Author’s response to reviews

Title: Early warnings and repayment plans: novel trial management methods for monitoring and managing data return rates in a multi-centre phase III randomised controlled trial with paper case report forms

Authors:

William Cragg (w.cragg@leeds.ac.uk)
Fay Cafferty (f.cafferty@ucl.ac.uk)
Carlos Diaz-Montana (c.diaz@ucl.ac.uk)
Elizabeth James (e.james@ucl.ac.uk)
Johnathan Joffe (johnathan.joffe@cht.nhs.uk)
Monica Mascarenhas (monica.mascarenhasjones@gmail.com)
Victoria Yorke-Edwards (v.yorke-edwards@ucl.ac.uk)

Version: 1 Date: 26 Feb 2019

Author’s response to reviews:

[See also uploaded file for more user-friendly layout]


We thank both reviewers for their time and detailed comments on the manuscript, which have helped refine and enrich the content for a wider audience. See below for a point-by-point response to each comment.

Reviewer #1:

# Comment

Response and any changes to manuscript

1
Reviewer #1: The authors make a very good case for the need a) for more evidence-based methods for improving data returns, and b) for making these improvements sustainable across the life of a study; all too often we approach data returns as binary problem, or in the worst case as emergency exercises ahead of data cut-off.

N/a

2

The approach is presented and discussed very clearly, and its applications and limitations are laid out very thoroughly; the paper also eloquently accounts for the less quantifiable aspects of managing such processes. The paper merits consideration by a wider trial management audience. My suggestions below are largely minor and should be viewed as suggestions to clarify and potentially enrich some of the findings.

N/a

3

The "repayment plan" approach strikes me as a key element of the presented method. Perhaps it would merit a little more detail, in either the Methods or Results section, e.g. by illustrating this with a practical example.

There is a brief example in the Methods section (“We approached centres with a proposed portion of the overdue forms to send within a certain timeframe, e.g. 20 forms within two weeks.”)

We hope this is a reasonable balance between giving enough detail to allow replication and not expanding the word count too much; in response to the other reviewer’s comments we have added in some clarification of the process, including that we would, after the initial agreed timeframe, review the site’s DRR again and agree another target if need be.

4

The Results section is somewhat sparse compared to other sections; I would be interested here in the structure of data reminders and general communications. For instance, you say that 10 sites were identified with diminishing DRR rates, and that they were contacted initially saying only that they hadn’t submitted forms recently. Were all these sites contacted again at every subsequent extract / CTU meeting, or only once a change in trend had emerged? Could you give some (perhaps more descriptive, qualitative) detail on how these communications progressed? Given that I would expect you to be in communication with the remaining 25 sites in some form, this would help to isolate the nature of the "intervention" a bit more clearly. It might also serve to
illustrate how sustainable the intervention was during its limited time of application, and to help
the reader assess how much effort is involved in producing

a DRR increase per site.

These are all sensible suggestions. In response we have made the following changes:

- In Methods, we have clarified that centres without any obvious issues were not contacted
specifically about data return (i.e. the intervention is fairly easy to isolate)

- In Results, we have added: “Apart from discussion arising from this initial contact, these
centres were not contacted again prior to subsequent CTU team meeting review.”

You ask whether all the good-but-falling sites were contacted again at the next CTU meeting; the
answer is that they would be if there was still a problem. Hopefully this does not need explaining
further in the manuscript, i.e. that each review was based on current data.

5

While this may have been outside the remit of the project, but I wonder if any effect on data
quality (not just completeness) was seen, e.g. in terms of number of data queries required. My
expectation would be that as sites that are lagging behind are not required to submit all overdue
data immediately, this results in less of a rush to submit data and might contribute to relative
better data quality; are the authors able to comment?

It would not be straightforward to give statistics about other aspects of data quality, and therefore
we agree that it is outside of the scope of this work. Number of data queries, for example, is
flawed because not all data queries address data quality issues (e.g. they are also used to flag data
that may well be correct, but indicate a deviation from the protocol, or out-of-normal-range data
that are ultimately found to be correct).

We have added a sentence near the end of the Discussion: "It is beyond the scope of this work to
explore the effects of our methods on other aspects of data quality (e.g. accuracy of data
provided), but this could be included in future work in this area."

6

In the discussion, the authors mention the increased efforts required for communicating with
sites e.g. through telephone. I wonder how these efforts compare with the more traditional, static
model of data collection reminders?

