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)	Preferences for centralised emergency medical services: a discrete choice experiment
AUTHORS	Bhattarai, Nawaraj; Mcmeekin, Peter; Price, Christopher I; Vale, Luke

VERSION 1 – REVIEW

REVIEWER	Gerard FitzGerald Queensland University of Technology, Brisbane Australia
REVIEW RETURNED	06-May-2019

GENERAL COMMENTS	The aim of this paper is to test preferences in regard to trade offs between quality of care and access to care that may support an agenda to centralize complex emergency care. The research is well constructed and the findings and analysis clearly presented.

REVIEWER	Dr Caroline S Clarke University College London, UK
REVIEW RETURNED	30-May-2019

GENERAL COMMENTS	Thanks very much for the opportunity to review this paper. I am a health economist working on a wide range of clinical trials and service reorganisations, and have done a small amount of work in DCEs. **Abstract (page 2)** Could include where participants were recruited from in the abstract (if space). **Background (page 4)** The wording of the first sentence seems to suggest that "clinical care standards" are a threat to the NHS. Is this what is meant? Perhaps re-phrase. **Page 4** "A radical and transformative change is essential for the NHS to maintain safety and quality" This is perhaps true, but it's possible that more funding could also provide part of the solution. Could be worth mentioned this here as well. **Page 4** "Incorporating patient views in healthcare policy decisions may improve the uptake and efficiency of services." This is true, but more importantly it should be noted that including patient/public views also leads to better research on treatments and service provision. Doing PPI work is not just about raising awareness among service users,

it's also about raising awareness among researchers of other points of view that can improve the quality of their work.

Methods (page 5)
"The identified attributes were assigned levels that were realistically applicable within the UK NHS.[25]"
This reference is 10 years old. Who decided what levels were realistic? Was it discussed with clinicians, patients, etc.?

Page 6
Typo – should 2128 be 2048?

Page 7
Pretesting and piloting – was any PPI work done here? Slightly ironic to have no patient/public input into a study on public and patient preferences…

Page 7-8
"The sample represented a general population whose EMS had recently been centralised."
Why was this population used? What impact could this have had on the results? Please discuss.

Page 8, 10
On page 8 it mentions "voluntary participation". Is this the same as "informed consent"?
Later on page 10 it says that informed consent was not sought – why not?

Page 10
"No patient or public were involved in the design, conduct and reporting of this research. Nevertheless, the priors generated from a pilot study conducted in a sample of general population informed the design of the DCE survey."
Please can you describe the pilot study and the priors mentioned and explain how they informed the design.

Page 11
Please note that the EQ-5D is not a measure of health, it is a measure of health-related quality of life, which is a different thing. Please change this.

Page 11
"The general health measures (EQ-VAS=75.5, EQ-5D-5L=0.77) is similar to that estimated for the UK general population above 55 years of age.[41]"
Please state here what is the estimated utility (quality of life) score for the UK general population above 55 years of age.

Table 2
Why are the age categories presented in this way? It might be more usual to present the median and range (or mean and SD) of age in a single row.

Table 2
I am not sure that 2 decimal places are required here. One (or none) would be sufficient.

Table 2
Same comment in this table about "self-reported health" – this is not

what the EQ-5D measures (the 5 questions or the VAS). Please change this.

Table 2
For the emergency experience, does it have to have been they who suffered the emergency, or could it have been a friend/family member? Would be interesting to know what they were asked.

Table 3
For the regressions where the variables were treated as continuous, what values did you use? E.g., for travel time, was it 15 mins, 60 mins, 90 mins, and then 180 mins? Or were other assumptions made for the "less than" and "more than" categories? Please explicitly state this for all variables. Without explicitly stating this, it could be assumed that you have just assigned values of 1, 2, 3, 4 for the 4 levels, which would probably not be appropriate in most cases – please state what assumptions you have made.

Table 3
Please make sure that the labelling of the attributes is unambiguous regarding what the categories actually mean. E.g. for the "survey mode", does this mean paper vs. online? What are the categories for each? Same for "outpatient follow-up" – what are the values used for each category?

Table 3
For the paper vs. online point – was there a difference in the demographics of those filling it in online and those filling it in on paper? How was this dealt with? Please explain more fully what this variable was for and what it meant.

