PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Randomised controlled non-inferiority trial of primary care-based facilitated access to an alcohol reduction website
AUTHORS	Wallace, Paul; Struzzo, Pierliugio; Della Vedova, Roberto; Scafuri, Francesca; Tersar, Costanza; Lygidakis, Charilaos; Mc Gregor, Richard; Scafato, Emanuele; Hunter, Rachael; Freemantle, Nick

VERSION 1 – REVIEW

REVIEWER	John Cunningham Centre for Addiction and Mental Health, Canada
REVIEW RETURNED	20-Oct-2016

GENERAL COMMENTS	The paper presents an important study. The design is elegant and constructed in a fashion to have the potential to answer the research question. The authors may wish to consider the following points. As the authors point out, there are two primary limitations to the study. The first is that the majority of participants scored less than the cut-off point on the primary outcome measure at baseline. This means that there is the potential of a floor effect in the study in that there is limited opportunity for them to display any improvement. This is particularly problematic in a non-inferiority trial. The second limitation is that the primary outcome variable is systematically, and differentially, related to the intervention conditions in the study such that participants in one condition are more likely to score higher on the outcome variable at follow-up than participants in the other condition. Further, this systematic difference on the outcome variable acts in favour of the intervention under study in the non-inferiority trial. While serious, these limitations could be addressed in a fashion that would allow for useful information to be obtained from this trial. Unfortunately, the authors seem to be using the justification of needing to employ the analysis plan as pre-specified in the protocol as a means to draw conclusions from the trial that are perhaps unmerited. This is most clear in the Abstract where the first sentence of the Interpretation is that the trial shows clear evidence of non-inferiority for the Internet intervention compared to the face-to-face intervention. However, given the limitations of the trial, this interpretation, which stems directly from the conduct of the pre-specified outcome analysis, is almost certainly not supportable. My suggestion is that the authors modify the manuscript and structure it in such a way that, while the pre-specified analysis is conducted, more attention is paid to the secondary analyses (also

	briefly presented in the current manuscript) that allows for what is probably a more accurate depiction of the outcome. This would mean expanding the details of the secondary analyses presented in the manuscript, providing a candid discussion of how the secondary analyses are suggestive of the fact that the Internet intervention may, in fact, be judged as producing an inferior outcome to the face-to-face intervention, and then modify the conclusions drawn from the findings accordingly.

REVIEWER	Anne H Berman Anne H Berman, PhD Associate Professor of Clinical Psychology Karolinska Institutet Department of Clinical Neuroscience Center for Psychiatry Research http://ki.se/en/people/anberm Professor Paul Wallace is a guest researcher in my research group and is co-supervisor for PhD candidate Christopher Sundström, MSc. At this writing, Professor Wallace and I have no publications or funded grants in common.
REVIEW RETURNED	04-Nov-2016

GENERAL COMMENTS	This article reports on a study with significance for implementation of brief interventions for problematic alcohol use in primary care. The study was designed as a non-inferiority trial (although the trial registration is somewhat unclear on this point) and the results indicate general non-inferiority of a web-based intervention in comparison to face-to-face brief intervention, in a sample of primary care patients with a relatively low level of risky use. Overall, the study is well-described and the article is well-reasoned. I do however have some comments which the authors should address prior to acceptance. The numbering of the comments follows the questions to reviewers. 1. The research question is clear. 2. Abstract. The research question should be clearly stated in the Background section. In the Findings section, the results are reported very specifically in terms of recruiting and participant flow. However, the evidence for non-inferiority is reported in general terms, and I am missing information on the extent to which participants reduced their risky use in both groups (see below regarding results in the manuscript). The statement that post hoc analyses raise “important questions of interpretation” is vague and should be made more concrete regarding what the authors actually mean by this. 3. The study design focuses on facilitated access (FA) and also demonstrates non-inferiority. However, it is not clear exactly how much time/effort the study design actually saved the GPs, since the BI mostly lasted less than 5 minutes, and the FA recommended that the patient “discuss their web experience” with their GP. GPs were also able to upload personalization features for their patients (photo, tailored messages etc). It would be helpful if the long-term potential of FA interventions were addressed in the introduction as well as in the discussion from the resource perspective. For example, there could be advantages in terms of offering patients evidence-based, consistent psychoeducational intervention content, where the saving of resources may lie in treatment efficacy and accessibility rather than actual time spent for the GP (for some inspiration see Hedman,

