J. Jpn. Soc. Comp. Statist., **30.2**(2018), 65–74 DOI:10.5183/jjscs.1707001_236

THE PERFORMANCE OF RANDOMIZATION METHODS IN CONSIDERATION OF PROGNOSTIC FACTORS FOR SMALL-SIZE CLINICAL TRIALS: A SIMULATION STUDY

Kanae Takahashi* **and Kouji Yamamoto***

ABSTRACT

The performance of randomization methods in consideration of the impact of a prognostic factor that has an interaction and baseline characteristics that have no effect on the outcome has not been clarified, especially for small sized clinical trials. We conducted numerical simulations to identify the difference in behaviour of the empirical power and the empirical type 1 error rate among some randomization methods and statistical analyses when we use a prognostic factor that has an interaction or baseline characteristics that have no effect on the outcome for small sized randomized controlled trials. The empirical power was higher when using a prognostic factor that had an interaction. Also, by using stratified blocked randomization (ST) or minimization (MI) with the multiple regression, the empirical power was further increased. On the other hand, the empirical power was lower when using baseline characteristics that had no effect on the outcome. We recommend conducting ST or MI, multiple regression and using a prognostic factor that has an interaction in small-size randomized controlled trials.

1. Introduction

Clinical trials are the most definitive method of determining whether an intervention has the postulated effect (Friedman *et al.*, 2015). An important component of clinical trials is randomization, which is a technique used for allocation of patients to either the experimental treatment(s) group or the control group. Randomization promotes comparability among the study groups with respect to not only known covariates but also unknown important covariates, and the act of randomization provides a probabilistic basis for an inference from the observed results when considered in reference to all possible results (Rosenberger and Lachin, 2016), and it provides a precise and unbiased estimate of the intervention's effect.

There are several methods of randomization. Complete randomization (CP), permuted block design (PB), Stratified blocked randomization (ST), and Minimization (MI) are relatively commonly used. In the case that the target sample size in each group is established but the final sample size is not known with certainty, the randomization procedure is complete randomization (CP), analogous to tossing a fair coin. Here the sample size in each group is a binomially distributed random variable (Lachin, 1988). It can be executed easily, but a severe imbalance in numbers and baseline characteristics between treatment and control groups may occur by chance, especially in small clinical trials.

Permuted block design (PB) involves randomizing patients to treatment groups in sequential blocks. In the simplest case of constant block size, there are two treatments and

*Department of Medical Statistics, Graduate School of Medicine, Osaka City University, Osaka, Japan
Key words: minimization; permuted block design; simulation study; small sized clinical trial; stratified blocked randomization

m patients per treatment within each block of size $2m$. Furthermore, there are B blocks with a total sample size $N{=}2mB$, provided that all blocks are filled, in which case the total numbers assigned to each treatment are equal, Bm (Matts and Lachin, 1988). PB can easily be executed and each sample size for the treatment(s) group and control group is well-balanced, even in a small sized clinical trial. However, PB does not always ensure the balance of prognostic factors between groups.

Stratified randomization is a two-stage procedure in which patients who were enrolled in a clinical trial are first grouped into strata according to prognostic factors. Within each stratum, patients are then assigned to the experimental treatment(s) group or control group (Kernan *et al.*, 1999). Stratified "blocked" randomization (ST) is one of the stratified randomization methods that uses PB within each stratum. ST is useful to avoid the imbalance of prognostic factors, but if the study sample size is small and there are many prognostic factors used for stratification, most strata will have very few patients and a critical imbalance in numbers and prognostic factors between groups may occur.

Minimization (MI) (Taves, 1974; Pocock and Simon, 1975) is one of the covariate-adaptive randomization methods. For this method, a patient is allocated to whichever group minimizes the total imbalance of all prognostic factors of patients who have already been recruited to all of the groups. Many variables can be taken into consideration for a small sized clinical trial because MI focuses on the total imbalance of all prognostic factors. However, there are some issues with using MI (Scott *et al.*, 2002). For example, a constantly updated centralized system is required because the allocation of each new patient entering the trial depends on the details of the previous patients entered being kept up to date, and the use of many prognostic factors and extensions to the standard method (such as the use of weighting factors with different probabilities) can be costly.

