Design and Assessment of Representative Hybrid Clinical Trials using Health Recommender System

Nafis Neehal[1,*], Vibha Anand[2] and Kristin P. Bennett[3]

[1]Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY, 12180 USA
[2]Biomedical AI-Healthcare and Life Sciences, IBM T.J. Watson Research Center, Cambridge, MA, 02141 USA
[3]The Institute of Data Exploration and Applications and the Mathematical Sciences and Computer Science Departments, Rensselaer Polytechnic Institute, Troy, NY, 12180 USA

Abstract

Incorporating real-world data (RWD) into clinical trials can enhance trial efficiency, diversity, and generalizability. This paper introduces the Framework for Research in Synthetic Control Arms (FRESCA), which uses a novel Recommender System combined with Equity Adjustment strategies to design and evaluate Representative Hybrid Clinical Trials (HCTs). FRESCA employs a novel matching algorithm through its recommendation system to select suitable patients from RWD while ensuring that the selected population is representative of the target demographic. This dual approach improves both patient selection and trial outcomes by balancing statistical appropriateness and equity. Simulations based on data from two existing randomized clinical trials (RCTs) show that using FRESCA to recommend patients from RWD and apply equity adjustments enhances internal validity and generalizability. Our analysis indicates that combining matching and equity adjustments yields more accurate treatment effect estimates and fair population representation, even with reduced RCT control group sizes. In contrast, using either method alone may result in biased outcomes. The flexibility of FRESCA to simulate various HCT scenarios makes it a valuable tool for advancing equitable and efficient clinical trial designs.

Keywords

Causal Inference, Equity, Hybrid Clinical Trials, Randomized Clinical Trial, Recommender Systems

1. Introduction

Enhancing the efficiency, diversity, and generalizability of clinical trials can be achieved by incorporating real-world data (RWD) [1]. This study introduces the Framework for Research in Synthetic Control Arms (FRESCA), which combines a novel Recommender System with strategies for Equity Adjustment to design and assess representative Hybrid Clinical Trials (HCTs). Synthetic control patients are patients created from pre-existing de-identified datasets, used to mimic the characteristics of a real control group in clinical trials. Synthetic control arms (SCAs) are especially useful in trials for rare diseases, where finding enough "in-trial" concurrent controls (CCs) can be difficult due to ethical and practical concerns [2] [3]. To address these challenges, HCTs use hybrid control arms (HCAs) that combine both concurrent and synthetic controls. FRESCA uses health recommender systems based on propensity score matching to recommend patients from external RWD who are suitable for inclusion in the trial, creating a hybrid population.

The health recommender system in FRESCA identifies patients from RWD who closely match the characteristics of those in the trial, enhancing the statistical power and reducing variance without extending the trial duration or increasing costs. However, integrating RWD with randomized control trial (RCT) data is challenging due to differences in their distributions [4]. To overcome this, FRESCA first uses its health recommender system to select appropriate patients and then applies equity adjustments to ensure the trial population accurately represents the target demographic. This combined approach

HealthRecSys'24: The 6th Workshop on Health Recommender Systems co-located with ACM RecSys 2024
*Corresponding author.
neehan@rpi.edu (N. Neehal); anand@us.ibm.com (V. Anand); bennek@rpi.edu (K. P. Bennett)
https://nafis-neehal.github.io/ (N. Neehal); https://research.ibm.com/people/vibha-anand (V. Anand); https://faculty.rpi.edu/kristin-bennett (K. P. Bennett)
0000-0001-7015-8039 (N. Neehal); 0000-0001-8605-5712 (V. Anand); 0000-0002-8782-105X (K. P. Bennett)

improves patient selection and ensures trial results are both statistically robust and demographically representative.

Our research aims to develop methods for conducting equitable HCTs and to provide a framework for evaluating them. In this context, equity means that the trial participants should represent a broader target population [5]. While RCTs can provide unbiased estimates for their specific cohorts, they often fail to represent larger, more diverse target populations. Researchers use data from these target populations to adjust RCT samples, ensuring that all relevant subgroups are included [6] [7]. Ensuring equity in hybrid trials is essential for generalizability and is a key focus for institutions like the NIH and FDA [8]. Our approach demonstrates that combining a health recommender system with equity adjustments creates a more balanced and representative trial population than using either method alone.

