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)	HIV testing history and preferences for future tests among gay men, bisexual men and other MSM in England: results from a cross-sectional study
AUTHORS	Witzel, T Charles; Melendez-Torres, G.J.; Hickson, Ford; Weatherburn, Peter

VERSION 1 - REVIEW

REVIEWER	Paul Flowers Glasgow Caledonian University Scotland Also conducting research around HIV testing I do collaborate with SIGMA research on various projects and plan to continue to do so in the future.
REVIEW RETURNED	26-Feb-2016

GENERAL COMMENTS	I have some concerns about the overall quality of the manuscript, which the previous reviewers had identified (rather harshly) but it does need proof reading and could be improved in terms of tightening the language throughout. Table 4, for example, is split by preceding tables, references are poorly spelled. To me, the overall analytic plan underpinning the data presented, doesnt quite match the main narrative that binds the paper together. The discussion particularly seems slightly removed from the results presented. The fundamental idea behind the manuscript is great and I was very excited to have the opportunity to review it, as it promises to be highly relevant for future testing policy and practice in England and elsewhere. I do how ever have a few concerns, at least one of them major but most of them minor I worry the authors will find some of my comments annoying given not all of them were identified within the first round of reviews. But I am really keen to see the most being made out of this key data set. Major I think there is an assumption implicit within the manuscript, that implies that the UK guidance directly shapes testing behaviours, or that modes of delivery are the only interventions that seek to intervene with barriers to testing. I think being more explicit about the interventions which have been delivered in the recent past which fill in the middle ground between UK policy and MSM themselves would be helpful. It would also be useful to acknowledge the importance of psychological barriers to testing which explain far more variance in testing behaviours than changes to structural barriers (such as expanding the repertoire of test delivery), so I think

it needs to be clear that on quite a fundamental level, despite innovations in testing delivery, psychological barriers to testing still need to be reduced. Secondary questions of which testing option suits who best (the arena of this paper) then come in to play.

My most major concern relates to the operationalisation of the BHIVA/PHE guidance which as I understand it has two important elements, one: all MSM test annually (addressed and critiqued within the manuscript), and two that, those having sex UAI with new or casual partners test more often (every three months). This latter element is not addressed within the manuscript but is fairly fundamental. between page 3 and 4 this distinction is lost, and the lack of analysis with sexual behaviour variables means this remains unexplored. The paper would be so much more useful and citeable if it also included analyses of this second issue. Then a critique of BHIVA/PHE policy seems apt. Im not sure if the team collected data about frequency/regularity of testing but it looks like a proxy measure could be used from the 5 categories of time frame for most recent test. I think even an analysis of those higher risk MSM and their testing patterns and testing preferences would be more valuable than some of the current manuscript with its only partial analytic focus and dialogue with testing recommendations. I hope this makes sense and am not quite sure why previous reviewers did not pick this point up.

We have sometimes used comparisons of ever tested/tested within the last 12 months and tested within the previous 3 months as a way of trying to address some of these issues, doing the same within the whole sample, or within those men reporting new/casual UAI partners might be worth exploring too

Another key issue is the approach to missing data, unless I have missed it, i couldnt see how missing data is reported and if analytic strategies have been adopted to deal with it. I know in some of the on-line surveys I am involved in missing data levels can be quite high.

Minor
Page 3, explain 'test and treat' model
Page 3, lines 41-43, 'treads an awkward line' seems an odd phrase and I am not sure that the two elements of UK policy are so problematic. The two groups are not exclcusive and static and we'd expect to see variability across the life span in terms of risk.
page 4, line 22, 'their service' seems and odd phrase to use.
Page 8. line 30, given the % of never testers, it would be interesting to hear the authors views on intentions to test in the future.
Page 12, lines 7-15, I got a bit lost in these lines, the text mentions testing frequency (but see above comments) I don think this was measured, so is it a proxy measure? Also given the cross-sectional nature of the data are the conclusions about age trends in HIV testing not being confounded with both generational issues and indeed with the temporal perspectives of the HIV epidemic. My head hurt a bit from trying to think this through but I dont think the data do support the claims being made here. Maybe clarifying the interpretation would help?

Pg 15, 'important educational component' seems to miss the point about knowledge not being sufficient to change behaviour. As mentioned above, and core to potential intervention design,

acknowledging the psychological barriers to testing is vital (Lots of reviews highlight this, or here in the UK, some of my own work does this. Flowers P, Knussen C, McDaid L, Li J (2013). Has testing been normalised?: An analysis of changes in barriers to HIV testing among men who have sex with men between 2000 and 2010 in Scotland, UK: HIV Medicine, Feb;14 (2):92-8. DOI: 10.1111/j.1468-1293.2012.01041. We've got some online data that supports these findings too and reflects many of the demographic associations report here too.
page 12, line 57 - was previous testing site assessed? GIven the range of testing options? Given the centrality of habit in predicting future behaviour it would be good to hear more of this if it was collected

page 13, lines 12-17, community based rapid HIV testing were most popular within the groups most likely to test. Which variable is used to reflect likelihood of testing? I couldn't work this out .
page 14, 'service delivery puzzle' sounds a little odd.
the final sentence before limitations, doesnt seem to follow on from anything. Given the analysis doesnt look at sexual behaviour it feels an odd way to end the paragraph.

To summaries, for me the paper as it stands still has great potential to be an influential and useful contribution to the field but a number of problems remain to be addressed. To me, a more coherent analytic focus is needed that matches one of the main research questions; dialogue with the BHIVA/PHE guidance with the approach to key analysis (and incorporating an assessment of the second clause for men at some risk), i think the paper needs some further detail regarding which interventions have been delivered within England and some indication of their content (in relation to barriers to testing), I think the paper does need to address the psychological and social barriers to testing that persist and will remain obstinate, despite further choices in testing options. The manuscript overall needs to be improved with spell checks and some reigning in of the interpretations of the associations reported.