The selection of a randomization method is important at the study design stage. There have been several researches that evaluated the performance of these methods (Therneau, 1993; Weir and Lees, 2003; Hagino *et al.*, 2004; Brown *et al.*, 2005; Toorawa *et al.*, 2009; Xiao *et al.*, 2011; Kahan and Morris, 2013). The performance of randomization methods may be related to trial sample size, but there are only a few studies of this focused on small-size clinical trials. Furthermore, the impact of an interaction of the treatment and a prognostic factor for randomization and statistical analysis was not considered so much in previous research. In addition, for early phase clinical trials, baseline characteristics that have no effect on the outcome may be used as prognostic factors because there may be few existing studies that could be helpful for selecting prognostic factors. Thus, the objective of this paper is to compare the performance of the existing randomization methods when using a prognostic factor that has an interaction or baseline characteristics that have no effect on the outcome by simulations and to suggest the recommended randomization method for small-size randomized controlled trials. For simplicity of description, this paper has only considered the comparison of two treatments, although the principles may be easily extended to more than two treatments.

2. Methods

We performed a simulation study to examine the difference in behaviour of the empirical power and the empirical type 1 error rate among some randomization methods and some statistical analyses when we use a prognostic factor that has an interaction or baseline characteristics that have no effect on the outcome.

We considered a small-size randomized controlled trial in which patients were allocated

to either the treatment group or the control group. The total sample size N was assumed to be 30, 40 and 50. We generated simulation data by using the following model:

$$Y_i = \mu + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_{61} X_{61i} + \beta_{62} X_{62i} + \beta_7 X_{7i} + \beta_8 X_{1i} X_{7i} + \varepsilon_i$$

where Y_i is the continuous outcome from the ith patient, X_{1i} is a binomial variable that represents the group, and $X_{2i}, \ldots, X_{7i}$ are baseline characteristics that have effects on the outcome. X_{2i}, X_{3i}, X_{4i} and X_{7i} are binomial variables distributed as Bernoulli distributions. X_{5i} is a continuous variable distributed as a normal distribution with mean=5, variance=5. X_{6i} is a categorical variable distributed as a multinomial distribution with three categories. The probability of each category is $p_1 = p_2 = p_3 = 1/3$. X_{61i} and X_{62i} are dummy variables. $X_{61i} = 1$ if $X_{6i} = 1$, and otherwise equals zero; $X_{62i} = 1$ if $X_{6i} = 2$, and otherwise equals zero. In order to consider the effect of unknown prognostic factors, we defined X_{5i} and X_{6i} as unknown prognostic factors that are not used in randomization and statistical analysis. Apart from them, Z_i is a baseline characteristic that has no effect on the outcome. All prognostic factors and Z_i are independent. ε_i is a random variable that has mean=0 and variance=1. We set the coefficients as $\beta_2 = \beta_3 = \beta_4 = \beta_{61} = \beta_7 = 10$, $\beta_5 = 1$ and $\beta_{62} = 15$. The treatment effect is set as $\beta_1 = 10$ for sample size $N = 30$, $\beta_1 = 8$ for N=40 and $\beta_1 = 7$ for N=50. We determined β_1 so that the power of Student's t test with CP is about 80%. The interaction effect β_8 is set to 6.

To compare the performance of randomization methods described above, we generated X_{1i} by using one of four randomization methods: CP, PB, ST and MI. The number of baseline characteristics considered in ST and MI was $N/15$ (Harrell, 2001, pp. 53–85). The prognostic factors used in ST or MI are shown in Table 1. In PB and ST, the block size was set to four. In MI, the probabilities for allocation was set to 0.80 because some articles have recommended an allocation probability of about 0.80 (Brown *et al.*, 2005; Toorawa *et al.*, 2009). The weight of all baseline characteristics was defined as one to simplify the result. We generated 1,000,000 data sets of 30, 40 and 50 patients for each randomization scenario.

