EasyChair Preprint

№ 11981

RCT Simulations for Enhanced Tuberculosis Clinical Trial Design

Aditya Mavle, Aarushi Wagh, Atharva Nagmoti and Nikhil Gupte

February 7, 2024

RCT Simulations for Enhanced Tuberculosis Clinical Trial Design

Aditya Mavle
Dept. of Electronics and Telecommunication
Pune Institute of Computer Technology
adityamavle@gmail.com

Aarushi Wagh
Dept. of Electronics and Telecommunication
Pune Institute of Computer Technology
aarushiwagh16@gmail.com

Atharva Nagmoti
Dept. of Information Technology
Pune Institute of Computer Technology
atharvanagmoti@gmail.com

Dr. Nikhil Gupte
JHU-BJGMC Clinical Trials Unit (CTU)
Johns Hopkins Medicine
ngupte1@jhmi.edu

***Abstract*—Clinical trials, pivotal for evaluating medical interventions, are undergoing a shift with Clinical Trial Simulation (CTS). CTS optimizes trial design, providing a cost-effective, evidence-based approach to drug development and intervention testing. The primary objective of this study is to use the power of simulations to evaluate the effectiveness of behavioral interventions, specifically addressing alcohol, depression, and tobacco as risk factors, in enhancing tuberculosis treatment outcomes. Implementing a robust power analysis framework and introducing the concept of "resimulations", our study enhances statistical rigor and reliability across various intervention scenarios. This offers a versatile approach applicable to diverse research areas, providing a comprehensive framework to design and simulate Randomized Controlled Trials (RCTs), enabling reliable robust clinical studies.**

Index Terms*—*Biostatistics*, *Simulations*, *Clinical Trial Design*, *Randomized Controlled Trials*, *Power Analysis*, *Sample Size Optimization*, *Comorbidity*, *Tuberculosis*, *Intervention Testing

I. Introduction

Tuberculosis (TB), a global threat, has encountered setbacks due to COVID-19, leading to an 18% decline in TB diagnoses from 2019 to 2020 [1]. Despite this, the World Health Organization (WHO) estimates that over one million children develop TB annually, constituting 12% of the global TB burden [2]. TB is intricately linked with co-morbidities like HIV, diabetes, alcohol abuse, smoking, and depression, posing challenges in diagnosis and treatment [3]. Concurrent pathologies, along with alcohol dependence and smoking, contribute to widespread drug resistance and low adherence to TB therapy [4] [5]. Conducting clinical trials to observe the effect of interventions for such comorbidities can help us improve treatment outcomes [6]. Noteworthy clinical trials (Study 31/A5349, TRUNCATE-TB, SHINE, STREAM, NiX-TB, ZeNix, TB-PRACTECAL) have shown promise in shortening TB treatment duration [7].

However, clinical trials for testing novel treatments and drugs face several challenges, notably high logistical and financial costs associated with their execution [8] [9]. These costs can be a significant barrier, frequently restricting the breadth and depth of experimental research [7]. Ethical considerations in clinical trials are paramount, particularly in balancing patient welfare with scientific inquiry, ensuring informed consent, and addressing the morality of placebo use in the control group [4]. Based on these challenges, approaches to ensure treatment efficacy in clinical trials are of paramount importance. A critical aspect in this context is sample size estimation, which is crucial for achieving statistical significance and ensuring that the trial results are reliable [10]. An undersized study may subject participants to needless and possibly harmful treatments without yielding valuable knowledge, while an oversized study risks involving too many subjects in such treatments [11]. Frequently, sample size calculations in clinical trials are poorly conducted and reported [12], leading to reproducibility and accuracy issues. This often results in under-powered studies that fail to conclusively demonstrate treatment effectiveness [13]

Another challenge in estimating treatment effectiveness in clinical trials is the presence of multiple risk factors associated with a disease, such as in tuberculosis, where diabetes, smoking, alcohol use, and other drug use significantly influence trial outcomes and interpretations, complicating the analysis and management of the disease [6]. Furthermore, comorbidities often lead to worse health outcomes and an increased load on healthcare infrastructure. Interventions for patients with comorbidities in primary care require complex, tailored strategies. A review by Smith et al. (2012) [14] indicates limited overall effectiveness but suggests interventions targeting specific risk factors or functional difficulties may show improved outcomes. These considerations highlight the need for meticulous preclinical trial design studies that incorporate these factors, thereby enhancing the relevance and applicability of clinical trial findings in real-world settings.

