2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Using prognostic factors in the design and analysis of clinical trials: a simulation study

Supervisor :
Prof. dr. Tomasz BURZYKOWSKI
Prof. dr. Marc BUYSE

Arsenio Nhacolo

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

universiteit hasselt
KNOWLEDGE IN ACTION

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

universiteit hasselt | Maastricht University

2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Using prognostic factors in the design and analysis of clinical trials: a simulation study

Supervisor :
Prof. dr. Tomasz BURZYKOWSKI
Prof. dr. Marc BUYSE

Arsenio Nhacolo

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

universiteit hasselt | Maastricht University

HASSELT UNIVERSITY

Abstract

Faculty of Sciences
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Master of Science in Biostatistics

Using prognostic factors in the design and analysis of clinical trials: a simulation study

by Arsénio Quingue Nhacolo

In many diseases, prognostic factors are known to have a profound influence on patient outcomes. These factors can be taken into account when conducting and analysing a randomized trial aimed at comparing two treatments. In this project, three treatment allocation methods, Simple Randomization (SR), Permuted Blocks within Strata (PB) and Minimization (M), were compared in respect to the power of log-rank test (asymptotic and re-randomization), the bias in treatment effect estimation, predictability of treatment allocation, and imbalance. The conclusion was that there is only a slight gain in power for PB and M as compared to SR. This gain appears to be more pronounced in trials with small sample sizes and almost negligible in large trials. PB tends to yield higher power as compared to M when few prognostic factors are used, while for large number of factors M performs better. Accounting for prognostic factors in the test leads to a considerable gain in power, but it reaches a point where adding more factors does not increase power or even reduces it. The number of prognostic factors to account for in the analysis depends on sample size, with bigger trials tending to benefit more from inclusion of more factors. In general, both asymptotic and re-randomization log-rank tests lead to similar power, although re-randomization tends to be more powerful when the test statistic is not stratified. The overall bias is similar in three allocation methods, with the variability being higher for SR as compared to PB and M. The bias variability is lower for PB as compared to M when few prognostic factors are used, and higher otherwise. SR is the least predictable, followed by PB, with M being the more predictable. The awareness of the allocation algorithm by the guesser increases considerably the predictability for PB and M. M is a clear winner when it comes to overall and marginal balance, while PB is the best performer for within strata balance.

Acknowledgements

I would like to thank my supervisors, Prof. dr. Marc Buyse and Prof. dr. Tomasz Burzykowski, for their attention and guidance during this project.

A special thanks to VLIR-UOS for the financial support, without which my Master Programme wouldn't have been possible.

Contents

List of Figures

List of Tables

Abbreviations

C	Control Group
CAP	Cyclophosphamide, Doxorubicin, and Cisplatin
CI	Confidence Interval
CP	Cyclophosphamide plus Cisplatin
E	Experimental Group
FDA	Food and Drug Administration
FIGO	International Federation of Gynecology and Obstetrics
$\boldsymbol{H_C}$	Hazard in Control Group
$\boldsymbol{H_E}$	Hazard in Experimental Group
HR	Hazard Ratio
M	Minimization
MT	Median Survival Time
$\boldsymbol{MT_C}$	Median Survival Time in Control Group
$\boldsymbol{MT_E}$	Median Survival Time in Experimental Group
MTR	Median Survival Time Ratio
N	Sample Size
PB	Permuted Blocks within Strata
PH	Proportional Hazards
RSLR	Re-randomization Stratified Log-rank Test
RULR	Re-randomization Unstratified Log-rank Test
SD	Standard Deviation
SLR	Asymptotic Stratified Log-rank Test
SR	Simple Randomization
ULR	Asymptotic Unstratified Log-rank Test

To the loving memory of my father, Quingue Nhacolo, and to my mother, Celeste Nhancule, who despite many difficulties did their best to provide education to their children.

1

Introduction

In many diseases, prognostic factors are known to have a profound influence on patient outcomes. These factors can be taken into account when conducting and analysing a randomized trial aimed at comparing two treatments. These factors can be used at the time of allocating treatments to patients entering the trial, and they can also be used in the analysis of the trial results. However, the decision on what procedure to use for allocating patients to treatment arms in comparative clinical trials is frequently made with only minor deliberation, and this decision may ultimately impact the trial inference, credibility, and even validity of the trial analysis (Pond, 2011). There is little practical guidance available to help trialists on deciding whether it is worthwhile to account for prognostic factors when allocating treatments to patients and when analysing the data, and if so, how many factors can be accounted for, and what methods are to be preferred.

The purpose of this work was to compare different methods of treatment allocation and of analysis. Three commonly used methods of treatment allocation were compared: simple randomization (without allowance for prognostic factors), permuted blocks within strata, and minimization. Three common methods of analysis, using log-rank test statistic for the comparison of survival curves, were compared: an unadjusted comparison of the randomized groups using the asymptotic distribution of the test statistic, an adjusted comparison of the randomized groups, taking prognostic factors into account (stratified log-rank test), and a re-randomization test based on a large number of re-randomized trials to generate an empirical distribution of the test statistic under the null hypothesis. The combinations of these allocation methods and analysis methods were compared with respect to power of the statistical test. The allocation methods were also compared in respect to predictability of treatment allocation, bias in the estimated treatment effect, and imbalance. The comparisons were done via simulations conducted using actual clinical trial data.

The description of the data and the methodology are presented in Section 2, the results in in Section 3, the discussion and conclusion in Section 4.

2

Materials and Methods

2.1 Data

The data, on which simulations were based, are from a meta-analysis (The Ovarian Cancer Meta-Analysis Project, 1991) involving four randomized clinical trials comparing Cyclophosphamide plus Cisplatin (CP) versus Cyclophosphamide, Doxorubicin, and Cisplatin (CAP) in the treatment of ovarian carcinoma. The data set contained 1198 patients in total. Only 1011 patients, that had no missing values in all the variables of interest, were considered in the current project. The variables of interest were the treatment group, baseline clinical characteristics: age (a), performance status (p), extent of residual disease after debulking surgery (r), histological grade (g) and FIGO stage (s), the outcomes: survival time (in years) and survival status.

2.2 Exploratory Data Analysis

The distribution of the baseline clinical characteristics was checked via tabulations. To get the estimate of treatment effect (hazard ratio) and to have an idea of prognostic importance of the patient's baseline clinical characteristics , a Cox proportional hazard model (Cox, 1972) having treatment and baseline clinical characteristics as covariates and stratified by trial was fitted. The proportional hazard assumption was checked using a test based on Schoenfeld residuals (Grambsch and Therneau, 1994). The model fitting was done solely as exploratory tool which will help in further steps, it is by no means intended to be a meta-analysis. Readers interested in the results of the meta-analysis can refer to The Ovarian Cancer Meta-Analysis Project (1991). An extensive analysis of the same data was also done by Dubois (2001).

2.3 Simulations

Before simulations, a new survival time variable, $TIMEC$, was created in the original dataset. For the patients belonging to the control group (C), $TIMEC$ was set equal to original survival time, and for patients in the experimental group (E) it was set to the original time multiplied by the hazard ratio (HR) estimated from the Cox model described in Section 2.2 (the treatment indicator in this model refer to the treatment that the patient had actually received, not the treatment allocated in the simulations). This was done in order to "cancel" the treatment effect. Then in the simulated trials, the survival time was set equal to $TIMEC$ for the patients that were assigned to control group, and equal to $TIMEC$ divided by the desired HR for patients assigned to experimental group. This approach is reasonable if HR (H_E/H_C) is constant over time (i.e., proportional hazards) and the median survival time (MT) ratio (MT_C/MT_E) is approximately equal to HR, which is the case for negative exponential survival distributions (see Cortés et al., 2014; Machin et al., 2009, 2006; Peace, 2009). To check whether the survival function in each treatment group follows a negative exponential distribution, a exponential model was fitted to Kaplan-Meier survival estimates (Machin et al., 2006), using least squares non-linear regression (Bates and Watts, 2007). The model was specified as

$$S(t) = e^{-\lambda t}$$

where $S(t)$ is survival at time t and λ the hazard rate.
The censoring indicator (survival status) was kept unchanged.

For comparability, trials with sample size N simulated using different treatment allocation methods were based on the same set of samples of size N, randomly selected (with replacement) from the original data set. For each N, 1000 trials were generated. The considered N's were 100, 200 and 500, with HR of 0.5, 0.6 and 0.7, respectively.

2.3.1 Power

Different sets of patient's baseline clinical characteristics were considered as prognostic factors when allocating treatment and when testing for the difference in survival curves. For each N and HR, 1000 trials were generated for each of the allocation methods: Simple Randomization (SR), Permuted Blocks within Strata (PB) and Minimization (M). For each trial, the unstratified and stratified log-rank test statistics were calculated, then p-value was calculated based on their asymptotic theoretical distribution (χ^2_1) and

on their empirical distribution (obtained via 500 re-randomizations[1]). The power was calculated as the proportion of trials that had p-value less than 0.05 under the alternative hypothesis (HR < 1).

The considered sets of prognostic factors were: a; a and r; a, r and g; a, p, r and g; a, p, r, g and s. The details on how the allocation methods and the re-randomization test were implemented follow below.

Simple Randomization

Considered the most basic allocation method, Simple Randomization consists in assigning patients to different treatment arms with equal probability, regardless of all other considerations (Pond, 2011). Trials were generated by simply assigning to each patient a treatment randomly selected, with equal probability, from the set of two.

Permuted Blocks within Strata

With this method, the patient allocation list is grouped into blocks of size 2k (assuming 1:1 allocation) within each stratum, with k patients in each block assigned to each treatment (Pond, 2011). The strata are formed by the combination of levels of prognostic factors in consideration. The block size was set to 4 and, therefore, there were six possible permutations of treatments in a block: CCEE, EECC, ECEC, CECE, ECCE and CEEC. In each stratum, a permutation was selected at random and patients belonging to that stratum were sequentially assigned to treatment according to the selected permutation. Upon completion of the block, a new permutation was again randomly selected for the next patients. This process continued until the desired N was attained.
Permuted Blocks method has advantage of forcing periodic balance in the number of patients allocated to each treatment group within every stratum, and has disadvantage of predictability of last treatment(s) in a block (Buyse, 2000).

Minimization

Minimization is covariate-adaptive randomization method, it is among the dynamic allocation methods family proposed by Pocock and Simon (1975). As opposed to PB, it attempts to achieve treatment balance on several subject characteristics simultaneously– not within separate strata (Buyse, 2000). Its implementation consists in, using some

[1]Re-randomization test is computer intensive. Due to time constraints, power estimation were based on only 500 trials with 500 re-randomizations. This is clearly stated on the tables where results are presented.

measure of imbalance, assigning higher allocation probability P to the treatment that would minimize the total imbalance. There are many versions of the minimization algorithm (see Begg and Iglewicz, 1980; Heritier et al., 2005; Perry et al., 2010; Pocock and Simon, 1975; Taves, 1974; Wei, 1977). The algorithm used here consisted in getting the categories of prognostic factors of a new patient, then count the number of patients that are already in each of these categories for each treatment group (note that a patient can be counted more than once), then sum them up. Then treatment is randomly assigned, with probability of 0.9 for the treatment group with lower total. When the totals are equal (which is always the case when enrolling the first patient), the treatments are allocated with equal probability. This algorithm is similar to the one described by Buyse (2000), and it is equivalent to the one proposed by Pocock and Simon (1975) that uses variance as measure of imbalance (McEntegart, 2003).

