Reviewer's report

Title: Consistency between stated and revealed preferences: a Discrete Choice Experiment and a behavioral experiment on vaccination behavior compared.

Version: 2
Date: 13 December 2013

Reviewer: Fran E Carroll

Reviewer's report:

This is an interesting paper exploring the consistency of stated and revealed preferences for Hep-B vaccination of young babies.

I am not a modeling expert and as such am not able to comment on the statistical model component of the manuscript (Analyses; Step 1: computing individual probability on data DCE)

Major Compulsory Revisions:

BACKGROUND

(1) Please clarify the meaning of the 7th and 8th sentences in this section (beginning "Because either vaccination..." and "The current study was conducted...") - I was not clear on the relevance of the sentence regarding acceptance of vaccination against HPV in this context. Please also clarify why it is ..."no longer straightforward that people accept all vaccines..."

METHODS

The structure of the methods section needs some alteration - it appears a little mixed in the order of the techniques used.

(2a) I would rather see the description of the attributes and levels first in this section. This enables the reader to gain an understanding of the factors deemed important for inclusion and where they came from.

(2b) On that note, there is no mention of the levels used and how they came to be selected - please describe how and why the levels for each attribute were chosen.

(2c) The authors state that the attributes were based on 'focus group interviews, literature review and expert interviews'. There is no subsequent mention of these 'expert interviews' and it appears from the following text that the attributes were in fact all gained from previous literature rather than any research conducted by the authors themselves for the purpose of this study. Please clarify where the attributes came from and as stated before, justify the selection of the associated levels.

Final paragraph of 'Design of the choices' section

(3a) This paragraph is difficult to understand - it is not clear why the 144 scenarios (or profiles) were not paired with another using the foldover or
modulo-arithmetic technique, thereby giving 144 choice sets to present to participants. The design could then have been blocked so participants received one of several versions of the questionnaire.

(3b) Also, it is not clear how the authors got ‘2000 unique choice sets’ or why the questionnaire was then blocked so participants received only 4 choice sets each (i.e. does this mean there were 500 versions of the questionnaire?)

‘Design of DCE, introduction of scenarios’ section

(4) The first paragraph where the authors are referring to participants’ certainty is not very easy to understand. My eventual interpretation having looked at Appendix 1 is that participants’ were asked how certain they were that they would have their child vaccinated given the scenario they had just selected. However, this is not overly clear and clarity would ensure readers’ do not interpret this as referring to the participants’ certainty about their overall preference for/against vaccines in general.

ANALYSES

‘Step 1: computing individual probability on data DCE’

(5) Clarity needed on the ‘opt out’ identified from participants’ responses to the certainty question. Why did the authors select 6 as the cut-off? Why not just provide an ‘opt-out’ question in the DCE questionnaire in the first place?

(6) As stated above, further clarity on the certainty of the preference in this first paragraph would be beneficial.

‘Step 2: comparing the predictions and actual behavior’

(7) It is not clear where the 0.6 probabilities have come from in the second paragraph – please provide further explanation for this.

RESULTS

(8) The author’s should check the respondent numbers they are quoting. If 10 participants were excluded, the total should be 896 (906-10)? It should also be explained why participants were excluded for having missing demographic data, but not for having missing DCE data.

(9) The second paragraph makes references to the demographics of the sample, but there is no table provided to give this information. Authors refer to religion affecting participants’ opinion on vaccination, but do not explain how that is the case? Similarly for ‘positive previous experience of vaccination’ – does this refer to their own children having vaccinations previously? Other peoples’ children? More clarity needed here – and a table with demographic data would be useful.

‘Results step 1: the model on DCE data’

(10) Second and third paragraphs in this section make reference to the beta coefficients, but I am not clear the authors have interpreted them adequately in comparison to each other. For example, there is reference to a coefficient of 0.20 demonstrating that particular attribute as ‘less important’ than other attributes, despite there being others with the same coefficient. A coefficient of 0.38 for
‘knowledge of risk of side effects’ is described as ‘negligible’, yet in the same sentence, the same coefficient (0.38) for ‘knowledge of a priori risk of hepatitis B’ is described as showing that attribute as ‘important.’ Please check your interpretation and description of your results here.

DISCUSSION
(11) The penultimate paragraph in this section (where the sentence begins: “From the questionnaire, we have evidence”) – this appears to read as new results being introduced in the discussion. Unsure this is the best place for this – either include these as results or better incorporate the message into the discussion.
(12) The final paragraph of this section is not clear at all. This paragraph needs rewriting to ensure an understandable message for the reader.

Minor Essential Revisions:
(13) Authors should check throughout the manuscript the correct spelling of ‘multinomial’ – it is often referred to as ‘multinominal.’

BACKGROUND
(14) Second paragraph; fifth sentence: should read: “At the time of the study, the hepatitis B vaccine had not yet been introduced...”
(15) Penultimate paragraph; sentence beginning: “At the time of the study (Feb-Apr 2011)” – there is a mention of ‘risk groups’ – this has not been stated before. Please clarify what constitutes a ‘risk group’ in this context.
(16) The following sentence beginning: “The hepatitis B vaccine is part...” – use the word ‘consisting’, not ‘existing’.
(17) The final sentence of this paragraph should read: “The aim of this paper is to study the congruence between parents’ stated and revealed preferences for whether to vaccinate...”

METHODS
Design of DCE, introduction of the scenarios
(18) Second paragraph, first sentence beginning “The DCE was part of...” – it is not clear what the meaning of the second part of this sentence is (“...and the individual DCE attributes on the respondent level.”) Please reword this.