Clinical trial simulations for Randomized Controlled

Trials (RCTs) present a novel method for optimizing trial designs and resource allocation. By simulating RCTs before execution, researchers can effectively determine optimal population sizes and distribution, reducing costs and improving the efficiency of trials. This approach allows for refined study designs and ensures the reliability of trial results.

India, bearing the highest global TB burden and ranking second in tobacco consumption, along with significant alcohol use, prompts a critical examination of the intricate relationship between these factors and TB treatment outcomes [6]. Recognizing the lack of clarity in this domain, our study aims to fill this gap by conducting a simulation study. We analyze the impact of comorbidity features, exploring scenarios where TB patients solely consume alcohol, solely take tobacco, have poor mental health, or certain combinations of these conditions. Implementing interventions for smoking cessation, alcohol reduction, and mental health symptom management, we strive to optimize sample sizes and design a multi-factorial RCT before implementation. Assumptions include an overall TB treatment failure rate of 12%-15%, with interventions expected to reduce this rate by 50 %. The goal is to design a comprehensive trial, evaluating main intervention effects through participant comparisons with and without interventions. Beyond the scope of individual participant outcomes, this endeavor aims to create a dynamic platform for healthcare practitioners to experiment with different risk factor interventions, explore the diverse impacts of comorbidities across various use cases, and effectively plan clinical studies.

II. Relevant Work

Sample size optimization is crucial for ensuring reliable evidence in clinical trials, guiding treatment efficacy assessments, and clinical decision-making [10]. Typically, trials involve two groups: a case group receiving the intervention and a placebo group receiving an inactive treatment. Traditional methods rely on power calculations, often using mathematical formulas or Monte Carlo simulations [15] [16]. However, complex trial designs pose challenges addressed by various methods like SimSam and MLPowSim packages [17] [18]. Multi-objective optimization algorithms like NSGA-II offer theoretical solutions but may be impractical for demanding problems [19]. Surrogate models, particularly Gaussian process regression with efficient global optimization algorithms, address these challenges effectively [20]. Randomization, a common method in trials, controls bias, with simple randomization ensuring complete randomness but potentially leading to uneven group sizes, especially in smaller studies [21]. Cluster randomized trials, randomizing individuals in groups, offer advantages in healthcare evaluations and interventions, making higher-level randomization more feasible and practical [21].

Comorbidity, defined as the presence of additional conditions alongside a primary one [22], can impact treatment efficacy by introducing competing risks and interactions, including drug-drug, drug-disease, and disease-disease interactions, thereby affecting the risk-benefit balance [23] [24] [25]. In a cross-sectional analysis, Hanlon et al. (2019) [26] investigated the prevalence of comorbidity and multimorbidity among 122,969 participants in 116 industry-funded trials of novel drug therapies across 22 index conditions. They compared these findings with comorbidity data from 2.3 million community-dwelling patients in the UK, providing insights into the representation of individuals with comorbidities in clinical trial settings [26]. Unger et al. (2019) [27] investigated the association between comorbidities and participation in cancer clinical trials among patients with breast, colorectal, lung, or prostate cancer [27]. Using data from a national survey, Charlson et al. (1987) [28] examined the decision-making process regarding trial participation in relation to 18 comorbid conditions. Logistic regression analysis revealed insights into how the presence of comorbidities influenced discussions about trials, trial offers, and actual participation, shedding light on factors impacting patient engagement in clinical research [27].

Previous research has highlighted the critical importance of sample size calculation in the context of clinical trial design, emphasizing its integral role in ensuring the statistical power needed to detect significance [10]. Underpowered studies risk inconclusive results, jeopardizing the overall success of the research endeavor. The complexity of sample size determination is compounded by the competing objectives of maximizing statistical power and minimizing budgetary constraints. The determination of optimal sample size is a pivotal step in crafting a research protocol, striking a delicate balance between maximizing statistical power and minimizing costs associated with involving too many participants. It remains a universal challenge. Although work has been done statistically to answer the contradictions of sample size determination, a comprehensive solution that incorporates this facet in the designing of RCTs has not yet been extensively researched. [29] [15] [30]

Designing RCTs before their actual implementation serves as a strategic tool to test novel trial designs, understand trial requirements, and optimize resource allocation. PyTrial, a machine learning framework, is tailored for clinical trials, featuring over 30 algorithms across six tasks [31]. Its trial patient simulation task uses generative models to create synthetic clinical trial data, closely mimicking real patient records. Another module predicts patient outcomes, aiding in evaluating treatments while minimizing risks. While PyTrial offers a robust solution for clinical trial design and execution, it lacks capabilities for testing multiple interventions on a single synthetic dataset.