Re-randomization test

Re-randomization test has been considered as a robust alternative to the traditional population model-based methods for analysing randomized clinical trials (Han et al., 2013). The re-randomization tests are frequently requested by the Food and Drug Administration (FDA) to confirm the results from standard tests, specially when a trial used a dynamic allocation method (McEntegart, 2003). The test consists in simulating large number of trials by re-allocating treatment to the patients (the response is held fixed) using the same allocation method as in the original trial. The p-value is the proportion of simulated trials with test statistic equal to, or greater than, the one of the observed trial. This argument does not depend on the test statistic nor on the distribution of the response (Gabriel and Hsu, 1981). Unlike the permutation test, re-randomization test doesn't give exact p-value, the desired accuracy can be achieved by increasing the number of simulated trials (Morgan and Rubin, 2012).

2.3.2 Bias

The distribution of bias was compared across the three allocation methods. The same set of trials of Section 2.3.1 was used here. The bias was calculated as the difference of the natural logarithm of the estimated HR ($\hat{\beta}$) and the natural logarithm of the true HR (β), i.e, $bias = \hat{\beta} - \beta$. The following Cox PH model (Cox, 1972) was used to calculated $\hat{\beta}$:

$$h_E(t) = h_c(t)e^{\hat{\beta}T}$$

where $h_E(t)$ and $h_c(t)$ are, respectively, the hazard rates of experimental and control groups at time t, and T is the treatment indicator (0 for control group and 1 for experimental group).

2.3.3 Predictability

Following the suggestion by Blackwell and Hodges (1957), the predictability of the next treatment allocation was measured as the probability of a correct guess when the investigator uses an optimal guessing strategy. For designs that assign more probability to the under-represented treatment, this consists of picking the treatment that has been observed fewer times, without preference in case of a tie (Antognini and Giovagnoli, 2015). There are other guessing strategies discussed by various authors, see for example Brown et al. (2005); Hills et al. (2009); McPherson et al. (2013).
In the present project, assuming that the guesser has access to the information about all patients previously accrued, two scenarios were considered for the same guessing strategy:

1. *The guesser doesn't know the allocation algorithm*
   In this case he/she would pick a treatment at random, if both treatment groups have equal number of patients, or pick the treatment group with less patients otherwise.

2. *The guesser knows the allocation algorithm*
   Here he/she would proceed as in scenario 1 for Simple Randomization. For Permuted Blocks, the guesser would get the number of patients (ns) that are in the stratum the next patient belongs to, get the number of patients in current block (nb) by taking the remainder of the division of ns by block size, then pick a treatment at random if $nb = 0$ or pick the treatment that occurs less (in the current block) otherwise. This would always result in correct guesses for the last treatment in a block. In case Minimization, the guesser would sum up the number of patients in the categories of prognostic factors of the next patient in each treatment group (a patient can be counted more than once) , then pick a treatment at random, if totals are equal in both treatment groups, or pick the treatment with lower total otherwise.

2.3.4 Imbalance

The definition of imbalance used here was the same used by Dubois (2001), in which the imbalance between the two treatment groups in a trial was calculated as the absolute

difference of the number of patients in the two groups. When considering prognostic factors, this absolute difference was calculated in each stratum formed by cross-classification of the factors, then the maximum was taken as the value of imbalance. The comparison of the allocation methods was done using the mean and standard deviation of the imbalance.

The code for all simulations and analyses was written in `R` (R Core Team, 2015).

3

Results

3.1 Exploratory Data Analysis

The distribution of the baseline clinical characteristics, shown in Table 3.1, reveals that the majority (43.8%) of the patients were under 55 years of age. The most frequent performance status category was $ECOG0$ (64.3%), followed by $ECOG1$ (26.1%). Over 82% of the patients had residual disease, only 17.2% had histological grade classified as good, and the FIGO stage III was the most frequent (83.1%).

The test of hazards proportionality on the Cox PH model (described in Section 2.2), failed to reject the null hypothesis of proportional hazards ($p-value = 0.3346$), meaning that the treatment HR was constant over time. The results of the model are presented is Table 3.2. The treatment HR estimate was 0.8394, and it was statistically significant (as its 95% confidence interval did not contain 1). The results also suggested that all baseline characteristics (except FIGO stage) were of prognostic importance.

The median survival times in Control and Experimental groups were 2.06 and 2.44 years, respectively, yielding a median survival time ratio (MTR) of 0.8443, which was very close to the HR. Figure A.1 shows the plots of Kaplan-Meier survival curve and its fitted values from a negative exponential model. It can be seen that the model fits the survival reasonably well in both treatment groups, suggesting that the survival follows a negative exponential distribution. The estimate of the parameter λ (hazard rate) was 0.2858 in Control group and 0.2306 in Experimental group, yielding a HR of 0.8068, which is to the one estimated from Cox PH above. These findings make it reasonable to generate treatment effect in the simulated trials by dividing the baseline survival time in the Experimental group by the desired HR (see Section 2.3).

TABLE 3.1: Distibution of patients' baseline clinical characteristics

Total number of patients: 1198				
Characteristic	**Description**	**Value**	**N**	**%**
a	Age	<55	525	43.8
		55-65	419	35.0
		>65	248	20.7
		Missing	6	0.5
p	Performance status	ECOG 0	770	64.3
		ECOG 1	313	26.1
		ECOG 2 or worse	59	4.9
		Missing	56	4.7
r	Extent of residual disease	No residual disease	191	15.9
		≤2 cm	463	38.6
		>2 cm	523	43.7
		Missing	21	1.8
g	Histological grade	Good	206	17.2
		Intermediate	402	33.6
		Poor	463	38.6
		Missing	127	10.6
s	FIGO stage	I	7	0.6
		II	25	2.1
		III	996	83.1
		IV	157	13.1
		Missing	13	1.1

3.2 Simulations

Since the original treatment allocation was ignored when simulating trials, the treatment effect was "cancelled" before simulations took place by multiplying the survival time in Experimental group by the HR estimated from Cox PH model in Section 3.1. This was done so that all patients have the same distribution of baseline survival time. Then, in the simulated trials, the treatment effect was generated dividing this baseline survival time in the Experimental group by the desired HR (see Section 2.3 and last paragraph of Section 3.1 for reasoning). The results of the model re-fitted after "cancelling" the treatment effect are presented in Table A.1. They showed a statistically non-significant treatment HR of 0.969, while the prognostic importance of baseline characteristics remained as before.

TABLE 3.2: Hazard ratio estimated from Cox PH model fitted on the original data. The model was stratified by trial and had all baseline characteristics as covariates.

				95% CI Limits	
Covariate	**Reference cat.**	**Category**	**Estimate**	**Lower**	**Upper**
Treatment	CP (Control)	CAP (Exper.)	0.8394	0.7317	0.9629
a	<55	55-65	1.1609	0.9920	1.3586
		>65	1.4861	1.2359	1.7869
p	ECOG 0	ECOG 1	1.4599	1.2428	1.7149
		ECOG 2 or worse	1.8958	1.3679	2.6274
r	No residual disease	≤2 cm	1.4605	1.1639	1.8326
		>2 cm	2.8656	2.1726	3.7797
g	Good	Intermediate	1.3959	1.1303	1.7239
		Poor	1.4891	1.2095	1.8334
s	I	II	0.7728	0.2202	2.7120
		III	0.8192	0.2509	2.6749
		IV	1.0793	0.3253	3.5802

3.2.1 Power

The results are presented in Tables 3.3, A.2 and A.3. Graphical illustration of some aspects is done in Figures 3.1, 3.2 and 3.3. Looking at the allocation methods, there was a slight gain of power when taking in to account the prognostic factors during the randomization (see Tables 3.3 and A.2), but this gain was less important or almost non-existent when the sample size was large (see Table A.3). When there was a gain, Permuted Blocks within Strata (PB) tended to perform better than Minimization (M) when low number of prognostic factors is considered, while M tended to be better otherwise. It was not clear which number (set) of factors maximizes the power. Unlike in randomization, taking in to account the prognostic factors in the analysis resulted in a considerable gain of power. For instance, in trials of size 200 (Table 3.3), the power when no factor is taken into account in both randomization and analysis is about 0.56 and the maximum power obtained when taking into account factors in randomization was only about 0.63, while in analysis it was 0.73. The maximum benefit was obtained when 2, 3 and all (5) factors were considered in the analysis in trials of size 100, 200 and 500, respectively. However, using all the 5 factors was still better than ignoring them all in trials of size 200, while for trials of size 100 it was worse.

Despite the tendency of higher power in unstratified tests, in general the re-randomization tests yielded power similar to that of the respective asymptotic (parametric) tests. The

scatter plot of the power of asymptotic versus re-randomization tests, shown in Figure 3.3, revealed high correlation of their power. This was confirmed by high coefficient of linear correlation, which had value of 0.94.

Simulation aiming at studying the behaviour of the p-value of re-randomization test as a function of the number of re-randomizations showed that p-value starts to stabilize when the number of re-randomizations approaches 500 (see Figure A.2). Some minor fluctuations were still seen up to around 5000 re-randomizations. Therefore, in absence of time and/or computing resources constraints, the preferable number of re-randomization would 5000 or more.

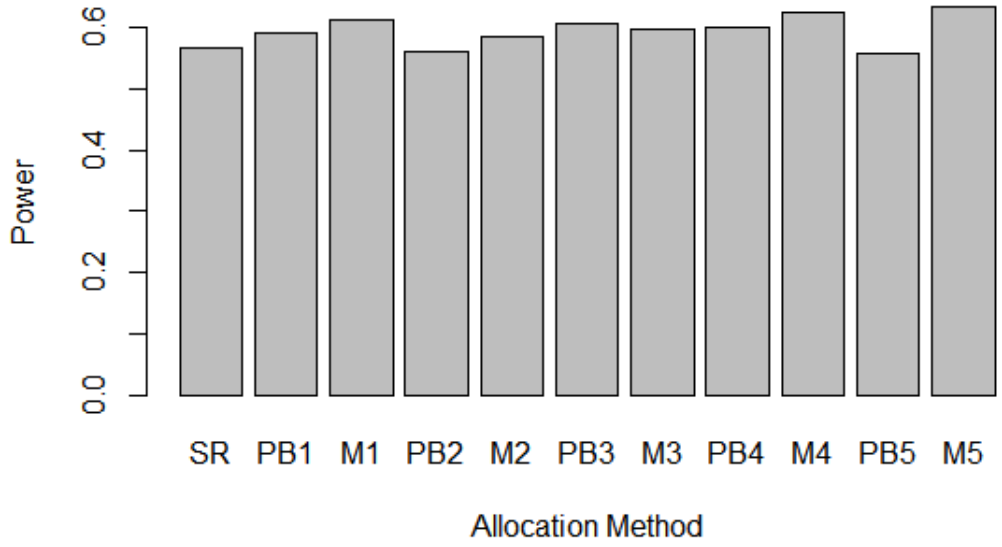


FIGURE 3.1: Power of asymptotic unstratified log-rank test in different allocations, based on 500 trials of size 200. The allocation methods are simple randomization (SR), permuted blocks within strata (PB) with blocks of size 4, and Minimization (M) with allocation probability of 0.9. The number attached to the abbreviation of allocation method represents the set of stratifying prognostic factors (1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s).

