

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

TITLE (PROVISIONAL)	Patient trade-offs between continuity and access in primary care interprofessional teaching clinics in Canada: a cross-sectional survey using discrete choice experiment
AUTHORS	Oliver, Doug; Deal, Ken; Howard, Michelle; Qian, Helen; Agarwal, Gina; Guenter, Dale

VERSION 1 - REVIEW

REVIEWER	Matthew Quaife London School of Hygiene and Tropical Medicine, United Kingdom
REVIEW RETURNED	18-Jun-2018

GENERAL COMMENTS	<p>Primary care patient trade-offs between continuity and access in interprofessional teaching clinics: a cross-sectional survey using discrete choice experiment</p> <p>This study seeks to elicit patient preferences for different elements of primary care health service provision, specifically in teaching clinics in Canada. The study found that patients valued minimising time to appointment the most, alongside being able to book their own physician.</p> <p>In general, this is an interesting and informative study. I have indicated some additional information is required before publication, particularly on the design and reporting of the results and background characteristics of the individuals and clinics included. The introduction gives a clear overview of the rationale and use of the study.</p> <p>The authors comment on attribute selection through a FGD then expert shortlisting. Was this FGD published elsewhere? It would be useful to know, if not, the range of topics in the topic guide or how varied discussion was about things which were not included in the final shortlist. How many potential wtributes were in the list which was reduced to six? Some more detail on this part of the process would be good.</p> <p>Is reference 31 the most up to date contribution on the restrictions of the MNL? The comparative benefits and drawbacks of HB over MNL could be discussed more, perhaps in relation to other options such as mixed logit or latent class models. I would have anticipated seeing some equations to be estiamted or a clearer description of how this methods works, despite the broad readership of the journal this is important to understand.</p> <p>In addition, the methods used for simulation are not described in the methods section, please add these and note why the two simulation profiles were chosen</p>
-------------------------	--

	<p>In the inclusion criteria, was “able to read English well enough to complete the questionnaire” assessed?</p> <p>The ISPOR guidelines cited support the inclusion of unidimensional attributes – web v phone booking and length of time is a composite attribute. Please comment on how this could have affected responses and the efficiency of the design.</p> <p>Why was an analysis not conducted interacting the health states with the individual attribute levels? This heterogeneity will drive the variation in the simulations, but it is not apparent what attributes are driving this. Additionally, are you able to recover SEs/CIs for these percentages?</p> <p>Presumably, this is a pooled analysis of web based and paper DCEs? Did the authors test for differences in scale or part worths between these two responses? Unless I have missed it, the number of people responding by either method is not included</p> <p>I would have expected a table of descriptive statistics of the individuals who took part in the survey – the strobe checklist notes this in p7 par 3, but this page reference is different to that in the manuscript I have seen. The paragraph describing this in the results section does not give a sufficient overview of this information. Where possible, a comparison of key variables (e.g. age, SES, ethnicity) e.g. with census data, or just those registered to each surgery would be useful to assess generalisability</p> <p>On a related note, sampling was a convenience sample from two clinics. What was the proportion from each clinic and via online/paper survey from each clinic? How were the two clinics selected?</p> <p>Smaller comment</p> <p>p. 12 29 onwards, I am not sure capitalisation of the attributes makes reading this easier</p>
--	--

REVIEWER	<p>Mattijs Lambooj National Institute of Public health and the Environment (RIVM), the Netherlands</p>
REVIEW RETURNED	<p>16-Aug-2018</p>

GENERAL COMMENTS	<p>Review of “ Primary care patient trade-offs between continuity and access in interprofessional teaching clinics: a cross-sectional survey using discrete choice experiment”.</p> <p>The paper presents the application of a DCE amongst patients of two locations of hospitals, aiming to quantify the preferences of these patients concerning their visits to medical staff. The researchers were especially interested in the relative importance of seeing a familiar health care professional versus timely or quick access to a health care professional. The most important conclusion is that patients prefer faster access to a health care provider over speaking to a familiar provider.</p> <p>But when the medical condition does not require urgent attention, people are willing to wait to meet their own doctor.</p> <p>Overall impression: the information of the results is presented unclearly, it is therefore hard to judge the results and its implications.</p>
-------------------------	--

