An Effective Visualization-Driven Tool for Power and Sample Size Estimation in Two-Stage General Randomized Basket Trial Designs

Desmond Zeya Chen[1,2,3], Sahil S. Patel[1,4,5], Aoqi Xie[1,2], Jiayin Chen[6], David Castle[7,8], Clement Ma[1,2,*]

Affiliations

1. Centre for Addiction and Mental Health, Toronto, Ontario, Canada.
2. Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada.
3. Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, ON, Canada.
4. Department of Statistical Sciences, Faculty of Arts and Science, University of Toronto, Toronto, Ontario, Canada.
5. Department of Statistical, North Carolina State University, Raleigh, North Carolina, United States.
6. Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada.
7. Department of Psychiatry, University of Tasmania, Tasmania, Australia.
8. Centre for Mental Health Service Innovation, Statewide Mental Health Service, Tasmania, Australia.

*Correspondence: clement.ma@camh.ca

Abstract: Basket trial designs with interim analysis have gained significant attention due to their adaptability, flexibility, and scalability. In response to the need for user-friendly tools that enhance the real-world applicability of these designs, we developed a web-based interface aimed at facilitating two-stage basket trial designs. Built using R Shiny, the tool was rigorously validated for output consistency by comparing it to an established R pipeline. Additionally, user testing was conducted to ensure the interface is intuitive and easy to use. The result is a freely accessible tool that provides effective and convenient visualizations for general basket trial designs with interim analysis, available at https://desmondzeyachen.shinyapps.io/AdaptiveTwoStageBasketTrialFeb14/. Future improvements may further expand the tool's capabilities to accommodate the increasing complexity of trial designs needed by the research community.

Keywords: Shiny App, Basket Trial, Interim Analysis, visualization, novel trial designs, Adaptive designs.

1. Introduction

A master protocol trial is a state-of-the-art clinical trial design that innovatively aggregates the evaluation of multiple treatments or disease subtypes[1]. Its flexibility, adaptability, and effectiveness aided it gaining popularity since the onset of the millennium[2]. After the first master protocol managed the Imatinib Target Exploration Consortium Study B2225 basket trial design[3,4], there has been a growing number of such trials being implemented[5,6]. Basket trials with biomarker enrichment have pioneered new frontiers in personalized medicine and thoroughly revolutionized the randomized trial design[7].

Prior work from Chen et al[8] has laid the statistical foundation for an adaptive basket trial design, it exquisitely addresses removal of given baskets based on interim results, while still accurately quantifies a trail-wide type 1 error rate, and a guidance for power. However, the impracticality of equal basket size assumption diminished its applicability. Such assumption is rarely fulfilled in real-word settings, considering the disproportionately distributed nature of disease subtypes prevalences among the population[9]. Following trial design guidance with such assumption may leads to a conservative, under-powered sample size estimate. After extensive investigation, we found an analytical solution to overcome this pitfall and developed a complementary software pipeline for efficient implementation and analysis[10].

1.1 Study Objectives and Existing Work

We aim to provide trial researchers with support in proposal writing, trial monitoring, and trial execution. Majority of trial design software facilitating multiplicity through multi-arm multi-stage (MAMS) designs, only a few explicitly address basket trial designs. Berry Consultants developed a commercial software called FACTS[11] (the “Fixed and Adaptive Clinical Trial Simulator”) , which offers simulation-based analysis with highly flexible options across different trial classes. FACTS can simulate basket trial designs with early futility and interim analysis. However, sample size estimation was not supported, as researchers must search through the range of sample sizes manually. Among open-source packages and pipelines, the Integrated Platform for Designing Clinical Trials (https://trialdesign.org) currently features two basket design trial R pipelines based on Bayesian framework. One of these proposes a calibration to the existing Bayesian hierarchical approach[12], significantly reducing type I error rate inflation compared to the uncalibrated version. The other, a Bayesian latent subgroup design[13], clusters cancer types in basket trials into responsive and non-responsive subgroups, resulting in higher power and controlled type I error rates compared to standard approaches. Additionally, a frequentist pipeline that analytically solves for early futility of baskets is available in the supplemental section of Chen et al[8]. Although this overview highlights several essential tools and pipelines, it is important to recognize that there may be additional software options for basket trial design that are not included. Hence, to further enhance the software applicability and user experience, and to present a holistic elucidation of

proposed trial design, we developed a web-based interactive R Shinny App[14,15], to fill the methodological gap in two-stage general basket trial design[16].

