there are too many objectives in the present study, and too many results which are reported. This makes the paper hard to read and we come to forget the main objective of the study. Thus:

- We have actually two trials: one in a urban area, the other one in a rural area
- We have actually 3 interventions (site of biospecimen collection, fasting, reminder), and Figure 1 is absolutely necessary to understand the randomization design
- Individual factors associated to the response rate are indeed of interest, but this comes down to report both the results of a randomized trial and of an observational study
- I do not well understand the relevance of results reported in Table 4:
  - Why comparing height, weight, blood pressure measurements between dedicated clinics and pathology services? This is in no way “data quality”.
  - What is the relevance of assessing the difference between self-reported and measured height and weight?
  - Worse: what is the relevance of reporting correlation between self-reported and measured height and weight?

In the end, doing so makes the paper long (discussion section is 6 page long!) and hard to read. I would strongly suggest to focus on the results of their randomized trials (and indeed, actually, there are two randomized trials, cf supra)
and to report them in agreement with the CONSORT Statement and its extensions.

Specific comments
1) The conclusion of the abstract does not provide enough information.
2) The end of the “Background” section explicitly reports the hierarchy of the research questions investigated in this study:
   o First question: influence of the site of data collection
   o Second question: influence of the fasting status
   o Third question: influence of a reminder letter
Therefore, I think the statistical analysis plan should follow this hierarchy, which is not presently the case.
3) Secondary aims listed in the end of the “Background” section should better be dropped from the present paper, for greater clarity. Moreover, I do not well understand how is assessed the “feasibility of collection of biospecimens and data on physical measures”. Is this assessed through the primary outcome?
4) I would strongly suggest reporting results as if two independent randomized trials had been conducted: one in an urban area, and the other in a rural area. I would add that sample size calculation had not been performed considering data of these two areas would be pooled.
5) Participant information: I supposed participant consented for the collection, but were not informed of the trial hypothesis. If such, this should be mentioned. Indeed, not fully informing participants of the study hypothesis is fundamental in the present study, since it is the only way of having some form of blinding in the present study. So a complete reporting of this point is necessary. Moreover, we should have a section talking about blinding in this study.
6) For greater clarity, I would suggest having a “Design” section explaining the randomization strategy, in agreement with the hierarchical objectives of the study. Of note, this section is advised in the CONSORT guidelines.
7) Page 9-10: there is a complete description of how blood specimen were collected and processed for storage but, in the end, quality of blood specimens is only assessed through the number of aliquots. So I wonder whether such a precision is necessary.
8) Data were collected during a one month period. Why having chosen such a short delay?
9) I do not well understand the sample size calculation. The hypothesis is a RR
of 1.3, and a 30% response rate is expected.

If 30% is the response rate in the control group, the expected response rate in the experimental group is 39%, and the sample size required (for a chi-square test, although I recognize that the main analysis was to fit a log binomial model) is then 437 per group.

If 30% is the overall mean response rate, it comes down that the response rate in the control group is 26% and the expected response rate in the experimental group is 34%. The sample size required is the 514 per group.

So precisions are expected.

Otherwise, authors never took into account multiplicity in statistical analysis. This point must be discussed, and authors have to explain us how they dealt with it.

10) The result section should be reported in agreement with the objective hierarchy previously specified.

In the end, I do think this paper is of interest, because of its originality and because of its relevance. However, it should be re-written to be more concise.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.