Monte Carlo Simulations (MCS) are used in Randomized Controlled Trials (RCTs) for their capability to employ random sampling and statistical models for predicting and analyzing outcomes [32] [33] [16] [34]. A key study by

Goldenholz's et al. (2017) [16] on epilepsy RCTs using MCS illustrates its effectiveness in optimizing trial design, highlighting how trial parameters influence costs and placebo responses. However, a limitation of MCS is its reliance on accurate data models to reflect real-world complexities, which can impact the generalizability of simulation results.

Recent collaborative efforts have been undertaken to develop a robust R function for simulating a two-group parallel-arm randomized controlled trial (RCT) with interim analyses [35]. The primary focus of the simulation was on binary outcomes, specifically exploring the probability of death for each group. The flexibility of the simulation design allowed for tuning various parameters, including the total number of participants, the frequency and schedule of interim analyses, the chosen group-sequential design, as well as conventional trial analysis parameters such as the significance level (alpha), the type of test (1-sided vs. 2-sided), and others. The function computes essential statistics like odds ratios, confidence limits, p-values, and the number of successes for each specified interim analysis. While this method offers flexibility in tuning parameters, its applicability is confined to specific trial scenarios and outcomes. The efficiency and scalability of the R code may also be subject to improvement, considering alternative coding practices.

III. Methodology

A. Sample Study Description

We consider a Tuberculosis simulation study, with the primary objective of evaluating the effectiveness of behavioral interventions targeting risk factors of alcohol use, tobacco use, and mental health issues in improving tuberculosis treatment outcomes. This study population is conceptualized as a cohort of individuals undergoing TB treatment, with treatment outcomes categorized as either "0" for adverse outcomes (such as treatment failure, death, or recurrence) or "1" for positive outcomes signifying recovery.

Addressing the complexity of multiple risk factors and their various overlaps in the study population, we consider single as well as multi-morbidity risk factor interventions in our study. Our simulations explore the interplay of multiple intervention groups dividing the population into eight intervention groups. To determine the optimal sample sizes for each treatment versus control group, we conduct statistical power analysis, aiming for a statistical power above 80 %. We implement recursive resampling to achieve these optimized sample sizes, ensuring the power analysis criteria are met. Our study hypothesizes that comorbidity interventions improve treatment outcomes, a theory we aim to validate.

Through this simulation study, we aim to validate our Randomized Controlled Trial (RCT) design framework and demonstrate its applicability across various clinical scenarios.

B. Data Collection

In our study, we leveraged data from the Geneva WHO Global TB Reports spanning 2020 to 2022 [2] [1] [36] to establish a comprehensive dataset of TB patients worldwide, accessed through the WHO TB Report app. This rich source provided region-specific information on TB patients, offering insights into population distribution and comorbidity prevalence, such as the gender-based disparity in alcohol consumption. This data formed the foundation for our study, allowing us to simulate realistic scenarios mirroring real-world conditions.

Furthermore, we employed correlation analysis and K-means clustering techniques to identify key population variables essential for our simulation model. K-means clustering enabled us to categorize data points into distinct clusters, facilitating a deeper understanding of population characteristics and comorbidity patterns. Correlation analysis provided insights into the relationships between intervention effectiveness and statistical distribution, guiding our selection of variables for simulation modeling. This data-driven approach ensured that we created a representative base population, necessary for precisely assessing intervention efficacy and guiding the design of our simulation framework.

C. Simulation Modeling

For our simulation framework, we consider several parameters describing the population's characteristics, such as total sample size, demographic information (age, gender, BMI, education), percentage of people with risk factor conditions (alcoholism, depression, tobacco use), percentage overlaps between conditions, intervention status, and treatment outcomes. The framework simulates a dataset of clinical trial data based on input parameters and percentage distributions.