Table 1: Variables used in randomization and regression analysis

Sample size	Variables used in ST, MI	Variables used in regression analysis	Scenario
N=30, 40	X_3 and X_2	X_1 and X_3	A
	X_3 and X_7	X_1 and X_3	B
		X_1 and X_7	C
	X_3 and Z	X_1 and X_3	D
		X_1 and Z	E
N=50	X_4, X_3 and X_2	X_1, X_4 and X_3	A
	X_4, X_3 and X_7	X_1, X_4 and X_3	B
		X_1, X_4 and X_7	C
	X_4, X_3 and Z	X_1, X_4 and X_3	D
		X_1, X_4 and Z	E

To assess the difference in the performance of randomization methods between statistical analysis techniques, we conducted three analyses: Student's t-test, permutation test, and multiple regression which adjusts the effect of prognostic factors in a regression model. We defined the number of independent variables used in multiple regression as $N/15$ in the simulations (Harrell, 2001, pp. 53–85), and those are shown in Table 1. The empirical power

and the empirical type 1 error rate are defined as performance measures. The empirical power is calculated as the percentage of significant results out of 1,000,000 samples given an effect size of 10, 8, and 7, and the empirical type 1 error rate is such a percentage given an effect size of 0.

3. Results

3.1. Using a prognostic factor that has an interaction

First, we consider scenario A, B and C. Figure 1 and 2 show the empirical power and the empirical type 1 error rate when using prognostic factors in randomization and statistical analysis (scenario A), using a prognostic factor that has an interaction in randomization (scenario B) or using a prognostic factor that has an interaction in randomization and statistical analysis (scenario C). For Student's t-test and permutation test, the results of scenario B and C are summarized because variables used in randomization for these scenarios are the same. Figure 1 shows that the empirical power of ST and MI are higher than CP and PB regardless of the sample size, prognostic factors used in randomization and statistical analysis methods. The empirical power increases when the prognostic factor that has an interaction is used in randomization regardless of statistical analysis methods. There was no big difference in the empirical power between ST and MI. Regarding the type 1 error rate, Figure 2 shows that CP and PB gave a valid empirical type 1 error rate of 0.05 regardless of the sample size, variables used in randomization and statistical analysis methods. On the other hand, ST and MI gave a lower type 1 error rate than the nominal type 1 error rate regardless of the sample size, variables used in randomization and statistical analysis methods.

Figure 3 and 4 show the empirical power and the empirical type 1 error rate for scenario A, B and C. Figure 3 shows that the empirical power is further increased when the prognostic factor that has an interaction is used in randomization and statistical analysis. There was no difference in the empirical type 1 error rate in these three scenarios.

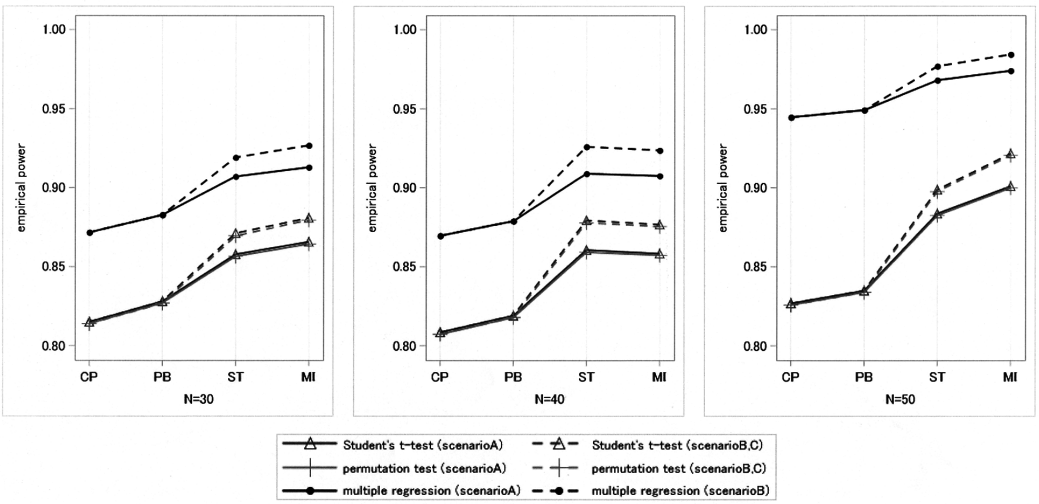

Fig. 1: Empirical power: when a prognostic factor that has an interaction is used in randomization