Table 1
Distribution of Protected Attributes in FRESCA Cohorts and Biased External Controls in ALLHAT and SPRINT trials along with NHANES Target Subgroup Rates

Attributes	ALLHAT			SPRINT			NHANES
	TA (N=8116)	CC (N=4000)	Biased EC (N=9762)	TA (N=4234)	CC (N=2000)	Biased EC (N=2200)	Target Rate
Age Group							
40-59	1549 (19.1%)	772 (19.3%)	5841 (59.8%)	923 (21.8%)	438 (21.9%)	334 (15.2%)	31.2%
59+	6567 (80.9%)	3228 (80.7%)	3921 (40.2%)	3311 (78.2%)	1562 (78.1%)	1866 (84.8%)	68.8%
Gender							
Female	3728 (45.9%)	1875 (46.9%)	2752 (28.2%)	1499 (35.4%)	691 (34.6%)	631 (28.7%)	55.4%
Male	4388 (54.1%)	2125 (53.1%)	7010 (71.8%)	2735 (64.6%)	1309 (65.5%)	1569 (71.3%)	44.6%
Race or Ethnicity							
Hispanic	1389 (17.1%)	740 (18.5%)	1671 (17.1%)	479 (11.3%)	225 (11.3%)	389 (17.7%)	10.0%
NH Asian	84 (1.0%)	41 (1.0%)	693 (7.1%)	42 (1.0%)	15 (0.8%)	138 (6.3%)	3.9%
NH Black	2525 (31.1%)	1265 (31.6%)	1937 (19.8%)	1232 (29.1%)	616 (30.8%)	468 (21.3%)	12.0%
NH White	4050 (49.9%)	1921 (48.0%)	5285 (54.1%)	2451 (57.9%)	1128 (56.4%)	1172 (53.3%)	69.3%
Other	68 (0.8%)	33 (0.8%)	176 (1.8%)	30 (0.7%)	16 (0.8%)	33 (1.5%)	4.8%

FRESCA was developed to design and evaluate these methods [9]. Before FRESCA, existing methods for HCTs did not explicitly address equity in patient selection. FRESCA uses real clinical trial data to simulate hypothetical trials, applying its methods to scenarios based on real RCTs, such as Systolic Blood Pressure Intervention Trial (SPRINT) [10] and Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) [11]. By integrating health recommender systems and equity adjustment techniques, FRESCA ensures that the selected patients not only meet statistical criteria but also represent the demographic characteristics of the target population. We define protected subgroups based on age, gender, and race/ethnicity, using NHANES data [12] to estimate the rates for these subgroups in a simulated target population. We will explore the use of additional protected attributes in future work. FRESCA includes five main functions: generating cohorts, simulating scenarios, calculating target subgroup rates, estimating treatment effects and equity, and providing a final assessment. It can evaluate any HCT method, including those that combine health recommender systems with equity adjustments. Detailed descriptions of the FRESCA framework and the trial configurations are provided in sections 2.4 and 2.5.

This paper makes several key contributions-

- It identifies equity issues in HCTs and proposes solutions to improve the representativeness of trial populations.
- It introduces an enhanced FRESCA framework with a modular architecture that supports multi-trial, multi-outcome, and multi-metric assessments for creating and evaluating HCTs.
- It evaluates HCT methods that combine health recommender systems based on novel matching algorithm with equity adjustments, showing that the best results come from combining propensity score matching with IPF [13] weighting.
- It demonstrates that using both health recommender systems and equity adjustments results in more equitable populations and precise estimates of the Population Hazard Ratio (PHR), even

with smaller CC sizes.

- It shows that variations in the sizes of treatment and control groups significantly affect the precision of treatment effect estimates, and that a balanced use of recommended synthetic controls is essential for accuracy.

There are several strategies for integrating synthetic control populations into trial populations. Most approaches use propensity score matching to select suitable external controls. Some methods focus on matching based on treatment propensity [4, 14, 15], while others use propensity to predict trial participation [3, 2, 1]. Bayesian approaches are also used to incorporate synthetic controls into trials [16, 17], but they typically do not consider equity adjustments for a target population. Future work with FRESCA could explore these and other methods to develop new strategies for equitable synthetic control arms and assess their effectiveness. Further theoretical exploration of these and other HCT algorithms is also planned.

