Power and sample size calculation for non-inferiority trials with treatment switching in intention-to-treat analysis comparing RMSTs

Austin Shih
Vanderbilt University
Chih-Yuan Hsu
chih-yuan.hsu@vumc.org
Vanderbilt University Medical Center
Yu Shyr
Vanderbilt University Medical Center

Research Article

Keywords: crossover, intention-to-treat analysis, non-inferiority trials, restricted mean survival time

Posted Date: December 12th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-5418253/v1

Additional Declarations: No competing interests reported.

Power and sample size calculation for non-inferiority trials with treatment switching in intention-to-treat analysis comparing RMSTs

Austin Shih[1,#], Chih-Yuan Hsu[2,3,#], and Yu Shyr[2,3,*]

[1] Department of Mathematics, Vanderbilt University, Nashville, TN, USA

[2] Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA

[3] Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

[#]Austin Shih and Chih-Yuan Hsu contributed equally to this work

[*]Correspondence: Yu Shyr: yu.shyr@vumc.org

Abstract

Background: Difference in Restricted Mean Survival Time (DRMST) has attracted attention and is increasingly used in non-inferiority (NI) trials because of its superior power in detecting treatment effects compared to hazard ratio. However, when treatment switching (also known as crossover) occurs, the widely used intention-to-treat (ITT) analysis can underpower or overpower NI trials.

Methods: We propose a simulation-based approach, named *nifts,* to calculate powers and determine the necessary sample size to achieve a desired power for non-inferiority trials that allow treatment switching, in ITT analysis using DRMST.

Results: Real-world and simulated examples are used to illustrate the proposed method and examine how switching probability, switching time, the relative effectiveness of treatments, allocation ratio, and even time distribution influence powers and sample sizes. Our simulation study shows that switching time and switching probability decrease or increase powers and sample sizes compared to those in the scenarios without treatment switching. A shorter switching time and a higher switching probability amplify the magnitude of these changes. The direction of the change in powers and sample sizes depends on the relative effectiveness of the treatments. When $m_2/m_1>1$, power decreases and sample size increases, while $m_2/m_1<1$ leads to the opposite effect, where m_1 and m_2 are the median survivals in the control and experimental groups, respectively.

Conclusions: This simulation-based approach offers a preview of how treatment switching can influence powers and sample sizes in NI trials, providing investigators with useful information before conducting the trials. *nifts* is freely available at https://github.com/cyhsuTN/nifts.

Keywords

crossover; intention-to-treat analysis; non-inferiority trials; restricted mean survival time

1 Introduction

A randomized controlled trial (RCT) is regarded as the gold standard for assessing the effectiveness of new treatments. Among the various types of RCTs, a non-inferiority (NI) trial aims to demonstrate that a new treatment is not significantly worse than an existing one, while potentially offering additional benefits such as fewer side effects or lower costs. One increasingly popular approach for evaluating treatment effects in NI trials with time-to-event outcomes is to compare restricted mean survival times (RMST) between treatment groups [1-3]. RMST provides a straightforward summary by averaging survival times up to a specified time point [4] and does not rely on the proportional hazards (PH) assumption, which is frequently violated in clinical trials [5]. As a result, Royston and Parmar [6] suggested using the difference in RMSTs (DRMST) between treatment groups as an alternative to the hazard ratio (HR) for designing randomized trials with time-to-event outcomes, including power and sample size calculations. Furthermore, DRMST has greater power in detecting treatment effects compared to HR, even under the PH assumption [7, 8]. Methods for determining powers and sample sizes in NI trials using DRMST have been proposed [9, 10].

In RCTs, including NI trials, treatment switching from the control group to the experimental group may occur due to ethical concerns or other reasons [11, 12]. This switch may happen when a disease progresses, when healthcare providers believe the patient's prognosis will improve with the experimental treatment, or when patients prefer the new treatment due to perceived benefits such as fewer side effects or greater convenience [11, 13]. However, treatment switching can confound the results of intention-to-treat (ITT) analysis, making it difficult to determine the true treatment effect. ITT analysis includes all participants with randomization and compares their responses to determine the treatment effect according to the initially assigned treatment groups,

regardless of what treatment they received. This may potentially lead to underpowered trials and inconclusive results [12]. An alternative approach is per-protocol analysis that excludes participants who switch treatments from the analysis. Nevertheless, this can heavily bias the results if there is a significant difference in prognosis between the included and excluded participants, particularly if the treatment switching is associated with prognostic variables [14]. Therefore, ITT analysis is still often used in the final analysis. Deng et al. [15] proposed a simulation-based approach to preview power reduction and sample sizes required in superiority trials with treatment switching in ITT analysis using the logrank test.