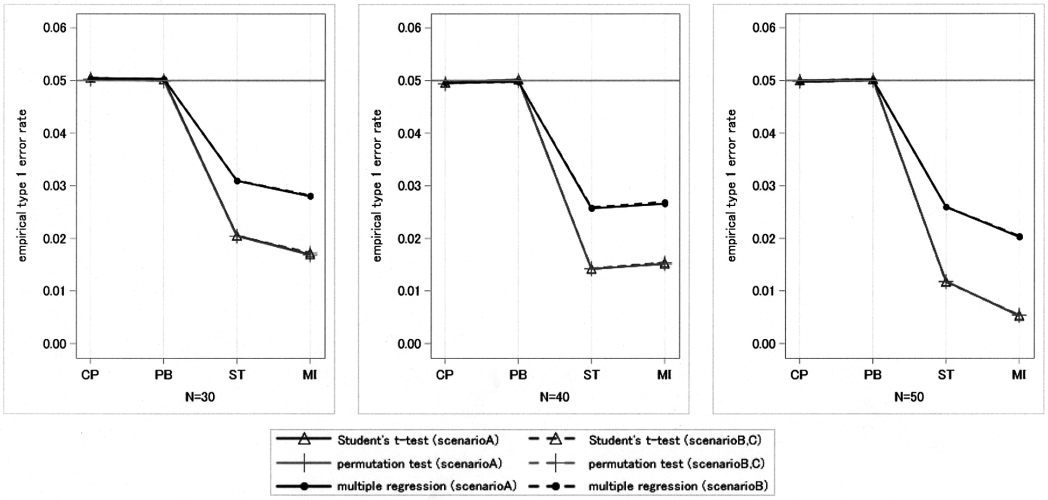

Fig. 2: Empirical type 1 error rate: when a prognostic factor that has an interaction is used in randomization

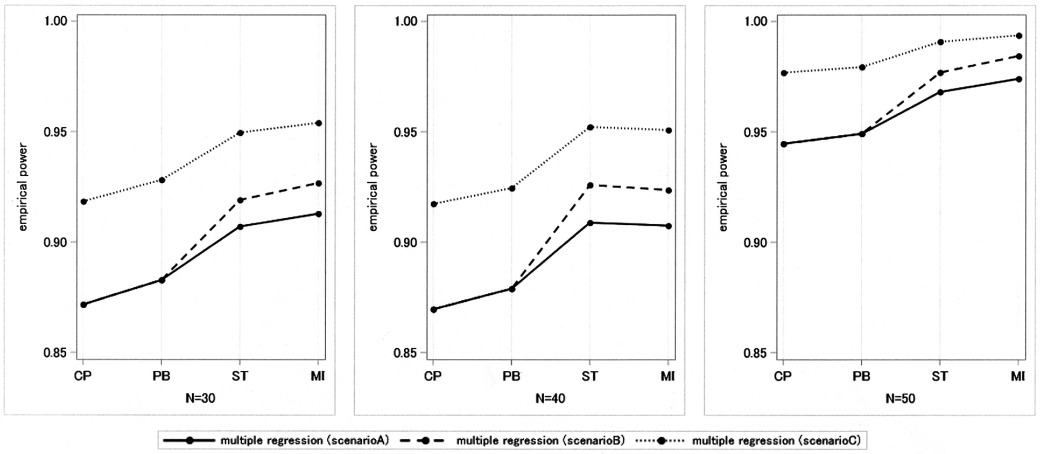

Fig. 3: Empirical power: when a prognostic factor that has an interaction is used in randomization and statistical analysis