2. Methodology

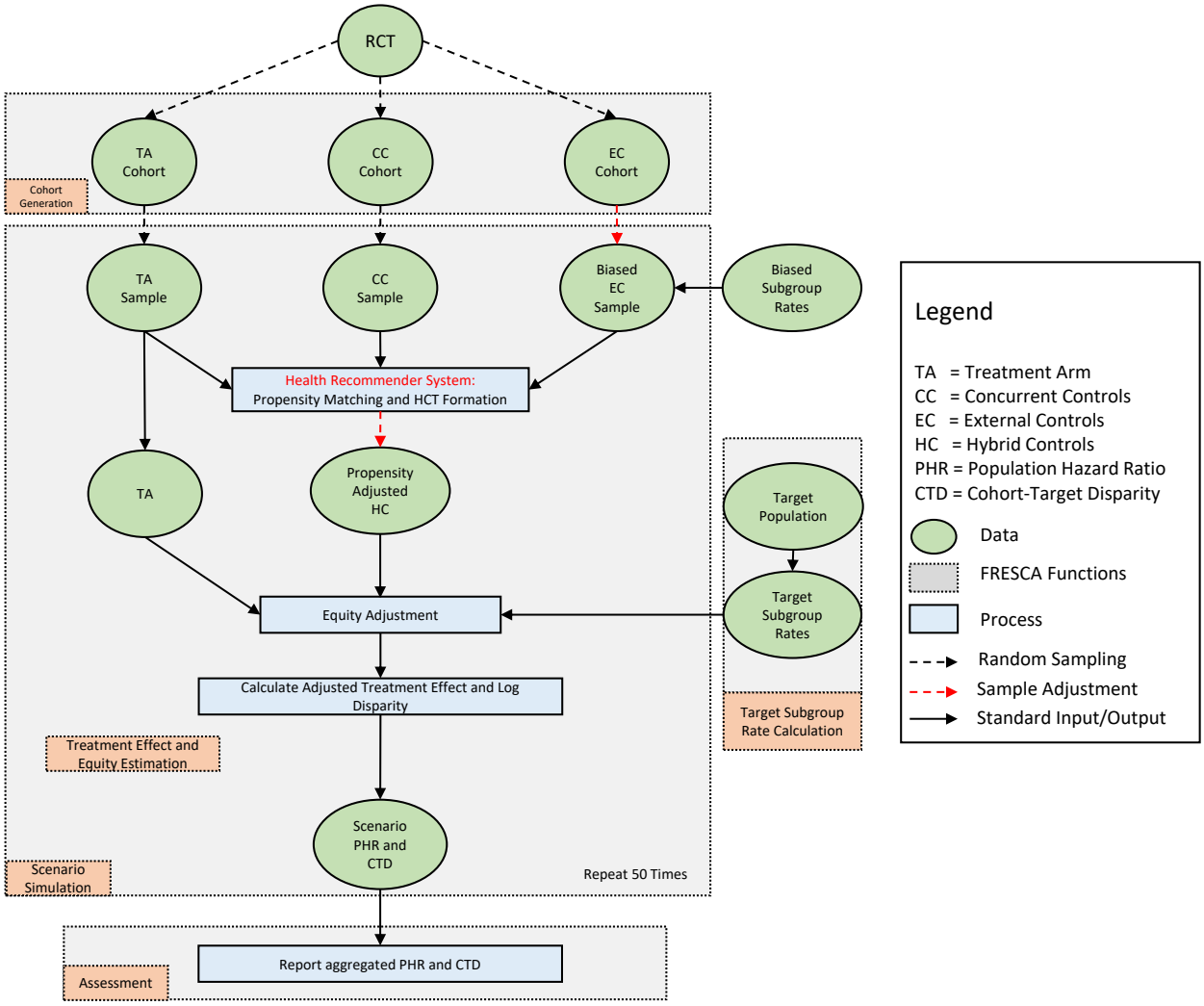


Figure 1: The FRESCA framework for hybrid clinical trials has five main functions. It utilizes a health recommender system for the "Propensity Matching and HC Formation" process and supports any standard method for distribution adjustment in the "Equity Adjustment" process.

2.1. Problem Definition

We define the problem using the potential outcomes framework from Neehal et al. [9]. Let Y_{st}^i represent the potential outcome for subject i in sample s under treatment t, where $s = 0$ is the target population, $s = 1$ the RCT population, $s = 2$ the RWD population, and $s = 3$ an adjusted sample combining RCT and RWD data. The Sample Hazard Ratio (SHR) in the RCT is $SHR = E(\text{effect}(Y_{11}, Y_{10})|S = 1)$, where $\text{effect}(Y_{11}, Y_{10})$ is the difference in treatment effects between treated and control groups. The Population Hazard Ratio (PHR) for the target population ($s = 0$) is defined as $PHR = E(\text{effect}(Y_{01}, Y_{00})|S = 0)$, representing the expected treatment effect in the target population. Two main issues arise: (1) *Equity*—the RCT may not represent the target population, leading to biased estimates ($PHR \neq SHR$), and (2) *Sample Size*—insufficient patients in the control group may require synthetic controls (SCs) from RWD. To accurately estimate PHR, we use a health recommender system with propensity score matching to augment RCT data with SCs, and then perform appropriate equity adjustment, forming an "equity-adjusted" sample ($s = 3$).

2.2. Data

We define the target population using the nationally representative hypertensive cohort from the National Health and Nutrition Examination Survey (NHANES) 2015-2016 [12]. Representativeness in the RCTs is assessed based on three protected attributes: Gender (Male, Female), Age Group (40-59, 59+), and Race/Ethnicity (Non-Hispanic Black, Non-Hispanic White, Non-Hispanic Asian, Hispanic, Other). These age groups align with the inclusion criteria of the SPRINT and ALLHAT hypertension studies. Target rates for each subgroup are calculated using survey-weighted analysis of US subjects aged 40+ with hypertension.