1.2 Target Audiences and Application Scope

Our tool is designed to support two-stage general randomized basket trials. Unlike conventional trials that focus on a single disease at a time, basket trials assess the efficacy of a single treatment across multiple homogenous diseases, significantly reducing the required sample size. The two-stage setup adds flexibility by allowing unresponsive disease baskets to be pruned at the interim stage, reallocating the remaining recruitment quota to responsive baskets. This early futility check ensures that resources are not wasted, improving the overall efficiency of the trial.

Initially developed for oncology studies, where it has been particularly effective in drug development for cancers and tumors, this design has become a widely adopted tool in various clinical research areas. Beyond oncology, the basket trial design is increasingly used in other therapeutic areas, including rare diseases and precision medicine.

The tool is particularly valuable for clinical trial researchers and biostatistical support teams due to its accessibility, intuitive usage, and convenience. Its user-friendly interface simplifies the complex process of designing and analyzing basket trials, making it easier for researchers to implement this advanced trial design without needing extensive statistical training.

2. Implementation

2.1 Statistical Methods

The null hypothesis for the design is that none of the baskets are active, while the alternative hypothesis is that only pre-specified baskets are active. As described by Chen et al[8] and Patel et al[10] , the trial's overall Type I error rate is calculated as the sum of rejection probabilities under the null hypothesis for different sets of baskets passing the interim analysis. The rejection probability is assessed based on whether all baskets pass through both the interim and final stages. Similarly, the power is estimated using the same sum of rejection probability function but under the alternative hypothesis, where the rejection region for pre-specified active baskets is adjusted according to the non-centrality parameter given by Chen et al[8].

Analytically solving the sample size formula can be algebraically tedious, so we propose estimating it using a root-finding approach based on the power function. Unequal basket size ratios can be directly estimated from disease prevalence ratios, meaning that only the trial-wide sample size needs to be determined. Given the monotonic relationship between power and sample size, a recursive binary search[17] algorithm is more efficient than computationally expensive iterative loops. This algorithm significantly improves the responsiveness of the app, making it faster and more efficient.

2.2 Variable Definitions and Input Parameters

Both calculations require the user to provide crucial trial characteristics (Figure 1) such as the standardized effect size (Δ) for each basket. Investigators can estimate this ratio based on subtype prevalence from published results. Additionally, the proportion of participants accrued at interim analysis (t) is needed. This is the cutoff point for interim analysis, expressed as a percentage of the total participants recruited, and should be consistent across all baskets. The interim significance level (αt) is also required, which is the designated α level for the interim analysis at proportion t. Any basket with a p-value exceeding this threshold will be dropped. The trial-wide significance level (α) must also be provided; this is the designated α level for the entire trial, representing the overall false positive rate.