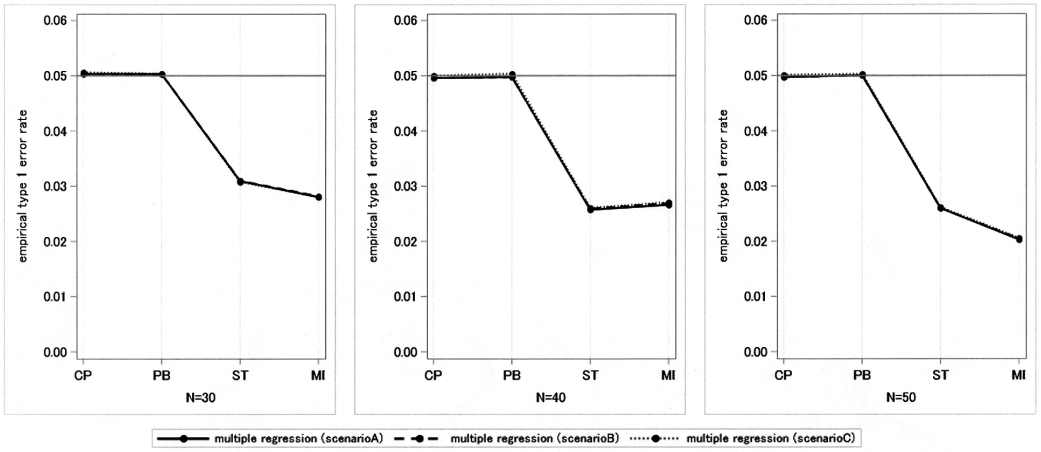

Fig. 4: Empirical type 1 error rate: when a prognostic factor that has an interaction is used in randomization and statistical analysis

3.2. Using baseline characteristics that have no effect on the outcome

Figure 5 and 6 show the empirical power and the empirical type 1 error rate when using prognostic factors in randomization and statistical analysis (scenario A), using baseline characteristics that have no effect on the outcome in randomization (scenario D) or using baseline characteristics that have no effect on the outcome in randomization and statistical analysis (scenario E). For Student's t-test and permutation test, the results of scenario D and E are summarized because variables used in randomization for these scenarios are the same. Figure 5 shows that the empirical power decreases when baseline characteristics that have no effect on the outcome is used in randomization. Regarding the type 1 error rate,

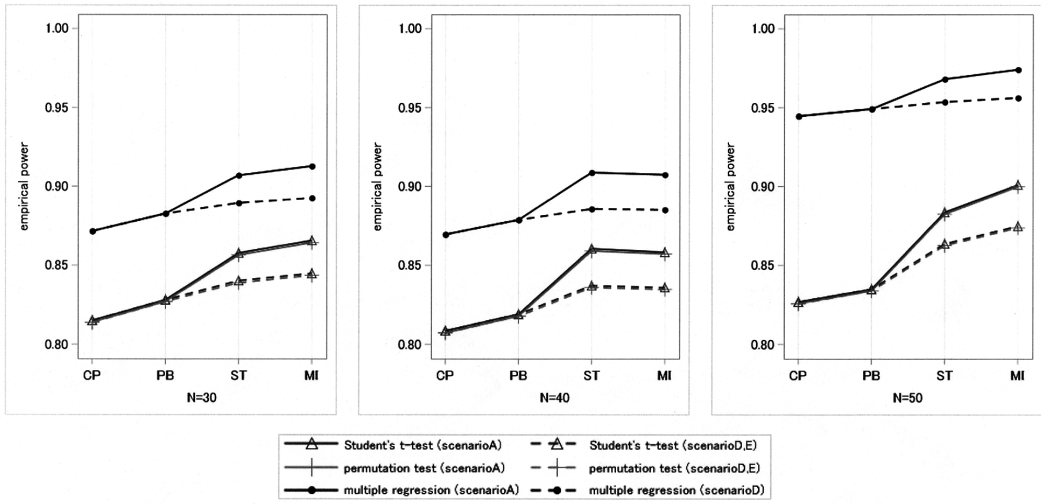

Fig. 5: Empirical power: when baseline characteristics that have no effect on the outcome is used in randomization

