

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Prioritising patients for bariatric surgery: building public preferences from a discrete choice experiment into public policy
<b>AUTHORS</b>	Whitty, Jennifer; Ratcliffe, Julie; Kendall, Elizabeth; Burton, Paul; Wilson, Andrew; Littlejohns, Peter; Harris, Paul; Krinks, Rachael; Scuffham, Paul

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Gema Frühbeck Dept. of Endocrinology & Nutrition, Clinica Universidad de Navarra CIBEROBN IdiSNA Pamplona, Spain
<b>REVIEW RETURNED</b>	15-Jul-2015

<b>GENERAL COMMENTS</b>	<p>The authors address a very interesting topic, namely the prioritization of patients for bariatric surgery based on public preferences from a discrete chose experiment performed in a supposedly representative sample of Australian adults. This represents a very timely issue given the obesity epidemic and the difficulty in offering the surgical therapy to all eligible candidates. At the same time, it is a complex issue were evidence gained so far needs to be clearly explained in order to be able to perform well-informed decisions.</p> <p>In my opinion, the manuscript would benefit from addressing the following relevant aspects:</p> <p>- Methods:</p> <p>Bariatric surgery is a general concept that involves diverse procedureds that differ in their effectiveness and complexity. In Australia the laparoscopic adjustable gastric banding is very popular and one of the most frequently performed techniques (as opposed to what happens in the USA and Europe where gastric bypass and sleeve gastrectomy are more frequently applied). Did individuals have information on the different techniques or was “bariatric surgery” used as a global term? Did the individuals have any clue to the different types of surgical approaches and their effectiveness? This is important given that the diverse techniques have their own ideal applications according to patient characteristics.</p> <p>How was effectiveness, i.e. chance of maintaining weight loss estimated? There is quite a huge inter-individual variability in the response to weight loss following bariatric surgery (which is not easy to predict). This is a limitation that needs to be clarified in the manuscript since the “chance of maintaining weight loss” that is</p>
-------------------------	---

used as one of the criteria to make the decision in reality is almost impossible to predict and, thus, an information that we do not have in real life. In fact, in relation to success prediction family support of the patient undergoing bariatric surgery and the presence of a favourable environment at home and work have been shown to be more important than other biological variables such as a high BMI.

Page 11, lines 22-36: the authors mention that quotas were used to ensure that the target sample was representative by age and gender for each State. Was this also applied to educational level and socioeconomic status? Table 2 shows differences in some of these aspects but it is not stated if they reached statistical significance. The text in page 12 (lines 50-55) also highlights the lower educational level, higher degree of unemployment, lower household income and lower health status of the study sample as compared to the general population. This needs to be clarified.

- Results:

Page 13, lines 25-28: what is referred to as Table 2 should be Table 3, while what is indicated as Table 3 should be Table 4; this needs to be corrected also in page 27 and 28, respectively.

Page 14, lines 30-31: what is referred to as Table 4 should be Table 5; this needs to be corrected also in page 29.

- Discussion:

In my view, this section needs to be expanded to comment in more detail the findings, in particular, to contextualize them with current clinical practice and evidence obtained so far in long-term outcomes of bariatric surgery.

For instance, currently more functional eligibility and success criteria are being considered (recently reviewed in Frühbeck G. Bariatric and metabolic surgery: a shift in eligibility and success criteria. *Nat Rev Endocrinol* 2015, 9 June 2015 [ePub ahead of print]) and this needs to be mentioned since the inadequacy of BMI as a primary criterion for the surgical selection of candidates has been highlighted.