3.2.2 Bias

Results on bias, based on 1000 simulated trials, are presented in Table 3.4 and graphically displayed in Figures 3.4 and B.1. For each combination of sample size and effect size, the overall bias was similar in all allocation methods. The variability (SD) was slightly higher for Simple Randomization. Both overall bias and variability decreased with increasing sample size. The variability tended to be higher for Minimization as compared to Permuted Blocks within Strata when a small number of stratifying factors was used (1 and 2), and similar when 3, 4 and 5 factors were used.

TABLE 3.3: Power based on 500 trials of size 200, assuming HR of 0.6, and with 500 re-randomizations for re-randomization tests. URL is unstratified log-rank test, SLR are the stratified ones (the ending cardinal indicates the set of stratifying factors, as specified in column Factor). Their re-randomization counterparts take prefix R. Allocation methods (Alloc.) are simple randomization (SR), permuted blocks within strata (PB), and Minimization (M). The prognostic factors are age (a), performance status (p), extent of residual disease (r), histological grade (g) and FIGO stage (s). The assumed significance level was $\alpha = 0.05$ for the power calculations. For the allocation methods, blocks of size 4 were used in PB, and a stochastic method of variances (with probability of 0.9) was used for M.

		Tests											
Factors	**Alloc.**	**ULR**	**SLR1**	**SLR2**	**SLR3**	**SLR4**	**SLR5**	**RULR**	**RSLR1**	**RSLR2**	**RSLR3**	**RSLR4**	**RSLR5**
	SR	0.568	0.666	0.680	0.734	0.738	0.712	0.562	0.648	0.678	0.714	0.730	0.718
1 (r)	PB	0.592	0.688	0.728	0.748	0.730	0.746	0.646	0.678	0.726	0.758	0.730	0.746
	M	0.614	0.672	0.704	0.758	0.738	0.710	0.644	0.662	0.694	0.760	0.716	0.712
2 (a,r)	PB	0.560	0.694	0.734	0.772	0.746	0.730	0.634	0.686	0.732	0.766	0.740	0.724
	M	0.586	0.680	0.704	0.742	0.732	0.716	0.628	0.672	0.702	0.736	0.706	0.716
3 (a,r,g)	PB	0.606	0.694	0.734	0.800	0.762	0.748	0.666	0.700	0.726	0.798	0.758	0.758
	M	0.598	0.686	0.710	0.776	0.750	0.754	0.650	0.696	0.706	0.762	0.746	0.754
4 (a,p,r,g)	PB	0.600	0.684	0.728	0.814	0.816	0.800	0.676	0.710	0.744	0.818	0.818	0.800
	M	0.626	0.728	0.756	0.782	0.766	0.756	0.690	0.732	0.762	0.792	0.756	0.758
5 (a,p,r,g,s)	PB	0.558	0.644	0.714	0.780	0.808	0.790	0.618	0.682	0.728	0.780	0.798	0.790
	M	0.634	0.696	0.718	0.740	0.744	0.746	0.686	0.718	0.732	0.752	0.744	0.734

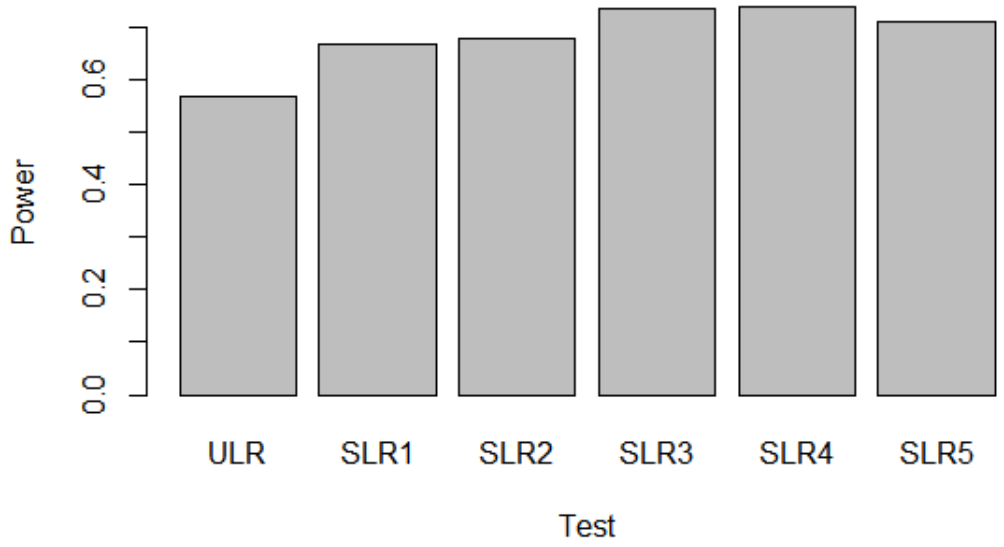


FIGURE 3.2: Power of different asymptotic tests, based on 500 trials of size 200 generated with Simple Randomization. The tests are Unstratified Log-rank (ULR) and Stratified Log-rank (SLR). The number attached to the abbreviation of the test represents the set of stratifying prognostic factors (1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s).

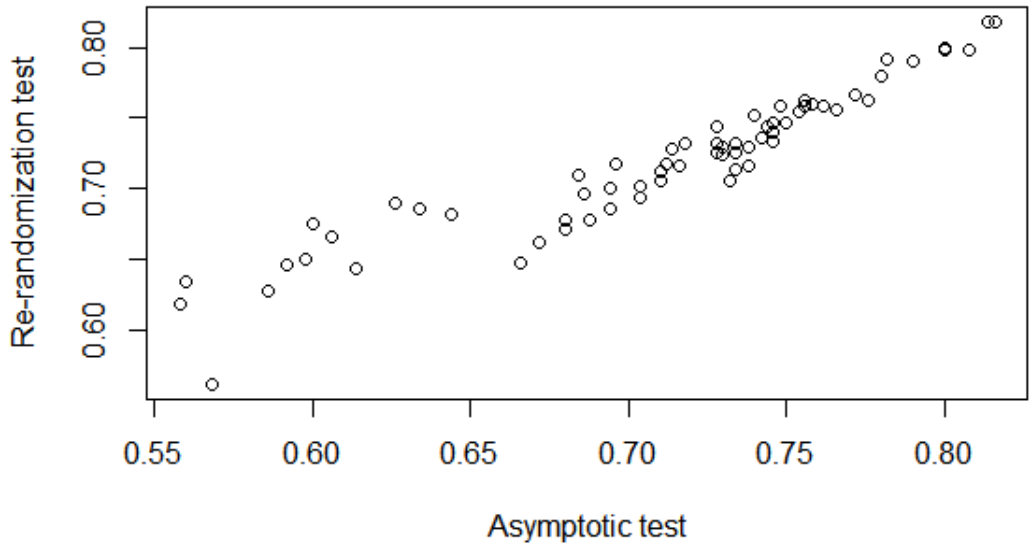


FIGURE 3.3: Power of asymptotic versus the corresponding re-randomization tests, based 5500 trials of size 200 from all allocation methods. The linear correlation was 0.94.

TABLE 3.4: Bias in log HR scale, based on 1000 trials. Allocation methods (Al.) are simple randomization (SR), permuted blocks within strata (PB), and Minimization (M). The prognostic factors are age (a), performance status (p), extent of residual disease (r), histological grade (g) and FIGO stage (s).

		N=100, β=ln(0.5)		N=200, β=ln(0.6)		N=500, β=ln(0.7)	
Factors	**Al.**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
	SR	0.213967	0.242014	0.159744	0.161159	0.115837	0.100037
1 (r)	PB	0.207950	0.215979	0.157611	0.148879	0.114662	0.091186
	M	0.215792	0.223348	0.165353	0.151851	0.118411	0.095699
2 (a,r)	PB	0.216174	0.208086	0.166372	0.149443	0.117665	0.092343
	M	0.216404	0.225901	0.164672	0.151280	0.117287	0.094032
3 (a,r,g)	PB	0.226095	0.212609	0.164829	0.139684	0.120007	0.091623
	M	0.195410	0.213855	0.161981	0.148631	0.115961	0.092758
4 (a,p,r,g)	PB	0.208525	0.211220	0.163247	0.147971	0.118640	0.092659
	M	0.218096	0.211020	0.160491	0.139893	0.123892	0.089880
5 (a,p,r,g,s)	PB	0.204755	0.215961	0.166580	0.145735	0.117543	0.088301
	M	0.218977	0.211191	0.161520	0.144947	0.118346	0.090346

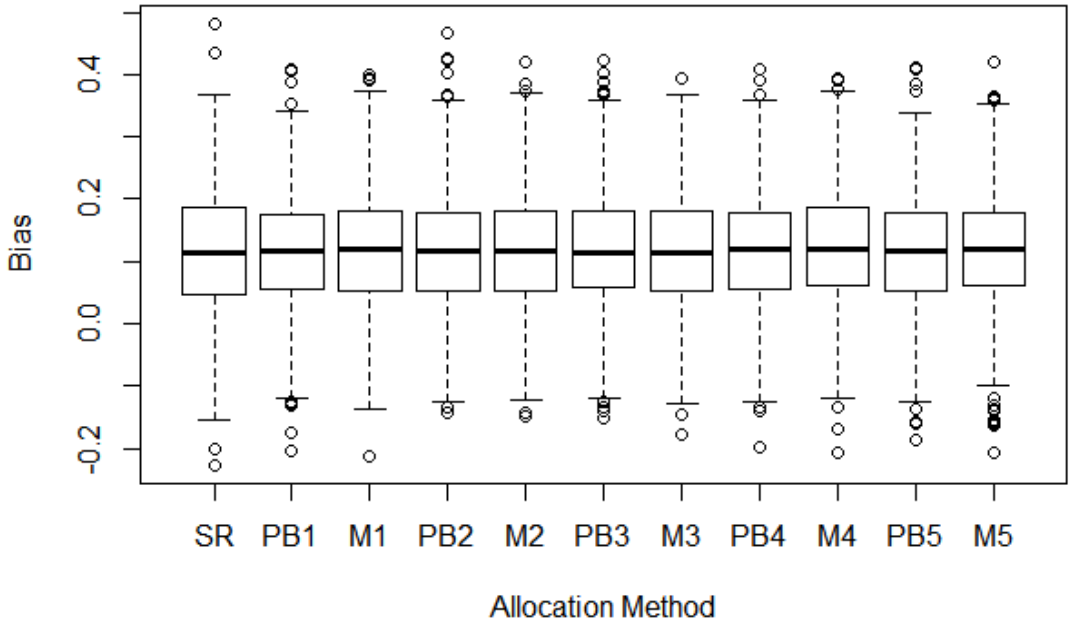


FIGURE 3.4: Bias in log HR scale, based on 1000 trials of size 500. The allocation methods are simple randomization (SR), permuted blocks within strata (PB) with blocks of size 4, and Minimization (M) with allocation probability of 0.9. The number attached to the abbreviation of allocation method represents the set of stratifying prognostic factors (1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s).