TABLE I: Description of Variables in Population Data

Variable	Type	Distribution
Age	Continuous	Range: 18 to 60; Mean of 35; Std Deviation of 15
Gender	Categorical	Male-Female 50-50 split
BMI	Categorical	Three Buckets (10%,50%, 30%)
Alcoholism, Smoking, Depression	Binary Predictor variables	Base Assumption (20%,25%, 20%)
Intervention	Categorical Variable	17 Unique Values, intervention status
Treatment Outcomes	Binary Response variable	0-Adverse Outcomes 1-Good Outcomes

We modeled the distribution of demographic variables based on our analysis of the WHO Global TB Report Data as mentioned in the previous section. The 3 risk factors of

alcohol, depression, and tobacco usage were modeled as binary predictor variables, reflecting risk factors in individuals based on specified population percentages and overlaps(initial sample 20%, 25%, 20%). Additionally, the percentage overlap parameters model the overlap of multiple risk factors in the population and calculate the inclusion-exclusion principle. The population's risk factor status had seven unique categories, accounting for individual and overlapping cases of the three risk factors. For a given risk factor status, interventions were administered by a 50-50 split, indicating an equal chance for an individual with a particular condition to receive one of the available interventions. Intervention status is a categorical variable with 17 unique values, describing the types of interventions for risk factors that a patient receives.

TABLE II: Comorbidity Groups and Intervention Groups

Comorbidity	**A**	**D**	**T**	**Intervention Type**	**Treatment vs Control**
0	0	0	0	Unaffected	UNAFFECTED
1	1	0	0	Alcoholism	A/NA
2	0	1	0	Depression	D/ND
3	0	0	1	Tobacco Use	T/NT
4	1	1	0	Alcoholism+Depression	AD/NAD
5	1	0	1	Alcoholism+Tobacco Use	AT/NAT
6	0	1	1	Depression+Tobacco Use	DT/NDT
7	1	1	1	Alcoholism+Depression+Tobacco Use	ADT/NADT

Treatment outcomes in our study are binary and determined by the interaction of risk factors and interventions. They are randomly assigned based on the probability of success for each risk factor and intervention combination, with these probabilities influencing the likelihood of a positive treatment response. For instance, a person with alcoholism receiving an intervention has a high probability of a good treatment outcome. However, if they also have depression without intervention, the probability decreases. A person with alcoholism and no intervention has a lower probability of a good outcome. The interplay of these variables creates a nuanced understanding of treatment outcomes based on the presence of risk factors and corresponding interventions.

D. Power Analysis

Following population simulation, we conduct a t-test power analysis. This analysis is crucial for determining the likelihood of observing the effects of the interventions on the improvement of treatment outcomes in a real-world population. By testing the null hypothesis, we can establish whether any observed differences are statistically significant [37]. This analysis aids in ensuring our study is adequately powered to detect the interventions' effects.

In our study, we identified seven distinct risk factor statuses(including individual risk factors, their dual combinations, and the collective combination of all three). For each status, individuals were split into two intervention statuses: those receiving the intervention (treatment) and those not (control). This created a total of 14 intervention statuses. Our power analysis compared each risk factor condition's treatment group with its corresponding control group, across all seven conditions, to evaluate our study's ability to detect the intervention's true effect on each risk factor.

E. Resimulations: Optimized resampling

In maintaining the integrity of our simulation outcomes, every power analysis case must achieve a statistical power above the crucial threshold of 0.80. This standard is key to rejecting the null hypothesis, correctly identifying true effects, and substantiating our results [37]. To address this, we implemented an optimized resimulation approach. Our recursive method precisely calculates statistical power for different treatment-control groups, targeting a minimum power of 0.75 for credible results. Should any group not meet this benchmark, we iteratively add more samples, thus engaging in a recursive cycle of sample resampling and power reevaluation. We also introduced a factor of randomization to ensure that the dataset encompassed nearly ideal numbers for each subgroup. While not essential for all parameters, randomization is crucial for certain factors to ensure simulation precision [21]. Our innovative backtracking technique dynamically modifies sample sizes until each treatment-control pair reaches the preferred power level. This iterative sample size refinement provides statistical consistency across various intervention scenarios.

IV. Results

Our simulation function was employed to test and validate a range of population scenarios, each varying in size and the distribution of key feature characteristics. We performed a power analysis of the simulated data's respective treatment/control groups and evaluated their statistical power.

The resimulation algorithm effectively managed to conserve the percentage distribution of the intervention groups in the initial population even after increasing the total number of samples. It also has a successful impact on increasing the power per treatment/control group of the population above the threshold value, successfully achieving its outcome of enhancing the statistical power and reliability of the study's findings [37]. This iterative recursion approach ensures that each treatment-control pair attains the desired power level.

