

1 **Power and sample sizes estimation in clinical trials with treatment switching in**
2 **intention-to-treat analysis: a simulation study**

3

4 Lejun Deng^{1#}, Chih-Yuan Hsu^{2,3#}, Yu Shyr^{2,3*}

5

6 ¹Montgomery Bell Academy, Nashville, TN, 37205, USA

7 ²Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN 37232, USA

8 ³Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN 37232,

9 USA

10 *corresponding authors: Yu Shyr (yu.shyr@vumc.org)

11 # These authors contributed equally.

12

13

14 **Abstract**

15 **Background:** Treatment switching, also called crossover, is common in clinical trials because of
16 ethical concerns or other reasons. When it occurs and the primary objective is to identify treatment
17 effects, the most widely used intention-to-treat analysis may lead to underpowered trials.
18 Therefore, we plan to propose an approach to preview power reductions and to estimate sample
19 sizes required for the desired power when treatment switching occurs in the intention-to-treat
20 analysis.

21 **Methods:** We proposed a simulation-based approach and developed an R package to attain the
22 purpose.

23 **Results:** We simulated a number of randomized trials incorporating treatment switching and
24 investigated the impact of the relative effectiveness of the experimental treatment to the control,
25 the switching probability, the allocation ratio, and the switching time on power reductions and
26 sample sizes estimation. The switching probability and the switching time are key determinants
27 for significant power decreasing and thus sample sizes surging to maintain the desired power. The
28 sample sizes required in randomized trials absence of treatment switching varied from around
29 three-fourths to one-twelfths of the sample sizes required in randomized trials allowing treatment
30 switching as the probability of switched patients increased. The power reductions and sample sizes
31 increase with the decrease of switching time.

32 **Conclusions:** The simulation-based approach not only provides a preview for power declining but
33 also calculates the required sample size to achieve an expected power in the intention-to-treat
34 analysis when treatment switching occurs. It will provide researchers and clinicians with useful
35 information before randomized controlled trials are conducted.

36 37 **Keywords**

38 Crossover; intention-to-treat analysis; randomized controlled trials; treatment switching

39 40 **1. Background**

41 Clinical trials are often used to assess the effectiveness of new treatments. The most reliable
42 clinical trial is the randomized controlled trials (RCT). In an RCT, patients are randomly assigned
43 to the control group, where they receive a placebo or an existing standard treatment, or the
44 experimental group, where they receive the experimental treatment. For treatments against
45 diseases like cancer, the effectiveness of the new treatment is determined by comparing the

46 observed survival time of patients in the experimental group to those in the control group. The
47 power of a statistical test, which is the probability to correctly tell if the experimental treatment is
48 better when it is truly better, is extremely important. Being able to correctly determine the
49 effectiveness of an experimental treatment gives justification for its production, allows for better
50 utilization of resources, and helps those who will benefit from the experimental treatment.

51
52 A common method used to interpret the data from a RCT is the intention-to-treat analysis (ITT).
53 The ITT analysis includes all patients with randomization in statistical analysis and compares their
54 responses to determine the effectiveness of the experimental treatment according to the treatment
55 group that was initially assigned to them, regardless of what treatment they actually received.
56 Patients, however, do not always comply with the treatment they were assigned. Because of ethical
57 concerns or other reasons, they may switch treatments from the control group to the experimental
58 group. For example, it may occur when a disease progresses or when the healthcare provider
59 believes the patient's prognosis will improve with the experimental treatment. When treatment
60 switching, also called crossover, is permitted, the ITT analysis is confounded, which decreases the
61 power of a statistical test. A simple and alternative approach, such as the per-protocol analysis,
62 excludes the subjects who switch treatments from the statistical analysis. Nevertheless, this may
63 heavily bias the analysis results if the subjects included and excluded in the analysis significantly
64 differ in prognosis, i.e., treatment switching is associated with prognostic variables [1].

65
66 Various statistical methods dealing with treatment switching have been proposed. Law and Kaldor
67 [2] proposed the adjusted Cox model by splitting the study population into four groups according
68 to their initial and final treatment groups and by assuming different hazard functions with time

69 dependent covariates defined by the time of switching treatment for the four group. Loeys and
70 Goetghebeur [3] proposed the causal proportional hazards estimator by assuming that patients in
71 one arm complete their treatment while the patients in the other arm either completely fulfill their
72 treatment or does not fulfill their treatment at all. Robins and Tsiatis [4] proposed correcting for
73 non-compliance by using the rank preserving structural accelerated failure time models (RPSFT).
74 Based on the RPSFT, Branson and Whitehead [5] and Zhang and Chen [6] provided iterative
75 parametric estimation and modified iterative parametric estimation methods, respectively, for fast
76 and reliably estimating the treatment effect. Morden et al. [1] and Latimer et al. [7, 8] performed
77 simulations studies to compare several adjustment methods mentioned above.