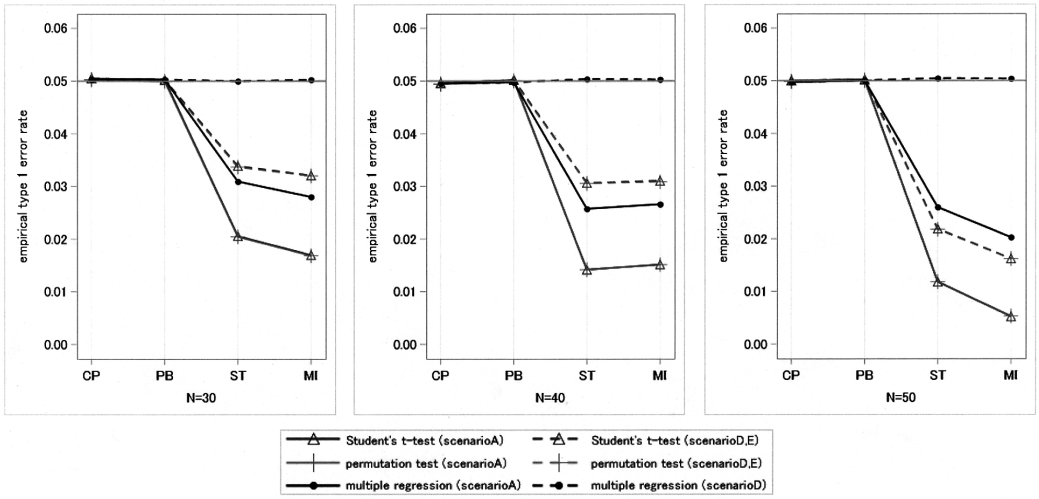

Fig. 6: Empirical type 1 error rate: when baseline characteristics that have no effect on the outcome is used in the randomization

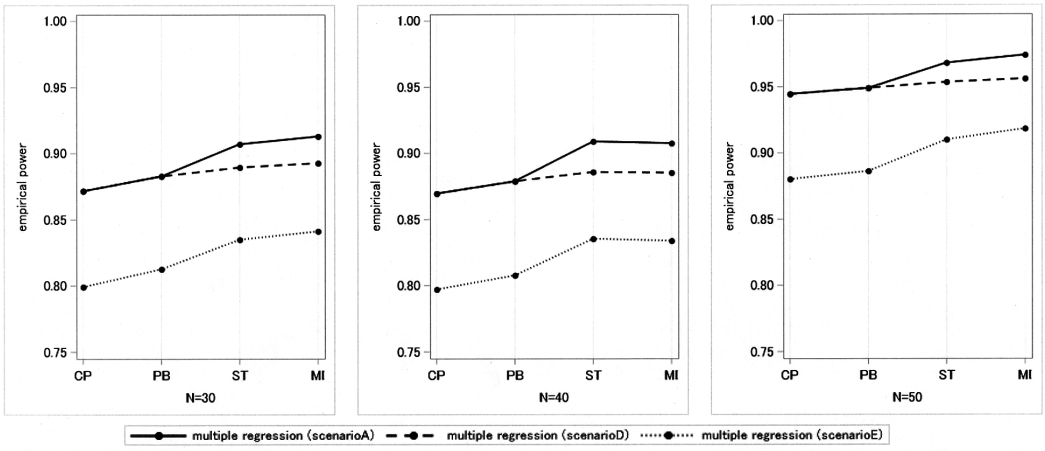

Fig. 7: Empirical power: when baseline characteristics that have no effect on the outcome is used in the randomization and statistical analysis