We do not have good data to report on resourcing – this is something a more robust evaluation
could explore.
However, we have mentioned resources (as you say) in the Discussion. We have amended slightly to highlight a contrast with more basic system: “…although this [liaising with centres] may be somewhat more effort than simple reminders for centres with unacceptable DRR, we consider the time worth spending…”

In response to the other reviewer’s comments, we have also confirmed that the new processes did not require any extra ’manpower’ for the trial – this is hopefully enlightening regarding the resourcing question.

7

Conclusions: I fully agree that the method presented merits further exploration. Perhaps the approach would even lend itself to a potential SWAT (study within a trial), randomising between the different data reminder methods?

True; added in mention of this approach and appropriate reference.

8

The abstract would benefit from a little more context: word count permitting, I would suggest adding e.g. that the particular trial entailed a very long-term follow-up, and that the 10 sites identified for DRR intervention were among 35 total.

Agree and amended.

9

Escalation policy: Are you able to quantify what would have constituted "consistently" (line 178) or "prolonged" (line 179) to warrant an escalation?

This would be based on discussion in the CTU team and a decision to escalate, rather than a set threshold. We have added into the first of the relevant sentences: “…in the view of the CTU team…” – which is hopefully adequate.

Reviewer #2:

# Comment

Response and any changes to manuscript
The authors have identified an issue related to data collection timelines and its delay on analysis and dissemination of results.

N/a

The authors have assessed the return rates of visit specific paper case report forms to a central data entry centre from a multi-centre phase III Non-CTIMP randomised controlled trial from April 2015 to September 2016. To set return rate targets, the team, reviewed pre-study return rates, set site targets and implemented communication intervention/feedback and measured return metrics per site. The aim was to increase return rates, as well as develop an escalation process in the event that return rates failed to meet set targets.

N/a

While this is a limited study, it does address a possible management tool to assess return rates to trigger pre-defined target communication. Thereby enabling trial and data managers to concentrate their efforts on sites that require assistance.

N/a

The work is small-scale and exploratory, related to one multi-center Non-CTIMP with limited, participant visits over an 18 month review period. Although, the authors have identified the need for additional research to generate evidence based methods, which could include these elements in due course.

N/a

The repayment plans described in the title are not currently evident in the manuscript. If repayment plans were part of the study design, these should be described in the methodology, and results. If this was a typographical error, this should be removed.

Apologies for the lack of clarity. The ‘repayment plan’ idea refers to the ‘targets for improvement’ that are mentioned in the abstract and main paper. Instead of asking for all outstanding data immediately (as had been done previously, and as, we believe, is done most commonly in this situation), we would request some of the outstanding data within a short timeframe – enough data to cause the data return rate to rise – after which point we would review again and set another target. This approach, along with the review of data across time, is what we believe make our methods novel (we do not know of any other descriptions of similar methods) and why these two elements in particular are mentioned in the article title.

Action taken in response to this comment: the section in Methods where this is described (paragraph beginning with “Based on…”) now links explicitly to the term ‘repayment plan’, and clarifies that the target-setting was repeated until the DRR was acceptable. In the abstract we have clarified that the target-setting would lead to “gradual” improvement, rather than being a one-off action.

2. The authors should review the manuscript, checking grammar and style writing, tenses, single/plural issues thereby ensuring the standard required for publication.

We have re-read the whole manuscript and made any amendments as appropriate; if there are additional, specific changes to make, please do highlight these for our attention.

3. Line 7. While this study is a NON_CTIMP, the authors should address Good Clinical Practice and introduce the concept in the Background and its relationship with data quality and validity. The authors should also address the responsibilities investigator to ensure the accuracy, completeness, legibility, and timeliness of the data reported to the sponsor in the CRFs and in all required reports.

We have now added a reference to Good Clinical Practice at the start of the Background to put the issue of data completeness in that context; the same sentence also makes reference to the direction of reporting, i.e. from investigator to sponsor. See responses elsewhere about ‘data quality’.
4. Throughout the manuscript there are numerous; for examples, which makes for a jarred reading experience.

Suggest the authors use; terms, including, embracing, comprising etc. In addition, these portray assumptions, with limited literature to support. Consider adding support, if available.

Good point; there were 10 uses of ‘for example’, and in the new draft there are 3, and the mid-sentence construction has been avoided. In reviewing these affected sentences, we have not found statements that we believe require extra supporting references, but if there is anything specific, please let us know.

18

5. Line 3. The introduction of Table 1, is premature, prior to introducing concepts. Consider relocating to

Methods section.