The above table presents different initial population use cases, showing their initial average statistical power and the final sample size, along with the final average power achieved, exceeding the critical threshold of 0.80. The relative difference between the final sample size after resimulation can be attributed to randomization [21] due to the resampling that we

TABLE III: Initial vs Final Power

Initial Sample Size	Initial Average Power	Final Sample Size	Final Average Power
10000	0.60	16460	0.89
12000	0.63	16240	0.87
15000	0.79	16120	0.86
20000	0.92	20180	0.92

have implemented. The variation in final sample sizes after resimulation is due to the randomization inherent in our re-sampling method. When the initial power already surpasses the 0.80 threshold, the function minimally iterates, demonstrating its reliability across various initial sample scenarios.

TABLE IV: Initial vs Final Size after Resimulation

Initial Sample Size	Risk Factor Distribution	Risk Factor Overlap	Final Sample Size
10000	8%-8%-8%	4%-4%-4%-3%	16460
12000	8%-8%-8%	4%-4%-4%-3%	16240
10000	8%-10%-8%	4%-4%-5%-3%	17760
10000	10%-10%-10%	5%-5%-5%-3%	13560
20000	8%-8%-8%	4%-4%-4%-3%	20180

The simulation results indicate that larger final sample sizes are necessary when risk factors are initially distributed at low and equal percentages. Conversely, smaller final sample sizes are required when these factors are high and equal. The largest final sample sizes are observed when risk factors are distributed at high but unequal percentages.

These results underscore a critical aspect of clinical trial design: the initial distribution of risk factors influences the necessary scale of a study. A balanced distribution of high-risk factors can optimize sample size [38], while a skewed distribution necessitates a larger cohort to ensure statistical significance. This insight is crucial for researchers planning trials, as it affects resource allocation, trial duration, and the overall feasibility of detecting meaningful treatment effects.

V. Discussion

Our study introduces a comprehensive approach to the simulation, optimization, and design of randomized controlled trials (RCTs) for pulmonary tuberculosis treatment. A function to generate synthetic clinical trial datasets is developed. These datasets simulate a broad range of population characteristics, considering factors such as condition prevalence (alcoholism, depression, tobacco use), demographic variables (age, gender, BMI, education), treatment interventions, and outcomes.

The power analysis framework we established plays a pivotal role in determining the necessary sample size for achieving a statistical power of at least 0.80. We introduce the concept of "resimulations," addressing the critical need for statistical power in clinical research. Resimulations offer an iterative approach to achieving the desired statistical power for each treatment-control pair, ensuring a reliable study design.

Our solution extends beyond the framework and is implemented as a Web application. This web application automates the RCT population design process, enabling researchers to input study details and receive a re-simulated dataset with optimized statistical power and sample sizes, eliminating manual calculations.

In the future, the method can be scaled up, accommodating an increased number of interventions and supporting various trial designs. It can also facilitate the design of clinical trials for testing new drug therapies by simulating diverse patient populations with varying comorbidities and demographic characteristics. This enables researchers to assess the efficacy and safety of experimental drugs more comprehensively before advancing to costly and time-consuming clinical trials. In the world of public health interventions, such as vaccination campaigns or disease prevention programs, our methodology allows for the simulation of intervention impacts on different population groups. By integrating demographic data and risk factors, organizations can make informed decisions about resource allocation and implementation strategies, maximizing the effectiveness of public health initiatives.

References

[1] J. Chakaya, E. Petersen, R. Nantanda, B. N. Mungai, G. B. Migliori, F. Amanullah, P. Lungu, F. Ntoumi, N. Kumarasamy, M. Maeurer *et al.*, "The who global tuberculosis 2021 report–not so good news and turning the tide back to end tb," *International Journal of Infectious Diseases*, vol. 124, pp. S26–S29, 2022.

[2] J. Chakaya, M. Khan, F. Ntoumi, E. Aklillu, R. Fatima, P. Mwaba, N. Kapata, S. Mfinanga, S. E. Hasnain, P. D. Katoto *et al.*, "Global tuberculosis report 2020–reflections on the global tb burden, treatment and prevention efforts," *International journal of infectious diseases*, vol. 113, pp. S7–S12, 2021.

[3] G. Cáceres, R. Calderon, and C. Ugarte-Gil, "Tuberculosis and comorbidities: treatment challenges in patients with comorbid diabetes mellitus and depression," *Therapeutic Advances in Infectious Disease*, vol. 9, p. 20499361221095831, 2022.