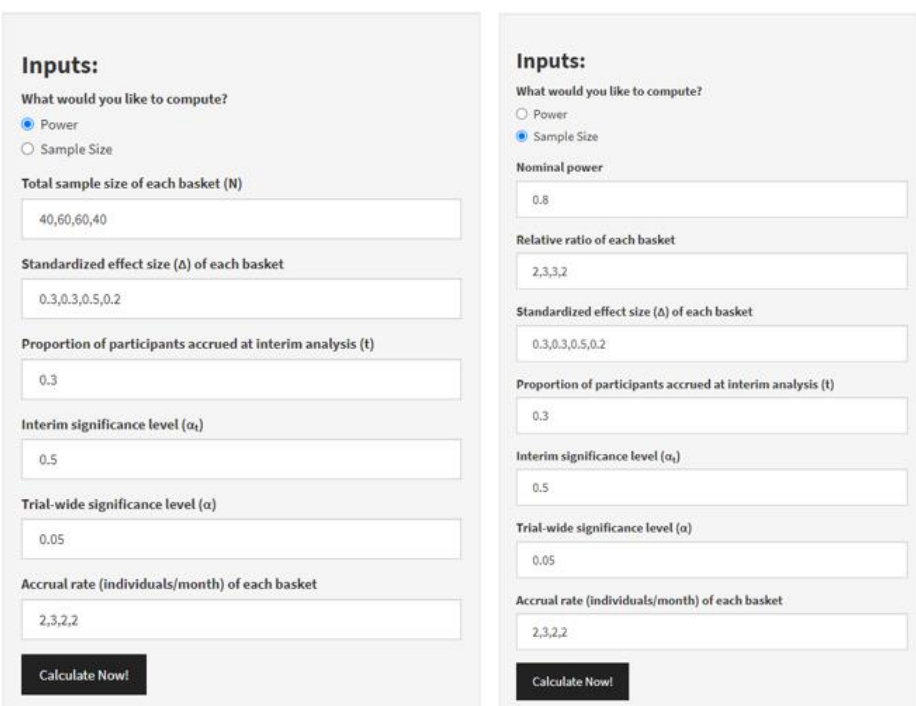

Figure 1: Input Parameters Required for Power or Sample Size Calculation

For sample size calculation, the sample size per basket is required. Conversely, for calculating the sample size per basket, the nominal power (designated 1−β level for the entire trial, or one minus the expected type II error rate) and the relative ratio of sample size in each basket must be provided. These are the primary parameters that directly influence the calculation, while the number of individuals recruited per month in each basket only affects the trial duration estimation.

2.3 System Architecture and User Interface Design

Having cloud computing power freely provided by the R Shiny[15] server, the local machine is only responsible for capturing inputs and rendering outputs on the user interface, requiring minimal hardware resources. The remote server handles heavy-duty statistical computing tasks. The main tab of the tool consists of two sections: the input panel and the output panel. Power and sample size calculations are unified through an initial question that asks users which type of calculation they intend the tool to perform. The corresponding trial characteristics are then required for the calculation of their choice. On the output panel, an interactive trial design plot is displayed above the text summary. If incorrect inputs are provided, error messages will appear, masking all outputs. Additionally, a 'Readme' tab is available, offering an example, detailed explanations of the interface, and the statistical methods used, helping users gain hands-on experience.

2.4 Operational Workflow and Error Detection

Users can input trial characteristics with just a few clicks and text entries. The app is initiated by clicking the 'Calculate Now!' button. All input parameters are sent to the backend, where they are checked for missing data or syntax errors before being passed into the function. The function then verifies that the sample sizes are whole numbers within baskets at each stage. If any errors are found, they are promptly flagged for correction. Once all error checks are passed, the power or sample size is computed. Both summary statistics and trial characteristics inputs are compiled into a plot to visually represent the proposed trial. A text summary of the plot is displayed below for additional clarity. Users could then download the plot and text summary for their intended use.\

2.5 Validation, Reproducibility, and User Testing

To enhance reproducibility, we conducted extensive validation analysis and user testing. App's outputs were systematically compared with results generated by the original R scripts used in the underlying statistical calculations from Patel et al[10]. This cross-verification ensured that the app replicates the established methods accurately. A variety of edge cases were tested, including extreme values of parameters and atypical trial configurations, to confirm the app handles all inputs robustly and responsively. This is crucial for identifying any scenarios where the app will fail and imposing corresponding error checkers as countermeasures. Additionally, we carefully examined the app’s error reporting mechanisms to ensure they provide clear and actionable feedback to users. This included verifying that the app correctly identifies and flags incorrect or missing

inputs before proceeding with calculations. Finally, we conducted user acceptance testing with potential end-users like clinicians and biostatisticians to validate the app's functionality in real-world scenarios. Participants were asked to replicate an example provided in the questionnaire, specifically for testing reproducibility. Feedback from these tests was incorporated to refine the app's interface and usability.