3.2.3 Predictability

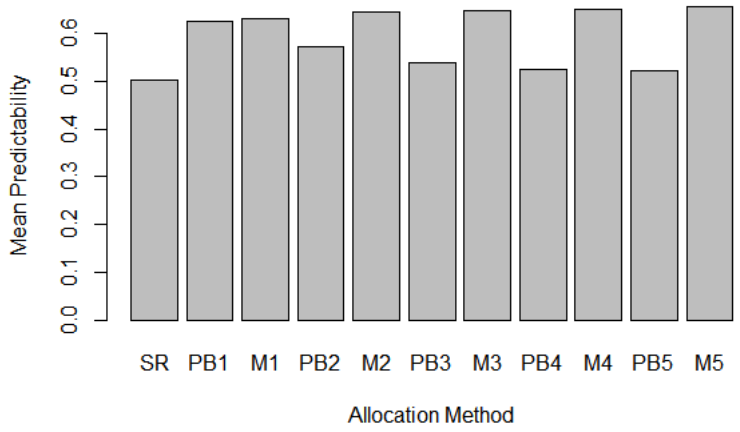
The Simple Randomization was the least predictable allocation method, with an overall predictability of 0.5 for the different sample sizes, irrespective of whether the guesser was aware of the allocation algorithm or not. The Permuted Blocks within Strata followed, with predictability that decreased with increasing number of stratifying factors. When the guesser was not aware of the allocation algorithm the predictability estimates were, respectively, 0.663, 0.570, 0.539, 0.525 and 0.52 when 1, 2, 3, 4 and 5 stratifying factors were used in trials of size 200. For the same sample size and when the guesser knew the allocation algorithm, the values were 0.704, 0.700, 0.684, 0.657 and 0.638, showing that predictability was higher is this case, but still decreased with increasing number of stratifying factors as before. Similar trends were seen in other sample sizes. When fixing the set of stratifying factors, the overall predictability tended to have slightly higher values for larger sample sizes. Minimization was the most predictable allocation method, with overall predictability taking values 0.630, 0.644, 0.647, 649, and 0.654, when the guesser was not aware of the allocation algorithm, and 0.719, 0.800, 0.832, 0.848 and 0.856, when the guesser was aware of the algorithm, for 1, 2, 3, 4 and 5 stratifying factors respectively, and sample size of 200. Unlike with PB, the predictability increased with increasing number of stratifying factors and, like with PB, it was higher when the guesser was aware of the algorithm. The sample size did not have any considerable impact. Detailed results are shown in Table 3.5 and in Figures 3.5, 3.6, C.1, C.2 and C.3.

TABLE 3.5: Mean predictability, based on 1000 trials of size N, in scenarios where the guesser knows and doesn't know the allocation algorithm. The allocation methods (Al.) are simple randomization (SR), permuted blocks within strata (PB) with blocks of size 4, and Minimization (M) with allocation probability of 0.9. The prognostic factors are age (a), performance status (p), extent of residual disease (r), histological grade (g) and FIGO stage (s).

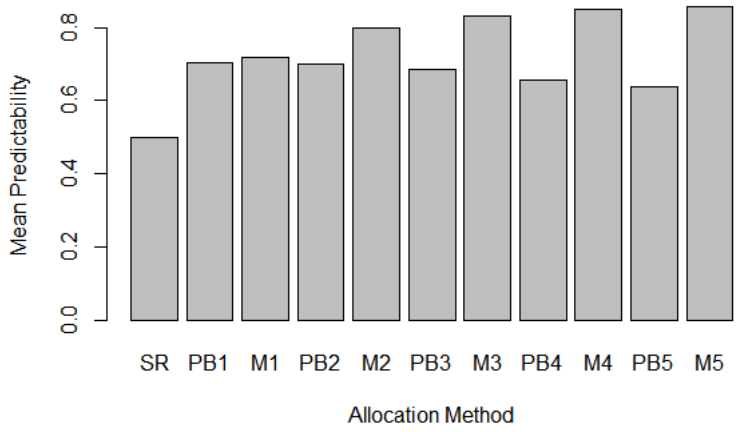
		Algorithm unknown			**Algorithm known**		
Factors	**Al.**	**N=100**	**N=200**	**N=500**	**N=100**	**N=200**	**N=500**
	SR	0.50330	0.50068	0.49960	0.50345	0.50064	0.49958
1 (r)	PB	0.62246	0.62315	0.62372	0.70073	0.70411	0.70737
	M	0.63069	0.63026	0.63133	0.72039	0.71974	0.72177
2 (a,r)	PB	0.56918	0.57026	0.57168	0.69163	0.70040	0.70554
	M	0.64321	0.64353	0.64418	0.79757	0.79986	0.80221
3 (a,r,g)	PB	0.53592	0.53856	0.54188	0.66281	0.68422	0.69881
	M	0.64677	0.64664	0.64682	0.82687	0.83155	0.83210
4 (a,p,r,g)	PB	0.52442	0.52500	0.52597	0.62423	0.65729	0.68303
	M	0.65177	0.64927	0.64880	0.84329	0.84772	0.84900
5 (a,p,r,g,s)	PB	0.51722	0.52000	0.52083	0.60352	0.63810	0.67199
	M	0.66074	0.65432	0.64768	0.85525	0.85564	0.85790

3.2.4 Imbalance

The results of imbalance are shown in Tables 3.6, D.1 and D.2. Plots of overall imbalance are presented in Figures 3.7, D.1 and D.2. In terms of overall balance, there was a lot of gain by taking in to account the prognostic factors. For instance, in trials of size 200 (Table 3.6) the overall imbalance when factors are ignored (Simple Randomization) had mean of about 11 with standard deviation of 8.6, these values dropped respectively to 1 and 1.2 when one factor was taken into account. Permuted Blocks (PB) and Minimization (M) performed equally when only factor was used, but when more factors were added PB worsened while M improved. For example, in trials of size 200 the mean overall imbalance for PB and M were, respectively, 1.1 and 1.0 for one factor and, 6.3 and 0.8 for 5 factors. So, the clear winner in terms of overall imbalance is M. When it comes



(a) Guesser doesn't know the allocation algorithm



(b) Guesser knows the allocation algorithm

FIGURE 3.5: Mean predictability, based on 1000 trials of size 200. The allocation methods are simple randomization (SR), permuted blocks within strata (PB) with blocks of size 4, and Minimization (M) with allocation probability of 0.9. The number attached to the abbreviation of allocation method represents the set of stratifying prognostic factors (1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s).

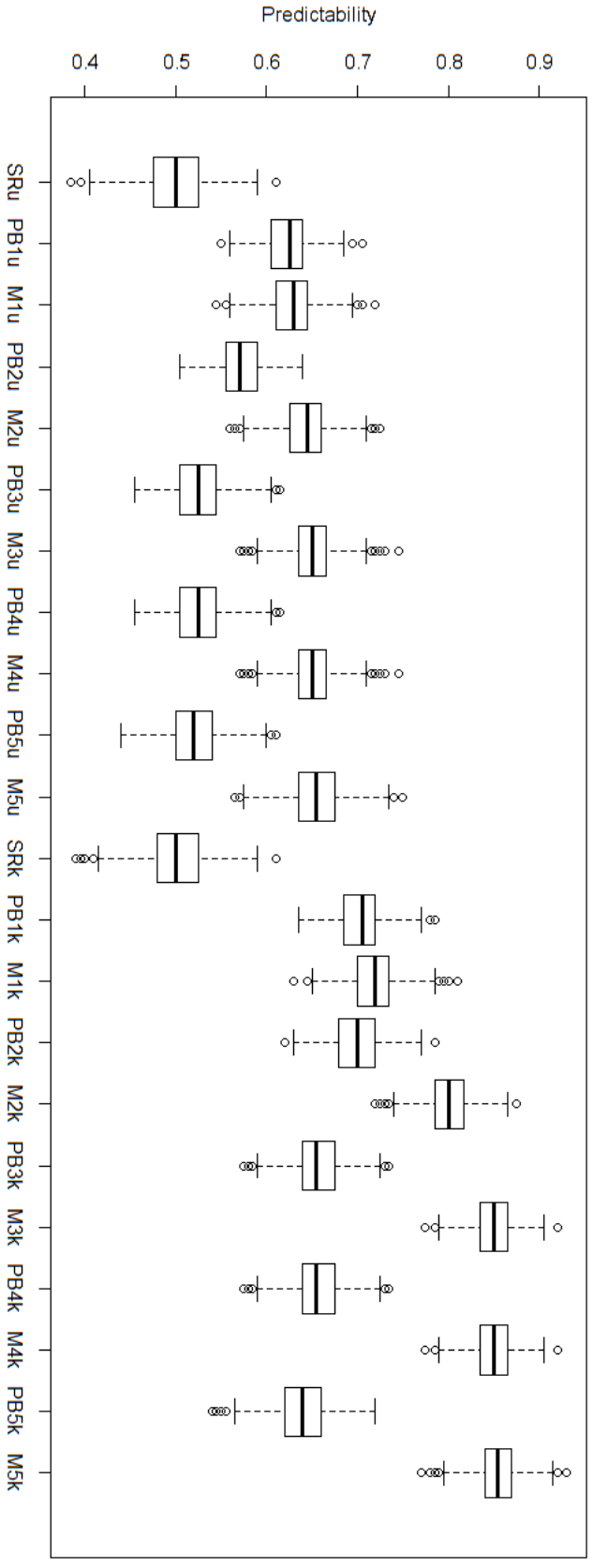


FIGURE 3.6: Predictability, based on 1000 trials of size 200. The allocation methods are simple randomization (SR), permuted blocks within strata (PB) with blocks of size 4, and Minimization (M) with allocation probability of 0.9. The number attached to the abbreviation of allocation method represents the set of stratifying prognostic factors (1:*r*; 2:*a*,*r*; 3:*a*,*r*,*g*; 4:*a*,*p*,*r*,*g*; 5:*a*,*p*,*r*,*g*,*s*), and the ending letter indicates whether the guesser knows (k) or doesn't know (u) the allocation algorithm.

to within strata imbalance, the trend reversed in favour of PB. M was only better in marginal balance (when only one strata defining factor was used), but for strata defined by cross-classification of more than one factor, PB was better. Similar trends were seen in trials of sizes 100 and 500.

TABLE 3.6: Imbalance, based on 1000 trials of size 200. **F** represent the set of prognostic factors taken in to account (0: none, 1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s). **A** is the treatment allocation method. **I** stands for imbalance, with the number attached to it representing the set of prognostic factors (**F**) defining the strata in which imbalances were calculated. For instance, **I0** is the overall imbalance (no factor).