[4] V. Antonio-Arques, J. Franch-Nadal, A. Moreno-Martinez, J. Real, À. Orcau, D. Mauricio, M. Mata-Cases, J. Julve, E. Navas Mendez, R. Puig Treserra *et al.*, "Subjects with diabetes mellitus are at increased risk for developing tuberculosis: a cohort study in an inner-city district of barcelona (spain)," *Frontiers in Public Health*, vol. 10, p. 789952, 2022.

[5] A. Starshinova, I. Dovgalyk, E. Belyaeva, A. Glushkova, N. Osipov, and D. Kudlay, "Efficacy of tuberculosis treatment in patients with drug-resistant tuberculosis with the use of bedaquiline: the experience of the russian federation," *Antibiotics*, vol. 11, no. 11, p. 1622, 2022.

[6] B. E. Thomas, K. Thiruvengadam, D. Kadam, S. Ovung, S. Sivakumar, S. V. Bala Yogendra Shivakumar, M. Paradkar, N. Gupte, N. Suryavanshi, C. Dolla *et al.*, "Smoking, alcohol use disorder and tuberculosis treatment outcomes: A dual co-morbidity burden that cannot be ignored," *PLoS One*, vol. 14, no. 7, p. e0220507, 2019.

[7] I. Motta, M. Boeree, D. Chesov, K. Dheda, G. Günther, C. R. Horsburgh Jr, Y. Kherabi, C. Lange, C. Lienhardt, H. M. McIlleron *et al.*, "Recent advances in the treatment of tuberculosis," *Clinical Microbiology and Infection*, 2023.

[8] A. A. Seyhan, "Lost in translation: the valley of death across preclinical and clinical divide–identification of problems and overcoming obstacles," *Translational Medicine Communications*, vol. 4, no. 1, pp. 1–19, 2019.

[9] H.-H. Wong, A. Jessup, A. Sertkaya, A. Birkenbach, A. Berlind, and J. Eyraud, "Examination of clinical trial costs and barriers for drug development final," *Office of the Assistant Secretary for Planning and Evaluation, US Department of Health & Human Services*, pp. 1–92, 2014.

[10] R. J. Tamanna, M. I. Alam, A. Hossain, and M. H. R. Khan, "On sample size calculation in testing treatment efficacy in clinical trials," *Biometrical Letters*, vol. 58, no. 2, pp. 133–147, 2021.

[11] D. G. Altman, "Statistics and ethics in medical research: Iii how large a sample?" *British medical journal*, vol. 281, no. 6251, p. 1336, 1980.

[12] B. Copsey, J. Y. Thompson, K. Vadher, U. Ali, S. J. Dutton, R. Fitzpatrick, S. E. Lamb, and J. A. Cook, "Sample size calculations are poorly conducted and reported in many randomized trials of hip and knee osteoarthritis: results of a systematic review," *Journal of Clinical Epidemiology*, vol. 104, pp. 52–61, 2018.

[13] D. B. Fogel, "Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review," *Contemporary clinical trials communications*, vol. 11, pp. 156–164, 2018.

[14] S. M. Smith, H. Soubhi, M. Fortin, C. Hudon, and T. O'Dowd, "Managing patients with multimorbidity: systematic review of interventions in primary care and community settings," *Bmj*, vol. 345, 2012.

[15] D. T. Wilson, R. Hooper, J. Brown, A. J. Farrin, and R. E. Walwyn, "Efficient and flexible simulation-based sample size determination for clinical trials with multiple design parameters," *Statistical methods in medical research*, vol. 30, no. 3, pp. 799–815, 2021.

[16] D. M. Goldenholz, J. Tharayil, R. Moss, E. Myers, and W. H. Theodore, "Monte carlo simulations of randomized clinical trials in epilepsy," *Annals of Clinical and Translational Neurology*, vol. 4, no. 8, pp. 544–552, 2017.

[17] R. Hooper, "Versatile sample-size calculation using simulation," *The Stata Journal*, vol. 13, no. 1, pp. 21–38, 2013.

[18] W. J. Browne, M. G. Lahi, and R. M. Parker, "A guide to sample size calculations for random effect models via simulation and the mlpowsim software package," *Bristol, United Kingdom: University of Bristol*, 2009.

[19] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.