3. Results

Our two-stage general basket design visualization tool is hosted at (https://desmondzeyachen.shinyapps.io/AdaptiveTwoStageBasketTrialFeb14/).

To ensure broad accessibility, the app is designed to be fully responsive, allowing seamless operation on desktops, tablets, or even smartphones. The intuitive user interface (Figure 2) is built to cater to users with varying levels of statistical knowledge, ensuring that even those unfamiliar with advanced statistical methods can easily navigate the tool. The interactive trial design figure can be easily downloaded as a JPEG file, and the design summary provides a clear, technical explanation of each characteristic and the summary statistics of the proposed trial design.

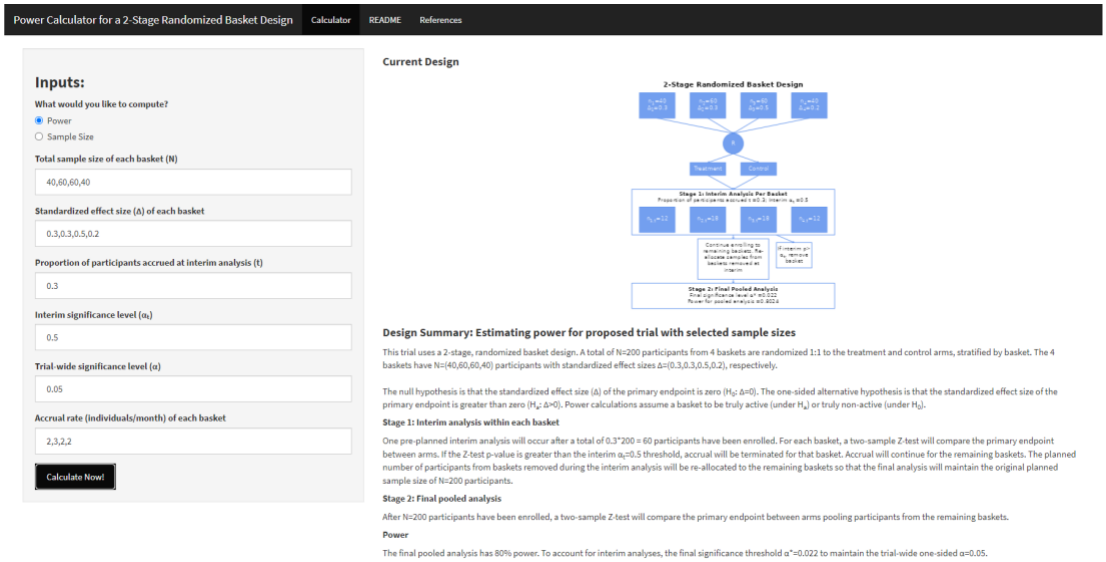

Figure 2: App Interface

3.1 User Interaction Feedback and Usability

Three users participated in the testing of the Shiny app, their feedback focused on ease of use, functionality, layout, and performance. All users found the app easy to navigate, especially with the help of default examples and detailed guidance provided in the Readme tab. Our original goal was to make the tool accessible to users with limited statistical training, aiming to simplify clinical trial design. However, one user noted that the use of highly specialized statistical language in the input variable names and explanations could create barriers for our target audience, potentially undermining our objective. We addressed this issue by incorporating additional explanations for each input from a trial design perspective, clarifying 'What does this mean in clinical trial terminology?' and 'How could you decide on a value for it?'

The app was also presented at an internal meeting, where feedback related to sample size estimation functionality was raised. In response, we enhanced the app by adding features that support sample size estimation directly within the tool, further improving its utility for users designing clinical trials.