Page 15, lines 55-59: the authors mention that the importance of identifying criteria to prioritise surgery were related to choosing those that were “the most obese, those with co-morbidities, and those most likely to benefit clinically”. However, the identification of those patients that will benefit the most does not depend directly on the degree of BMI. The same is applicable to comorbidities. In fact, it is debatable whether these patients are the ones who will benefit the most from the surgical approach. In patients with advanced evolution of their comorbidities and where end-organ damage is already present, the clinical benefit from surgery may be much more reduced than in a lower degree of obesity and comorbidity development. In a public-funded health-care system, with limited access and resources, prioritising patients with the greatest need is prudent. Moreover, prioritizing patients with the highest BMI and comorbidity for bariatric surgery will imply selection of individuals with the greatest anaesthetic risk to undergo a complex surgical procedure. On the other hand, increased severity of some comorbidities might result in a reduced likelihood of improvements occurring once end-organ damage or established end-stage

	<p>disabilities have been reached. Thus, prioritizing these patients might mean that an opportunity to use bariatric surgery as a more preventative measure in patients who could really attain full recovery is missed. In addition, given the lack of knowledge on the natural history of the diverse obesity states, it is challenging to identify which patients will indeed progress to higher functional impairment and which patients will remain stable for a given BMI.</p> <p>The recognition that the benefits of bariatric surgery extend well beyond weight loss provided the clinical rationale for the emergence of metabolic surgery. This type of surgery is primarily intended to treat T2DM and is offered to patients with obesity who have a lower BMI than those who are eligible for bariatric surgery (Fried, M. et al. Interdisciplinary European guidelines on metabolic and bariatric surgery. Obesity Surgery 2014;24: 42-55). Thus, during the last years the eligibility criteria have been extended, which suggests that the surgical approach can be advisable in patients with a BMI &lt;35 kg/m<sup>2</sup> and inadequately controlled T2DM. Moreover, other anthropometric measures different to the BMI better identify patients at a high cardiometabolic risk and, therefore, should be preferably considered (Gomez-Ambrosi, J. et al. Cardiometabolic profile related to body adiposity identifies patients eligible for bariatric surgery more accurately than BMI. Obesity Surgery 2015[22 January 2015 ePub ahead of print]). The limitations of the study to address these aspects need to be incorporated to the Discussion.</p> <p>Do the authors have information about the real knowledge of their participants about obesity, its associated comorbidities, drivers and obstacles to loose weight? In order to make informed decisions individuals should have a good understanding or at least enough knowledge. At least in Europe, it has been shown by a survey performed in 7 different countries (see survey at the European Association for the Study of Obesity EASO website) that there are very frequent misconceptions about the risks that obesity poses to individuals as well as contemplating obesity as a gateway to ill health (Frühbeck, G. et al. Obesity: the gateway to ill health - an EASO position statement on a rising public health, clinical and scientific challenge in Europe. Obes Facts 2013;6: 117-20).</p> <p>While involving the general population in public policy decisions may be interesting, in order to provide a balanced view, I would recommend the authors to comment and include the above-mentioned aspects together with the suggested references, which are very relevant to adequate eligibility of patients to maximize outcomes and cost-effectiveness.</p>
--	--

<b>REVIEWER</b>	Seda Erdem University of Stirling, UK
<b>REVIEW RETURNED</b>	20-Jul-2015

<b>GENERAL COMMENTS</b>	<p>This paper investigates public preferences for priority setting in the context of bariatric surgery using a Discrete Choice Experiment (DCE). Overall, the paper is not well motivated and it suffers from the lack of explanations on key points, including the following:</p> <ol style="list-style-type: none"> <li>1. Why is it important to look at the prioritisation from a public perspective, but not from patients' perspective? How can the results help policy-makers?</li> </ol>
-------------------------	--

2. Why is the main focus “average preferences”, whereas the paper also includes results from a latent class model? I find this quite strange and confusing. I think the LC model provides more insight into the topic and it should be the main model, and the paper should be re-orientated around it. Another important thing is that the LC model should be explicitly formulated and/or mentioned clearly, and the result table should be provided!

3. How do the results change with individuals’ characteristics? – in particular with their health status, whether they think they have a healthy lifestyle, and personal experience, etc.

4. Comorbidity issue needs more thoughts and discussions. Perhaps, at the end of the results section, a scenario analysis that presents choice predictions would help understand the prioritisation under comorbidity.

5. The LC model should be adjusted for scale (error variance) differences. There have been several papers on this (e.g., Magidson & Vermunt, 2007; Salisbury & Feinberg & by Fiebig et al., 2010). So, the authors need to re-estimate the models to take differences in scale into account to allow us to draw proper conclusions about the results.