In this study, we propose a simulation-based approach, named *nifts,* to determine power and sample size in NI trials that involve treatment switching when comparing RMSTs between two treatment groups in ITT analysis. To accelerate the computation of sample sizes, a monotonic smoothing technique is employed to estimate the power trend as sample sizes increase [16]. We utilize both real-world and simulated examples to illustrate the proposed method and examine how switching probability, switching time, the relative effectiveness of treatments, allocation ratio, and even time distribution influence power and sample sizes. *nifts* is freely available at https://github.com/cyhsuTN/nifts.

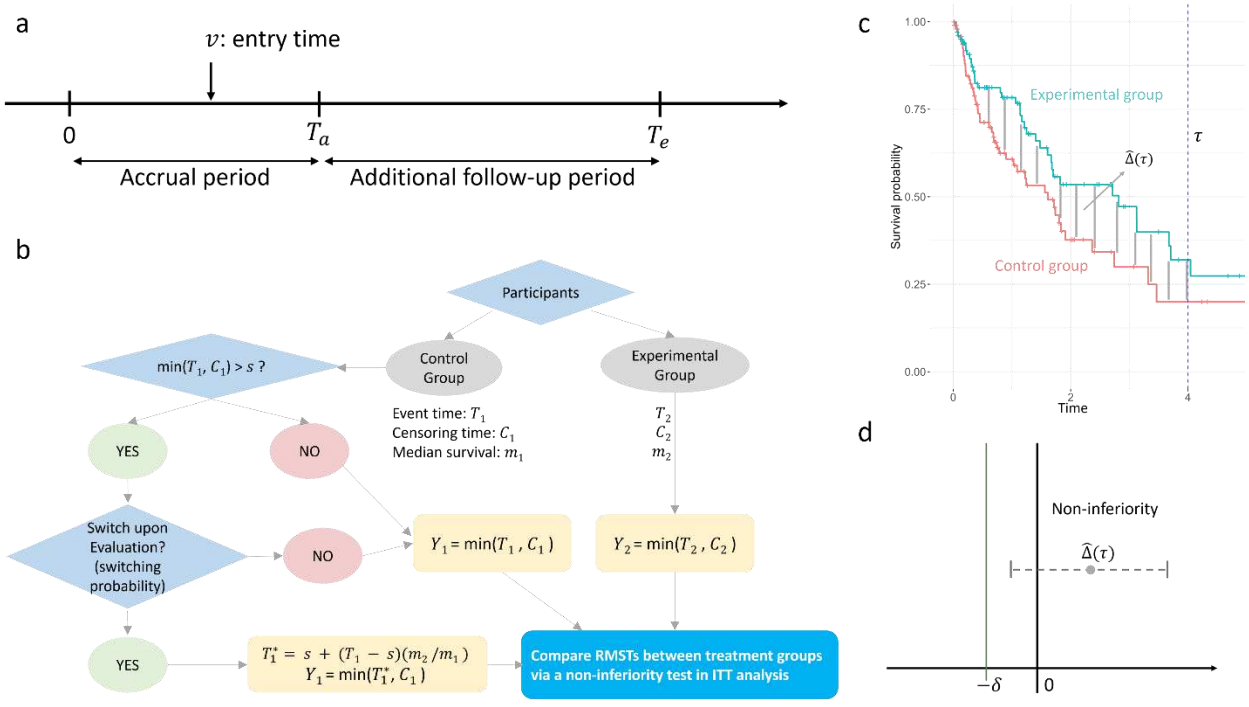

Figure 1: An overview of *nifts*. (a) Accrual time and trial duration. (b) Treatment switching. (c) Difference in RMSTs between two treatment groups. (d) Non-inferiority holds if the lower bound of DRMST is larger than $-\delta$.

2 Methods

2.1 Non-inferiority Trials using DRMST

Denote the survival functions for the control group and the experimental group by $S_1(t)$ and $S_2(t)$, respectively. The restricted mean survival times (RMST) at a specified time τ ($\tau > 0$) for the two groups are defined as $R_i(\tau) = \int_0^\tau S_i(t)\, dt$, i = 1 and 2. The difference in RMSTs between the two groups (DRMST) is given by $\Delta(\tau) = R_2(\tau) - R_1(\tau)$. The estimate for $\Delta(\tau)$ is $\hat{\Delta}(\tau) = \hat{R}_2(\tau) - \hat{R}_1(\tau)$, where $\hat{R}_i(\tau) = \int_0^\tau \hat{S}_i(t)\, dt$ and $\hat{S}_i(t)$ is the Kaplan-Meier estimate for $S_i(t)$. The aim of a non-inferiority trial using DRMST is to test $H_0: \Delta(\tau) \le -\delta$ vs $H_1: \Delta(\tau) > -\delta$, where $\delta > 0$ is a prespecified margin. If

