Statistical Science
2023, Vol. 38, No. 2, 216–218
https://doi.org/10.1214/23-STS865C
Main article: https://doi.org/10.1214/22-STS865

Comment: Advancing Clinical Trials with Novel Designs and Implementations

Lorenzo Trippa and Yanxun Xu

We extend our congratulations to Robertson and coauthors [12] for their comprehensive overview of response adaptive randomization (RAR) in clinical trials and insightful comparative analyses. Their contribution is noteworthy for clearly demonstrating the diversity of RAR designs and algorithms that utilize the available data to update randomization probabilities throughout the enrollment period. For instance, several forms of Bayesian RAR tend to accelerate the enrollment to the most promising arms, while other RAR strategies vary the randomization probabilities to enhance trial power compared to fixed or balanced randomization. The authors accurately highlight the wide range of statistical designs based on RAR, each with different benefits and risks compared to balanced randomization.

The significance of showing marked differences of the operating characteristics of various RAR designs lies in the fact that the views of stakeholders, including biostatisticians and clinical trial investigators, on the key aspects of RAR are often influenced by their experience with a single algorithm, a narrow subset of RAR designs explored in a publication or a few RAR clinical trials with favorable or unfavorable patient allocation. Such biases can obstruct the development of effective clinical trial designs.

Additionally, it is essential to consider the broad range of settings where RAR and adaptive trial designs can be utilized in clinical research. Context-specific evaluations of RAR designs are necessary as their advantages and disadvantages over traditional randomization vary with study-specific factors such as the expected accrual rate or the risk of time trends of the enrolled population. To fully assess the potential of RAR and adaptive trial designs, it is crucial to account for these context-specific considerations.

In our experience with adaptive platform trials [1, 5, 19], it is necessary to consider both operational/implementation complexity and potential efficiencies over fixed randomization. A relevant aspect of RAR that is often overlooked is the possibility of reducing the number of patients allocated to inferior arms. Early in the trial, data may suggest that an experimental therapy has lower efficacy compared to the control. Despite this, the experimental arm is not dropped from the study at this stage because of insufficient evidence. However, frequent interim data analyses and a preplanned mechanism, such as RAR, which allows for reducing or temporarily halting enrollment to the experimental arm until additional outcome information (e.g., survival data) becomes available, along with standard early stopping rules, can reduce the number of patients exposed to inferior treatments compared to less complex designs with fixed randomization. The RAR algorithm in a multiarm study can be customized to control the number of patients exposed to potentially inferior or toxic arms.

Lorenzo Trippa is Associate Professor, Dana-Farber Cancer Institute and Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA (e-mail: ltrippa@jimmy.harvard.edu). Yanxun Xu is Associate Professor, Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, Maryland, USA (e-mail: yanxun.xu@jhu.edu).

A Positive Perspective on the Future Use of Adaptive Trial Designs

Despite significant attention paid to RAR algorithms and adaptive trial designs in academic literature, their application in clinical trials is still limited, as noted by Robertson and colleagues. Predicting which trial designs will be implemented in the next decades is challenging, but we maintain a positive outlook on the future ability of clinical research across various areas, including oncology, to leverage the advancements in RAR designs and accelerate the development of new and effective treatments. Here, we discuss three motivations.

Collaboration-Centered Strategies for the Development of New Treatments

Our optimism is driven by the growing interest and utilization of multiarm and platform trial designs. These designs offer greater efficiency by evaluating multiple experimental treatments with a shared control arm, instead of conducting separate two-arm RCTs with similar control groups. Both multiarm trials and platform trials can employ fixed randomization or RAR. In both cases, they achieve substantial efficiency gains compared to traditional two-arm RCTs [14].

Collaboration among pharmaceutical companies, biostatisticians and other stakeholders in conducting clinical studies of multiple experimental treatments presents an

opportunity to identify innovative trial designs and economic models. Literature has shown that the benefits of using Bayesian RAR and more generally adaptive decisions increases with the number of treatments being tested [2]. A collaborative environment can also drive further elements of innovation, such as new economic models with shared risks and profits for pharmaceutical companies and investors [11], or new adaptive designs taking into account cost-effectiveness considerations [7]. Companies, funding agencies and investigators can share costs for several tasks such as developing curated real-world data sets or new predictive models. This can lead to improvements of RAR and other decision-making during the drug development process.

The transition from fragmented clinical research environments with trials that evaluate a single experimental therapeutic to collaboration-focused platform trials and other methods for sharing information and resources (see, e.g., Kotecha et al. [10]) opens new opportunities for successful implementations of RAR and more broadly, adaptive designs.