We use data from SPRINT and ALLHAT trials available from BioLINCC [18]. For ALLHAT, we focus on the Amlodipine vs. Clorthalidone group, as the results are similar for the Lisinopril vs. Clorthalidone group. After preprocessing, SPRINT includes 4,234 treated and 4,200 control patients, and ALLHAT includes 8,116 treated and 13,762 control patients. The primary outcome for SPRINT is a composite of major cardiovascular events, while for ALLHAT, the outcome is heart failure. Figure 1 illustrates how FRESCA divides the RCT data into Treatment Arm (TA), Concurrent Controls (CC), and External Controls (EC) cohorts. Table 1 presents the distribution of protected attributes in the RCT data and the Biased External Controls for both SPRINT and ALLHAT, showing their differences from the target NHANES population.

The NHANES surveillance data and clinical trial data from BioLINCC were used with appropriate approvals: BioLINCC approved ALLHAT and SPRINT data use under case 123537, and NHANES data is freely available and exempt from human subjects research regulations per Rensselaer Polytechnic Institute IRB 1863.

2.3. Adjustment Methods and Assessment Metrics

For balancing distributions between synthetic control and trial populations through the recommender system, we employ propensity score matching using the "MatchIt" R package [19]. Iterative Proportional Fitting (IPF) via the "IPFR" R package [20] is used for equity adjustment and Biased EC Cohort formation.

Treatment effects are assessed using Cox's Proportional Hazards Regression to estimate the Population Hazard Ratio (PHR). The "ground-truth" Target PHR is estimated by equity adjustment on the entire RCT dataset. For equity assessment, we use a variant of log disparity (LD)[5]:

$$\log\left\{\frac{odds(g(x)=1|y'=1)}{odds(g(x)=1|y=1)}\right\} \tag{1}$$

where $g(x)$ is the protected group, y' is the observed cohort, and y is the target population. Absolute LD values between 0 and 0.22 are considered equitable [5]. We introduce Cohort-Target Disparity (CTD) as the mean of median absolute LD values across simulated runs, calculated for subgroups defined by age,

race, and gender. This provides a comprehensive measure of demographic representativeness between the study cohort and target population.

2.4. FRESCA Framework

FRESCA integrates a health recommender system to select suitable synthetic controls from external data and combine them with RCT data, forming a hybrid control population. This framework uses propensity score matching for patient selection and equity adjustments to ensure accurate estimates of the Population Hazard Ratio (PHR) while maintaining representativeness for any target population. FRESCA provides tools to assess the effectiveness of these methods, as shown in Fig 1, and comprises five main functions. We demonstrate FRESCA's application using the SPRINT and ALLHAT trials with NHANES as the target population, but the framework is flexible and can be adapted to any RCT, RWD, or target population.

2.4.1. Cohort Generation

In the Cohort Generation phase, FRESCA employs its health recommender system to generate three cohorts: TA (Treatment Arm), CC (Concurrent Controls), and EC (External Controls). The TA and CC cohorts are derived from RCT data, representing the treatment and control groups, respectively. The EC cohort is sourced from real-world data, providing a pool of synthetic controls recommended by the health recommender system to supplement the RCT. These cohorts collectively form the basis for subsequent analyses.

2.4.2. Scenario Simulation

FRESCA facilitates the creation of diverse simulated trial scenarios to calculate adjusted PHRs and equity metrics. This involves two stages: first, generating unbiased samples from the TA and CC cohorts to simulate a randomized clinical trial; second, creating a Biased EC Sample from the EC cohort, reflecting biased real-world data as the source of synthetic controls. The health recommender system plays a key role in selecting these controls. Further details on simulation configurations are provided in section 2.5.

Table 2
Comparison of PHR and CTD across different trials, outcomes and methods. We show this for ALLHAT ($N_{TA} = 4000$, $N_{CC} = 2000$) and SPRINT ($N_{TA} = 2000$, $N_{CC} = 1000$) respectively. Symbol (†) in Cohort-Target Disparity column indicates measured CTD not being within equitable range ($CTD > 0.22$). Bold font indicates the best performing method.

Trial (Study)	Outcome Examined	Control Population	Adjustment Method	Target PHR [95% CI]	Estimated PHR [95% CI]	Cohort-Target Disparity [95% CI]
ALLHAT (Hypertension)	Secondary (Heart Failure)	CC	None	1.38 [1.36, 1.41]	1.39 [1.36, 1.43]	0.89 [0.84, 0.94]†
		Hybrid	NC Matching		1.42 [1.37, 1.48]	0.87 [0.81, 0.94]†
		Hybrid	Propensity Matching + IPF Sampling		1.43 [1.32, 1.49]	0.03 [0.02, 0.04]
		Hybrid	Propensity Matching + IPF Weighting		**1.39 [1.33, 1.46]**	**0.04 [0.03, 0.05]**
SPRINT (Hypertension)	Primary	CC	None	0.79 [0.77, 0.82]	0.75 [0.73, 0.78]	0.91 [0.86, 0.97]†
		Hybrid	NC Matching		0.74 [0.72, 0.77]	0.89 [0.84, 0.96]†
		Hybrid	Propensity Matching + IPF Sampling		0.75 [0.67, 0.84]	0.01 [0.00, 0.01]
		Hybrid	Propensity Matching + IPF Weighting		**0.78 [0.74, 0.81]**	**0.04 [0.03, 0.05]**

2.4.3. Target Subgroup Rates Calculation

The target rates for each subgroup are calculated using a survey-weighted analysis of the desired target population from NHANES (e.g., adults with hypertension).

