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

<table>
<thead>
<tr>
<th>TITLE (PROVISIONAL)</th>
<th>Financial incentives to discontinue long-term benzodiazepine use: a discrete choice experiment investigating patient preferences and willingness to participate</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTHORS</td>
<td>Marti, Joachim; Bachhuber, Marcus; Feingold, Jordyn; Meads, David; Richards, Michael; Hennessy, S</td>
</tr>
</tbody>
</table>

### VERSION 1 - REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Eric Finkelstein</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Duke NUS Medical School, Singapore</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>14-Feb-2017</td>
</tr>
</tbody>
</table>

| GENERAL COMMENTS | The primary objectives of this study are to investigate the acceptability of financial incentives for initiating a medically supervised benzodiazepine discontinuation program and to identify program features that influence the willingness to participate. Although the low response rate greatly limits the generalizability of the results, using a discrete choice experiment to look at factors that influence uptake is likely of interest to some readers. However, several issues will need to be addressed before the manuscript is ready for publication. The authors should motivate specific hypotheses in the introduction and then revisit them in the Discussion section, including whether or not results are consistent with theory and/or with what other researchers have found. For example, is there any reason to believe participants would prefer lotteries over cash? How about a program that offers partial rewards for partial success? I would think that would be preferred. I am puzzled why this study would be exempt from IRB as it involves human subjects. Can the authors elaborate on that. I would also encourage the authors to look at best practice modelling for DCEs and few recent publications in health services journals. The models employed here are not up to current standard. Few DCE researchers today use conditional logit. Mixed logit or latent class models generally perform superior in goodness of fit. At a minimum the authors should justify why they chose such a model over more flexible alternatives. Also, given their design it would seem a bivariate probit model would be an appropriate choice given that would allow for simultaneously modelling both tradeoffs and uptake. Another approach is to use an alternative specific constant in your model to allow for quantifying the minimum utility required for someone to choose a program over opting out. I would also like to see additional validity tests if possible. Did the authors test for attribute dominance? For the 30% who are 'out of the market' do their tradeoffs really mean anything? If they are truly different then a conditional logit model is unlikely to be the right model. |

---

**Note:** The table above contains all the necessary information from the document. The content includes the title, authors, and general comments from the review. The review highlights the primary objectives of the study, the limitations due to low response rate, and suggestions for improvement in methodology and hypothesis testing.
I have also never seen uptake modelled the way the authors have done it. Typically it is modelled in the conditional logit framework with interactions between participant characteristics and attribute levels. I believe that is a superior approach. In the discussion section, I do not think it is fair to say that ‘as only 28.4% of eligible patients agreed to participate and returned the survey, the real world enrollment rate among eligible patients might be closer to 14% (28% × 50%) to 22% (28% × 80%)’ I think we don’t know. Also, the comment “In other words, are the benefits to patients in terms of avoided health care costs and improved quality of life from discontinuing benzodiazepines large enough to justify a monetary investment?” needs to be couched in terms of an appropriate perspective. If it is a government, the first question is why they would do this at all. If not government, then who? And why? Is it to save money, to be cost effective? Something else? I also did not see the relevance of the Tannenbaum cite or the statement.

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Frans Zitman</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUMC, The Netherlands</td>
<td></td>
</tr>
</tbody>
</table>

| REVIEW RETURNED | 20-Feb-2017 |

<table>
<thead>
<tr>
<th>GENERAL COMMENTS</th>
</tr>
</thead>
</table>
| This paper studies the usefulness of financial incentives as help in the discontinuation of benzodiazepines. This an interesting, and indeed, new approach. As the authors state correctly, it investigates the willingness of patients to make use of this possibility, not its execution in clinical practice. I have some questions as can be seen in the review checklist above. I address them here in the order in which they occur in the paper:p4 44-57: I am surprised that it was allowed to identify potential subjects by the researchers although they are not their treating physicians. Is not that violating privacy rules?
| p4 48-52 You want a discontinuation program that is useful specifically in real long-term use, for many years. In this respect the inclusion criteria are rather lenient. Three prescriptions the last 12 months is not the typical long-term user. Fortunately, the duration of use in most participants was much longer.
| p4-5: The participants were asked to fill out the questionnaire at home? If so, how the investigators made sure that the approached person also was the one who filled out the questionnaire? Was it allowed to consult family members and friends?
| p 5, line 4: What is the difference between version A and B of the study questionnaire? What is the relation, if any, with the choice sets (p5, line 53-54)
| p7-58 to p8-6. The OR’s and CI’s in the text do not correspond with those in table 4.
| Should not it be discussed how specific the results may be for (the health system of) the USA?
| Financial incentives may increase the willingness to participate in a dose-reduction program. However, what will be the long-term effects? I think, this aspect should be addressed in the discussion. |
Reviewers:1

“The authors should motivate specific hypotheses in the introduction and then revisit them in the Discussion section, including whether or not results are consistent with theory and/or with what other researchers have found. For example, is there any reason to believe participants would prefer lotteries over cash? How about a program that offers partial rewards for partial success? I would think that would be preferred.”