Improved Evaluations of Candidate Trial Designs

The second reason for our optimism lies in the continuous advancements in methods and software for evaluating candidate study designs. The use of simulations to compare the operating characteristics of candidate designs under various scenarios has been a common practice for many years. The tools for anticipating the operating characteristics of a study design in a specific context (e.g., a phase II trial in asthma) have been greatly improved. With the advancement of computing infrastructure and innovative methods (such as those described by Golchi [8]), the analyst can rapidly assess the variations of the operating characteristics of interest across plausible scenarios.

Data sharing initiatives, such as Yoda [13] and Project Data Sphere [9], and disease-specific data collections, are useful for comparing trial designs in a specific clinical setting. The analyst can utilize real-world data sets or patient-level data from previous trials to define realistic simulation scenarios consistent with previous studies. Recently, in [16] and [15], we employed direct subsampling of data sets from completed clinical studies to simulate trials in glioblastoma and compare candidate study designs. We used a subsampling procedure, similar to bootstrapping, to define scenarios that reflect various aspects of previous clinical trials, including enrollment rates and the joint distribution of pretreatment clinical profiles and outcomes.

There are several other effective approaches for rigorous and context-specific comparisons. For example, Broglio et al. [3] prospectively planned the comparison of two candidate trial designs using the data generated by a randomized trial of acute stroke interventions. An ideal extension of this approach would be a prospectively planned evaluation of candidate designs, possibly developed by different research groups, to identify strengths and weaknesses based on data generated from multiple randomized trials. To summarize, new and improved methods for accurate, comprehensive and context-specific assessments of candidate designs are emerging.

Effective approaches for illustrating the operating characteristics of adaptive clinical trials are critical for allowing stakeholders, such as patient representatives, to evaluate a candidate trial design. A related and less explored problem is the rigorous evaluation of the impact of early-phase designs and preplanned interim decisions, such as RAR, on the subsequent phases of drug development. Indeed the design of an early phase trial directly impacts on the accuracy of treatment effect estimates, which then affects the decision to continue the drug development with subsequent trials and various aspects of their design, including the sample size and the eligibility criteria.

Auxiliary Outcomes, New Biomarkers and Improved Prediction Models

Our optimism is also motivated by the rapid advancements in various fields including biology, statistics and others, which we believe will improve RAR and other preplanned adaptive decisions in future clinical trials. For instance, innovative approaches to analyze imaging data have produced ground-breaking results in radiology [6]. Similarly, recent developments in biomarkers and assays have enabled scientists to predict clinical outcomes such as disease progression and survival in oncology and other areas of medicine. Novel biomarkers can be useful to identify patient subpopulations and modify the eligibility criteria at interim analyses in adaptive enrichment designs [18]. The integration of patient-level information using novel statistical and machine learning approaches is leading to improved accuracy in predicting clinical outcomes.

These advancements and joint models combining information available soon after randomization, such as circulating tumor cells, with primary outcomes could improve the prediction of the trial results and identify early during the study promising treatments and patient subpopulations. Multioutcome models (e.g., joint models of progression-free survival and overall survival in oncology) have been used to adapt the randomization probabilities and show promise in settings where treatment effects tend to be more pronounced on auxiliary outcomes compared to primary outcomes [4, 17]. Evaluating the performance of trial designs that use joint models incorporating predictive variables and primary outcomes for adaptive decisions is particularly challenging. Real-world data sets could be useful to assess the value of predictive variables and compare trial designs.

FUNDING

The first author was supported by the NIH Grant 1R01LM013352.

The second author was supported in part by NIH Grant 1R01MH128085 and NSF Grant 1918854.

REFERENCES

[1] ALEXANDER, B. M., TRIPPA, L., GAFFEY, S., ARRILLAGA-ROMANY, I. C., LEE, E. Q., RINNE, M. L., AHLUWALIA, M. S., COLMAN, H., FELL, G. et al. (2019). Individualized screening trial of innovative glioblastoma therapy (INSIGhT): A Bayesian adaptive platform trial to develop precision medicines for patients with glioblastoma. *JCO Precis. Oncol.* **3** 1–13.

[2] BERRY, D. A. (2011). Adaptive clinical trials: The promise and the caution. *J. Clin. Oncol.* **29** 606–609.

