

# 1 *Background*

## 1.1 Randomization in clinical trials

Research to identify effective and/or efficacious interventions requires balance between scientific and logistic constraints. Randomization in clinical trials allows one to infer causation from associations in the presence of an appropriate experimental design, appreciating the limitations in identifying true causal relationships. Complete randomization also ensures all variables, particularly those predictive of the outcome measurement (referred to as prognostic factors) have balanced distributions across groups, on average. However, the likelihood of imbalance increases as the trial sample size decreases or the number of measured prognostic variables increases.

Especially in small studies, at any given point in the randomization process there could be a substantial imbalance in the number of patients assigned to each group. Chance imbalance in important prognostic factors may be seen as impacting the credibility of observed treatment effect estimates. For this reason, restricted randomization techniques are often used to guarantee the sequential and overall imbalance in the number of participants in each group is controlled. We introduce restricted randomization and discuss their extension to control imbalance in prognostic factors by means of stratification or covariate adaptive randomization.

## 1.2 Restricted randomization

Blocked randomization is a widely accepted restricted randomization method that ensures current and overall imbalance in treatment assignments are controlled. Sequential balance is achieved by specifying a block size for which the sequence of (usually equal numbers of) treatment assignments is permuted and assigned to patients as they are enrolled. However, blocked randomization does not by itself guarantee overall balance in known prognostic variables of interest. For this reason, blocking is

often combined with stratification to make study groups comparable with regard to specified stratifying factors.

Stratified randomization is a procedure that separates the recruitment population into smaller subgroups (strata) where randomization, either complete or blocked, is performed. This property can be useful in multi-center trials, for instance, where it is of interest to account for between-center variability in patient outcomes due to unmeasured or unimportant factors *a priori*. While the randomization ratio is guaranteed to hold within pre-specified blocks of enrolled subjects, in small trials with many stratification factors one cannot assure accrued patients will fill the block for each subgroup, and randomization within strata alone will not ensure balance.

### 1.3 Adaptive randomization

A randomization method is static if the probability of treatment assignment is not conditional on information on patients already enrolled. In contrast, adaptive (or dynamic) randomization approaches control imbalance by dynamically altering the randomization probability based on accrued patient information. In this thesis we consider covariate adaptive randomization procedures, which are a natural comparison to static randomization strategies intended to control imbalance of baseline prognostic factors across treatment groups. These procedures have been increasingly used as an alternative to stratified block randomization, particularly in small scale clinical trials with many prognostic factors.

Initial developments in covariate adaptive methods aimed to reduce the probability of undesirable, albeit unlikely, allocation sequences which result in both overall treatment group imbalances and imbalances within subgroups defined by important prognostic factors. In this subsection we follow the historical development of covariate adaptive

approaches with a brief discussion of the characteristics and performance of a few selected methods.

Biased coin randomization introduced by Efron (1971) was the first randomization method to change the probability of assignment dynamically based on observed covariate values of accrued patients. Simple randomization is performed until the disparity reaches a prespecified limit, at which time the group with the least subjects is biased to have a greater probability of assignment.

Taves (1974) extended Efron's biased coin design to the context of small scale clinical trials, where it is of interest to constrain imbalance in multiple prognostic factors across treatment groups. Briefly, the method sequentially allocates incoming patients deterministically to the treatment category that minimizes the overall unweighted sum of covariate imbalance given the new assignment. The assignment is performed deterministically: assignment is randomized only when allocation to either treatment category results in the same imbalance. Pocock and Simon (1975) further generalized Taves' method to incorporate relative importance of prognostic factors by introducing weighting of covariate imbalances into the overall imbalance metric.

Signorini et. al (1993) extended earlier methods in order to induce balance both overall and within strata while avoiding investigator bias through unblinding. He proposed a tree-based method of dynamic balancing randomization (DBR) that evaluates imbalance for each prognostic factor in a nested fashion by their prespecified order of importance. The method flexibly allows for different levels of imbalance in different strata and ensures conditional balance, meaning that within each subgroup the ratio of treatment assignment is constrained within prespecified bounds. However, the method does not guarantee balanced group assignments will be achieved within each prognostic factor considered separately.

Heritier et. al (2005) modify Signorini et. al's DBR method to instead control

imbalance marginally within each prognostic factor. For each accrued patient, the potential imbalance for each treatment assignment is considered sequentially within each prognostic factor in decreasing order of importance. If the potential observed imbalance exceeds a prespecified threshold, assignment is performed deterministically (or forced) to the group which minimizes the imbalance. Heritier et. al suggested including non-deterministic allocation to reduce the number of forced allocations and prevent investigator unblinding.

Model-based approaches are another alternative approach to dynamic randomization, where the probability of treatment assignment is chosen to minimize the variance of the estimated treatment effect. Model-based methods can flexibly incorporate continuous prognostic factors without the need to dichotomize into groups, and can include interaction terms and balance prognostic factors even when the number of variables is large. Aickin (1998, 2001, 2009) proposed a model-based approach to covariate adaptive randomization, where a subjects' treatment assignment is based on maximizing the log-likelihood of the model.

## 1.4 Adjusting for prognostic factors in the analysis model

Statistical adjustment is another approach to addressing prognostic factors that may confound the relationship between an intervention (the predictor of interest) and the outcome measure. Oftentimes it is of interest to measure and adjust for known variables predictive of the outcome measurement in the analysis model. It may be scientifically meaningful to provide estimates for known confounders or if the effect may differ within subgroups. Statistically, adjustment can potentially reduce bias from confounding and increase precision. Statistical adjustment accounts for, but

does not control, imbalance in known prognostic factors predictive of the outcome measurement.

## 1.5 Analysis considerations under alternative randomization schemes

Analysis of trials using a covariate adaptive allocation (CAA) scheme must account for the randomization scheme to recover the precision gains conferred by inducing more balanced treatment groups with respect to chosen balancing factors. To obtain the correct variance term and significance level for the test statistic, one must consider all possible sequences of assignments which could have been made in repeated trials assuming no group differences in mean response. In most cases, ignoring the randomization procedure and using standard regression methods that implicitly assume complete randomization lead to larger variance estimates and conservative inference. Since CAA modifies the randomization scheme to induce similarity across treatment arms relative to within arms, the efficiency gain can be realized using a nonparametric re-randomization approach for estimating standard errors (Simon and Simon 2011). Briefly, observed values and entry order are fixed, treatment assignments are reshuffled and the test statistic computed for each permutation.

## 1.6 Aims

The goal of the thesis is to address in both the binary and continuous outcome setting if covariate adaptive allocation (CAA) followed by standard asymptotic tests yield valid inference, and if so, to quantify the gains in precision relative to complete randomization (SR) or stratified block randomization (SBR). We will compare Heritier's

modified DBR scheme to stratified block randomization and complete randomization, while comparing re-randomization based permutation tests to standard asymptotic tests in a simulation study. We will consider the setting of equal allocation to treatment assignment, no temporal trend (drift), binary predictors, and two outcome types (binary and continuous). Our objective is to identify any scenarios, if any, where minimization improves power relative to SBR or complete randomization. Contour plots of effect size by sample size will compare power across methods for various outcome types and conditions.

We are also interested if and when the answers to the above questions change when the effect size of prognostic factors is varied relative to the treatment effect, the baseline outcome prevalence varies from 10% to 50% in the binary outcome setting, inference on treatment effect is performed adjusting for none or all of the prognostic variables, and when the sample size is varied. It is well known that ignoring the minimization design tends to yield conservative inference, and that adjusting for covariates used in the randomization scheme (balancing factors) recovers type I error rate to nominal significance levels (Xu, Proschan, Lee 2016). Through comparison of estimated marginal and conditional treatment effects we seek to confirm this finding. We consider different sample sizes ranging from  $n=32$  to 96 to compare CAR to SBR as small scale trials are the setting in which alternative randomization methods are considered. We seek to identify the specific conditions by which CAR confers a precision advantage, if any, relative to other methods to offset the operational complexities involved in implementing an adaptive allocation procedure. Our intent is to provide guidance to clinical researchers for determining under what settings covariate adaptive allocation provides precision gains relative to competing approaches as well as which analysis method yields valid tests with the most power.

Chapter 2 will introduce the notation used throughout the thesis and discuss the

design of the simulation study in further detail. The tables of simulation results will be presented in Chapter 3 and the key observations will be discussed in Chapter 4.

## 1.7 Measures to evaluate aims

For each combination of randomization scheme and analysis approach, we assess validity by estimating the nominal significance level of the test under the null hypothesis. We evaluate precision by estimating power as a function of the true treatment effect size.

We evaluate accuracy by estimating any potential bias for all methods. We also estimate coverage probability of confidence intervals for model-based analysis approaches. However, we cannot evaluate coverage probability for permutation test quantile-based confidence intervals as they are estimated under the null hypothesis.

## 2 *Methods*

## 2.1 Data generation

Simulations were conducted of a two-arm randomized clinical trial with equal allocation (1:1 treatment:control). We conducted simulations for both continuous and binary outcome and covariate types, and varied the overall trial size ( $n$ ) from 32 to 96.

The outcome measure ( $Y$ ) was simulated with a marginal prevalence of 10% or 50% in the binary setting to evaluate the potential impact of low numbers of observed outcomes on inference. Continuous outcomes were simulated as normally distributed with constant variance, with mean as a linear combination of the treatment assignment ( $Z$ ) and pre-specified prognostic factors ( $X$ ).

Binary risk factors were simulated such that their marginal prevalence was either 25% or 50%. Continuous risk factors were generated under a standard normal distribution. The prognostic factors are modeled as independent. The effect sizes for treatment assignment ( $Z$ ) and prognostic variables ( $X$ ) were separately varied from none, low, medium, and high. Balancing factors refer to prognostic factors used in adaptive allocation procedures, for which it is desired to have comparability either within or between treatment groups. The exact type of balance desired informs the choice of imbalance metric minimized at each sequential allocation step. For instance, it may be of interest to ensure within-strata subgroups have approximately proportional treatment and control assignments (conditional balance), or that treatment groups are otherwise comparable with respect to pre-specified balancing factors (marginal balance).

Observed patient entry times occurred following a uniform distribution. In subsequent re-randomization analysis, patient entry order is considered fixed (see last subsection entitled 'Re-randomization inference').

## 2.2 Allocation procedures

For each simulated set of observed entry times and prognostic factors, treatment group assignments were determined using three allocation procedures: complete randomization, stratified permuted block randomization with fixed block sizes,<sup>1</sup> and an adapted form of covariate adaptive allocation proposed by Heritier et. al (2005).

For the covariate adaptive allocation procedure, the maximum imbalance of treatment to control assignments (overall and within strata defined by each balancing factor level, considered separately) was set to 2. The allocation biasing probability, or the probability of assigning patient to treatment minimizing the imbalance measure when prospective imbalance meets or exceeds a prespecified threshold, was chosen as 0.7 to minimize the effect of non-deterministic allocations on inference.

## 2.3 Varied conditions

The following table describes the conditions varied in the binary outcome setting.