Without the above points being dealt with, it is (literally) not possible to evaluate this paper properly.

Below are my comments to specific points:

1. P.9 line 8-26 briefly mentions about the importance of public perspective. This part needs more elaboration as it is one of the main motivations of the paper and currently it is poorly written. This part is also lacking a literature review (e.g., what have others done? Are there contradicting views? Etc). This section also seems to be lacking some references. Please include them.

2. P.10 line 51-56: Having a choice task repeated (19th choice task) does not necessarily measure the internal choice consistency. For example, if the choice task includes alternatives having dominant attribute-levels, it may be more obvious to have the same “dominant” (or inferior) choices. Then you’d expect consistency in responses to these two questions. However, if the choice task includes alternatives that are equally attractive to respondents (I.e., utility balance), then it may not result in the same choice in repeated questions. Therefore, excluding “unmatched” responses, therefore, is not necessarily the correct strategy. This should be addressed in the paper.

3. P. 11, Sample: This section should include the characteristics of the sample, not the results section. As it says, the results section is for the results of the analyses.

4. P. 11, Data Analysis: Line 46-51 needs an explanation on “why” this is the case.

5. P.16 Line 48, what are the other studies on this issue (age as a decision criterion), in addition to the (17)? The paper is citing this source quite commonly without giving other evidence/perspectives from the literature. This needs to be addressed.

	<p>6. P.16 line 34: This claim does not seem to be true (just a simple search on Google provides some research done on bariatric surgery and other weight loss programs using DCEs). For example, Reference (11) in the manuscript. So, this needs to be revisited.</p> <p>7. P.18 Line 26-35: I am not clear with that. Is this based on LC specification?</p> <p>8. Table 3: Explain how you calculated the “priority weights” in the main manuscript.</p> <p>9. Table 4: I might have missed it in the main body of the text. But if it is not mentioned, please do mention. It would also be helpful to indicate where to insert the tables and figures in the manuscript.</p> <p>10. The DCE in the paper uses seven attributes with varying levels. Have you tested respondents’ understanding of the scenarios?</p> <p>11. The DCE tasks ask respondents to evaluate the scenarios that apply to “others”, rather than themselves. How would this differ if they had to evaluate the scenarios for themselves? This could have been mentioned in the section where the paper promotes prioritisation from public perspectives. As more than half of the sample is overweight and obese, it’d be useful to analyse the results with respect to respondents’ characteristics, in particular, their BMI.</p> <p>12. Please present the results of LC model. You can present both MNL and LC model results in one table and discuss your results under both model specifications.</p> <p>13. Figure 2 is very small. Please enlarge it.</p> <p>14. It seems like “maintaining healthy life” is a dominant level. Did you check how many times this level appears in choice tasks included in your experimental design?</p>
--	---

### VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

We agree with the Reviewer that this is an important and timely topic.

**1. Methods: Information on different bariatric surgical techniques**

Participants were provided with general information in layperson terms on the overall effectiveness of bariatric surgery and also in the variation of expected outcomes in the survey material (Refer to attached survey that has been provided to the BMJ Open as online supplementary material). To keep the survey materials as understandable for the public as possible, we did not discriminate between the different surgical techniques, as this was outside of the scope of this survey. That is, we aimed to consider who should be prioritised for access to bariatric surgery as an interventional approach; we did not attempt to address the question of which type of surgery should be provided.

**2. Methods: Description of effectiveness attribute (chance of maintaining weight loss) is a limitation given the difficulty in predicting this outcome.**

Thank you for raising this point. We do not believe it will impede application of the priority weights, but

have added it as a potential limitation (p22):

“The implementation of the findings may be limited since application of the priority weights requires an ability to predict the category into which each patient fits for each of the attributes, before their treatment. This may be challenging for the attribute “chance of maintaining weight loss”, since effectiveness is difficult to predict a priori. Nevertheless, estimates of effectiveness are available in the international literature. Alternatively, if distinguishing likely effectiveness between potential patients is considered to be unreliable, this attribute could be excluded from the priority estimates for all potential patients.”