[20] M. J. Sasena, P. Papalambros, and P. Goovaerts, "Exploration of metamodeling sampling criteria for constrained global optimization," *Engineering optimization*, vol. 34, no. 3, pp. 263–278, 2002.

[21] K. Suresh, "An overview of randomization techniques: an unbiased assessment of outcome in clinical research," *Journal of human reproductive sciences*, vol. 4, no. 1, p. 8, 2011.

[22] J. M. Valderas, B. Starfield, B. Sibbald, C. Salisbury, and M. Roland, "Defining comorbidity: implications for understanding health and health services," *The Annals of Family Medicine*, vol. 7, no. 4, pp. 357–363, 2009.

[23] C. M. Boyd and D. M. Kent, "Evidence-based medicine and the hard problem of multimorbidity," *Journal of general internal medicine*, vol. 29, pp. 552–553, 2014.

[24] E. Wallace, C. Salisbury, B. Guthrie, C. Lewis, T. Fahey, and S. M. Smith, "Managing patients with multimorbidity in primary care," *Bmj*, vol. 350, 2015.

[25] M. Böhm, J. Pogue, I. Kindermann, J. Pöss, T. Koon, and S. Yusuf, "Effect of comorbidities on outcomes and angiotensin converting enzyme inhibitor effects in patients with predominantly left ventricular dysfunction and heart failure," *European Journal of Heart Failure*, vol. 16, no. 3, pp. 325–333, 2014.

[26] P. Hanlon, L. Hannigan, J. Rodriguez-Perez, C. Fischbacher, N. J. Welton, S. Dias, F. S. Mair, B. Guthrie, S. Wild, and D. A. McAllister, "Representation of people with comorbidity and multimorbidity in clinical trials of novel drug therapies: an individual-level participant data analysis," *BMC medicine*, vol. 17, pp. 1–12, 2019.

[27] J. M. Unger, D. L. Hershman, M. E. Fleury, and R. Vaidya, "Association of patient comorbid conditions with cancer clinical trial participation," *JAMA oncology*, vol. 5, no. 3, pp. 326–333, 2019.

[28] M. E. Charlson, P. Pompei, K. L. Ales, and C. R. MacKenzie, "A new method of classifying prognostic comorbidity in longitudinal studies: development and validation," *Journal of chronic diseases*, vol. 40, no. 5, pp. 373–383, 1987.

[29] I. Khan, S. Sarker, and A. Hackshaw, "Smaller sample sizes for phase ii trials based on exact tests with actual error rates by trading-off their nominal levels of significance and power," *British journal of cancer*, vol. 107, no. 11, pp. 1801–1809, 2012.

[30] J. Wittes, "Sample size calculations for randomized controlled trials," *Epidemiologic reviews*, vol. 24, no. 1, pp. 39–53, 2002.

[31] Z. Wang, B. Theodorou, T. Fu, C. Xiao, and J. Sun, "Pytrial: A comprehensive platform for artificial intelligence for drug development," *arXiv preprint arXiv:2306.04018*, 2023.

[32] D.-G. Chen and J. D. Chen, *Monte-Carlo simulation-based statistical modeling*. Springer, 2017.

[33] I. Abbas, J. Rovira, and J. Casanovas, "Clinical trial optimization: Monte carlo simulation markov model for planning clinical trials recruitment," *Contemporary clinical trials*, vol. 28, no. 3, pp. 220–231, 2007.

[34] S. Ankolekar, C. Mehta, R. Mukherjee, S. Hsiao, J. Smith, and T. Haddad, "Monte carlo simulation for trial design tool," *Principles and Practice of Clinical Trials*, pp. 1–23, 2020.

[35] I. Needleman, H. Worthington, D. Moher, K. Schulz, and D. G. Altman, "Improving the completeness and transparency of reports of randomized trials in oral health: the consort statement," *American journal of dentistry*, vol. 21, no. 1, p. 7, 2008.

[36] S. Bagcchi, "Who's global tuberculosis report 2022," *The Lancet Microbe*, vol. 4, no. 1, p. e20, 2023.

[37] K. Suresh and S. Chandrashekara, "Sample size estimation and power analysis for clinical research studies," *Journal of human reproductive sciences*, vol. 5, no. 1, p. 7, 2012.

[38] A. L. Whitehead, S. A. Julious, C. L. Cooper, and M. J. Campbell, "Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable," *Statistical methods in medical research*, vol. 25, no. 3, pp. 1057–1073, 2016.