	E., Ljótsson, B., & Lindefors, N. (2012). Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res, 12(6), 745-764. doi:10.1586/erp.12.67). Another topic that should be mentioned is the concept of ”blended” care; see, e.g., Ly, K. H., Topooco, N., Cederlund, H., Wallin, A., Bergström, J., Molander, O., Carlbring, P., Andersson, G. (2015). Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS ONE, 10(5), e0126559. 4. The methods are clearly described and the published study protocol is helpful in terms of replicability. • Establishment of whether or not patients met exclusion criteria should be clarified; under “Patients” the authors state that those “known to suffer” from one of several conditions were excluded, while under “Procedures” the authors state that subjects who screened positive were asked to self-report that they did not meet any exclusion criteria. This procedure seems less than rigorous and the rationale should be explained. • The description of the alcohol reduction website should include a description of the “thinker drinker record” since results on this are presented. The reader should not have to refer to a separate article to understand data presented in the current article. • GPS participated in a one-day training for the trial (see protocol article); did the training include guidance for “discussion of the website” aside from study procedures, interactive training in motivational interviewing as well as an introduction to the website? • The web intervention is referred to both as an alcohol intervention and, in one place, as a “lifestyle” intervention. Please correct or clarify. • Please note that the trial registration site is erroneously given as clinicaltrials.org, whereas the correct address is clinicaltrials.gov. 5. Research ethics and procedures are satisfactorily described. 6. Outcomes. An important point concerns the primary outcome measure, which was the proportion of participants scoring 8 or more points on the AUDIT. There seems to be a disconnect between the primary outcome measure and the recruitment threshold, which was based on a score of 4 or 5 on the AUDIT-C. I think the authors need to explain why they did not simply recruit participants with 8 or more points on the AUDIT – or else designate AUDIT-C scores as the primary outcome. 7. Regarding the description of statistical analysis, this is generally quite clear. • I am wondering however about the supportive analyses, which are described but not motivated (i.e., supportive of what?). • Also, the non-inferiority outcome is measured by the proportion of participants showing hazardous or harmful drinking, and the non-inferiority criterion is assessed via non-linear mixed models. This is fine, but perplexing that the results later report the proportions at BL and 3m follow-up, but not 12-month follow-up. • Instead, the mean for the AUDIT is reported in Figure 2. Surely, this does not align with the intentions outlined in the statistical analysis part. • Finally, the analyses are performed on an intention-to-treat basis, which would imply imputation of missing follow-up data. However, no mention is made of imputation procedures. Please explain.

	8. References are generally up-to-date. However, it is somewhat confusing with Roman numerals for numbering them. 9. Results. The results do not quite deliver the information expected based on the background, methods and research design, see item 10. 10. Specific comments on the results • Table 1 shows baseline characteristics, where I wonder regarding the ethnicity issue: I am curious about the distinction between Caucasian and Italian. Is not Caucasian an umbrella term including Italian? I suggest including Italian under Caucasian and explaining with a footnote, or else leaving the table the way it is but also with an explanatory footnote. • The educational level of the participants seems generally low to middle. Perhaps this is an aspect that should be commented on in the discussion, as text-based interventions can vary depending on educational level. • As noted above, the TDR reports do not add anything to the reader's understanding as we do not know what the TDR feature offers and the results on this are thus impossible to interpret. • Regarding reporting of proportions of hazardous/harmful drinkers, these should be reported at 12 months also (currently only BL and 3 months), as well as a weighted report of AUDIT-10 scores with question 10 removed. A weighted representation of AUDIT-10 scores without Q10 will facilitate comparison of AUDIT-10 outcomes with results from other trials using the AUDIT-10 as a main or even secondary outcome. This information is currently missing from the results and makes them difficult to interpret. • Regarding Tables 3-7, none of these report raw data but rather analyses assessing non-inferiority. Although this accords with the statistical plan, the presentation makes it difficult for the reader to understand what changes actually occurred in the alcohol consumption for the participants over time. Figures 2 and 3 partly answer this question but figure 2 is lacking since concrete numbers and standard deviations are not reported (actually, the proportions of participants with hazardous/harmful drinking should be reported, not necessarily AUDIT means). I would suggest reporting numerical outcomes and adding the analyses in columns to the side. This would reduce the number of tables and make it easier to take in the results. 11. The discussion discusses feasibility nicely and the level of patient engagement is impressive in relation to population-based studies on eBIs for alcohol, a sector of research that could well be mentioned in this context. As it is, the authors contrast their results with their own ODHIN trial, which the EFAR-EVG trial is an improvement upon. It would be helpful to understand how the results relate to general engagement in eBIs. An interesting question concerns the fact that GPs seem to have treated patients randomized to both FA and BI. In what way might this have affected the external validity of the trial? I.e., GPs in this trial may have had a higher level of training in delivering BIs for hazardous or harmful alcohol use, and this may have made the FA intervention more effective than it would be in a non-research setting. How do the authors suggest that GPs could be trained to use FA interventions at a level with the same – or higher – effectiveness as in this trial? 12. The limitations are adequately discussed, including the issue of