	<p>The paper could benefit from some editorial work, more links with literature can be made and (the presentation of) results can be more elaborated, since I found it hard to thoroughly understand all implications of the results due to unclear presentation of the results. In the discussion section, more links with societal implications can be made.</p> <p>I got the impression that the authors do not fully exploit all results of the study. I think that the paper has the potential to gain in strength if the results are described in more detail and the conclusions better reflect the results.</p> <p>Introduction</p> <p>What struck me in the first paragraph is that it appears to consist of five lines of text seemingly unrelated, summing up the practical relevance, scientific relevance, introduce an IT system to make appointments, ending in a statement that the implementation of the IT-system in Canada appears not to have conducted optimally, since the access of same day appointments of Canadian patients is worse than in other countries.</p> <p>This does not invite the reader to continue to read and I think that the paragraph would gain in strength if it would introduce the topic of the paper, rather than what is currently stated.</p> <p>The actual aim of the paper is stated much later e.g. here “To understand what relative value patients place on various aspects of the clinical encounter, the “discrete choice experiment (DCE) method” has been utilized extensively in health care research” and here “The objective of this study was to use DCE in an inter-professional academic setting to evaluate patients’ preferences for various attributes of access to their family practice clinic including preferences regarding staff physician, trainee physician and allied health professionals, and method of booking (telephone versus online) across different scenarios of reasons for seeking an appointment.”</p> <p>Even though the IT system mentioned in the first paragraph may be a tool to plan appointments, it is just instrumental to the main question of the paper. I would therefore skip the emphasis on the IT-tool and focus on the trade offs that patients were asked to make in the DCE.</p> <p>Methods</p> <p>Development of the questionnaire is done properly.</p> <p>Question: why did the researchers use a fractional factorial design instead of an efficient design?</p> <p>The authors make a rough estimation of needed sample size “a sample size of 300 to 500 subjects is generally considered adequate”. I too believe that these number per DCE are adequate. However, the researchers add a dimension to their DCE: they add 6 labels (clinical scenarios) to the choice sets and as such divide the sample into six. This means that overall 430 respondents were included. This means that per clinical scenario $430/6 \sim 72$ respondents were included. Could the researchers please explain to</p>
--	--

the reader that this yielded sufficient power for the analyses, even though it is much smaller than the lower boundary of the sample size stated to be needed?

Results

Here it becomes extremely difficult for me to judge the paper. The main results apparently are displayed in a figure 2. I do find a caption with figure to and below the document I see a figure containing box plots, that may be related to the caption.

Why did the authors chose not to present a table with parameters of the attributes? Presenting 6 models for six of the clinical scenarios? That would give the reader the opportunity to thoroughly study all results and distil whether they would yield similar conclusions as the authors of the paper.

The figure “importance of the attributes” is most informative to me of the results included in the paper. However, this figure only presents the mean of all six scenarios. I would find it interesting to see this figure for all six scenarios.

Conclusion

The authors state “results lend support to the notion that improved and timely access to primary care seems to be the leading priority for patients as wel”. This only holds for the overall results. In the final figure, I would conclude that if the medical condition does not require quick response, patients are willing to wait longer to see their own doctor.

This final result is more nuanced, and provides more opportunities to tailor the provided care to the needs of patients, given their situation.

The previous point may also explain why I do not understand the statement “These results help to demonstrate that an advanced access booking model does in fact target what many patients value most, across a number of health states – that is, timely access to their primary care team”.

I do not conclude, with the information that I have, that patients always want timely access to health care, as they appear to be willing to wait in some health states. Second, I would imagine that the patient does not really care which booking system is used by the organisation to schedule appointments, just as long as it gets done properly. So I wonder why the authors make this strong statement that is only partly supported by their results.

Minor points

Abstract

Results end with “were not strongly related to the clinical scenario.” This needs earlier introduction.

