**Project Title:** Increasing efficiency in mental health trials by developing statistical methods to overcome challenges in crossover designs

**Lead applicant:** Brennan Kahan

**Co-applicants:** Rafael Gafoor

**Introduction**

Crossover trials are a more efficient study design than the frequently used parallel group randomised controlled trial (RCT). In crossover trials, patients are randomly allocated to either receive treatment A then treatment B in periods 1 and 2 (possibly with a washout period in between) consecutively, or in the reverse order. Crossover trials are more efficient because they remove the between-subject variance that exists in a parallel arm trial, thus requiring much smaller sample sizes.

Mental health trials often have poor recruitment. This can lead to studies being underpowered, resulting in a poor evidence base. Because crossover trials require smaller sample sizes than parallel group RCTs, they could help to overcome the poor recruitment seen in mental health trials. A crossover design requires at least 50% fewer participants than a parallel group design (and could require 75% fewer in some cases).

However, a barrier to the use of crossover trials in mental health is that patients should be in the same health state at the start of both periods. In a trial of antidepressants, patients with a baseline depression score that is 30% lower at the start of their second period versus the first period, may respond differently to their next treatment. This is referred to as “carryover” and may introduce bias in the comparison between treatments. No current statistical methods are able to adequately address this issue. Because fluctuations in patients’ health status (such as in anxiety where the outcome is psychologically dependent) are common in mental health trials, crossover trial design may be less often used due to concerns over “carryover”.

We propose a novel method to address these imbalances which enables unbiased estimation while still maintaining efficiency advantages. We have investigated an approach which treats differences in baseline values across periods in individual patients as intercurrent event (an event emerging after randomisation which affects interpretation of the outcome) and handles it using a hypothetical strategy. This approach may facilitate unbiased estimation of what the treatment effect would have been had patients been in the same health state at the start of both periods, while still maintaining sample size savings compared to standard parallel group trials.

**Objectives**

* To determine whether our proposed approach does in fact provide unbiased estimation in mental health crossover trials affected by carryover
* To determine whether our proposed approach still provides strong efficiency gains compared to parallel group designs
* To compare our approach against other competing methods to evaluate whether its performance is superior in terms of bias, coverage, and precision

**Progress**

We have completed a project that used computer simulation to compare our proposed method against other competing methods for mental health crossover trials.

Briefly, we used two data generated mechanisms, the first where carryover was not present, and the second where it was. We included a scenario with no carryover to ensure our proposed method works well even when carryover is not present.

For both scenarios, we generated data from a crossover trial with a continuous outcome whose distribution was meant to mimic a depression or anxiety score with approximate range from 0-100. Entry into the trial required a baseline score greater than 30. We generated carryover by generating data such that some participants had a baseline score <30 at the start of period 2 (i.e. so they would not have been eligible to take part if they had had that score at the start of period 1; but because they were eligible at the start of period 1, they remain eligible for period 2, regardless of their baseline score), and including a treatment-by-baseline score interaction. Here, presence of a score <30 at the start of period 2 represents a form of observed carryover.

For the scenario with no carryover, we set the treatment-by-baseline score interaction term to 0 (i.e. the treatment effect was the same regardless of whether participants had carryover); and for the scenario with carryover, we set there to be a non-negligible treatment-by-baseline interaction, such that there was a treatment effect for participants with a score >30 (who were eligible), and no effect for participants with a score <30.

We used 15,000 replications for each scenario, and used a trial size of n=50.

We compared four methods of analysis:

1. A mixed-effects model, with a random intercept for participant (standard method)
2. Our novel method, which treats observed carryover as an intercurrent event and estimates a hypothetical strategy (i.e. what the treatment effect would have been, had no carryover occurred). We did this by (a) setting period 2 outcomes to missing for participants with observed carryover; and (b) fitting a mixed-effects model, which implicitly imputes what the period 2 outcomes would have been had the carryover not been present
3. A 2-stage analysis, where a test for carryover is performed, and a mixed-model is used if the test is not significant, and the period 2 outcomes are discarded if the test is significant and a linear regression model is used to analyse the period 1 data only (standard method).
4. Using only period 1 data (as though it were a parallel group trial)

We compared each method of analysis on three measures: (a) bias; (b) precision (as measured by the method’s empirical standard error); and (c) coverage.

Results can be seen in Figures 1 and 2. When there is no carryover (i.e. the treatment effect was the same regardless of whether baseline status changed between periods), all methods are unbiased, as expected. However, our novel method has much better precision than all competing methods, except for the standard mixed-model, however the difference is very small. Similarly, our method has good coverage.

When there is carryover (which is the scenario of primary interest), the standard methods of analysis (standard mixed-model, 2-stage analysis) are both notably biased; however, our method is unbiased, i.e. has successfully managed to correct for the carryover. Our method retains its high level of precision and good coverage.

Overall, our method appears to be a robust and reliable option for carryover trials with potential carryover, and performs as well as or better than other methods across both scenarios.

**Figure 1 – Simulation results (setting without carryover)**

A graph of a number and a number

AI-generated content may be incorrect.

**Figure 2 – Simulation results (setting with carryover)**

A graph of a number of people

AI-generated content may be incorrect.

Of note, while we had initially planned to use funding from this grant to obtain PPI feedback on our results, Co-I Rafael Gafoor has an existing PPI group for mental health studies which has an alternate source of funding. As a result, and due to tight timelines between completion of the simulation study and grant end date, we decided to meet with the PPI group after the grant end date, using Co-I Gafoor’s alternate source of funding.

**Outputs**

There are no outputs yet; a draft manuscript is in preparation.

**Future plans**

We intend to complete our draft manuscript and submit it for publication. Once the manuscript is complete, we will meet with Co-I Gafoor’s PPI group to solicit feedback on the results, and obtain their views on both the relevance of this approach for patient’s, as well as dissemination opportunities.

Finally, given our proposed method’s success in the simulation study, we may plan to seek further funding to develop the method further, to ensure it is sufficiently robust to allow it to be routinely used in the design and analysis of crossover trials (e.g. by performing a more comprehensive set of simulations; perform re-analyses of published crossover trials using this method; and seeking input from various stakeholders around their views of this method, and considering what further work would be required such that they would seek to use it in their own trials).