Agree, and moved.

19

6. The authors do not specifically address the concepts of Data Quality Management (DQM) and Data Management Plans in relation to this study. While the authors address data management and quality they do not reference the quality management process, which is a formal process for managing the quality, validity and integrity of the research data captured throughout the study from the time it is collected, stored and transformed (processed) through analysis and publication. With the DMP describing data to be collected; how and where the data are captured and stored; process for reporting and handling corrections; confidentiality and data sharing etc. Both these concepts should be clearly defined and addressed and referenced within the manuscript.

We note several reviewer comments asking for clarity around 'data quality'. While we agree the concept is fundamental to integrity of clinical trial results, there are other resources available discussing general aspects of data quality, and very little discussing the specific issue we've aimed to cover, i.e. how to maintain a complete dataset on an ongoing basis. We feel that to include additional content about other aspects of data quality is out of the clearly-defined scope of this work, would be detrimental to the focus of this paper, and expand the word count further without much added value.

However, we accept that there are some mentions of 'data quality' that perhaps invite the reader to wonder about definitions etc as per your questions, and we have clarified or removed these: 1)
in Background paragraph about risk-based monitoring and 2) in Methods in paragraph prior to title "Data return rate reporting".

In response to this particular comment: we agree it would give useful context to explain the other processes in place to collect and clean data in this trial. We have added in an additional Figure (as suggested elsewhere –

Figure 1 in new draft) to cover this. All trials at MRC CTU do indeed have Data Management Plans, which describe the measures in place to control quality of all stages of data collection and cleaning, up to database lock. We have also added in a sentence to confirm that a DMP is in place.

20

Abstract

Lines 40-42. Clarification required - "We also reported separately on forms that can and cannot be completed retrospectively" This is not clear, please revise.

Amended to: “The tool allowed us to distinguish between forms that can and cannot be completed retrospectively…”

21

Line 53. Grammar check required - "agreeing targets"

Perhaps this is an issue of British vs American English usage? (E.g. https://www.dailywritingtips.com/transitive-twist-on-agree/)

This seems correct in British English; no change made.

22

[Main paper]

Lines 2-4. The introductory sentence is disjointed and should be rewritten to introduce the subject matter.

Agree, and reworded, incorporating a link to Good Clinical Practice as advised elsewhere.

Amended to: “Complete and timely reporting of trial data from investigator to sponsor is a key process in Good
Clinical Practice in clinical trials.[1] There are various reasons why maintaining a complete dataset on an ongoing basis is important in trial management.”

As a result, terminology (including ‘data return rate’) not defined until Methods – but this fits anyway with the suggestion to move Table 1. Terminology accordingly adjusted in Introduction.

23

Line 4. Suggest that high data return rates are pertinent to all studies, not merely RCT. The authors can defined why of particular importance to RCTs; Non-CTMP and CTIMP.

Agree; amended to ‘clinical trials’, and abbreviation ‘RCT’ corrected throughout (and removed from abbreviation list).

24

Line 5 "a long delay’ and Line 25 :" short data return” - can the authors quantify the term 'long' and 'short', or is it merely a mismatch between trial activity data acquisition and data availability in DMS.

This is difficult because it would vary depending on the trial. In a trial like TRISST, we usually allow 28 days for data to be returned; in an early phase trial, daily data return might be desirable.

Amended the first case to ‘unreasonable’, which hopefully conveys the sense of ‘whatever is detrimental to the trial’s needs’, and the second to ‘delays in’, rather than ‘slow’.

25

Line 7 - Introduction of monitoring requires definition in relation to this manuscript. Is this a correct assumption?

Added in excerpt from ICH GCP definition of monitoring.

26

Line 7-8. Introduction of the consent of monitoring. Require further explanation.

As above.
Line 9. Suggest amending the word 'capability' - the oversight committees have capability of making decisions, however, their decisions may not informed due absence of data.

Agree; amended to “Oversight committees’ decision-making may be impaired by reviewing trial data that are not complete.”

Line 9. Consider singular and plural conditions in relation to return.

Based on the comment linked to line 24, we assume the reviewer is suggesting we amend all instances of ‘data return’ to ‘data returns’; we have done this, except where the phrase used is ‘data return rates’ or some other phrase where the plural is taken by some other word (e.g. ‘data return problems’). Hopefully this is an improvement.

L17. …"helping prepare the database for trial analyses." This sentence should address the need for data lock prior to analysis.