Discussion (page 19)
I struggled with the discussion of gender differences here. Is it possible that men in this study had stronger preferences for better/faster treatment because of cultural and social conditioning, whereas women had lower expectations/preferences for the same reason? I couldn't work out if the authors were suggesting that those with stronger preferences to be treated more quickly should in fact be treated more quickly ("it is recommended that future researchers and healthcare providers consider whether decisions and communication about centralising services should be sensitive to gender differences."). I would very strongly suggest that expectations based on gender should not inform this type of decision making, especially when there is a wider social context to be considered. Please consider re-phrasing this.

Page 18
"This study reported differences in preferences in patient related to gender and age."
What were those differences? Please discuss further as it seems to be the only other study to report gender differences that you mention.

Page 19
There is extensive discussion in "Implications of the study findings" about "choosing hospitals for emergency healthcare". Is this looking at A&E walk-ins, or at blue-light emergencies where the patient is e.g. unconscious? In the latter scenario you would not be making an active choice in the UK of which hospital to go to. Please re-phrase/explicitly address this and discuss what this means for the

analysis presented.

Page 19
"This assumes that the ambulance response would be the same, and that other
aspects of local services would not be affected"
Are these assumptions realistic? Please discuss in the paper. There would need to be some coordination with town/traffic planning services and consideration of where ambulance services are based.

Page 20
"Despite a high proportion of complete and correct responses in the DCE"
What do you mean by "correct responses"?

Further thoughts (1)
I couldn't see any mention or discussion of the number of levels (4 or 2) per attribute. Why was 4 chosen for most of them? Please discuss in the paper.

Further thoughts (2)
What happens if you simply sum the travel+waiting times, and use them as a single attribute? It's then the overall time from call to treatment, which might be a confounder in some way if people have that in the back of their mind when making their choices.

Further thoughts (3)
How is triage in the A&E department considered, i.e. severity of condition? What assumptions were made around this?

REVIEWER	Glenn Arendts
	University of Western Australia
REVIEW RETURNED	26-Jun-2019

GENERAL COMMENTS	This is a generally well constructed and interesting study, that I believe adds to the limited existing literature on the topic.
	I would suggest the following could be clarified by the authors to improve an otherwise sound paper:
	1. The use of attributes derived only from existing literature to make up the alternatives offered to respondents is limited. The risk is that an important attribute that is highly influential in choice was not modelled. Did the authors consider obtaining qualitative data from the Northumberland population to inform the experiment design? I note you mention one line in the limitation section about this, but I would think a more comprehensive discussion is needed.
	2. Random population based sampling is claimed, but the methods sound more like targeted sampling with individuals approached to participate. Please clarify and if particular the targeting was based on specific respondent characteristics
	3. Whilst understanding an efficient design was chosen, this does not preclude an a priori sample size calculation, which I missed in the manuscript. Was one done and if not, why not?
	4. It is not clear to me how the covariate of respondent age was treated in your analysis. As differences were found in preferences between paper and webbased participants, I am curious that there was no effect modification of age on attribute preference. Could you please better explain how you modelled age, so the reader can interpret the robustness of your conclusion that preferences do not vary by age? As it is the results seem counter to known influences of poor social mobility on access to healthcare for older people.

REVIEWER	Milena Pavlova Maastricht University
REVIEW RETURNED	09-Sep-2019

GENERAL COMMENTS	This is a well-written paper presenting a well-elaborated study. The relevance of the paper for the journal audience is clear. I have only two minor comments which need to be addressed by the authors. 1. The base for selecting the attributes for the DCE is clear: literature and quality standards. But it is unclear if all possible attributes found in these two sources, were included or a selection was made. And if yes, which attributes were selected and which not, and why? 2. The detailed DCE design (attribute levels per profile) is not included in the paper. It will be fine to have this as a supplementary material.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Gerard FitzGerald
The aim of this paper is to test preferences in regard to trade-offs between quality of care and access to care that may support an agenda to centralize complex emergency care. The research is well constructed and the findings and analysis clearly presented.

Thank you very much; we appreciate that you found our research well constructed and clearly presented.

Reviewer: 2
Reviewer Name: Dr Caroline S Clarke

Abstract (page 2)
Could include where participants were recruited from in the abstract (if space).