I have really enjoyed the manuscript but am not so sure that other people, less passionate about the topic, would be willing to interpret and reflect on what has been presented. I hope this is useful feedback and not too annoying, given I am suggesting new analytic strategies on a second round of reviewing.

REVIEWER	Damian Patrick Conway
	The Kirby Institute, University of New South Wales, Australia
REVIEW RETURNED	18-Apr-2016

| GENERAL COMMENTS | The authors have conducted an analysis of data from the 2014 Gay Men's Sex Survey (GMSS) in the UK. The title of the manuscript and the introduction of the paper claim that this analysis has been conducted to describe and assess the HIV testing history and future HIV testing preferences among gay, bisexual and other men who have sex with men (MSM) in England. As the authors state, assessing HIV testing patters among British MSM is a very important matter in the current HIV test and treat era. While the data presented in the paper does describe the HIV testing history and preferences of these men, from reading the paper it appears that the authors were also seeking to answer these questions: 1) Is the HIV testing |

history of English MSM consistent with national UK HIV testing guidelines? 2) Are the current UK HIV testing guidelines appropriate? The authors have sought to present their discussion and the interpretation of their findings around whether certain groups of MSM have sufficient behavioural risk to justify the current HIV testing guidelines. This is despite the fact that they have presented no behavioural data of any kind for the men in their survey, which is disappointing and surprising given that assessing the contemporary pattern of behavioural risk among MSM is a major focus of the GMSS (as stated on the GMSS website). The decision by the authors to exclude all behavioural data from the GMSS from this analysis significantly weakens their paper and their ability to interpret and put the findings in context regarding other work in this field. In addition, the paper is not well written with many grammatical errors and poorly constructed sentences, and there are issues with how the analysis was conducted and the results presented. For these reasons, I believe the paper is not suitable for publication in its current form and the manuscript requires proof reading and major revision before it is submitted again. To help make the paper suitable for publication, I suggest that the issue of the missing behavioural risk data be addressed in one of three ways: 1) Repeat the analysis with the behavioural risk data from GMSS included and interpret the findings accordingly; 2) Refer to a separate publicly available analysis which includes behavioural risk data on the same population in the same year; or 3) Revise and scale back the claims and statements about the behavioural risk of sub-groups of men in this analysis (and how that relates to the testing guidelines) that are made in the discussion and conclusion.

Some comments on specific sections of the paper follow.

ABSTRACT

Behavioural risk of MSM and the BHIVA guidelines are not mentioned in the first three sections of the abstract, yet they are the focus of the conclusion. The conclusion should only refer to the data that was presented in this analysis and given the authors did not present any data on behavioural risk, the conclusions they have made should be revised.

STRENGTHS AND LIMITATIONS OF THIS STUDY

Page 2, line 17. The aim of the study is stated here, rather than whether the study achieved the aim or not. Suggest the authors state whether the study achieved the aim.

Page 2, line 23. It is stated that the study provides a robust critique of the BHIVA guidelines. It does not. This was never stated as an aim of the study in the background or methods sections and, realistically, in order to criticise the BHIVA guidelines the authors would need to present the risk data for the MSM in the study. The lack of behavioural risk data is a clear limitation of this analysis, but it is not stated here. Depending on how the authors address this, they should include it as a limitation in this section if they do not include behavioural data when they revise the paper.

BACKGROUND

Page 3, line 32. The estimate given by PHE in the cited report of 14% of MSM being undiagnosed is a UK-wide figure.

Page 3, lines 35-46. It is not accurate to say that the BHIVA guidelines do not make any allowance for behavioural risk. The recommendation in the guidelines for repeat testing annually or more frequent testing in MSM is based on behavioural risk and the presence or otherwise of symptoms suggestive of HIV infection. Also, people at risk of HIV may not always recall or appreciate their own risk and their recollection of risk events may not be complete. Care providers may not always ask about risk events and clients

may not always report them. Hence, relying mainly on the perception or the reporting of risk to guide HIV testing is less than ideal. While having a blanket recommendation in public health guidelines for a specific risk group in the population may not account for the range of risk behaviour that occurs among sub-groups, there is a potential benefit to recommending that all people within a group have an annual test. Having such a recommendation can help to normalise testing and can facilitate testing where disclosure of potentially stigmatising behaviour could act as a barrier to testing.

Page 4, lines 17-33. It is wrongly stated that the recommendation for annual testing for MSM in the BHIVA guidelines is made regardless of risk. The BHIVA guidelines recommend annual or more frequent testing for MSM if ongoing high risk exposure or clinical symptoms suggestive of seroconversion are present. The reason given for not including behavioural risk data in this analysis is inadequate and not justified on the basis given.

METHODS

Page 4, line 45. Why was data from England only included in the analysis? The GMSS includes data from all over the UK, not just England. Explain why data from the other countries in the UK were excluded from this analysis. The BHIVA guidelines are UK-wide, so if this analysis has been inspired by these guidelines, then it should be UK-wide. Also, the HIV risk reduction messages published in the 2015 Public Health England report are not claimed to be England-specific. Further, state whether there is an agreement between the authors and other groups of researchers to restrict this analysis to England only.

Page 4, line 45. How was the eligibility of men for the survey defined? Was it on the basis of sexual identity or on the basis sexual behaviour or both?

Page 4, line 54. Describe the range of variables included in the survey and provide a copy of the text of the questions as a supplementary file. Explain why excluded variables were dropped from this analysis.

Page 4, line 57. BHIVA's guidelines are UK-wide; they are not specific to England.

Page 5, line 3. The reason given for not including behavioural risk data in this analysis is inadequate and not justified on the basis given. If the authors or the same research team are going to publish an analysis of the behavioural risk data in another paper or report that should be stated.