### 3. Methods: Quotas and representativeness of sample

Quotas were set by age and gender only; it was considered that setting quotas by additional characteristics would make numerous smaller target quota cells that might be challenging to complete. We are unable to statistically test representativeness, as our population norms were extracted from a variety of different sources, not all of which have the sample size for the population. In any case, we would argue that statistically testing the representativeness of the sample against the population is redundant. The proportion of participants in our sample clearly differed from the population on the characteristics highlighted – given the size of the population, these differences are likely to reach statistical significance. In fact, quite a small difference might reach statistical significance but yet be relatively meaningless in terms of its implication. Therefore, we have described the main differences that might indicate a non-representativeness of our sample based on the size of difference. We had previously discussed the limitations related to the representativeness of the sample, but have now strengthened this discussion in view of the Reviewer’s comments, by adding (p21):

“The sample differed descriptively from the Australian population on a number of characteristics (education level, employment status, household income and health status). Of these, only health status and education level were found to be associated with preference class in secondary latent class analyses (Supplementary material). Therefore, it is not known to what extent this may have impacted the representativeness of the overall sample preferences on these characteristics.”

### 4. Results: Table numbers

Apologies, the Table numbers have now been corrected.

5. Discussion: Section needs to be expanded to comment in more detail the findings, in particular, to contextualize them with current clinical practice and evidence obtained so far in long-term outcomes of bariatric surgery.

Thank you for providing your detailed comments and suggestions. We have amended our discussion accordingly, and in particular:

Page 18: We have amended the wording to reflect the points raised by the Reviewer:

“The indicated importance of these criteria, particularly a desire to prioritise the most severely obese and those with co-morbidities, are largely consistent with previous studies that suggest social preferences in other health priority setting contexts would prioritise those who are most severely affected by the condition being treated. 1 They are also largely consistent with existing obesity guidelines, which recommend the use of BMI and/or comorbidities as criteria for surgery.4,6”

We have added two new paragraphs discussing the issues raised by the Reviewer within the context of our study:

“The choice tasks given to respondents in this study were of necessity somewhat simplified to enable their administration to laypersons in a survey format. However, the clinical decision-making context around the appropriateness of bariatric surgery for specific individuals, and who would benefit most, is somewhat more complex. For example, the benefits of surgery may extend beyond weight loss and include metabolic outcomes, leading to the emergence of “metabolic surgery” which has differing therapeutic goals and a lower BMI criterion threshold, with some effects occurring independent of weight loss.<sup>2 3</sup> Thus, the potential criteria used in this DCE may not be the only criteria of clinical relevance for selection of individuals for surgery. The inadequacy of BMI as a primary clinical criterion for selection for surgery and potential of other clinical criteria to augment selection has been highlighted.<sup>2 4</sup> Further, those with a higher BMI and comorbidities such as diabetes, obstructive sleep apnoea and cardiac disease, may be at greater risk of adverse events from surgery.<sup>2</sup> Thus, the optimal selection of candidates for bariatric surgery from a clinical perspective so as to balance the benefits and risks of surgery is not straight forward. Nevertheless, despite these potential limitations, the current study focussed on prioritising individuals for surgery assuming surgery was considered to be clinically appropriate. Respondents were instructed in the survey to imagine that each of the potential surgery recipients had been clinically assessed as in equal need of surgery to manage their obesity. Thus, any “real world” clinical consideration around the benefit of surgery was held constant in each hypothetical choice and should not have impacted the hypothetical decisions.”

“To support their capacity to make decisions in the DCE, respondents were provided with some basic information on obesity, its consequences, and its management at the start of the survey. However, obesity and its management is a complex issue and although the pilot study suggested the survey was easy to understand, respondent understanding of the obesity information was not tested in the main survey. Further studies investigating public opinion for prioritising bariatric surgery using a Citizens’ Jury, which represents a deliberative approach in which participants are informed and can challenge experts before making recommendations on the issues, are planned as part of the parent study within which this DCE is undertaken. 32”

Reviewer: 2

General points

1. Why is it important to look at the prioritisation from a public perspective, but not from patients’ perspective? How can the results help policy-makers?