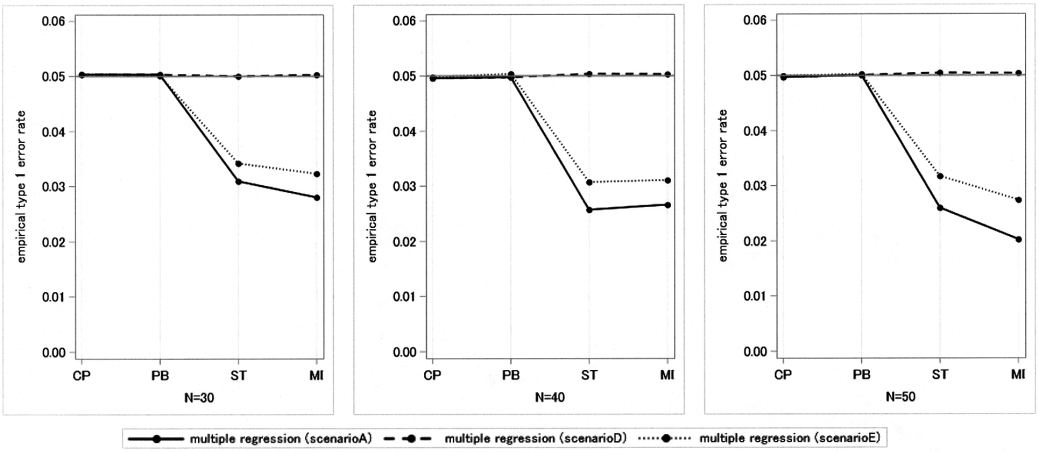

Fig. 8: Empirical type 1 error rate: when baseline characteristics that have no effect on the outcome is used in the randomization and statistical analysis

Figure 6 shows that CP and PB gave a valid empirical type 1 error rate of 0.05 regardless of the sample size, variables used in randomization and statistical analysis methods. On the other hand, ST and MI gave a lower type 1 error rate than the nominal type 1 error rate regardless of the sample size, variables used in randomization and statistical analysis methods except in the case of the multiple regression of scenario D.

Figure 7 and 8 show the empirical power and the empirical type 1 error rate for scenario A, D and E. Figure 7 shows that the empirical power is particularly decreased when the baseline characteristics that have no effect on the outcome is used in randomization and statistical analysis. Figure 8 shows that ST and MI gave a lower type 1 error rate than the nominal type 1 error rate in scenario D like scenario A.

4. Discussion

In the simulation study, we investigated the performance of the existing randomization methods when using (a) a prognostic factor that has an interaction, and (b) baseline characteristics that have no effect on the outcome.

The empirical power of ST and MI were higher than that of CP and PB, and the empirical type 1 error rate of ST and MI were lower than the nominal type 1 error rate, so it is considered that adjusting the distribution of prognostic factors positively in randomization is effective, especially when the sample size is small and it is difficult to execute multiple regression. Also, when using multiple regression model, the empirical power was very large compared to when using Student's t-test and a permutation test.

Regarding (a), the impact of an interaction in randomization and statistical analysis has not been clarified so much yet. This study has shown that the empirical power got even higher when using a prognostic factor that has an interaction in randomization and statistical analysis. Hence, it may be possible to cut down the sample size by considering a randomization method, statistical analysis method and prognostic factors that have interactions. This will make a great deal of sense to small sized clinical trials.

Regarding (b), whether the use of baseline characteristics that have no effect on the

outcome has an impact has not been investigated so much yet. This study has shown that the empirical power is lower when using such baseline characteristics in randomization and statistical analysis because it is considered that the addition of the baseline characteristics that have no effect on the outcome to ST or MI increases the imbalance in the distribution of other prognostic factors between the two groups and decreases the efficiency of the hypothesis testing. Therefore, it is considered as dangerous to include baseline characteristics easily because there is a possibility that the benefit of randomization method and statistical analysis cannot be received.

The simulation studies in this paper have shown that MI has a slightly better performance than ST. Some studies (Therneau, 1993; Weir and Lees, 2003; Hagino *et al.*, 2004; Toorawa *et al.*, 2009) have also shown the same tendency when the sample size is large. On the other hand, MI has some organizational issues (Scott *et al.*, 2002), and the performance of ST is not much inferior to MI. Therefore, we recommend not only MI but also ST by judging comprehensively from performance and simplicity of application.

One of the limitation of this paper is that we did not consider the site effect. This has been done to offer an easy interpretation, simplify the result, and we consider that small trials are often conducted monocentrically. Moreover, in a small multicenter randomized controlled trial, the number of patients per site will be very small, so it is probably difficult to conduct randomization including the site effect. However, we recommend accounting for the site as a prognostic factor in randomization and statistical analysis if the sample size per site is large enough and the site effect is larger than the other prognostic factors.