Page 5, line 14. Rephrase "to avoid the small-cell problem…"

Page 6, lines 5-23. No explanation is given here of how the authors managed missing data. Was it included or excluded in the analysis? How did the authors deal with participants who started, but did not finish a survey? Was their data included or did the authors apply filters to exclude data? Also, how was the study sample for this analysis defined? In the first paragraph of the results section, it is stated that there were two samples in the analysis based upon whether the participant provided answers to the two questions on their testing history – this does not make sense. There is one sample only. The number of men recruited to the survey (over 14,000 who were not known to be HIV-positive) is very large and it should still be possible to exclude data or apply filters and get meaningful results.

Page 6, lines 7-14. Rephrase. Suggest something like "we used logistic regression to assess the associations between demographic variables (age, sexual identity, ethnicity, educational attainment and migrancy) and the dependent variables of never having tested before and not having tested in the last 12 months…".

Page 6, line 16. Clarify what is meant by "understanding the key equity dimensions of access to healthcare".

Page 6, lines 20-28. The levels of significance for the p-values in the Wald and the Chi-square statistical tests are not stated.

Page 6, lines 30-32. If there are a lot of categories, then it may be helpful to conduct the tests and give this data as a supplementary table.

Page 6, line 35. State the manufacturer of the statistical package used for the analysis.

RESULTS

Page 6, lines 41-48. Clearly state here the total number of survey participants, and what the number and percentage was for those who: reported they were HIV-positive; had ever tested; and had tested in the last 12 months. Insert a flowchart which shows the total sample of survey participants and the men who were excluded from this analysis and why. It would also be helpful to give a brief overview here in text of the sample: n (%) identified as gay, a summary statistic for age, n (%) were white, n (%) had a high level of educational attainment, and n (%) were UK-born. The participant characteristics that are currently tabulated in Table 1 in the column "percent of the sample" would be better presented in their own table reporting participant characteristics, as is standard practice in papers reporting the outcomes of large surveys. Give the number for each line item with the percentage in brackets.

Page 6, line 52. See comment above. It does not make sense to present the survey participants in this analysis as two separate samples based on whether there is data available for two testing history questions.

Page 7, line 39. See comment above. It does not make sense to present the survey participants in this analysis as two separate samples based on whether there is data available for two testing history questions.

Page 7, line 18. Rephrase. Suggest "compared to white men, black men and men from other ethnic groups were less likely to have ever tested for HIV, while Asian men were not significantly different."

Page 8, line 21. Rephrase. Suggest "these men had the lowest reported interest in testing in GUM clinics of all groups…"

Page 8, lines 33-41. No statistical tests were performed for the data reported in this table, so it is not accurate to say that certain groups had a greater preference than other groups for a specific mode or location of testing (even with the very large sample size). Performing the tests would give a clearer picture regarding the data in this table.

Results in Tables 1 and 2. It would be clearer if the bivariate and multivariate analysis (the odds ratios and adjusted odds ratios) for each outcome variable were presented in separate tables. Suggest presenting the data for ever and never testing in one table and the data for having tested recently in the last 12 months or not in another table. In the first table, give the number and the percentage of subjects that have ever and never tested for each demographic variable category. Similarly, in the second table, give the number and the percentage of subjects that have and have not tested in the last 12 months for each demographic variable category.

Page 10, line 3. The top row of table 4 has become separated from that table.

Page 10, Table 3 on testing preferences. Give the number and the percentage in brackets for each line item in the table. Degrees of freedom not needed. In the fourth column on the right the p-values are given as being less than zero. Choose a significant figure to which p-values are to be quoted and stick to it in this table and the others. Suggest 0.01 or 0.001.

Page 11, Table 4. Units are not given – are they all percentages? Suggest giving both n and the percentage. Some cells have data that is quoted to one decimal place and some are quoted to two decimal places. Explain the abbreviations in the top row (e.g. GUM) as footnotes to the table. Suggest that significance tests are conducted for the data presented in this table and that the findings are presented as separate supplementary tables for each testing mode and location. Claims made about the data would be stronger if significance test were performed.

DISCUSSION

Page 11, lines 39-41. The authors state that their findings clearly show that the BHIVA guidelines are not being followed. Was assessing compliance with the guidelines an objective of their analysis? If so, why wasn't that stated in the background or methods? If not, why mention this in the first paragraph of the discussion? Also, are they suggesting that the guidelines should be rigidly followed by all MSM and care providers? Is it not possible and indeed preferable that MSM and care providers might use their common sense to decide when and how often to engage in testing based on their or the client's situation?

Page 11, lines 44-50. How does the demographic profile of the MSM in this survey compare to other surveys of MSM in the UK, in terms of the findings being generalisable to other MSM? Also, how do the findings of this analysis compare to the testing history data reported by MSM in the GMSS in 2010 and 2008? What is the trend in these data over time?

Page 11, line 53. English data was presented only, not UK data.

Page 11, line 55. Rephrase. Suggest "men under the age of 20 years were less likely…"

Page 12, lines 3-14. The authors make claims about the testing frequency of the subjects in the survey, despite the fact that testing frequency was not assessed in the presented data. Ever testing and recent testing were assessed, not testing frequency. Assessing frequency would require questions about how often men test within a specific time period.

Page 12, lines 25-38. Suggest the authors comment on the potential effect of stigma and discrimination on testing practices among non-gay identified men.

Page 12, lines 44-50. Rephrase. This text is confusing. Suggest the authors state that lower levels of educational attainment remained associated with never testing in multivariate analysis.

Page 13, lines 12-17. Providing community-based rapid testing for men with higher behavioural risk could be a very efficient use of resources. It all depends on the risk profile of the users of the service.

