

## Reviewer's report

**Title:** Rationale, design and conduct of a primary dental care based intervention to improve professional performance on routine oral examinations and the management of asymptomatic impacted third molars: a cluster randomised controlled trial with an incomplete block design, a trial protocol.

**Version: 2 Date:** 30 November 2006

**Reviewer:** Ian Steen

### Reviewer's report:

I have been asked by an editor to look specifically at the power calculation and comment on the power of the study to detect the effect specified. In order to comment on the sample size I need to first understand the study design.

This study is in fact two studies in one – each evaluating a different intervention:

1. Study 1; behaviour 1; intervention 1; outcome measure 1
2. Study 2; behaviour 2; intervention 2; outcome measure 2

Each intervention targets a particular behaviour and different outcome measures will be used to assess the impact of each intervention. GDPs will be assigned to the intervention arm in one of the studies and the control arm in the other.

Implicit in the design of this study is that intervention 1 should not influence behaviour 2 and that intervention 2 should not influence behaviour 1. Only if this is true will the study yield unbiased estimates of the two interventions.

For each study outcome appears to be assessed at the level of the individual patient. For the ROE study the outcome can be considered a binary measure of whether that patient is being appropriately managed (with respect to recall interval). For the MIM study I think that here too the outcome can be considered a binary measure of whether that patient is being appropriately managed (with respect to the patient's third molar) but I'm unclear about the role of the number of x-rays here.

I assume that in general dental practice a single identified practitioner is responsible for the management of an individual patient. A measure of behaviour for each practitioner will therefore be available.

The unit of randomisation is a critical issue. In the methods section it is stated that the unit of randomisation is an IQual-group. I assume that these groups are pre-existing and that there will no control over the allocation of GDPs to these groups. Unfortunately the sample size calculation is written as though the unit of randomisation is the GDP ("each arm should comprise 480 patients clustered within 24 GDPs"). The very high ICC of 0.29 would be more in line with what I expect from patients clustered within GDPs than either patients or practitioners clustered within IQual-groups.

The phrase "an effect size of 20%" is unclear. Given that we have a binary outcome, do the authors mean that the proportion of "successes" in one arm will be 20% greater than the proportion of "successes" in the other? If so, in order undertake such sample size calculation it would be necessary to specify the actual proportions in each arm. Using plausible values (and assuming that the cluster is the GDP and that on average there will be 20 patients per GDP) I am not able to replicate the figures given in the paper.

Perhaps a more natural way to approach the sample size calculation would be to consider GDPs nested within IQual-groups. There is a measure of behaviour for each GDP (based on a sample of 20 patients?). These measures would then be compared between the two arms of the trial taking into account the clustering of GDPs within IQual-groups. A design with two groups of 24 GDPs (corresponding to two groups of 3 IQual-groups) would yield 90% power to detect an effect size of 1 (a difference of 1 standard deviation) assuming a significance level of 5% and an ICC of 0.01. An effect size of 1 would be considered very large in the context of a study targeted at behaviour change. In my opinion an effect size of 0.5 or smaller would be more appropriate for this type of trial.

The above text is the extent of my report but I am unable to submit it without responding to some "tick box" questions that I am not qualified to answer such as "What next?". Please ignore responses to the following questions.

**What next?:** Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

**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.