Table 2.1: Simulation study conditions

Variable (notation)	Description	Associated parameters	Cardinality
Response (Y)	Outcome, binary or continuous	Marginal prevalence $p_Y \equiv Pr(Y = 1) = \{0.1, 0.5\}$	2 dichotomous, 1 continuous = 3
Treatment assignment (Z)	$Z = \{0, 1\}$	$p_{alloc} \equiv Pr(Z = 1) = 0.5$	1
Prognostic factors (X)	Binary: $X = \{0, 1\}$ Continuous: dichotomized by pop. median	$p_X \equiv Pr(X = 1) = \{0.25, 0.5\}$	2 dichotomous, 1 continuous = 3
Trial size (n)	Overall sample size	$n = \{32, 96\}$	2
Effect sizes ( $\beta_X$ and $\beta_Z$ ) <sup>2</sup>	$\beta = \begin{cases} \text{None: } \exp(\beta) = 1 \\ \text{Low: } \exp(\beta) = 1.1 \\ \text{High: } \exp(\beta) = 3 \end{cases}$	$\beta_Z = \{\log(1.0), \log(1.1), \log(3)\}$ $\beta_X = \{\log(1.1), \log(3)\}$	$3^*2 = 6$

<sup>1</sup>Block sizes are equal to overall trial size divided by number of strata, which are defined by all combinations of balancing factor levels.

<sup>2</sup>In the continuous outcome setting, the effect sizes are modified to represent comparable differences in means to the given odds ratios.

## 2.4 Analysis approaches

For each allocation procedure we estimated the treatment effect, adjusted and unadjusted for prognostic factors. To evaluate power, coverage, and type I error control, we report the associated linear and logistic regression model-based p-values and Wald-type confidence intervals, based on a two-sided type I error threshold of 0.05. For balancing purposes, continuous prognostic factors were first dichotomized by their population median, and later parameterized as continuous in the adjusted analysis.

The simulation model is of the form

$$g(E[Y_i|Z_i, \mathbf{X}_i]) = \beta_0 + \beta_Z \cdot Z_i + \beta_X \cdot \mathbf{X}_i. \quad (2.1)$$

the adjusted regression model is,

$$g(E[Y_i|Z_i, \mathbf{X}_i]) = \beta_0 + \beta_Z \cdot Z_i + \beta_X \cdot \mathbf{X}_i, \quad (2.2)$$

and the unadjusted model is

$$g(E[Y_i|Z_i]) = \gamma_0 + \gamma_Z \cdot Z_i, \quad (2.3)$$

where  $g()$  is the identity link when  $Y$  is continuous and the logit link when  $Y$  is binary.

### Re-randomization inference

To compare the bias and power based on re-randomization analysis to that based on standard regression techniques, we compute power, coverage, and level and conducted re-randomization based inference for each simulated trial following covariate adaptive allocation. Re-randomization is a permutation-based method for estimating uncertainty and follows from the generally accepted sentiment to ‘analyze as you randomize’. The approach considers the outcomes, prognostic factors, and observed entry time

as fixed and repeats the allocation procedure multiple times, each generating a new sequence of treatment assignments. Regression estimates are computed under each re-randomized treatment allocation sequence, and re-randomization based 95% confidence intervals are generated using the 2.5th and 97.5th quantile of the re-randomization-based treatment effect estimates. Re-randomization based p-values are estimated using the observed proportion of re-randomized allocation sequences yielding treatment effect estimates as or more extreme than the observed treatment effect.

Each simulation model configuration was simulated 5,000 trials, for which the re-randomization procedure was repeated 500 times.

### 3 *Simulation*

### 3.1 Batch 1: Binary Outcome, Binary Predictors

Table 3.1: Batch 1 (Binary Y, Binary X): Power

n	Pr( Y )	Pr( X )	bZ	bX	Model-based				Rerandomization		
					CR		SBR		CAA		CAA
					adj	unadj	adj	unadj	adj	unadj	adj
32	0.1	0.25	1.0	1.1	0.009	0.001	0.006	0.000	0.007	0.000	0.075
32	0.1	0.25	1.0	3.0	0.015	0.001	0.017	0.000	0.018	0.001	0.071
32	0.1	0.25	1.1	1.1	0.009	0.001	0.007	0.000	0.008	0.000	0.077
32	0.1	0.25	1.1	3.0	0.016	0.001	0.017	0.000	0.020	0.001	0.070
32	0.1	0.25	3.0	1.1	0.015	0.004	0.012	0.003	0.013	0.003	0.147
32	0.1	0.25	3.0	3.0	0.024	0.007	0.025	0.004	0.026	0.007	0.157
32	0.1	0.50	1.0	1.1	0.009	0.000	0.008	0.000	0.005	0.000	0.078
32	0.1	0.50	1.0	3.0	0.018	0.001	0.022	0.001	0.020	0.001	0.070
32	0.1	0.50	1.1	1.1	0.009	0.001	0.007	0.000	0.004	0.000	0.079
32	0.1	0.50	1.1	3.0	0.019	0.002	0.021	0.000	0.022	0.001	0.071
32	0.1	0.50	3.0	1.1	0.015	0.004	0.009	0.002	0.013	0.004	0.149
32	0.1	0.50	3.0	3.0	0.042	0.007	0.040	0.003	0.045	0.006	0.165
32	0.5	0.25	1.0	1.1	0.028	0.029	0.031	0.026	0.028	0.025	0.057
32	0.5	0.25	1.0	3.0	0.019	0.023	0.022	0.017	0.025	0.022	0.056
32	0.5	0.25	1.1	1.1	0.025	0.027	0.027	0.022	0.033	0.029	0.063
32	0.5	0.25	1.1	3.0	0.023	0.025	0.024	0.017	0.026	0.022	0.059
32	0.5	0.25	3.0	1.1	0.210	0.230	0.232	0.230	0.222	0.219	0.326
32	0.5	0.25	3.0	3.0	0.172	0.192	0.197	0.181	0.198	0.182	0.302
32	0.5	0.50	1.0	1.1	0.026	0.029	0.026	0.025	0.026	0.025	0.062
32	0.5	0.50	1.0	3.0	0.022	0.025	0.028	0.016	0.024	0.019	0.059
32	0.5	0.50	1.1	1.1	0.029	0.030	0.024	0.023	0.024	0.023	0.064
32	0.5	0.50	1.1	3.0	0.027	0.031	0.028	0.018	0.028	0.022	0.061
32	0.5	0.50	3.0	1.1	0.202	0.226	0.223	0.223	0.223	0.228	0.327
32	0.5	0.50	3.0	3.0	0.161	0.177	0.190	0.164	0.180	0.170	0.300
96	0.1	0.25	1.0	1.1	0.026	0.018	0.024	0.019	0.026	0.018	0.058
96	0.1	0.25	1.0	3.0	0.038	0.026	0.041	0.026	0.036	0.019	0.051
96	0.1	0.25	1.1	1.1	0.025	0.019	0.027	0.020	0.028	0.018	0.063
96	0.1	0.25	1.1	3.0	0.038	0.027	0.040	0.025	0.039	0.022	0.055
96	0.1	0.25	3.0	1.1	0.226	0.210	0.213	0.205	0.224	0.207	0.358
96	0.1	0.25	3.0	3.0	0.270	0.245	0.274	0.241	0.278	0.242	0.376
96	0.1	0.50	1.0	1.1	0.024	0.018	0.025	0.017	0.026	0.018	0.060

96	0.1	0.50	1.0	3.0	0.045	0.024	0.044	0.021	0.048	0.026	0.061
96	0.1	0.50	1.1	1.1	0.025	0.019	0.023	0.018	0.027	0.019	0.061
96	0.1	0.50	1.1	3.0	0.047	0.026	0.045	0.021	0.049	0.025	0.060
96	0.1	0.50	3.0	1.1	0.217	0.212	0.219	0.207	0.232	0.218	0.363
96	0.1	0.50	3.0	3.0	0.308	0.267	0.328	0.277	0.318	0.267	0.394
96	0.5	0.25	1.0	1.1	0.043	0.045	0.044	0.048	0.038	0.045	0.050
96	0.5	0.25	1.0	3.0	0.046	0.047	0.045	0.041	0.043	0.041	0.055
96	0.5	0.25	1.1	1.1	0.046	0.049	0.049	0.052	0.046	0.052	0.057
96	0.5	0.25	1.1	3.0	0.050	0.051	0.051	0.048	0.048	0.047	0.060
96	0.5	0.25	3.0	1.1	0.725	0.744	0.730	0.745	0.741	0.764	0.752
96	0.5	0.25	3.0	3.0	0.680	0.649	0.691	0.675	0.683	0.672	0.704
96	0.5	0.50	1.0	1.1	0.044	0.047	0.045	0.050	0.048	0.053	0.059
96	0.5	0.50	1.0	3.0	0.042	0.048	0.042	0.036	0.046	0.044	0.058
96	0.5	0.50	1.1	1.1	0.045	0.047	0.049	0.054	0.050	0.056	0.059
96	0.5	0.50	1.1	3.0	0.046	0.048	0.047	0.042	0.046	0.040	0.057
96	0.5	0.50	3.0	1.1	0.725	0.741	0.740	0.757	0.739	0.758	0.752
96	0.5	0.50	3.0	3.0	0.661	0.629	0.676	0.647	0.669	0.648	0.694

Table 3.2: Batch 1 (Binary Y, Binary X): Median bias

n	Pr( Y )	Pr( X )	bZ	bX	Model-based				Rerandomization	
					CR		SBR		CAA	
					adj	unadj	adj	unadj	adj	unadj
32	0.1	0.25	1.0	1.1	0.000	0.000	0.000	0.000	0.000	0.000
32	0.1	0.25	1.0	3.0	0.000	0.000	0.000	0.000	0.000	0.000
32	0.1	0.25	1.1	1.1	-0.019	-0.044	-0.095	-0.095	-0.065	-0.095
32	0.1	0.25	1.1	3.0	-0.007	0.038	-0.045	-0.095	-0.028	-0.095
32	0.1	0.25	3.0	1.1	0.297	0.167	0.196	0.143	0.237	0.143
32	0.1	0.25	3.0	3.0	0.316	0.065	0.227	0.000	0.282	0.000
32	0.1	0.50	1.0	1.1	0.000	0.000	0.000	0.000	0.000	0.000
32	0.1	0.50	1.0	3.0	0.000	0.000	0.000	0.000	0.000	0.000
32	0.1	0.50	1.1	1.1	-0.013	-0.065	-0.095	-0.095	-0.079	-0.095
32	0.1	0.50	1.1	3.0	0.028	0.022	-0.011	-0.095	-0.031	-0.095
32	0.1	0.50	3.0	1.1	0.289	0.143	0.173	0.065	0.284	0.143
32	0.1	0.50	3.0	3.0	0.283	0.065	0.222	0.059	0.229	0.059
32	0.5	0.25	1.0	1.1	0.012	0.000	0.000	0.000	0.000	0.000
32	0.5	0.25	1.0	3.0	0.001	0.000	0.000	0.000	-0.003	0.000
32	0.5	0.25	1.1	1.1	0.021	-0.008	0.008	-0.047	0.011	-0.031
32	0.5	0.25	1.1	3.0	0.030	-0.026	0.009	-0.047	-0.006	-0.047
32	0.5	0.25	3.0	1.1	0.131	0.031	0.145	0.031	0.128	-0.011
32	0.5	0.25	3.0	3.0	0.176	-0.049	0.117	-0.059	0.148	-0.059
32	0.5	0.50	1.0	1.1	-0.007	0.000	-0.008	0.000	0.012	0.000
32	0.5	0.50	1.0	3.0	-0.003	0.000	-0.014	0.000	0.000	0.000
32	0.5	0.50	1.1	1.1	-0.002	-0.031	-0.014	-0.047	0.012	-0.047
32	0.5	0.50	1.1	3.0	0.006	-0.047	-0.004	-0.047	-0.010	-0.080
32	0.5	0.50	3.0	1.1	0.136	0.031	0.168	0.031	0.133	0.000
32	0.5	0.50	3.0	3.0	0.144	-0.087	0.153	-0.077	0.149	-0.077
96	0.1	0.25	1.0	1.1	0.004	0.000	0.001	0.000	0.002	0.000
96	0.1	0.25	1.0	3.0	0.001	0.000	-0.002	0.000	-0.005	0.000
96	0.1	0.25	1.1	1.1	0.009	0.002	-0.004	-0.004	-0.007	-0.013
96	0.1	0.25	1.1	3.0	-0.001	-0.004	-0.001	-0.002	0.006	-0.013
96	0.1	0.25	3.0	1.1	0.081	0.047	0.084	0.047	0.077	0.047
96	0.1	0.25	3.0	3.0	0.065	-0.045	0.071	-0.047	0.062	-0.047
96	0.1	0.50	1.0	1.1	0.007	0.000	0.005	0.000	0.000	0.000
96	0.1	0.50	1.0	3.0	-0.001	0.000	0.018	0.000	0.001	0.000
96	0.1	0.50	1.1	1.1	0.009	0.000	0.002	-0.002	-0.009	-0.035
96	0.1	0.50	1.1	3.0	0.006	-0.013	0.017	0.002	0.011	0.000