78
79 However, sample sizes estimation in most RCTs is still based on the assumption of no treatment
80 switching and predetermined statistical tests, such as the logrank test in ITT analysis. Given the
81 sample sizes estimated from no-switching designs, the ITT analysis may underestimate the positive
82 treatment effect when treatment switching occurs. In this study, we proposed a simulation-based
83 approach, *PowerSwitchingTrial*, to preview power reductions and calculate sample sizes required
84 in RCTs allowing treatment switching. We investigated the impact of the relative effectiveness of
85 the experimental treatment to the control, the switching probability, the allocation ratio, and the
86 switching time on power reductions and sample sizes estimation. We found the switching
87 probability and the switching time are key determinants for power decreasing and thus sample
88 sizes increasing to maintain the desired power in ITT analysis. *PowerSwitchingTrial* is freely
89 available at <https://github.com/darwin-hub/PowerSwitchingTrial>.

90

91

92 **2. Methods**

93 **2.1 *PowerSwitchingTrial* overview**

94 *PowerSwitchingTrial* is a simulation-based R package, developed for exploring statistical powers
95 and sample sizes required in RCT trials allowing treatment switching when the logrank test is used
96 in ITT analysis. The flowchart of *PowerSwitchingTrial* is shown in Fig. 1. *PowerSwitchingTrial*
97 simulates a two-arm randomized clinical trial study with an accrual time of T_a and an additional
98 follow-up time of $T_e - T_a$, where T_e is the time from the start to the end of the trial. Participants
99 are assumed to enter the study uniformly during the accrual time period, and soon be randomly
100 assigned to one of the two treatment groups, the control group (TX_1) or the experimental group
101 (TX_2), with an allocation ratio of r ($r = TX_2/TX_1$). The survival time of subjects in the control and
102 experimental groups, denoted by T_1 and T_2 , is assumed to follow exponential distributions with
103 the median value of m_1 and m_2 , respectively. The censoring time of subjects in the two groups,
104 denoted by C_1 and C_2 , is defined as the time from randomization until the end of the trial if the
105 subjects do not experience the event of interest.

106
107 *PowerSwitchingTrial* allows subjects in the control group have a chance to switch to the
108 experimental group if predetermined conditions are satisfied. For example, if patients with cancer
109 are observed to have a disease progression before death (assume death is the event of interest),
110 they may switch from the standard treatment to the new treatment after disease progression. The
111 switching time, denoted by s , is defined as the time from randomization until treatment switching.
112 s is fixed or random, which depends on a practical application. For example, s will be fixed if
113 there is a time set in a clinical trial to switch treatments, while s will be random if the switching
114 time depends on disease progression. In the setting of random s , we assume $s = X \times T_1$, where

115 X is a random variable generated from a beta distribution with the parameters of a and b , namely
116 $\text{Beta}(a,b)$. The two parameters control the relationship between s and T_1 . Moreover,
117 *PowerSwitchingTrial* considers a parameter of p (switching probability) to define the probability
118 that a subject who qualifies for treatment switching will switch from TX_1 to TX_2 after evaluations
119 from physicians or health care professionals. The survival time for subjects starting from switching
120 is assumed to increase by a constant multiplier of A , based on the rank preserving structural failure
121 time model [4]. That is, the survival time of the subjects with treatment switching will be $T_1^* =$
122 $s + (T_1 - s) \times A$. Therefore, the observable survival time Y_1 for the subjects who are
123 originally assigned to TX_1 will be equal to $\min(T_1, C_1)$ without treatment switching, and
124 $\min(T_1^*, C_1)$ with treatment switching. The observable survival time Y_2 for the subjects in TX_2
125 will be equal to $\min(T_2, C_2)$. Finally, the logrank test is used to compare Y_1 and Y_2 .