The app's layout and functionality were deemed logical and well-organized, with users appreciating the clear instructions and intuitive design. Suggestions for enhanced result explanations and improved result visibility were carefully incorporated. In terms of performance, while the app generally worked well, users did experience occasional freezing or slower performance when working with larger sample sizes. These issues are common with the free version of the Shiny App server, and we plan to address them by either hosting the app on an internal server or upgrading to a premium account. All users expressed their willingness to use and recommend the app, citing its user-friendly interface and practical features. The user testing process, along with the feedback from the internal meeting, significantly improved the app’s usability and provided

valuable, unbiased feedback from potential users.

a. Comparative Analysis

Primary trial characteristics like the total sample size, interim sample size, and number of baskets are influenced by the interrelated inputs: the number of baskets, sample size per basket, and interim proportions. Selecting an interim alpha of 0.3 allows for effective

interim analysis without being overly conservative during partial recruitment. The trial-wide alpha is set at the conventional two-sided significance level of 0.025. We validate the Shiny App by comparing it against an R pipeline, using selected combinations of these inputs while controlling for effect sizes and alphas (Table 1).

Table 1: Comparative Analysis of the Tool vs R pipeline varying number of baskets and interim proportion

Number of Baskets	Size Per Basket	Interim Proportion	R code alpha*	Shiny App alpha*	R code Power	Shiny App Power
2	(60,60)	0.3	0.018037	0.018	0.765977	77%
2	(60,60)	0.5	0.014339	0.0143	0.789333	79%
3	(60,60,30)	0.3	0.0137	0.0137	0.860155	86%
3	(60,60,30)	0.5	0.01057	0.0106	0.872503	87%
4	(60,60,20,60)	0.3	0.011321	0.0113	0.942003	94%
4	(60,60,20,60)	0.5	0.008469	0.0085	0.948105	95%
5	(60,60,20,60,40)	0.3	0.009582	0.0096	0.973314	97%
5	(60,60,20,60,40)	0.5	0.006907	0.0069	0.976438	98%
6	(60,60,20,60,40,60)	0.3	0.008301	0.0083	0.991963	99%
6	(60,60,20,60,40,60)	0.5	0.005759	0.0058	0.993236	99%
7	(60,60,20,30,40,60,30)	0.3	0.007563	0.0076	0.992084	99%
7	(60,60,20,30,40,60,30)	0.5	0.005134	0.0051	0.993035	99%
8	(60,60,20,30,40,40,30,20)	0.3	0.006894	0.0069	0.992196	99%
8	(60,60,20,30,40,40,30,20)	0.5	0.004558	0.0046	0.99303	99%
9	(60,30,20,30,40,40,30,20,30)	0.3	0.006115	0.0061	0.992624	99%

9	(60,30,20,30,40,40,30,20,30)	0.5	0.00388	0.0039	0.993487	99%
10	(20,30,20,30,40,40,30,20,30,40)	0.3	0.005407	0.0054	0.99337	99%
10	(20,30,20,30,40,40,30,20,30,40)	0.5	0.003276	0.0033	0.994405	99%

Table 1: Assuming standardized effect size are all 0.5, all baskets are active interim alpha set at 0.3, and trial-wide alpha set at 0.025, alpha* is the recommended significance threshold of final analysis based on user provided trial wide alpha.

The alternative hypothesis is defined by user-specified effect sizes, with a value of 0 indicating that a basket will be pruned during interim analysis. To validate the tool under different alternative hypothesis scenarios, we varied the effect sizes per basket while keeping the sample size, number of baskets, interim proportion, and alphas constant (Table 2).

The results from the tool matched those generated by the R code after rounding, confirming that the tool correctly interprets the inputs and performs equivalently to the R code. This validation demonstrates the tool's robustness and reliability in replicating established statistical processes. The consistency between the tool and the R code suggests that users can confidently employ the tool for interim analysis and hypothesis testing in clinical trials.

