Thank you for considering our manuscript. We appreciate the reviewers' comments and have incorporated much of their feedback into the manuscript.

Reviewer: 1
Recommendation:
Comments:
Thanks for the opportunity to review this manuscript describing the results of a large, randomized clinical trial evaluating the effect of three HIV test consent methods on acceptance. Understanding the effect of various forms of consent on acceptance of HIV screening remains an important topic and I believe this study advances, to an extent, our understanding of this process. Acknowledged by the authors, their definition of opt-in and active-choice is subtle, and I believe most who think of opt-in are actually thinking of active-choice.

In the Discussion, the authors suggest clarifying opt-in to represent no testing unless a test is specifically requested by the patient; this seems to blur the distinction between identification of patients for testing and actual consent. As such, I have difficulty with this distinction as 'opt-in' refers to a form of consent and not a method for identifying someone (in this case, self-identification) for actual testing. Implicit in how the authors suggest thinking about opt-in is that testing is driven by the patient, although it isn't clear how this would work with various forms of nontargeted or targeted screening.

We recognize that our use of opt-in is a source of potential confusion. Our use of the terms opt-in, opt-out, and defaults are consistent with definitions used in other, broader literature spanning many different areas, such as organ donation, email marketing, and retirement savings. We believe that it is important that the terminology regarding defaults in HIV screening becomes consistent with that of other fields. As it stands in the consensus statement on nomenclature and definitions, opt-in is defined so as to include both what we present as opt-in as well as active-choice testing. We find a difference in test acceptance rates according to these treatment assignments that we believe is clinically significant; this difference is consistent with findings from research outside of HIV testing on the effects of defaults. Hence our recommendation to distinguish these two approaches to HIV test offers.

To minimize potential confusion, we have clarified what we mean by "opt-in testing" to be limited to the setting in which patients are made aware of testing, and informed that they can request a test. This stands in contrast to a scheme in which nothing is presented by the clinician but the patient can request a test, which we agree is better-characterized as patient-initiated testing. We have otherwise left our terminology as is, but would change it if the editors prefer.

The Introduction is too long and much of what's written could be moved to the Discussion. The authors should also describe their principal hypotheses of the study in the last paragraph of the Introduction. As such, the analytic plan and sample size should be aligned with the hypotheses. Additionally, this manuscript should be reported in accordance with CONSORT, and this should be explicitly stated.

We have removed some of our introductory statements and moved others to the discussion section. We have added our primary hypotheses in the final paragraph of the introduction, and these are now better aligned with the described methods. We have submitted our CONSORT checklist, and noted that the manuscript is prepared in accordance with CONSORT guidelines

I have a deep methodological concern related to consent and blinding. The authors state that patients were blinded until debriefed and after HIV testing and completion of the questionnaire; unfortunately, this section of the methods isn't clear, and implies, unconventionally, that consent was obtained after entry into the study. If so, isn't it possible (even likely) that post-randomization exclusions occurred, thus critically biasing participation? If, on the other hand, consent was obtained before testing, wouldn't this form of unblinding result in a critical flaw since understanding the goals of the study would likely influence acceptance? A much deeper understanding of the methods is warranted.

Similarly, the authors should provide more detail about how randomization actually occurred, and how allocation concealment was maintained? Without these critical features, it would be difficult to know the effect of the interventions.

The manuscript now contains further detail on critical aspects of the intervention. We randomized patients and exposed them to treatment assignments prior to study consent; this approach was approved by the relevant institutional review board. While this is unconventional, it was felt that this design would yield the most generalizable, least biased results. Approximately 80% of patients consented to study inclusion; patients who refused consent were excluded from the study analysis, but as indicated in Figure 1 the refusal rates were similar across arms.

Treatment assignments were not concealed from the research staff prior to approaching patients. We cannot exclude the possibility that research staff occasionally violated protocol and approached patients in a biased manner; however, given our consistent results across each research staff (see Figure 3) and successful randomization (see Table 1), we feel that it is quite unlikely that research staff selected patients in a manner that biased our results.

Other aspects of the methods are confusing as well. The authors describe reporting other results (elsewhere) where patients were randomized to the questionnaire (it isn't clear what's in the questionnaire), either before or after HIV testing, using a factorial design, and another sample of patients being compensated. This significantly confuses the description of the methods of this study, and I would suggest a thorough rewrite of the methods for clarity, consistency, and completeness.