Conclusion ends with “ [...]advanced access booking systems seem to target issues that patients value highly.” This is an example of an IT-system popping up in the text that appears to play an important role in the daily organisation of the place where data collection took place, but the reader gets too little information to fully grasp the importance and implications of the implementation of the IT-system at hand.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Matthew Quaife

1. The authors comment on attribute selection through a FGD then expert shortlisting. Was this FGD published elsewhere? It would be useful to know, if not, the range of topics in the topic guide or how varied discussion was about things which were not included in the final shortlist. How many potential attributes were in the list which was reduced to six? Some more detail on this part of the process would be good.

Response: Thank you for this suggestion. The FGD results have not been published, however we have added further information as requested. The relevant section of the methods now reads as: (pg 7-8)

We provided scenarios to practice team members in the focus group, to stimulate discussion about attributes. The scenarios reflected access attributes from the literature and that the research team felt were relevant to primary care: speed of appointment, appointment with regular clinician who knows the patient, type of provider (physician, nurse, resident). These attributes were validated by the focus group as very important to include. Participants also suggested an attribute relating to number of phone calls needed to reach the practice for an appointment which was felt to be a lower priority and not included in the DCE. We next described four scenarios that might affect patients' access preferences (new minor symptom, new urgent symptom, anxiety issues, routine check-up) and asked for input on these and for additional scenarios that would be relevant in the context of a family practice teaching centre. The additional level of online booking was added to the appointment booking method attribute, and for type of provider, the levels of family doctor, training doctor (resident) and nurse/nurse practitioner were recommended. The wording of attribute levels was also refined through discussions and expert judgment of the research team (including two physicians involved in implementing open access).

2. Is reference 31 the most up to date contribution on the restrictions of the MNL? The comparative benefits and drawbacks of HB over MNL could be discussed more, perhaps in relation to other options such as mixed logit or latent class models. I would have anticipated seeing some equations to be estimated or a clearer description of how this method works, despite the broad readership of the journal this is important to understand.

RESPONSE: As requested, we have added further rationale for the HB estimation method, as follows (pg 12, para 2):

...Hierarchical Bayes estimation (HB) of preference coefficients was chosen over multinomial logit (MNL) since HB largely overcomes the independence of irrelevant alternatives (IIA) issue of MNL[22] and provides preference coefficients for each individual respondent. Huber and Train[23] found that part-worth utility estimates produced by HB and mixed logit were not significantly different. HB uses the Metropolis Hastings Algorithm, a type of Markov chain Monte Carlo iterative procedure that analyzes individual choices at the lower model level using MNL and then analyzes the aggregated data at the upper level using multivariate normal methods. The initial burn-in phase was run with 20,000 iterations with 20,000 additional iterations used for estimating the part-worth utilities.

3. In addition, the methods used for simulation are not described in the methods section, please add these and note why the two simulation profiles were chosen

RESPONSE: As requested, we have added the following information in the methods section: (pg 13, top)

Simulations were conducted in Sawtooth Software SMRT using the randomized first choice simulation method. That method was chosen because it attempts to mimic the noise inherent in human decisions by automatically adding appropriate error to the levels of the attributes included in the simulation scenarios, plus an overall error term. We chose the simulation profiles to contain the three most important attributes to ensure a good split in shares-of-preferences and to provide a range of shares across the six scenarios.

4. In the inclusion criteria, was “able to read English well enough to complete the questionnaire” assessed?

RESPONSE: We did not conduct a formal assessment of the participants’ ability to read English. We have added this information pg 11 para 3.

5. The ISPOR guidelines cited support the inclusion of unidimensional attributes – web v phone booking and length of time is a composite attribute. Please comment on how this could have affected responses and the efficiency of the design.

