Missed opportunity: Efficient use of baseline can increase power and reduce bias in psychiatric crossover trials

Cristan Farmer, Dede Greenstein, and Carlos Zarate, Jr. Experimental Therapeutics and Pathophysiology Branch, Intramural Research Program, National Institute of Mental Health

Can baseline measurements be better leveraged to increase power and reduce bias in psychiatric crossover trials?

Inclusion of baseline measurements in analyses of RCT can increase power and reduce the required sample size by increasing the precision of the estimate of the drug effect.

The basic ANCOVA approach is not valid for crossover trials. Few studies have implemented best-practice recommendations for including baseline in analysis.

Depending on characteristics of the trial and outcome measure, efficient use of baseline is equivalent to increasing enrollment by up to 30%.

BACKGROUND

- The benefits of a crossover trial include that (a) the same number of observations may be obtained from fewer participants and (b) the same degree of precision in estimation may be obtained using fewer observations.
- In some cases, the inclusion of baseline measurements can further increase precision in crossover trials. If baseline scores are only weakly related to outcome scores, their inclusion *will not* increase precision.
- Several approaches to integrating baseline data are commonly used, including the analysis of change scores and the addition of baseline values as a covariate.
- However, the analysis of change scores does not usually increase the precision of the estimate relative to the omission of baseline altogether, and can be biased.
- While covarying baseline can increase precision, caution must be exercised because the treatment effect estimate is subject to cross-level bias. Kenward & Roger (2010)[1] published a solution that has not been widely adopted.
- In this poster, we:

Describe the Kenward & Roger best-practice recommendation (referred to as KR)

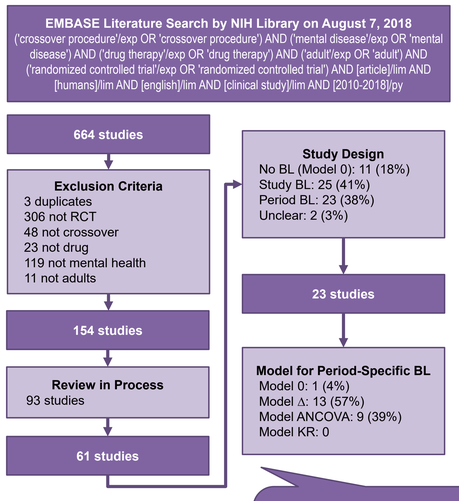
Formally evaluate the extent to which crossover trials in psychiatry have adopted the KR approach

Illustrate the potential benefits of the KR approach using archival crossover RCT data

LITERATURE REVIEW

- We performed a systematic review of the literature, including all RCTs in psychiatry published after 2010.
- Each study was coded for inclusion/exclusion criteria (see **Figure 1**). Studies which met these criteria were reviewed and the treatment of baseline was recorded. Here, we present initial results (n = 61), as coding is ongoing.

Figure 1. Results of literature review



SUMMARY | Few (if any) crossover studies in psychiatry use the KR approach.

BEST-PRACTICE RECOMMENDATION

SUMMARY | The ANCOVA approach causes bias, but this is abrogated by an additional term for the average of each participants' baselines.

- When baseline is entered as a single covariate in a model with a random subject effect, cross-level bias occurs because the model implies that the within-subject correlation between baseline and outcome is equivalent to the between-subject correlation. This is extremely unlikely.
- This results in an inflated estimate for the effect of baseline on outcome, leading to over-adjustment of the outcome and a biased estimate of the treatment effect.
- The KR solution is to allow for separate regression coefficients to be estimated for the within-subject and the between-subject effects of baseline. Two separate baseline terms are added to the model:

1. Period-specific baseline (as in ANCOVA; this is the within-subject effect)
2. Average of period-specific baselines per subject (this is the between-subject effect)

ILLUSTRATION OF METHODS

- To illustrate the KR approach relative to the most commonly observed methods found in our literature review (see **Figure 1**), we applied each method to archival data from three double-blind crossover RCTs performed at the NIMH.

1. Study 1: Trial[2] of AZD6765 in 22 participants with major depression.
2. Study 2: Combined trials[3,4] of ketamine in 41 participants with bipolar 1 disorder.
3. Study 3: Trial[5] of ketamine in 42 participants with major depression or bipolar 1 disorder.