$$\hat{\Delta}(\tau) - z_{1-\alpha}\mathrm{SE}\left(\hat{\Delta}(\tau)\right) > -\delta,$$

we reject the null hypothesis at a one-sided significance level of α and claim that non-inferiority holds (i.e., the experimental treatment is not significantly worse than the control). Here, $z_{1-\alpha}$ represents the $(1-\alpha)$th quantile of the standard normal distribution, and $\mathrm{SE}\left(\hat{\Delta}(\tau)\right) = \sqrt{\widehat{Var\left(\hat{R}_1(\tau)\right)} + \widehat{Var\left(\hat{R}_2(\tau)\right)}}$. $\widehat{Var\left(\hat{R}_i(\tau)\right)}$ is the estimate for the variance of $\hat{R}_i(\tau)$, whose explicit expression can be found in [9]. Both $\hat{R}_i(\tau)$ and $\widehat{Var\left(\hat{R}_i(\tau)\right)}$ can be calculated using the *survfit* function in the *survival* R package.

2.2 The Choice of Margins

In this study, we propose three options for selecting margins.

Preserved fraction of the RMST of the control group

We aim for $R_2(\tau)$ to maintain at least the preserved fraction, f_1, of the RMST of the control group, where $0 < f_1 < 1$. This means $R_2(\tau) > f_1 R_1(\tau)$. Thus, $\Delta(\tau) > -(1-f_1)R_1(\tau)$ and $\delta = (1-f_1)R_1(\tau)$.

Preserved fraction of the DRMST between the control and the placebo groups

In this option, we aim for the RMST of the experimental group to be better than the RMST of the placebo group, and the DRMST between the experimental and placebo groups to maintain at least the preserved fraction, f_2, of the DRMST between the control and placebo groups, where $0 < f_2 < 1$. This means $R_1(\tau) - R_0(\tau) > 0$ and $R_2(\tau) - R_0(\tau) > f_2\big(R_1(\tau) - R_0(\tau)\big)$, where $R_0(\tau) = \int_0^\tau S_0(t)\, dt$ and $S_0(t)$ is the survival function for the placebo group. Typically, $R_1(\tau) - R_0(\tau) > 0$ holds, so $\delta = (1 - f_2)\big(R_1(\tau) - R_0(\tau)\big)$.

Conversion from the hazard ratio

Given $S_1(t)$ and assuming proportional hazards, a margin $(1/\theta)$ for the hazard ratio (HR) of the experimental group to the control group can be converted to a margin for DRMST from $\mathrm{HR}_{21} < 1/\theta$ to $\Delta(\tau) > -\delta$, where $\delta = R_1(\tau) - R_\theta(\tau)$ with $R_\theta(\tau) = \int_0^\tau \big(S_1(t)\big)^{1/\theta}\, dt$ and $0 < \theta < 1$.

2.3 The Design Setting and Assumption

Denote the trial duration by $T_e > 0$ and the accrual time during which participants are recruited by $T_a \geq 0$. $T_e - T_a \geq 0$ is the additional follow-up time. Participants are assumed to enter the study uniformly during the accrual period, i.e., $v \sim U(0, T_a)$, where v is the entry time of a participant (Figure 1a). If $T_a = 0$, all participants are assumed to enter the study at its start. We assume participants are randomly assigned to the control group or the experimental group with an allocation ratio of r, where r is defined as the ratio of the participants in the experimental group to those in the control group.

Denote the survival times for participants in the control and experimental groups by T_1 and T_2, respectively, and assume T_1 and T_2 follow Weibull distributions with the same shape parameter but different scale parameters. The scale and shape parameters of the two Weibull distributions are determined by given median survivals of $m = m_1$ and m_2, and a survival rate at a specific time t in the control group. Specifically, the scale and shape parameters are obtained by solving the equations: $\mathrm{scale}\,(\log 2)^{1/\mathrm{shape}} = m$ and $\exp(-(t/\mathrm{scale})^{\mathrm{shape}}) = \text{survival rate}$.

Denote the censoring times for participants in the control and experimental groups by C_1 and C_2, defined as the duration from randomization to either dropping out of the trial or reaching the end

of the trial if participants don't experience the event of interest. Therefore, the censoring time comprises dropout censoring and administrative censoring, and its distributions can be formulated as follows [17]:

$$f(c|v) = d(c)I(0 < c < T_e - v) + \bar{D}(T_e - v)I(c = T_e - v),$$

where $d(c)$ and $\bar{D}(c)$ are the density function and survival function of the dropout censoring, respectively. $I(\cdot)$ is the indicator function. The dropout censoring is assumed to follow a uniform distribution $U(0, h)$, where h is determined by a given censoring rate of the control group under no treatment switching (see Supplementary Materials for details). For the scenario of no dropout censoring, we set $P(c = T_e - v|v) = 1$. Additionally, the distributions of the censoring times in the two groups are assumed to be the same.