RESPONSE: We appreciate the reviewer raising this point. Attribute 1 is a compound attribute. However, we had no desire to separately estimate the booking method (internet or phone) from the booking wait time (‘right now’, ‘less than 1 minute’ and ‘1 to 10 minutes’). Also, separating the appointment booking method from the time-to-book would create a situation where prohibitions would need to be introduced, thereby reducing the statistical quality of the design. There are only two formats for booking an appointment, internet and phone. Since internet is immediate, if the second part of this compound attribute were time (‘right now’, ‘less than 1 minute’ and ‘1 to 10 minutes’), only ‘right now’ could be applied to the internet and prohibitions would need to be used to ensure that the other two levels were not combined with the internet booking method. We have added a paragraph addressing this issue in the Discussion (page

19, para 3)

...The appointment booking method is a compound attribute of method and time to book the appointment. We had no desire to separately estimate the booking method (internet or phone) from the booking wait time (‘right now’, ‘less than 1 minute’ and ‘1 to 10 minutes’). Separating the appointment booking method from the time-to-book would have created a situation where prohibitions would have been needed to avoid unrealistic combinations of method and time, thereby reducing the statistical quality of the design. While some may desire to estimate each univariate attribute separately, this compound attribute best supported this research.

6. Why was an analysis not conducted interacting the health states with the individual attribute levels? This heterogeneity will drive the variation in the simulations, but it is not apparent what attributes are driving this. Additionally, are you able to recover SEs/CIs for these percentages?

RESPONSE: Thank you for this suggestion. We have now provided the PWUs for each randomized health scenario and in aggregate and the confidence intervals, as a supplementary file because the table is wide. (Supplementary File 1). We have also provided a new Figure 2 showing the relative importance of attributes by health state.

7. Presumably, this is a pooled analysis of web based and paper DCEs? Did the authors test for differences in scale or part worths between these two responses? Unless I have missed it, the number of people responding by either method is not included I would have expected a table of descriptive statistics

of the individuals who took part in the survey – the strobe checklist notes this in p7 par 3, but this page reference is different to that in the manuscript I have seen. The paragraph describing this in the results section does not give a sufficient overview of this information. Where possible, a comparison of key variables (e.g. age, SES, ethnicity) e.g. with census data, or just those registered to each surgery would be useful to assess generalisability

RESPONSE: Thank you for this suggestion. We have clarified the survey administration methods which we recognize were previously unclear (and corrected an error in the sample size in clinic A). The section now reads as (pg 11, para 4):

The questionnaire was created electronically (web-based) and self-administered via computer-assisted personal interviews (CAPI). Recruitment was done by a research assistant who approached patients in the waiting room (clinic A, n=53) while waiting to see their health care providers, and through emails to patients who had email addresses on file (clinic B, n=377). The research assistant initiated the CAPI questionnaire on her laptop for patients recruited in the waiting room of Clinic A and was available for questions.

We have also created a new Table 2 (page 16) which shows the demographic characteristics of the participants from the two recruitment methods. We feel it may not be useful to compare these patients to census data since we know the general population is somewhat different than people enrolled to a family practice.

Regarding the differences between the two groups of respondents, among the 18 attribute levels, 14 were not significantly different. The four attribute levels that were significantly different were in the two least impactful attributes, clinic waiting time and convenience of the clinic appointment. Since those two attributes have quite minor affects on the overall utilities, the two sources of subject responses were combined.

9. On a related note, sampling was a convenience sample from two clinics. What was the proportion from each clinic and via online/paper survey from each clinic? How were the two clinics selected?

Response: We have clarified the methods used in each clinic in the above comment. We have added the suggested information about selection of the clinics. (pg 11, para 2)

A convenience sample of patients was recruited in 2012 from two inter-professional family practice teaching clinics with which the researchers are affiliated, in Hamilton, Ontario, Canada.

10. Smaller comment

p. 12 | 29 onwards, I am not sure capitalisation of the attributes makes reading this easier

Response: Thank you for this comment, we have removed capitalisation.

Reviewer: 2

Reviewer Name: Mattijs Lambooj

Institution and Country: National Institute of Public health and the Environment (RIVM), the Netherlands

The paper presents the application of a DCE amongst patients of two locations of hospitals, aiming to quantify the preferences of these patients concerning their visits to medical staff. The researchers were especially interested in the relative importance of seeing a familiar health care professional versus timely or quick access to a health care professional. The most important conclusion is that patients prefer faster access to a health care provider over speaking to a familiar provider.