[3] BROGLIO, K., MEURER, W. J., DURKALSKI, V., PAULS, Q., CONNOR, J., BERRY, D., LEWIS, R. J., JOHNSTON, K. C. and BARSAN, W. G. (2022). Comparison of Bayesian vs frequentist adaptive trial design in the stroke hyperglycemia insulin network effort trial. *JAMA Netw. Open* **5** e2211616. https://doi.org/10.1001/jamanetworkopen.2022.11616

[4] BROGLIO, K. R. and BERRY, D. A. (2009). Detecting an overall survival benefit that is derived from progression-free survival. *J. Natl. Cancer Inst.* **101** 1642–1649. https://doi.org/10.1093/jnci/djp369

[5] CELLAMARE, M., VENTZ, S., BAUDIN, E., MITNICK, C. D. and TRIPPA, L. (2017). A Bayesian response-adaptive trial in tuberculosis: The endTB trial. *Clin. Trials* **14** 17–28. https://doi.org/10.1177/1740774516665090

[6] CHENG, P. M., MONTAGNON, E., YAMASHITA, R., PAN, I., CADRIN-CHENEVERT, A., PERDIGÓN ROMERO, F., CHARTRAND, G., KADOURY, S. and TANG, A. (2021). Deep learning: An update for radiologists. *Radiographics* **41** 1427–1445.

[7] FLIGHT, L., BRENNAN, A., CHICK, S. E., FORSTER, M., JULIOUS, S. and THARMANATHAN, P. (2022). Value-adaptive clinical trial designs for efficient delivery of research—actions, opportunities and challenges for publicly funded trials. Discussion paper, School of Health and Related Research, University of Sheffield.

[8] GOLCHI, S. (2022). Estimating design operating characteristics in Bayesian adaptive clinical trials. *Canad. J. Statist.* **50** 417–436. MR4429845 https://doi.org/10.1002/cjs.11699

[9] GREEN, A. K., REEDER-HAYES, K. E., CORTY, R. W., BASCH, E., MILOWSKY, M. I., DUSETZINA, S. B., BENNETT, A. V. and WOOD, W. A. (2015). The project data sphere initiative: Accelerating cancer research by sharing data. *Oncol.* **20** 464–e20.

[10] KOTECHA, G., VENTZ, S. and TRIPPA, L. (2023). Prospectively shared control data across concurrent randomised clinical trials. *Eur. J. Cancer* **181** 18–20. https://doi.org/10.1016/j.ejca.2022.11.038

[11] LO, A. W. and THAKOR, R. T. (2022). Financing biomedical innovation. *Annu. Rev. Financ. Econ.* **14** 231–270.

[12] ROBERTSON, D. S., LEE, K. M., LOPEZ-KOLKOVSKA, B. C. and VILLAR, S. S. (2023). Response-adaptive randomization in clinical trials: From myths to practical considerations. *Statist. Sci.*

[13] ROSS, J. S., WALDSTREICHER, J., BAMFORD, S., BERLIN, J. A., CHILDERS, K., DESAI, N. R., GAMBLE, G., GROSS, C. P., KUNTZ, R. et al. (2018). Overview and experience of the YODA project with clinical trial data sharing after 5 years. *Sci. Data* **5** 1–14.

[14] VENTZ, S., ALEXANDER, B. M., PARMIGIANI, G., GELBER, R. D. and TRIPPA, L. (2017). Designing clinical trials that accept new arms: An example in metastatic breast cancer. *J. Clin. Oncol.* **35** 3160–3168. https://doi.org/10.1200/JCO.2016.70.1169

[15] VENTZ, S., COMMENT, L., LOUV, B., RAHMAN, R., WEN, P. Y., ALEXANDER, B. M. and TRIPPA, L. (2022). The use of external control data for predictions and futility interim analyses in clinical trials. *Neuro-Oncol.* **24** 247–256. https://doi.org/10.1093/neuonc/noab141

[16] VENTZ, S., LAI, A., CLOUGHESY, T. F., WEN, P. Y., TRIPPA, L. and ALEXANDER, B. M. (2019). Design and evaluation of an external control arm using prior clinical trials and real-world data. *Clin. Cancer Res.* **25** 4993–5001.

[17] WANG, H. (2016). Response adaptive randomization using surrogate and primary endpoints. Ph.D. thesis, Virginia Commonwealth University, Richmond, VA.

[18] XU, Y., CONSTANTINE, F., YUAN, Y. and PRITCHETT, Y. L. (2020). ASIED: A Bayesian adaptive subgroup-identification enrichment design. *J. Biopharm. Statist.* **30** 623–638.

[19] XU, Y., MÜLLER, P., TSIMBERIDOU, A. M. and BERRY, D. (2019). A nonparametric Bayesian basket trial design. *Biom. J.* **61** 1160–1174. MR4013340 https://doi.org/10.1002/bimj.201700162