2.4 Treatment Switching

nifts allows participants in the control group to switch to the experimental group if certain predetermined conditions are met. For example, if patients with cancer have a disease progression before death (assume death is the event of interest), they may switch from the standard treatment to the new treatment after disease progression and evaluation by the investigators [11]. Denote the switching time by s, defined as the duration from randomization to the moment when a participant may switch, with a probability p_s. The switching probability (p_s) is the likelihood that a participant who qualifies for treatment switching will switch from the control group to the experimental group after evaluation by healthcare professionals.

Five options for the distributions of the switching time are provided (Table 1), as used in [15]. The first three options assume s is correlated with T_1, while the other two options assume s is not correlated with T_1. The parameters in the assumed distributions are determined based on the given values of r_s and ρ_s (See Supplementary Materials for details). $r_s = E(s)/E(T_1)$ denotes the ratio of the average switching time to the average survival time of the control group, and ρ_s denotes the correlation between s and T_1.

The survival time for participants starting from switching is assumed to increase by m_2/m_1, based on the rank preserving structural failure time model (RPSFTM) [18]. Thus, the survival time of the participants with treatment switching will be $T_1^* = s + (T_1 - s) \times (m_2/m_1)$. Therefore, the observable survival time Y_1 for the participants without and with treatment switching from the

control group to the experimental group will be $\min(T_1, C_1)$ and $\min(T_1^*, C_1)$, respectively. The observable survival time Y_2 for the participants in the experimental group will be $\min(T_2, C_2)$. Finally, a non-inferiority test for DRMST between the two samples $\{Y_1\}$ and $\{Y_2\}$ in ITT analysis is performed (Figure 1b-1d).

2.5 The Proposed Method

The proposed *nifts* includes two main functions: *calculate_power* and *calculate_size*. The first function calculates power and outputs the associated expected number of events in the control and experimental groups. The latter determines the required sample size to achieve a specified power. The required sample size is obtained by a monotonically increasing power curve to the sample sizes. This curve is estimated using a monotonic smoothing technique [16] based on a finite number of power points and sample sizes.

The *calculate_power* function includes 21 parameters to simulate various scenarios: n, r, m_1, m_2, *shape*, f_1, m_0, f_2, *margin*, p_s, r_s, ρ_s, *s.dist*, *censoring.rate*, T_a, T_e, τ, *one.sided.alpha*, *TXswitch*, *n_simulations*, and *seed*. n: sample size of the control group, r: allocation ratio, m_1 and m_2: median survivals, *shape*: shape parameter of the Weibull distributions for event times, f_1 and f_2: preserved fractions, m_0: median survival of the placebo group for calculating $R_0(\tau)$ if f_2 is given, *margin*: non-inferiority margin, p_s: switching probability, r_s: ratio of $E(s)$ to $E(T_1)$, ρ_s: correlation of s and T_1, *s.dist*: options for the distributions of switching time (*s.dist* = "unif", "beta", "gamma", "indepExp", or a numeric value), *censoring.rate*: censoring rate of the control group (*censoring.rate* = "AC.only" meaning administrative censoring only, or = a numeric value), T_a and T_e: accrual duration and trial duration, τ: prespecified time for RMST calculation, *one.sided.alpha*: one-sided significance level, *TXswitch*: direction of treatment switching (*TXswitch* = "1to2" (default) or "2to1"), *n_simulations*: number of simulations, and *seed*: simulation seed. When f_1 is provided, the first margin option is used. When f_2 and m_0 are provided, the second margin option is used. A customized margin is applied when a numeric *margin* is provided, for example, an RMST margin converted from an HR margin.

The *calculate_size* function uses the same parameters as *calculate_power* while adding 4 parameters n_L, n_U, B, *epwr*. The lower (n_L) and upper (n_U) bounds are minimum and maximum sample sizes users input when exploring sample sizes for a desired expected power (*epwr*). The

function divides the range of the bounds into B equal intervals and calculate the powers at $n = n_L + k \times w$, where $w = \text{round}((n_U - n_L)/B)$ and $k = 0, 1, 2, ..., B$. A shape constrained additive model [16] is employed to fit a monotonically increasing power curve to the sample sizes, from which the required sample size is determined.