96	0.1	0.50	3.0	1.1	0.085	0.054	0.070	0.047	0.093	0.054	0.093
96	0.1	0.50	3.0	3.0	0.062	-0.037	0.065	-0.036	0.077	-0.036	0.077
96	0.5	0.25	1.0	1.1	0.013	0.007	0.002	0.000	0.001	0.000	0.001
96	0.5	0.25	1.0	3.0	0.016	0.009	0.000	0.000	0.007	0.000	0.007
96	0.5	0.25	1.1	1.1	0.006	-0.001	-0.004	-0.011	-0.001	-0.011	-0.001
96	0.5	0.25	1.1	3.0	0.021	0.007	-0.002	-0.012	0.015	-0.010	0.015
96	0.5	0.25	3.0	1.1	0.039	0.013	0.037	0.013	0.038	0.013	0.038
96	0.5	0.25	3.0	3.0	0.046	-0.084	0.039	-0.078	0.036	-0.078	0.036
96	0.5	0.50	1.0	1.1	-0.008	-0.005	-0.005	0.000	0.005	0.000	0.005
96	0.5	0.50	1.0	3.0	-0.007	0.000	-0.008	0.000	0.005	0.000	0.005
96	0.5	0.50	1.1	1.1	-0.002	-0.008	-0.001	-0.009	0.001	-0.008	0.001
96	0.5	0.50	1.1	3.0	0.013	-0.007	0.011	-0.012	-0.002	-0.012	-0.002
96	0.5	0.50	3.0	1.1	0.042	0.013	0.041	0.017	0.037	0.013	0.037
96	0.5	0.50	3.0	3.0	0.037	-0.112	0.032	-0.143	0.041	-0.136	0.041

Table 3.3: Batch 1 (Binary Y, Binary X): Coverage probability

n	Pr( Y )	Pr( X )	bZ	bX	Model-based					
					CR		SBR		CAA	
					adj	unadj	adj	unadj	adj	unadj
32	0.1	0.25	1.0	1.1	0.989	0.999	0.993	1.000	0.991	0.999
32	0.1	0.25	1.0	3.0	0.983	0.998	0.979	0.999	0.979	0.999
32	0.1	0.25	1.1	1.1	0.989	0.999	0.992	1.000	0.990	0.999
32	0.1	0.25	1.1	3.0	0.982	0.998	0.979	0.999	0.978	0.999
32	0.1	0.25	3.0	1.1	0.958	0.968	0.962	0.971	0.957	0.968
32	0.1	0.25	3.0	3.0	0.969	0.979	0.959	0.975	0.957	0.975
32	0.1	0.50	1.0	1.1	0.988	0.999	0.991	0.999	0.993	0.998
32	0.1	0.50	1.0	3.0	0.979	0.998	0.974	0.998	0.975	0.998
32	0.1	0.50	1.1	1.1	0.988	0.999	0.992	0.999	0.992	0.999
32	0.1	0.50	1.1	3.0	0.977	0.997	0.975	0.999	0.974	0.998
32	0.1	0.50	3.0	1.1	0.954	0.966	0.959	0.967	0.954	0.967
32	0.1	0.50	3.0	3.0	0.956	0.977	0.960	0.977	0.958	0.976
32	0.5	0.25	1.0	1.1	0.963	0.954	0.958	0.948	0.964	0.956
32	0.5	0.25	1.0	3.0	0.973	0.960	0.967	0.959	0.965	0.957
32	0.5	0.25	1.1	1.1	0.965	0.964	0.966	0.968	0.959	0.961
32	0.5	0.25	1.1	3.0	0.969	0.967	0.966	0.971	0.965	0.967
32	0.5	0.25	3.0	1.1	0.967	0.962	0.966	0.961	0.967	0.962
32	0.5	0.25	3.0	3.0	0.975	0.967	0.978	0.972	0.969	0.969
32	0.5	0.50	1.0	1.1	0.964	0.955	0.964	0.954	0.963	0.953
32	0.5	0.50	1.0	3.0	0.969	0.956	0.964	0.968	0.969	0.965
32	0.5	0.50	1.1	1.1	0.963	0.960	0.969	0.968	0.969	0.968
32	0.5	0.50	1.1	3.0	0.967	0.960	0.966	0.975	0.962	0.968
32	0.5	0.50	3.0	1.1	0.969	0.963	0.968	0.965	0.967	0.961
32	0.5	0.50	3.0	3.0	0.973	0.959	0.972	0.966	0.973	0.968
96	0.1	0.25	1.0	1.1	0.973	0.979	0.974	0.980	0.973	0.980
96	0.1	0.25	1.0	3.0	0.959	0.971	0.956	0.971	0.960	0.979
96	0.1	0.25	1.1	1.1	0.974	0.978	0.974	0.979	0.971	0.980
96	0.1	0.25	1.1	3.0	0.962	0.971	0.959	0.974	0.959	0.978
96	0.1	0.25	3.0	1.1	0.972	0.974	0.975	0.977	0.972	0.974
96	0.1	0.25	3.0	3.0	0.963	0.966	0.968	0.972	0.962	0.969
96	0.1	0.50	1.0	1.1	0.974	0.980	0.974	0.981	0.971	0.980
96	0.1	0.50	1.0	3.0	0.951	0.974	0.953	0.977	0.950	0.972
96	0.1	0.50	1.1	1.1	0.972	0.978	0.974	0.980	0.972	0.977
96	0.1	0.50	1.1	3.0	0.950	0.972	0.955	0.977	0.949	0.973

96	0.1	0.50	3.0	1.1	0.974	0.974	0.979	0.978	0.976	0.974
96	0.1	0.50	3.0	3.0	0.961	0.966	0.958	0.970	0.961	0.974
96	0.5	0.25	1.0	1.1	0.954	0.953	0.952	0.950	0.958	0.954
96	0.5	0.25	1.0	3.0	0.951	0.951	0.952	0.957	0.955	0.957
96	0.5	0.25	1.1	1.1	0.957	0.956	0.954	0.952	0.956	0.952
96	0.5	0.25	1.1	3.0	0.950	0.952	0.952	0.957	0.953	0.957
96	0.5	0.25	3.0	1.1	0.955	0.956	0.954	0.954	0.957	0.954
96	0.5	0.25	3.0	3.0	0.952	0.944	0.952	0.952	0.957	0.951
96	0.5	0.50	1.0	1.1	0.953	0.951	0.952	0.948	0.949	0.946
96	0.5	0.50	1.0	3.0	0.955	0.950	0.956	0.963	0.951	0.956
96	0.5	0.50	1.1	1.1	0.956	0.955	0.951	0.950	0.954	0.951
96	0.5	0.50	1.1	3.0	0.960	0.952	0.956	0.964	0.956	0.961
96	0.5	0.50	3.0	1.1	0.952	0.953	0.953	0.953	0.952	0.954
96	0.5	0.50	3.0	3.0	0.953	0.937	0.954	0.953	0.950	0.946