But when the medical condition does not require urgent attention, people are willing to wait to meet their own doctor.

Overall impression: the information of the results is presented unclearly, it is therefore hard to judge the results and its implications.

11. The paper could benefit from some editorial work, more links with literature can be made and (the presentation of) results can be more elaborated, since I found it hard to thoroughly understand all implications of the results due to unclear presentation of the results.

12. In the discussion section, more links with societal implications can be made.

13. I got the impression that the authors do not fully exploit all results of the study. I think that the paper has the potential to gain in strength if the results are described in more detail and the conclusions better reflect the results.

Response: Thank you for these comments. We believe we have addressed them through the specific questions pertaining to Results and Discussion below.

Introduction

14. What struck me in the first paragraph is that it appears to consist of five lines of text seemingly unrelated, summing up the practical relevance, scientific relevance, introduce an IT system to make appointments, ending in a statement that the implementation of the IT-system in Canada appears not to have conducted optimally, since the access of same day appoints of Canadian patients is worse than in other countries. This does not invite the reader to continue to read and I think that the paragraph would gain in strength if it would introduce the topic of the paper, rather than what is currently stated. The actual aim of the paper is stated much later e.g. here "To understand what relative value patients place on various aspects of the clinical encounter, the "discrete choice experiment (DCE) method" has been utilized extensively in health care research" and here "The objective of this study was to use DCE in an inter-professional academic setting to evaluate patients' preferences for various attributes of access to their family practice clinic including preferences regarding staff physician, trainee physician and allied health professionals, and method of booking

(telephone versus online) across different scenarios of reasons for seeking an appointment." Even though the IT system mentioned in the first paragraph may be a tool to plan appointments, it is just instrumental to the main question of the paper. I would therefore skip the emphasis on the IT-tool and focus on the trade offs that patients were asked to make in the DCE.

Response: We agree that the Introduction required revision. We have extensively revised the Introduction to clarify the objective and highlighted the need for further understanding of how primary care patients values aspects of access.

Methods

Development of the questionnaire is done properly.

15. Question: why did the researchers use a fractional factorial design instead of an efficient design?

Response: As per the first reviewer's comments, we have elaborated on the rationale for the methods (pg 12, para 2). Full factorial designs are rarely, if ever, used in DCEs except with very few attributes, say two or three, and very few levels per attributes. Fractional factorial designs can be very efficient and have been shown to be more desirable than orthogonal design in some applications.

16. The authors make a rough estimation of needed sample size "a sample size of 300 to 500 subjects is generally considered adequate". I too believe that these number per DCE are adequate. However, the researchers add a dimension to their DCE: they add 6 labels |(clinical scenarios) to the choice sets and as such divide the sample into six. This means that overall 430 respondents were included. This means that per clinical scenario $430/6 \sim 72$ respondents were included. Could the researchers please explain to the reader that this yielded sufficient power for the analyses, even though it is much smaller than the lower boundary of the sample size stated to be needed?

Response: We agree this point needs more explanation. We have added the following wording to the methods (pg 11 para 1):

For a DCE study, a sample size of 300 to 500 subjects is generally considered adequate and Johnson's often-used rule-of-thumb calculates a sample of 100 for a DCE having our design specifications.[21]

For additional information, the complete sample of 430 subjects was analyzed by HB to obtain the partworth utilities. Then, the sample was split by clinical scenario. Differences among the 6 scenarios were tested statistically. We also conducted HB analysis on each randomized health scenario group, i.e., those subjects who completed the DCE for each specific scenario. As recognized by the reviewer, those sample sizes were rather small and the variances would be correspondingly larger. We statistically tested whether each set of PWUs estimated on the individual strata was significantly different from the PWUs per scenario that were subsetted after the HB analysis on the total sample. There were 19 attributes and 6 randomized health scenarios, making 114 statistical tests conducted (Mann-Whitney Wilcoxon Test). Among those 114 tests, only 8 or 4.6% were significantly different, although directionally very similar. Consequently, we decided to use the HB analysis that was conducted on the full sample with those respondents' PWUs split by the health scenario in effect for each participant.