5. Conclusion

The objective of this paper was to compare the performance of the existing randomization methods when using a prognostic factor that has an interaction or baseline characteristics that have no effect on the outcome by simulation and to suggest the recommended randomization method for small randomized controlled trials. This paper has revealed that ST and MI with a prognostic factor that has an interaction have good performance even when the sample size is small. In addition, it has been shown that performance is improved by applying multiple regression jointly to those methods. On the other hand, this paper has revealed that the use of baseline characteristics that have no effect on the outcome in randomization and statistical analysis have a bad effect in terms of empirical power. These new findings were obtained from the results of the numerical simulations of this study, and they are considered to contribute to the selection of randomization method with proper statistical analysis in clinical trials and the development of future research on randomization.

Consequently, we recommend conducting ST or MI, multiple regression and using a prognostic factor that has an interaction in small randomized controlled trials.

REFERENCES

Brown, S., Thorpe, H., Hawkins, K. and Brown, J. M. (2005). Minimization-reducing predictability for multi-centre trials whilst retaining balance within centre. *Statistics in Medicine,* **24**, 3715–3727.

Friedman, L. M., Furberg, C. D., DeMets, D. L., Reboussin, D. M. and Granger, C. B. (2015). Introduction to clinical trials. In *Fundamentals of Clinical Trials,* pp. 1–23, Cham: Springer International Publishing.

Hagino, A., Hamada, C., Yoshimura, I., Ohashi, Y., Sakamoto, J. and Nakazato, H. (2004). Statistical comparison of random allocation methods in cancer clinical trials. *Controlled Clinical Trials,* **25**, 572–584.

Harrell, F. E. (2001). Multivariable modeling strategies. In *Regression Modeling Strategies,* pp. 53–85, New York: Springer Science+Business Media.

Kahan, B. C. and Morris, T. P. (2013). Adjusting for multiple prognostic factors in the analysis of randomised trials. *BMC Medical Research Methodology,* **13**, 99.

Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M. and Horwitz, R. I. (1999). Stratified randomization for clinical trials. *Journal of Clinical Epidemiology,* **52**, 19–26.

Lachin, J. M. (1988). Properties of simple randomization in clinical trials. *Controlled Clinical Trials,* **9**, 312–326.

Matts, J. P. and Lachin, J. M. (1988). Properties of permuted-block randomization in clinical trials. *Controlled Clinical Trials,* **9**, 327–344.

McPherson, G. C., Campbell, M. K. and Elbourne, D. R. (2013). Investigating the relationship between predictability and imbalance in minimisation: a simulation study. *Trials,* **14**, 86.

Pocock, S. J. and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics,* **31**, 103–115.

Rosenberger, W. F. and Lachin, J. M. (2015). Randomization and the clinical trial. In *Randomization in Clinical Trials: Theory and Practice,* 2nd Edition, New York: Wiley.

Scott, N. W., McPherson, G. C., Ramsay, C. R. and Campbell, M. K. (2002). The method of minimization for allocation to clinical trials: a review. *Controlled Clinical Trials,* **23**, 662–674.

Taves, D. R. (1974). Minimization: a new method of assigning patients to treatment and control groups. *Clinical Pharmacology and Therapeutics,* **15**, 443–453.

Therneau, T. M. (1993). How many stratification factors are 'too many' to use in a randomization plan? *Controlled Clinical Trials,* **14**, 98–108.

Toorawa, R., Adena, M., Donovan, M., Jones, S. and Conlon, J. (2009). Use of simulation to compare the performance of minimization with stratified blocked randomization. *Pharmaceutical Statistics,* **8**, 264–278.

Weir, C. J. and Lees, K. R. (2003). Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. *Statistics in Medicine,* **22**, 705–726.

Xiao, L., Lavori, P. W., Wilson, S. R. and Ma, J. (2011). Comparison of dynamic block randomization and minimization in randomized trials: a simulation study. *Clinical Trials,* **8**, 59–69.

(Received: July 19, 2017, Accepted: February 14, 2018)