Table 3.4: Batch 1 (Binary Y, Binary X): Power, subsetted

Avg. nsims	n	Pr( Y )	Pr( X )	bZ	bX	Model-based				Rerandomization		
						CR		SBR		CAA		CAA
						adj	unadj	adj	unadj	adj	unadj	adj
3452 (68.9%)	32	0.1	0.25	1.0	1.1	0.006	0.001	0.003	0.000	0.003	0.001	0.017
3818 (76.2%)	32	0.1	0.25	1.0	3.0	0.008	0.001	0.007	0.000	0.009	0.001	0.025
3451 (68.9%)	32	0.1	0.25	1.1	1.1	0.007	0.001	0.003	0.000	0.003	0.000	0.017
3804 (75.9%)	32	0.1	0.25	1.1	3.0	0.008	0.001	0.008	0.001	0.010	0.001	0.024
3062 (61.1%)	32	0.1	0.25	3.0	1.1	0.018	0.007	0.012	0.005	0.012	0.005	0.044
3408 (68%)	32	0.1	0.25	3.0	3.0	0.024	0.010	0.022	0.006	0.025	0.010	0.065
3445 (68.8%)	32	0.1	0.50	1.0	1.1	0.009	0.001	0.009	0.000	0.005	0.000	0.021
3868 (77.2%)	32	0.1	0.50	1.0	3.0	0.021	0.002	0.025	0.001	0.022	0.001	0.022
3443 (68.7%)	32	0.1	0.50	1.1	1.1	0.009	0.001	0.008	0.001	0.004	0.000	0.021
3879 (77.4%)	32	0.1	0.50	1.1	3.0	0.021	0.002	0.024	0.001	0.024	0.001	0.022
3053 (60.9%)	32	0.1	0.50	3.0	1.1	0.019	0.007	0.013	0.003	0.018	0.006	0.046
3482 (69.5%)	32	0.1	0.50	3.0	3.0	0.055	0.010	0.052	0.004	0.057	0.009	0.068
5008 (100%)	32	0.5	0.25	1.0	1.1	0.028	0.029	0.031	0.026	0.028	0.025	0.057
5009 (100%)	32	0.5	0.25	1.0	3.0	0.019	0.023	0.022	0.017	0.025	0.022	0.056
5009 (100%)	32	0.5	0.25	1.1	1.1	0.025	0.027	0.027	0.022	0.033	0.029	0.063
5009 (100%)	32	0.5	0.25	1.1	3.0	0.023	0.025	0.024	0.017	0.026	0.022	0.059
5001 (99.8%)	32	0.5	0.25	3.0	1.1	0.210	0.231	0.232	0.230	0.222	0.219	0.326
5001 (99.8%)	32	0.5	0.25	3.0	3.0	0.173	0.192	0.197	0.182	0.199	0.182	0.301
5008 (100%)	32	0.5	0.50	1.0	1.1	0.026	0.029	0.026	0.025	0.026	0.025	0.062
5009 (100%)	32	0.5	0.50	1.0	3.0	0.022	0.025	0.028	0.016	0.024	0.019	0.059
5009 (100%)	32	0.5	0.50	1.1	1.1	0.029	0.030	0.024	0.023	0.024	0.023	0.064
5009 (100%)	32	0.5	0.50	1.1	3.0	0.027	0.031	0.028	0.018	0.028	0.022	0.061
5000 (99.8%)	32	0.5	0.50	3.0	1.1	0.203	0.227	0.223	0.224	0.223	0.228	0.326
5004 (99.9%)	32	0.5	0.50	3.0	3.0	0.161	0.177	0.190	0.164	0.180	0.170	0.299
4944 (98.7%)	96	0.1	0.25	1.0	1.1	0.026	0.018	0.024	0.019	0.026	0.018	0.050
4985 (99.5%)	96	0.1	0.25	1.0	3.0	0.039	0.026	0.041	0.026	0.036	0.019	0.049
4939 (98.6%)	96	0.1	0.25	1.1	1.1	0.026	0.019	0.027	0.020	0.028	0.018	0.054
4983 (99.5%)	96	0.1	0.25	1.1	3.0	0.038	0.027	0.040	0.025	0.039	0.022	0.052
4726 (94.3%)	96	0.1	0.25	3.0	1.1	0.238	0.222	0.227	0.218	0.238	0.220	0.324
4877 (97.3%)	96	0.1	0.25	3.0	3.0	0.277	0.252	0.282	0.248	0.285	0.248	0.360
4950 (98.8%)	96	0.1	0.50	1.0	1.1	0.024	0.018	0.025	0.017	0.027	0.018	0.053
4992 (99.6%)	96	0.1	0.50	1.0	3.0	0.045	0.024	0.044	0.021	0.048	0.026	0.058
4947 (98.7%)	96	0.1	0.50	1.1	1.1	0.025	0.019	0.023	0.018	0.027	0.019	0.054
4991 (99.6%)	96	0.1	0.50	1.1	3.0	0.047	0.026	0.045	0.021	0.049	0.025	0.057

4752 (94.9%)	96	0.1	0.50	3.0	1.1	0.230	0.224	0.231	0.218	0.244	0.230	0.334
4899 (97.8%)	96	0.1	0.50	3.0	3.0	0.316	0.274	0.334	0.282	0.325	0.273	0.381
5010 (100%)	96	0.5	0.25	1.0	1.1	0.043	0.045	0.044	0.048	0.038	0.045	0.050
5010 (100%)	96	0.5	0.25	1.0	3.0	0.046	0.047	0.045	0.041	0.043	0.041	0.055
5010 (100%)	96	0.5	0.25	1.1	1.1	0.046	0.049	0.049	0.052	0.046	0.052	0.057
5010 (100%)	96	0.5	0.25	1.1	3.0	0.050	0.051	0.051	0.048	0.048	0.047	0.060
5010 (100%)	96	0.5	0.25	3.0	1.1	0.725	0.744	0.730	0.745	0.741	0.764	0.752
5010 (100%)	96	0.5	0.25	3.0	3.0	0.680	0.649	0.691	0.675	0.683	0.672	0.704
5010 (100%)	96	0.5	0.50	1.0	1.1	0.044	0.047	0.045	0.050	0.048	0.053	0.059
5010 (100%)	96	0.5	0.50	1.0	3.0	0.042	0.048	0.042	0.036	0.046	0.044	0.058
5010 (100%)	96	0.5	0.50	1.1	1.1	0.045	0.047	0.049	0.054	0.050	0.056	0.059
5010 (100%)	96	0.5	0.50	1.1	3.0	0.046	0.048	0.047	0.042	0.046	0.040	0.057
5010 (100%)	96	0.5	0.50	3.0	1.1	0.725	0.741	0.740	0.757	0.739	0.758	0.752
5010 (100%)	96	0.5	0.50	3.0	3.0	0.661	0.629	0.676	0.647	0.669	0.648	0.694

## 3.2 Batch 2: Binary Outcome, Continuous Predictors

Table 3.5: Batch 2 (Binary Y, Continuous X): Power

n	Pr( Y )	bZ	bX	Model-based				Rerandomization		
				CR		SBR		CAA		CAA
				adj	unadj	adj	unadj	adj	unadj	adj
32	0.1	1.0	1.1	0.068	0.001	0.073	0.000	0.068	0.001	0.061
32	0.1	1.0	3.0	0.126	0.004	0.126	0.002	0.125	0.005	0.061
32	0.1	1.1	1.1	0.067	0.001	0.074	0.000	0.069	0.001	0.062
32	0.1	1.1	3.0	0.124	0.004	0.123	0.003	0.128	0.006	0.060
32	0.1	3.0	1.1	0.067	0.006	0.068	0.003	0.067	0.005	0.137
32	0.1	3.0	3.0	0.177	0.021	0.178	0.012	0.179	0.031	0.146
32	0.5	1.0	1.1	0.030	0.028	0.029	0.026	0.027	0.026	0.054
32	0.5	1.0	3.0	0.050	0.029	0.046	0.013	0.046	0.025	0.054
32	0.5	1.1	1.1	0.028	0.028	0.026	0.023	0.027	0.026	0.057
32	0.5	1.1	3.0	0.051	0.029	0.050	0.014	0.050	0.029	0.055
32	0.5	3.0	1.1	0.219	0.235	0.242	0.235	0.216	0.236	0.289
32	0.5	3.0	3.0	0.181	0.116	0.192	0.097	0.174	0.115	0.210
96	0.1	1.0	1.1	0.025	0.018	0.022	0.016	0.026	0.021	0.063
96	0.1	1.0	3.0	0.065	0.032	0.057	0.024	0.062	0.040	0.061
96	0.1	1.1	1.1	0.028	0.019	0.023	0.017	0.027	0.024	0.068
96	0.1	1.1	3.0	0.065	0.035	0.063	0.026	0.064	0.043	0.064
96	0.1	3.0	1.1	0.215	0.204	0.218	0.214	0.209	0.201	0.346
96	0.1	3.0	3.0	0.390	0.272	0.400	0.268	0.392	0.284	0.398
96	0.5	1.0	1.1	0.045	0.049	0.046	0.053	0.043	0.046	0.063
96	0.5	1.0	3.0	0.049	0.047	0.050	0.035	0.046	0.052	0.065
96	0.5	1.1	1.1	0.047	0.050	0.050	0.055	0.053	0.057	0.073
96	0.5	1.1	3.0	0.056	0.053	0.055	0.037	0.051	0.059	0.069
96	0.5	3.0	1.1	0.718	0.727	0.721	0.742	0.714	0.727	0.718
96	0.5	3.0	3.0	0.568	0.447	0.560	0.442	0.562	0.444	0.563

Table 3.6: Batch 2 (Binary Y, Continuous X): Median bias

n	Pr( Y )	bZ	bX	Model-based				Rerandomization		
				CR		SBR		CAA		CAA
				adj	unadj	adj	unadj	adj	unadj	adj
32	0.1	1.0	1.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
32	0.1	1.0	3.0	0.014	0.000	-0.019	0.000	0.047	0.000	0.047
32	0.1	1.1	1.1	-0.048	-0.065	-0.065	-0.095	0.005	-0.044	0.005
32	0.1	1.1	3.0	0.031	-0.049	0.007	-0.095	0.044	0.022	0.044
32	0.1	3.0	1.1	0.354	0.143	0.311	0.143	0.366	0.143	0.366
32	0.1	3.0	3.0	0.332	-0.198	0.332	-0.236	0.354	-0.198	0.354
32	0.5	1.0	1.1	0.012	0.000	-0.014	0.000	-0.012	0.000	-0.012
32	0.5	1.0	3.0	0.017	0.000	-0.002	0.000	0.016	0.000	0.016
32	0.5	1.1	1.1	0.011	-0.026	0.010	-0.031	0.001	-0.031	0.001
32	0.5	1.1	3.0	0.047	-0.031	0.040	-0.080	0.021	-0.029	0.021
32	0.5	3.0	1.1	0.144	0.036	0.175	0.041	0.152	0.031	0.152
32	0.5	3.0	3.0	0.158	-0.310	0.194	-0.325	0.165	-0.310	0.165
96	0.1	1.0	1.1	0.014	0.000	0.003	0.000	0.002	0.000	0.002
96	0.1	1.0	3.0	0.013	0.000	-0.004	0.000	0.013	0.000	0.013
96	0.1	1.1	1.1	0.010	0.004	0.002	-0.002	0.015	0.009	0.015
96	0.1	1.1	3.0	0.015	-0.026	-0.001	-0.044	0.027	-0.012	0.027
96	0.1	3.0	1.1	0.076	0.039	0.090	0.047	0.075	0.039	0.075
96	0.1	3.0	3.0	0.075	-0.251	0.078	-0.260	0.078	-0.234	0.078
96	0.5	1.0	1.1	0.009	0.000	-0.014	-0.002	0.001	0.000	0.001
96	0.5	1.0	3.0	-0.001	0.000	-0.006	0.000	0.000	0.000	0.000
96	0.5	1.1	1.1	0.013	-0.001	-0.014	-0.012	0.013	0.000	0.013
96	0.5	1.1	3.0	0.008	-0.015	0.000	-0.018	0.005	-0.015	0.005
96	0.5	3.0	1.1	0.037	0.011	0.026	0.013	0.033	0.011	0.033
96	0.5	3.0	3.0	0.057	-0.323	0.034	-0.336	0.061	-0.323	0.061