Page 13, lines 33-46. These findings by themselves do not show that the guidelines are inappropriate. The authors did not assess or report the behavioural risk of the men in the survey, so whether the guidelines are inappropriate or not remains to be seen. Younger men and non-gay identifying MSM have been shown to have higher behavioural risk in studies internationally, and given their risk in this survey has not been reported, it is not appropriate to suggest here that they should not be a focus for testing programs.

Page 13, line 49. "These groups are not reflective of those most likely to have HIV in the UK…" Provide citations for this statement.

Page 13, lines 55-58. Reference is made to a modelling analysis where targeting men with higher behavioural risk is presented as a better option to testing all men with the same frequency. How is it possible to know who has higher risk without reporting behavioural risk? The authors have chosen not to report risk data, despite it

	being collected by the survey, and then spend a considerable amount of their discussion speculating about the risk of groups of men in their survey.
	Page 14, lines 16-18. Whether or not self-testing is more or less expensive than facility-based testing methods depends on from which perspective the costs are assessed, the price of self-tests and whether self-testing is subsidised.
	LIMITATIONS
	In addition to the limitations listed, not reporting behavioural risk for the men in the survey, despite that being collected by the survey, is the most important limitation. Also, the analysis reports data from England only, which limits its generalisability to MSM UK-wide.
	CONCLUSIONS
	The conclusion section should only refer to the data that was presented in this analysis and given the authors did not present any data on behavioural risk, the conclusions they have made about behavioural risk should be revised. Also, the conclusion refers to the appropriateness of the BHIVA and PHE testing guidelines despite this not being given as an aim or objective of the analysis.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

I have some concerns about the overall quality of the manuscript, which the previous reviewers had identified (rather harshly) but it does need proof reading and could be improved in terms of tightening the language throughout. Table 4, for example, is split by preceding tables, references are poorly spelled. To me, the overall analytic plan underpinning the data presented, doesnt quite match the main narrative that binds the paper together. The discussion particularly seems slightly removed from the results presented.

The fundamental idea behind the manuscript is great and I was very excited to have the opportunity to review it, as it promises to be highly relevant for future testing policy and practice in England and elsewhere. I do how ever have a few concerns, at least one of them major but most of them minor I worry the authors will find some of my comments annoying given not all of them were identified within the first round of reviews. But I am really keen to see the most being made out of this key data set.

We thank the reviewer for taking time to provide such helpful comments. We have edited the paper and provided new analyses which we hope will satisfy the majority of his concerns. In particular we have now included a sexual behaviour variable alongside our demographic ones in both univariate and multivariate analyses which should respond to his most fundamental concern. We have also paid more attention to the psychological barriers to testing and efforts to counter these in our background section.

Major
I think there is an assumption implicit within the manuscript, that implies that the UK guidance directly shapes testing behaviours, or that modes of delivery are the only interventions that seek to intervene with barriers to testing. I think being more explicit about the interventions which have been delivered in the recent past which fill in the middle ground between UK policy and MSM themselves would be helpful. It would also be useful to acknowledge the importance of psychological barriers to testing which explain far more variance in testing behaviours than changes to structural barriers (such as expanding the repertoire of test delivery), so I think it needs to be clear that on quite a fundamental level, despite innovations in testing delivery, psychological barriers to testing still need to be reduced. Secondary questions of which testing option suits who best (the arena of this paper) then come in to

play.

We thank the reviewer for this very helpful comment. Having reflected on our rationale for producing this manuscript we agree with Dr. Flowers and paid greater attention to health promotion efforts to boost testing. This appears in our third paragraph of the background section which now reads:

Factors mediating MSM's decisions to test are complex and varied. Significant barriers to encouraging MSM to test for HIV exist, particularly in relation to psychosocial needs and negative emotional responses to testing [5]. In the UK policy makers have focused on creating demand for testing services through demand side interventions including policy and health promotion interventions addressing some of these factors, and by encouraging regular testing through national campaigns [6]. Efforts to boost supply have also been central in attempts to increase the rate and frequency of testing among MSM. Models of delivery for HIV testing have also evolved with a key aim of reducing barriers to testing among most at risk populations (for a comprehensive discussion on barriers to HIV testing see [7]). While hospital-based outpatient HIV testing remains the norm, public health provision has focused increasingly on providing a wider range of settings for HIV testing. Initial expansion focused on opt-out as opposed to opt-in protocols in clinics, and providing HIV testing services within the community [7, 8] and, more recently, providing opportunities for self-administered testing methods including self-sampling and self-testing [9].
This has also been drawn out in our discussion section and our interpretation of our findings.

My most major concern relates to the operationalisation of the BHIVA/PHE guidance which as I understand it has two important elements, one: all MSM test annually (addressed and critiqued within the manuscript), and two that, those having sex UAI with new or casual partners test more often (every three months). This latter element is not addressed within the manuscript but is fairly fundamental. between page 3 and 4 this distinction is lost, and the lack of analysis with sexual behaviour variables means this remains unexplored. The paper would be so much more useful and citeable if it also included analyses of this second issue. Then a critique of BHIVA/PHE policy seems apt. Im not sure if the team collected data about frequency/regularity of testing but it looks like a proxy measure could be used from the 5 categories of time frame for most recent test. I think even an analysis of those higher risk MSM and their testing patterns and testing preferences would be more valuable than some of the current manuscript with its only partial analytic focus and dialogue with testing recommendations. I hope this makes sense and am not quite sure why previous reviewers did not pick this point up.

We have sometimes used comparisons of ever tested/tested within the last 12 months and tested within the previous 3 months as a way of trying to address some of these issues, doing the same within the whole sample, or within those men reporting new/casual UAI partners might be worth exploring too

We thank the reviewer for this helpful suggestion. While our initial intention was to critique the 12 month recommendation of the guidelines, we agree with both reviewers that including a behavioural variable is more useful for our intended audience. We also feel that including these variables allows us to engage more fully with the guidelines as the reviewer suggests. We therefore have included a variable representing whether or not men reported having two or more non-steady sex partners with whom condoms were not used for anal sex and have re-run our multivariate models with these included. We have also produced a new analysis examining testing behaviour in the preceding six months for these men. Updated results exploring these can be found in tables 1-3 and throughout the results section.