Thank you. We have now included this in the abstract, which now reads "Participants were a randomly sampled general population, aged 16 years or above recruited from Healthwatch Northumberland network database of lay members and from clinical contact with Northumbria Health Care NHS Foundation Trust via the Patient Experience Team".

Background (page 4)
The wording of the first sentence seems to suggest that "clinical care standards" are a threat to the NHS. Is this what is meant? Perhaps re-phrase.

Thank you for noticing this error. We have now corrected this and the first sentence now reads as "The National Health Services (NHS) in the United Kingdom (UK) faces growing service demands and costs which threaten its sustainability and financial stability".

Page 4
"A radical and transformative change is essential for the NHS to maintain safety and quality" This is perhaps true, but it's possible that more funding could also provide part of the solution. Could be

worth mentioned this here as well.

Agree. We have now added a sentence, which reads "Increased funding could be a part of the solution but the NHS budget is limited".

Page 4
"Incorporating patient views in healthcare policy decisions may improve the uptake and efficiency of services." This is true, but more importantly it should be noted that including patient/public views also leads to better research on treatments and service provision. Doing PPI work is not just about raising awareness among service users, it's also about raising awareness among researchers of other points of view that can improve the quality of their work.

Agree. We have now added a sentence, which reads "Including patient/public views may also lead to better quality research on treatments and service provision".

Methods (page 5)
"The identified attributes were assigned levels that were realistically applicable within the UK NHS.[25]"
This reference is 10 years old. Who decided what levels were realistic? Was it discussed with clinicians, patients, etc.?

We agree the reference is 10 years old, but relevant research in this field is quite rare and this was the most recent research we could find. However, these levels were discussed within the research team consisting of a clinician with considerable number of years of experience in this area. We have now reworded the sentence as follows:

"After discussions within the research team which also consisted of an experienced senior clinician working in emergency medicine, the identified attributes were assigned discrete levels that were likely to be applicable within the UK NHS and closely reflected the reality[25] whilst keeping the design as simple as possible".

Page 6
Typo – should 2128 be 2048?
Yes, this was a typo- thank you very much. Corrected it.

Page 7
Pretesting and piloting – was any PPI work done here? Slightly ironic to have no patient/public input into a study on public and patient preferences…
We agree there was no PPI work done pre-survey. We have stated this in the "Patient and public involvement" section. However the attributes selected were ones which could occur in the real world as a result of the described service reconfiguration. We have corrected a sentence on a limitation in the Strengths and Limitation section, which now reads as follows:

"Furthermore, the attributes used in the DCE were taken from the literature solely and were not based of findings from qualitative research[49], nor was the choice of attributes and levels informed by any patient and public involvement (PPI). Consequently, some other important attributes may have been left out."

Page 7-8
"The sample represented a general population whose EMS had recently been centralised."
Why was this population used? What impact could this have had on the results? Please discuss.
Thank you. This work was a part of broader research, which also looked into cost-effectiveness of

centralisation of emergency medical services in a real world setting, and it was necessary the sample for this DCE research to be taken from the same setting where the EMS had recently been centralised. This could have had positively impacted our results as preferences from participants with recent experience of centralisation could be more representative of potential users of newly built centralised emergency hospital. We have discussed this and refer to the first two sentences of "Strengths and limitation" section. The first two sentences read as follows:

"The findings of this study should be interpreted in light of some strengths and limitations. We attempted to study a wide cross section of a local unselected population to represent the preferences of potential users for a newly built centralised emergency hospital. However, due to the contact databases which were available for us to invite participants, it only represents specific population groups registered with the Healthwatch or Northumbria Healthcare NHS Foundation Trust, and sampling bias cannot be ruled out."

Page 8, 10
On page 8 it mentions "voluntary participation". Is this the same as "informed consent"?
Later on page 10 it says that informed consent was not sought – why not?

Thank you for your question. The survey questionnaire was completed either online or in post, therefore because of the nature of the survey it was not possible for us to take a written consent before proceeding to complete the survey. Therefore, we indicated in the Ethics Approval section that returning a completed questionnaire was an indication of consent.

Page 10
"No patient or public were involved in the design, conduct and reporting of this research. Nevertheless, the priors generated from a pilot study conducted in a sample of general population informed the design of the DCE survey."
Please can you describe the pilot study and the priors mentioned and explain how they informed the design.