126

127 **2.2 Two main functions of *PowerSwitchingTrial***

128 *PowerSwitchingTrial* provides two main functions, *LogRankTestMix2PowerMedian* and
129 *LogRankTestMix2Nmedian*. *LogRankTestMix2PowerMedian* calculates the power, given sample
130 sizes, the difference in the survival time between two groups, and the other parameters.
131 *LogRankTestMix2NMedian* estimates sample sizes, given an expected power to detect the
132 difference in the survival time between two groups, and the other required parameters.

133

134 *LogRankTestMix2PowerMedian*, a function to calculate the power, includes 13 input parameters
135 to simulate a number of scenarios, n , m_1 , m_2 , r , T_a , T_e , p , α , random , s , a , b , and reps . n is
136 the number of participants in the control group, m_1 , m_2 , r , T_a , T_e , and the switching probability
137 (p) were defined in Section 2.1, α is the significance level, and reps is the number of computer

138 simulations. The *random* is a logic indicator, FALSE or TRUE. If *random* = FALSE, the switching
139 time is fixed and *s* is required; if *random* = TRUE, the switching time is random and the
140 parameters of *a* and *b* in the beta distribution are required. The function will return the power, that
141 is, the average number of rejecting the null hypothesis which suggests no difference in survival
142 time between the two treatment groups across *reps* simulations.

143
144 *LogRankTestMix2Nmedian* is a function to calculate the sample sizes. *LogRankTestMix2NMedian*
145 uses the same parameters as *LogRankTestMix2PowerMedian*, except the parameter of *n* that is
146 replaced by *power*, *lower*, and *upper*. Lower and upper is the minimum and maximum sample
147 sizes that users set to explore. The function uses the bisection method to find the sample size (*n*)
148 required between the initial *lower* and *upper* bounds.

149

150

151 **3. Results**

152 A number of scenarios were generated to explore the effect of treatment switching on powers and
153 sample sizes estimation in ITT analysis, which included the relative different effectiveness of the
154 experimental treatment to the control, the different switching probability, the different allocation
155 ratio, and the fixed/random switching time. In each scenario, the accrual time T_a was set to 3, the
156 study time T_e was set to 5, and m_1 was set to 1. Under no treatment switching, the expected power
157 to detect the survival difference between two treatment groups was set as 0.8 and the significance
158 level was set as 0.05. The number of computer simulations was set at 5,000.

159

160 **3.1 Effect of the relative effectiveness on power and sample sizes estimation**

161 We first investigated the impact of the relative effectiveness of the experimental treatment to the
162 control (m_2/m_1) on power and sample sizes estimation, which ranged at 1.25, 1.5, 1.75, and 2
163 (Supplementary Tables s1-s3). We found the power decreased and the ratio of sample sizes with
164 treatment switching to those absence of switching (n_s/n_{no-s}) increased slightly with the rising
165 relative effectiveness (Fig. 2). For example, given $r = 1$, $p = 0.4$, and $s = 0.5$, the ratio
166 increased from 1.94 to 2.07 and the power decreased from 0.52 to 0.50 when the relative
167 effectiveness grew from 1.25 to 2 (Supplementary Table s1).

168

169 **3.2 Effect of the switching probability on power and sample sizes estimation**

170 We further evaluated the effect the switching probability (p) on power and sample sizes
171 estimation, which was set at 0.2, 0.4, 0.6, 0.8, and 1.0. We found the power decreased and the ratio
172 of sample sizes (n_s/n_{no-s}) increased significantly with the enlarging switching probability (Fig.
173 3). For example, the ratio of sample sizes increased from ~ 1.4 to ~ 11.6 and the power decreased
174 from ~ 0.7 to ~ 0.1 When the switching probability increased from 0.2 to 1.0 (Supplementary Table
175 s1). The result indicates that the switching probability is critical for power and sample sizes
176 estimation. A small change of the switching probability would require much more sample sizes to
177 maintain the expected power.

178

179 **3.3 Effect of the allocation ratio on power and sample sizes estimation**

180 Moreover, we examined the effect of the allocation ratio (r) on power and sample sizes estimation,
181 which was set at 1 and 2. The sample size required for the control group decreased as the allocation
182 ratio increased from 1 to 2, the total sample size increased though. The ratio of sample sizes
183 (n_s/n_{no-s}) increased slightly but the power didn't show a clear increasing or decreasing pattern.

184 For example, under $s = 0.5$, given $p = 0.2$, the power with $r = 1$ was larger than that with
185 $r = 2$. In contrast, the power with $r = 1$ was smaller than that with $r = 2$ given $p = 1$
186 (Supplementary Table s1).

