Author's response to reviews

Title: An Empirical Comparison of Methods for Analyzing Correlated Data from a Discrete Choice Survey to Elicit Patient Preference for Colorectal Cancer Screening

Authors:

Ji Cheng (chengj2@mcmaster.ca)
Eleanor Pullenayegum (pullena@mcmaster.ca)
Deborah A Marshall (damarsha@ucalgary.ca)
John K Marshall (marshlj@mcmaster.ca)
Lehana Thabane (thabanl@mcmaster.ca)

Version: 2 Date: 28 October 2011

Author's response to reviews: see over
October 21, 2011

Anastasios Koutsos, PhD.
Senior Assistant Editor
BMC-series journals

Dear Dr Koutsos

MS: 2026795082556474-An Empirical Comparison of Methods for Analyzing Correlated Data from a Discrete Choice Survey to Elicit Patient Preference for Colorectal Cancer Screening

Thank you for your email with reviewers’ comments on our paper. All comments are very constructive and helpful. We have revised the paper in accordance with the suggested revisions. Please find attached a revised version of the manuscript with all changes highlighted using a YELLOW marker. Below are point-by-point responses to the reviewers’ comments. In addition, we have shortened the abstract to meet the 350 word limit. Finally, we have added some competing interest statement at the end of the manuscript.

We hope that this revised version is now suitable for publication in BMC Medical Research Methodology.

We look forward to hear from you.

Sincerely

Lehana Thabane, PhD
Professor, Department of Clinical Epidemiology and Biostatistics
McMaster University
Responses to reviewers’ comments
We thank the reviewers for their careful review of the manuscript and for their constructive and insightful comments. We feel that these have led to substantive improvement in presentation. Below are our responses to the specific issues raised by the reviewers.

Comments from Martina Vandebroek

Major remarks:
1. Although the journal does not aim to publish articles describing scientific methods or techniques, more details about the different models and estimation procedures are needed to fully understand the paper.
   RESPONSE: We have added more details on describing the models including the covariance structure for random-effects model, the assumptions for multinomial logistic and probit models. We also added some key references for readers. (see page 10, 11, 12)

2. Nothing is said about the correlation structure in the random effects probit model. Especially the difference between the correlation among the 3 categories in the multinomial probit model and in the binary probit model deserves some attention (this should become clear if the models are described better).
   RESPONSE: For the GEE models (logistic and probit), exchangeable correlation structure was used. For random-effects models (logistic, probit and multinomial logistic), a constant covariance structure was assumed. We have added more details regarding the correlation structure in the method section. (see page 10)

3. I suggest to include simple ML estimation of the conditional logit and the multinomial model as the results will very likely be similar and it is much easier to apply.
   RESPONSE: In our analysis, we treated the nominal (three-point) outcomes as case-specific data which means that we fitted a separate binary logit for each pair of outcome categories. Conditional logit model using alternative-specific data structure is often used in analyzing choice data. However, in our case, we are not interested in comparing test A and test B. Therefore, we decided not to include conditional logit model in this paper.

4. It is not clear how exactly the “relative importance of choice between test A and test B”? is defined in case the attributes have more than 2 levels. It also seems strange to me to talk in terms of test A and test B as these are generic profiles, wouldn’t it be better to talk about paired comparisons?…
   RESPONSE: We used the “relative importance” of the attributes to assess the relative contribution of each attribute on participants’ decision making. This criterion measured the percent change between the log-likelihood value of the full model and the log-likelihood value after removing one specific attribute out of the model. The higher the percentage, the bigger the contribution. Because the entire attribute is removed when obtained the change of log-likelihood, the number of levels of the attribute did not affect the calculation.
Test A and B were a pair of combination containing different levels of 6 attributes. They
were presented to the participants as two possible CRC screening tests (Appendix B) for
each choice task in the survey questionnaire. Therefore, we used the same name in
reporting the analysis.

5. It must be made clear that the interest is in average preference levels only, as even
when a random effects model is fitted and individual preferences can be obtained,
the main focus seems to be on the means.
RESPONSE: We have explained the difference on the estimates between random-effects
models and other models in the method section. (see page 10)

5. The references should be checked carefully (for instance the page numbers of
reference 25 are wrong) and the references should be more relevant (for instance for
the clustered robust standard error, the only references that are given are to a not
related paper in the American Statistician and one to a
technical report of the software used).
RESPONSE: For the reference 25, the correct one is On the So-called “Huber Sandwich
Estimator” and “Robust Standard Errors” by Freedman. We have corrected it and
rechecked all other references. In the revised version, it became reference 28 (see page
9).