We note that there is a reference to database lock in the main text; however, we have also now added this to the abstract, as the word count allows it.

Line 24. Suggest rewrite to; The central’s team's ability to identify issues in a timely manner in relation to patient safety or protocol adherence is also reduced by slow data returns.

“Central’s team’s…” does not seem correct to us; we suggest a compromise that hopefully improves the sentence: “The CTU trial team’s ability to spot patient safety or protocol adherence problems in a timely manner is also reduced by delays in data returns.”

Line 25. a long delay' - can the authors quantify the term 'long', or is it merely a mismatch between data from activities performed not being recorded in the DMS.

Addressed in relation to reviewer comment, above.
Lines 8, 24-28 - Concept of monitoring. The authors are requested to introduce the concept of monitoring

Addressed in relation to reviewer comment, above.

33

Line 27. Suggest amending "on-site -monitoring and increased use.. " to

Unsure what change is being suggested; please clarify (if change still required).

34

Line 28 "The usefulness of these is largely dependent on having" to "The usefulness of these are largely dependent upon having"…the authors are requested to review the whole manuscript and edit singular/plural grammatical issues

‘Usefulness’ is a singular/uncountable noun, and should take a singular verb; no change made.

35

Line 29. Concept of good quality date - authors require to define good quality data. What does it look like, how is it assessed?

See response above about 'data quality'; now removed the mention from this place.

36

Line 32. The authors require to outline what is novel about their methods in relation to current knowledge of the subject matter.

As mentioned in response to a comment above, we believe the novel features are 1) reviewing DRR change over time at each site, and 2) a flexible, collaborative approach to working with sites to improve DRR by agreeing short-term improvement targets. As well as being highlighted in the title, explained in the Methods and reiterated at the start of the Discussion, we hope that additional changes made in response to other comments have made this clearer.

37

Line 34. Is there a reference to support this assumption?
We agree that this is a supposition only; softened the wording to align with this.

38

Line 39. Grammar check required - "time to data being available"
Amended to: “…time to data availability…”

39

Line 45. The authors should consider why pCRF are not completely replaced by EDCS; relevant to many investigator trial, cost/time/validation methods/systems etc.

We feel it is beyond the scope of this paper to go into detail about the reasons why paper CRFs may not have been completely replaced in all settings; there are various possible reasons which would perhaps warrant a separate discussion paper.

In the case of TRISST, the trial was set up in 2006 before trials at MRC CTU started using electronic data capture more routinely; it would have been a significant undertaking to change the data collection method part-way through, requiring not only systems changes, but also major process changes at both participating NHS centres, as well as at the CTU. More generally, the adoption of electronic data capture across our trials has been a prolonged process for these reasons.

However, as we argue in the Discussion, the same challenges apply to both systems of data collection, i.e. how to ensure a complete dataset on an ongoing basis. Thus, the methods described apply equally to both settings.

No change made to the manuscript.

40

Lines 53 -54. Review grammar in relation to "i.e. thresholds to colour acceptable centres green…"

For the avoidance of doubt, “to colour” here is intended as a verb. Nonetheless we have aimed to clarify by re-wording: “i.e. thresholds used to assign acceptable centres a green label…”

41

Line 56. Consider amending "slipping" to deteriorating, falling, declining etc.
Amended to “falling”.

42

Line 88. Grammar correction required. "by, at most, one full-time data". There is only one, so most is not applicable.

By ‘at most’, we mean that at times, the trial’s data has been managed by part-time staff, or staff with time split between TRISST and other projects. Amended to ‘full-time equivalent’, which is hopefully clearer and language familiar to most.

43

Line 90. Clarification in relation to "sent." Sent by email, post. If post, were they copies with the original at site?

Amended to “…posted to the CTU (with a copy retained at the centre)” – which hopefully clarifies this issue.

44

The authors should consider a flow diagram depicting the CRF requirements per visits/study timeline, and the interaction with CTU. Indicating best case and worst case scenario timelines.

We have now added a new Figure 1, to give context of data management and related quality assurance processes in TRISST, and associated timeline allowances. This may not be entirely what the reviewer was suggesting, but hopefully helps clarify our expectations of our trial centres, and the various quality assurance processes we have in place alongside the methods described in this work.

45

Line 98-99. Grammar check required. "we expect" to we would expect

Amended to “we would expect”.

46

Line 103. Clarification related to percentages listed; "was around 92%, and had generally been between 85% and 90%". How can it be greater than the maximum of 90%.
“Generally” here is intended to mean “not always, but usually”. Amended to “…at most Trial Management Group (TMG) and CTU team reviews had been between 85% and 90%...”.