2.4.4. Treatment Effect and Equity Estimation

Once the scenario samples are created and target subgroup rates are determined, the next step is to construct an equity-adjusted HCT population and estimate the treatment effect. FRESCA utilizes a

health recommender system based on propensity score matching to recommend suitable synthetic controls (SCs) from the Biased EC Sample. A binary logistic regression model, incorporating TA, CC, and Biased EC samples, generates propensity scores to select SCs, thereby forming a matched Hybrid Control Arm (HCA). Equity adjustments are then applied to both the TA and HCA cohorts using the Iterative Proportional Fitting (IPF) technique to better align them with the target population. This process results in specific weight vectors, W_{IPF_TA} for the TA and W_{IPF_HCA} for HCA. Unlike the previously used approach [9], where these weights were used to generate random samples, we now directly compute the weighted and equity-adjusted treatment effect and equity value (LD) using these weights.

2.4.5. Assessment

In the Assessment phase, FRESCA evaluates various HCT construction methods by combining the health recommender system for propensity matching with two types of IPF-based equity adjustments: weighted and sampling. Baseline scenarios without any adjustments or inclusion of SC are also evaluated and compared with the NC Matching technique [14], with results summarized in Table 2. To assess the precision of PHR estimations, FRESCA compares them with a "ground-truth" target PHR, derived from equity adjustments applied to the complete RCT dataset (e.g., SPRINT/ALLHAT) using all treated and control subjects. The data set is divided into treated and control cohorts, bootstrapped to match the sizes, and adjusted to align with the NHANES population. The target PHR is calculated as an average across all bootstrap samples and scenarios. Equity is evaluated by checking if the Cohort-Target Disparity (CTD) falls within the [0, 0.22] range, adhering to the 80% rule [5].

2.5. Simulation of HCT Scenarios

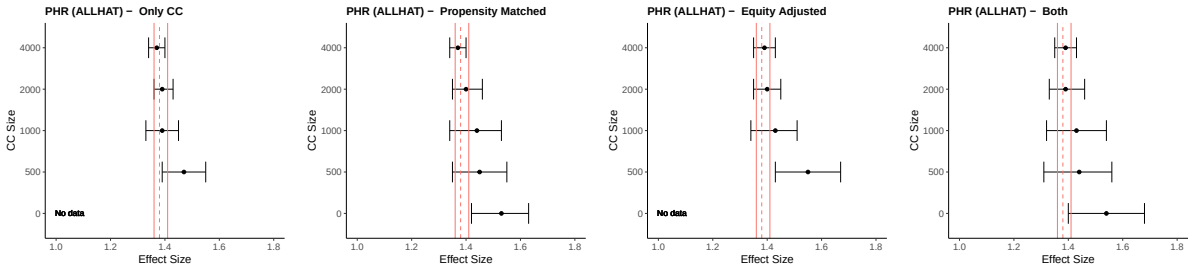
FRESCA simulates various clinical trial scenarios to evaluate the effects of different trial design parameters on outcomes using a health recommender system for patient selection.

2.5.1. ALLHAT

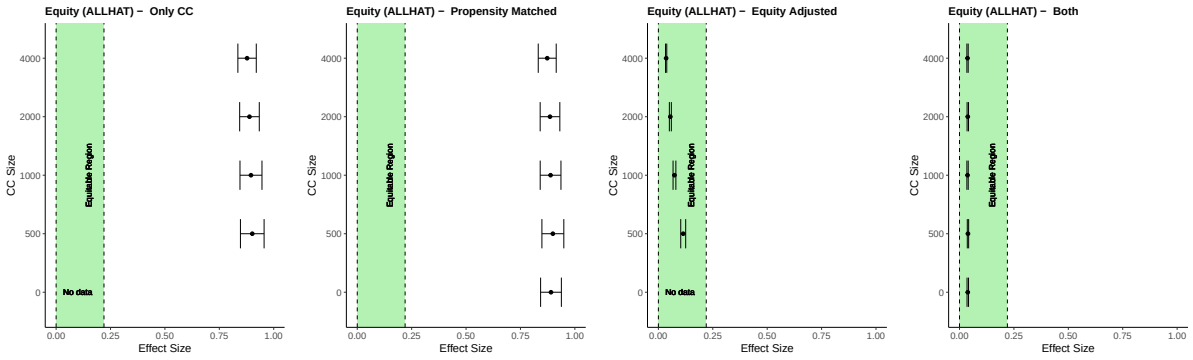
For ALLHAT, FRESCA creates a Concurrent Control (CC) Cohort by selecting 4,000 individuals from the ALLHAT trial's original control group, leaving 9,762 in the External Control (EC) Cohort. The Treatment Arm (TA) remains unchanged with 8,116 participants. During simulation, biases are introduced in the EC Cohort using the Iterative Proportional Fitting (IPF) method to reflect biased subgroup rates for age, gender, and race, as well as smoker status, depression history, and HDLC history (Table 1). FRESCA explores different experimental scenarios with varying sample sizes for TA (N_{TA}=4000) and CC (N_{CC}= 0, 500, 1000, 2000, 4000), conducting 50 bootstrap simulations for each scenario. The mean Population Hazard Ratio (PHR) and Cohort-Target Disparity (CTD) are calculated for each scenario, along with a 95% Confidence Interval.