187

188 **3.4 Effect of the switching time on power and sample sizes estimation**

189 Finally, we explored the effect of the switching time on power and sample sizes estimation. The
190 switching time (s) was either set at a fixed value or generated by random. The switching time was
191 set at 0.5 and 1 if fixed, otherwise the switching time was generated by a product of survival time
192 of the control group and a random variable from a beta distribution with shape parameters of $a =$
193 2 and $b = 2$. The mean of the random switching time was 0.5. Overall, the ratios of sample sizes
194 (n_s/n_{no-s}) using random switching time were smaller than those using the fixed switching time
195 of 0.5 (Fig. 4 and Supplementary Tables s1 and s2). The ratio of sample sizes (n_s/n_{no-s}) increased
196 and power decreased significantly when the fixed switching time moved to an earlier time point
197 (from 1 to 0.5), especially when the switching probability p is greater than 0.6 (Fig.4 and
198 Supplementary Table s1 and s3). For example, the ratio of sample sizes decreased from ~ 11 to ~ 3
199 when the switching time moved from 0.5 to 1 under $p = 1$, $r = 1$, and $m_2/m_1 = 1.5$ (Fig. 4).
200 The results indicated that the switching time is also a key factor for power reductions and sample
201 sizes increase in clinical trials with treatment switching.

202

203 **4. Discussion**

204 Our simulation study showed a standard ITT analysis failing to consider treatment switching
205 resulted in a significant reduction of statistical powers especially when the switching probability
206 is high and the switching time is early, which indicated that a much larger sample size was required

207 to maintain the expected power. The sample sizes required in RCTs absence of treatment switching
208 varied from around three-fourths to one-twelfths of the sample sizes required in RCTs allowing
209 treatment switching as the probability of switched patients increased. The growth of the relative
210 effectiveness and the switching probability increased the ratio of samples and decreased the power,
211 while the later switching time decreased the ratio of sample sizes and increased the power. The
212 allocation ratio was slightly associated with the increasing sample sizes but was not associated
213 with the decreasing power.

214
215 *PowerSwitchingTrial* assumed the effects of the experimental treatment were the same for the
216 subjects switched from TX_1 to TX_2 and those initially assigned to TX_2 , which is the same as the
217 “common treatment effect” assumption made by RPSFTM [8]. In some cases, the assumption may
218 be problematic. The subjects switched from TX_1 to TX_2 may be associated with worse survival.
219 Properly adjusting the accelerated factor A may be useful to fit the scenario. Multiplying a constant
220 that is less than 1 to A may be a solution, but it is difficult to determine the constant value before
221 clinical trials even if we can borrow the information from previous similar studies.

222
223 The assumption of exponential distributions for survival time also limits the flexibility and practice
224 of *PowerSwitchingTrial*. Survival times do not always follow exponential distributions well.
225 Weibull or gamma distribution may be alternatives to exponential distributions. However, it is not
226 easy for clinicians to determine a non-exponential distribution with at least two parameters.
227 Median survival or mean survival is still commonly used, which leads to favoring of exponential
228 distributions. In addition, *PowerSwitchingTrial* assumed the censoring time depended only on the
229 entry time and study time although other factors that would cause censoring include a small

230 proportion of patients dropping out from the study and a loss of follow up. This was due to no need
231 for additionally assuming a probability distribution for the censoring time.

232

233 The current version of *PowerSwitchingTrial* is built for two-arm RCTs. It can be extended to multi-
234 arm RCTs, straightforwardly but not simply. More parameters need to be determined in multi-arm
235 clinical trial designs. We plan to extend *PowerSwitchingTrial* for multi-arm RCTs in the near
236 future.

237

238 **5. Conclusions**

239 In this study, we proposed a simulation-based approach, *PowerSwitchingTrial*, to evaluate
240 statistical powers and sample sizes required in RCTs allowing switching in ITT analysis when the
241 logrank test is used. The approach not only provides a preview for power declining but also
242 calculates the required sample size to achieve an expected power when treatment switching occurs.
243 It will provide researchers and clinicians with useful information before RCTs are conducted.

244

245

246 **Abbreviations**

247 **ITT:** intention-to-treat

248 **RCT:** randomized controlled trials

249 **RPSFT:** rank preserving structural accelerated failure time models

250

251 **Declarations**

252 **Ethics approval and consent to participate**

253 Not applicable.

254 **Consent for publication**

255 Not applicable.

256 **Availability of data and materials**

257 Additional file 1: Supplementary Tables s1-s3.

258 **Competing interests**

259 The authors declare that they have no competing interests.

260 **Funding**

261 This work was supported by the National Institutes of Health [P30CA068485, U24CA163056,
262 U24CA213274, P50CA236733, P50CA098131, U54CA163072].