Table 3.7: Batch 2 (Binary Y, Continuous X): Coverage probability

n	Pr( Y )	bZ	bX	Model-based					
				CR		SBR		CAA	
				adj	unadj	adj	unadj	adj	unadj
32	0.1	1.0	1.1	0.930	0.999	0.924	1.000	0.929	0.999
32	0.1	1.0	3.0	0.862	0.993	0.860	0.997	0.863	0.991
32	0.1	1.1	1.1	0.932	0.999	0.925	1.000	0.930	0.999
32	0.1	1.1	3.0	0.867	0.994	0.865	0.996	0.861	0.991
32	0.1	3.0	1.1	0.913	0.970	0.909	0.970	0.910	0.966
32	0.1	3.0	3.0	0.868	0.967	0.872	0.978	0.870	0.964
32	0.5	1.0	1.1	0.960	0.956	0.958	0.948	0.965	0.960
32	0.5	1.0	3.0	0.940	0.955	0.941	0.971	0.943	0.960
32	0.5	1.1	1.1	0.963	0.962	0.965	0.968	0.963	0.962
32	0.5	1.1	3.0	0.938	0.960	0.939	0.978	0.941	0.962
32	0.5	3.0	1.1	0.966	0.964	0.966	0.963	0.966	0.964
32	0.5	3.0	3.0	0.935	0.940	0.941	0.961	0.934	0.940
96	0.1	1.0	1.1	0.973	0.980	0.976	0.981	0.971	0.977
96	0.1	1.0	3.0	0.931	0.968	0.939	0.975	0.933	0.958
96	0.1	1.1	1.1	0.972	0.980	0.976	0.982	0.970	0.976
96	0.1	1.1	3.0	0.934	0.967	0.937	0.973	0.933	0.954
96	0.1	3.0	1.1	0.978	0.978	0.980	0.978	0.976	0.978
96	0.1	3.0	3.0	0.935	0.927	0.934	0.933	0.932	0.917
96	0.5	1.0	1.1	0.951	0.947	0.949	0.945	0.954	0.951
96	0.5	1.0	3.0	0.947	0.951	0.945	0.963	0.951	0.945
96	0.5	1.1	1.1	0.954	0.951	0.954	0.951	0.949	0.947
96	0.5	1.1	3.0	0.943	0.951	0.943	0.963	0.953	0.940
96	0.5	3.0	1.1	0.951	0.952	0.954	0.954	0.951	0.951
96	0.5	3.0	3.0	0.947	0.879	0.946	0.891	0.944	0.874

Table 3.8: Batch 2 (Binary Y, Continuous X): Power, subsetted

Avg. nsims	n	Pr( Y )	bZ	bX	Model-based				Rerandomization		
					CR		SBR		CAA		
					adj	unadj	adj	unadj	adj	unadj	
3469 (69.2%)	32	0.1	1.0	1.1	0.011	0.001	0.011	0.000	0.012	0.001	0.014
4320 (86.2%)	32	0.1	1.0	3.0	0.058	0.004	0.060	0.002	0.059	0.006	0.027
3462 (69.1%)	32	0.1	1.1	1.1	0.011	0.001	0.012	0.000	0.013	0.001	0.014
4318 (86.2%)	32	0.1	1.1	3.0	0.057	0.005	0.060	0.003	0.063	0.006	0.027
3048 (60.8%)	32	0.1	3.0	1.1	0.025	0.009	0.019	0.005	0.028	0.008	0.037
4123 (82.3%)	32	0.1	3.0	3.0	0.127	0.025	0.131	0.014	0.134	0.038	0.088
5008 (100%)	32	0.5	1.0	1.1	0.030	0.028	0.029	0.026	0.027	0.026	0.054
4999 (99.8%)	32	0.5	1.0	3.0	0.047	0.029	0.043	0.013	0.044	0.025	0.054
5010 (100%)	32	0.5	1.1	1.1	0.028	0.028	0.026	0.023	0.027	0.026	0.057
4996 (99.7%)	32	0.5	1.1	3.0	0.048	0.029	0.048	0.014	0.047	0.028	0.054
4999 (99.8%)	32	0.5	3.0	1.1	0.219	0.235	0.242	0.235	0.216	0.236	0.288
4982 (99.4%)	32	0.5	3.0	3.0	0.175	0.116	0.190	0.097	0.169	0.114	0.206
4934 (98.5%)	96	0.1	1.0	1.1	0.025	0.018	0.022	0.017	0.027	0.021	0.054
5009 (100%)	96	0.1	1.0	3.0	0.065	0.032	0.057	0.024	0.062	0.040	0.061
4929 (98.4%)	96	0.1	1.1	1.1	0.028	0.019	0.023	0.018	0.027	0.024	0.058
5009 (100%)	96	0.1	1.1	3.0	0.066	0.035	0.063	0.026	0.064	0.043	0.064
4745 (94.7%)	96	0.1	3.0	1.1	0.227	0.215	0.229	0.225	0.221	0.213	0.313
5001 (99.8%)	96	0.1	3.0	3.0	0.391	0.272	0.400	0.269	0.392	0.284	0.397
5010 (100%)	96	0.5	1.0	1.1	0.045	0.049	0.046	0.053	0.043	0.046	0.063
5010 (100%)	96	0.5	1.0	3.0	0.049	0.047	0.050	0.035	0.046	0.052	0.065
5010 (100%)	96	0.5	1.1	1.1	0.047	0.050	0.050	0.055	0.053	0.057	0.073
5010 (100%)	96	0.5	1.1	3.0	0.056	0.053	0.055	0.037	0.051	0.059	0.069
5010 (100%)	96	0.5	3.0	1.1	0.718	0.727	0.721	0.742	0.714	0.727	0.718
5010 (100%)	96	0.5	3.0	3.0	0.568	0.447	0.560	0.442	0.562	0.444	0.563

### 3.3 Batch 3: Continuous Outcome, Binary Predictors

Table 3.9: Batch 3 (Continuous Y, Binary X): Power

Table 3.10: Batch 3 (Continuous Y, Binary X): Median bias

n	Pr( X )	bZ	bX	Model-based				Rerandomization		
				CR		SBR		CAA		CAA
				adj	unadj	adj	unadj	adj	unadj	adj
32	0.25	1	1.1	0.005	0.005	0.009	0.008	0.003	-0.003	0.003
32	0.25	1	3.0	0.005	-0.003	0.009	-0.003	0.003	-0.089	0.003
32	0.25	3	1.1	0.005	0.005	0.009	0.008	0.003	-0.003	0.003
32	0.25	3	3.0	0.005	-0.003	0.009	-0.003	0.003	-0.089	0.003
32	0.50	1	1.1	0.005	0.006	0.003	-0.003	0.004	0.005	0.004
32	0.50	1	3.0	0.005	-0.006	0.003	0.001	0.004	0.045	0.004
32	0.50	3	1.1	0.005	0.006	0.003	-0.003	0.004	0.005	0.004
32	0.50	3	3.0	0.005	-0.006	0.003	0.001	0.004	0.045	0.004
96	0.25	1	1.1	-0.002	-0.006	0.001	0.000	-0.002	-0.010	-0.002
96	0.25	1	3.0	-0.002	-0.007	0.001	0.001	-0.002	-0.142	-0.002
96	0.25	3	1.1	-0.002	-0.006	0.001	0.000	-0.002	-0.010	-0.002
96	0.25	3	3.0	-0.002	-0.007	0.001	0.001	-0.002	-0.142	-0.002
96	0.50	1	1.1	-0.004	-0.004	0.001	0.002	-0.002	-0.004	-0.002
96	0.50	1	3.0	-0.004	0.001	0.001	0.003	-0.002	0.018	-0.002
96	0.50	3	1.1	-0.004	-0.004	0.001	0.002	-0.002	-0.004	-0.002
96	0.50	3	3.0	-0.004	0.001	0.001	0.003	-0.002	0.018	-0.002

Table 3.11: Batch 3 (Continuous Y, Binary X): Coverage probability

n	Pr( X )	bZ	bX	Model-based					
				CR		SBR		CAA	
				adj	unadj	adj	unadj	adj	unadj
32	0.25	1	1.1	0.937	0.937	0.933	0.934	0.938	0.940
32	0.25	1	3.0	0.937	0.945	0.933	0.982	0.938	0.931
32	0.25	3	1.1	0.937	0.937	0.933	0.934	0.938	0.940
32	0.25	3	3.0	0.937	0.945	0.933	0.982	0.938	0.931
32	0.50	1	1.1	0.935	0.935	0.941	0.941	0.934	0.935
32	0.50	1	3.0	0.935	0.941	0.941	0.996	0.934	0.928
32	0.50	3	1.1	0.935	0.935	0.941	0.941	0.934	0.935
32	0.50	3	3.0	0.935	0.941	0.941	0.996	0.934	0.928
96	0.25	1	1.1	0.948	0.950	0.950	0.948	0.951	0.947
96	0.25	1	3.0	0.948	0.943	0.950	0.987	0.951	0.924
96	0.25	3	1.1	0.948	0.950	0.950	0.948	0.951	0.947
96	0.25	3	3.0	0.948	0.943	0.950	0.987	0.951	0.924
96	0.50	1	1.1	0.948	0.947	0.955	0.957	0.951	0.951
96	0.50	1	3.0	0.948	0.942	0.955	0.997	0.951	0.901
96	0.50	3	1.1	0.948	0.947	0.955	0.957	0.951	0.951
96	0.50	3	3.0	0.948	0.942	0.955	0.997	0.951	0.901

## 3.4 Batch 4: Continuous Outcome, Continuous Predictors

Table 3.12: Batch 4 (Continuous Y, Continuous X): Power

n	bZ	bX	Model-based				Rerandomization			
			CR		SBR		CAA		CAA	
			adj	unadj	adj	unadj	adj	unadj	adj	adj
32	1.0	1.1	0.051	0.048	0.052	0.054	0.051	0.048	0.054	
32	1.0	3.0	0.051	0.049	0.052	0.007	0.051	0.049	0.054	
32	1.1	1.1	0.060	0.059	0.061	0.062	0.060	0.059	0.058	
32	1.1	3.0	0.060	0.051	0.061	0.008	0.060	0.051	0.058	
32	3.0	1.1	0.999	0.999	0.999	0.999	0.999	0.999	0.999	
32	3.0	3.0	0.999	0.435	0.999	0.419	0.999	0.435	0.999	
96	1.0	1.1	0.052	0.051	0.049	0.049	0.052	0.051	0.054	
96	1.0	3.0	0.052	0.049	0.049	0.005	0.052	0.049	0.054	
96	1.1	1.1	0.079	0.079	0.074	0.074	0.079	0.079	0.084	
96	1.1	3.0	0.079	0.052	0.074	0.006	0.079	0.052	0.084	
96	3.0	1.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
96	3.0	3.0	1.000	0.895	1.000	0.961	1.000	0.895	1.000	

Table 3.13: Batch 4 (Continuous Y, Continuous X): Median bias

n	bZ	bX	Model-based				Rerandomization		
			CR		SBR		CAA		
			adj	unadj	adj	unadj	adj	unadj	
32	1.0	1.1	-0.013	-0.009	0.002	-0.006	-0.013	-0.009	-0.013
32	1.0	3.0	-0.013	-0.019	0.002	-0.002	-0.013	-0.019	-0.013
32	1.1	1.1	-0.013	-0.009	0.002	-0.006	-0.013	-0.009	-0.013
32	1.1	3.0	-0.013	-0.019	0.002	-0.002	-0.013	-0.019	-0.013
32	3.0	1.1	-0.013	-0.009	0.002	-0.006	-0.013	-0.009	-0.013
32	3.0	3.0	-0.013	-0.019	0.002	-0.002	-0.013	-0.019	-0.013
96	1.0	1.1	-0.003	-0.004	0.002	0.003	-0.003	-0.004	-0.003
96	1.0	3.0	-0.003	-0.013	0.002	-0.002	-0.003	-0.013	-0.003
96	1.1	1.1	-0.003	-0.004	0.002	0.003	-0.003	-0.004	-0.003
96	1.1	3.0	-0.003	-0.013	0.002	-0.002	-0.003	-0.013	-0.003
96	3.0	1.1	-0.003	-0.004	0.002	0.003	-0.003	-0.004	-0.003
96	3.0	3.0	-0.003	-0.013	0.002	-0.002	-0.003	-0.013	-0.003