2.5.2. SPRINT

For SPRINT, FRESCA selects a CC Cohort of 2,000 from the control group (total N=4200), leaving 2,200 in the EC Cohort, with the TA consisting of 4,234 participants. IPF is used to adjust for biases in three protected attributes (Table 1) and additional factors such as Framingham Risk Score and Cardiovascular Disease (CVD) History. Simulations use a TA sample size (N_{TA}=2000) and vary CC sample sizes (N_{CC}=0, 500, 1000, 1500, 2000), running 50 bootstrap simulations per scenario. Results include mean PHR, CTD, and 95% Confidence Intervals, as reported in the final assessment.



(a) PHR Estimates measured using ALLHAT trial data



(b) Cohort-Target Disparity measured using ALLHAT trial data

Figure 2: Variation of PHR Estimates and Cohort-Target Disparity in ALLHAT Trial with Different CC Sizes. This figure illustrates the influence of various CC sizes (N_{CC} = 0, 500, 1000, 2000, 4000) on the PHR estimates and equity measures for ALLHAT trial with a treatment arm size of N_{TA} = 4000. In panel (a), the target PHR is demarcated by two solid red lines, encompassing the red dashed line representing the 95% confidence interval. Panel (b) features a green shaded area delineated by two black dashed lines, indicating the range considered equitable.

3. Results and Discussion

3.1. Performance comparison of different methods for creating HCTs

Our study evaluates methods for creating HCTs using two metrics as described above: PHR and CTD. We analyze methods that combine propensity score matching with Iterative Proportional Fitting (IPF) equity-adjustment methods (weighting or random sampling) and compare them to two baseline scenarios: (i) no adjustments applied to the CC population, and (ii) SC added to trial via NC Matching [14] algorithm. The results for the ALLHAT and SPRINT trials are detailed in Table 2. Our analysis reveals two key findings: (i) *Necessity of both Propensity and Equity Adjustments:* PHR estimates without equity adjustments (either by IPF weighting or sampling) are typically inequitable as evident by CTD values, and they may produce biased treatment effects, as evident in high CTD values, and (ii) *Superiority of Weighting over Sampling:* Using IPF Weighting for equity adjustment improves the accuracy and consistency of PHR estimates compared to IPF Sampling. These findings, particularly the need for comprehensive adjustments and the effectiveness of sample weighting, have been observed consistently across multiple trials and outcomes, demonstrating the robustness of the FRESCA framework.

3.2. Examination of Variation in CC Size on PHR and CTD estimation

We studied the effect of varying CC population sizes on the estimated PHR and CTD in HCT. We examined four methods: (i) no adjustments, (ii) only propensity matching, (iii) only equity adjustment, and (iv) both propensity and equity adjustments. Fig 2 shows the results on ALLHAT for CC sizes varying from 0 to 4000. Two main findings emerged - (i) *Benefits of Synthetic Controls for Limited Data:* In scenarios with smaller CC sizes, missing subgroups were compensated by incorporating SC. This strategy, especially with both propensity and equity adjustments, yielded accurate estimates, and (ii)

Table 3
The effect and significance of several trial design parameters in predicting the bias in PHR estimation. Here the bias is defined to be the squared deviation of the estimated PHR from the target PHR. Star (*) symbol represents a significant effect with $p<0.05$.

Predictors of Linear Model	Estimate	2.5%	97.5%	P Value
TA Size	0.022	0.009	0.036	0.001*
CC Size	-0.134	-0.157	-0.110	0.000*
Cohort-Target Disparity (CTD)	-0.062	-0.183	0.059	0.316
Cohort-RCT Disparity (CRD)	0.160	0.021	0.298	0.024*
Controls (Only Equity)	-0.132	-0.264	-0.001	0.048*
Controls (Both)	-0.089	-0.222	0.043	0.186
Controls (Only Propensity)	0.035	0.020	0.049	0.000*

PHR Accuracy and Acceptable Equity with Reduced CC Population Size: Reducing the CC population size by 50% (from 4000 to 2000) still produced PHR estimates close to the target PHR. However, a larger CC size is preferable for lower estimation variance, indicating a trade-off for trial designers. These patterns were consistent across both SPRINT and ALLHAT trials, affirming the robustness of our findings. We only show results for weighting equity adjustment for the ALLHAT trial for brevity, and additional results are available in the Supplementary.