47

Line 106. How and who decided that 80% return rate was an amber flag for a centre requiring support to increase data return?

80% had emerged as a suggested threshold based on experiences in this and previous trials at the MRC CTU, as a level that is achievable for most centres and so would tend to highlight those centres experiencing difficulties in most settings. Some trials may use higher thresholds if their trial risks dictate it, or if they are nearer to the time of database lock and analysis. Manuscript amended to reflect this.

For clarity, we have also added a sentence here to confirm that we would aim for 100% data completeness by the time of database lock.

48

Line 109. Clarification required. “in each report”. Please clarify the meaning of this sentence.

Clarified to ‘each TMG report’ – i.e. for each TMG meeting as mentioned in the previous sentence.

49

Line 111. As the TMG only met every 6 months, there was potentially only 3 opportunities for metrics to be assessed during the 18 month CRF review period. Therefore, four centres (11%) had <80% DRR in more than 1 meeting? Please clarify the importance of this statement.

This section is describing DRR management and figures in the trial overall. We have added ‘since the start of the trial’ to highlight this. Specific figures for the period of interest are included in the Results section.

50

Line 112. Grammar check required. Please consider rewriting the sentence amending the concept of "better quality" to improved, and to outline how completeness is different from quality? Suggest defining the metrics used to assess quality and provide relevant results/discussions in the appropriate sections.
As mentioned above, we have removed this reference to data quality as it is not relevant to the scope of this work.

For improved clarity, we've removed 'better' to leave simply "As data completeness was identified…"

51

Line 114. Suggest outlining previous processes and how the new processes were determined as novel. In the discussion the authors will have to address

We have added in a sentence above Table 1 to clarify previous processes, i.e. that sites below the threshold at a review point were contacted and asked for outstanding CRFs. All the sections that follow describe what was novel, with explicit contrast to previous processes where necessary, e.g. new manuscript line 180 (tracked change version line 198): “As we had done previously, we notified centres with DRR <80%, but…[w]e also began to contact centres with ≥80% data returns but consistently falling” and line 186 (tracked change line 204), “Based on previous lack of success at some centres in simply asking for all outstanding data…”

52

Line 115. Developed a more comprehensive process for handling DRR than had been used previously. Outline what the process was before how the new processes were novel. In the discussion a comparison is required outlining the improvement metrics in relation to resource/capacity of DM investment into the new processes. The authors should also discuss future professing via automatic reminders to sites.

In response to other reviewer comments, we hope that both the previous processes and the novelty of the new processes are now clearer.

We do not have detailed data on how much time the new processes took in comparison to the previous processes; however, there was no change in the amount of staff time given to the trial. We have added in this point to the

Discussion.

The manuscript already suggests that automating some aspects of these methods would be a useful development – see line 292 (312 in the tracked change version) of the new manuscript.

53

Line 119. Grammar check required. "Our first aim was to be able to better visualise…”
Agree; amended to simpler: “Our first aim was to visualise change…”

54

Line 126. The authors should expand on the concept of "The tool underwent testing prior to use". How and what was done - validation process should be outlined.

Clarification added that testing involved checking that “the calculations were correct for each site”.

55

Line 127 -129. In contract to data integrity and GCP expectations - the process outlined to assess data appears to be open to potential errors. The authors should address the validation processes used to ensure no errors during copying and pasting. In addition, the authors should address this as a current limitation and outline potential technology advances that can be employed going forward.

Although we agree that systems validation is a crucial aspect of ensuring data integrity, regulators allow for validation to be proportionate to the risks of a given system (reference: section 14 of MHRA’s Grey Guide). In this case, the risks are very low: if the calculation were incorrect for a given site, the worst outcome would be that a site would get a message that is incorrect (e.g. asking them to send in data when they have no data outstanding).

It is important to remember that the automated DRR report (mentioned in the same paragraph) was our primary reporting tool for DRR, with the spreadsheet as a useful added tool. The automated DRR report was more rigorously tested in accordance with MRC CTU SOPs.

Although there is a small risk of copy and paste error, the tool is designed to minimise this. Copying and pasting is one single step (i.e. one copy, one paste, no manipulation required). It has clear and concise instructions explaining exactly what to do, including exactly what data to copy and paste, and where to paste it. We have added a sentence to the manuscript to mention these instructions.

