Author's response to reviews

Title: How many mailouts? could attempts to increase the response rate in the Iraq war cohort study be counter-productive?

Authors:

Anne A.R. Tate (rosemary.tate@iop.kcl.ac.uk)
Margaret M Jones (m.jones@kcl.ac.uk)
Lisa L Hull (Lisa.Hull@iop.kcl.ac.uk)
Nicola NT Fear (Nicola.Fear@iop.kcl.ac.uk)
Roberto R Rona (Roberto.Rona@iop.kcl.ac.uk)
Simon S Wessely (Simon.Wessely@iop.kcl.ac.uk)
Matthew Hotopf (matthew.hotopf@iop.kcl.ac.uk)

Version: 2 Date: 22 August 2007

Author's response to reviews: see over
Dear Editor,

MS: 1287094085147648 How many mailouts? Could attempts to increase the response rate in the Iraq war cohort study be counter-productive?

Thank-you for considering the above manuscript for inclusion in the BMC Research Methodology. We have made the required revisions and have provided detailed replies to the reviewers' comments. We would like to thank the reviewers for their careful reading of the manuscript and helpful comments, and we feel that paper has benefited as a result.

We look forward to your response.

With best wishes,

Rosemary Tate

General Clarifications
We have rewritten the abstract to make the aims and methods clearer.
We have rewritten the last paragraph of the background to make the aims and methods clearer.
We have completely restructured the Methods section and rewritten more of it in the first person as requested by a reviewer
We have moved most of the section on simulations to “additional material”
We have moved parts of the Results section to the Discussion.
Response to reviewers

Reviewer 1. Timothy Lash
We are grateful for the Dr Lash’s comments. As requested as we have made the following changes:

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

1 The verification of the Stang et al hypothesis is somewhat reduced by the lack of response bias, the relatively small amount of misclassification, and the lack of real validity data by which to judge classification rates (they use a measure of agreement). Although these points are touched upon by the authors, I believe a stronger statement about these limitations is in order.

The main aim of this paper was to assess bias in the Iraq war study and ascertain if increasing the number of mailouts led to increased bias in the relative risk estimates. We used Stang’s model to help us achieve this aim, and also to see how closely real-life data fitted his model, but verification of this model was only a secondary aim. We apologise if the aims were not specified clearly enough. In an attempt to redress this we have rewritten the abstract to make this point much more clearly and have dropped reference to Stang’s work in the abstract.

2. In the first paragraph of the background section, the authors say there is a "considerable literature" and then cite only one paper, which is not a review. The statement and citations need to be brought into agreement.

This comment is a bit difficult to understand since citation 1 (Edwards P, et al.: Increasing response rates to postal questionnaires: systematic review) is a review paper. However, we take the point that it is not a review of whether attempts to increase response are worthwhile, and have amended the sentences as follows: “Poor response is a major source of concern in epidemiological surveys, and much effort is often spent on chasing up initial non-responders [1] with the implicit assumption that a higher response rate is associated with a more representative sample and hence lower bias.”

3. In the same paragraph, the hope is a more complete sample, and the assumption is that a more complete sample will be more representative.
We have amended this sentence - see above

4. I do not agree with the authors’ contention that bias is larger for rare outcomes. That depends on the nature of the classification problem. For example, with perfect specificity and non-differential sensitivity of outcome classification, the risk ratio is expected to be unbiased regardless of how common the outcome.
We have removed that statement.

5. In the investigation of misclassification by response wave, they say the worst case scenario hopefully provides an upper bound. If it is the worst case scenario, doesn’t it
be definition provide an upper bound?
We have removed the word ‘hopefully’

6. In the third to last paragraph of the conclusion, both "not differential" and "non-differential" in the last two sentences should be "differential" to be consistent with the paragraph's message as I read it.
Thank-you very much for pointing this out, we have corrected this.

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. Throughout, replace “in order to” with "to".
Done
2. Throughout, always provide an antecedent following "this."
Done
3. Insert "independent" after "non-differential" in the first paragraph of the background, as independent errors is also required to expect the bias toward the null.
Done
4. The methods section would be more easily read if it were written in first person active voice.
We have rewritten much of this section in the first person.

