Author’s response to reviews

Title: Wound Healing In Surgery for Trauma (WHIST): statistical analysis plan for a randomised controlled trial comparing standard wound management with negative pressure wound therapy

Authors:

Ruth Knight (ruth.knight@csm.ox.ac.uk)
Louise Spoors (louise.spoors@ndorms.ox.ac.uk)
Matthew Costa (matthew.costa@ndorms.ox.ac.uk)
Susan Dutton (susan.dutton@csm.ox.ac.uk)

Version: 1 Date: 10 Dec 2018

Author’s response to reviews:

We would like to thank all the reviewers for their detailed review and comments. We have addressed each of the comments, and believe that answering these has enabled us to improve the clarity of the article.

Reviewer #1: Clearly written and appropriate statistical analysis for the soon to be completed WHIST multi-site RCT comparing negative pressure wound care to standard care in lower limb fractures. There are only a few clarifications I would like to see addressed:

- Page 11, paragraph 207 - Only wound location is discussed as a potential covariate to investigate in secondary analysis given that the primary outcome is different between treatment groups. In reality, won't the investigators also consider other covariate effects from Table 2 such as antibiotic prophylaxis, how closed, etc?

Response: We anticipate that all the patients having surgery as part of WHIST will have antibiotic prophylaxis. The type of antibiotic varies by Trust, according to local microbiology guidelines, but we do not feel that including this as a variable in the analysis is likely to add anything. Similarly we are collecting the type of wound closure, but given the lack of evidence that wound closure influences the rate of deep infection we do not propose to include this as a variable in the analysis.

- In the primary analysis of deep infection rate @30 days, all adjustments are made by random effects (site, open vs closed, ISS level, age, and gender) but why are open v closed, ISS, age and gender included as fixed effects in the secondary outcome models? Also, why are age and gender included in the models at all?
Response: Open versus closed fractures, ISS level, age and gender will be included as fixed effects in all models. Centre is included as a random effect due to the number of recruiting centres, and the fact that we would expect the impact of these to be random variation. We have clarified this in the text.

Age and gender are included in the model since there is some evidence that older men have worse outcomes after major trauma. We have explained this in the text.

- Page 12, line 233 "non-parametric equivalent" - assuming Rank Sum but please specify.

Response: The non-parametric equivalent which we will use is the Mann-Whitney U-test. We have clarified this in the text.

- Page 12-13, lines 235-239 It is not clear how you plan to use the coefficients from the mixed effects linear regression model to compute AUC scores or estimate recovery time. Please include more details explaining this analysis. If switching to binary outcomes of DRI and EQ-5D please indicate threshold values as have been included in the description of the other secondary outcomes.

Response: We have re-written this section to more clearly describe the methods which will be used. In addition, the reference provided (Bell et al., 2014) describes these methods in detail and explains their missing data properties. The parameter estimates will be combined using the lincom command in Stata.

- Secondary (binary?) outcome POSAS is not discussed in the analysis section.

Response: Many thanks to the reviewer for pointing out this inconsistency in terminology used – the POSAS is the patient-reported scar assessment which will be treated as continuous. We have clarified this in the text.

Reviewer #2: With the change of the CDC definition up to 90 days for metal implants, what proportion of cases will that encompass? I would think the rate of infection could increase simply because the time to monitor them will increase. Would this change your initial assumptions used to calculate sample size?

Response: Our primary outcome remains deep infection at 30 days, so in this regard there is no concern regarding the sample size assumptions. As the CDC definition of deep infection was updated during the trial we will also report deep infection up to 90 days as a secondary outcome in support of the primary analysis, so that the results of this trial can be compared to future studies using the updated definition.
Sample size considerations: one-sided or two-doses alpha level?

Response: Two sided, we have updated text to reflect this.

Line 205: What is a sufficient number of deaths in the 30 days?

Response: We have clarified in the text that a sufficient number of deaths equates to at least 5% of participants.

Reviewer #3: TRLS-D-18-00703_reviewer

The manuscript for the SAP is comprehensive, clear and easy to follow. Some comments:

1. P. 4, line 70: Says: "Superiority"

Comment: In my view this is not a superiority trial as the alternative hypothesis includes treatment differences is both directions.

Response: The trial is indeed a superiority trial. In a superiority trial we aim to show that one of the treatments compared is superior to the other, thus an alternative hypothesis including treatment differences in both directions is appropriate.

2. P. 7, line 147:

Unclear how you handle multiple testing. I don’t think stating alternative analyses as "supporting" is convincing argument for ignoring multiplicity. I can live with not doing multiplicity correction but I would appreciate a better rational.