Another key issue is the approach to missing data, unless I have missed it, i couldnt see how missing data is reported and if analytic strategies have been adopted to deal with it. I know in some of the on-

line surveys I am involved in missing data levels can be quite high.

We have now also included information on our approach to missing data in this manuscript. Levels of missing data for the variables which we examine are low, and therefore we have not attempted to use corrective mechanisms. The section describing this now reads:
Missing data across all variables was less than five percent of observations. We decided this level of missing data was acceptable for a community based cross-sectional survey and did not make attempts to use corrective statistical mechanisms.

Minor
Page 3, explain 'test and treat' model
We thank the reviewer for this suggestion, we have now clarified the paragraph which now reads:
Both globally and in the UK, HIV prevention is moving toward a test and treat model. With This approach evolved from the with recognition that the majority of new HIV infections among men who have sex with men (MSM) are passed from those who are unaware of their infection, and that treating HIV positive individuals early drastically reduces their infectivity [1, 2], sexual Sexual health promotion now has a major focus on reducing the time between infection and diagnosis through increasing rates of testing, as well as providing earlier HIV treatment to those found to be positive [1, – 2]. Essentially this approach emphasises reducing the amount of undiagnosed HIV within the population in order to reduce community viral load (and therefore onward transmission) while also preventing illness in individuals who have HIV [3].

Page 3, lines 41-43, 'treads an awkward line' seems an odd phrase and I am not sure that the two elements of UK policy are so problematic. The two groups are not exclcusive and static and we'd expect to see variability across the life span in terms of risk.

This section has now been deleted as part of our editing process.

page 4, line 22, 'their service' seems and odd phrase to use.

We thank the reviewer for this suggestion. The sentence now reads:
We identify which groups of MSM in England are less likely to have tested for HIV and their service preferences for future tests by testing model, in order to inform health promotion programmes.

Page 8. line 30, given the % of never testers, it would be interesting to hear the authors views on intentions to test in the future.

We have now provided this, indicating that this demonstrates that testing is highly acceptable to men even if they have not tested.

Page 12, lines 7-15, I got a bit lost in these lines, the text mentions testing frequency (but see above comments) I don think this was measured, so is it a proxy measure? Also given the cross-sectional nature of the data are the conclusions about age trends in HIV testing not being confounded with both generational issues and indeed with the temporal perspectives of the HIV epidemic. My head hurt a bit from trying to think this through but I dont think the data do support the claims being made here. Maybe clarifying the interpretation would help?

We have now deleted this sentence as we agree that our data does not support this interpretation.

Pg 15, 'important educational component' seems to miss the point about knowledge not being sufficient to change behaviour. As mentioned above, and core to potential intervention design, acknowledging the psychological barriers to testing is vital (Lots of reviews highlight this, or here in

the UK, some of my own work does this. Flowers P, Knussen C, McDaid L, Li J (2013). Has testing been normalised?: An analysis of changes in barriers to HIV testing among men who have sex with men between 2000 and 2010 in Scotland, UK: HIV Medicine, Feb;14 (2):92-8. DOI: 10.1111/j.1468-1293.2012.01041. We've got some online data that supports these findings too and reflects many of the demographic associations report here too.

We thank the reviewer for this suggestion and have now included a sentence expanding on this point.

page 12, line 57 - was previous testing site assessed? GIven the range of testing options? Given the centrality of habit in predicting future behaviour it would be good to hear more of this if it was collected

We thank the reviewer for this comment, upon reflection we have decided not to include this information. While this is certainly a component of testing decision making, we felt it better to restrict this analysis to prospective intention. Further, we feel that we are already presenting a significant amount of data and including this will detract from the overall focus of the manuscript.

page 13, lines 12-17, community based rapid HIV testing were most popular within the groups most likely to test. Which variable is used to reflect likelihood of testing? I couldn't work this out .

We thank the reviewer for this comment and the opportunity to clarify this in our manuscript. The section now reads:
Importantly, community based rapid HIV testing services were most popular with the demographic groups who were most likely to have ever tested (and to a lesser degree among men reporting higher risk), indicating that expanding access to these services is unlikely to be efficient if policy goals include meeting the testing needs of men that would not otherwise be testing or expanding testing to those most likely to have undiagnosed HIV.

page 14, 'service delivery puzzle' sounds a little odd.

We have now changed this to service mix which we hope is clearer.

the final sentence before limitations, doesnt seem to follow on from anything. Given the analysis doesnt look at sexual behaviour it feels an odd way to end the paragraph.
We thank the reviewer for this helpful suggestion. We feel that in responding to Dr Flowers' first comment about psychological barriers to testing, and by editing this paragraph that it is much clearer.

To summaries, for me the paper as it stands still has great potential to be an influential and useful contribution to the field but a number of problems remain to be addressed. To me, a more coherent analytic focus is needed that matches one of the main research questions; dialogue with the BHIVA/PHE guidance with the approach to key analysis (and incorporating an assessment of the second clause for men at some risk), i think the paper needs some further detail regarding which interventions have been delivered within England and some indication of their content (in relation to barriers to testing), I think the paper does need to address the psychological and social barriers to testing that persist and will remain obstinate, despite further choices in testing options. The manuscript overall needs to be improved with spell checks and some reigning in of the interpretations of the associations reported.

I have really enjoyed the manuscript but am not so sure that other people, less passionate about the topic, would be willing to interpret and reflect on what has been presented. I hope this is useful feedback and not too annoying, given I am suggesting new analytic strategies on a second round of reviewing.

