Author’s response to reviews

Title: Trial Forge Guidance 1: What is a Study Within A Trial (SWAT)?

Authors:

Shaun Treweek (streweek@mac.com)
Simon Bevan (simon.bevan@nihr.ac.uk)
Peter Bower (peter.bower@manchester.ac.uk)
Marion Campbell (m.k.campbell@abdn.ac.uk)
Jacquie Christie (jacquie.a.christie@gsk.com)
Mike Clarke (m.clarke@qub.ac.uk)
Clive Collett (clive.collett@nhs.net)
Seonaidh Cotton (s.c.cotton@abdn.ac.uk)
Declan Devane (declan.devane@nuigalway.ie)
Adel El Feky (a.elfeky@abdn.ac.uk)
Ella Flemyng (ella.flemyng@biomedcentral.com)
Sandra Galvin (sandra.galvin@nuigalway.ie)
Heidi Gardner (heidi.gardner.10@aberdeen.ac.uk)
Katie Gillies (k.gillies@abdn.ac.uk)
Jan Jansen (jjansen@uabmc.edu)
Roberta Littleford (R.Littleford@dundee.ac.uk)
Adwoa Parker (adwoa.parker@york.ac.uk)
Craig Ramsay (c.r.ramsay@abdn.ac.uk)
Lynne Restrup (lrestrup@btinternet.com)
Frank Sullivan (fms20@st-andrews.ac.uk)
Author’s response to reviews:

Response to Editorial comments on TRLS-D-17-00779

2/2/2018

**************

- Please move the declaration section to above the reference section.
  Response: Done.

- Please move the abbreviations section to above the declarations section.
  Response: Done.

- Please mention each author individually in your Authors’ Contributions sub-section. We suggest the following kind of format (please use initials to refer to each author's contribution): “AB carried out the molecular genetic studies, participated in the sequence alignment and drafted the manuscript. JY carried out the immunoassays. MT participated in the sequence alignment. ES participated in the design of the study and performed the statistical analysis. FG conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.”
  Response: Done.
Response to comments on TRLS-D-17-00779

2/2/2018

[***NOTE: the responses below are also appended as a formatted version at the end of the pdf generated by the submission system, which may be easier to read.***]

Thanks for the reviewers’ comments on our manuscript, our responses to them are below. We have made the changes to the manuscript itself using track changes so that they can be easily seen; we have also submitted a clean version with all our tracked changes accepted.

Reviewer #1

1. Page 9; Line 17; in the sentence ending ‘….with the bespoke PIL compared to a standard PIL.’…for clarity, consider adding at the end of the sentence 'before adapting its use in a future trial.'

We agree. We have added the following text to the end of the sentence:

‘…before using it in a future trial.’

2. How was the approach to the evaluation coordinated and collaborative? Are you referring to conducting the meta-analysis, or did the various researchers conducting these SWATs come together in some way? It's not entirely clear what the coordinated collaboration actually was?

Good point. We have added some new text to the start of this sentence (marked in red):

‘By approaching investigators, encouraging them to embed an evaluation of the two types of PIL into their trials and then coordinating the analysis of data from those trials that did, the START program’s coordinated, collaborative approach of embedding a SWAT evaluation in trials involving over 6600 people now provides an evidence base for researchers trying to decide on whether to develop a bespoke PIL for their trial.’

3. In the examples of questions that could be addressed in SWATs; third bullet point - is this a 2 stage PIL (rather than a 2-component), i.e. giving short version first, then long, or is this a three arm SWAT? (short, long and standard PIL) - needs a little clarity.

The intention is that the example is a 2-stage PIL rather than a three-arm SWAT. We have changed the text of the bullet point to:
‘Evaluating the effect on recruitment and retention of 2-stage participant information leaflet (i.e. the leaflet is delivered to participants in two parts: a short ‘key points’ version together with a longer version containing more detail) compared with a standard, single-stage leaflet.’

4. Also fourth bullet point; please explain for the reader what is meant by 'formal site selection questionnaires' - not sure what this means.

This is unclear and we have decided to change the example to something a bit easier to explain in a bullet item:

‘Evaluating the effect on data quality of providing site staff with face-to-face data entry training compared with Skype or video-conference training.’