Table 2: Comparative Analysis of the Tool vs R pipeline varying effect size per basket

Effect Size Per Basket	Truly Active Vector	R code alpha*	Shiny App alpha*	R code Power	Shiny App Power
(0.2,0.2,0)	(1,1,0)	0.01370037	0.0137	0.19221983	19%
(0.2,0.2,0.2)	(1,1,1)	0.01369794	0.0137	0.23586268	24%

(0.5,0.2,0)	(1,1,0)	0.01370189	0.0137	0.55850379	56%
(0.5,0.2,0.2)	(1,1,1)	0.01369941	0.0137	0.58211621	58%
(0.5,0.5,0)	(1,1,0)	0.01370145	0.0137	0.78413515	78%
(0.5,0.5,0.2)	(1,1,1)	0.01370213	0.0137	0.79281031	79%
(0.5,0.5,0.5)	(1,1,1)	0.0137017	0.0137	0.8601982	86%

Table 2: Varying effect size & whether the basket is truly active of (60,60,30). Interim proportion set at 0.3, interim alpha set at 0.3, and trial-wide alpha set at 0.025. Truly active vector, the corresponding baskets that pass interim analysis, deduced via effect size per basket that user provided.

4. Discussion

Our open-source calculator is compiled to run on browser while computing were done remotely at RShiny[15] server. Both power and sample size calculation functions were integrated into the same interface. Instead of manually constructing the pipeline, calibrating inputs, rephrasing outputs so its human readable, and plotting the design; our visualization tool automates all above tasks with just few simple steps. It provides users with an integrated, statistics enriched trial design figure and detailed textual exposition of the trial design alongside statistical results. These outputs can be conveniently assimilated as presented.

This user-friendly approach not only saves time but also reduces the potential for human error in trial design, ensuring a higher degree of accuracy and consistency. The tool's ability to produce ready-to-assimilate outputs enhances its practical value, allowing for

immediate integration into trial planning and reporting.

As clinical research moves towards more complex trial designs, basket trials have become increasingly popular due to their efficiency in exploring the effects of interventions across multiple disease types. These designs allow clinicians to investigate multiple patient populations simultaneously, improving resource allocation and potentially accelerating the discovery of effective treatments. Our tool fits seamlessly into this evolving landscape, providing an accessible means for researchers to engage with these complex designs without the need for extensive statistical training. By supporting the development and execution of basket trials, the tool also holds potential for educational purposes, enabling clinicians and researchers to independently explore and visualize these innovative trial structures. This flexibility, combined with the potential to expand into multi-stage and multi-arm trial designs, underscores the importance of developing tools that can keep pace with modern research needs.

Additionally, as the use of basket trials continues to expand, our tool is positioned to evolve alongside these developments, with potential future enhancements to support multi-stage designs, multi-arm trials, and other innovative trial methodologies. Gathering user feedback will be critical in guiding these developments to ensure the tool continues to meet the evolving needs of the research community.

Author Contributions

D.Z.C. and C.M. conceptualized the study. C.M. supervised the study. D.Z.C. and C.M. drafted the manuscript. D.Z.C. performed the analyses, developed the visualization software, and summarized the results. S.S.P. developed the base R code. S.S.P., A.X., J.C., and D.C. contributed to the investigation by conducting user testing of the application. All authors reviewed and edited the manuscript.

Declaration of interests

The author(s) declare no conflicts of interest.

Acknowledgements

We would like to thank Marcos Sanches and the rest of the biostatistics core at CAMH for their insightful questions during the internal meeting, which played a key role in shaping the development of this application. We would like to acknowledge the support of the University of Toronto Fellowship, provided by the University of Toronto for funding this research.

Software and Web Resources

The visualization tool was developed using R[14] version 4.3.2. The implementation involved several R packages, including shiny[15] (version 1.8.0), shinythemes[18] (version 1.2.0), ggplot2[19] (version 3.4.4), ggtext[20] (version 0.1.2), and ggforce[21] (version 0.4.1).