We have edited our methods section to clarify the study protocol, particularly the questionnaire-related aspects that are directly relevant to this manuscript.

It seems the research assistants may have offered the HIV testing, using one of the three approaches, but this isn't clear as well. Although the investigators used standard scripting, they did not explicitly describe who actually performed the offer.
It would be useful to better understand how risk assessment occurred and whether patients knew their risk profile before being offered testing. Figure 2 is fascinating in that patients at increasing levels of risk were also more likely to accept testing, but it isn't clear if this was a product of them

knowing their risk.

Research staff offered the HIV tests to patients in all treatment groups. We have clarified this portion of the manuscript.

No risk assessment was done during the course of the patient's care. Risk-related data were obtained from the questionnaire, but research staff did not see this information, and no feedback was given to patients regarding their level of risk. It is possible that by performing the questionnaire they reflected upon their risk. However, we found that patients at increasing levels of risk were more likely to test whether they were offered the test before or after completing the questionnaire. We present these data as a figure in the appendix.

We appreciate the authors' use of the Denver HIV Risk Score (DHRS) in quantifying risk; they should know a larger, national validation of the score was recently published in JAIDS; also, it would be useful to know how the investigators defined 'low', 'intermediate', and 'high' risk in the context of the DHRS.

Thank you for alerting us to this new literature. We classified patients with a score less than 20 as low risk, those scoring 20-39 as intermediate risk, and those scoring 40 or higher as high risk (we have added these definitions to our methods section). There are of course no objectively "correct" cutoffs; we have chosen cutoffs that we believe are of most relevance to a broad audience of interested clinicians and policymakers.

Reviewer: 2

Recommendation:

Comments:
Thank you for the opportunity to review your manuscript. It is helpful to the HIV testing research community that you attempted to directly compare three different types of test "offering" in the emergency medicine setting.

I have several strong concerns about the manuscript and the study, as follows:

1. Few clinicians and researchers would agree with the terminology for HIV testing "offers" as you presented them. After the "opt-out" approach was recommended for HIV testing in US healthcare settings by the CDC, "opt-in" was distinguished from "opt-out" by the removal of the need to voice a response to a testing "offer". Your term of "active choice" is what most people in the scientific community would use as the "opt-in" approach. That is, patients are offered testing and a reply of yes or no is elicited. So, I would remove the term "active choice" as this is not a term used by any other group. Instead, label your current "active choice" group as "opt in" ("Would you like a test today?") and label your "opt-in" group ("You can let me, your nurse, or your doctor know if you'd like a test today.") as a more appropriate descriptor, such as "patient-invited choice".

Please see our response to Reviewer 1's first comment. Our terminology may be novel as applied to the specific case of HIV screening, but it is in accordance with what the larger scientific community understands these terms to mean. Furthermore, we believe that our empirical results of differences between active choice and the other arms validate our ex-ante hypothesis that these are distinct approaches that can have meaningfully different effects on test acceptance.

2. The fundamental principle behind "opt-out" is that patients "offered" HIV testing in this manner is that they can acquiesce to testing---so, I would remove any mention in your manuscript about "patient-centered" approaches to testing. The not publicly stated point about the "opt-out" approach is that it makes testing more likely because patients are less likely to actively say no to testing if they are compelled to do so. So, this cannot rightly be termed as something "patient-centered" as you presented this concept.

We agree that opt-out testing is likely to result in some people testing due to acquiescence, or some other similar mechanism. We do not present opt-out testing as a patient-centered approach; nor do we imply that testing guidelines present it as such. Rather, we state that active-choice testing may be considered to be the most patient-centered approach to testing.

Patient-centered decision making is a separate but related concept that we feel our study helps inform. Our results highlight the degree to which patient-stated preferences are malleable by as little as a one sentence change in wording. We thus caution those attempting to infer patient preferences from observed behaviors (in medical settings) to take care when doing so.

3. Your description of the research about HIV testing in the emergency department is over simplified in many regards. The testing approach (opt-out or opt-in) has been assessed in different studies, but perhaps not in a head-to-head trial as you have tried to do. Also, the impact of other approaches and interventions to increasing testing as well as the concept of acquiescence in testing has been assessed in some regards in emergency department-based

studies.