During the period this manuscript describes, there were no such errors that we are aware of. This was adequately controlled, and so we do not consider it a limitation per se, but nonetheless the Discussion does already suggest that an automated version of this reporting tool would be a useful development.

56
Line 135 - 138. Interesting granularity between green and red. It would be advantageous for the reader to understanding what each would indicate if a continued trend and what migrating and escalation processes would be required to rectify intervention points.

In the first paragraph of the section ‘Initial management of highlighted centres’, we have explained how centres in different categories would be managed; however, we have now linked more explicitly back to these different categories, with hopefully better clarity as a result.

57

Line 150-151. While this is a correct assumption. The study design could also have had an impact. As the study and clinic visits could align, depending upon the organizational skills of the study team, CRN or trial specific RN, issues with staff change over, the patient outcome questionnaires may not have been readily amiable at the time of the patient visits. This should be addressed as a possible causal impact.

OK; added sentence: “Alternatively, this could indicate that the protocol’s data collection processes are not feasible.”

58

Related Figure 3. CRF return rate for patient -reported outcomes. This requires to be addressed by the authors in relation to the low number of completion at some sites. What remedial actions were instigated to improve this data?

Some sites had problems earlier in the trial with return of patient reported outcome questionnaires, partly due to patient refusal and partly due to organisational issues at the site. We can perhaps expect return of these questionnaires to be lower than data collection for the trial in general due to the possibility of patient refusal.

Nonetheless, we had engaged with all the affected sites – involving principle investigators as required - to help understand the cause of their problems, and to help resolve them.

By the time of the data shown in Figure 4 (formerly Figure 3), there were no significant current problems with questionnaire return at any sites but – as mentioned in the text – as the questionnaires cannot be completed retrospectively, the DRR for the questionnaires at some sites would likely remain low for the remainder of the trial. We have added some text at the end of the Methods section “Data return rate reporting” to explain why some sites in Figure 4 have low DRR for PROs, and what was done about it.

59
…"hypothesised that this approach might be better received by centres". The methodology of 'novel trial management methods for monitoring" did not include discussions with trial centres?

Although we accept that a site’s point of view can be illuminating, in this case we did not consult our sites about our new DRR methods – as we are the sponsor of the study, we would usually take sole responsibility for developing processes around trial and data management.

No change to manuscript.

Line 173. …" some room for negotiation, as long as the plan would result in improved DRR over time" how was this managed as this would not be known prior to negotiation…this was prospective.

This was simply a matter of estimating how many CRFs would be required over the agreed time (i.e. linked to how many patient visits will be due) and asking for more CRFs to be returned than that; the DRR will then inevitably rise, if they meet that target. This is explicit in the text – see around line 190 in the new draft (208 in tracked change version).

Line 176-188. Escalation Process. Was it ever implemented, if so how successful was it? Did the sites know there would be escalation if they did not reach the pre-defined targets? How were sites informed of their targets?

Hopefully it is now clear enough in the manuscript that we did not have to use the escalation plan during the time period described; this is indeed given as a limitation around line 318 of the new manuscript (338 in tracked changes).

We did not specifically tell sites about the new escalation plan, but it seems reasonable to expect that they would expect some escalation on our part if they do not send data, as this is routine. As mentioned in the Methods, the concept of escalation was not new in the trial, only that we “formalised” what had been more ad hoc escalation previously.

Targets – once agreed – were confirmed to sites in writing (i.e. email). We have added in that detail to the relevant section in Methods (around line 194 in the new, clean manuscript; 213 in tracked change version).
Line 190-191. Evaluation. The authors should either consider removing this and document analysis with the results, or provide a description of the evaluation, methods chosen.

We have not been clear enough that this was a post-hoc exercise, and the text at this point in the manuscript has been amended to clarify this (as well as the abstract). Our analyses are purely descriptive (as we admit several times that our evidence is limited or preliminary), so we feel detailed description of analysis methods is not necessary.

63

Line 194-195. Rationale for stopping the analysis. The authors should consider providing additional details. Did this have a limitation to data collection/analysis etc?

Hopefully we have been clearer now about this being a post-hoc analysis. The processes were stopped because of staff changes on the trial, as explained in the manuscript. Part of the motivation for our paper is that there is next to no evidence to support processes for maintaining complete data; different staff are, as a result, entitled to implement processes as per their own professional judgement, as long as they are in line with MRC CTU guidance about suitable processes. The fact that the new methods described in this paper were used for a period but not before or after is useful in that it gives us a little insight – albeit not free from possible bias – into their possible impact.