3 Results

3.1 Parameters Setting via Real-World Examples

The first example is an open-label phase III trial comparing survival benefits in patients with chemotherapy-refractory metastatic colorectal cancer, who were randomly assigned to either panitumumab + best supportive care (BSC) or BSC alone [11]. A total of 231 patients were randomly assigned to panitumumab + BSC, and 232 to BSC alone. Among the BSC alone patients, 85% experienced disease progression, and 76% switched to panitumumab + BSC after evaluation by the investigator. Thus, the switch probability p_s was 0.89 (= 0.76/0.85). We use the trial scenario (ClinicalTrial.gov: NCT00113763) to illustrate the proposed method for power and sample size calculation in NI trials with treatment switching when using DRMST in ITT analysis. We set $T_a = 0$, $T_e = 26$ (in months), $n = 232$ with a 1:1 allocation ratio ($r = 1$), and a censoring rate of 0.05 for BSC alone group. We compare the RMSTs of overall survival at $\tau = 12$ (in months) between the two groups with a preserved fraction of $f_1 = 0.8$ (i.e., margin = $0.2R_1(\tau)$), and assume Weibull distributions with $m_1 = 6.0$ and $m_2 = 6.4$ and *shape* = 1 for event time. If there were no treatment switching, the power at $n = 232$ could reach 90% at a one-sided significance level of 0.005 in this setting.

Next, we examine the changes in powers and required sample sizes when treatment switching occurs with a switch probability of $p_s = 0.89$. We assume *s.dist* = "gamma" or "indepExp" with $r_s = 0.3$ (= 1.96/6.4, the ratio of $E(s) = 1.96$ (the reported mean PFS) to $E(T_1) = 6.4$). For *s.dist* = "gamma", we assume $\rho_s = 0.1$, 0.3, 0.5, 0.7, and 0.9 to model low to high correlations between progression-free survival and overall survival. The resulting powers at $n = 232$ range from 0.775 to 0.833, which are less than 0.9, and the required sample sizes to achieve the power of 0.9 range from 284 to 308 (Table 2).

The second example is a non-inferiority trial involving 1,234 women with early-stage breast cancer who have undergone breast-conserving surgery [13]. This trial compares hypofractionated radiotherapy to standard radiotherapy for preventing local recurrence of invasive breast cancer. Between April 1993 and September 1996, 622 and 612 patients were randomly assigned to hypofractionated radiotherapy and standard radiotherapy, respectively, and were followed up to 12 years (T_a = 3.5, T_e = 12, and r = 1) with 7.9% dropout censoring. Among the patients randomized to hypofractionated radiotherapy, 1.2% selected standard radiotherapy instead (p_s = 0.012, s = 0, and *TXswitch* = "2to1") [9]. Given the assumption of a 7% 5-year local recurrence rate for standard radiotherapy [13], we assume Weibull distributions with $m_1 = m_2 = -5\log(2)/\log(0.93) = 47.8$ and *shape* = 1 for event time, and a dropout censoring rate of 4% (about a half of 7.9%) for standard radiotherapy, i.e., *censoring.rate* = 0.902 (including 86.2% administrative censoring). Based on the hypofractionated radiotherapy is not worse than the standard radiotherapy by 5% in local recurrence-free survival at 5 years, the HR margin is $1/\theta = \log(0.88)/\log(0.93) = 1.762$.

We compare the RMSTs at τ = 5.75 and 10 (corresponding to two analysis times in [13]) between the two radiotherapy groups. The DRMST margins, converted from the HR margin, are 0.169 and 0.484, respectively. With a one-sided significance level of 0.05 and a power of 0.9, the required sample sizes n in the standard radiotherapy group are 550 for τ = 5.75 and 376 for τ = 10.

3.2 Simulation Scenarios

Various simulations are conducted to examine the impact of treatment switching on power and sample size estimation in NI trials using DRMST in ITT analysis. These simulations consider different relative effectiveness of the experimental versus the control group (m_2/m_1 = 1.1 and 0.9), switching times (r_s = 0.5 and 0.25) and switching probabilities (p_s = 0.2 and 0.4), event time distributions (Weibull distributions with *shape* = 1, 0.75, and 1.25), distributions of switching time (*s.dist* = "unif", "beta", "gamma", and "indepExp"), and allocation ratios (r = 1 and 2). For each scenario, we set T_a = 3, T_e = 5, τ = 5, m_1 = 1, ρ_s = 0.775, *censoring.rate* = 0.2, and *n_simulations* = 5000. Also, m_0 = 0.5 and f_2 = 0.5 are used for calculating $R_0(\tau)$ and the DRMST margin, i.e., the margin equals $0.5\left(R_1(\tau) - R_0(\tau)\right)$. When r_s = 0.5, ρ_s = 0.775, and event times follow exponential distributions, the results of assuming *s.dist* = "beta" will be similar to those of

assuming *s.dist* = "unif" because the shape1 and shape2 parameters in the beta distributions are close to 1. The one-sided significance level is set at 0.025.