Thank you. We have now revised the sentences in the Questionnaire design section as follows:
"The generation of efficient design requires a priori knowledge of attributes used in the choice model.[28] The prior[28] estimates of attribute coefficients used in the final efficient design were derived from a pilot survey (see below)".

We have now reported the priors in supplementary file: Table S1 and have amended the last two sentences in "Pretesting and piloting" section in the manuscript as follows:
The attribute coefficients generated from the pilot were used as priors to generate the final questionnaire design as described earlier. The priors generated from the pilot are presented in the supplementary file: Table S1.

Page 11
Please note that the EQ-5D is not a measure of health, it is a measure of health-related quality of life, which is a different thing. Please change this.
We agree, thank you very much. We have amended this. The sentence now reads "The health related quality of life (EQ-VAS=75.5, EQ-5D-5L=0.77) is similar to that estimated for the UK general population above 55 years of age (which is EQ-VAS=77.6; EQ-5D=0.77). "

Page 11

"The general health measures (EQ-VAS=75.5, EQ-5D-5L=0.77) is similar to that estimated for the UK general population above 55 years of age.[41]"
Please state here what is the estimated utility (quality of life) score for the UK general population above 55 years of age.

Thank you. We have now added this information as shown in the sentence above.

Table 2
Why are the age categories presented in this way? It might be more usual to present the median and range (or mean and SD) of age in a single row.
Thanks, we agree, but we presented the age categories to show the age composition of the study participants- reflecting on the fact that the people above 50 are more likely to attend the emergency medical services. The focus on summary statistics such as median and mean would not have as clearly illustrated this.

Table 2
I am not sure that 2 decimal places are required here. One (or none) would be sufficient.

Agree. We have now changed this to 1 decimal places.

Table 2
Same comment in this table about "self-reported health" – this is not what the EQ-5D measures (the 5 questions or the VAS). Please change this.

Thanks. We have replaced this with "Health related quality of life".

Table 2
For the emergency experience, does it have to have been they who suffered the emergency, or could it have been a friend/family member? Would be interesting to know what they were asked.

Thanks for this question. It was the emergency experience of the patient. The question related to their emergency visit in the last 12 months, if any.

Table 3
For the regressions where the variables were treated as continuous, what values did you use? E.g., for travel time, was it 15 mins, 60 mins, 90 mins, and then 180 mins? Or were other assumptions made for the "less than" and "more than" categories? Please explicitly state this for all variables. Without explicitly stating this, it could be assumed that you have just assigned values of 1, 2, 3, 4 for the 4 levels, which would probably not be appropriate in most cases – please state what assumptions you have made.

Thank you. We have now clarified this in the text in section "Regression analysis of the DCE data" with the sentence "In the travel time and waiting time attributes where the levels had "less than" or "more than" categories, only the number of minutes were used, for example 30 minutes for "less than 30 minutes". The outpatient follow up was coded as "0" for local hospital and "1" for a distant hospital"

Table 3
Please make sure that the labelling of the attributes is unambiguous regarding what the categories actually mean. E.g. for the "survey mode", does this mean paper vs. online? What are the categories for each? Same for "outpatient follow-up" – what are the values used for each category?

Thank you. We have now added "Survey Mode refers to paper vs web based" in the footnote of

Table3. The way outpatient follow up was coded is clarified in the text as mentioned above.

Table 3
For the paper vs. online point – was there a difference in the demographics of those filling it in online and those filling it in on paper? How was this dealt with? Please explain more fully what this variable was for and what it meant.

Yes, there were higher proportion of older people in the postal group. We have now added the following text in the "Strengths and limitation" section
"One of the strengths of this survey lies in the fact that both survey modes- web based and postal (paper) were used enabling us to increase the representation of a wider cross section of population. However, experimenting with the different approaches for survey administration was beyond the scope of this study, use of different approaches might have introduced a response bias because of the systematic differences (for example proportion of older people) between the respondents in each approach".

Discussion (page 19)
I struggled with the discussion of gender differences here. Is it possible that men in this study had stronger preferences for better/faster treatment because of cultural and social conditioning, whereas women had lower expectations/preferences for the same reason? I couldn't work out if the authors were suggesting that those with stronger preferences to be treated more quickly should in fact be treated more quickly ("it is recommended that future researchers and healthcare providers consider whether decisions and communication about centralising services should be sensitive to gender differences."). I would very strongly suggest that expectations based on gender should not inform this type of decision making, especially when there is a wider social context to be considered. Please consider re-phrasing this.