	the AUDIT-C inclusion criterion. However, the authors should add an explanation of how they came to recruit based on one measure but to measure outcome based on another (AUDIT-10). Also, the authors should motivate why they suggest the TLFB as an alternative outcome measure in future trials. What limitations in the current trial would this remedy? 13-15. These points are satisfactory.

REVIEWER	Robert M West University
REVIEW RETURNED	20-Apr-2017

GENERAL COMMENTS	I restrict my comments to statistical aspects reported in the manuscript and label my comments by the authors headings: Strengths and weaknesses of this study I am concerned that the reader may be filed by the claim that follow up rates exceeded 90% at 3 months and 80% at 12 months. These figures are supported by the results but only after restricting to those who received the intervention. For example only 304 of 416 (73.1%) received the brief intervention. The authors should make it clear that their rates apply to those who receive an intervention and that there are many who are randomised but are lost. Procedures It would be helpful if it is clearly stated that randomisation is at the patient level since this is a strength of the design. Outcomes It should be stated that the primary outcome is hazardous drinking AT 3 MONTHS. This is stated in the protocol paper and I have inferred that it is this outcome that is used in this manuscript. The authors might comment on the loss of power from dichotomisation of the AUDIT score and the clinical reason behind this. See further comments below. Statistical analysis The primary analysis without accounting for clustering of the brief intervention within GP. Since the brief intervention is down to the GP to deliver (after the group 1 dat training) there could be variation due to how this is delivered. This should at least be investigated to rule it out as a concern. This is the major revision part of my recommendation. This has implications throughout the reporting in the manuscript. The sample size calculation does not reflect clustering and as a consequence the trial could be under-powered once clustering is accounted for even with 1000 patients. Note too that there is large variation (1-89) between cluster sizes which further drives up the sample size required. The authors do include Supportive 3 analysis which has a continuous outcome and a random intercept for GP. Is this random intercept also used for the facilitated access? Should the intervention group be regarded as a single cluster? - since the web intervention is delivered uniformly. It is interesting to note that the analysis which accounts for clustering shows a non-significant effect despite being more powerful due to a continuous outcome rather than a dichotomised one. This raises

	major concerns about lack of accounting for clustering within the primary analysis.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: John Cunningham
Institution and Country: Centre for Addiction and Mental Health, Canada
Please state any competing interests: None declared

Please leave your comments for the authors below

The paper presents an important study. The design is elegant and constructed in a fashion to have the potential to answer the research question. The authors may wish to consider the following points.

*We appreciate the reviewer's helpful comments about the elegance of the study design, and are happy to address the points raised.

As the authors point out, there are two primary limitations to the study. The first is that the majority of participants scored less than the cut-off point on the primary outcome measure at baseline. This means that there is the potential of a floor effect in the study in that there is limited opportunity for them to display any improvement. This is particularly problematic in a non-inferiority trial. The second limitation is that the primary outcome variable is systematically, and differentially, related to the intervention conditions in the study such that participants in one condition are more likely to score higher on the outcome variable at follow-up than participants in the other condition. Further, this systematic difference on the outcome variable acts in favour of the intervention under study in the non-inferiority trial.