Effect of relative effectiveness on power and sample size

Treatment switching results in a decrease in power when $m_2 > m_1$ ($m_2/m_1 = 1.1$) and an increase when $m_2 < m_1$ ($m_2/m_1 = 0.9$). Consequently, this corresponds an increase and decrease in the ratio (n/n_{ns}) of sample sizes with treatment switching (n) to those without switching (n_{ns}), respectively (Table 3). For example, at $p_s = 0.2$ and *s.dist* = "unif", the power is 0.776 at $n_{ns} = 158$ and $n/n_{ns} = 1.044$ when $m_2/m_1 = 1.1$, while the power is 0.859 at $n_{ns} = 656$ and $n/n_{ns} = 0.886$ when $m_2/m_1 = 0.9$. Similar changes in powers and sample sizes are observed for other distributions of switching time. The powers decrease to between 0.764 and 0.783 when $m_2/m_1 = 1.1$ and increase to between 0.849 and 0.875 when $m_2/m_1 = 0.9$. The ratios of sample sizes increase to between 1.019 and 1.082 when $m_2/m_1 = 1.1$ and decrease to between 0.849 and 0.886 when $m_2/m_1 = 0.9$.

Effect of switching probability on power and sample size

When p_s increases to 0.4, the magnitude of changes in powers and sample sizes increases. Across four distributions of switching time, when $m_2/m_1 = 1.1$, the powers decrease to a range of 0.739 and 0.757, and when $m_2/m_1 = 0.9$, the power increase to a range of 0.879 and 0.907. The ratios of sample sizes increase to a range of 1.127 and 1.184 when $m_2/m_1 = 1.1$ and decrease to a range of 0.706 and 0.788 when $m_2/m_1 = 0.9$.

Effect of switching time on power and sample size

When r_s is reduced from 0.5 to 0.25, indicating a shorter switching time, the magnitude of changes in powers and sample sizes increases (Table 4). Comparing the results at $p_s = 0.4$ in Table 4 with those above, across the three distributions of switching time, when $m_2/m_1 = 1.1$, the powers decrease to a range of 0.720 and 0.726, and when $m_2/m_1 = 0.9$, the power increase to a range of 0.915 and 0.929. The ratios of sample sizes rise to a range of 1.203 and 1.228 when $m_2/m_1 = 1.1$ and fall to a range of 0.671 and 0.698 when $m_2/m_1 = 0.9$.

We also adjust the shape parameters in Weibull distributions to assess the impact of different event time distributions (Supplementary Figure s1). The changes in powers are similar and there is no

significant trend (Supplementary Tables s1, s2, and Table 3). The ratios of sample sizes slightly increase and decrease with the shape values when m_2/m_1 = 1.1 and 0.9, respectively, except for *s.dist* = "indepExp". However, the required sample sizes vary significantly, decreasing with the shape values. In addition, when we change the allocation ratio from 1 to 2, the change patterns are similar, but more total sample sizes ($n(r+1)$) are needed (Supplementary Table s3).

4 Discussion

Our simulation study shows that switching time and switching probability can decrease or increase power and sample sizes compared to those in the scenarios without treatment switching. A shorter switching time and a higher switching probability amplify the magnitude of these changes. Whether power and sample sizes decrease or increase depends on the relative effectiveness. When m_2/m_1>1, powers decrease and sample sizes increase, while m_2/m_1<1 leads to the opposite result. When $m_1 = m_2$, treatment switching does not impact power and sample sizes. The changes in powers and sample sizes are not sensitive to the choice of the distributions of switch time. To accelerate the computation of sample sizes, we employ a monotonic smoothing technique [16] to model the power trend as sample sizes increase. The powers at the sample size estimated by the power curve exhibit a bias of less than 2% from the expected power.

nifts assumes the effects of the experimental treatment are the same (common treatment effect, made by RPSFTM [18]) for participants initially in the experimental group and those who switch from the control group to the experimental group. This assumption may be problematic, as participants who switch from the control group to the experimental group may have worse survival outcomes. Properly adjusting the accelerated factor m_2/m_1 could help fit the scenario. Multiplying m_2/m_1 by a constant less than 1 might be a solution, but determining this constant value before clinical trials is challenging, even with information from previous similar studies.

In this study, we assume event times follow Weibull distributions rather more flexible distributions like generalized gamma distributions that can fit more real-world scenarios. This is because determining the three parameters for the latter can be challenging for investigators. Besides, median survival times and hazard ratios are still commonly used for power and sample size calculations, so we ultimately choose Weibull distributions that satisfy the proportional hazards assumption. *nifts* will help users calculate the scale and shape parameters required for Weibull

distributions when provided the median survivals of two treatment groups and a survival rate at a specific time in the control group.

5 Conclusions

We propose a simulation-based approach, *nifts,* for power and sample size calculation in NI trials with treatment switching when comparing the RMSTs of two treatment groups in ITT analysis. This approach offers a preview of how treatment switching can influence powers and sample sizes in NI trials, providing investigators with useful information before conducting the trials.