Table 3.14: Batch 4 (Continuous Y, Continuous X): Coverage probability

Model-based									
n	bZ	bX	CR		SBR		CAA		
			adj	unadj	adj	unadj	adj	unadj	
32	1.0	1.1	0.938	0.941	0.939	0.937	0.938	0.941	
32	1.0	3.0	0.938	0.940	0.939	0.991	0.938	0.940	
32	1.1	1.1	0.938	0.941	0.939	0.937	0.938	0.941	
32	1.1	3.0	0.938	0.940	0.939	0.991	0.938	0.940	
32	3.0	1.1	0.938	0.941	0.939	0.937	0.938	0.941	
32	3.0	3.0	0.938	0.940	0.939	0.991	0.938	0.940	
96	1.0	1.1	0.946	0.947	0.949	0.948	0.946	0.947	
96	1.0	3.0	0.946	0.949	0.949	0.994	0.946	0.949	
96	1.1	1.1	0.946	0.947	0.949	0.948	0.946	0.947	
96	1.1	3.0	0.946	0.949	0.949	0.994	0.946	0.949	
96	3.0	1.1	0.946	0.947	0.949	0.948	0.946	0.947	
96	3.0	3.0	0.946	0.949	0.949	0.994	0.946	0.949	

### 3.5 Figures: Power, all settings

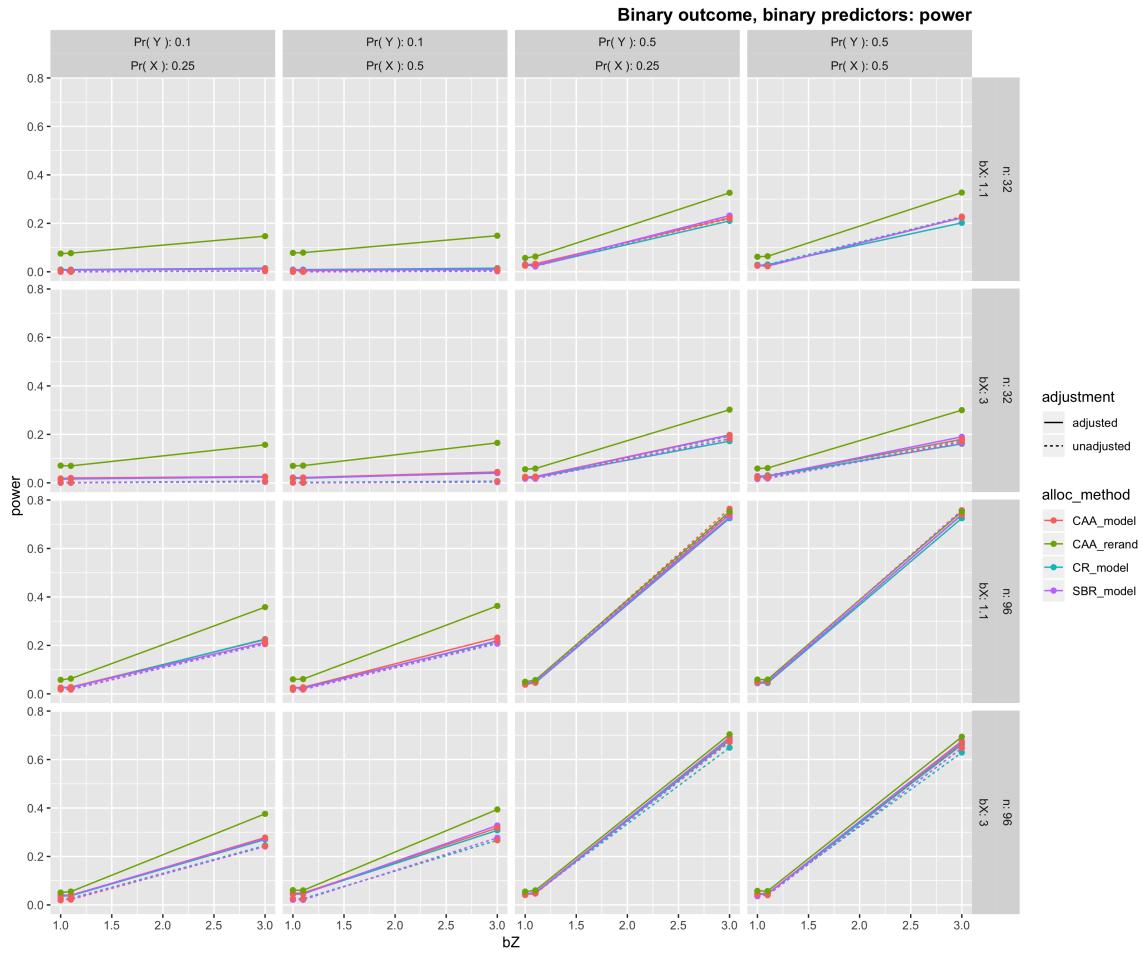


Figure 3.1: Batch 1: Power

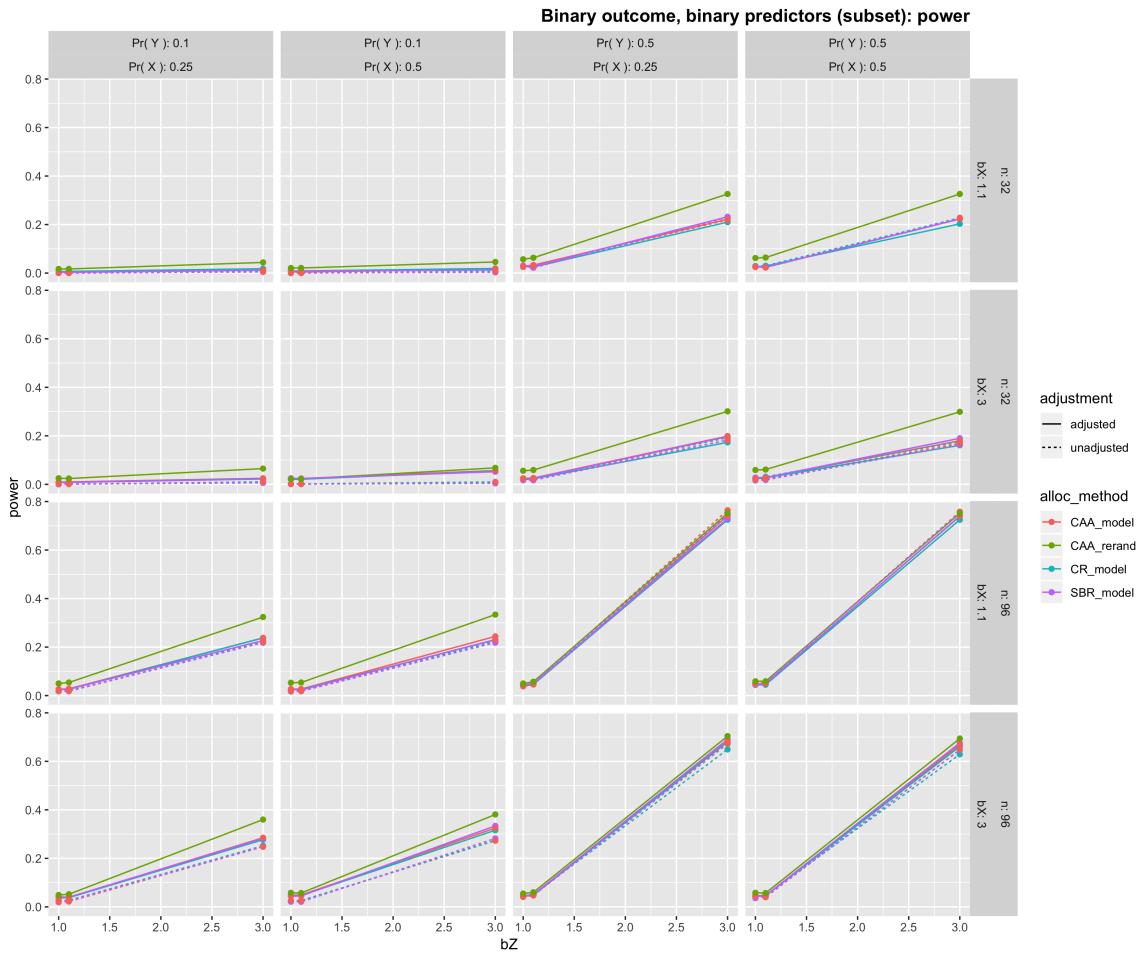


Figure 3.2: Batch 1 subset: Power

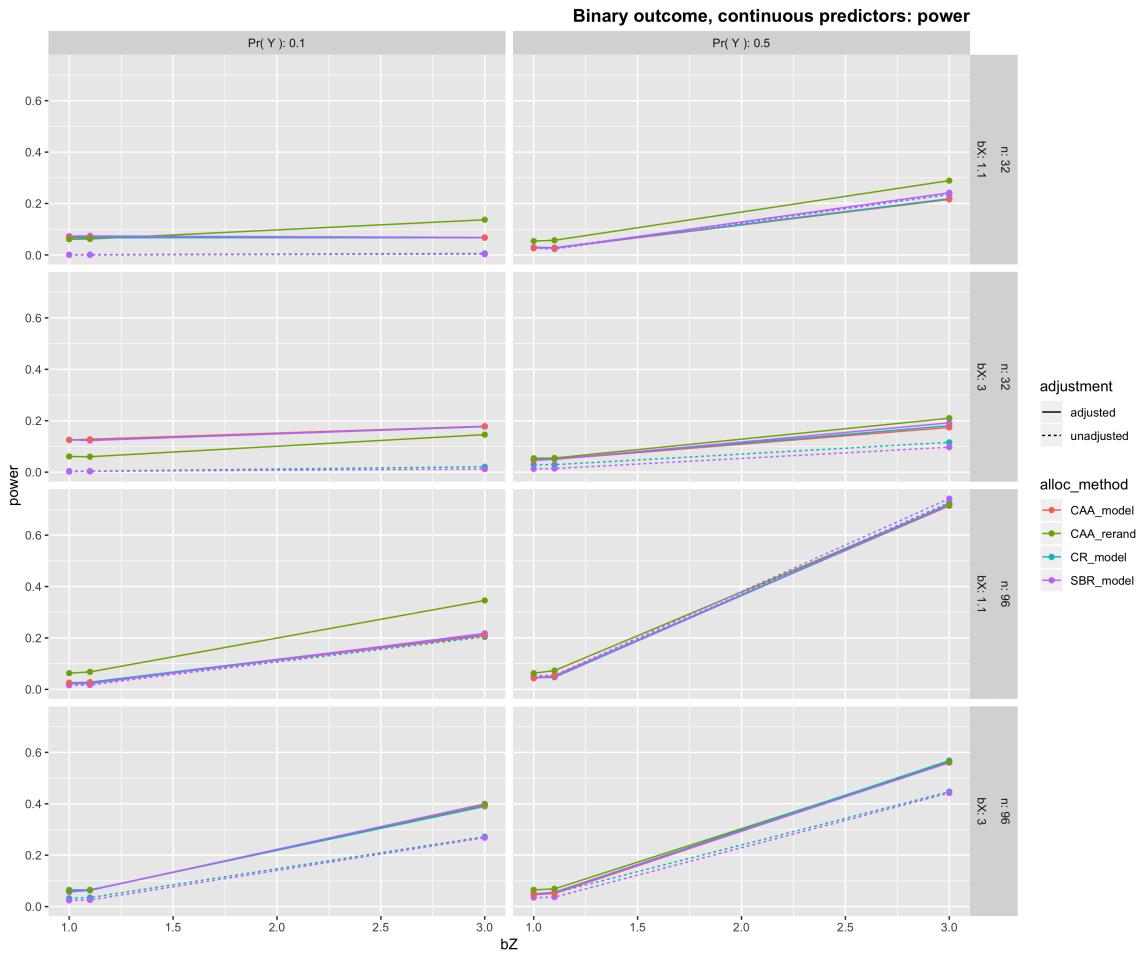


Figure 3.3: Batch 2: Power