Minor remarks:
1. As a binary response is also a categorical response, it is not always clear what
data are modeled in which way.
RESPONSE: we have replaced “categorical” with “nominal” in the entire paper. We used
binary outcome to refer two-level outcome and nominal outcome to refer three-level
outcome. (see Abstract, pages 5, 7, 10, 11, 15)

2. As there is not just one way to check the rationality of the observed choices, more
details are needed about this.
RESPONSE: In the original design, two warm-up scenarios were given in the
questionnaire. In these two scenarios, one alternative was dominant over another by
assigned with all favourable attributes. If participants failed to choose the dominant
alternative for the warm-up questions, we considered them failed the rationality test.
Because our objective of this paper is to compare statistical models on analyzing
clustered DCE data, we decided to exclude the participants who failed the rationality test
to reduce the potential noise. We have described the rationality test in the result section.
(see page 13)

3. Nowadays, a lot of researchers focus on optimal designs for discrete choice
experiments taking into account some prior knowledge of the parameters which
leads to much more efficient designs. As for several of the attributes in the study, it
is very well known what the sign of the parameters will be, better designs can be
found. A remark about this would be welcome.
RESPONSE: The original study started in 2002. At that time, very information was
available to guide the design. We used a qualitative study to inform the choice of
attributes and a pilot study was used to decide the number of choice task for each participant. Regarding to the optimal design, D-efficiency was used to choose design with smallest standard error. SAS Optex procedure was also used to choose the most efficient design based on the generalized linear model. We have commented the facts in the method section. (see page 7)

4. It is would be useful to give the 4 designs that were used in the appendix. Were no differences found between the results of the different designs?

RESPONSE: We are not very clear which “4 designs” reviewer refers to in the appendix. If reviewer asks about four procedures (Stool, Scoop, CT and Enema and X-ray) listed in Appendix A, they are related to fecal occult blood testing (FOBT), flexible sigmoidoscopy (SIG), colonoscopy (COL) and double-contrast barium enema (DCBE). However, because the objective of the original study was to identify the key attributes of CRC screening programme, no labelled test was given in the questionnaire. Our results showed that most models agreed that participants preferred a screening test using stool sample, having neither pain nor special preparation, with a 100% sensitivity and 70% specificity and only required very small cost.

Minor issues not for publication: first line of second paragraph in discussion: verb Missing
RESPONSE: The verb has been added. (see page 16)

Comments from Axel C. Mühlbacher

Major Compulsory Revisions

Decision context: Differences are obtained trough the specific design of the survey. This survey used a two-stage design. The opt-out alternative is which is often seen as a more realistic approach to the decision-making context. First respondents are asked to make decisions based on the six treatment attributes. After having made this decision, they are given the chance to “opt-out”. This design is excellent to predict the uptake of screening test (more realistic would be the integration of the opt-alternative).

In the context of the research question stated above, this specific experiment might lead to two different decisions by the respondents. The decisions that are based on the treatment characteristics will provide information about preferences for different treatment characteristics. The second decision could be seen as a statement regarding the approval of the screening itself. If respondents in principle do not favor screening tests, the decisions are more or less independent of the decision in stage one. Therefore it might be interesting for the reader to have more information on the populations voting for or against the screening (15 % always opting out and 48 % always taking the screening test). Please analyze attitudes and/or sociodemographic variables for the populations who choose to opt out (page 14). Do the results vary in these subgroups?
RESPONSE: We have conducted a subgroup analysis for the stage-one data (assess the preference according to treatment characteristics) including the participants who always chose to opt-out. The results are similar to those using entire data: the direction and magnitude of the estimates were similar.

Minor Essential Revisions

1. Respondents: The total number of subjects analyzed was 468. Each subject completed 10 choices. The design was blocked into 4 different surveys. On average 117 respondents in each block. Please provide the actual number of respondents for each block. Power calculations before conducting the research study are almost impossible; still the reader of this article might be interested if the number of respondents was sufficient to answer the research question. Please discuss this in the method section or as a limitation in the discussion section.
RESPONSE: We have added the actual numbers of participant for 4 blocks in the result section: 105, 124, 120 and 119 (see page 13). We didn’t find the information of sample size calculation in the original design. Because post hoc power calculation is no recommended by most research guidelines, we decided no to pursue this type of analysis. However, we addressed this limitation in the method section when we described the original design to readers. (see page 7)

2. Cost attribute: Table 2a and 2b show inexplicable coefficients for the price attribute. The middle level ($250) is least relevant. Some more explanation for the reader would help. How could this be explained?
RESPONSE: We agree that the results of cost attribute in Tables 2a and 2b are difficult to explain. These could be the result of the violation of the linear utility function assumption. We have added a paragraph in the discuss section to discuss the reasons. (see page 17,18)

3. Sensitivity: Also the coefficients show that respondents favor low sensitivity. Some more explanation for the reader would help. How could this be explained?
RESPONSE: This might be because most participants in our study preferred FOBT method, which has relatively low sensitivity comparing to other CRC screening modalities. It seems that participants were not willing to trade other test characteristics such as stool sample with the test accuracy. (see page 16)

4. Page 16: Conclusions are, that small within-cluster correlation can result in the use of a simple logistic regression. High within-cluster correlation is seen as a reason for sensitivity analysis. Would it be possible to give more recommendations on how a systematic approach would look like (as a final summary of the work done in this paper)?
RESPONSE: We have added the recommendations of investigating intra-class correlation before fitting statistical model. We recommended simple logistic or probit model when little within-cluster correlation present. We also added the recommendations for reporting the results from sensitivity analysis. It is impossible to answer which model should be used from this empirical study. Further simulation studies are needed on comparing estimates to the true parameters. (see page 18, 19)
Discretionary Revisions

1. Report results: The complexity of the many statistical models makes it difficult to read and understand. For the clarity of the method section and the result section it might be beneficial to drop statistical methods (eg. 1 or 2 from the binary outcomes).
RESPONSE: we have noticed the number of models discussed in this paper. However, because all models are commonly used in DCE analysis, we decided to keep them.