		Mean (SD)					
F	**A**	**I0**	**I1**	**I2**	**I3**	**I4**	**I5**
0	SR	11.2 (8.6)	10.8 (5.3)	9.0 (3.0)	7.0 (1.8)	6.0 (1.62)	5.5 (1.47)
1	PB	1.1 (1.2)	1.1 (0.6)	6.8 (2.3)	6.3 (1.6)	5.6 (1.40)	5.3 (1.29)
	M	1.0 (1.2)	1.0 (0.7)	6.6 (2.4)	6.3 (1.7)	5.6 (1.43)	5.3 (1.33)
2	PB	2.1 (1.7)	2.0 (0.9)	1.5 (0.5)	5.1 (1.4)	5.0 (1.22)	4.8 (1.17)
	M	0.9 (1.1)	1.3 (0.7)	4.8 (1.8)	5.8 (1.6)	5.3 (1.37)	5.1 (1.24)
3	PB	3.6 (3.0)	3.5 (1.6)	2.8 (0.8)	1.9 (0.3)	3.7 (0.90)	3.8 (0.97)
	M	0.8 (1.1)	1.5 (0.8)	4.9 (2.0)	5.5 (1.5)	5.1 (1.33)	4.9 (1.20)
4	PB	5.5 (4.3)	5.3 (2.4)	4.2 (1.2)	3.1 (0.7)	2.0 (0.13)	2.8 (0.82)
	M	0.8 (1.1)	1.7 (0.9)	5.0 (1.9)	5.5 (1.4)	5.1 (1.26)	4.9 (1.17)
5	PB	6.3 (4.8)	6.1 (2.8)	5.0 (1.5)	3.6 (0.9)	2.7 (0.60)	2.0 (0.03)
	M	0.8 (1.0)	1.9 (1.0)	5.0 (2.0)	5.5 (1.5)	5.0 (1.27)	4.8 (1.17)

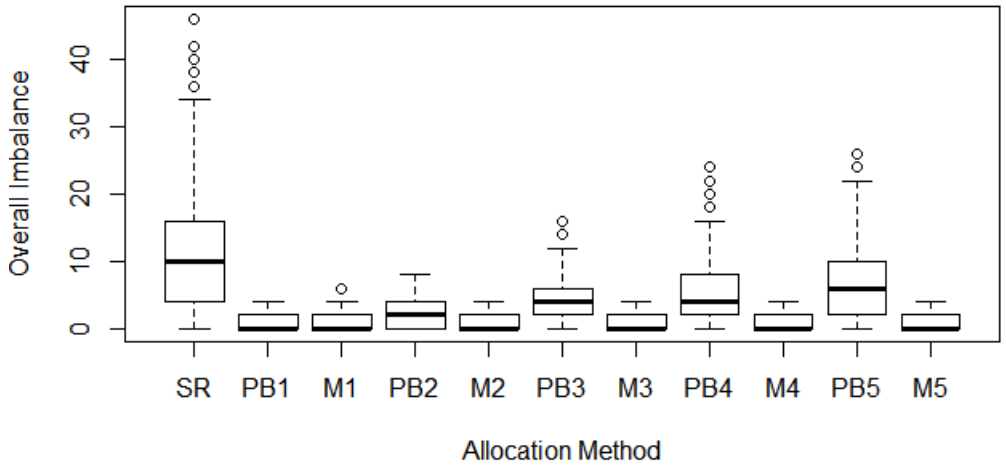


FIGURE 3.7: Overall imbalance, based on 1000 trials of size 200. The numbers attached to allocation methods represent the set of prognostic factors taken in to account (1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s).

4

Discussion and Conclusion

In this project, three treatment allocation methods, Simple Randomization (SR), Permuted Blocks within Strata (PB) and Minimization (M), were compared in respect to the power of log-rank test (asymptotic and re-randomization), the bias in treatment effect estimation, predictability of treatment allocation, and imbalance. The conclusion was that there is only a slight gain in power for PB and M as compared to SR. This gain appears to be more pronounced in trials with small sample sizes and almost negligible in large trials. PB tends to yield higher power as compared to M when few prognostic factors are used, while for large number of factors M performs better. Accounting for prognostic factors in the test leads to a considerable gain in power, but it reaches a point where adding more factors does not increase power or even reduces it. As seen in the results, the number of prognostic factors to account for in the analysis depends on sample size, with bigger trials tending to benefit more from inclusion of more factors. In general, both asymptotic and re-randomization log-rank tests lead to similar power, although re-randomization tends to be more powerful when the test statistic is not stratified. Although an acceptable stability of p-value is achieved with 500 re-randomizations, a great p-value stability is attained with around 5000 re-randomizations or more. The overall bias is similar in three allocation methods, with the variability being higher for SR as compared to PB and M. The bias variability is lower for PB as compared to M when few prognostic factors are used, and higher otherwise. Regarding predictability, SR is the least predictable, followed by PB, with M being the more predictable. The awareness of the allocation algorithm by the guesser increases considerably the predictability for PB and M, with SR not being affected. There is a considerable gain in balance with PB and M as compared to SR. M is a clear winner when it comes to overall and marginal balance, while PB is the best performer for within strata balance.

The low gain in power on using prognostic factors when allocating treatment might be due to fact that SR (which ignores the factors) tend to have less imbalance when the sample size is large and, as stated by Lachin et al. (1988), the effects of treatment imbalances on power are trivial unless the imbalances are substantial. It is argued that taking in to account the prognostic factors in randomization is mainly advantageous for a small trial and has negligible advantages in terms of power or efficiency in a large trial, say $N > 100$ (Lachin et al., 1988).

It is not clear whether it is always the case that asymptotic and re-randomization tests lead to similar power, as found in this project. Edgington and Onghena (2007) state that the p-values provided by the two tests can sometimes be very similar or different, and this depends on many factors. Some of the factors they mention are the absolute sample sizes, the relative sample sizes for the different treatments, the sample distribution shape, the number of tied measurements, and the type of statistical test.

The similarity of overall bias (estimated using unadjusted Cox PH model) across the allocation methods might indicate that using different allocation methods has no major impact on bias. The observed bias might be due to other reasons. One of the reasons might be the lack of adjustment for prognostic factors in the Cox PH model. Even in randomized trials, omitting important prognostic covariates from this model leads to biased estimates of treatment effect (Gail et al., 1984; Hauck et al., 1998). It is also known that Cox PH model produces biased estimates of treatment effect if the proportional hazards assumption is violated (Gandrud, 2015). Although this assumption was met in the original data set, there is no guarantee that this was the case in all simulated trials.

The estimates of predictability for SR and for PB when the guesser is aware of the allocation algorithm are in line with the theoretical results discussed by Antognini and Giovagnoli (2015). The higher predictability of Minimization even when the guesser is not aware of the algorithm might be because of higher marginal balance that it produces in each allocation step. This observation suggests that in a multi-centre trial using minimization, investigators should not be aware of the total number of patients already allocated to each treatment arm, since such knowledge could result in an unacceptably large number of allocations being correctly predicted.

This project had some limitations. There were only 5 prognostic factors in the original dataset and, therefore, the effect of over-stratification in both randomization and testing could not be studied properly. However, in practice, it is unusual that more than 5 factors have a strong prognostic impact on the outcome of interest. In addition, a multivariate score can be calculated that includes as many factors as desired, and the treatment allocation can use this score instead of separate factors. Due to time/resource

constraints, power calculation were based on only 500 trials, with the same number of re-randomization for calculating the p-values of re-randomization test. For similar reasons, many scenarios were not considered. It could have been useful, for instance, to vary block sizes in PB, allocation probabilities in M, and to consider other test statistics, to see how these parameters could impact the results.

Appendix A

Additional Results for Power

TABLE A.1: Hazard ratio estimated from Cox PH model fitted on the original data after "cancelling" treatment effect. The model was stratified by trial and had all baseline characteristics as covariates.

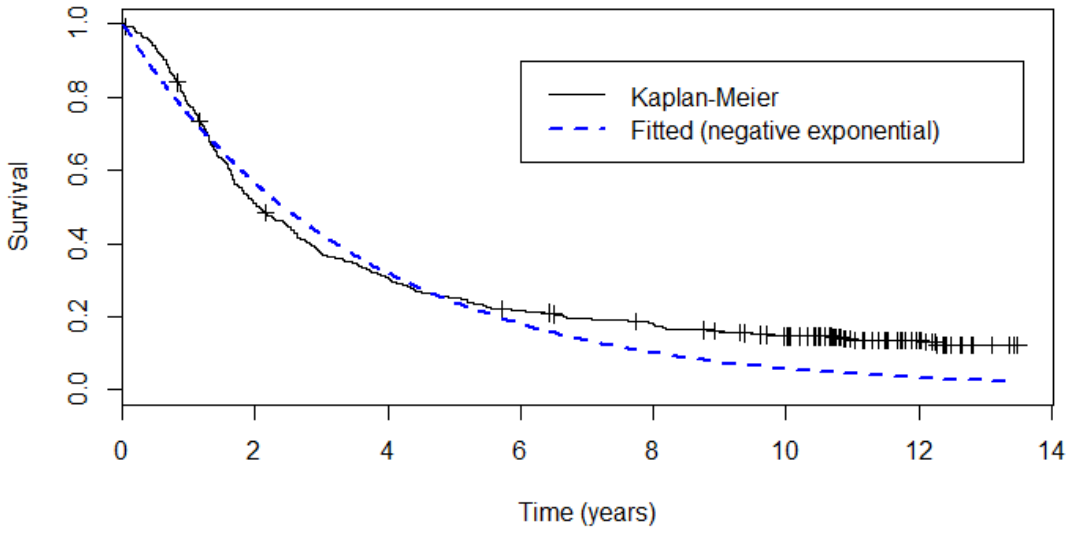
				95% CI Limits	
Covariate	**Reference cat.**	**Category**	**Estimate**	**Lower**	**Upper**
Treatment	CP (Control)	CAP (Exper.)	0.9690	0.8446	1.1120
a	<55	55-65	1.1601	0.9913	1.3580
		>65	1.4767	1.2281	1.7760
p	ECOG 0	ECOG 1	1.4718	1.2529	1.7290
		ECOG 2 or worse	1.9267	1.3899	2.6710
r	No residual disease	≤2 cm	1.4660	1.1678	1.8400
		>2 cm	2.8834	2.1856	3.8040
g	Good	Intermediate	1.3961	1.1303	1.7250
		Poor	1.4932	1.2125	1.8390
s	I	II	0.7911	0.2252	2.7780
		III	0.8311	0.2544	2.7150
		IV	1.0925	0.3291	3.6270

TABLE A.2: Power based on 500 trials of size 100, assuming HR of 0.5, and with 500 re-randomizations for re-randomization tests. URL is unstratified log-rank test, SLR are the stratified ones (the ending cardinal indicates the set of stratifying factors, as specified in column Factor). Their re-randomization counterparts take prefixe R. Allocation methods (Alloc.) are simple randomization (SR), permuted blocks within strata (PB), and Minimization (M). The prognostic factors are age (a), performance status (p), extent of residual disease (r), histological grade (g) and FIGO stage (s). The assumed significance level was $\alpha = 0.05$ for the power calculations. For the allocation methods, blocks of size 4 were used in PB, and a stochastic method of variances (with probability of 0.9) was used for M.