No change to manuscript.

64

Line 197-198. The authors should present potential reasons why it reached a peak at this point; related to study design, due to DM measures being implemented, was there a lag time and this was the impact zone, etc.?

This is presented in our Results section as some (admittedly limited) evidence of the effectiveness of our new data return rate monitoring methods.

It is hopefully clear that it is being stated here for this purpose, and the various limitations we present in the Discussion should be taken into account by the reader when considering our conclusions – as well as any further speculations they may make about other explanations for our results.

65

Line 203. Clarification required - the new process. What was the new process, the title states novel trail management methods - what were they? The authors should address how communication with sites the need to address an issue is novel.
Hopefully this is now addressed in responses to other comments, above. We have been clear that we did contact problem centres in previous practice (e.g. line 180 [198 in tracked changes], “As we had done previously, we notified centres with DRR <80%...”) – what was new was a) the more nuanced contact based on the trend data, and b) the negotiation and target-setting with sites, rather than asking for all outstanding data immediately. These are, we believe, adequately described and highlighted in the Methods.

66

Table 1 outlines multiple metrics, but they are not addressed independently in the results section. How each/or a combination changed over time or in response to 'collaborative' telephone reminder.

To clarify: the terms in Table 1 are not really ‘metrics’, but ‘statuses’ of forms in terms of whether or not we expect to have received them and, if so, whether or not we have received them. For a trial manager, the statuses dictate what action to take, e.g. “scheduled” means not due yet, so no action to take; “expected” means we are allowing the trial centre time to complete and return the form, and will not take any action; “Overdue” means we need to chase. We feel the list of definitions/explanations is a helpful paradigm and a useful additional output of our work in its own right, as well as supporting the content of the manuscript.

For the purpose of the paper itself, and the trial overall, the statistic we care about most is the proportion of forms received, so that is what has been reported in the results.

No change to manuscript.

67

Neither the results nor the discussion addresses the concept of data quality measures. The authors should address the quality of data prior to managing CRF targets, was quality maintained, reduced or improved as a result of setting and delivering to target?

As mentioned elsewhere, the focus of this paper is on data completeness in terms of CRFs returned. We have clarified in the 4th from last paragraph in Discussion that "...additional processes are required to address other aspects of data quality and integrity."

In response to the other reviewer's comments, we have also confirmed, at the end of the same paragraph, that "It is beyond the scope of this work to explore and effects of our methods on other aspects of data quality (e.g. accuracy of data provided), but this could be included in future work in this area."
Lines 245-247. "Centres might not be contacted about corrective actions, for example, if they have already notified the CTU team that they currently have resourcing issues." The authors require to address this. In site agreements for RCT and a requirement for GCP (RCT) it is a legal responsibility there is adequate recourses available. In addition, HRA, Statement of Activates are sign-off to ensure there is site resource/capacity. As part of trial management these issues need to be addressed, not dismissed.

From time to time, temporary resourcing issues may be unavoidable; e.g. if staff members are suddenly on long-term sick leave. It would be damaging to our relationship with our sites if we were not understanding about this – after all, similar things may happen in the CTU. The intended meaning here is that if, for instance, one week before reviewing the DRR figures, a site had notified us about such resourcing issues, then the DRR figures show that the site seems concerning, we might hold off contacting the site for a couple of weeks rather than sending the usual message as if they hadn’t already given us an explanation.

Perhaps what could be clearer in the manuscript is that we are talking about a delay in contacting such sites, rather than permanent postponement, and that we are talking about temporary resourcing problems, not long-term issues, which of course would be dealt with more seriously. We have amended the manuscript accordingly.

Lines 238-240. Are the authors suggesting that the new staff on the trial were the stopping rule for evaluation?

Independent decision to change the processes. What were those processes and who decided that this was appropriate at that stage of the study? With reference to Figure 1b. Could the authors describe the impact on the new preferences and the number of centres who fell below the 80% target?

Hopefully it is now clearer that our analyses are post-hoc. The decision to change processes (reverting to previous processes, in fact) was not a stopping rule of any kind. Figure 2 (formerly Figure 1) highlights – with admittedly limited evidence – that DRR issues seemed to increase again after the more detailed processes were stopped.

No change to manuscript.
Line 263. Grammar check required - "This would require storing of each report snapshot within the report" Please clarify.

The intended meaning here is that the reporting system would need to store data, as opposed to simply reporting it from the trial database – at the time of this work it could not do this. Amended to “…storing of data from each report snapshot within the reporting platform” in case this helps clarify.