Results

17. Here it becomes extremely difficult for me to judge the paper. The main results apparently are displayed in a figure 2. I do find a caption with figure to and below the document I see a figure containing box plots, that may be related to the caption.

Response: We agree the results require clearer presentation. We have replaced the figure with a new version showing the overall results for relative importance of the attributes overall (new Figure 2) showing the relative importance of attributes across the six health states.

18. Why did the authors chose not to present a table with parameters of the attributes? Presenting 6 models for six of the clinical scenarios? That would give the reader the opportunity to thoroughly study all results and distil whether they would yield similar conclusions as the authors of the paper.

Response: Thank you for this suggestion. We hope that the Supplementary File 1 will provide the detail requested. These results show the aggregate PWUs for the 430 respondents and the PWUs when split by health scenario. With these results the reader can now discern the range of influences of the

attributes over the health scenarios. For example, the PWU for 'knowing the health care provider well' is higher for a routine check-up than for a new cold.

19. The figure "importance of the attributes" is most informative to me of the results included in the paper. However, this figure only presents the mean of all six scenarios. I would find it interesting to see this figure for all six scenarios.

RESPONSE: As per the above comment, the Supplementary File 1 provides detailed information.

There are several interesting comparisons, one of which is the much higher importance placed on Provider Level and Familiarity with provider by those in the Routine Check-up scenario and the higher emphasis on Time-to-Appointment among those who were in the New Sudden Pain scenario. These comparisons are, of course, in line with the PWUs and make a great deal of sense.

Conclusion

20. The authors state "results lend support to the notion that improved and timely access to primary care seems to be the leading priority for patients as well". This only holds for the overall results. In the final figure, I would conclude that if the medical condition does not require quick response, patients are willing to wait longer to see their own doctor. This final result is more nuanced, and provides more opportunities to tailor the provided care to the needs of patients, given their situation.

Response: Thank you for this comment. We agree that the conclusions should be more nuanced and have revised them (pg 20 para 3 and Abstract):

In conclusion, patients preferred timely access to care over all other attributes for the majority of health scenarios tested in this study. In other words, patients seeking care for sudden pain, new cold-like illness or other episodic ailments are willing to trade-off continuity for the offer of a timely appointment. The exception to this rule is the scenario of a patient booking for a routine check-up where they prefer to see the provider with which they are most familiar. These results support the notion that advanced access booking models which hold most, but not all appointment spots for same day access match up well with patient preferences over a vast array of clinical scenarios.

Abstract conclusion: These results support the notion that advanced access booking systems which leave the majority of appointment spots for same day access and still leave a few for continuity (check-up) bookings, align well with trends in patient preferences.

21. The previous point may also explain why I do not understand the statement "These results help to demonstrate that an advanced access booking model does in fact target what many patients value most, across a number of health states – that is, timely access to their primary care team". I do not conclude, with the information that I have, that patients always want timely access to health care, as they appear to be willing to wait in some health states. Second, I would imagine that the patient does not really care which booking system is used by the organisation to schedule appointments, just as long as it gets done properly. So I wonder why the authors make this strong statement that is only partly supported by their results.

Response: We hope that the revisions to the conclusions have addressed this comment.

Minor points Abstract

22. Results end with "were not strongly related to the clinical scenario." This needs earlier introduction.

Response: This sentence in the abstract has been re-written as:

Patients showed a significant preference ($p < 0.02$) for their own physician for booking of routine annual check-ups and other logical preferences across attributes overall and by clinical scenario.

23. Conclusion ends with “[...]advanced access booking systems seem to target issues that patients value highly.” This is an example of an IT-system popping up in the text that appears to play an important role in the daily organisation of the place where data collection took place, but the reader gets too little information to fully grasp the importance and implications of the implementation of the IT-system at hand.

Response: As per above comments on the conclusions, we have revised.