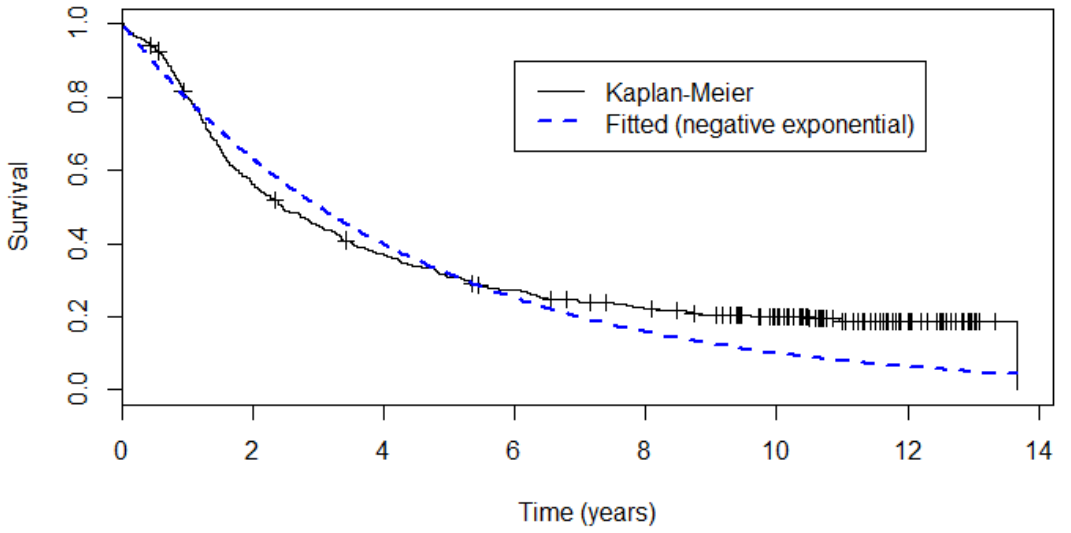
		Tests											
Factors	**Alloc.**	**ULR**	**SLR1**	**SLR2**	**SLR3**	**SLR4**	**SLR5**	**RULR**	**RSLR1**	**RSLR2**	**RSLR3**	**RSLR4**	**RSLR5**
	SR	0.562	0.640	0.678	0.630	0.548	0.496	0.542	0.622	0.658	0.634	0.540	0.488
1 (r)	PB	0.578	0.648	0.664	0.666	0.578	0.528	0.596	0.646	0.652	0.656	0.570	0.524
	M	0.592	0.670	0.666	0.662	0.564	0.514	0.608	0.654	0.642	0.666	0.548	0.510
2 (a,r)	PB	0.560	0.694	0.734	0.772	0.746	0.730	0.634	0.686	0.732	0.766	0.740	0.724
	M	0.578	0.650	0.670	0.658	0.568	0.516	0.610	0.644	0.644	0.656	0.554	0.512
3 (a,r,g)	PB	0.528	0.644	0.670	0.698	0.616	0.560	0.580	0.646	0.664	0.684	0.614	0.548
	M	0.598	0.702	0.694	0.668	0.630	0.582	0.638	0.696	0.700	0.662	0.610	0.564
4 (a,p,r,g)	PB	0.556	0.682	0.690	0.702	0.652	0.630	0.630	0.692	0.694	0.706	0.658	0.616
	M	0.564	0.660	0.668	0.642	0.532	0.498	0.622	0.662	0.652	0.638	0.506	0.482
5 (a,p,r,g,s)	PB	0.592	0.658	0.680	0.700	0.660	0.634	0.616	0.684	0.678	0.692	0.656	0.618
	M	0.558	0.626	0.642	0.624	0.576	0.552	0.606	0.640	0.644	0.606	0.564	0.538

TABLE A.3: Power based on 500 trials of size 500, assuming HR of 0.7, and with 500 re-randomizations for re-randomization tests. URL is unstratified log-rank test, SLR are the stratified ones (the ending cardinal indicates the set of stratifying factors, as specified in column Factor). Their re-randomization counterparts take prefixe R. Allocation methods (Alloc.) are simple randomization (SR), permuted blocks within strata (PB), and Minimization (M). The prognostic factors are age (a), performance status (p), extent of residual disease (r), histological grade (g) and FIGO stage (s). The assumed significance level was $\alpha = 0.05$ for the power calculations. For the allocation methods, blocks of size 4 were used in PB, and a stochastic method of variances (with probability of 0.9) was used for M.

		Tests											
Factors	**Alloc.**	**ULR**	**SLR1**	**SLR2**	**SLR3**	**SLR4**	**SLR5**	**RULR**	**RSLR1**	**RSLR2**	**RSLR3**	**RSLR4**	**RSLR5**
	SR	0.678	0.776	0.816	0.876	0.918	0.944	0.672	0.768	0.820	0.886	0.924	0.936
1 (r)	PB	0.692	0.770	0.790	0.872	0.918	0.934	0.734	0.766	0.780	0.870	0.914	0.930
	M	0.680	0.746	0.784	0.862	0.920	0.938	0.722	0.750	0.786	0.860	0.918	0.938
2 (a,r)	PB	0.666	0.760	0.818	0.878	0.908	0.932	0.724	0.764	0.812	0.872	0.908	0.930
	M	0.674	0.758	0.804	0.870	0.904	0.926	0.728	0.756	0.804	0.864	0.898	0.930
3 (a,r,g)	PB	0.666	0.758	0.808	0.874	0.916	0.938	0.736	0.756	0.806	0.866	0.912	0.936
	M	0.700	0.768	0.804	0.864	0.896	0.926	0.736	0.774	0.798	0.860	0.898	0.922
4 (a,p,r,g)	PB	0.670	0.732	0.792	0.870	0.946	0.956	0.734	0.760	0.802	0.868	0.938	0.952
	M	0.696	0.754	0.798	0.882	0.916	0.928	0.742	0.768	0.804	0.876	0.910	0.924
5 (a,p,r,g,s)	PB	0.688	0.772	0.816	0.886	0.946	0.962	0.740	0.804	0.828	0.900	0.946	0.964
	M	0.678	0.776	0.802	0.894	0.940	0.940	0.752	0.792	0.816	0.900	0.930	0.942

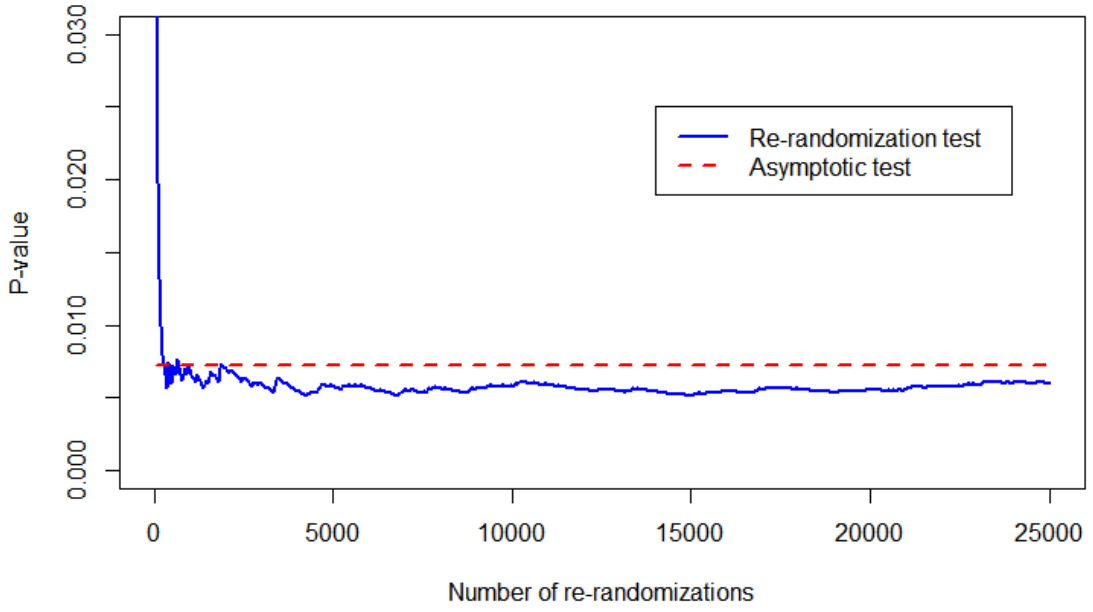


(a) Control group (CP)

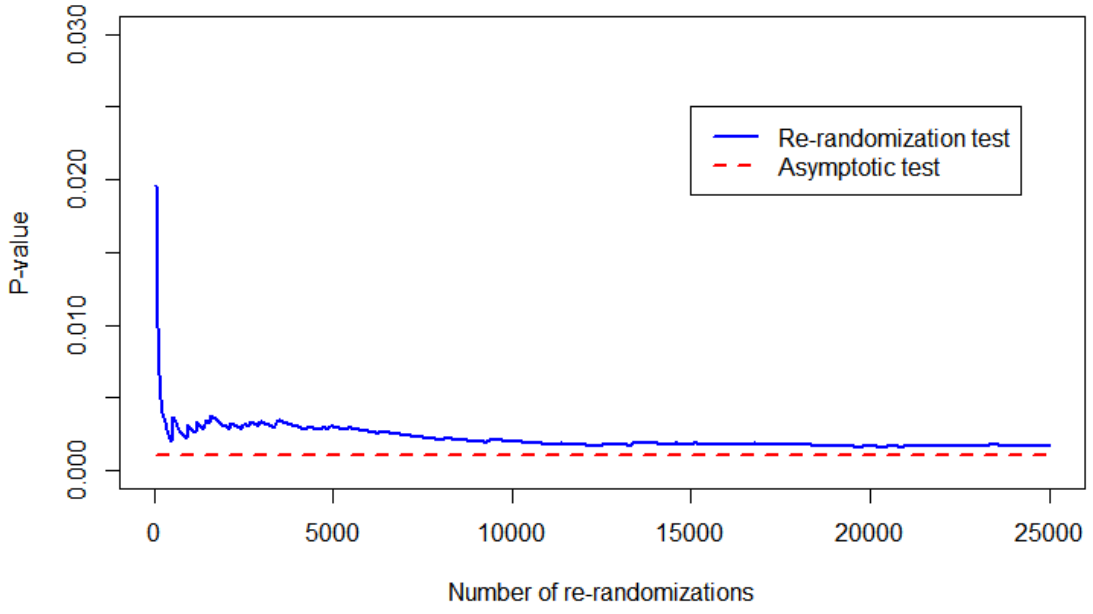


(b) Experimental group (CAP)

FIGURE A.1: Kaplan-Meier and fitted exponential survival. The estimated hazard rate λ from the negative exponential model was 0.2858 in control group and 0.2306 in experimental group.



(a) Unstratified log-rank statistic

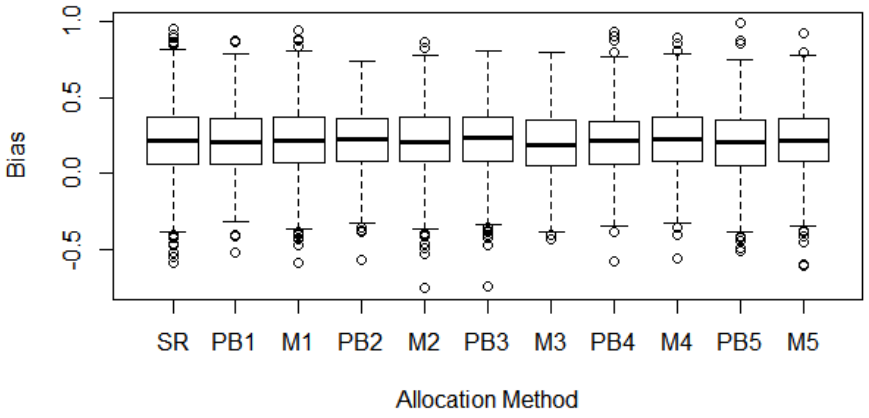


(b) Stratified log-rank statistic

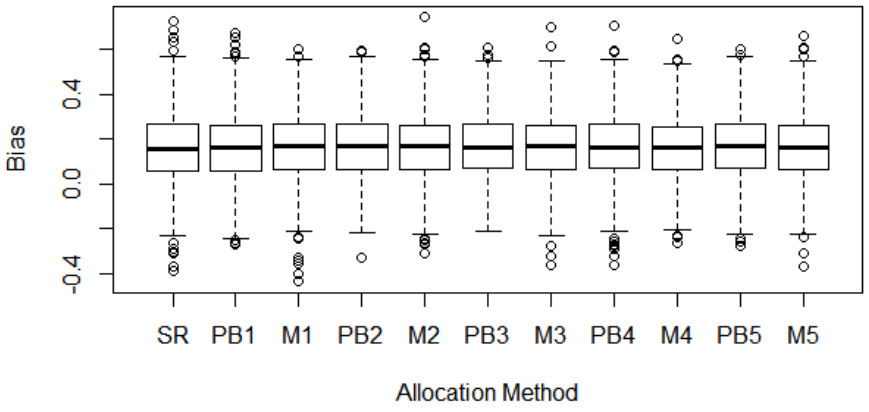
FIGURE A.2: Re-randomization p-value stability – p-value as function of the number of re-randomizations based on a trial of size 200, with HR of 0.6, generated using PB allocation method, with block of size 4. The stratifying factor was the extent of residual disease (r).