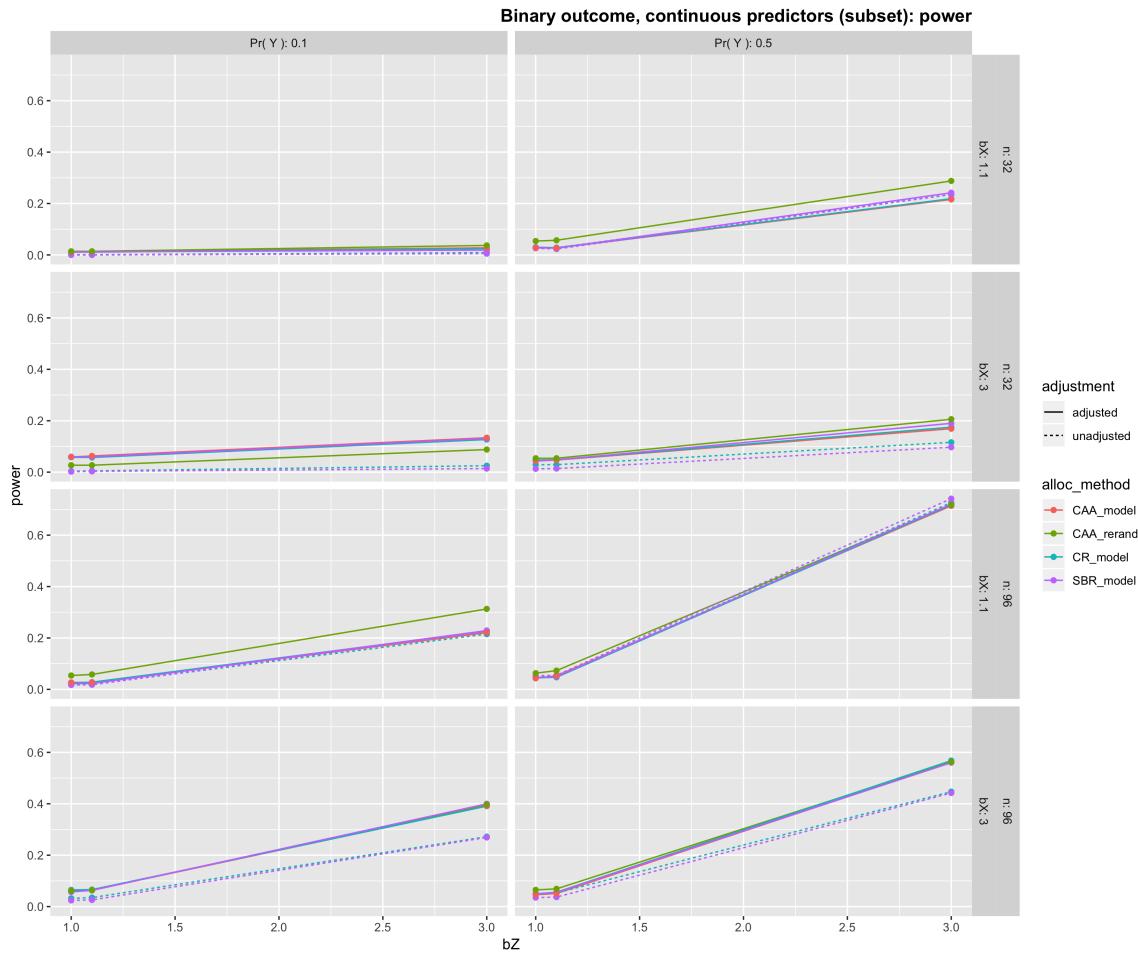


Figure 3.4: Batch 2 subset: Power

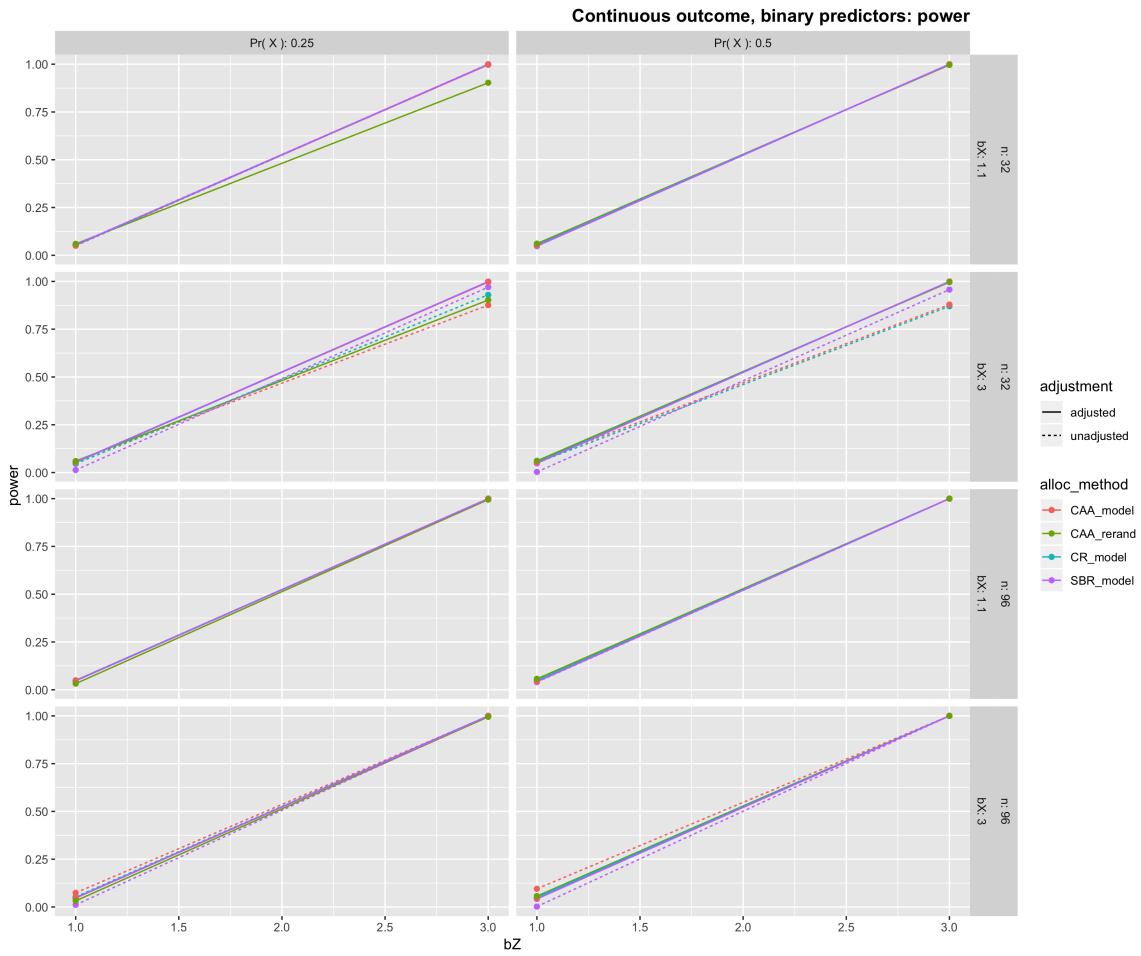


Figure 3.5: Batch 3: Power

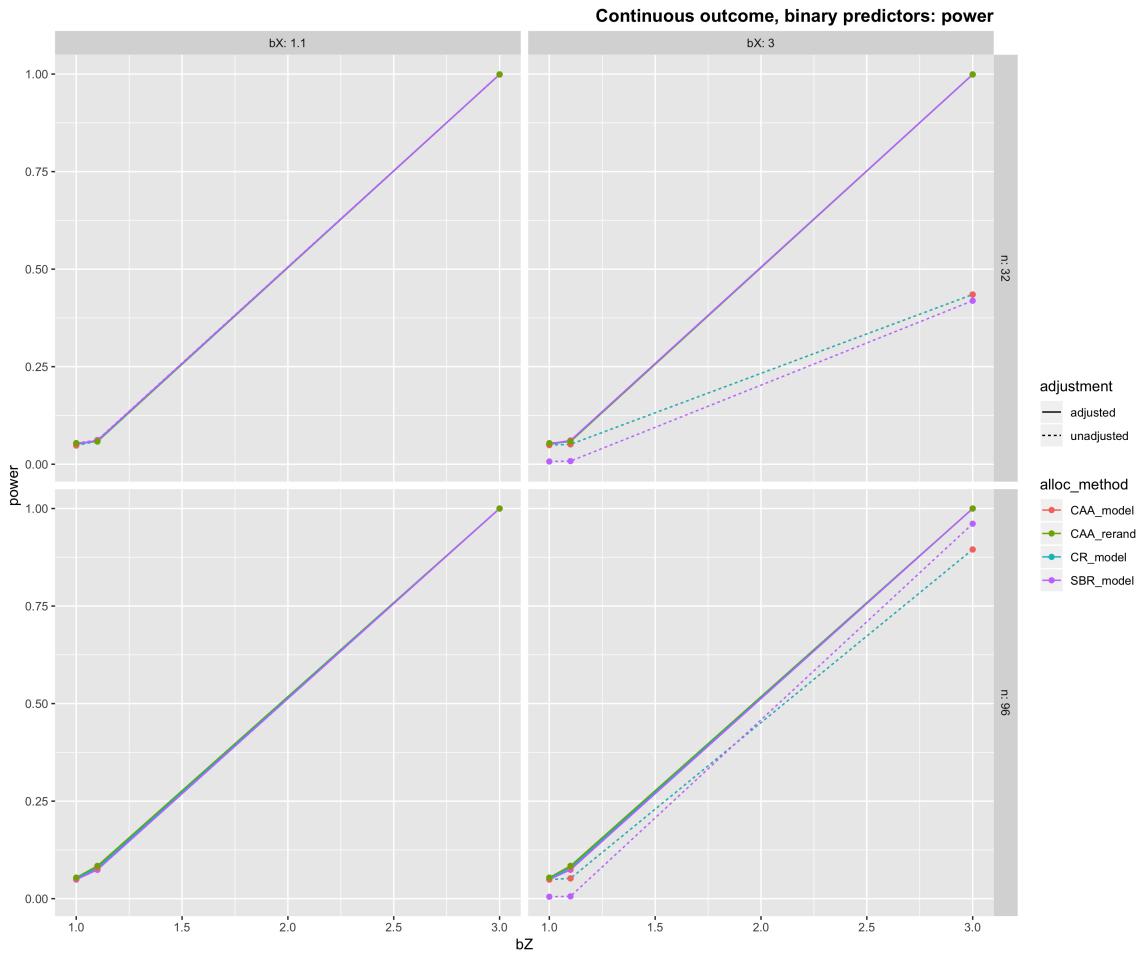


Figure 3.6: Batch 4: Power

## 4 Results

## 4.1 Binary outcome setting

### Note on subsetted simulations

Simulations with low observed event rates in the binary outcome model could result in large estimated treatment effects and standard errors, often corresponding to cases where the `glm()` algorithm did not converge. For this reason, subsetted results are presented for simulation outcomes where both the `glm()` algorithm converged and at least one event was observed in each treatment arm. Subsetting on these two criterion led to excluding greater than 30 percent of simulations in low sample size scenarios. This issue occurs more frequently when sample size is small ( $n=32$ ), outcome prevalence ( $Pr(Y)$ ) is small, and for small treatment and prognostic factor effect sizes ( $bZ$  and  $bX$ , respectively).