3.3. Examining effects of multiple factors for predicting PHR estimation accuracy

We analyzed factors affecting the accuracy of PHR estimates in ALLHAT using a linear model with seven predictors. Bias in PHR estimation was quantified as the squared deviation from the target PHR. Predictors included the size of the treatment arm (TA Size), the control group size (CC Size), Cohort-Target Disparity (CTD), and Cohort-RCT Disparity (CRD), which measures the distribution differences between control populations and the RCT population. Controls were categorized by adjustment type: propensity, equity, both, or none. Results in Table 3 showed that TA Size, CC Size, and CRD significantly predicted PHR bias. Key findings include: (i) *Larger CC Size Reduces Bias:* Increasing CC size lowers bias, favoring a larger control group directly recruited over synthetic controls; (ii) *Impact of Adjustments:* "Only Equity" and "Propensity and Equity" adjustments reduce bias compared to the "Only CC" category, while "Only Propensity" adjustments increase bias, highlighting the importance of equity adjustments for accurate PHR estimates.

3.4. Examining the effect of TA and CC Size ratio for a Fixed Size Recruitment Trial

We examined the balance between TA and CC population sizes when supplemented by SC in clinical trials with a fixed recruitment size. Using ALLHAT and SPRINT trial data, we maintained a total recruited participant cap of 4000, varying TA and CC sizes with corresponding SC adjustments. Four

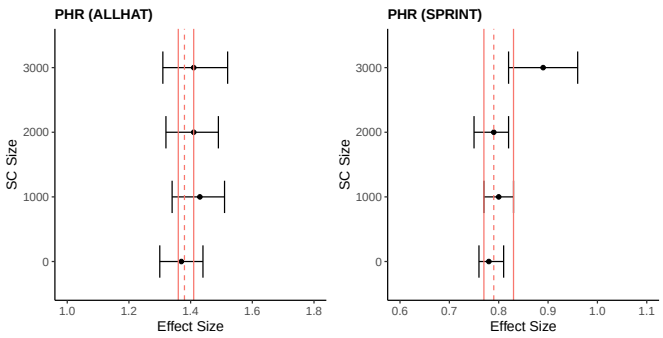


Figure 3: Examining the influence of varied TA, CC, and SC Sizes on PHR Estimation in Fixed Sized Recruitment Trials.

scenarios were considered with TA sizes of 3500, 3000, 2500, 2000 and CC sizes of 500, 1000, 1500, 2000, inversely adjusting SC sizes 3000, 2000, 1000, 0. The PHR estimates from these scenarios are shown in Figure 3. Key findings include: (i) The variance of PHR estimates increases with SC size, affecting the stability of treatment effect estimation, and (ii) The PHR estimate can significantly shift with a highly imbalanced ratio of real to synthetic data; especially observed in some scenarios with substantially high SC size. This investigation therefore highlights the importance of carefully balancing the ratio of CC and SC patients in HC to ensure accurate treatment effect estimates and avoiding erroneous conclusions about a trial's efficacy.

4. Conclusion

FRESCA offers a major advancement in equitable HCT methods and serves as a valuable tool for future research. It creates realistic HCT scenarios, using a health recommender system for propensity score matching and equity adjustments to provide more precise and equitable PHR estimates. Our simulations suggest that fewer patients may be needed to achieve results similar to full trials, but further research is required to determine the optimal balance of synthetic and concurrent controls in fixed-size trials. Future work will involve testing FRESCA with more realistic EHR data, with additional protected attributes and exploring the optimal size for CC recruitment during RCT design. Additionally, developing strategies that integrate matching and equity adjustments in a single step could enhance efficiency and reduce variance. These areas present opportunities for further refinement, making FRESCA a significant step forward in hybrid clinical research with potential for ongoing improvement.

Acknowledgments

This work was partially funded by IBM Research. This manuscript was prepared using SPRINT and ALLHAT study research materials obtained from the NHLBI Biologic Specimen and Data Repository Information Coordinating Center and does not necessarily reflect the opinions or views of SPRINT, ALLHAT or NHLBI.

References

[1] A. Sachdeva, R. C. Tiwari, S. Guha, A novel approach to augment single-arm clinical studies with real-world data, Journal of Biopharmaceutical Statistics 32 (2022) 141–157. doi:10.1080/10543406.2021.2011902.

