Retention: Electronic prompts (ID Ret2)

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 the electronic prompts?

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 electronic prompt should form part of their retention strategy:

i) effect on questionnaire response rate (retention)
ii) cost
iii) participant irritation at receiving an electronic prompt

Applying the five criteria

Outcome availability– Data are only available for questionnaire response rate.

GRADE– The overall GRADE certainty in the evidence is moderate. ***Criterion met*** (the GRADE certainty in the evidence for all essential outcomes is lower than 'high').

Cumulative evidence– There are only three 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** – All three host trials were done in the UK between 2011 and 2015 by York Trials Unit, with a total of 710 participants. The trials focused on migraine (men and women, 19-65 years); back pain (men and women, 20-65 years); and COPD (men and women, 35-82).
- **I** – The host trial intervention in the migraine trial was evaluating a food elimination diet. The back pain trial evaluated yoga and the COPD trial was looking at case finding to identify people at risk.
- **C** - The host trial comparator in all three trials was usual care. In other words, no food elimination, no yoga or no case finding.
- **O** – All studies measured retention to the host trial. Underlying retention ranged from 61% (COPD trial) to 86% (back pain trial). None started with really poor retention.
- **T** – All three trials are relatively recent and the behaviour of responding to the electronic prompt is unlikely to be more or less receptive in 2020 than in 2011. It

is possible that electronic alternatives to SMS and email may work better for some segments of the community. Not all people have access to a mobile 'phone or computer.

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– The individual participant has little to gain from the intervention. The participant may potentially be annoyed at the wording of the prompt although anecdotal information from the authors of all three evaluations suggest that no or very few participants react negatively, certainly not enough to complain (personal communication). Some participants did need reassurance that mobile telephone numbers would not be shared with companies that might then cold-call them. ***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 modest increase in response rates. The potential disadvantage to the host trial is the costs of providing electronic prompts. This is uncertain at present although additional costs are likely to be low. In the UK, the cost of sending automatic SMS prompts is around £0.08 per message (from https://www.jclinepi.com/article/S0895-4356(15)00024-4/fulltext). Email is essentially free per message. Setting up automatic systems will have a cost, which is unclear. ***Criterion met*** (the balance of benefit and disadvantage to those running the host trial is not clear).

Considering the responses across all five criteria leads us to conclude that further evaluation of electronic prompts is needed. Priority should be given to evaluation in trials that:

- Are drug trials.
- Are done outside the UK.
- Are expected to have underlying retention below 60%.

Additionally, collecting cost information would be useful, as would be documenting the participants' views on the electronic prompt.

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-analyses 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 consent[24].

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.