Appendix B

Additional Results for Bias



(a) Sample size: 100

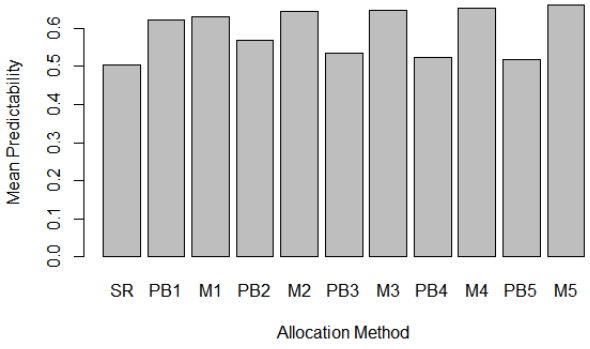


(b) Sample size: 200

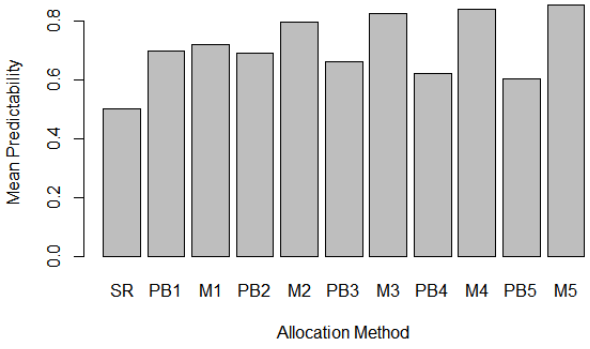
FIGURE B.1: Bias in log HR scale, based on 1000 trials. The allocation methods are simple randomization (SR), permuted blocks within strata (PB) with blocks of size 4, and Minimization (M) with allocation probability of 0.9. The number attached to the abbreviation of allocation method represents the set of stratifying prognostic factors (1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s).

Appendix C

Additional Results for Predictability

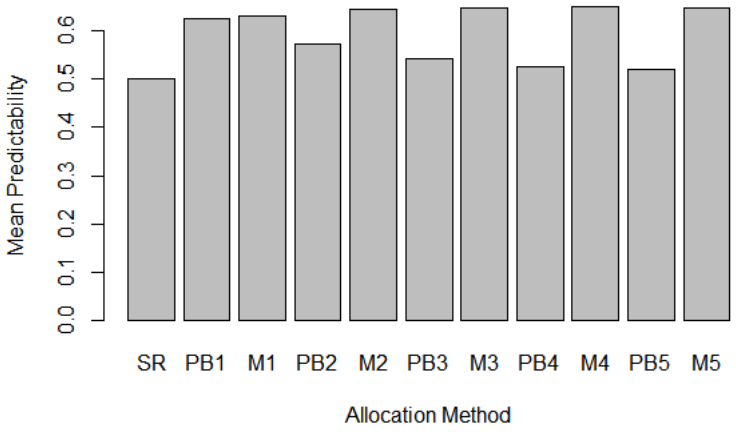


(a) Guesser doesn't know the allocation algorithm

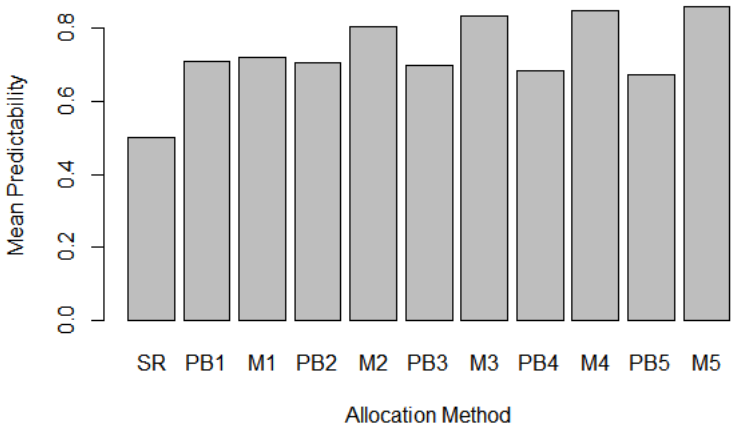


(b) Guesser knows the allocation algorithm

FIGURE C.1: Mean predictability, based on 1000 trials of size 100. The allocation methods are simple randomization (SR), permuted blocks within strata (PB) with blocks of size 4, and Minimization (M) with allocation probability of 0.9. The number attached to the abbreviation of allocation method represents the set of stratifying prognostic factors (1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s).

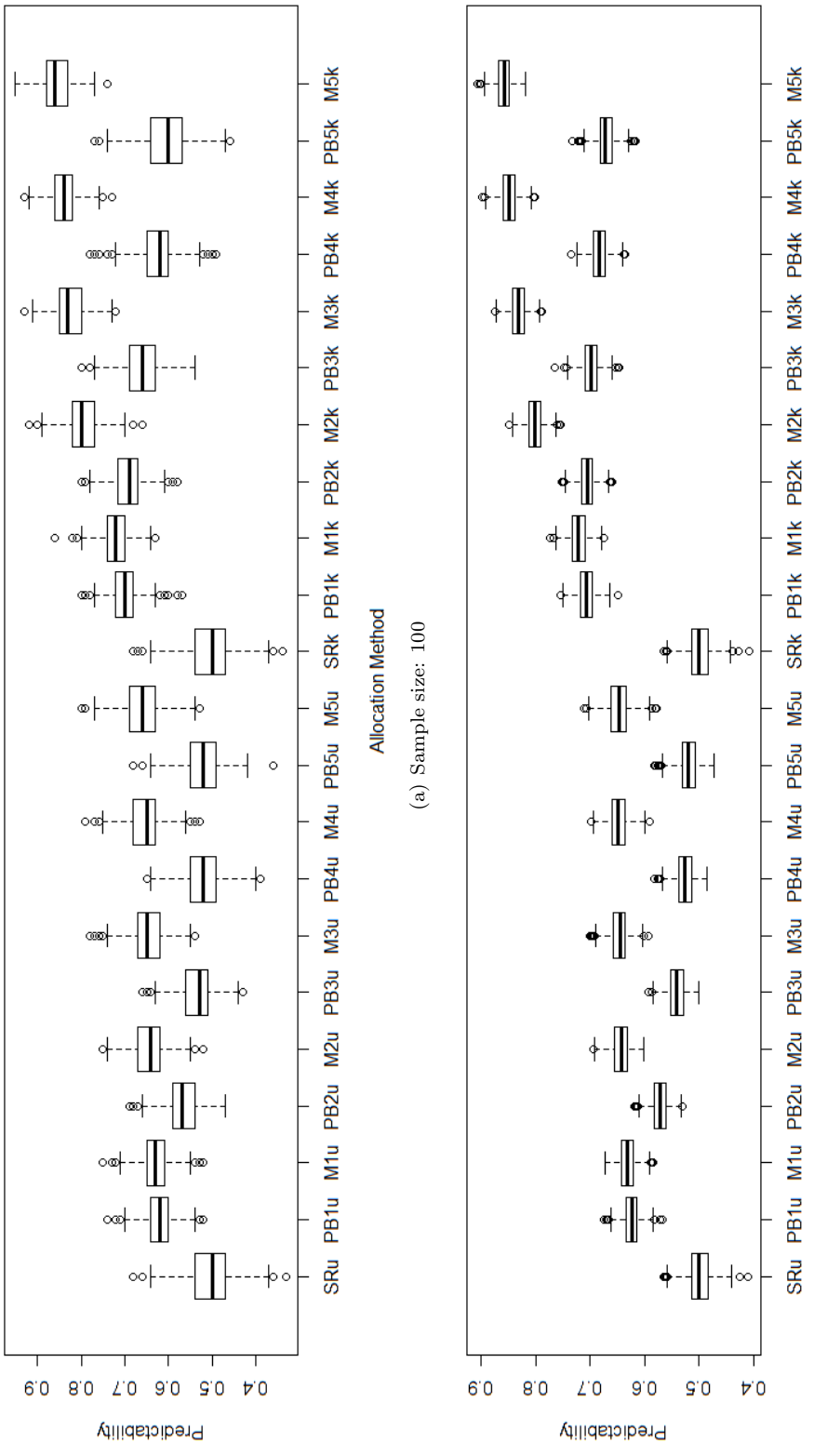


(a) Guesser doesn't know the allocation algorithm



(b) Guesser knows the allocation algorithm

FIGURE C.2: Mean predictability, based on 1000 trials of size 500. The allocation methods are simple randomization (SR), permuted blocks within strata (PB) with blocks of size 4, and Minimization (M) with allocation probability of 0.9. The number attached to the abbreviation of allocation method represents the set of stratifying prognostic factors (1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s).



(a) Sample size: 100

(b) Sample size: 500

FIGURE C.3: Predictability, based on 1000 trials. The allocation methods are simple randomization (SR), permuted blocks within strata (PB) with blocks of size 4, and Minimization (M) with allocation probability of 0.9. The number attached to the abbreviation of allocation method represents the set of stratifying prognostic factors (1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s), and the ending letter indicates whether the guesser knows (k) or doesn't know (u) the allocation algorithm.

Appendix D

Additional Results for Imbalance

TABLE D.1: Imbalance, based on 1000 trials of size 100. **F** represents the set of prognostic factors taken in to account (0: none, 1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s). **A** is the treatment allocation method. **I** stands for imbalance, with the number attached to it representing the set of prognostic factors (**F**) defining the strata in which imbalances were calculated. For instance, **I0** is the overall imbalance (no factor).

