Methodological Comments: McFarlane et al

Helena Chmura Kraemer*

Department of Psychiatry and Behavioral Sciences, Stanford University (Emerita), Palo Alto, CA

*To whom correspondence should be addressed; Department of Psychiatry and Behavioral Sciences, Stanford University (Emerita), 1116 Forest Avenue, Palo Alto, CA 94301, US; tel: 650-328-7564; e-mail: hckhome@pacbell.net

Does the McFarlane et al study1 provide a model for practical trials? Unfortunately, not. The methodological problems are here reviewed and used as a springboard to set out methodological criteria that might define such a model.

Key words: practical trials/methodology/criteria

A “practical trial” is one that seeks to answer those questions clinicians need answered to make the best possible clinical decisions for patients. Does the McFarlane et al study1 provide a model for practical trials? Unfortunately, not. The authors themselves give warning that the results are likely to be invalid except under special circumstances.1

A reminder of the design: A baseline variable is used to identify the “high risk” and “low risk” patients, with “high risk” assigned to treatment (T) and “low risk” to control (C). Two straight lines with the same slope are fitted to the observed outcomes and are used to extrapolate what the responses of “low risk” to T and “high risk” to C would have been. The vertical distance between these 2 parallel lines indicates the treatment effect.

Figure 1a illustrates those circumstances (parallel linear trajectories) on which correct inferences are based. However, in figure 1b, the T and C have the same nonlinear trajectory and linear extrapolation leads to concluding that T is superior to C (T > C) when T and C are clinically equivalent (T ≈ C). In figure 1c, the lines are not parallel, and here extrapolation leads to concluding that T ≈ C, when in fact T > C for some and T < C for others. Nonsignificant tests of linearity and interaction do not support the contention that these assumptions hold: Absence of proof (nonsignificance) is not proof of absence. Nonsignificant results generally mean inadequate power to detect nonlinearity or interaction.

At initiation of any randomized clinical trial (RCT), there is rationale and justification to believe that T may be better than C. The point of a RCT is to prove beyond reasonable doubt that T is better than C. Here, the conclusion can only continue to be that T may be better than C. What characterizes an ideal practical trial?

1. Clinical equipoise: Reasonable doubt of what the right answer is and a design that will disturb such equipoise.2
2. A patient sample representative of the population clinicians need to deal with.
3. Credible C, a treatment clinicians might choose if T were not available.
5. A single primary outcome that reflects the benefit to harm balance in individual patients (avoiding multiple testing).3–5
6. If an observational trial, use of propensity6–8 or other such analyses to remove as much of the bias as possible. If a RCT, an approach consistent with RCT methodology. Any crucial mathematical assumptions must be empirically supported.
7. Adequate power to detect clinically significant treatment effects.
8. Emphasis on clinically interpretable effect sizes, not on P values.
9. After the primary analysis is done, exploratory moderator analysis to identify the subgroups, if any, for whom T > C, T ≈ C, T < C.9–11

In short, the ideal practical trial is designed, not to avoid the rigors of standard RCTs, but to enhance methodology to provide practical guidance to clinicians in order to reduce the burden of health problems.

Acknowledgment

The authors have declared that there are no conflicts of interest in relation to the subject of this study.
Fig. 1. Three hypothetical situations with solid lines indicating the true outcome curves for T and C as a function of the selected baseline variable, the dotted lines indicate extrapolations that would be used with the parallel line assumption. The vertical axis is the outcome and the horizontal axis the baseline factor used for assignment (cutpoint here at 50). Only the right side (baseline over 50 here) of the T response curve and the left side of the C response curve are actually observed. The vertical separation of the dotted lines indicates the estimated treatment effect size using this design. The separation between the solid lines indicates the true treatment effect, which may not be the same for patients with differing baselines.

References