Response: We have clarified in the text that the conclusion of the trial will be based on the primary analysis of the primary outcome, and that additional analyses of this outcome will be sensitivity analyses to investigate the robustness of these results. In addition, we have clarified that conclusions of the trial will not be based on the analysis of secondary outcomes.

3. P. 11, line 196:

You do in general use a random effect for centre. How many centres are planned to participate? In there are several centers I think using a random effect is good. If the centers are few, a fixed effect may be more appropriate.

Response: The trial recruited from 24 study sites (see trial status), therefore we considered including centre as a random effect was most appropriate.
4. P. 12, line 230:

Explore the possibility of using quantile regression if normality assumption is questionable.

Response: We have clarified in the text that we plan to use the Mann-Whitney U-test as a non-parametric equivalent in the event that the assumption of normality does not hold and transformation to achieve normality is not suitable.

5. P. 13, line 250:

Please consider Poisson/Negative Binomial regression when patients have several complications (count data)

Response: Many thanks for this suggestion. In the context of this trial, we do not anticipate that sufficient numbers of events, or individuals with more than one event, will be observed for this type of analysis to be performed.

6. P. 13, line 258:

How is time to events recorded? Are they reliable? What is the mode of recording date of complications (hospital admission date for example?)

Response: Thank you very much for this question. Time to event is not well recorded. For some complications participants are asked when they experienced these symptoms, and for complications which are also SAEs the date will be recorded. For these reasons, we do not envisage a time to event analysis will be appropriate; however, we have left it in the SAP as this analysis is mentioned in the published protocol.

Reviewer #4: The current manuscript describes a proposed statistical analysis plan for assessing a reduction in the rate of deep surgical site infection post major trauma. This study is of high relevance to the field, which currently appears to lack studies which have been appropriately performed statistically. The manuscript is clearly written, but I suggest to incorporate the following points to further improve it:

1) Could the authors supply the actual power calculation they have performed to define their sample size to be 615 (or 770 when accounting for dropout) per arm. This would be appropriate given they will be using mixed-effect models, and do not appear to have ran simulation studies to determine the sample size.

Response: The sample size calculation was based on a test of two independent proportions, we have clarified this in the text.
2) The authors should elaborate on "major protocol deviations" (R143-146). They do give an example which seems clear to be deemed as such, but apart from that the extent of deviation seems arbitrarily defined and thus yields a somewhat vague exclusion criterion.

Response: Thank you for the comment. We have clarified that major protocol deviations are expected to consist of: those who did not satisfy the eligibility criteria; those who did not receive their allocated treatment; and those for whom insufficient data are available on the primary outcome.

3) I would suggest to rewrite the phrase "a sufficient number of deaths" on R205-206, and perhaps even define a fraction of the sample size to be "sufficient".

Response: We have clarified in the text that a sufficient number of deaths equates to at least 5% of participants.

4) For analysis of the secondary outcomes the authors describe they will be assessing the relevant variables for approximate normality (R218-219), and if absent will be transformed to achieve so (R230-232). However, given they will be fitting mixed models for these outcomes (R220-223), there is no need to have them be normally distributed. What is, however, explicitly necessary for these mixed models to make sense, is whether the resulting residuals are normally distributed. Mixed models assume the outcome to be normally distributed, conditional on the random effects. If the authors wish to deposit a statistically sound paper, they should address this issue.

Response: It is planned to assess the appropriateness of the assumption of approximate normality by plotting the residuals from the mixed models. Many thanks to the reviewer for highlighting that this was not clear in the text, we have updated the text accordingly.

5) The authors will include interaction effects between time and treatment in the secondary analyses (R225-226) if trends over time appear appropriate to do so. The authors should define appropriate here.

Response: We have clarified in the text that trends over time for each intervention arm will be plotted, and if there are substantial differences between these then interactions between treatment and time will be included in the model.

Reviewer #5: A very detailed and thoughtful description of an analysis plan for a multi-centre parallel-group randomized superiority trial evaluating NPWT in lower-limb traumatic fractures. A couple of minor comments:
1) Which dataset will be used to compare AUC summary statistics for DRI and EQ-5D between groups - ITT with imputed or available data sets?

Response: The ITT population with available data will be used, we have clarified this in the text.

2) In the adjusted models, age and gender are included without any justification; certainly given the size of the trial would not expect a difference in these baseline covariates by random chance, and no evidence is presented that suggests gender or age in this patient population is associated with deep SSI rates. Conversely, it is conceivable QOL or disability measures may be associated with age and/or gender. Given the predicted event rate, I do not believe the inclusion of these two covariates would adversely alter (or improve) estimation of the treatment effect on the primary outcome. Please comment.

Response: Age and gender are included in the model since there is some evidence that older men have worse outcomes after major trauma. We have explained this in the text.