Author’s response to reviews

Title: Efficiency of two-phase methods with focus on a planned population-based case-control study on air pollution and stroke

Authors:

Anna Oudin (anna.oudin@med.lu.se)
Jonas Bjork (jonas bjork@skane.se)
Ulf Stromberg (ulf.stromberg@med.lu.se)

Version: 4 Date: 25 July 2007

Author’s response to reviews: see over
Dear Editors,
We hereby resubmit the manuscript *Efficiency of two-phase methods with focus on a planned population-based case-control study on air pollution and stroke*.

We think that the comments of the reviewer were initiated and insightful, and they have lead to improvement of the paper. Below, we have addressed the comments, one by one.

Sincerely,
Anna Oudin, Jonas Björk, Ulf Strömberg

1. The Background/Introduction does not really link up with the substantial literature on 2-phase design and analysis: it mainly reviews the literature on air pollution instead. But this is a methodology paper! An additional document is entitled "overview of 2-phase methods": why is material from this document not incorporated into the background? This is the traditional way of explaining what is already known so that the new method in a research paper can be justified. Without that we cannot tell what is new.

   Authors: We chose to separate the background on two-phase methods from the background of health effects of air pollution because it was suggested by the reviewers that the background was too long, and that the background of two-phase methods was too technical and not necessary for the total understanding of the paper. It certainly has beneficial aspects to include the background of two-phase methods in the general background, but we think that although this is a methodology paper, it is built up around an example of health effects of air pollution, which we think that many of the readers of environmental health will be interested in without special interest in the history of two-phase methods. We think that presenting the background of two-phase methods in a separate section (which is easily allowed for with online publishing), to be read by methodologically initiated and interested readers is a rather appealing solution.

2. Importance of design
The issue of how subjects come to be included in the 2nd phase of the simulations is not clear nor is the general importance of this question acknowledged.
(a) The introduction should clarify that, in general, a non-random design or even stratified random designs with incorrect analysis will lead to bias. Also the issue of self-selection needs to be differentiated from active selection by researchers as part of a study design. In the latter case some data is missing (randomly) by design and bias can be avoided; in the former it is not and bias may be unavoidable. ‘Participation’ seems to be used to mean both things in this paper – leading to confusion. Mention of ‘participation’ in the Abstract and p 14 seems misleading – it seems to imply that two-phase studies could overcome the problem of bias caused by self-selection.

   The discussion does eventually pick up these issues (cf participation bias, p13) but it would be clearer if the general principles of 2-phase designs had already been set out in the Introduction - ie which combinations of design and analysis method are valid? In general a 2-phase design, where data is MAR - see below- accompanied by an analysis which acknowledges the design will be unbiased.

   Authors: We agree that the paper would benefit from a clearer definition of what kind of participation/selection bias can be avoided with two-phase methods, and also a clearer distinguishing of self-selection from selection from design.
It was previously mentioned in the section “Overview of two-phase methods”, that “Two-phase studies are often more efficient than traditional designs and may account for bias due to varying participation rates, or varying sampling fractions, across areas.”, which is true; if the participation differs between areas, a stratified design might be used to avoid the bias stemming from self-selection. Bias due to differential participation between areas could thereby be avoided. As the reviewer points out, however, a two-phase design does not per se reduce all types of bias. This has been further clarified in the “Overview of two-phase designs” in the sentence “If participation bias is present within areas, however, a two-phase design can not be used to reduce that type of bias”.

(b). The simulations
Is the random sampling plan (p10) stratified by disease status only or by disease status and area? The abstract Conclusion refers to varying participation rates across areas. This is nowhere mentioned in the Background. The Methods seem to imply that selection probabilities vary across areas but they do not give any information.

Authors: The random sampling is stratified by disease status only, and not by area. We have added: “although in this study we do not incorporate any such differences in participation rates” to page 10.

Discussion (p13) ‘Missing at random (MAR)’ assumption is NOT violated if selection probabilities vary by area although the Missing Completely at Random (MCAR) assumption will be (cf Little and Rubin usage of these terms).

Authors: The reviewer is right, we changed Missing at Random to Missing Completely at Random.

It is earlier stated that all of the analysis methods are unbiased for the design used in the simulations but the discussion refers to the potential bias of Method 1 (presumably for designs other than those simulated) – more details are needed.

Authors: We clarified the sentence by adding “In our setting, participation bias is not present, but generally” to the beginning of the sentence.

3. Description of the design of the study (as implied by the simulations) should precede the description of the analysis methods since choice of correct methods depends on design. It would be helpful to distinguish between 2-phase design and 2-phase analysis. The idea of a 2-phase design is not new; the authors’ proposed ‘method’ is a method for analysis of certain 2-phase study designs.

Authors: We agree with the referee and have restructured the method section so that the description of the simulations precede the description of the methods. We have also been more careful to distinguish between two-phase design and two-phase analysis through the paper.

P5, line 10 should read: “we simulate data from a two-phase design”.

Authors: we changed the sentence according to the referee’s suggestion.

P10, paragraph 5, 1st sentence. Distributions of what?
We think that the reviewer considers page 8 instead of page 10, and we agree with her that it should be specified. We changed the sentence to “In each replication the subjects were selected randomly, so that the distributions of the exposure and co-variables among the second-phase cases and controls would reflect those distributions in the population.”

P12, last part of paragraph 1 beginning ”In Table 3….”. I cannot understand the point being made here.
Authors: We agree that that sentence was unclear. We decided to cut it since it does not contribute much to the understanding of the paper.