*In relation to the two primary limitations, we agree with the points made though in relation to the first we would point out that the trial was based upon real world inclusion criteria - it is standard practice in many countries for clinicians to identify risky drinkers using the AUDIT-C and so the trial provides an answer to the real world question. However, from an explanatory perspective the study was limited by the unexpected low scores on the primary outcome measure (AUDIT) at baseline. The use of AUDIT C cut points of 5 for men and 4 for women would have been expected to lead to the inclusion of substantially higher proportions of hazardous and harmful drinkers as defined by a score of 8 or more on the AUDIT (Rumpf HJ, Hapke U, Meyer C, John U. Screening for alcohol use disorders and at-risk drinking in the general population: psychometric performance of three questionnaires. Alcohol Alcohol. 2002;37:261–8 ; Dawson DA, Grant BF, Stinson FS, Zhou Y. Effectiveness of the derived Alcohol Use Disorders Identification Test (AUDIT-C) in screening for alcohol use disorders and risk drinking in the US general population. Alcohol Clin Exp Res. 2005;29:844–54).The AUDIT C has been validated in Italian populations and was found to perform similarly (Struzzo P, De Faccio, S. Moscatelli, E. Identificazione precoce dei bevitori a rischio in Assistenza Sanitaria Primaria in Italia: adattamento del Questionario AUDIT e verifica dell'efficacia dell'uso dello short-AUDIT nel contest nazionale. Bollettino delle Farmacodipendenze e Alcolismo. 2006 http://www.unicri.it/wwk/publications/dacp/index.php). However a recent paper has suggested that higher cut points could be used more reliably for the identification of risky drinkers (Khadjesari et al. Addict Sci Clin Pract (2017) 12:2 DOI 10.1186/s13722-016-0066-5)

*We have identified that the AUDIT, the primary outcome measure, appears to have been systematically and differentially related to the intervention conditions, acting in favour of the condition under investigation in the trial. However, we were not aware of this at the design stage of the trial, when we selected the AUDIT because this measure had been validated for use as an outcome measure in primary care populations (Bradley KA, McDonell MB, Bush K, Kivlahan DR, Diehr P, Fihn SD. The AUDIT alcohol consumption questions: reliability, validity, and responsiveness to change in older male primary care patients. Alcohol Clin Exp Res 1998;22:1842-9), and widely used in a number of other important trials of brief intervention where the same considerations would have applied (eg: BMJ 2013; 346 doi: https://doi.org/10.1136/bmj.e8501: Effectiveness of screening and brief alcohol intervention in primary care (SIPS trial): pragmatic cluster randomised controlled trial)

*We first became aware of the systematically higher scores on question 10 of the AUDIT in the face to face group only after the analysis of the data after the end of follow-up to the trial. We agree and identified that this has almost certainly introduced bias to our findings and fully accept the need to recognise this in the interpretation of the findings.

While serious, these limitations could be addressed in a fashion that would allow for useful information to be obtained from this trial. Unfortunately, the authors seem to be using the justification of needing to employ the analysis plan as pre-specified in the protocol as a means to draw conclusions from the trial that are perhaps unmerited. This is most clear in the Abstract where the first sentence of the Interpretation is that the trial shows clear evidence of non-inferiority for the Internet intervention compared to the face-to-face intervention. However, given the limitations of the trial, this interpretation, which stems directly from the conduct of the pre-specified outcome analysis, is almost certainly not supportable.

My suggestion is that the authors modify the manuscript and structure it in such a way that, while the pre-specified analysis is conducted, more attention is paid to the secondary analyses (also briefly presented in the current manuscript) that allows for what is probably a more accurate depiction of the outcome. This would mean expanding the details of the secondary analyses presented in the manuscript, providing a candid discussion of how the secondary analyses are suggestive of the fact that the Internet intervention may, in fact, be judged as producing an inferior outcome to the face-to-face intervention, and then modify the conclusions drawn from the findings accordingly.

*We are in agreement with the reviewer's proposal and have re-structured the paper accordingly to give greater prominence to the secondary supportive "post hoc" analyses which were carried out to address the problems detailed above. We have also modified our interpretation of the findings as suggested, to give greater weight to the findings of the secondary analyses which do not support non-inferiority of facilitated access.

Reviewer: 2
Reviewer Name: Anne H Berman
Institution and Country: Anne H Berman, PhD Associate Professor of Clinical Psychology Karolinska Institutet
Department of Clinical Neuroscience Center for Psychiatry Research http://ki.se/en/people/anberm
Please state any competing interests: Professor Paul Wallace is a guest researcher in my research group and is co-supervisor for PhD candidate Christopher Sundström, MSc. At this writing, Professor Wallace and have no publications or funded grants in common.

Please leave your comments for the authors below
Review of bmjopen-2016-014576

This article reports on a study with significance for implementation of brief interventions for problematic alcohol use in primary care. The study was designed as a non-inferiority trial (although the trial registration is somewhat unclear on this point)and the results indicate general non-inferiority of a web-based intervention in comparison to face-to-face brief intervention, in a sample of primary care patients with a relatively low level of risky use. Overall, the study is well-described and the article is well-reasoned. I do however have some comments which the authors should address prior to acceptance. The numbering of the comments follows the questions to reviewers.

