Statistical Science
2023, Vol. 38, No. 2, 209–211
https://doi.org/10.1214/23-STS865A
Main article: https://doi.org/10.1214/22-STS865
© Institute of Mathematical Statistics, 2023

Comment: A Quarter Century of Methodological Research in Response-Adaptive Randomization

Anastasia Ivanova and William F. Rosenberger

1. INTRODUCTION

We congratulate David Robertson and his colleagues for the first *Statistical Science* paper reviewing general response-adaptive randomization procedures since Rosenberger (1996) 26 years ago. In this discussion, we briefly revisit Rosenberger (1996) and describe a few examples under which response-adaptive randomization should be considered for use in practice.

Response-adaptive randomization has been a challenging area in which to conduct research, not because of any difficulties in publishing methodology, but rather the biases that have been present since the 1980s and are still flourishing today. Some of these have been pointed out and debunked by Robertson et al. We start with two points:

1. The ECMO trial, already described in Robertson et al., was a highly atypical clinical trial with 11 patients, conducted 37 years ago during the Reagan Administration, 20 years before smart phones. Any relevance to this trial today concerns ethics, not statistics. We know why this ECMO trial failed, and there have been more than three decades of methodological work in RAR procedures since then to remedy this failure. This work has included assurances that trials are sufficiently powered, have adequate sample sizes and do not create massive imbalances with treatment arms with no or almost no patients. Even so, a recent article Proschan and Evans, 2020 cited ECMO as "a bad start" and a compelling reason to "resist the temptation of response-adaptive randomization." Let us move on as scientists.

2. Anyone can find a RAR procedure that does not work under specific circumstances and run a simulation showing it. It is certainly valuable to know when RAR procedures do not work. But it is *unscientific* to generalize to all RAR procedures, or classes of RAR procedures and draw conclusions that RAR procedures are "discouraged" or that one should "just say no" (e.g., Korn and Freidlin, 2011). Robertson et al. demonstrate clearly that such procedures *can work* under certain circumstances, and like all statistical methods, have unique settings in which particular procedures should be preferred or not used.

Anastasia Ivanova is Professor of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27516, USA (e-mail: aivanova@biostat.unc.edu). William F. Rosenberger is University Professor, Department of Statistics, George Mason University, Fairfax, Virginia 22030, USA (e-mail: wrosenbe@gmu.edu).

2. COMPARING TO ROSENBERGER (1996)

One looks back at the naïvete of the RAR procedures explored in Rosenberger (1996) and recognizes the magnitude of methodological work that has been achieved in the past quarter century. Only half the paper concerns RAR for phase III clinical trials, and the only procedures explored are based on the generalized Pólya urn, and treatment effect mappings, which would today be considered too ad hoc, and have too much variability to use in practice.

One thinks back to this era, and recognizes that the concepts of platform trials, baskets and buckets and enrichment were not even in the lexicon, and precision medicine was just at its infancy. While not necessarily in the realm of RAR, all of these trial designs are adaptive in some sense.

The optimal allocation approach by Rosenberger et al. (2001), the characterization of the variability of the RAR procedure (Hu and Rosenberger, 2003), followed by the concept of asymptotically best procedures Hu, Rosenberger and Zhang, 2006, led to the template favored by Hu and Rosenberger in their 2006 book *The Theory of Response-Adaptive Randomization*. The improved efficiency of these designs made much of the work in the 1980s and 1990s irrelevant. Extensions with Bayesian priors are given by Xiao, Liu and Hu (2017) and Proper and Murray (2022).

On the other side of the Atlantic, Villar and her colleagues were reviving interest in multiarmed bandits and discovering new ways of conducting clinical trials using them (see Villar, Bowden and Wason, 2015a, for a *Statistical Science* review). Villar and her colleagues take the approach of using the Gittins index and randomization to assign more patients to the better treatment (see Villar, Wason and Bowden, 2015b).

Both of these approaches deal with the issues of the optimal allocation to achieve some objective (e.g., higher