5. Re the possible cost of SWAT; where did the 5-10k figures come from?

These figures were based on our experience of running SWATs. We have made it clear that this is an experience-based figure rather than any sort of formal evaluation by changing the text to:

‘SWATs need not be expensive; our experience is that many are likely to cost between £5000 and £10,000.’

6. In Box 1; randomisation section. While you acknowledge randomisation is not always necessary; I'm not sure its fair to suggest that this will weaken results. What should be emphasised here is that different designs can be used to address SWAT questions. I think this section would benefit from expanding to include the possibility of using diverse designs, including qualitative studies.

We have changed the Randomisation bullet list to make it clear that whether randomisation is needed depends on the research question. Where the intention is to measure effect we remain clear that randomisation is the preferred approach, while recognising that this may not always be possible. We have though added a new sentence to say that where measuring effect is not the focus, randomisation is likely to be inappropriate. Reviewer #2 also had a comment about this bullet point. The new text is:

‘Whether randomisation is needed depends on the question being asked. If the intention is to evaluate the effect of alternative ways of doing a trial process, then the alternatives being compared should be allocated at random. This may not always be possible and another allocation method can be used (e.g. before and after the new alternative) but in most cases this will weaken confidence in the results. However, if the question being asked is not focused on
measuring effect sizes (e.g. it could be concerned with understanding why something is done the way it is) then randomisation is likely to be inappropriate and other qualitative methods would be required. Randomisation is not a defining feature of a SWAT.

Randomisation can be by a separate process to that used for the host trial randomisation.’

7. Similarly, under the analysis section; need to refer to qualitative analyses and other possible types.

We have added a new bullet point to the Analysis section, which is given below. We haven’t mentioned ‘other possible types’ because our existing first bullet point does say that analysis ‘might be simple’, which isn’t really suggesting that there is only one type of SWAT and our new second bullet point now makes it clear that qualitative questions need to be handled in ways appropriate to the qualitative question.

‘SWATs exploring qualitative questions about how a trial process is delivered, organised or perceived will be analysed using a suitable qualitative analysis method.’

8. Under the SWAT Repository section; might be good to expand and state the application form for registering a SWAT is accessible on line and provide the precise link.

Done.

Reviewer #2

1. Typo on page 8, line 12/13: "…this is that the evidence…"

Corrected.

2. Box 1: Under Randomisation - "In some cases randomization may not be appropriate or possible…” - could you give an example for this?

"…and another allocation method can be used but…” - do you have any specific recommendation for this case?

Reviewer #1 also had some questions about the Randomisation bullet point (see Reviewer #1, comment 6). We have now expanded this section and the new text does, we think, address
Reviewer #2’s request for examples by saying randomisation is best for questions of effect but that, for example, would be inappropriate for qualitative questions. We also give the example of before-after as an alternative allocation method for effect questions that may sometimes be the only approach possible but which would weaken our confidence in the results compared with random allocation.

3. Box 1: Under Analysis - Would a sample size calculation for a SWAT make sense to see how many patients/trials will be needed to answer the SWAT question with sufficient statistical power? Do you have any thoughts on Minimum Important Differences for recruitment and/or retention?

Sample sizes are a tricky issue for SWATs because the SWAT is constrained by the size of the host trial; we imagine that in most cases the size of a host trial will not be changed to help a SWAT. We have added some text to the Analysis heading in the Box to mention sample sizes and the likely host trial constraint. We also make it clear that many individual SWATs will be underpowered but that they are designed for meta-analysis. The next text is:

‘Sample size calculations for SWATs can be done in the usual way using estimates of minimum important differences that the investigators or others consider appropriate. The size of a SWAT is constrained by the host trial. The size of a recruitment SWAT will generally be larger than the host trial sample size (the constraint is the size of the patient population approached, not recruited). Other SWATs (such as those on retention) will be limited to the actual host trial sample. It is highly unlikely that the size of the host trial will be changed for the benefit of a SWAT. SWATs are designed for future meta-analysis. In other words, while an individual SWAT may be underpowered, a meta-analysis of several well-done SWATs evaluating the same intervention and following the same protocol can provide compelling evidence for trial process decision making. As with all meta-analysis, judgements need to be made about whether it is sensible to combine studies done in different populations, disease areas and settings. This issue will be the topic of future Trial Forge Guidance.’