1. The research question is clear

2. Abstract. The research question should be clearly stated in the Background section.
In the Findings section, the results are reported very specifically in terms of recruiting and participant flow. However, the evidence for non-inferiority is reported in general terms, and I am missing information on the extent to which participants reduced their risky use in both groups (see below regarding results in the manuscript). The statement that post hoc analyses raise “important questions of interpretation” is vague and should be made more concrete regarding what the authors actually mean by this.

*We have now stated the research question clearly in the background question.

*We have provided further information on the extent to which participants reduced their risky used in both groups. We have given more prominence to the findings of the secondary analyses and made it clear that they fail to provide robust support for non inferiority for facilitated access.

3. The study design focuses on facilitated access (FA) and also demonstrates non-inferiority. However, it is not clear exactly how much time/effort the study design actually saved the GPs, since the BI mostly lasted less than 5 minutes, and the FA recommended that the patient “discuss their web experience” with their GP.

*The time taken by the GPs to deliver facilitated access was generally substantially less than that taken to deliver BI, and the offer to discuss their web experience was rarely taken up by the patients. These aspects of the trial are dealt with comprehensively in the accompanying paper dealing with the health economic evaluation of the trial (Hunter et al, Cost effectiveness analysis of EFAR FVG. Submitted to BMJ Open 4TH October 2016 and currently under review: manuscript ID is bmjopen-2016-014577)

GPs were also able to upload personalization features for their patients (photo, tailored messages etc). It would be helpful if the long-term potential of FA interventions were addressed in the introduction as well as in the discussion from the resource perspective. For example, there could be advantages in terms of offering patients evidence-based, consistent psychoeducational intervention content, where the saving of resources may lie in treatment efficacy and accessibility rather than actual time spent for the GP (for some inspiration see Hedman, E., Ljótsson, B., & Lindefors, N. (2012). Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res, 12(6), 745-764. doi:10.1586/erp.12.67). Another topic that should be mentioned is the concept of ”blended” care; see, e.g., Ly, K. H., Topooco, N., Cederlund, H., Wallin, A., Bergström, J., Molander, O., Carlbring, P., Andersson, G. (2015). Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS ONE, 10(5), e0126559.

*We accept the reviewer's suggestion that other e-interventions such as blended care might offer additional benefits, but this issue was beyond the scope of the trial.

4. The methods are clearly described and the published study protocol is helpful in terms of replicability.
• Establishment of whether or not patients met exclusion criteria should be clarified; under "Patients" the authors state that those "known to suffer" from one of several conditions were excluded, while under "Procedures" the authors state that subjects who screened positive were asked to self-report that they did not meet any exclusion criteria. This procedure seems less than rigorous and the rationale should be explained.

*We have amended this section to make it clear that the establishment of whether or not patients met the exclusion criteria was carried out by the GPs directly in their own office prior to offer of facilitated access. We included the online request to patients who screened positive to confirm that they did not meet any of the exclusion criteria simply as a further check.

• The description of the alcohol reduction website should include a description of the "thinker drinker record" since results on this are presented. The reader should not have to refer to a separate article to understand data presented in the current article.

*The thinker drinker record is described in the paper referenced in this section (Linke, S, McCambridge J, Khadjesari Z, Wallace P, Murray, E. Development of a psychologically enhanced interactive online intervention for hazardous drinking. Alcohol and Alcoholism 2008;43:(6):669-674;) However, we accept the reviewer's comment and have also included a brief description in the methods section.

• GPs participated in a one-day training for the trial (see protocol article); did the training include guidance for "discussion of the website" aside from study procedures, interactive training in motivational interviewing as well as an introduction to the website?

*We can confirm that the training included brief guidance on how to actively encourage patient to access the website

• The web intervention is referred to both as an alcohol intervention and, in one place, as a "lifestyle" intervention. Please correct or clarify.

*We have amended the manuscript to clarify this issue

• Please note that the trial registration site is erroneously given as clinicaltrials.org, whereas the correct address is clinicaltrials.gov.

*Thank you. We have amended the manuscript to correct this error.

5. Research ethics and procedures are satisfactorily described.

6. Outcomes. An important point concerns the primary outcome measure, which was the proportion of participants scoring 8 or more points on the AUDIT. There seems to be a disconnect between the primary outcome measure and the recruitment threshold, which was based on a score of 4 or 5 on the AUDIT-C. I think the authors need to explain why they did not simply recruit participants with 8 or more points on the AUDIT – or else designate AUDIT-C scores as the primary outcome.