Reviewer: 2

Reviewer Name

Damian Patrick Conway

Institution and Country

We thank the reviewer for his very thorough and insightful review of our paper. We have made significant amendments and have included behavioural risk data. We have provided responses to his comments below, and in some places referenced our above response to reviewer one for brevity's sake.

Please leave your comments for the authors below
The authors have conducted an analysis of data from the 2014 Gay Men's Sex Survey (GMSS) in the UK. The title of the manuscript and the introduction of the paper claim that this analysis has been conducted to describe and assess the HIV testing history and future HIV testing preferences among gay, bisexual and other men who have sex with men (MSM) in England. As the authors state, assessing HIV testing patters among British MSM is a very important matter in the current HIV test and treat era. While the data presented in the paper does describe the HIV testing history and preferences of these men, from reading the paper it appears that the authors were also seeking to answer these questions: 1) Is the HIV testing history of English MSM consistent with national UK HIV testing guidelines? 2) Are the current UK HIV testing guidelines appropriate? The authors have sought to present their discussion and the interpretation of their findings around whether certain groups of MSM have sufficient behavioural risk to justify the current HIV testing guidelines. This is despite the fact that they have presented no behavioural data of any kind for the men in their survey, which is disappointing and surprising given that assessing the contemporary pattern of behavioural risk among MSM is a major focus of the GMSS (as stated on the GMSS website). The decision by the authors to exclude all behavioural data from the GMSS from this analysis significantly weakens their paper and their ability to interpret and put the findings in context regarding other work in this field. In addition, the paper is not well written with many grammatical errors and poorly constructed sentences, and there are issues with how the analysis was conducted and the results presented. For these reasons, I believe the paper is not suitable for publication in its current form and the manuscript requires proof reading and major revision before it is submitted again. To help make the paper suitable for publication, I suggest that the issue of the missing behavioural risk data be addressed in one of three ways: 1) Repeat the analysis with the behavioural risk data from GMSS included and interpret the findings accordingly; 2) Refer to a separate publicly available analysis which includes behavioural risk data on the same population in the same year; or 3) Revise and scale back the claims and statements about the behavioural risk of sub-groups of men in this analysis (and how that relates to the testing guidelines) that are made in the discussion and conclusion.

We thank the reviewer for this overarching comment and providing us the opportunity to reflect on our aspirations for this paper. We have taken this opportunity to provide additional analyses incorporating a risk variable which we hope should satisfy this concern. For a full description of our approach please see comment two to reviewer one.

Some comments on specific sections of the paper follow.
ABSTRACT
Behavioural risk of MSM and the BHIVA guidelines are not mentioned in the first three sections of the abstract, yet they are the focus of the conclusion. The conclusion should only refer to the data that was presented in this analysis and given the authors did not present any data on behavioural risk, the

conclusions they have made should be revised.

We thank the reviewer for this comment and have now included the policy in the abstract. We have also included behavioural data as requested.

STRENGTHS AND LIMITATIONS OF THIS STUDY
Page 2, line 17. The aim of the study is stated here, rather than whether the study achieved the aim or not. Suggest the authors state whether the study achieved the aim.
We thank the reviewer for this comment. This section now reads:
This study identifies which groups of MSM in England are less likely to have tested for HIV ever and in the preceding 12 months and their service preferences for future tests.

Page 2, line 23. It is stated that the study provides a robust critique of the BHIVA guidelines. It does not. This was never stated as an aim of the study in the background or methods sections and, realistically, in order to criticise the BHIVA guidelines the authors would need to present the risk data for the MSM in the study. The lack of behavioural risk data is a clear limitation of this analysis, but it is not stated here. Depending on how the authors address this, they should include it as a limitation in this section if they do not include behavioural data when they revise the paper.

We feel that in the light of changes made to this manuscript that the study now does provide a robust critique of the guidelines.

BACKGROUND
Page 3, line 32. The estimate given by PHE in the cited report of 14% of MSM being undiagnosed is a UK-wide figure.

We thank the reviewer for this comment, we have now clarified that this is a UK wide figure.

Page 3, lines 35-46. It is not accurate to say that the BHIVA guidelines do not make any allowance for behavioural risk. The recommendation in the guidelines for repeat testing annually or more frequent testing in MSM is based on behavioural risk and the presence or otherwise of symptoms suggestive of HIV infection. Also, people at risk of HIV may not always recall or appreciate their own risk and their recollection of risk events may not be complete. Care providers may not always ask about risk events and clients may not always report them. Hence, relying mainly on the perception or the reporting of risk to guide HIV testing is less than ideal. While having a blanket recommendation in public health guidelines for a specific risk group in the population may not account for the range of risk behaviour that occurs among sub-groups, there is a potential benefit to recommending that all people within a group have an annual test. Having such a recommendation can help to normalise testing and can facilitate testing where disclosure of potentially stigmatising behaviour could act as a barrier to testing.

We feel that this has now been addressed by our inclusion of behavioural variables.

Page 4, lines 17-33. It is wrongly stated that the recommendation for annual testing for MSM in the BHIVA guidelines is made regardless of risk. The BHIVA guidelines recommend annual or more frequent testing for MSM if ongoing high risk exposure or clinical symptoms suggestive of seroconversion are present. The reason given for not including behavioural risk data in this analysis is inadequate and not justified on the basis given.

We feel that this has now been addressed by our inclusion of behavioural variables.

METHODS

Page 4, line 45. Why was data from England only included in the analysis? The GMSS includes data from all over the UK, not just England. Explain why data from the other countries in the UK were excluded from this analysis. The BHIVA guidelines are UK-wide, so if this analysis has been inspired by these guidelines, then it should be UK-wide. Also, the HIV risk reduction messages published in the 2015 Public Health England report are not claimed to be England-specific. Further, state whether there is an agreement between the authors and other groups of researchers to restrict this analysis to England only.