Abbreviations

HR: hazard ratio

PH: proportional hazards

ITT: intention-to-treat

RCT: randomized controlled trials

NI: non-inferiority

RMST: restricted mean survival time

DRMST: difference in restricted mean survival times

RPSFTM: rank preserving structural accelerated failure time models

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Additional file 1: Supplementary Material. *nifts* is freely available at https://github.com/cyhsuTN/nifts.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Institutes of Health [P30 CA068485, U2C CA233291, R01 CA252964, and U54 CA260560, CA163072].

Authors' contributions

AS and CYH were major contributors in writing the manuscript and developing the R package. YS was the major contributor in the conception and design of the work.

Acknowledgements

Not applicable.

References

1. Uno H, Wittes J, Fu H, Solomon SD, Claggett B, Tian L, et al. Alternatives to hazard ratios for comparing the efficacy or safety of therapies in noninferiority studies. *Ann Intern Med*. 2015;163:127–134.
2. Cheng D, Pak K and Wei LJ. Demonstrating noninferiority of accelerated radiotherapy with panitumumab vs standard radiotherapy with cisplatin in locoregionally advanced squamous cell head and neck carcinoma. *JAMA Oncol*. 2017;3:1430–1431.
3. Kim DH, Uno H and Wei LJ. Restricted mean survival time as a measure to interpret clinical trial results. *JAMA Cardiol*. 2017;2:1179–1180.
4. Irwin JO. The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. *Journal of Hygiene* 1949;47:188-189.
5. Royston, P. and Parmar, M.K.B. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med*. 2011;30:2409-2421.
6. Royston, P. and Parmar, M.K.B. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol*. 2013;13:152.
7. Freidlin B, Hu C, Korn EL. Are restricted mean survival time methods especially useful for Noninferiority Trials. *Clin Trials*. 2021;18:188–196.
8. Quartagno M. et al. A comparison of different population-level summary measures for randomised trials with time-to-event outcomes, with a focus on non-inferiority trials. *Clin Trials*. 2023;20:594–602.

9. Weir IR, Trinquart L. Design of non-inferiority randomized trials using the difference in Restricted Mean Survival Times. *Clin Trials*. 2018;15: 499-508.
10. Phadnis MA, Mayo MS. Sample size calculations for noninferiority trials for time-to-event data using the concept of proportional time. *J Appl Stat*. 2020;48:1009-1032.
11. Van Cutsem E, Peeters M, Siena S, Humblet Y, Hendlisz A, Neyns B, et al. Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy refractory metastatic colorectal cancer. *J Clin Oncol*. 2007;25:1658–64.
12. Mo Y, Lim C, Watson JA, White NJ, Cooper BS. Non-adherence in non-inferiority trials: pitfalls and recommendations. *BMJ*. 2020;370:m2215.
13. Whelan TJ, Pignol JP, Levine MN, Julian JA, MacKenzie R, Parpia S, et al. Long-term results of hypofractionated radiation therapy for breast cancer. *N Engl J Med*. 2010;362:513-20.
14. Morden JP, Lambert PC, Latimer N, Abrams KR, Wailoo AJ. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. *BMC Med Res Methodol*. 2011;11:4.
15. Deng L., Hsu C.-Y., Shyr Y. Power and sample sizes estimation in clinical trials with treatment switching in intention-to-treat analysis: a simulation study. *BMC Med Res Methodol.* 2023;23:49.
16. Pya N, Wood SN. Shape constrained additive models. *Stat Comput.* 2015; 25:543–55.
17. Hsu C.-Y., Chen C.-H., Hsu K.-N., Lu Y.-H. A useful design utilizing the information fraction in a group sequential clinical trial with censored survival data. *Biometrics.* 2019;75:133-143.
18. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models, *Communications in Statistics - Theory and Methods*. 1991;20:2609-2631.

Table 1. Five options for the distributions of switching time are provided.

	Options	**Property**
s is correlated with T_1 Assume $s = XT_1$ and X is independent of T_1	$X \sim U(0,1)$	$s < T_1$
	$X \sim \text{Beta}(\text{shape}_1 = a, \text{shape}_2 = b)$ $r_s = a/(a+b)$ and $\rho_s = \left\{\frac{a}{a+b}\sqrt{Var(T_1)}\right\} / \left\{\left(\frac{a}{a+b}\right)^2 Var(T_1) + \frac{ab}{(a+b)^2(a+b+1)} E(T_1^2)\right\}^{1/2}$	$s < T_1$
	$X \sim \text{Gamma}(\text{shape} = a, \text{rate} = b)$ $r_s = a/b$ and $\rho_s = \left\{\frac{a}{b}\sqrt{Var(T_1)}\right\} / \left\{\left(\frac{a}{b}\right)^2 Var(T_1) + \frac{a}{b^2} E(T_1^2)\right\}^{1/2}$	
s is not correlated with T_1	$s \sim \text{Exponential}(\text{rate} = b)$ $b = \left(r_s E(T_1)\right)^{-1}$	
	s is a specific time. e.g., $s = 0$ denotes the switch occurs at the start of the study	