*We accept the reviewer's comments as the study was unexpectedly limited by the low scores on the AUDIT 10 at baseline. The use of AUDIT C cut points of 5 for men and 4 for women was expected to lead to the inclusion of substantially higher proportions of hazardous and harmful drinkers as defined by a score of 8 or more on the AUDIT (Rumpf HJ, Hapke U, Meyer C, John U. Screening for alcohol use disorders and at-risk drinking in the general population: psychometric performance of three questionnaires. Alcohol Alcohol. 2002;37:261–8 ; Dawson DA, Grant BF, Stinson FS, Zhou Y. Effectiveness of the derived Alcohol Use Disorders Identification Test (AUDIT-C) in screening for alcohol use disorders and risk drinking in the US general population. Alcohol Clin Exp Res. 2005;29:844–54).The AUDIT C has been validated in Italian populations and was found to perform similarly (Struzzo P, De Faccio, S. Moscatelli, E. Identificazione precoce dei bevitori a rischio in Assistenza Sanitaria Primaria in Italia: adattamento del Questionario AUDIT e verifica dell'efficacia dell'uso dello short-AUDIT nel contest nazionale. Bollettino delle Farmacodipendenze e Alcolismo. 2006 http://www.unicri.it/wwk/publications/dacp/index.php).

*We have addressed the suggestion made by the reviewer by undertaking a secondary analysis using the AUDIT-C as the outcome measure.

7. Regarding the description of statistical analysis, this is generally quite clear.
• I am wondering however about the supportive analyses, which are described but not motivated (i.e., supportive of what?).

*The term "supportive analyses" refers to those analyses undertaken in order to determine whether the primary pre-specified analysis is robust. Some of these were pre-specified and others (identified) were undertaken as a result of the identified challenges.

• Also, the non-inferiority outcome is measured by the proportion of participants showing hazardous or harmful drinking, and the non-inferiority criterion is assessed via non-linear mixed models. This is fine, but perplexing that the results later report the proportions at BL and 3m follow-up, but not 12-month follow-up.

*The reason for reporting the findings only at baseline and 3 months is that the primary outcome measurement point was pre-specified as three months following recruitment.

• Instead, the mean for the AUDIT is reported in Figure 2. Surely, this does not align with the intentions outlined in the statistical analysis part.

*We agree with the reviewer and have removed this figure.

• Finally, the analyses are performed on an intention-to-treat basis, which would imply imputation of missing follow-up data. However, no mention is made of imputation procedures. Please explain.

*No imputation procedures were used. The intention to treat principle, as described in ICH E9, makes no reference to missing data. It is merely the approach to analysis where all patients are included on the basis of the group to which they were randomised, independent of whether or not they received the specified intervention.

8. References are generally up-to-date. However, it is somewhat confusing with Roman numerals for numbering them.

*We have reviewed the numerals used for the references which are now numeric.

9. Results. The results do not quite deliver the information expected based on the background, methods and research design, see item 10.

10. Specific comments on the results

• Table 1 shows baseline characteristics, where I wonder regarding the ethnicity issue: I am curious about the distinction between Caucasian and Italian. Is not Caucasian an umbrella term including Italian? I suggest including Italian under Caucasian and explaining with a footnote, or else leaving the table the way it is but also with an explanatory footnote.

*We acknowledge this inconsistency and have revised the table accordingly.

• The educational level of the participants seems generally low to middle. Perhaps this is an aspect that should be commented on in the discussion, as text-based interventions can vary depending on educational level.

*The reviewer's point is valid and the manuscript has been revised accordingly to reflect this. However, we would point out that any effects of educational level on outcome would have been experienced similarly by both the FA and facilitated access groups

• As noted above, the TDR reports do not add anything to the reader's understanding as we do not know what the TDR feature offers and the results on this are thus impossible to interpret.

*As described above, we have revised the manuscript to clarify the nature of the TDR

• Regarding reporting of proportions of hazardous/harmful drinkers, these should be reported at 12 months also (currently only BL and 3 months), as well as a weighted report of AUDIT-10 scores with question 10 removed. A weighted representation of AUDIT-10 scores without Q10 will facilitate comparison of AUDIT-10 outcomes with results from other trials using the AUDIT-10 as a main or even secondary outcome. This information is currently missing from the results and makes them difficult to interpret.

*As described above, the primary outcome measurement point was at three months, hence the inclusion of these results only in presenting the results of the primary analysis.

