



Retention: Open trials (ID Rec1)

General

This document uses the five criteria listed in Trial Forge Guidance 2 'How to decide if a further Study Within A Trial (SWAT) is needed' (<https://doi.org/10.1186/s13063-019-3980-5>). The criteria are listed in Appendix 1 at the end of this document.

Do we need more evaluations of open trials?

Yes.

Why do we need more evaluations and in what sort of host trial?

A trial team is likely to consider information about the following essential when deciding whether a further evaluation of an open trial compared to blinded trial should form part of their recruitment strategy:

- i) effect on the internal validity of the host trial
- ii) effect on recruitment
- iii) cost

Applying the five criteria

Outcome availability– Data are available for i) and ii).

GRADE– The overall GRADE certainty in the evidence is high. *Criterion not met* (the GRADE certainty in the evidence for all essential outcomes is not lower than 'high').

Cumulative evidence– There are only two trials and it seems too early to claim that the cumulative meta-analysis has converged. *Criterion met* (the effect estimate for each essential outcome has not converged).

Context– The PICOT for the available evidence is:

- **P** – One 2004 host trial was done in the UK and involved 538 participants aged 70 years or over, attending a fracture clinic or orthopaedic ward. The second trial, also 2004, was done in Estonia and involved 4295 postmenopausal women aged 50 to 64.
- **I** – The host trial intervention in the UK trial was vitamin D. The Estonian trial evaluated hormone replacement therapy. Both are cheap, easy to obtain drugs.
- **C** – The UK trial comparator was calcium, no tablets or placebo. The Estonian trial comparator was no treatment or placebo.
- **O** – Both studies measured recruitment to the host trial. Underlying recruitment was 37% in the UK trial and 65% in the Estonian trial.
- **T** – Both trials are now old although it is doubtful that the impact or otherwise of blinding on human behaviour has changed a great deal since then.

Considering the above, leads to ***Criterion partially met*** (a new evaluation is likely to contain several elements in the PICOT that are importantly different to those in the three existing evaluations).

Balance- participants- Randomisation to unknown treatments is widely cited (e.g. <https://doi.org/10.1002/14651858.MR000045>) as a barrier for participants to agree to taking part in trials. It is likely that potential participants' preference is to know what treatment they are receiving, i.e. in favour of open trials. Knowing the treatment could bring a downside (e.g. anxiety) but probably no more than not knowing which treatment was being taken. ***Criterion not met*** (the balance of benefit and disadvantage to participants in the new host trial and/or SWAT is clear).

Balance- host trial- The benefit to the host trial is a large increase in recruitment. The potential disadvantage to the host trial is an adverse effect on the integrity of the trial (i.e. the internal validity). Blinding is a key component of internal validity and routine use of open trials is likely to undermine the integrity of at least some trials, especially those where outcomes are subjective, or participants and staff are able to alter the intervention delivery by knowing what treatment is being given.

Many trials however are already open because it is impossible to blind them (e.g. many lifestyle change or educational trials, most surgical trials). Measures can be put in place to protect integrity, to a large degree at least, generally through blinded outcome assessment and/or objective outcomes. Recent evidence (<https://www.bmj.com/content/368/bmj.l6802.long>) found no average difference in estimated treatment effect between trials with and without blinded patients, healthcare providers, or outcome assessors. These results could reflect that blinding is less important than often believed, or weaknesses in this piece of work. Blinding was recommended where feasible.

To what extent internal validity is affected is a judgement but this is a familiar judgement to trial methodologists, statisticians and trialists. ***Criterion not met*** (the balance of benefit and disadvantage to those running the host trial is clear).

Considering the responses across all five criteria leads us to conclude that further evaluation of open trials might be needed but is probably not a priority. Any further evaluations should:

- Do not involve cheap, easy to obtain drugs.
- Be outside the UK or Estonia.
- Are expected to have underlying recruitment below 40%.

Additionally, collecting cost information would be useful though not essential.

Appendix 1

The five Trial Forge Guidance 2 criteria for deciding when a new evaluation of a SWAT intervention is needed (from <https://doi.org/10.1186/s13063-019-3980-5>).

The five proposed criteria for deciding whether the intervention needs another evaluation in a SWAT. The more criteria that are met, the more likely we are to conclude that further evaluation in a SWAT is appropriate.

1. *GRADE*: the GRADE certainty in the evidence for all key outcomes is lower than 'high'.ⁱ
2. *Cumulated evidence*: the cumulative meta-analysis shows that the effect estimate for each outcome essential to make an informed decision has not converged.^{ii, iii}
3. *Context*: the range of host trial contexts evaluated to date does not translate easily to the context of the proposed SWAT^{iv}. For the proposed SWAT consider PICOT:
 - P – is the population in the host trial so different from those already included that the current evidence does not provide sufficient certainty?
 - I – are the health interventions in the host trial so different from those already included that the current evidence does not provide sufficient certainty?
 - C – is the comparator in the host trial so different from those already included that the current evidence does not provide sufficient certainty?
 - O – is the SWAT outcome(s) so different to those used in the existing evaluations that that the current evidence does not provide sufficient certainty?
 - T – in the time since the existing evaluations were done, have regulatory, technological or societal changes made those evaluations less relevant?
4. *Balance– participants*: the balance of benefit and disadvantage to participants in the host trial and/or the SWAT is not clear^v.
5. *Balance– host trial*: the balance of benefit and disadvantage to the new host trial is not clear^{vi}.

Notes

- i. A GRADE assessment of 'high' means that we are confident that the true effect lies close to the estimate of effect coming from the cumulative meta-analysis. In Cochrane's deliberations as to when to close a Cochrane Review (<https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.ED000107/full>), the collaboration chose not to require 'high' GRADE certainty in the evidence because it was felt that this may not always be achievable. Although we recognise the pragmatic nature of this, we recommend 'high' in our criteria because SWATs are usually simple studies for which it should be possible to generate high certainty evidence. We will, however, keep this criterion under review to consider whether it needs relaxing.
- ii. This is a judgement that depends on the behaviour of the effect estimates and on whether the confidence intervals include the threshold for an important benefit (or disadvantage). For example, if there is drift in the effect estimates of a meta-analysis but the confidence intervals around the estimates are consistently above what you think is an important benefit (or below a relevant disadvantage) then the cumulative meta-analysis can be judged to have converged despite movement in the effect estimates. For more on GRADE see <http://www.gradeworkinggroup.org>.
- iii. A cumulative meta-analysis requires the same outcomes to have been measured in the same way in the studies to be combined. Most SWAT protocols specify just one or perhaps two outcomes, which reduces the scope for different outcomes between evaluations. Tighter specification of outcomes on SWAT protocols would help even more (e.g. retention sounds simple but could mean the proportion of participants who remain in the trial, the proportion who return a form, or the proportion who fully complete all forms). Core outcome sets for trial processes may help and this is being done in ELICIT for interventions to improve informed consent²⁴.
- iv. This is to provide reassurance about the applicability of the result to different types of trials. Care is needed to avoid a default position of insisting on an evaluation in every conceivable context. In other words, is there any reason to believe that the intervention would *not* work in your context given the contexts already studied?

It is possible that evidence from SWATs will eventually splinter off to focus specifically on certain contexts but, for now, we suggest pooling evaluations of the same intervention because there are so few SWAT evaluations of any intervention and this pooling will provide a basic foundation on which to build.

- v. Where there may be no conceivable benefit or disadvantage for participants, they should be considered as balanced.
- vi. A benefit might be that the host trial recruits faster, or its data quality is improved. Examples of disadvantages might be that there are added costs to the host trial, or that a new task is introduced into the workload of trial managers.