		Mean (SD)					
F	**A**	**I0**	**I1**	**I2**	**I3**	**I4**	**I5**
0	SR	7.7 (6.2)	7.7 (3.6)	6.4 (2.2)	5.0 (1.4)	4.3 (1.15)	4.0 (1.04)
1	PB	1.1 (1.2)	1.1 (0.6)	4.8 (1.7)	4.5 (1.2)	4.0 (1.02)	3.7 (0.98)
	M	1.1 (1.2)	1.0 (0.7)	4.8 (1.7)	4.5 (1.3)	4.0 (1.04)	3.8 (0.99)
2	PB	2.1 (1.7)	2.1 (1.0)	1.6 (0.5)	3.7 (1.0)	3.5 (0.90)	3.4 (0.85)
	M	0.9 (1.1)	1.3 (0.8)	3.4 (1.3)	4.2 (1.2)	3.8 (0.96)	3.6 (0.92)
3	PB	3.7 (2.9)	3.5 (1.6)	2.8 (0.8)	1.9 (0.3)	2.7 (0.70)	2.8 (0.73)
	M	0.9 (1.1)	1.6 (0.8)	3.7 (1.4)	3.8 (1.0)	3.6 (0.92)	3.5 (0.89)
4	PB	5.0 (4.0)	4.9 (2.3)	3.9 (1.2)	2.8 (0.7)	2.0 (0.17)	2.2 (0.46)
	M	0.9 (1.1)	1.6 (0.9)	3.6 (1.4)	3.8 (1.0)	3.5 (0.94)	3.4 (0.87)
5	PB	5.6 (4.2)	5.3 (2.4)	4.4 (1.4)	3.2 (0.8)	2.4 (0.56)	2.0 (0.11)
	M	0.8 (1.1)	1.8 (1.0)	3.7 (1.3)	3.9 (1.1)	3.6 (0.92)	3.4 (0.85)

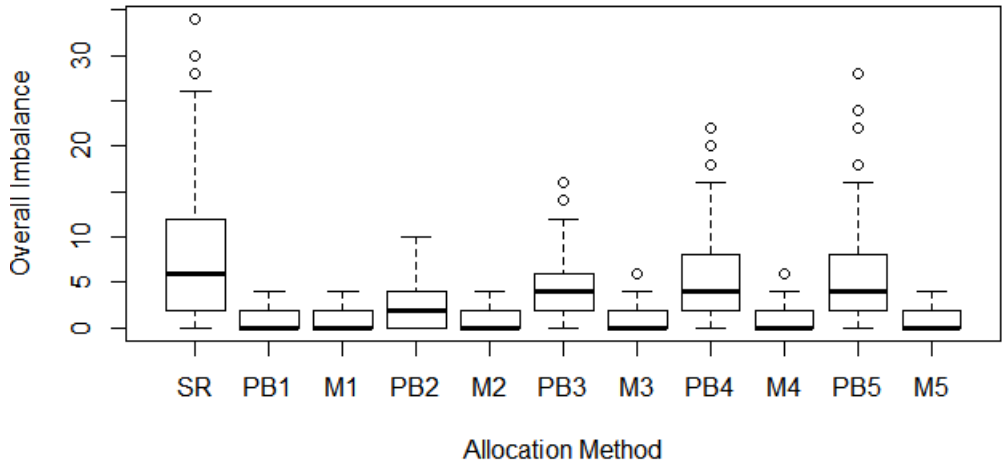


FIGURE D.1: Overall imbalance, based on 1000 trials of size 100. The numbers attached to allocation methods represent the set of prognostic factors taken in to account (1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s).

TABLE D.2: Imbalance, based on 1000 trials of size 500. **F** represents the set of prognostic factors taken in to account (0: none, 1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s). **A** is the treatment allocation method. **I** stands for imbalance, with the number attached to it representing the set of prognostic factors (**F**) defining the strata in which imbalances were calculated. For instance, **I0** is the overall imbalance (no factor).

		Mean (SD)					
F	**A**	**I0**	**I1**	**I2**	**I3**	**I4**	**I5**
0	SR	18.0 (13.2)	17.4 (8.1)	14.4 (4.6)	11.2 (3.0)	9.5 (2.42)	8.8 (2.21)
1	PB	1.1 (1.2)	1.1 (0.6)	10.6 (3.9)	10.0 (2.7)	8.8 (2.23)	8.2 (2.06)
	M	1.1 (1.1)	1.0 (0.6)	10.7 (4.0)	10.0 (2.6)	8.8 (2.14)	8.3 (1.97)
2	PB	2.1 (1.7)	2.1 (1.0)	1.5 (0.5)	8.1 (2.2)	7.9 (1.94)	7.6 (1.85)
	M	0.9 (1.1)	1.3 (0.8)	7.5 (3.0)	9.3 (2.5)	8.5 (2.13)	8.0 (1.92)
3	PB	3.5 (2.8)	3.6 (1.7)	2.8 (0.8)	1.9 (0.3)	5.5 (1.37)	5.9 (1.41)
	M	0.9 (1.1)	1.5 (0.8)	7.7 (3.0)	8.5 (2.3)	8.1 (2.01)	7.7 (1.96)
4	PB	5.8 (4.4)	5.6 (2.5)	4.5 (1.3)	3.2 (0.7)	2.0 (0.06)	4.0 (1.11)
	M	0.8 (1.1)	1.7 (0.9)	7.6 (2.9)	8.4 (2.2)	7.9 (1.92)	7.5 (1.78)
5	PB	7.1 (5.5)	6.9 (3.1)	5.6 (1.7)	4.1 (0.9)	3.0 (0.58)	2.0 (0.03)
	M	0.9 (1.1)	1.8 (1.0)	7.8 (3.0)	8.6 (2.4)	8.0 (1.96)	7.5 (1.80)

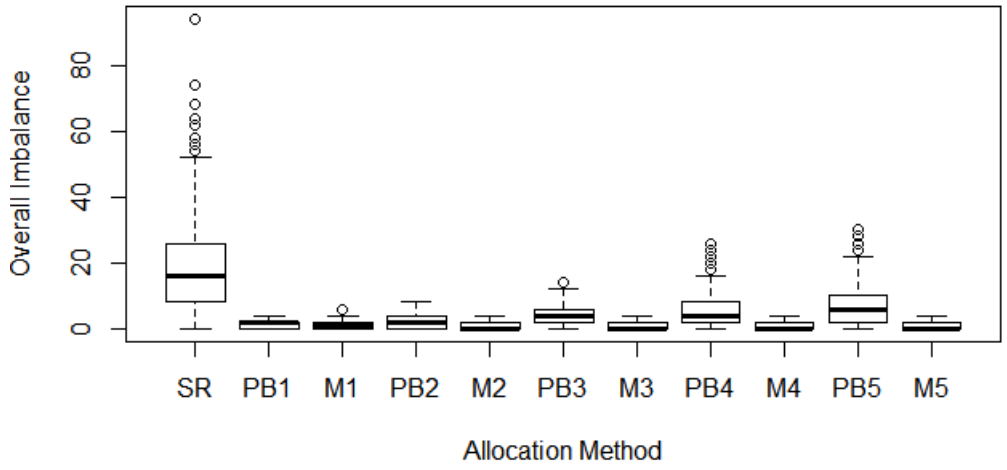


FIGURE D.2: Overall imbalance, based on 1000 trials of size 500. The numbers attached to allocation methods represent the set of prognostic factors taken in to account (1:r; 2:a,r; 3:a,r,g; 4:a,p,r,g; 5:a,p,r,g,s).

Bibliography

Antognini, A. B. and Giovagnoli, A. (2015). *Adaptive Designs for Sequential Treatment Allocation*, volume 73 of *Chapman & Hall/CRC Biostatistics Series*. CRC Press, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.

Bates, D. M. and Watts, D. G. (2007). *Nonlinear Regression Analysis and Its Applications*, volume 696 of *Wiley Series in Probability and Statistics*. John Wiley & Sons, Inc.

Begg, C. B. and Iglewicz, B. (1980). A treatment allocation procedure for sequential clinical trials. *Biometrics*, 36(1):81–90.

Blackwell, D. and Hodges, J. L. (1957). Design for the control of selection bias. *The Annals of Mathematical Statistics*, 28(2):449–460.

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

Buyse, M. (2000). Centralized treatment allocation in comparative clinical trials. *Applied Clinical Trials*, 9(6):32–37.

Cortés, J., González, J. A., Campbell, M. J., and Cobo, E. (2014). A hazard ratio was estimated by a ratio of median survival times, but with considerable uncertainty. *Journal of Clinical Epidemiology*, 67:1172–1177.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society*, 34(2):187–220.

Dubois, C. (2001). Choosing a treatment allocation method for the randomized clinical trial: Results of simulations in advanced ovarian cancer. Master's thesis, Center For Statistics - Limburg Universitair Centrum.

Edgington, E. S. and Onghena, P. (2007). *Randomization Tests*. Chapman & Hall/CRC, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742, fourth edition.

Gabriel, K. R. and Hsu, C.-F. (1981). Evaluation of the power of re-randomization tests, with application to weather modification experiments. Technical Report 81/11, Department of Statistics and Division of Biostatistics, University of Rochester, Rochester, NY 14627 USA.

Gail, M. H., Wieand, S., and Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, 71(3):431–444.

Gandrud, C. (2015). simph: An r package for illustrating estimates from cox proportional hazard models including for interactive and nonlinear effects. *Journal of Statistical Software*, 65:1–20.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81:515–526.

Han, B., Yu, M., and McEntegart, D. (2013). Weighted re-randomization tests for minimization with unbalanced allocation. *Pharmaceutical Statistics*, 12:243–253.

Hauck, W. W., Anderson, S., and Marcus, S. M. (1998). Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Controlled Clinical Trials*, 19:249–256.

Heritier, S., Gebski, V., and Pillai, A. (2005). Dynamic balancing randomization in controlled clinical trials. *Statistics in Medicine*, 24:3729–3741.

Hills, R. K., Gray, R., and Wheatley, K. (2009). Balancing treatment allocations by clinician or center in randomized trials allows unacceptable levels of treatment prediction. *Journal of Evidence-Based Medicine*, 2:196–204.

Lachin, J. M., Matts, J. P., and Wei, L. (1988). Randomization in clinical trials: Conclusions and recommendations. *Controlled Clinical Trials*, 9:365–374.

Machin, D., Campbell, M. J., Tan, S. B., and Tan, S. H. (2009). *Sample Size Tables for Clinical Studies*. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK, third edition.

Machin, D., Cheung, Y. B., and Parmar, M. K. (2006). *Survival Analysis: A Practical Approach*. John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, second edition.

McEntegart, D. J. (2003). The pursuit of balance using stratified and dynamic randomization techniques: An overview. *Drug Information Journal*, 37(3):293–308.

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.

Morgan, K. L. and Rubin, D. B. (2012). Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 40(2):1263–1282.

Peace, K. E. (2009). *Design and Analysis of Clinical Trials with Time-to-Event Endpoints*. Chapman & Hall/CRC Biostatistics Series. Chapman & Hall/CRC, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.

Perry, M., Faes, M., Reelick, M. F., Rikkert, M. G. O., and Borm, G. F. (2010). Study-wise minimization: A treatment allocation method that improves balance among treatment groups and makes allocation unpredictable. *Journal of Clinical Epidemiology*, 63:11181122.

Pocock, S. J. and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, 31(1):103–115.

Pond, G. (2011). Statistical issues in the use of dynamic allocation methods for balancing baseline covariate. *British Journal of Cancer*, 104:1711–1715.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Taves, D. (1974). Minimization: A new method of assigning patients to treatment and control groups. *Clinical Pharmacology and Therapeutics*, 15(5):443–453.

The Ovarian Cancer Meta-Analysis Project (1991). Cyclophosphamide plus cisplatin versus cyclophosphamide, doxorubicin, and cisplatin chemotherapy of ovarian carcinoma: a meta-analysis. *Journal of Clinical Oncology*, 9:1668–1674.

Wei, L.-J. (1977). A class of designs for sequential clinical trials. *Journal of the American Statistical Association*, 72(358):382–386.

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Using prognostic factors in the design and analysis of clinical trials: a simulation study

Richting: **Master of Statistics-Biostatistics**
Jaar: **2015**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Nhacolo, Arsenio

Datum: **1/09/2015**