*In response to the reviewer's proposal to report weighted representation of the AUDIT scores without Q 10, we would point out that we have indeed used a "weighted" cut point of 7 (rather than 8) in order to define hazardous/harmful drinkers, and present our findings accordingly.

• Regarding Tables 3-7, none of these report raw data but rather analyses assessing non-inferiority. Although this accords with the statistical plan, the presentation makes it difficult for the reader to understand what changes actually occurred in the alcohol consumption for the participants over time. Figures 2 and 3 partly answer this question but figure 2 is lacking since concrete numbers and standard deviations are not reported (actually, the proportions of participants with hazardous/harmful drinking should be reported, not necessarily AUDIT means). I would suggest reporting numerical outcomes and adding the analyses in columns to the side. This would reduce the number of tables and make it easier to take in the results.

*We have revised the presentation of the results in line with the reviewer's comments and have included a table with the proportions of patients with hazardous/harmful drinking at all three measurement points of the trial. We have also removed the figure reporting the AUDIT mean scores. Regarding tables 3-7, we accept the reviewer's observation that these report the results of the analyses rather than presenting actual data. However we would respectfully point out that that presentation of these tables was pre-specified in our analysis plan, but that we have in addition identified and explained problems which arose in following the plan, and have also presented post hoc analyses addressing these.

11. The discussion discusses feasibility nicely and the level of patient engagement is impressive in relation to population-based studies on eBIs for alcohol, a sector of research that could well be mentioned in this context. As it is, the authors contrast their results with their own ODHIN trial, which the EFAR-EVG trial is an improvement upon. It would be helpful to understand how the results relate to general engagement in eBIs.

*In the ODHIN trial, the GPs undertook all the screening using paper forms rather referring their patients to website. Additionally, the training and familiarisation with the BI website offered to the GPs was almost certainly less rigorous than in the EFAR- FVG trial

An interesting question concerns the fact that GPs seem to have treated patients randomized to both FA and BI. In what way might this have affected the external validity of the trial? I.e., GPs in this trial may have had a higher level of training in delivering BIs for hazardous or harmful alcohol use, and this may have made the FA intervention more effective than it would be in a non-research setting. How do the authors suggest that GPs could be trained to use FA interventions at a level with the same – or higher – effectiveness as in this trial?

*The GPs in the trial received additional training both in the delivery of BIs face to face and via FA. While this may have reduced the external validity of the trial, brief training in the use of FA interventions could readily be delivered under "real world" conditions

12. The limitations are adequately discussed, including the issue of the AUDIT-C inclusion criterion. However, the authors should add an explanation of how they came to recruit based on one measure but to measure outcome based on another (AUDIT-10). Also, the authors should motivate why they suggest the TLFB as an alternative outcome measure in future trials. What limitations in the current trial would this remedy?

*We have already described the reasons for the selection of patients using the AUDIT C (see 6: Outcomes)
We initially considered using the TLFB, but decided against using it in order to increase acceptability and minimise reactivity of assessment

13-15. These points are satisfactory.

Reviewer: 3
Reviewer Name: Robert M West
Institution and Country: University of Leeds, UK
Please state any competing interests: None declared

Please leave your comments for the authors below

I restrict my comments to statistical aspects reported in the manuscript and label my comments by the authors headings:

Strengths and weaknesses of this study
I am concerned that the reader may be filed by the claim that follow up rates exceeded 90% at 3 months and 80% at 12 months. These figures are supported by the results but only after restricting to those who received the intervention. For example only 304 of 416 (73.1%) received the brief intervention. The authors should make it clear that their rates apply to those who receive an intervention and that there are many who are randomised but are lost.

*We would respectfully point out that the follow-up rates cited in the paper refer to the percentage of all participants randomised to the trial, irrespective of whether or not they received the intervention. As shown in the CONSORT diagram reproduced below, the total numbers of trial participants lost to follow up were relatively small.

Procedures

It would be helpful if it is clearly stated that randomisation is at the patient level since this is a strength of the design.

*We have revised the manuscript to clarify this point

Outcomes
It should be stated that the primary outcome is hazardous drinking AT 3 MONTHS. This is stated in the protocol paper and I have inferred that it is this outcome that is used in this manuscript. The authors might comment on the loss of power from dichotomisation of the AUDIT score and the clinical reason behind this. See further comments below.

*We have revised the manuscript to make it clear that the primary outcome is hazardous drinking at 3 months. We have also included reference to the loss of power from dichotomisation.