71

Line 288. Clarification required - "higher levels of the escalation plan". Earlier in the manuscript the authors note that the escalation plan was not required. The authors are requested to clarify, if and to what degree the escalation plan was used.

Agree; manuscript amended.

72

Line 292. "there are no significant incentives to help ensure completeness of follow-up data" - breach to GCP and research governance framework. The authors should address the legal and best practice guidelines representing the foundation of data integrity and address the issues in the manuscript.

The point we are making here is the comparison with recruitment, i.e. sites are directly incentivised to recruit, with regular reporting of their recruitment figures, while a similar system is not in place for reporting of follow-up data. However, you are right that there is not a complete absence of any reason for sites to send in follow-up data – we have added in mentions of the standards you refer to here. We have also highlighted (with the addition of the word “ongoing”) that the key issue is, as in the whole paper, about ensuring ongoing completeness of data, not just at the end of the study.

73

Line 293-296. The authors should address the issue of secondary outcomes in TRISST and the likelihood they will be analysed and published.

We did note in the submitted manuscript that ‘this should not be considered acceptable practice’, and does not apply to TRISST or other trials at MRC CTU at UCL. However, we feel this point is probably unhelpfully detracting from the main argument we are trying to make, which is that trialists should only collect data that are definitely necessary. The manuscript has been simplified to focus on this point.
Line 335. The authors are requested to delete sentence related to publication and add the emphasis of further research required.

Removed final clause of sentence (“…and we hope that publication of this paper will encourage other CTUs to do so.”) and reordered paragraph in order to retain logical flow.

75

Line 312. The authors should address the idea that CRF is over a tear overdue. Unlikely to receive data, staff changes, access to patient notes, etc. This timeline is unacceptable, and is a trial management failing.

In the context of TRISST – with over 600 patients and, towards the end of the follow-up schedule, study visits 6 months apart – this is possible. From our experience of running other multicentre phase III trials with long follow-up periods, a year delay to receiving a CRF is not common, but not impossible. Hence this method suggested in the text to look specifically for these outliers.

However, for the purpose of wider generalisability, we have amended to 6 months in the manuscript.

76

The authors should consider per site, central data entry issues. The level of responsibility and relevance for site staff.

We have already mentioned central data entry – i.e. at CTU – see line 361 in the new draft (382 in tracked change copy). We have added a line to refer to possible factors at site: “Data completeness may well be impacted by the way centres are organised and resourced, but it is beyond the scope of this paper to explore such factors.”

77

While the DRR and timelines are specific for this trial what inferences, if any, can be made regarding CTIMPs in relation to reporting AEs/SAEs?

In sentences about how acceptability thresholds may vary, added in references to the thresholds being different depending on trial characteristics (e.g. use of IMP) and, within trials, the type of data (e.g. adverse event data may be prioritised).
Its generalizability to studies with additional visits, assessments per protocol and medicinal intervention are not addressed, especially in relation to the need for in-time safety reporting.

We do address the question of generalisability to larger studies at the end of the Discussion – to say that we cannot claim our results can be generalised to a larger study without extra evidence. We have added that the same can be said for ‘studies with greater safety reporting requirements’. Also relevant to the question of ‘in-time safety reporting’ is the section in Discussion about how our method does not address unscheduled data reporting (including SAEs) and suggesting other methods that could do this role. No further changes made to manuscript.

79

The authors are requested to address the limitation of the current studies in relation to these elements.

See comments above.

80

Figures

Figure 1c. With reference to the addition of a timelines in relation to patient visits and CRF completion, this would highlight a decrease in the percentage of CRFs for the stage of the study. Not sure what benefit the cumulative recruitment line gives to understanding. To present context a percentage of expected vs actual CRFs would perhaps be more informative.

We acknowledge in the main text (as a limitation) that the number of CRFs expected per month at this stage in the trial was decreasing.

We present cumulative recruitment as it a) shows when recruitment finished and b) explains some of the trial activity at an earlier stage, therefore giving context to the data return rate figures in 1a and 1b.

The number of expected CRFs per month is preferable to the cumulative number expected over the trial as it gives a clearer indication of when the busiest times are in the trial – as we would have to manage data around the time when they are expected.

No change made to Figure.
Figure 3. CRF return rate for patient-reported outcomes. This requires to be addressed by the authors in relation to the low number of completion at some sites. What remedial actions were instigated to improve this data?

Already addressed in response to a comment, above.