This paper takes the normative position that it is the preferences of the public, rather than patients per se, that are relevant for informing priority setting decisions. This is generally the accepted approach in health economics (in the context of priority setting). We have expanded the paragraph addressing this in the discussion to do so more explicitly, by adding (p20):

“This study takes the normative position that it is the preferences of the public, rather than individuals with a specific condition, that are relevant for informing priority setting decisions. Moreover, for health services funded by taxation of the public, the public are a key stakeholder in how those funds are used. Therefore, the public’s perspective is important for allocating funds to specific services. Whilst this is an accepted approach in health economics in the context of priority setting,<sup>17</sup> the exploratory latent class analysis in this study suggests that the preferences of an obese population around priorities for bariatric surgery may differ to those of the general public.”

2. Why is the main focus “average preferences”, whereas the paper also includes results from a latent class model? I think the LC model provides more insight into the topic and it should be the main model, and the paper should be re-orientated around it. Another important thing is that the LC model should be explicitly formulated and/or mentioned clearly, and the result table should be provided!

Given our focus on overall average public preferences (refer point 1 above for justification), a latent class model (which explores heterogeneity and does not present an “average” preference across the sample on which to base priority weights) is not an appropriate primary analytic model. However, we agree with the reviewer that the latent class model does provide some additional insights and indeed we have referred to these in the paper. Therefore, to retain the tight focus of the paper on public preferences whilst still being able to refer to the insights provided by the latent class model, we made the decision to put the latent class model in supplementary material. The results table of the latent class model was also provided in the supplementary material. If the Editor considers it more appropriate we would be happy to move the latent class model presented in the supplementary appendix to be in the main text, alongside the primary MNL model. However, we feel the simplicity and focus of the paper is better maintained by having it as additional material.

3. How do the results change with individuals’ characteristics? – in particular with their health status, whether they think they have a healthy lifestyle, and personal experience, etc.

As discussed above, this is outside the primary focus of the paper on average public preferences for policy purposes. Only BMI, History of weight loss surgery, AQL utility score, and education level were significantly associated with class membership at the 5% level in the final latent class model ( $p \leq 0.05$ ). This is described in the Supplementary material.

4. Comorbidity issue needs more thoughts and discussions. Perhaps, at the end of the results section, a scenario analysis would help understand the prioritisation under comorbidity.

Comorbidity is included in the case scenarios in Table 5, but was termed Obesity-related conditions – we have now highlighted this. Apologies for any confusion.

5. The LC model should be adjusted for scale (error variance) differences. There have been several papers on this (e.g., Magidson & Vermunt, 2007; Salisbury & Feinberg & by Fiebig et al., 2010). So, the authors need to re-estimate the models to take differences in scale into account to allow us to draw proper conclusions about the results.

Any differences in scale of the error variance become pertinent when the analyst wishes either to combine two or more datasets, or to draw conclusions in directly comparing the coefficients/findings from two or more models (i.e. asking whether and differences observed in coefficients are caused by true difference in preferences, differences in scale, or both). Neither of these of these contexts applies here. Additionally, the papers conclusions are predominantly based on the MNL model, and not the latent class model (which was exploratory), as discussed elsewhere.

Below are my comments to specific points:

1. P.9 line 8-26: More elaboration and referencing to the literature required on the importance of the public perspective.

We have now substantially revised the third paragraph of the Introduction and added further references, as suggested.

2. P.10 line 51-56: Repeated choice task (19th choice task) does not necessarily measure the internal

choice consistency.

In our study, there was not a clearly identifiable dominant choice alternative for any of the choice sets. For example, some individuals may prefer to treat 20 year olds, some may prefer 35 year olds, some may prefer 50 year olds, some may not be driven by the age criterion. Whilst we agree with the reviewer that profiles would be harder to distinguish when they are close to being utility balanced than when this is not the case, we still consider a repeat choice to be an indicator for internal consistency. We have however changed our manuscript to describe this as a less strict internal consistency “indicator” rather than “test”.