References:

1. Redman MW, Allegra CJ. The Master Protocol Concept. *Semin Oncol*. Oct 2015;42(5):724-30. doi:10.1053/j.seminoncol.2015.07.009
2. Cunanan KM, Iasonos A, Shen R, Begg CB, Gonen M. An efficient basket trial design. *Stat Med*. May 10 2017;36(10):1568-1579. doi:10.1002/sim.7227
3. McArthur GA, Demetri GD, van Oosterom A, et al. Molecular and clinical analysis of locally advanced dermatofibrosarcoma protuberans treated with imatinib: Imatinib Target Exploration Consortium Study B2225. *J Clin Oncol*. Feb 1 2005;23(4):866-73. doi:10.1200/JCO.2005.07.088
4. Heinrich MC, Joensuu H, Demetri GD, et al. Phase II, open-label study evaluating the activity of imatinib in treating life-threatening malignancies known to be associated with imatinib-sensitive tyrosine kinases. *Clin Cancer Res*. May 1 2008;14(9):2717-25. doi:10.1158/1078-0432.CCR-07-4575
5. Park JJH, Siden E, Zoratti MJ, et al. Systematic review of basket trials, umbrella trials, and platform trials: a landscape analysis of master protocols. *Trials*. Sep 18 2019;20(1):572. doi:10.1186/s13063-019-3664-1
6. Cummings J, Montes A, Kamboj S, Cacho JF. The role of basket trials in drug development for neurodegenerative disorders. *Alzheimers Res Ther*. May 25 2022;14(1):73. doi:10.1186/s13195-022-01015-6
7. Fountzilas E, Tsimberidou AM, Vo HH, Kurzrock R. Clinical trial design in the era of precision medicine. *Genome Med*. Aug 31 2022;14(1):101. doi:10.1186/s13073-022-01102-1
8. Chen C ea. Statistical Design and Considerations of a Phase 3 Basket Trial for Simultaneous Investigation of Multiple Tumor Types in One Study. *Statistics in Biopharmaceutical Research*. 2016;8(3):248-257. doi:10.1080/19466315.2016.1193044
9. Wang M, Spiegelman D, Kuchiba A, et al. Statistical methods for studying disease subtype heterogeneity. *Stat Med*. Feb 28 2016;35(5):782-800. doi:10.1002/sim.6793
10. Patel SS. Randomized Basket Trial with an Interim Analysis (RaBIt) and Applications in Mental Health. 2024
11. Berry Consultants. Fixed and adaptive clinical trial simulator. Accessed 08-13, 2024. https://www.berryconsultants.com/wp-content/uploads/2014/09/FACTS-Overview.pdf
12. Chu Y, Yuan Y. A Bayesian basket trial design using a calibrated Bayesian hierarchical model. *Clin Trials*. Apr 2018;15(2):149-158. doi:10.1177/1740774518755122
13. Yiyi Chu YY. BLAST: Bayesian Latent Subgroup Design for Basket Trials Accounting for Patient Heterogeneity. *Journal of the Royal Statistical Society Series C: Applied Statistics*. 2018;67(3):723–740.
14. RCoreTeam. R: A language and environment for statistical computing. https://www.R-project.org/
15. *Shiny App: Easy web applications in R*. Version 1.8.0. 2013. http://www.rstudio.com/shiny/
16. Kasim A, Bean N, Hendriksen SJ, Chen TT, Zhou H, Psioda MA. Basket trials in oncology: a systematic review of practices and methods, comparative analysis of innovative methods, and an appraisal of a missed opportunity. *Front Oncol*. 2023;13:1266286. doi:10.3389/fonc.2023.1266286
17. D.H.Lehmer. *Proc Symp Appl Math*. 1960;10:180-181.
18. Chang W. shinythemes: Themes for Shiny. https://CRAN.R-project.org/package=shinythemes
19. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2016. https://ggplot2.tidyverse.org
20. Wiernik COWaBM. ggtext: Improved Text Rendering Support for 'ggplot2'. https://CRAN.R-project.org/package=ggtext
21. Pedersen TL. ggforce: Accelerating 'ggplot2'. https://CRAN.R-project.org/package=ggforce

314