Reviewer's report

Title: The stepped wedge trial design: literature review

Version: 2 Date: 21 September 2006

Reviewer: George Borm

Reviewer's report:

General
Due to its conciseness and its clear and interesting message, it was a pleasure to review your article. I only have minor comments.

-------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

None

-------------------------------------------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1) I very much appreciate your concise writing, but the convincing motivations of the stepped wedge design are kept at an abstract level (ethical reasons; logistical, practical financial constraints). I think this is fine when the subject is well-known and/or if an abundance of other literature discussing this is available. Since your aim is to promote a relatively unknown design, I think a motivating example would help. In the present form it is only near the end (page 10) that I could give “flesh and blood” to these motivations. In my view it would already suffice to include few words e.g. vaccinations as example of ethical reasons and only one vaccination team having to serve several distant areas as example of logical/practical/financial constraints.

2) In your sentence (p. 10) “Reported motivations were ethical (does not require withholding the intervention) in four studies [14 16, 18, 19].” the connection between the first part of the sentence and the part between parentheses is unclear to me: do you mean that it was unethical to withhold the intervention from some of the participants/clusters?

3) The purport of the (first part of the) sentence (p.10) “The authors reported a design to use an RCT for evaluation in four studies [7, 12, 16, 20] and …” is unclear to me. Do you mean that in four studies one intended to make a comparison control-intervention or in four studies one mentioned use of randomisation or ..?

-------------------------------------------------------------------------------

Discretionary Revisions (which the author can choose to ignore)

4) Figure 1 suggests that each experimental unit (clusters or participants) is measured in both conditions and, in general, not as much times in the one condition as in the other. The text referring to Fig 1 (Background: end of second paragraph) states that “Data analysis … involves comparison of the data points in the control section of the wedge with those in the intervention section”. While this is sufficiently generally formulated to be true, I wonder whether this analysis should account for correlation within the experimental unit (typically in the cohort design of course) but perhaps also when different participants are sampled at different times, but still from a “common environment” which induces a correlation too. You do mention time trends, which also can induce correlation where one unit followed in time is concerned, but not this aspect.

5) Your sentence (p. 5) “The Department for Education and Skills ruled out a cluster trial where deprived areas would be randomised to either receive the intervention or act as controls, since to intervene in some areas but not in others was judged unacceptable” puzzled me: your sentence suggests that all areas are deprived, but I guess only the ones that are randomised to act as controls are deprived?

6) Do I understand from your sentence on p. 11 “ ….Priestly et al. [20], with the latter using matched
randomised pairs where the wards in each pair have the same amount of time in either the control or intervention groups." That one ward in the pair switches from intervention to control in the end (some kind of cross-over design)? Or do you mean that some pairs get the intervention late, while others early?

7) In the discussion section I personally miss a discussion of possible/suspected pitfalls in the analysis of stepped wedge designs. If you have ideas about this and share these, it would make your case for a statistical model in your next article stronger.

8) You mention that in the conclusions that it is “questionable whether studies with only two steps should be considered as stepped wedge designs.” Could you give some flesh and blood to your hesitations?

What next?: Accept after minor essential revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:

I declare that I have no competing interests