Table 2. Powers and required sample sizes in a NI trial allowing treatment switching with a switch probability of $p_s = 0.89$ when using DRMST in ITT analysis.

	s.dist = "gamma"					*s.dist* = "indepExp"
ρ_s	0.1	0.3	0.5	0.7	0.9	0
Power at a one-sided significance level of 0.005 with $n = 232$ and $r = 1$	0.775	0.807	0.808	0.815	0.808	0.833
Required sample sizes (n) to achieve the power of 0.9 at the one-sided significance level of 0.005	308	299	293	290	292	284

Table 3. Required sample sizes (n) and powers at n_{ns} with r_s = 0.5, shape = 1, and r = 1, where n_{ns} denotes the sample size under no treatment switching, given a power of 0.8 and a one-sided significance level of 0.025. E1 and E2 are the expected number of events in the control and experimental groups.

m_2/m_1 = 1.1 n_{ns} = 158; E1 = 126.4; E2 = 122.5		*s.dist*			
		unif	beta	gamma	indepExp
p_s = 0.2	n	165	171	161	170
	E1	131.6	136.4	128.4	135.4
	E2	127.9	132.6	124.8	131.8
	n/n_{ns}	1.044	1.082	1.019	1.076
	Power at n_{ns}	0.776	0.783	0.775	0.764
	Power at n	0.790	0.813	0.782	0.795
p_s = 0.4	n	178	179	179	187
	E1	141.5	142.3	142.2	148.3
	E2	138.1	138.9	138.8	145.1
	n/n_{ns}	1.127	1.133	1.133	1.184
	Power at n_{ns}	0.757	0.753	0.745	0.739
	Power at n	0.789	0.801	0.802	0.801
m_2/m_1 = 0.9 n_{ns} = 656; E1 = 524.8; E2 = 541.0		*s.dist*			
		unif	beta	gamma	indepExp
p_s = 0.2	n	581	581	578	557
	E1	466.2	466.3	464.0	447.6
	E2	479.4	479.3	476.8	459.5
	n/n_{ns}	0.886	0.886	0.881	0.849
	Power at n_{ns}	0.859	0.851	0.849	0.875
	Power at n	0.807	0.801	0.810	0.799
p_s = 0.4	n	513	517	508	463
	E1	412.9	416.2	409.0	373.7
	E2	423.2	426.5	419.1	381.9
	n/n_{ns}	0.782	0.788	0.774	0.706
	Power at n_{ns}	0.879	0.885	0.899	0.907
	Power at n	0.806	0.805	0.796	0.782

Table 4. Required sample sizes (n) and powers at n_{ns} with r_s = 0.25, *shape* = 1 and r = 1, where n_{ns} denotes the sample size under no treatment switching, given a power of 0.8 and a one-sided significance level of 0.025. E1 and E2 are the expected number of events in the control and experimental groups.

m_2/m_1 = 1.1 n_{ns} = 158; E1 = 126.4; E2 = 122.5		*s.dist*			
		unif	beta	gamma	indepExp
p_s = 0.2	n	-	178	167	172
	E1	-	141.7	133.1	136.9
	E2	-	138.2	129.4	133.4
	n/n_{ns}	-	1.127	1.057	1.089
	Power at n_{ns}	-	0.763	0.764	0.759
	Power at n	-	0.804	0.788	0.802
p_s = 0.4	n	-	193	190	194
	E1	-	152.9	150.6	153.6
	E2	-	149.8	147.3	150.5
	n/n_{ns}	-	1.222	1.203	1.228
	Power at n_{ns}	-	0.720	0.722	0.722
	Power at n	-	0.800	0.806	0.797
m_2/m_1 = 0.9 n_{ns} = 656; E1 = 524.8; E2 = 541.0		*s.dist*			
		unif	beta	gamma	indepExp
p_s = 0.2	n	-	546	539	536
	E1	-	438.8	433.3	431.1
	E2	-	450.4	444.6	442.2
	n/n_{ns}	-	0.832	0.822	0.817
	Power at n_{ns}	-	0.860	0.865	0.878
	Power at n	-	0.789	0.807	0.809
p_s = 0.4	n	-	458	453	440
	E1	-	370.0	365.9	355.7
	E2	-	377.8	373.7	362.9
	n/n_{ns}	-	0.698	0.691	0.671
	Power at n_{ns}	-	0.915	0.929	0.924
	Power at n	-	0.815	0.800	0.788

- *s.dist* = "unif" does not satisfy r_s = 0.25.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- niftsSupplementaryF.docx