2. Number of statistical methods:
Page 4: McFadden was introducing the basics of the statistical methods in 1974. As far as I know two papers of Louvriere and Woodworth (1983) and Louvrier and Hensher (1982) introduced discrete choice experiments in Marketing.
RESPONSE: we have revised the sentence regarding the literature and added Louvriere and Woodworth (1983) and Louvrier and Hensher (1982)’s papers as reference. (see page 4)

3. Page 8: Random utility theory usually refers to indirect utility functions (indirect utility due to price and income).
RESPONSE: Some papers used the term of “random utility theory (RUT)” and some papers used the term of “random utility model”. We followed the definition in the Louviere, Flynn and Carson’s paper (2010), used “random utility theory”. The corresponding paper has been added as the reference (see page 7 and reference 25).

4. The conclusion could refer to findings in light of other studies.
RESPONSE: We compared our findings regarding to participants preference of CRC screening task to previous studies. More details have been added in the discussion section. (see page 16)

Comments from celia berchi

1. Methods
   a - The alternative statistical techniques are well described but insufficiently compared each other. It would be helpful for readers if the models were classified according to their theoretical relevance or frequency of use in existing literature.
Response: We classified statistical models according to the type of outcomes: binary, nominal and bivariiate. After incorporating with the available methods of dealing with intra-class correlation, each simple model derived to several forms. To allow readers tracking the messages easier, we added a paragraph just under the sub-title of “statistical methods” to outline how the discussions of the models were organized. (see page 8)

   b – I think that the third and fourth paragraphs of the discussion comprise background elements and should be set in the methods.
RESPONSE: We agree that the two paragraphs containing some background information, but it is also a part of discussion on what should be done on dealing with intra-class correlation. Particularly, the fourth paragraph is related to our conclusion on the need of a
simulation study. Therefore, we decided to keep them in the discussion section. We have added a sentence to link our study to the other studies. (see page 16)

2. Results
There are two much tables and figures. I suggest deleting Figure 2 and Table 1 and its associate paragraph which are not essential.
RESPONSE: For Table 1, the reviewers provided conflicting advice, and therefore we decided to keep this table. For Figure 2, because a flow chart of the study sample is recommended by several guidelines for reporting clinical studies including CONSORT Statement, STROBE Statement, and etc, and therefore we decided to keep this chart for readers who may be interested in knowing how the patient selection was done for study.

3. Discussion-conclusion
a - Results should be further compared with those of previous theoretical and/or empirical studies if possible.
RESPONSE: We compared our findings regarding to the participants’ preference of CRC screening test to previous studies. More details have been added in the discuss section (see page 16)

b - The conclusion suffers from lack of a clear indication of how to best deal with model selection issue when sensitivity analysis shows significant differences according to the model used.
RESPONSE: This is a good point. However, our study may not be suited to address this issue. A simulation study would be the most appropriate approach to provide insight about model selection. We have provided some description to this effect in the conclusions (see page 18, 19).

Minor issues not for publication
1 - There are some spelling mistakes and typographical errors to correct.
RESPONSE: The paper has been proofread independently by a third party. We have made every attempt to correct all spelling mistakes and typos.

Methods:
- 2nd paragraph: there is an extra bracket to DCBE
RESPONSE: corrected (see page 6)
- 3rd paragraph: I suggest replacing orthogonally by orthogonality?
RESPONSE: replaced (see page 7)
- 4th paragraph (RUT): delete “s” from “components”
RESPONSE: deleted (see page 8)
- last paragraph: delete “s” from “each attributes”
RESPONSE: deleted (see page 12)
- Please define the acronym “GEE”
RESPONSE: GEE was defined in the first paragraph of “Statistical methods” subsection where it was first introduced.

Results:
- 4th paragraph: delete “the results” from “the results the # coefficients”
RESPONSE: deleted (see page 14)
- last paragraph: add “s” to “six attribute”
RESPONSE: suggestion adopted (see page 15)
Discussion:
- 1st paragraph, line 5: add “cancer” to “colorectal screening test”
RESPONSE: Suggestion adopted (see page 15)

2 - Authors should check the references Methods, paragraph entitled “multinomial logistic model”: check the reference
RESPONSE: all references have been checked and all errors have been corrected.

References:
- Reference 5 is incomplete
RESPONSE: checked and corrected (after revisions it became reference 7)
- Reference 21: correct the authors names which are Zwerina and Kuhfeld
RESPONSE: corrected (after revisions it became reference 23)
- Reference 33 should be revised
RESPONSE: revised (after revisions it became reference 38)