We have now included additional references to key related ED-based studies, although a comprehensive literature review is beyond the scope of this manuscript. We agree that testing programs have included both interventions designed to increase testing rates as well as practices that may unintentionally affect testing (e.g., timing and location of offer). We appreciate that these factors likely have large effects, and thus sought to control those factors and vary only scripts used to offer testing.

4. Your use of a questionnaire prior to presenting testing is in itself an intervention. Please review other studies, perhaps not head-to-head studies, from the emergency medicine setting for HIV testing that have employed questionnaires and their probable impact on HIV testing acceptance. It would be helpful to show a comparison of testing by questionnaire (before testing "offer", after "offer", no questionnaire, etc.) by testing group.

We shared your concern that the questionnaire itself is an intervention, and thus designed this study with half of patients assigned to the questionnaire prior to the test offer, and the other half to after the test offer. We present the treatment effects according to questionnaire timing in Appendix Figure 1.

We sought to balance two concerns: 1) the questionnaire itself is an intervention and could thus affect test acceptance rates, and 2) patients could modify their questionnaire responses based on their acceptance or refusal of the test (they could, for example, answer the questionnaire in a way that rationalizes their test decision). In the former case, it is an interesting empiric question whether answering questions about HIV-related behaviors impacts one's decision to test. Patients at low risk may decide that testing is not worthwhile, or could decide to test to be as certain as possible even if they were previously not concerned about their risk. Similarly, patients at high risk could realize the importance of testing and test at higher rates, or could become anxious about the real possibility of being diagnosed with HIV and thus avoid testing. While our study does not conclusively answer these questions, it does add to the existing literature, which we have now briefly acknowledged and cited key articles in our discussion section (Strengths and Weaknesses).

5. Your use of "defaults" and other unusual terminology throughout the manuscript is confusing. Besides the problem with terminology mentioned above, the term "defaults" does not add meaning to the manuscript and makes the comparisons performed seem different than what they truly were. I do not agree that you conducted a study with a simple wording change. These statements throughout the manuscript underestimate the impact of what was presented to patients. Your intervention was far more than a word change; the actions required by patients were what differed. I strongly recommend that you change the manuscript throughout to instead emphasize that the study was not about word changes, but changes in what patients were required to do in response to different approaches to bringing up HIV testing to patients in the emergency medicine setting.

As also discussed in our responses to above comments, we do believe that our analysis does yield important findings for the selection of defaults. We agree that

the different wording changes require different patient action, and that both are important. Respectfully, we maintain that the key contribution of the manuscript is how these seemingly minor differences in wording trigger these important differences in patient behavior as measured in our dependent variable, and we believe that this is what is most important to emphasize as we contribute to the larger body of knowledge in this area.

6. You must acknowledge that your patient population is unique. It is a population that likely is well experienced with HIV testing, particularly given that an extremely high proportion of patients were men-who-have-sex-with-men----far more than many other US emergency departments. As such, there undoubtedly was a difference in how they responded to your testing "offers" than would be seen in other emergency departments.

We have added more discussion of some of the unique characteristics of our patients and setting in our strengths and weaknesses section. We hypothesize that these characteristics likely impact the overall test acceptance rates, but should not greatly affect the differential response to treatment assignments.

7. As you hinted at, but should state more strongly, the variations in uptake across HIV testing studies in emergency departments are due to the testing systems used and study design. It oversimplifies matters to state that your study was about differences in word changes in how patients were presented with being tested for HIV. You presented, but might not be showing in all details, a system of testing that included aspects such as who offered testing, when it was offered, under what context (informing patients that testing would be available), the utilization of questionnaires that assessed risk (which is an intervention that impacts testing) the type of testing performed, etc. I would re-arrange the methods section to describe ALL aspects of the testing process from start to finish in one section (labeled as such) so that it can become clear to others how your study and approach differed, and how this affected your observed outcomes.

We agree strongly that the details regarding how tests are offered can be crucial to test acceptance rates, and have modified our wording to reflect this. In addition, we have added more detail to our methods section to help clarify the testing system within which we tested our three treatment arms.