Statistical analysis
The primary analysis without accounting for clustering of the brief intervention within GP. Since the brief intervention is down to the GP to deliver (after the group 1 dat training) there could be variation due to how this is delivered. This should at least be investigated to rule it out as a concern. This is the major revision part of my recommendation. This has implications throughout the reporting in the manuscript.

The sample size calculation does not reflect clustering and as a consequence the trial could be under-powered once clustering is accounted for even with 1000 patients. Note too that there is large variation (1-89) between cluster sizes which further drives up the sample size required.

*We agree with the reviewer's suggestion that there was almost certainly variation in the way in which face to face BI was delivered by the different GPs participating in the trial. However, since randomisation in the trial took place at the individual patient level rather than cluster randomisation, we addressed these effects by including general practices as random intercept terms in the generalised non-linear mixed models used to compare proportions of risky drinkers in the two groups as is orthodox.

The reviewer seems to have been confused by the final supportive analysis which examined the effect of intervention on the continuous audit 10 score rather than examining the proportion of risky drinkers. We have revised the footnotes and paper to make the explanation clearer.

The authors do include Supportive 3 analysis which has a continuous outcome and a random intercept for GP. Is this random intercept also used for the facilitated access? Should the intervention group be regarded as a single cluster? - since the web intervention is delivered uniformly.

*As stated in the methods, the supportive 3 analysis (in common with the primary analysis) used a random intercept term for a GP practice. This term was applied to both experimental conditions.

It is interesting to note that the analysis which accounts for clustering shows a non-significant effect despite being more powerful due to a continuous outcome rather than a dichotomised one. This raises major concerns about lack of accounting for clustering within the primary analysis.

*In all analyses, consideration for centre effects was applied for both groups via some form of random effect at the practice level. The referee is correct that therapist level effects are important, and these have been addressed in the analyses.

VERSION 2 – REVIEW

REVIEWER	John Cunningham Centre for Addiction and Mental Health
REVIEW RETURNED	24-Jul-2017

GENERAL COMMENTS	No additional comments. Concerns raised in first review adequately addressed.

REVIEWER	Robert M West University of Leeds UK
REVIEW RETURNED	28-Jul-2017

GENERAL COMMENTS	This is an informative piece of work which is well presented. There is a very minor point regarding the abstract. "9080 patients of whom 3841 (84·8%)" omits the step that "4529 (49·9 %) patients logging on". My comments are focussed on the statistical analysis which is generally to a good standard of statistical practice. The authors were careful and correct to include a random effect for GP for the reasons they state in the manuscript. It is possible however that clustering due to the GP could be different in the BI arm than in the FA arm. In the BI arm, the GP clearly has an influence delivering a face-to-face intervention. In the FA arm also, the GP can customise messages. The clustering effect could differ by arm. The authors might at least discuss this under limitations if not undertake another supportive analysis. All of the tables and figures provided contribute useful information. It is pleasing to note that the authors have gone to lengths to explore the implications of the trial results.

	The low level of drop out is reassuring. There might have been an analysis to predict drop out based on baseline data: drop out might be associated with drinking behaviour with risky drinkers dropping out and so reducing the rate of risky drinking at 3 and 12 months.

VERSION 2 – AUTHOR RESPONSE

We note the points raised by Reviewer 3, Robert West and are pleased to confirm that we have responded as follows:

Comment 1: There is a very minor point regarding the abstract. "9080 patients of whom 3841 (84·8%)" omits the step that "4529 (49·9 %) patients logging on".

Response:

WE ACCEPT THIS OBSERVATION AND HAVE INSERTED THE ADDITIONAL STEP IN THE ABSTRACT

Comment 2: It is possible however that clustering due to the GP could be different in the BI arm than in the FA arm. In the BI arm, the GP clearly has an influence delivering a face-to-face intervention. In the FA arm also, the GP can customise messages. The clustering effect could differ by arm. The authors might at least discuss this under limitations if not undertake another supportive analysis.

Response:

WE ARE GRATEFUL FOR THIS FEEDBACK AND HAVE INCLUDED CONSIDERATION OF THE ISSUES RAISED IN THE DISCUSSION AS FOLLOWS: "It could be argued that differences in the intervention could lead to difference in the clustering within practices, however our assumption a priori, based upon substantial relevant experience, was that shared practice characteristics were likely to be the dominant factors in our analysis. Given the overall nature of the results and their interpretation we have not undertaken further supportive analyses on this question."

We trust that these minor revisions will meet with your approval and that it will now be possible to proceed to publication of the paper.