The English National Health Service operates independently from the health services of Wales, Scotland and Nothern Ireland. Therefore health interventions tend to target only one of these countries at their most broad. We have restricted our analysis to England to reflect this and have provided a reference to support this approach.

Page 4, line 45. How was the eligibility of men for the survey defined? Was it on the basis of sexual identity or on the basis sexual behaviour or both?

We thank the reviewer for this comment. We have now included this in our text which reads:
The survey recruited men who reported attraction to other men who were aged 16 or older and living in England between August and November of 2014.

Page 4, line 54. Describe the range of variables included in the survey and provide a copy of the text of the questions as a supplementary file. Explain why excluded variables were dropped from this analysis.

We thank the reviewer for this comment. We have decided upon reflection not to provide an explanation of why variables were dropped from this analysis as we believe this is made implicit through our aims and objectives. The questionnaire itself is open access and is freely available on Sigma Research's website (http://sigmaresearch.org.uk/files/GMSS_2014_Questionnaire.pdf)

Page 4, line 57. BHIVA's guidelines are UK-wide; they are not specific to England.

We thank the reviewer for this comment, we have removed this section however as we have now included behavioural variables in this analysis.

Page 5, line 3. The reason given for not including behavioural risk data in this analysis is inadequate and not justified on the basis given. If the authors or the same research team are going to publish an analysis of the behavioural risk data in another paper or report that should be stated.

We have now included behavioural analysis in this manuscript.

Page 5, line 14. Rephrase "to avoid the small-cell problem…"

Thank you for this comment. We have rephrased this section as per the reviewer's request. It now reads: Ethnicity was recoded from standard UK ethnicity codes into four categories to avoid issues with having many categories with small numbers of observations…

Page 6, lines 5-23. No explanation is given here of how the authors managed missing data. Was it included or excluded in the analysis? How did the authors deal with participants who started, but did not finish a survey? Was their data included or did the authors apply filters to exclude data? Also, how was the study sample for this analysis defined? In the first paragraph of the results section, it is stated that there were two samples in the analysis based upon whether the participant provided answers to the two questions on their testing history – this does not make sense. There is one sample only. The

number of men recruited to the survey (over 14,000 who were not known to be HIV-positive) is very large and it should still be possible to exclude data or apply filters and get meaningful results.

We have now provided this information. The section in question now reads: Missing data across all variables was less than five percent of observations. We decided this level of missing data was acceptable for a community based cross-sectional survey and did not make attempts to use corrective statistical mechanisms.

Page 6, lines 7-14. Rephrase. Suggest something like "we used logistic regression to assess the associations between demographic variables (age, sexual identity, ethnicity, educational attainment and migrancy) and the dependent variables of never having tested before and not having tested in the last 12 months…".

We thank the reviewer for this suggestion. The section now reads: We estimated a set of logistic regression models to understand how demographic and behavioural variables were associated with the dependent variables of not having tested for HIV in the past 12 months and never having tested for HIV.

Page 6, line 16. Clarify what is meant by "understanding the key equity dimensions of access to healthcare".

We have provided this information and provided a key reference.

Page 6, lines 20-28. The levels of significance for the p-values in the Wald and the Chi-square statistical tests are not stated.
We provide this detail in the results.
Page 6, lines 30-32. If there are a lot of categories, then it may be helpful to conduct the tests and give this data as a supplementary table.
We thank the reviewer for this comment, however we have made a decision to keep this data as a main table in our manuscript as we believe it is key.

Page 6, line 35. State the manufacturer of the statistical package used for the analysis.

We have now clarified that Stata 13 is made by Statacorp.

RESULTS
Page 6, lines 41-48. Clearly state here the total number of survey participants, and what the number and percentage was for those who: reported they were HIV-positive; had ever tested; and had tested in the last 12 months. Insert a flowchart which shows the total sample of survey participants and the men who were excluded from this analysis and why. It would also be helpful to give a brief overview here in text of the sample: n (%) identified as gay, a summary statistic for age, n (%) were white, n (%) had a high level of educational attainment, and n (%) were UK-born. The participant characteristics that are currently tabulated in Table 1 in the column "percent of the sample" would be better presented in their own table reporting participant characteristics, as is standard practice in papers reporting the outcomes of large surveys. Give the number for each line item with the percentage in brackets.

We thank the reviewer for this helpful comment. We have now produced a flowchart demonstrating our participant exclusions. We have decided not to include a further table describing these demographic details as we feel they are better presented in our current tables.

Page 6, line 52. See comment above. It does not make sense to present the survey participants in

this analysis as two separate samples based on whether there is data available for two testing history questions.

Page 7, line 39. See comment above. It does not make sense to present the survey participants in this analysis as two separate samples based on whether there is data available for two testing history questions.

We have now deleted this sentence.

Page 7, line 18. Rephrase. Suggest "compared to white men, black men and men from other ethnic groups were less likely to have ever tested for HIV, while Asian men were not significantly different."

We thank the reviewer for this comment. While we initially made the change, we have reverted to the original which is more consistent with the language in the rest of the paper.

Page 8, line 21. Rephrase. Suggest "these men had the lowest reported interest in testing in GUM clinics of all groups…"

We thank the reviewer for this comment. This now reads: These men had the lowest reported interest in testing in GUM settings of all groups included in this analysis.

Page 8, lines 33-41. No statistical tests were performed for the data reported in this table, so it is not accurate to say that certain groups had a greater preference than other groups for a specific mode or location of testing (even with the very large sample size). Performing the tests would give a clearer picture regarding the data in this table.

Upon consideration we have decided not to perform statistical tests on this section of the data because of the very high number of categories. We have therefore softened our language changing 'observed' to 'reported' which we feel is more accurate.