No, we did not mean that those with stronger preferences to be treated more quickly should in fact be treated more quickly. Instead, we wanted to say that women and men may have differences in preferences for emergency health care and service provision should not be driven by the preferences of one particular gender.

We have now rephrased this sentence as follows:

"Whilst this finding may specifically reflect this cohort, it is recommended that future researchers and healthcare providers consider any communication about centralising services should be sensitive to gender differences whilst not allowing decisions about service provision to be driven by the preferences of one particular gender especially without clearly understanding why these differences exist".

Page 18
"This study reported differences in preferences in patient related to gender and age."
What were those differences? Please discuss further as it seems to be the only other study to report gender differences that you mention.

We have now added further information in the following text in the "Comparison with other studies" section:

"This study reported differences in preferences in patient related to gender and age suggesting that patients are more likely to prefer stay at their local hospitals as their age increases and males are more likely to choose to move to a non-local alternative hospital than females".

Page 19

There is extensive discussion in "Implications of the study findings" about "choosing hospitals for emergency healthcare". Is this looking at A&E walk-ins, or at blue-light emergencies where the patient is e.g. unconscious? In the latter scenario you would not be making an active choice in the UK of which hospital to go to. Please re-phrase/explicitly address this and discuss what this means for the analysis presented.

Thank you for raising this. We have added the following text towards the end of the "Implications of the study findings" section:

"However, our survey was framed to look at choices for emergency ambulance admissions and we caution that these findings may not reflect the preferences in "blue-light" emergencies where the patient is not in a position to make the choice of which hospital to go to. Nevertheless, the preferences observed are for planning services and not for making decision about immediate care in an emergency."

Page 19

"This assumes that the ambulance response would be the same, and that other aspects of local services would not be affected"
Are these assumptions realistic? Please discuss in the paper. There would need to be some coordination with town/traffic planning services and consideration of where ambulance services are based.

Thank you for bringing this to our attention. We now say the following:
"This assumes that the ambulance response would be the same, and that other aspects of local services would not be affected, which may not be realistic but exploring this with the DCE would have added additional complexity that may make the tool too difficult to complete. However, healthcare centralisation planners should also carefully how best to work with town and traffic planning services to help optimalise services or at the very least ensure that other system constraints do not remove any potential benefits of centralisation."

Page 20

"Despite a high proportion of complete and correct responses in the DCE"
What do you mean by "correct responses"?

It meant that the survey was completed correctly, not that the answers were "correct". We have now corrected this. The sentence now reads as follows:
"Despite a high proportion of complete responses in the DCE……"

Further thoughts (1)

I couldn't see any mention or discussion of the number of levels (4 or 2) per attribute. Why was 4 chosen for most of them? Please discuss in the paper.

This was done to done to closely reflect reality with respect to the variation in levels and to keep the design as simple as possible so that respondents can make a trade-off easily. , but still provide more information than a binary choice. We used the current attribute levels to ensure that we follow desirable design criteria such as level balance, minimum overlap of the levels etc, and had fewer number of choice sets thereby reducing cognitive burden (in making trade-offs) and complexity. We have amended the sentence in the "Attributes and level" section as follows:

"After discussions within the research team which also consists of an experienced clinician, the

identified attributes were assigned levels that were realistically assumed to be applicable within the UK NHS and closely reflected the reality[25]. The design was kept as simple as possible so that respondents can make a trade-off easily".

Further thoughts (2)
What happens if you simply sum the travel+waiting times, and use them as a single attribute? It's then the overall time from call to treatment, which might be a confounder in some way if people have that in the back of their mind when making their choices.

This is what we did not look into but an individual when faced with a choice question can readily see the overall time. We had considered merging these at design stage but decided to decompose then into two attributes. This allowed us to understand separate aspects, which is important because it is quite plausible that travel times are reduced but treatment times increase. We do not know whether our results were confounded by the way attributes were framed and how the respondents understood the attributes. We have reworded this in the attribute-framing example as follows:

"However, the way the attributes were framed could possibly have influenced the choices made by respondents,[50-52] and it is unclear whether choices would remain the same, if the attributes were framed in the other way, e.g. would preferences have been different if "risk of dying" was framed as "chance of survival"? or if "travel time" and "waiting time" were used as a single attribute of "call to treatment time"? But, merging the travel time and waiting time as a single attribute at the design stage would have reduced the explanatory power of the DCE"

Further thoughts (3)
How is triage in the A&E department considered, i.e. severity of condition? What assumptions were made around this?