5. It’s not clear to this reviewer why there is a "(contradictory)" parenthetical in the second paragraph of the section called "investigation of responders."

We have removed this word.
Reviewer: Phil Edwards
We also thank Dr Edwards for his comments and have implemented the requested changes as follows:

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

METHODS
1) The simulation material could be simplified and use a clearer sub-heading – e.g. “Simulation of the effect of misclassification.” If the simulation methods and the formulae presented are described elsewhere (e.g. reference 6), then the methods should be summarised in two or three sentences, and then the formulae moved to an appendix, or possibly removed altogether. (This also applies to Table 1.)
We have moved most of this section to “Additional Material”

2) All statistical methods used (e.g. Chi-squared test) should be described under the heading “statistical analysis”.
Done

RESULTS
3) A Spearman rank correlation is given in the results, but no mention of using this method is given in the methods section.
Done

4) Page 8: the sentence beginning “Since the main aim was to assess...” goes on to describe a method used. This should be in the Methods section.
We considered moving this to the methods section. However, whilst we appreciate the point being made, on reflection with my co authors we feel that this sentence is better left in the results section, since the decision to use this measure was based on the first section of the results. We decided to use this measure because it was non-differential between TELIC and Era, but we did not know this until we had done the analysis. We have moved the sentence to the section on simulations in order to emphasise this point.

5) There is discussion in the Results section that should be moved to the discussion (e.g. first sentence “In common with most surveys..”; second paragraph of ‘Investigation of responders’ “..indicating that misclassification of health outcomes may be present...”).
We have moved the first sentence to the discussion section. We have removed “..indicating that misclassification of health outcomes may be present...”

CONCLUSIONS
6) The section headed “Conclusions” is actually a discussion and should be renamed as such.
Done
7) The opening sentence should be a statement of the principal findings, not a reiteration of the study aims.
We have removed the first sentence of the conclusions.
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

8) Abstract: add “disorder” after “stress” in the full PTSD.
Done

Discretionary Revisions (which the author can choose to ignore)

9) The discussion shouldn’t refer to the results of Table 6. This should be described in the results section.

We have removed the reference to the table
We also thank Dr Johnson for his comments and have implemented the requested changes as follows:

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

I would encourage the author(s) to think about the degree to which one of both of their misclassification measures may have itself been biased by respondent behaviors, particularly acquiescent response styles.

Thank-you for this suggestion. We already checked to see if the proportion of discrepancies was greater in some groups than in others, but the only related factors were Service and deployment (as we reported in the paper). Following your suggestion we investigated the proportion of extreme answers to the four SF36 health questions and found that “right-ticking” of all four SF36 health questions is significantly more likely (p=0.0005) in ERA than TELIC, with 0.9% (n=49) of ERA vs. 0.3% (n=16) of TELIC ticking definitely false for all 4 questions. Conversely, TELIC are more likely (p=0.03) to have missing answers to at least one of these questions than ERA (3.8% vs. 3.0%).

We looked for other predictors of right –ticking and the only other one we could find was service Navy least likely). Left-ticking was almost non-existent (n=4).

This backs up the findings that we have already explained in the paper that discrepancies are more likely in ERA and that missing answers are more likely in TELIC. We have changed the sentences describing the differences based on your suggestions.

“There were slightly fewer discrepancies in the TELIC cohort; 10.9% TELIC vs. 12.6% Era (p=0.01), mainly due to the lower percentage of TELIC personnel answering "definitely false to both questions (1.7% vs. 3.3% ). These differences held after adjustment for the other only factors found to be related to discrepancies, i.e. lower rank, and Service (the Army had the highest percentage). However, the percentage with missing answers to both questions was significantly (p=0.02) greater for TELIC than Era (2.1 vs. 1.5).

And added this sentence to the Discussion

… even though the differences cancelled each other to some extent. This might be due to the fact that personnel deployed on TELIC 1 take slightly more care in answering the questions, but have more doubts how to complete them.

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

The paper could be considerably sharpened by turning it over to a good copy editor.

Following this and the other reviewers’ comments We have restructured the Methods and Results section and rewritten the abstract. We hope that it now reads much more clearly