[2] J. Harton, B. Segal, R. Mamtani, N. Mitra, R. A. Hubbard, Combining real-world and randomized control trial data using data-adaptive weighting via the on-trial score, Statistics in Biopharmaceutical Research (2022) 1–13. doi:10.1080/19466315.2022.2071982.

[3] X. Yin, P. S. Mishra-Kalyan, R. Sridhara, M. D. Stewart, E. A. Stuart, R. C. Davi, Exploring the potential of external control arms created from patient level data: a case study in non-small cell lung cancer, Journal of Biopharmaceutical Statistics 32 (2022) 204–218. doi:10.1080/10543406.2021.2011901.

[4] E. A. Stuart, D. B. Rubin, Matching with multiple control groups with adjustment for group differences, Journal of Educational and Behavioral Statistics 33 (2008) 279–306. doi:10.3102/1076998607306.

[5] M. Qi, O. Cahan, M. A. Foreman, D. M. Gruen, A. K. Das, K. P. Bennett, Quantifying representativeness in randomized clinical trials using machine learning fairness metrics, JAMIA Open 4 (2021) ooab077. doi:10.1093/jamiaopen/ooab077.

[6] E. Hartman, R. Grieve, R. Ramsahai, J. S. Sekhon, From sample average treatment effect to population average treatment effect on the treated: combining experimental with observational studies to estimate population treatment effects, Journal of the Royal Statistical Society. Series A (Statistics in Society) (2015) 757–778. doi:10.1111/rssa.12094.

[7] A. Y. Ling, M. E. Montez-Rath, P. Carita, K. J. Chandross, L. Lucats, Z. Meng, B. Sebastien, K. Kapphahn, M. Desai, An overview of current methods for real-world applications to generalize or transport clinical trial findings to target populations of interest, Epidemiology 34 (2023) 627–636. doi:10.1097/EDE.0000000000001633.

[8] J. Petkovic, J. Jull, M. Yoganathan, O. Dewidar, S. Baird, J. M. Grimshaw, K. A. Johansson, E. Kristjansson, J. McGowan, D. Moher, et al., Reporting of health equity considerations in cluster and individually randomized trials, Trials 21 (2020) 1–12. doi:10.1186/s13063-020-4223-5.

[9] N. Neehal, V. Anand, K. P. Bennett, Framework for research in equitable synthetic control arms, in: AMIA Annual Symposium Proceedings, volume 2023, American Medical Informatics Association, 2023, p. 530.

[10] J. T. Wright Jr, et al., A randomized trial of intensive versus standard blood-pressure control, New England Journal of Medicine 373 (2015) 2103–2116. doi:10.1056/NEJMoa1511939.

[11] C. D. Furberg, et al., Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The antihypertensive and lipid-lowering treatment to prevent heart attack trial (allhat), JAMA 288 (2002) 2981–2997. doi:10.1001/jama.288.23.2981.

[12] Centers for Disease Control and Prevention, National health and nutrition examination survey (nhanes) data, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2023. URL: https://www.cdc.gov/nchs/nhanes/index.htm, accessed on: October 2023.

[13] W. E. Deming, F. F. Stephan, On a least squares adjustment of a sampled frequency table when the expected marginal totals are known, The Annals of Mathematical Statistics 11 (1940) 427–444. doi:10.1214/aoms/1177731829.

[14] J. Yuan, J. Liu, R. Zhu, Y. Lu, U. Palm, Design of randomized controlled confirmatory trials using historical control data to augment sample size for concurrent controls, Journal of Biopharmaceutical Statistics 29 (2019) 558–573. doi:10.1080/10543406.2018.1559853.

[15] Y. Liu, B. Lu, J. Foster, Y. Zhang, Z. J. Zhong, M.-H. Chen, P. Sun, Matching design for augmenting the control arm of a randomized controlled trial using real-world data, Journal of Biopharmaceutical Statistics 32 (2022) 124–140. doi:10.1080/10543406.2021.2011900.

[16] K. Viele, et al., Use of historical control data for assessing treatment effects in clinical trials, Pharmaceutical statistics 13 (2014) 41–54. doi:10.1002/pst.1589.

[17] X. Pang, et al., A bayesian alternative to synthetic control for comparative case studies, Political Analysis 30 (2022) 269–288. doi:10.1017/pan.2021.22.

[18] National Heart, Lung, and Blood Institute, Biologic specimen and data repository information coordinating center, https://biolincc.nhlbi.nih.gov/home/, 2024. Accessed September 22, 2024.

[19] D. E. Ho, K. Imai, G. King, E. A. Stuart, Matchit: Nonparametric preprocessing for parametric causal inference, Journal of Statistical Software 42 (2011) 1–28. doi:10.18637/jss.v042.i08.

[20] K. Ward, IPFR: List Balancing for Reweighting and Population Synthesis, 2020. URL: https://CRAN.R-project.org/package=ipfr, r package version 1.0.2.