We did not delete respondents who were considered not to be internally consistency (Lancsar & Louviere have argued strongly against deleting such respondents 5). We decided a priori not to include the 19th choice set for any respondent (consistent or not), as the 19th choice set was a repeat of an earlier choice set, and was not part of the design. We have clarified this in the methods (p12).

3. P. 11, Sample: This (Methods) section should include the characteristics of the sample, not the results section.

With respect, we disagree with the reviewer. This comment would appear to relate to a field specific difference in conventional reporting approaches. We are aware that in some fields (e.g. economics) sample characteristics may be conventionally reported in the methods section. In the health field it is more usual to report the sampling methods in the methods section, and the final achieved sample characteristic in the results section, as we have done.

4. P. 11, Data Analysis: Line 46-51 needs an explanation on “why” this is the case.

The focus on average public preferences has been justified elsewhere (refer our response to Reviewer 2 General Points 1 and 2).

5. P.16 Line 48: References missing for age as a decision criterion, in addition to the (17).

We had cited a systematic review that refers to the previous studies of relevance to the “age” criterion and discussed their inconsistencies. However, at the Reviewer’s request, we have now added citations to the original studies in addition to the review, and also to some relevant recently published studies (Erdem & Thompson 2014 6; Whitty et al., 2014 7).

6. P.16 line 34: This claim does not seem to be true ... For example, Reference (11) in the manuscript.

Reference 11 in the original manuscript (Scuffham et al., 2014) is a protocol (methods) paper for a body of research within which the current study sits. It does not present any results. Our extensive searching on Google and Pubmed does not reveal any previous preference-based study (e.g. using DCE) to elicit public priorities for bariatric surgery.

- Our own previous work presents background papers and protocols on this issue, not results.8 9
- Roux et al., 2004 and Muhlbacher & Bethge 2013 elicited patient own preferences (not public priorities) for weight loss programmes (not surgery) 10 11
- Erdem & Thompson 2014 elicited public preferences for healthcare innovation (described generically, not bariatric surgery specifically) for target populations that included people with obesity – thus they explore priorities for managing obesity relative to other health conditions 6
- Ryan et al., 2015 elicited preferences for lifestyle interventions 12

We have made minor modifications to clarify the text to state:

“Nevertheless, we are not aware of any previous preference study that has attempted to quantify priorities for bariatric surgery from a public perspective,…”

7. P.18 Line 26-35: I am not clear with that. Is this based on LC specification?

This sentence has been removed in response to other feedback, so no amendment is required.

8. Table 3: Explain how you calculated the “priority weights” in the main manuscript.

The explanation of this in the methods section has now been expanded, by adding:

“The marginal rate of substitution (and therefore priority weight) for each criterion was estimated by dividing the marginal utility for that criterion level by the marginal utility for effectiveness. For example, the weight for prioritising an individual with “very severe obesity” rather than “obesity” is equal to the difference between the coefficients from the MNL model between these two attribute levels, divided by the coefficient for a one percentage increase in the chance of maintaining weight loss (i.e. priority weight =  $((0.28751 - (-0.30626)) / 0.01530) = 38.80850$ ; from results presented in Tables 3 and 4; calculations performed prior to rounding of decimal places).”

9. Table 4 not mentioned in the main text.

Apologies, there was an error in Table numbering, which has now been corrected. Table 4 is now mentioned in the main text.

10. The DCE in the paper uses seven attributes with varying levels. Have you tested respondents' understanding of the scenarios?

A pilot study (n=20) which was undertaken face to face and with a qualitative component indicated that respondents understood and could complete all choice scenarios. This is explained in the manuscript (p12):

“Extensive pilot testing was undertaken to confirm the face validity of the instrument, prior to main data collection. This involved face-to-face completion of the survey by an adult convenience sample (n=20), with qualitative exploration of understanding of the instrument along with estimation of a preliminary choice model.”