Thank you for the suggestion. We now discuss this in more details in the introduction, especially on page 4.

“I am puzzled why this study would be exempt from IRB as it involves human subjects. Can the authors elaborate on that.”

The study has been exempt from IRB. We have made the corresponding letter available to the Editor.

The study was considered under exemption category 2 (i.e. research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects’ responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation)

“I would also encourage the authors to look at best practice modelling for DCEs and few recent publications in health services journals. The models employed here are not up to current standard. Few DCE researchers today use conditional logit. Mixed logit or latent class models generally perform superior in goodness of fit. At a minimum the authors should justify why they chose such a model over more flexible alternatives. Also, given their design it would seem a bivariate probit model would be an appropriate choice given that would allow for simultaneously modelling both tradeoffs and uptake. Another approach is to use an alternative specific constant in your model to allow for quantifying the minimum utility required for someone to choose a program over opting out.”

Many thanks for these very useful suggestions. We have decided to go beyond conditional logit and now estimate a latent class logit model to explore heterogeneity. Also, our models now include an ASC for opting-out.

“I would also like to see additional validity tests if possible. Did the authors test for attribute dominance?
For the 30% who are ‘out of the market’ do their tradeoffs really mean anything? If they are truly different then a conditional logit model is unlikely to be the right model. I have also never seen uptake modelled the way the authors have done it. Typically it is modelled in the conditional logit framework with interactions between participant characteristics and attribute levels. I believe that is a superior approach.”

Thanks for this suggestion. As mentioned above, we estimated a latent class logit mode and identified 2 classes of respondents. Class 1 respondents, who represent about 35% of the sample, are characterized as “non-traders”. Our LC model results also show the impact of individual characteristics on probability to be a “non-trader” (i.e. Class 1). Regarding validity test, we have checked for attribute dominance and have identified 3 respondents who always chose the alternative with the highest incentive amount.
“In the discussion section, I do not think it is fair to say that ‘as only 28.4% of eligible patients agreed to participate and returned the survey, the real world enrollment rate among eligible patients might be closer to 14% (28% × 50%) to 22% (28% × 80%).’ I think we don’t know.”

We have modified the text accordingly.

“Also, the comment “In other words, are the benefits to patients in terms of avoided health care costs and improved quality of life from discontinuing benzodiazepines large enough to justify a monetary investment?” needs to be couched in terms of an appropriate perspective. If it is a government, the first question is why they would do this at all. If not government, then who? And why? Is it to save money, to be cost effective? Something else? I also did not see the relevance of the Tannenbaum cite or the statement.”

We now clarify that this should be understood in a health system perspective. The Tannenbaum study is used as an example of cost-effectiveness study that took long-term costs of a similar product into account – and not only short-term clinical effectiveness – and whose results might justify earlier discontinuation.

Reviewer: 2

“p4 44-57: I am surprised that it was allowed to identify potential subjects by the researchers although they are not their treating physicians. Is not that violating privacy rules?”

We have attached the IRB exemption letter for this study.

“p4 48-52 You want a discontinuation program that is useful specifically in real long-term use, for many years. In this respect the inclusion criteria are rather lenient. Three prescriptions the last 12 months is not the typical long-term user. Fortunately, the duration of use in most participants was much longer.”

Thanks for pointing this out. We modeled the inclusion criteria for this survey on the inclusion criteria that might be used for an actual trial. Three months is a widely-used definition of long-term use, and an intervention might be more effective among those who are early in the course of chronic use. Nevertheless, as the reviewer points out, when using such a duration criterion, the eligible population consists mostly of people who have been using benzodiazepines for much longer than three months.

“p4-5: The participants were asked to fill out the questionnaire at home? If so, how the investigators made sure that the approached person also was the one who filled out the questionnaire? Was it allowed to consult family members and friends?”

Yes, respondents were asked to fill the questionnaire at home. We cannot be certain that the approached person actually filled out the questionnaire and now acknowledge this limitation in the paper.

“p 5, line 4: What is the difference between version A and B of the study questionnaire? What is the relation, if any, with the choice sets (p5, line 53-54)”

The two blocks are part of the experimental design. A block design was preferred to limit the number of choice set each respondent had to complete. Respondents were randomly allocated to one of the two blocks.
“p7-58 to p8-6. The OR's and CI's in the text do not correspond with those in table 4.”