Study population
(19) Again, there is reference to ‘risk groups’ in this paragraph – as before, please clarify what this means.
(20) Final sentence should read: “The parents did not receive specific additional information on hepatitis B, which is in line with current standards within the NIP.”
(21) The final sentence refers to ‘elaborate description of the process of data collection’ in another paper. Although elaborate detail is not needed here, it would be useful to give a brief description on how data were collected – postal questionnaire presumably? Local/national area? All parents/first time parents
only? These kind of details would give context to the study.

Observation of revealed preferences

(22) Second sentence of this section refers to the DCE and the questionnaire – were there two separate documents distributed or are these one and the same? Please clarify this.

(23) Second paragraph beginning “those parents who agreed...” – please clarify the 5 or 6-valent option. Could parents decline a vaccination altogether? Is the standard 5-valent vaccine offered to all parents anyway? The authors state that those who opted for the 6-valent were offered a voucher – was this also the case for those opting for the 5-valent? A little more clarification here about the process for these 333 respondents would be useful.

ANALYSES

(24) First sentence – authors state that the answers given were recoded. Why was this? And what were they recoded into?

(25) Second sentence should read: “The answer of how certain the respondent was that their child would be vaccinated after making...”

(26) I don’t feel that Table 1’s details of the coding and variable names adds anything or is necessary. I recommend removing this information from this table and just using it to describe the attributes and levels. Amend manuscript text appropriately.

(27) The second paragraph after the first equation should read: “The model in the equation...”

(28) Sentence beginning ...“the effect of the vaccination is known...” should go on to read “...the National Institute of Health advises use of the vaccine...”

(30) Does ‘p’ refer to the probability per choice set? Does ‘V’ refer to observed choice? Please clarify your use of these terms

‘Step 2: comparing the predictions with actual behaviour’

(31) Add “…given the scenario presented” to the end of the first sentence so that it reads: “The above results in the estimated probability that the respondent will have his/her child vaccinated against hepatitis B given the scenario presented”

RESULTS

(32) Authors should check the formatting of large numbers in the text – use commas if necessary to separate e.g. 7,132

(33) Remove the text “…with a minimum age of 16 and a maximum of 48 years.” And just replace with “(range 16-48 years)”

(34) 81% of the respondents “were” female

(35) Check spacing at the beginning of the third paragraph (beginning “333 respondents...”). This sentence should read: “…in the second portion of the study where they were given the opportunity...”

(36) The penultimate sentence of this section really belongs to the Methods
Results step 1: the model on DCE data

(37) Authors refer to 3,566 choice sets, whereas earlier in the manuscript there were 2000? Please clarify and check this.

(38) Third sentence – replace ‘de’ with ‘the’. Spelling of standard ‘deviations’.

(39) Authors should check the text referring to the second column of Table 2. It appears there is some duplication in the description in the third and fourth sentences of this paragraph.

(40) Second paragraph, third sentence should read: “Knowledge of the risk of....and knowledge of the a priori risk...”

(41) Authors should check the beta coefficient quoted for an acquaintance reporting hospitalisation – in the text is shown as -0.21, but table refers to -0.20.

Results step 2: comparing the stated and revealed preferences

(42) Authors state the average probability of the sample to choose a vaccination is 73%. This is a newly introduced figure – where has it come from?

(43) Fourth sentence should read: “...computing how many of the sample (n=252) have a probability...”

(44) The second paragraph in this section beginning “The terms ‘positive predictive value...” should be moved to the end of the results section before the Discussion.

(45) I am unsure of the need for Table 4 given that the data is described in the text. I would recommend removing this table as it currently is a duplication of what is already in the text.

(46) Final sentence before Discussion should read “...implying that the probability parameter of the stated preferences is a poor predictor...”

CONCLUSIONS AND DISCUSSIONS

Overall, the structure of this section could be improved. Focussing on ‘main results’, ‘findings from previous research’, ‘strengths and weaknesses’ and ‘implications for practice/future research’ could offer a clearer way to lay out this section.

(47) I recommend altering the heading of this section simply to ‘Discussion’ – there is already a ‘Conclusions’ heading at the end of the manuscript.

(48) First sentence should be reworded: “The aim of this paper was to study the predictive value of the DCE-technique for comparing stated with revealed preferences among parents deciding whether to vaccinate their newborn child against hepatitis B.”

(49) Second sentence mentions ‘young’ parents – this has not been described elsewhere in the manuscript – as listed previously, more detail about the demographics of the sample would be helpful and would provide justification for this comment.
(50) Final sentence of first paragraph should read: “in 78% of the respondents, their actual behaviour was predicted correctly by the results of the stated preference DCE.”

(51) First/second sentences of second paragraph should read: “…it does not provide a definite answer with regard to the external validity of DCEs. The behaviour in this study was very asymmetric…”

(52) Remove “than vaccination behaviour in newborn children” from end of that sentence.

(53) Final sentence of that paragraph should read: “…could for instance be tested in HPV vaccination, where acceptance levels are around 50%…”

(54) Authors should clarify the meaning of the third paragraph – the message here is not clear at all.

(55) In the fourth paragraph, the authors refer to “comparing the two groups”, but it is not clear which two groups are being referred to.

(56) Penultimate sentence of fourth paragraph should read: “…us to study revealed preferences were more highly educated than the other group.”

(57) The authors go on to describe 13% and 11% of parents who did and did not get the vaccine respectively – please clarify how these figures were achieved.

CONCLUSIONS
(58) Final sentence should read: “To further study the relationship between stated…”

TABLES
- Please add a table detailing the sociodemographic characteristics of the sample
- Table 1: No need for the variable names and coding. Just use this table to show the attributes and associated levels.
- Table 2: Please address the formatting of this table – it is difficult to easily see the attributes and levels here
- Table 4: Do not feel this table is necessary. Recommend removing and just referring to results in the text.

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

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:
I declare that I have no competing interests.