8. You must acknowledge that other emergency medicine studies have not shown that testing is related to self-reported or measured risk. Your results are somewhat anomalous in that regard, so you cannot overstate your findings.

We agree that our study adds something new to this literature. These variable results suggest that further work may be needed.

9. I would not use the Denver risk assessment, as this might not be appropriate for your study and has not be universally accepted as an risk assessment instrument.

Given the different preferences of our reviewers, we have retained the Denver Risk Score and defer to the editor to suggest which risk-classification scheme we use. Our tables in the appendix demonstrate that our results are robust to different classifications. Appendix Table 2 presents the results using a scheme where patients are simply categorized as reporting any risky behaviors or none.

10. Please be certain that the scripts that you presented about what staff said to patients are the COMPLETE script. True word differences and presentations and solicitation of responses are the real differences that can impact testing uptake. For example, did the staff say anything else? How did they acknowledge the response? What if the patient had a question? Was the script prefaced by some statement or statements that affected uptake? I have rarely found studies that are truly "opt-out" in nature. For example, after an "opt-out" statement is made, a follow question is asked that changes the approach to "opt-in" or something else is added, such as "Do you have questions?" or "What would you like to do?". Please explicate your entire script.

We present the complete script. Research staff indeed introduced themselves prior to the script but were trained to do so in a naturalistic manner that did not vary by treatment assignment. They also answered questions from patients, and again were trained to do so in a manner consistent with the treatment assignment but in a naturalistic conversational manner. Given their training, periodic meetings, and oversight from our research coordinator, we are confident that research staff did not routinely deviate from protocol. Even if some research staff did sometimes stray from the script, the analysis we conduct is an intent-to-treat analysis. Perhaps more importantly, if research staff did deviate from the script or add concluding questions, this would serve to bias both opt-in and opt-out toward active choice, and would thus imply that if anything, our findings of differences across arms are understated.

11. The evidence is really lacking that using an "opt-out" approach is better. Please modify your entire manuscript to acknowledge that some studies show that it works better, some do not. But---as noted above, the differences are likely due to the details of the testing approach and study design, as well as the study population.

We agree that this evidence is lacking and note this as a motivation for our study. We now cite several more studies in the introduction and note in particular conflicting results from two studies from the same institution. Further, we do note the importance of the details surrounding the test offer, and for this reason our study is designed to hold all of those variables constant across treatment assignments.

12. Abstract: no need to provide p-values when you provide the 95% CIs.

We have removed these p-values.

13. Abstract (and later in results): your interaction term interpretation of OR 0.43 is unclear because it is unclear what is being compared.

We have removed this statement from the abstract and explained it more clearly in the results section.

14. Abstract: your conclusions are overstated. You conducted one study that showed a differential in testing uptake based on variations on how testing was presented to patients and their role in initiating testing. Global statements about the superiority of one approach vs. another and patient preferences do not follow as logical conclusions from the study you conducted.

We have removed the noted global statement, which had been in the final sentence of the conclusion.

15. Introduction: remove "universal" since this does not mean "non-targeted"

Thank you for flagging the inaccurate wording; we have removed the term "universal."

16. Introduction: "opt-out" is an approach, but is not a testing system. It is one of the CDC's recommendations for altering testing. This term frequently gets lumped with the rest of CDC's 2006 recommendations (e.g., non-targeted testing, removal of specific signed consent, etc.).

We agree with this distinction. We have clarified to better emphasize our contribution of analyzing the difference in test acceptance rates according to patient assignment to three types of test offers while holding all else constant; we do not use the term "system."

17. Introduction: as noted before, testing uptake across emergency department studies are all over the map, whether they involve "opt-in" or "opt-out" or whatever approaches and components, so it is not accurate to say that studies including "opt-out" generally have higher testing uptake.

We agree that uptake rates vary. Our reading of the literature finds that opt-out schemes are generally though not universally associated with higher testing rates, but note that studies finding this association are limited due to the presence of many potential confounders.

18. Much of the introduction can be simplified and focused on the true objectives of the study: investigating how uptake of HIV screening differed under three conditions. Much of the conjecture can be removed. Instead, present facts about what is currently known or not known about HIV testing in this setting and how this study attempted to add to the current knowledge.

We have eliminated some parts from the introduction and moved others to our discussion section.