Results in Tables 1 and 2. It would be clearer if the bivariate and multivariate analysis (the odds ratios and adjusted odds ratios) for each outcome variable were presented in separate tables. Suggest presenting the data for ever and never testing in one table and the data for having tested recently in the last 12 months or not in another table. In the first table, give the number and the percentage of subjects that have ever and never tested for each demographic variable category. Similarly, in the second table, give the number and the percentage of subjects that have and have not tested in the last 12 months for each demographic variable category.
We thank the reviewer for this comment but feel that the way that the data is currently presented is clear and is more useful than the changes suggested.

Page 10, line 3. The top row of table 4 has become separated from that table.
This has now been corrected.

Page 10, Table 3 on testing preferences. Give the number and the percentage in brackets for each line item in the table. Degrees of freedom not needed. In the fourth column on the right the p-values are given as being less than zero. Choose a significant figure to which p-values are to be quoted and stick to it in this table and the others. Suggest 0.01 or 0.001.

We thank the reviewer for this comment. We have now removed degrees of freedom from this table and have harmonised reporting of p-values across all tables where relevant.

Page 11, Table 4. Units are not given – are they all percentages? Suggest giving both n and the percentage. Some cells have data that is quoted to one decimal place and some are quoted to two

decimal places. Explain the abbreviations in the top row (e.g. GUM) as footnotes to the table. Suggest that significance tests are conducted for the data presented in this table and that the findings are presented as separate supplementary tables for each testing mode and location. Claims made about the data would be stronger if significance test were performed.

We thank the reviewer for this comment. We have now made it clear that all data in this table are in percentages. We have also harmonised the percentages.

As stated above, we have decided not to include significance tests and have correspondingly softened our interpretations.

DISCUSSION
Page 11, lines 39-41. The authors state that their findings clearly show that the BHIVA guidelines are not being followed. Was assessing compliance with the guidelines an objective of their analysis? If so, why wasn't that stated in the background or methods? If not, why mention this in the first paragraph of the discussion? Also, are they suggesting that the guidelines should be rigidly followed by all MSM and care providers? Is it not possible and indeed preferable that MSM and care providers might use their common sense to decide when and how often to engage in testing based on their or the client's situation?

We feel that this comment has now been addressed through our inclusion of behavioural data and our changes to interpretation.

Page 11, lines 44-50. How does the demographic profile of the MSM in this survey compare to other surveys of MSM in the UK, in terms of the findings being generalisable to other MSM? Also, how do the findings of this analysis compare to the testing history data reported by MSM in the GMSS in 2010 and 2008? What is the trend in these data over time?

We have now made it clear in the text that the reference provided discusses this detail in more depth. We feel that this discussion is not relevant here given its availability elsewhere.

Page 11, line 53. English data was presented only, not UK data.

This has been amended.

Page 11, line 55. Rephrase. Suggest "men under the age of 20 years were less likely…"

Thank you- we have now amended.

Page 12, lines 3-14. The authors make claims about the testing frequency of the subjects in the survey, despite the fact that testing frequency was not assessed in the presented data. Ever testing and recent testing were assessed, not testing frequency. Assessing frequency would require questions about how often men test within a specific time period.

We thanks the author for this comment. This claim has now been deleted.

Page 12, lines 25-38. Suggest the authors comment on the potential effect of stigma and discrimination on testing practices among non-gay identified men.

In our editing of the manuscript we have now provided greater focus on the psychological barriers to testing in both the background and discussion sections.

Page 12, lines 44-50. Rephrase. This text is confusing. Suggest the authors state that lower levels of educational attainment remained associated with never testing in multivariate analysis.

We have now provided further comment interpreting these results and feel that this has been dealt with.

Page 13, lines 33-46. These findings by themselves do not show that the guidelines are inappropriate. The authors did not assess or report the behavioural risk of the men in the survey, so whether the guidelines are inappropriate or not remains to be seen. Younger men and non-gay identifying MSM have been shown to have higher behavioural risk in studies internationally, and given their risk in this survey has not been reported, it is not appropriate to suggest here that they should not be a focus for testing programs.

This has now been dealt with following our inclusion of risk behaviour data.

Page 13, line 49. "These groups are not reflective of those most likely to have HIV in the UK…" Provide citations for this statement.

We have provided citations as requested.

Page 13, lines 55-58. Reference is made to a modelling analysis where targeting men with higher behavioural risk is presented as a better option to testing all men with the same frequency. How is it possible to know who has higher risk without reporting behavioural risk? The authors have chosen not to report risk data, despite it being collected by the survey, and then spend a considerable amount of their discussion speculating about the risk of groups of men in their survey.

This has now been dealt with following our inclusion of risk behaviour data.

Page 14, lines 16-18. Whether or not self-testing is more or less expensive than facility-based testing methods depends on from which perspective the costs are assessed, the price of self-tests and whether self-testing is subsidised.

We have now clarified that we mean per unit cost when discussing costs of testing options here.

LIMITATIONS
In addition to the limitations listed, not reporting behavioural risk for the men in the survey, despite that being collected by the survey, is the most important limitation. Also, the analysis reports data from England only, which limits its generalisability to MSM UK-wide.

This has now been dealt with following our inclusion of risk behaviour data.

CONCLUSIONS
The conclusion section should only refer to the data that was presented in this analysis and given the authors did not present any data on behavioural risk, the conclusions they have made about behavioural risk should be revised. Also, the conclusion refers to the appropriateness of the BHIVA and PHE testing guidelines despite this not being given as an aim or objective of the analysis.
This has now been dealt with following our inclusion of risk behaviour data.

VERSION 2 – REVIEW

REVIEWER	Paul Flowers Glasgow Caledonian University Scotland Similar research conducted in Scotland
REVIEW RETURNED	25-Jun-2016

GENERAL COMMENTS	Really enjoyed the revised manuscript and it is very useful