263 **Authors' contributions**

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

266 **Acknowledgements**

267 Not applicable.

268

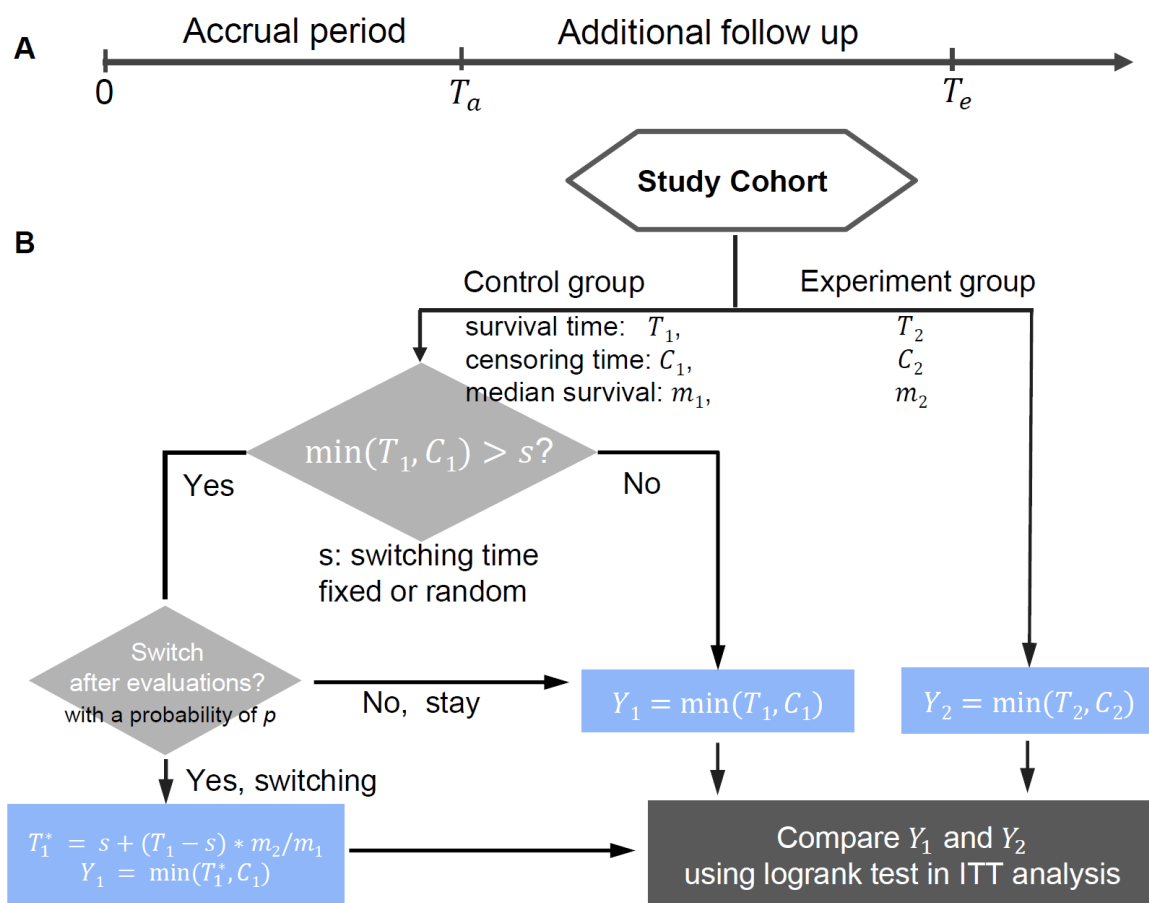
269 **References**

- 270 1. Morden, JP, Lambert PC, Latimer N, Abrams, KR, Wailoo, AJ. Assessing methods for dealing
271 with treatment switching in randomised controlled trials: a simulation study. BMC Medical
272 Research Methodology. 2011; 11, 4.
- 273 2. Law, MG, Kaldor, JM. Survival analyses of randomized clinical trials adjusted for patients
274 who switch treatments. Statistics in Medicine. 1996; 15: 2069-2076.
- 275 3. Loeys, T, Goetghebeur, E. A Causal Proportional Hazards Estimator for the Effect of
276 Treatment Actually Received in a Randomized Trial with All-or-Nothing Compliance.
277 Biometrics. 2003; 59: 100-105.
- 278 4. Robins, JM, Tsiatis, AA. Correcting for non-compliance in randomized trials using rank
279 preserving structural failure time models, Communications in Statistics - Theory and Methods.
280 1991; 20, 2609-2631.
- 281 5. Branson M, Whitehead J. Estimating a treatment effect in survival studies in which patients
282 switch treatment, Statistics in Medicine. 2002; 21, 2449-2463.
- 283 6. Zhang J, Chen, C. Correcting treatment effect for treatment switching in randomized oncology
284 trials with a modified iterative parametric estimation method. Statistics in Medicine. 2016; 35,
285 3690-3703.
- 286 7. Latimer, NR, Abrams, KR, Lambert, PC, Crowther, MJ, Wailoo, AJ, Morden, JP, et al.
287 Adjusting for treatment switching in randomised controlled trials – a simulation study and a
288 simplified two-stage method. Statistical Methods in Medical Research. 2017; 26, 724-751.
- 289 8. Latimer, NR, Abrams, KR, Lambert, PC, Morden, JP, Crowther, MJ. Assessing methods for
290 dealing with treatment switching in clinical trials: A follow-up simulation study. Statistical
291 Methods in Medical Research. 2018; 27, 765-784.