While we mention minimum important differences, there are no widely accepted differences for recruitment, retention or any trial process really. Recruitment and retention are also likely to be some of the easier outcomes to define and operationalise than others that may be measured in a SWAT, such as data quality or participant burden. Also, for trial processes, it is easy to imagine that what is an acceptable difference will vary by trial question, intervention and setting and, probably, by the resources available to a particular trial. We are thinking about these issues because they are important but we don’t think this is the paper to raise them. This leads nicely to the Reviewer’s last question in fact.
4. Trial Forge Guidance 1…" - As an interested reader I am wondering whether this article is the first in a series of planned articles? If yes, could you add 1-2 sentences briefly explaining to readers the purpose of the series and the content of the following articles of this series?

It is our intention to produce guidance of the sort described in this paper through Trial Forge. The next one is likely to be on when to start a new evaluation of a SWAT, which does cover some of the sample size issues the Reviewer raised in comment 3. We are not intending a fixed series though so we do not have, say, five articles in mind and that’s the end. We see this as an ongoing guidance process to help methodologists and others make decisions about trial process evidence.

This intention, however, was not clear in the original draft as noted by the Reviewer. We have added the following text at the end of the Introduction:

‘This paper is the first Trial Forge Guidance document and there will be more Guidance documents in the future, each providing what we hope is clear help and guidance around an issue relevant to improving the evidence base for trial decision-making. Trial methodologists and other stakeholders will be consulted to determine the topic areas and scope for future guidance.’

Editorial comments

1. In several parts in the article, it is said that SWATs on the same kind of question can be object of systematic reviews and meta-analyses. This aspect does require some additional discussion. Meta-analysing SWATs would often encounter the same kinds of issues as a meta-analysis of studies on the same medical intervention but done in different populations/diseases/settings. Judgement on the “exchangeability” of these studies is required, as well as assessment of possible sources of heterogeneity.

This is an excellent point and is an issue we are grappling with right now because it is central to how SWATs work. As Reviewer #2 guessed, we have plans for more Trial Forge Guidance documents and the next one will be on when to start (and stop) a SWAT evaluation and the thorny issue at the heart of that is the degree to which it is sensible to combine studies done in different populations, disease areas and settings.

We have a draft of that Guidance document but we are not yet agreed on how best to address the context issues the Editor raises and we need some more discussion within the Trial Forge group. We have added two sentences to the end of the Analysis section of the Box to raise the issue but say we will give more guidance in the future:

‘As with all meta-analysis, judgements need to be made about whether it is sensible to combine studies done in different populations, disease areas and settings. This issue will be the topic of future Trial Forge Guidance.’
2. Please, check for consistency throughout the paper in the use of SWAT versus SWATs when appropriate

Done.

3. On page 9, the wording "evaluating a SWAT in a trial" sounds odd, i.e., in “The SWAT has been already evaluated in several trials” (lines 17-19), or “…embedding a SWAT evaluation in trials”. Is a SWAT evaluation that is embedded in a trial, or a specific trial method (e.g. bespoke PIL versus standard PIL) that is evaluated using a SWAT (which already means "a study embedded in a trial")?

We understand the Editor’s point and it is indeed odd if we unpack the S, W, A and T to ‘The Study Within A Trial has already been evaluated in several trials.’ when reading or speaking about the SWAT. However, people tend not to unpack the letters and the word ‘SWAT’ just becomes a noun for the particular evaluation.

So, while we completely understand and accept the Editor’s point, the way ‘SWAT’ is used is more as a noun than an abbreviation and we prefer to keep things in the paper the way they are because it reflect current usage and, additionally, saying ‘The SWAT has been evaluated in several trials’ while a bit odd if unpacked, does emphasises the need for SWATs to be evaluated in several trials, which would be lost in ‘The SWAT has been evaluated’.

Two author additions

While addressing the above comments, we have made two additional changes ourselves, both to the Ethics section of Box 1. The new text is below, marked in red:

‘Ethics

Ethical approval guidelines and regulations for conducting research in humans vary between countries. Depending on the specific SWAT protocol being evaluated, it is advised that the researcher checks national guidance and discusses whether ethical approval is required with their institutional or local ethical committee.

‘SWATs are generally low risk and it is rare for them to impose additional burden or risk on participants and consequently it will not usually be necessary to get individual consent from
participants. Indeed, in many cases individual consent may not be appropriate. It may confuse patients as to what they are consenting to, and may impact on their behaviour if they are aware that different recruitment methods are being tested, confounding the evaluation.’