Thanks, this was assumed as an average time irrespective of the severity of the condition

Reviewer: 3
Reviewer Name: Glenn Arendts
This is a generally well constructed and interesting study, that I believe adds to the limited existing literature on the topic.

Thank you

I would suggest the following could be clarified by the authors to improve an otherwise sound paper:

1. The use of attributes derived only from existing literature to make up the alternatives offered to respondents is limited. The risk is that an important attribute that is highly influential in choice was not modelled. Did the authors consider obtaining qualitative data from the Northumberland population to inform the experiment design? I note you mention one line in the limitation section about this, but I would think a more comprehensive discussion is needed.

Thank you. It was a limitation but the attributes used are still representative of the impact of service change and had clinical validity.

2. Random population based sampling is claimed, but the methods sound more like targeted sampling with individuals approached to participate. Please clarify and if particular the targeting was based on

specific respondent characteristics

The research was a part of broader research assessing centralisation of emergency medical services. It was necessary for us to base this survey in a setting where actual centralisation of emergency medical services had recently occurred. However, we did not approach participants on specific characteristics. We acknowledge that because of our approach of approaching participants in databases of named organisations to take part in the survey, sampling bias cannot be ruled out.

3. Whilst understanding an efficient design was chosen, this does not preclude an a priori sample size calculation, which I missed in the manuscript. Was one done and if not, why not?

Thank you; a priori sample size was generated, somehow this was missed in the manuscript. We now add the following text in "Sample" section:

"Sample size estimation methods in healthcare DCE studies are currently developing[32]. Therefore, the sample required for this study was estimated following the rule of thumb suggested by Johnson and Orme[32-34]using the equation N>500L/TA where L is the largest number of levels for any of the choice attributes, T is the number of choice sets and A is the number of alternatives assessed. The required minimum sample was estimated as 100 respondents. However, much larger sample was targeted to allow for heterogeneity between respondents."

4. It is not clear to me how the covariate of respondent age was treated in your analysis. As differences were found in preferences between paper and web based participants, I am curious that there was no effect modification of age on attribute preference. Could you please better explain how you modelled age, so the reader can interpret the robustness of your conclusion that preferences do not vary by age? As it is the results seem counter to known influences of poor social mobility on access to healthcare for older people.

Age was modelled as a continuous variable and we agree it was surprising to see that there was no evidence that age had an influence on the preferences. It is possible that this is because everyone viewed emergency health need in the same way irrespective of age. However, our finding is of 'no evidence of; may reflect a lack of statistical power to detect a difference. At the same time we understand that the differences between preferences between paper and web based participants could be because of the possible systematic differences between the participant characteristics in each of the survey modes. However, the two survey modes were used to widen participation and hence we find it unsurprising that there is a difference in characteristics of participants. We would contend that given the purposive reason for using both modes it wold be more surprising if there were no differences in participant characteristics.

Reviewer: 4
Reviewer Name: Milena Pavlova
This is a well-written paper presenting a well-elaborated study. The relevance of the paper for the journal audience is clear.

I have only two minor comments which need to be addressed by the authors.
1. The base for selecting the attributes for the DCE is clear: literature and quality standards. But it is unclear if all possible attributes found in these two sources, were included or a selection was made. And if yes, which attributes were selected and which not, and why?
Thank you, we have stated in the "Attributes and level" section that it was essential the attributes reflected the key performance measures commonly used in emergency medical services and key quality indicators in NHS England. We now add the following text:
"A long list of attributes and their levels were identified from the literature, but the list was shortened

based on their relative importance for the NHS "

2. The detailed DCE design (attribute levels per profile) is not included in the paper. It will be fine to have this as a supplementary material.
We have now included the blocks of choice sets used in the study in the supplementary file.

VERSION 2 – REVIEW

REVIEWER	M.Pavlova Maastricht University, The Netherlands
REVIEW RETURNED	11-Oct-2019

GENERAL COMMENTS	All my comments are well addressed.