292

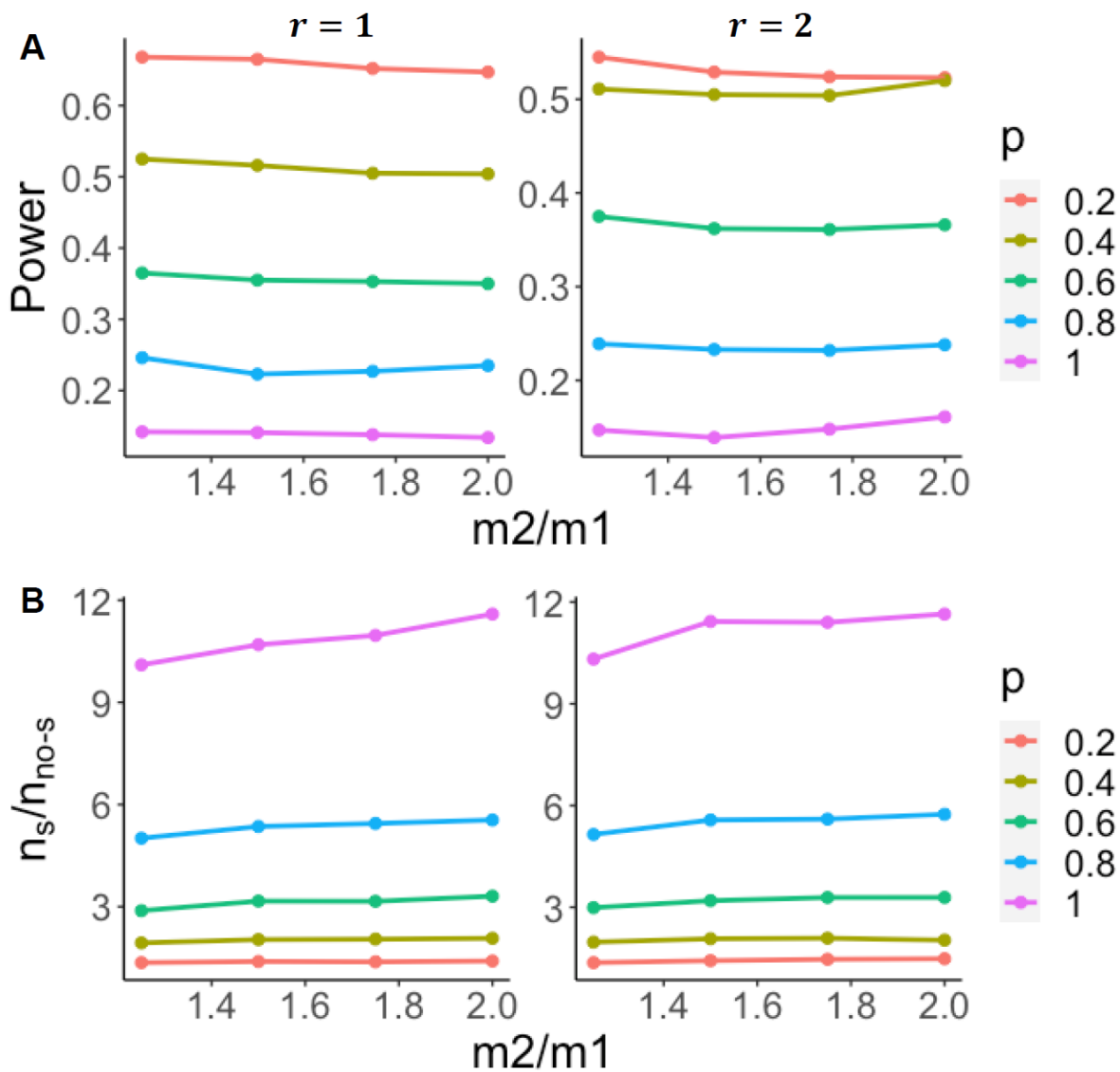
293

294



295