11. The DCE tasks ask respondents to evaluate the scenarios that apply to “others”, rather than themselves. How would this differ if they had to evaluate the scenarios for themselves? This could have been mentioned in the section where the paper promotes prioritisation from public perspectives. As more than half of the sample is overweight and obese, it'd be useful to analyse the results with respect to respondents' characteristics, in particular, their BMI.

Asking respondents to value and/or prioritise health interventions in a social decision-making context as is the case here has been widely applied in the literature (see for example Whitty et al., The Patient 2014, which was ref 17 in the original manuscript submission). Asking respondents to prioritise bariatric surgery between patients would be non-sensical if applied to themselves. I think the reviewer may possibly be referring to the idea that people with obesity may have different preferences for prioritising surgery to those without; whilst outside the primary focus of the paper, this is nevertheless explored in the paper via the latent class model in the supplementary material.

12. Please present the results of LC model. You can present both MNL and LC model results in one table and discuss your results under both model specifications.

Refer to our response to this Reviewer, particularly to general points 1 and 2. The LC model has been presented in the Supplementary material. If the Editor considers it appropriate to present the LC model in the main paper alongside the MNL model despite this being an exploratory analysis in a paper that has a focus on average public preferences (predicted by the MNL model), we are happy to do so.

13. Please enlarge Figure 2.

Figure 2 has now been enlarged.

14. It seems like “maintaining healthy life” is a dominant level. Did you check how many times this level appears in choice tasks included in your experimental design?

The design was level balanced – both levels for the “maintaining a healthy lifestyle” attribute appeared an equal number of times in the choice tasks.

#### References used in response

1. Whitty JA, Lancsar E, Rixon K, et al. A systematic review of stated preference studies reporting public preferences for healthcare priority setting. *The Patient* 2014;7(4):365-86.
2. Fruhbeck G. Bariatric and metabolic surgery: a shift in eligibility and success criteria. *Nat Rev Endocrinol* 2015;11(8):465-77.
3. Fried M, Yumuk V, Oppert JM, et al. Interdisciplinary European guidelines on metabolic and bariatric surgery. *Obes Surg* 2014;24(1):42-55.
4. Gómez-Ambrosi J, Moncada R, Víctor V, et al. Cardiometabolic Profile Related to Body Adiposity Identifies Patients Eligible for Bariatric Surgery More Accurately than BMI. *Obes Surg* 2015;25:1594-603.
5. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Econ* 2006;15(8):797-811.
6. Erdem S, Thompson C. Prioritising health service innovation investments using public preferences: a discrete choice experiment. *BMC Health Serv Res* 2014;14:360.
7. Whitty JA, Ratcliffe J, Chen G, et al. Australian Public Preferences for the Funding of New Health Technologies: A Comparison of Discrete Choice and Profile Case Best-Worst Scaling Methods. *Med Decis Making* 2014;34:638–54.
8. Scuffham PA, Ratcliffe J, Kendall E, et al. Engaging the public in healthcare decision-making: quantifying preferences for healthcare through citizens' juries. *BMJ open* 2014;4(5):e005437.
9. Whitty JA, Burton P, Kendall E, et al. Harnessing the potential to quantify public preferences for healthcare priorities through citizens' juries. *International journal of health policy and management* 2014;3(2):57-62.
10. Muhlbacher A, Bethge S. Preferences of overweight and obese patients for weight loss programmes: a discrete-choice experiment. *Int J Integr Care* 2013;13:e034.
11. Roux L, Ubach C, Donaldson C, et al. Valuing the benefits of weight loss programs: an application of the discrete choice experiment. *Obes Res* 2004;12(8):1342-51.
12. Ryan M, Yi D, Avenell A, et al. Gaining pounds by losing pounds: preferences for lifestyle interventions to reduce obesity. *Health Econ Policy Law* 2015;10(2):161-82.

## VERSION 2 – REVIEW

<b>REVIEWER</b>	Gema Frühbeck Clinica Univ de Navarra, CIBEROBN, IdiSNA, Pamplona, Spain
<b>REVIEW RETURNED</b>	28-Aug-2015

<b>GENERAL COMMENTS</b>	The reviewer completed the checklist but made no further comments.
-------------------------	--