19. Methods: please remove the term "convenience sample" since this does not have a uniform definition. Simply describe your patient selection process.

We have removed this term and instead describe the process.

20. Methods: instead of in the results section, state in the methods section when the study was conducted.

We have moved the study dates to the methods section.

21. Methods: if you are not going to present data about the monetary incentives, then remove it from the methods section as it is irrelevant.

We chose to mention the other arms of this large trial so as not to mislead the readers about the context for this study on defaults. Furthermore, future readers may be aware of other manuscripts (in process) from the larger trial that do analyze incentives, and we believe that it is important to clarify that the present manuscript focuses only on those with no incentives (as readers could be concerned that the interpretation of results might have been different if we had also included individuals from the incentive arms).

22. Methods: I think that it's simpler to say that consent for including the data in this analysis was obtained after testing was presented to patients (or the questionnaire). Tell us how many people declined to have their data included.

We have reworded our methods section to help make our protocol clearer. We report in Figure 1 the number of patients in each treatment assignment who declined inclusion.

23. Methods: I would not presume risk data on patients in the study. Just conduct a sub analysis on the portion of patients for whom this information is available. Too much conjecture is made here for imputation techniques to be used.

Thank you for raising this point. We have added to Table 2 the results from this sub-analysis (columns 4 and 5). Because different readers may have different preferences regarding imputation versus complete cases analysis, we believe that it is helpful to show both methods in order to reassure readers that the results do not vary across these two approaches to missingness.

24. Methods: your sample and power size section is confusing. Why did you choose the difference you chose? What was the basis of that choice? The rest of that paragraph is also confusing. Please revise.

We have revised this paragraph to help make this clearer.

25. Results: please remove any statements about findings being "statistically significant" or attempting to qualify them.

We have removed these descriptors.

26. How and when did you record a patient's response to the testing approach? How many said yes to testing, but did not get tested?

Patient response was recorded by the research staff immediately after the offer was made. Occasionally patients were unexpectedly discharged from the emergency department prior to blood draw blood draw; this was highly unusual though, especially since the vast majority of patients had their blood drawn for other reasons anyway.

27. Will you be presenting data on how many were HIV+ by testing approach?

Our study was not powered to detect differences in HIV-positive test results across testing approaches, thus we do not present these data.

28. Methods: remove the section about 3% testing during "usual care" to the discussion as a point of discussion. It's not comparable data and does not follow from the study methods.

We have moved this to the discussion section. We do not wish to suggest that this statistic is part of our study, but raise it only as a point of reference for discussion.

29. Methods: condense much of your regression discussion to main points. You should simplify the presentation to provide the primary results without regression, and then the adjusted results under different sub-populations and models.

We have simplified our presentation of the unadjusted and adjusted results. Rates are presented for unadjusted analyses, and for adjusted analyses we now present the corresponding differences in rates derived from the regressions

30. Methods and discussion: you oversimplify the importance of the "personal touch" variation of staff members in getting patients to be tested for HIV. It is likely that despite all of your efforts that staff members varied the script. (Cite other emergency medicine HIV testing studies on this topic, too.) Or, there is some "personal touch" that makes some staff members more successful in getting people to agree to be tested.

Please see also response to point 10 and 11. We agree that there is likely a "personal touch," which could be due to deviations from the script or something else particular to the staff member even if the script is adhered to; however, we did not set out to test the variability across research staff. We present Figure 3 in order to demonstrate that the associations between test offer and response rate are generally consistent across all staff members. Just as we expect variation in test acceptance rates for different testing systems, we expect different rates for different people offering the test. In the Discussion: Meaning of the Study section we cite additional studies that have investigated the relationship between person offering the test and test acceptance rates.

31. Discussion: please note prior points about acknowledging that you evaluated a testing "system" and not just word variation. In particular, informing patients about testing undoubtedly made them think about testing as an option which interacted with how testing uptake was presented to them and what they should do to get tested.

We agree that the action of informing patients about testing could itself increase test rates. Likewise, informing patients that testing is available for all patients (rather than only those with risk factors) could increase testing rates, as could a host of many other seemingly-small differences. Thus, we held these variables constant and varied only the test offer. That is, we created a single system and tested variations in wording within this system, but did not assess the system itself compared to another testing system. We suspect that system variations matter quite substantially (see previous responses 11 and 16). For this reason, we held these variables constant across treatment assignments.