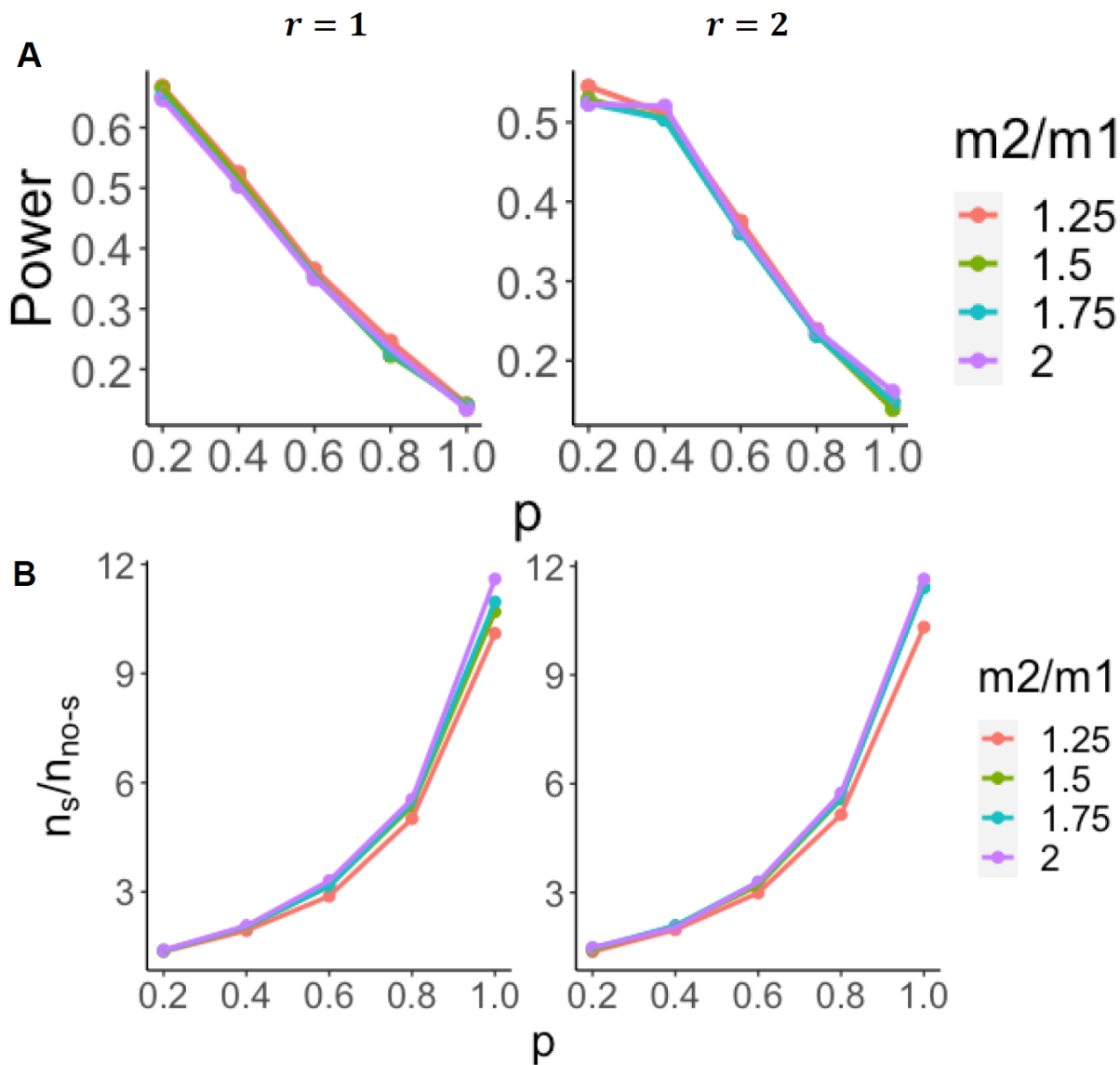
296 **Figure 1. A flowchart of *PowerSwitchingTrial*.**