As the analysis has changed, this is no longer relevant.

Should not it be discussed how specific the results may be for (the health system of) the USA?

Financial incentives may increase the willingness to participate in a dose-reduction program. However, what will be the long-term effects? I think, this aspect should be addressed in the discussion.

Thank you for the suggestion. This is now addressed in the discussion section.

VERSION 2 – REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Eric Finkelstein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duke-NUS Medical School Singapore</td>
<td></td>
</tr>
</tbody>
</table>

| REVIEW RETURNED  | 08-May-2017                             |

<table>
<thead>
<tr>
<th>GENERAL COMMENTS</th>
<th>This version of the manuscript is much improved. I appreciate the authors willingness to address my and the other reviewer's concerns. I have no more concerns with the methods/results but have a few more comments with the framing of this study that I would like the authors to consider. These concerns are as follows:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) In general I find that many studies overplay the angle of behavioral economics, when the interventions are in fact, just applications of standard economic theory. PP2 in the introduction is such a case. Giving people money to make a behavior change works regardless of present bias, repeated choices, or any other failure of rationality. This is why supply curves slope up. I would suggest reframing the second pp to focus on standard economic theory and then the third pp can speak to how behavioral economics may help make standard incentive programs better by taking advantage of where people may get it wrong, in terms of non-optimizing behavior.</td>
<td></td>
</tr>
<tr>
<td>2) The additions on the top of page 4 is a bit unclear. Standard economic theory would say that individuals maximize the expected value of discounted future utility. If individuals are risk averse, they still may gain higher utility by opting for the certainty equivalent. This is quite different from what you wrote and from the idea of maximizing expected payouts. So again one can generate your hypothesis without having to call on behavioral economics. There may be additional behavioral econ reasons, but you should first fix these additions as they are incorrect as written.</td>
<td></td>
</tr>
<tr>
<td>3) On page 4 line 9, you state that ‘surprisingly little is known’. What exactly is known? Later you state ‘almost no work exploring….’. What is there? In the introduction or discussion section you should compare your results to others who have studied similar questions, even if not identical.</td>
<td></td>
</tr>
<tr>
<td>4) The term ‘optimal’ is unclear in ‘optimal incentive structure’. I am sure what is optimal from an enrollment perspective is not optimal from an effectiveness or cost effectiveness perspective. Need to clarify what you mean here.</td>
<td></td>
</tr>
<tr>
<td>5) ‘may not necessarily be perceived as unhealthy by patients’. Is this relevant?</td>
<td></td>
</tr>
<tr>
<td>6) You used one DCE, not plural.</td>
<td></td>
</tr>
<tr>
<td>7) I am not sure that the latent class membership allows for including...</td>
<td></td>
</tr>
</tbody>
</table>
individual characteristics directly in the model. I do not think that is correct. I think what it allows is to model the likelihood of being in a particular class as a function of individual characteristics. Please check this and clarify the language as needed.

8) Might be friends or relatives who help them take the survey.

9) Is the goal of a health system perspective to reduce health care costs? That may be true but I would focus more on cost effectiveness, which I think is more consistent with how most interventions would be evaluated.

REVIEWER
Frans G. Zitman
Department of Psychiatry
Leiden University Medical Center (LUMC)
Leiden
The Netherlands

REVIEW RETURNED
06-May-2017

GENERAL COMMENTS
To facilitate the interpretation of my comments, I give them below as a reaction to the comments the authors gave to my earlier comments:

1) p4 44-57 pages and lines in original version): I am surprised that it was allowed to identify potential subjects by the researchers although they are not their treating physicians. Is not that violating privacy rules?“

We have attached the IRB exemption letter for this study.

Apart from providing an exemption letter the authors should add an explanation in the text. If both reviewers ask questions on this topic, it is to be expected that readers of the paper will also be puzzled with respect to this subject.

2) p4 48-52 You want a discontinuation program that is useful specifically in real long-term use, for many years. In this respect the inclusion criteria are rather lenient. Three prescriptions the last 12 months is not the typical long-term user. Fortunately, the duration of use in most participants was much longer.”

Thanks for pointing this out. We modeled the inclusion criteria for this survey on the inclusion criteria that might be used for an actual trial. Three months is a widely-used definition of long-term use, and an intervention might be more effective among those who are early in the course of chronic use. Nevertheless, as the reviewer points out, when using such a duration criterion, the eligible population consists mostly of people who have been using benzodiazepines for much longer than three months.

No further comment.