32. Discussion: if you really didn't use the Denver Risk score, then don't use it in your study. Simply look at individual risk behaviors as factors that could have affected testing uptake.

Please see above responses regarding risk classification. Appendix Table 2 presents the results from a risk classification scheme that categorizes patients into those reporting risky behaviors and those who report no risks.

33. Keep in mind that failure to find an interaction effect in most cases is not due to the absence of an effect, but is due to lack of power to detect one.

Agreed, we cannot definitely say that we have ruled out any of these interactions. We now more clearly present the interaction effects sizes can and confidence intervals to allow for a clearer interpretation.

34. Discussion: it is well beyond the scope of your manuscript to conjecture why patients declined testing and whether or not stigma, emotions, or other factors affected their decision to test or not to test. Other than the stated reason of not believing that they are at risk, practical matters such as time spent in the emergency department and pain keep people from getting tested.

We agree that our study is unable to identify the reasons that patients declined testing; it is simply not designed to do so. However, we believe that for the purpose of stimulating related future work it is appropriate to mention some of our hypotheses regarding why patients declined testing in our discussion section.

35. Discussion: I do not agree, given the usual use of the terms of "opt-in" vs. "opt-out" with your terminology change. It is irrelevant to the manuscript and is beyond the scope of your study.

Please see above responses, noting that our use of these terms is consistent with their use in other domains in the broader scientific literature from which they have been borrowed.

36. Figure 1: might be helpful to show percentages in the boxes as a percentage of the prior box, or as appropriate to the chosen denominator, to help put the raw numbers in perspective

We appreciate this suggestion, which we have also debated. Currently we have not added the percentages due to the fact that the figure is already “busy,” however we would be happy to add them if recommended by the editors.

37. Figure 2: as above; remove Denver risk comparisons

Please see above responses regarding risk scoring.

38. Figure 3: really illustrates big variations in uptake by staff members!

Agreed, as noted in the manuscript

39. Table 2 and supplemental tables: SEs are not helpful with ORs. Provide 95% CIs instead. For this table, the 1, 2, and 3 at the top are not clear---clarify that these are univariable and different

multivariable models

These have been clarified.

Reviewer: 3

Recommendation:

Comments:
Statistical Review

BMJ.2015.025090

Patient choice in opt-in, active-choice, and opt-out HIV screening: a randomized controlled trial.

The paper reports a randomized trial undertaken in San Francisco comparing 3 different verbal approaches made to patients in an emergency room concerning offering a rapid HIV test. The trial also appears to have had comparisons with offering monetary incentives, but the results of this analysis is not reported (and the patients offered monetary incentives are not included). There is a further randomization according to the order in which parts of the study (the questionnaire and the test offer) were made to the patient which is not reported in this paper either.

1) There is some jargon which could be removed to made the paper more accessible, particularly from the abstract. Phrases such as “choice architecture”, “active choice” and “test offer scripts” may not be well known.

We have minimized our use of language that may not be familiar to readers. As noted in response to reviewers 1 and 2, we retain key language that we felt was important to put the paper in the context of the larger scientific literature in related domains.

2) The paper does not describe the method by which the randomization list was created (were they computer generated random numbers? was any blocking used?) and how the allocations were made to consecutive patients. From the protocol it appears that the allocation list was pinned to a board and thus there was no concealment of the allocations. As failure to conceal allocations has been shown to be a source of bias in randomized trials, it is important to fully describe the process by which patients were placed in order on this list, and any checks made to ensure that the there was no attempt to subvert the order of the allocations. These are essential components on the CONSORT checklist, and would advise the authors to properly check that their report fully adheres to this statement.

As noted in the response to reviewer 1’s third comment, this is a potential weakness in our study. We report our process, per the CONSORT checklist, to ensure that the reader is fully informed as to our methods.

3) Was any process used to assure the fidelity of the intervention? i.e. were there any checks put in place to ensure that the staff stuck to the script?

The project manager and lead investigator trained staff to ensure initial fidelity, and then worked with staff regularly to reinforce the protocol . Please also see our response to Reviewer 2, point 10.