297

298 **Figure 2. Effect of relative effectiveness on power (A) and sample sizes estimation (B). *s* =**

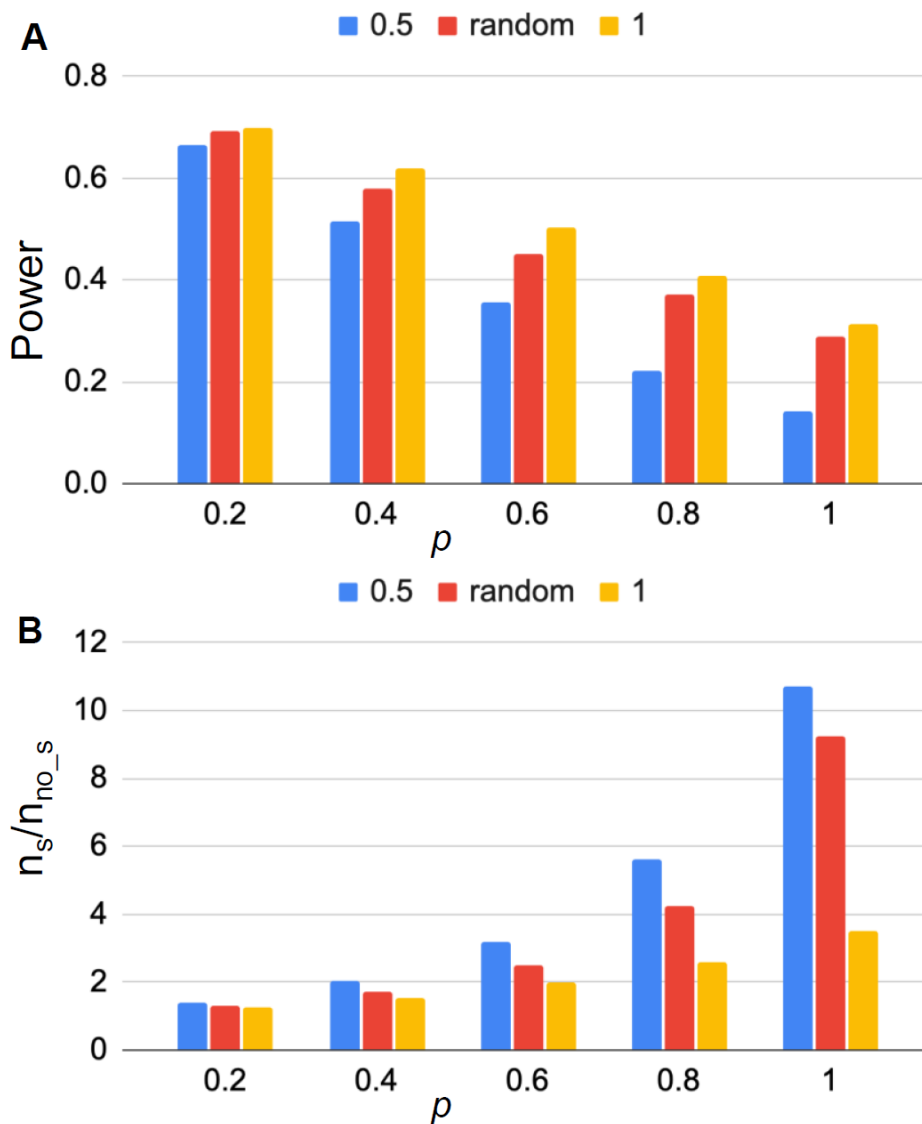
299 0.5.



300

301 **Figure 3. Effect of switching probability on power (A) and sample sizes estimation (B). $s =$**

302 0.5.



303

304 **Figure 4. Effect of switching time on power (A) and sample sizes estimation (B). $s = 0.5$**

305 (blue), $s = 1$ (orange) or s is random based on a beta distribution with mean value of 0.5 (red)

306 under $r = 1$ and $m_2/m_1 = 1.5$.

307