3) p4-5: The participants were asked to fill out the questionnaire at home? If so, how the investigators made sure that the approached person also was the one who filled out the questionnaire? Was it allowed to consult family members and friends?“

Yes, respondents were asked to fill the questionnaire at home. We cannot be certain that the approached person actually filled out the questionnaire and now acknowledge this limitation in the paper.
4) “p 5, line 4: What is the difference between version A and B of the study questionnaire? What is the relation, if any, with the choice sets (p5, line 53-54)”

The two blocks are part of the experimental design. A block design was preferred to limit the number of choice set each respondent had to complete. Respondents were randomly allocated to one of the two blocks.

When reading the design of the choice experiment it still is not clear to me where the A and B version consists of. Is it the two blocks of 6 choice sets?

5) p7-58 to p8-6. The OR’s and CI’s in the text do not correspond with those in table 4.”

As the analysis has changed, this is no longer relevant.

6) Should not it be discussed how specific the results may be for (the health system of) the USA?

OK

7) Financial incentives may increase the willingness to participate in a dose-reduction program. However, what will be the long-term effects? I think, this aspect should be addressed in the discussion.

OK

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2

Apart from providing an exemption letter the authors should add an explanation in the text. If both reviewers ask questions on this topic, it is to be expected that readers of the paper will also be puzzled with respect to this subject.

Thanks for the suggestion. We have added an explanation on page 5.

When reading the design of the choice experiment it still is not clear to me where the A and B version consists of. Is it the two blocks of 6 choice sets?

Thanks. We have clarified this in the Data Collection and Design of the choice experiment sections.

Reviewer: 1

This version of the manuscript is much improved. I appreciate the authors willingness to address my and the other reviewer’s concerns. I have no more concerns with the methods/results but have a few more comments with the framing of this study that I would like the authors to consider. These concerns are as follows:

1) In general I find that many studies overplay the angle of behavioral economics, when the
interventions are in fact, just applications of standard economic theory. PP2 in the introduction is such a case. Giving people money to make a behavior change works regardless of present bias, repeated choices, or any other failure of rationality. This is why supply curves slope up. I would suggest reframing the second pp to focus on standard economic theory and then the third pp can speak to how behavioral economics may help make standard incentive programs better by taking advantage of where people may get it wrong, in terms of non-optimizing behavior.

Good point. We have modified PP2 and 3 accordingly.

2) The additions on the top of page 4 is a bit unclear. Standard economic theory would say that individuals maximize the expected value of discounted future utility. If individuals are risk averse, they still may gain higher utility by opting for the certainty equivalent. This is quite different from what you wrote and from the idea of maximizing expected payouts. So again one can generate your hypothesis without having to call on behavioral economics. There may be additional behavioral econ reasons, but you should first fix these additions as they are incorrect as written.

Thanks for these suggestions, we have modified the text accordingly (p. 3-4).

3) On page 4 line 9, you state that ‘surprisingly little is known’. What exactly is known? Later you state ‘almost no work exploring….’. What is there? In the introduction or discussion section you should compare your results to others who have studied similar questions, even if not identical.

We have discussed additional recent work in the introduction (page 4).

4) The term ‘optimal’ is unclear in ‘optimal incentive structure’. I am sure what is optimal from an enrolment perspective is not optimal from an effectiveness or cost effectiveness perspective. Need to clarify what you mean here.

We have amended the text to clarify.

5) ‘may not necessarily be perceived as unhealthy by patients’. Is this relevant?

We have dropped this sentence.

6) You used one DCE, not plural.

Ok.

7) I am not sure that the latent class membership allows for including individual characteristics directly in the model. I do not think that is correct. I think what it allows is to model the likelihood of being in a particular class as a function of individual characteristics. Please check this and clarify the language as needed.

Thanks. We have clarified the language.

8) Might be friends or relatives who help them take the survey.

This is now acknowledged.

9) Is the goal of a health system perspective to reduce health care costs? That may be true but I would focus more on cost effectiveness, which I think is more consistent with how most interventions
would be evaluated.

We agree with this. In the conclusion, our statement “are the benefits to patients in terms of avoided health care costs and improved health outcomes from discontinuing benzodiazepines large enough to justify a monetary investment?” refers to cost-effectiveness and not only to reduction in costs.

**VERSION 3 – REVIEW**

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Frans Zitman</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Leiden University Medical Center</td>
</tr>
<tr>
<td></td>
<td>Leiden</td>
</tr>
<tr>
<td></td>
<td>The Netherlands</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>28-May-2017</td>
</tr>
</tbody>
</table>

**GENERAL COMMENTS**

The changes made based on my second round of comments are sufficient.