4) Could the authors clarify the order of the consent process, and receiving the test results?

We have reworded to clarify our description of the consent process and added a description of the delivery of test results.

5) Retrospective consent designs are rare, and the authors should be congratulated for completing one. It would be interesting for the authors to comment on the responses that they received from the participants concerning the retrospective process, whether it was found to generally be acceptable.

We generally had a favorable response to this design. Patients did not appear upset or overly surprised that we might be collecting data for research purposes, and no complaints were lodged. Approximately 1/5 in each treatment assignment did not consent to the study, but we do not know to what extent the retrospective consent influenced this consent rate.

6) The sample size calculation is based on a 2.3% difference in test acceptance rates. Can the authors please explain why this figure was chosen? Also, please state what event rates the planned sample size was based on, and what the original sample size actually was. You need a much larger sample size to detect a difference between 50% and 52.3% than between 10% and 12.3%.

Our power calculations were designed for the larger trial for which this is a subset. We hoped to recruit 18,000 patients; due mainly to a significantly higher proportion of patients not eligible for inclusion, recruitment fell short of that. We assumed a baseline acceptance rate of 50%. We powered our study to detect a 5 percentage point difference between opt-in and opt-out treatments with near 100% power. We felt this was the minimum difference that would be clinically important. This corresponded to a 2.3 percentage point difference between either opt-in and active choice, or between active choice and opt-out testing (under our hypothesis that active choice testing would yield a test rate intermediate to opt-in and opt-out schemes).

7) The study was stopped early. Could the authors please explain how this decision was made? Particularly could they explain whether it was based informed by an analysis of the findings at that point in time?

Our study ended at the end of our grant-funded period. We overestimated our expected monthly enrollment by overestimating the pool of eligible patients (most were excluded due to age). Our endpoint is not a result of a decision to stop based on any interim analysis or attained goal.

8) A challenge with the retrospective consent design are post randomization drop-outs. The validity of the design is compromised where the drop-out rate is high, and where it differs between arms. No comment is made about the comparability of the drop-out rates between arms. It is possible to derive the rates from Figure 1, but it deserves comment in the text.

We have added the number and percentage in each arm who consented to inclusion in the study.

9) Results page 10 last paragraph report the test rates with 95% confidence intervals, and differences between event rates with 95% confidence intervals. This is the most informative way

in which the results of the trial are presented – far more accessible than presentation of odds ratios. I would encourage the authors to present these findings as the key results in the abstract as well as in this section (in place of the odds ratios).

We have rewritten our abstract and results section with a focus on reporting differences in test acceptance rates expressed as percentage point differences.

10) Results page 11 first paragraph includes a result from a follow-up study after the trial concluded. There is no method for this part of the results – it refers to data from outside of the trial as it is described, and I would suggest that such an observation really belongs in the discussion if in the paper at all.

We have moved this to the discussion section. We do not wish to suggest that this statistic is part of our study, but raise it only as a point of reference for discussion.

11) The description of low medium and high risk is not reported in the paper. I would encourage the authors to add a simple box which describes how these classifications are made.

We have added our number cutoffs for low-, medium-, and high-risk categories. Given our space and figure limitations, we report the number values in the appendix.

12) The analysis of testing rates by risk factors is very poorly presented, and I would encourage the authors to greatly improve this part of the paper, including Table 2 and all Tables in the supplementary material. The key problem is that there is no presentation of the testing rates for any of the groups described – data are only presented as odds ratios throughout which makes them inaccessible to mere mortals – particularly once you start quoting interaction terms. Please give us the raw data, and the testing rate with 95% confidence intervals. Whilst the logistic regression models are helpful to know whether the differences between groups are statistically significant or explicable by chance, they mask the magnitude of the effects. Some of this data is presented in Figure 2.

Thank you for this feedback. We now show the raw (unadjusted) rates in Figure 1 and unadjusted rate differences with confidence intervals in Table 2. We also now report results of linear probability models as percentage point differences in rates (we also test this using logistic regression models using "margins" commands in Stata, see Appendix), which we agree are much more interpretable.

13) I would also request that the "test of interaction" in the abstract is enhanced by reporting the event rates.

We have removed this statistic from the abstract.

14) Figures 2 and 3 have no label for the y-axis

We have corrected these omissions.