VERSION 2 – REVIEW

REVIEWER	Matthew Quaife LSHTM, UK
REVIEW RETURNED	17-Oct-2018

GENERAL COMMENTS	The authors have addressed my concerns, and in my view the is ready for publication. Nice work all.
-------------------------	---

REVIEWER	Mattijs Lambooi RIVM: National Institute of Public Health and the Environment. The Netherlands
REVIEW RETURNED	26-Oct-2018

GENERAL COMMENTS	<p>Review Patient trade-offs between continuity and access in primary care interprofessional teaching clinics: a cross sectional survey using discrete choice experiment.</p> <p>General: I would like to thank the authors for the revision of the paper. It is much nicer to read and clearer to understand what they have done. This causes other questions to arise. Sorry about that, but I think the paper still can improve with some changes.</p> <p>Abstract. Abstract is clear</p> <p>Introduction. The introduction has improved considerably. The problem and research aim are presented clearly.</p> <p>Page 5 Lines 51 and following: “Speed of access and continuity with the same clinician are commonly studied” statements such as this are usually supported by sources from the literature. I would suggest sot do so here as well.</p>
-------------------------	---

The following text in de paragraph nicely illustrates the added value of DCE over a regular survey considering this research question.

The authors state that they followed the ISPOR guidelines in development of attributes and levels.(page 7 lines 25 and following). The following description is indeed in line with the steps as described in the ISPOR guideline. My personal view is that the ISPOR guidelines are too limited in their requirements on description of attribute development. Since presentation of the scenario's, including the wording of level and attributes are essential for the results of the DCE, it should be described more thoroughly in all DCE-papers. Given that this way to present the development, and that an example of the choice sets are included in the paper, I will not ask the authors to change it. But I do want to use this opportunity to express my concerns on common practice in the DCE community.

Statistical methods.

I have no experience with Hierarchical Bayes estimations. I do not feel capable to judge whether this is done properly.

Sample.

The descriptives of the sample are presented clearly. Since it is a convenience sample, I think it would add value to the paper to study whether the sample differs from average demographics of the patient population in the clinics. Either the two clinics used for this study or nationwide numbers, which of the two is more accessible to the authors. This would give readers a feel for the representativeness of the sample.

P 15 lines 54 and 55 "Aggregate importance differences were statistically different.", combined with the statement (P11 lines 3-4) "...a sample size of 300 to 500 subjects is generally..." combined with (p13, line 46 "430 fully complete and usable responses" make me suspicious about the statistical power available in the sample to draw conclusions of importance differences on the sample sizes of each of the six disease states.

The authors do not state that for each disease state, the importance difference are statistically different. However, figure Relative importances of attributes by scenario presents these differences, and the authors suggest that the severity of the health state affects the relative importance of appointment time.

I appreciate adding Supplementary file 1, where the readers can conclude that most effects exceed the 95% thresholds. But the paper needs to present more evidence to show that the relative importance weights differ significantly by health state of the scenarios. Considering that the authors do not mention this result in their discussion and conclusion, they may want to consider not splitting the results to disease specific cases, and just present the aggregate results. I think the added value of that information is obvious, and it removes a weakness from the paper.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

COMMENT #2:

The introduction has improved considerably. The problem and research aim are presented clearly.

Page 5 Lines 51 and following: “Speed of access and continuity with the same clinician are commonly studied” statements such as this are usually supported by sources from the literature. I would suggest not do so here as well.

RESPONSE 2: We agree, these statements need referencing. Several citations have been added. [Bower et al., 2003; Gerard et al., 2008; Turner et al., 2007; Haggerty et al., 2003; Saultz et al., 2004; Cheraghi-Sohi et al., 2008; Uijen et al., 2012]

COMMENT #3:

The following text in the paragraph nicely illustrates the added value of DCE over a regular survey considering this research question.

The authors state that they followed the ISPOR guidelines in development of attributes and levels. (page 7 lines 25 and following). The following description is indeed in line with the steps as described in the ISPOR guideline. My personal view is that the ISPOR guidelines are too limited in their requirements on description of attribute development. Since presentation of the scenario's, including the wording of level and attributes are essential for the results of the DCE, it should be described more thoroughly in all DCE-papers. Given that this way to present the development, and that an example of the choice sets are included in the paper, I will not ask the authors to change it. But I do want to use this opportunity to express my concerns on common practice in the DCE community.