- MADRS Total Scores from timepoints baseline (-60 minutes), 230 minutes, Day 1, and Day 3 were analyzed as the outcome measure for all trials. Fixed effects included drug, time, and drug*time (age, sex, and infusion added as covariates). Degrees of freedom were adjusted using the Kenward-Roger formula. Within-subject residual covariance was modeled by drug with an unstructured matrix per infusion, plus a random intercept.
- We present the treatment effect estimates at Day 1 (in units of MADRS Total Score), and relative changes in sample size estimates based on the standard error of the treatment effect estimate for four models:

1. Model 0: Baseline data excluded from analysis.
2. Model Δ: Change from period-specific baseline analyzed as outcome.
3. Model ANCOVA: Period-specific baseline entered as covariate.
4. Model KR: Period-specific and subject-average baseline entered as covariates.

SUMMARY | Whether and how baseline is used can alter effect size estimates. The KR model yielded the most precise estimates, which translates to smaller required sample sizes (see Figure 3).

Figure 2. Comparison of drug effect size (and standard errors) for each model

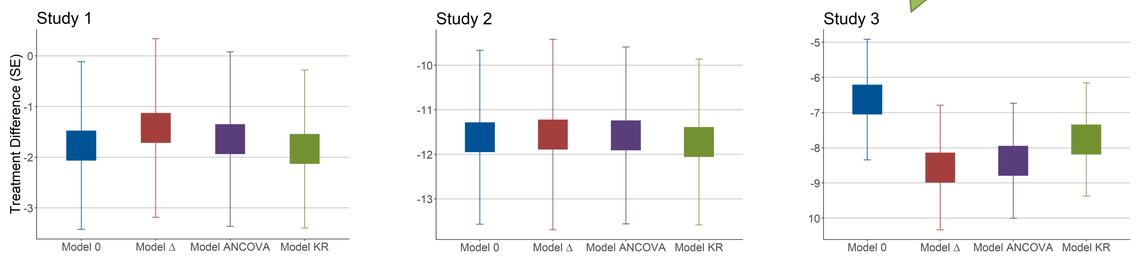
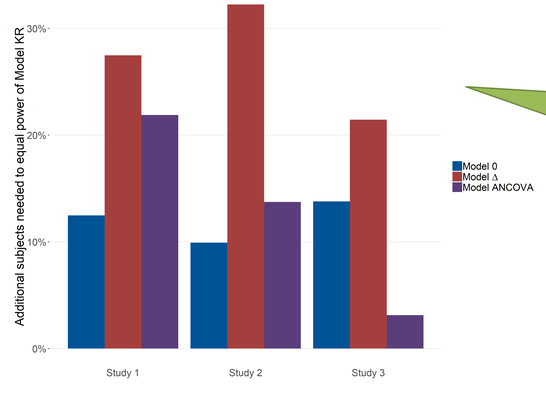


Figure 3. Sample size required for models of baseline analysis compared to KR



SUMMARY | All three other models require a larger sample size (up to +30%) than the KR approach. Model Δ affords less power than every other approach.

DISCUSSION & CONCLUSIONS

- The most common approach to using baseline in psychiatry crossover trials appears to be as a change score. This is the least efficient approach of those reviewed, and may yield biased estimates of the treatment effect.
- The inclusion of baseline as a covariate in crossover trials is intended to improve the precision of the treatment effect, but it may result in a biased estimate.
- The addition of the average baseline value can eliminate this bias and yield more accurate estimates of the treatment effect and more power to detect this effect.

REFERENCES

1. Kenward, M. & Roger, J. (2010), doi.org/10.1093/biostatistics/kxp046
2. Zarate et al. (2013), doi.org/10.1016/j.biopsych.2012.10.019
3. Zarate et al. (2012), doi.org/10.1016/j.biopsych.2011.12.010
4. Diazgranados et al. (2010), doi:10.1001/archgenpsychiatry.2010.90
5. Nugent et al. (2018), doi.org/10.1038/s41380-018-0028-2

CONTACT INFORMATION

Cristan Farmer, Ph.D.
farmerca@mail.nih.gov

National Institute of Mental Health