RESPONSE 3: Thank you for this comment, it will be taken into consideration in any of our future research using this method.

COMMENT #4

Sample.

The descriptives of the sample are presented clearly. Since it is a convenience sample, I think it would add value to the paper to study whether the sample differs from average demographics of the patient population in the clinics. Either the two clinics used for this study or nationwide numbers, whichever of the two is more accessible to the authors. This would give readers a feel for the representativeness of the sample.

RESPONSE 4: We have added the mean ages and gender distribution of the two clinics to the Methods in describing the two clinics (Pg 11, para 2), and in Table 2 have added the comparable statistics (where available) for the city from the latest available national census.

The average age of patients is 48.2 years and 52.5% are female in clinic A, and average age is 45.4 years and 56.3% are female at clinic B. (Pg 14. Para 1)

COMMENT #5

P 15 lines 54 and 55 “Aggregate importance differences were statistically different.”, combined with the statement (P11 lines 3-4) “..a sample size of 300 to 500 subjects is generally...”. combined with (p13, line 46 “430 fully complete and usable responses” make me suspicious about the statistical power available in the sample to draw conclusions of importance differences on the sample sizes of each of the six disease states.

The authors do not state that for each disease state, the importance difference are statistically different. However, figure Relative importances of attributes by scenario presents these differences, and the authors suggest that the severity of the health state affects the relative importance of appointment time.

I appreciate adding Supplementary file 1, where the readers can conclude that most effects exceed the 95% thresholds. But the paper needs to present more evidence to show that the relative importance weights differ significantly by health state of the scenarios. Considering that the authors do not mention this result in their discussion and conclusion, they may want to consider not splitting the results to disease specific cases, and just present the aggregate results. I think the added value of that information is obvious, and it removes a weakness from the paper.

RESPONSE 5: We agree that further supporting information on this analysis was needed. We have included the statistical tests on the effects of health scenario on the relative importances of the attributes in a new Supplementary Table 2. We used MANOVA to test statistical significance and present a measure of effect size (eta-squared or partial-eta squared; Cohen, 1988). We have altered the Results and Discussion, and feel that the messages in the Conclusions are inclusive of these results.

The relative importance of the 6 attributes for each of the randomized health scenarios is presented in Figure 2 and the effect sizes are shown in Supplementary File 2. There was significant variation over all six attributes and across the six health scenarios (MANOVA, Wilk’s lambda = 0.694, p-value < 0.0001) indicating a range of different responses under the various health conditions. The relative importance of time-to-appointment, waiting room time, familiarity with provider and provider level varied significantly over the 6 health scenarios. Using Cohen’s guide to effect sizes as represented by eta-squared (or partial eta-squared, but equal here), the effect size of health scenario can be considered large for time-to-appointment, between medium and large for familiarity with provider, between small and medium for waiting room time, appointment convenience and provider level and small for method of booking appointment. The relative importance of time-to-appointment was statistically significantly less ($p < 0.05$) for those responding to the routine check-up scenario than all others. The relative importance of familiarity with provider was statistically significantly greater ($p < 0.05$) for those responding to the routine check-up scenario than for those responding to new cold and new sudden pain and was numerically greater than all others. (RESULTS- Page 16, para 1).

....On the other hand, there was variability in importance by the health state presented. The relative importance of familiarity with the provider was greater in the context of a routine check-up compared to a new cold and new sudden pain. This finding makes sense since most people are not in a rush to have the routine annual check-up but do like to see their regular health provider for continuity. (DISCUSSION Pg 18, para 2).

VERSION 3 – REVIEW

REVIEWER	Mattijs Lambooj National Institute of Public Health and the Environment, the netherlands
REVIEW RETURNED	22-Jan-2019

GENERAL COMMENTS	All of my concerns that I expressed are adressed appropriately.
-------------------------	---