### Binary predictors

Power is substantially higher using CAA followed by re-randomization when compared to all other methods, including CAA followed by model-based inference. The modest increase in power is especially apparent in low sample size settings and low marginal outcome prevalence. Adjusting for prognostic factors used in the balancing method consistently increased power relative to unadjusted methods in almost all considered settings.

The power benefit from using CAA followed by re-randomization becomes diminished relative to all other allocation procedures when sample size increases and marginal outcome prevalence increases (see Figure 3.1, bottom right quadrant). Even at large sample sizes re-randomization has greater power than model-based approaches, particularly when outcome prevalence is low.

Median bias (Table 3.2) is appreciably low and approaches zero when the estimated

treatment effect is null or when trial size is large ( $n=96$ ). The bias results from `glm()` algorithm returning inflated estimates when no valid odds ratio estimate exists, *i.e.* when the denominator of the odds ratio is zero because there are no events in the untreated arm. This happens more often when considering within-strata odds ratios, which is why the median bias for adjusted estimates, though still small, is larger on average when compared to unadjusted estimates.

Coverage probability nears 100 percent at low outcome prevalence and low trial size, likely from standard errors being too large with low event rates. Coverage probability nears 95 percent when outcome prevalence is high and when sample size is large ( $n=96$ ).

Subsetted power results show better type I error control for CAA followed by re-randomization overall (compare Table 3.1 rows 1,2 with Table 3.4, rows 1,2). The modest power increase of CAA with re-randomization relative to other allocation procedures using model-based analysis is consistent in large and small sample sizes and high and low outcome prevalences.

## Continuous predictors

Both CAA and SBR with model-based inference have higher power but fail to control size. In low outcome prevalence ( $\text{Pr}(Y)=0.1$ ) under strong prognostic factor effect size ( $bX=3$ ) with a type I error threshold of 0.05, the power under the null setting is greater than 0.10 (see Table 3.5, rows 1,2). As discussed in the previous section, the tendency for `glm()` to return large estimates and standard errors given certain simulation outcomes results in spuriously small p-values. Looking at the subsetted results in Table 3.8, power is highest with re-randomization or at least comparable to model-based approaches that adjust for prognostic factors.

Considering model-based approaches, adjusted estimates are consistently more

powerful than unadjusted estimates. Adjusting for prognostic factors used in the balancing method controls median bias. Median bias tended to be less than 10 percent and decreased with increasing sample size.

The coverage probability at low sample sizes was inflated due to low event rates, as described in the previous section on binary predictors. All adjusted methods have coverage probabilities near 95 percent in the large sample size setting.

## 4.2 Continuous outcome setting

### Binary predictors

All methods perform comparably with respect to power. There is not as much of an advantage to re-randomization compared to model-based analysis approaches in the continuous outcome setting.

Median bias was tiny and coverage probability approached 95 percent for all methods and all conditions. An exception is when prognostic factor effect size is large, stratified block randomization was used, and covariates were not adjusted for.

### Continuous predictors

Adjusting for prognostic factors did not considerably increase power, except when prognostic factor effect size is large. As in all simulated conditions with a continuous outcome, re-randomization following allocation does not confer an additional advantage in terms of increased power.

## 5 *Discussion*

## 5.1 Discussion

We sought to address whether covariate-adaptive allocation followed by re-randomization conferred any statistical power advantage relative to complete randomization, stratified block randomization, and covariate-adaptive allocation using model-based inference.

### Review of Results

We should address the following topics:

1. Does adjusting for balancing factors affect validity, or power? (short and long answer: yes).
  - a) Forsythe (1987) showed Type 1 error is preserved if analysis adjusts for balancing factors
  - b) Birkett (1985) shows omitting balancing factors in linear models leads to conservative Type 1 error
  - c) Shao et. al (2010) show test procedure valid for complete randomization is valid for CAR provided correct model specification, including balancing factors in model.
2. Does re-randomization following CAA confer a power advantage relative to complete randomization?
3. Does re-randomization following CAA confer a power advantage relative to model-based analysis following CAA?

### Limitations

We would have liked to address the following weak points in our approach:

1. **Model drift** to assess advantage of re-randomization analysis to model-based methods on reducing bias and increasing power

2. Considered **varying allocation method imbalance parameters** for balancing prognostic factors. Reporting metrics on "cosmetic balance" would highlight differences between randomization approaches and main practical reason behind implementing constrained randomization approaches.
3. Consider **different non-continuous outcome settings** where constrained randomization approaches are commonly used (e.g. Poisson regression, proportional hazards regression).
4. Varying **number of prognostic factors** and adjusting for a subset of prognostic factors. Larger numbers of predictors is the case where stratification fails to provide balance and dynamic hierarchical schemes such as Heritier et. al 2005 can control balance.
5. **Adjusting for continuous prognostic factors** (dichotomized for allocation) as **continuous**
6. **Confidence intervals for re-randomization are not valid for parameter estimation**, as they are generated under the null hypothesis
7. Considered **model-based approaches** that would pick allocations that minimize the variance of the estimated treatment effect

## Potential next steps

If we had more time, we would have liked to:

1. Model the impact of drift
2. Vary imbalance parameters for all constrained randomization methods
3. Consider count and time-to-event settings
4. Implement approaches for confidence intervals for treatment effect under alternative hypotheses (bootstrap, etc.)

## Concluding remarks

We conclude with the following remarks:

1. Adjusting for balancing factors is necessary to guarantee validity of analysis approach, model-based or otherwise
2. Re-randomization following any allocation method is recommended under certain settings:
  - Under drift: to reduce bias and power
  - To increase power relative to model-based approaches by recovering precision gain from balancing procedure
3. There is considerable value in considering the type of balance desired and choosing an appropriate allocation method.
  - Conditional or marginal balance: conditional balance may be desired if subgroup effects are of interest, marginal balance may be desired if comparability across treatment arms is most important
  - CAR methods can control treatment arm imbalance within group-level covariates where stratification would yield too small sample sizes within strata.
4. Balancing distribution of prognostic factors not guaranteed to increase precision in non-continuous outcome settings
  - Equal balance across treatment arms will not necessarily guarantee variance of estimated treatment effect is minimized
5. Implementing a random element in the constrained allocation procedure is recommended; to prevent overly restricting candidate set size of allocation sequences (see Kuznetsova 2009 for refs, incl. ICH 1998; CPMP 2003)

## 5.2 Addressing reader comments

**Mike L: Address if power of re-randomization due to inflated Type I error**

Potential for increased power of rerandomization in small sample sizes could be in part due to the poor type 1 error control.

For binary outcomes and predictors, CAA with rerandomization has slightly inflated type I error of 0.075 and 0.071 when n=32, outcome prevalence is 10%, and with low and high prognostic factor effect size, respectively. In contrast, all model-based methods had type 1 error below 0.018 in the same setting regardless of allocation method (Table 3.1 Rows 1 and 2). Increasing outcome prevalence to 50% lowers the type 1 error of rerandomization following CAA to 0.057 and 0.056 (see Table 3.1 Rows 13, 14), whereas increasing prognostic factor effect size had no impact on type 1 error control.

Considering subsetted results in Table 3.4 under the same settings show CAA with re-randomization has type 1 error of 0.017 and 0.025 for n=32, outcome prevalence of 10%, with low and high prognostic factor effect size, respectively. Again, all model-based methods had type 1 error below 1% with a specified type 1 error threshold of 5%. Increasing outcome prevalence to 50% from 10%, the type 1 error rate is 0.057 and 0.056 and nears the nominal rate (see Table 3.4 Rows 13, 14).

For binary outcomes and continuous predictors, CAA with rerandomization has type I error of 0.061 when n=32, outcome prevalence is 10%, and for both low and high prognostic factor effect size. Compared to model based methods, their adjusted type 1 errors are uniformly higher and increase to above 12.5% when prognostic factor effect size is increased from low to high (see Table 3.5, Rows 1,2). At large sample sizes, model-based methods have type 1 error near half the nominal 0.05 rate while re-randomization has slightly inflated type 1 error of 0.063 (see Table 3.5, Rows 13,

14).

For the continuous outcome settings, rerandomization following CAA has comparable type 1 error rates to model-based methods.

1. show CAA with re-randomization has type 1 error of 0.075 and 0.071 for n=32, outcome prevalence of 10%, with low and high prognostic factor effect size, respectively.
  - a) Increasing outcome prevalence to 50% from 10% lowers type 1 error rate to 0.057 and 0.056 (see Table 3.1 Rows 13, 14)
  - b) Increasing prognostic factor prevalence to 50% from 25% (with outcome prevalence of 10%) does not affect type 1 error rate (0.078 and 0.070, respectively: see Table 3.1 Rows 7,8)
  - c) Further increasing outcome prevalence to 50% from 10% lowers type 1 error rate to 0.062 and 0.059 (see Table 3.1 Rows 19, 20)
2. Subsetted results on Table 3.4 Rows 1 and 2 show CAA with re-randomization has type 1 error of 0.017 and 0.025 for n=32, outcome prevalence of 10%, with low and high prognostic factor effect size, respectively.
  - a) Increasing outcome prevalence to 50% from 10% increases type 1 error rate to the nominal rate of 0.057 and 0.056 (see Table 3.4 Rows 13, 14)
  - b) Increasing prognostic factor prevalence to 50% from 25% (with outcome prevalence of 10%) does not affect type 1 error rate (0.021 and 0.022, respectively: see Table 3.4 Rows 7,8)
  - c) Further increasing outcome prevalence to 50% from 10% increases type 1 error rate to 0.062 and 0.059 (see Table 3.4 Rows 19, 20)

## 6 *Limitations*

Due to many unforeseen factors, we ran into these complications:

## 6.1 Simulation management

For each simulation, we had to keep track of the following:

- Prognostic factors matrix  $\mathbf{X}$
- Entry time vector  $\mathbf{T}$
- Allocation sequence vectors  $\mathbf{Z}$ , each with their associated
- Outcome measure vector  $\mathbf{Y}$
- Model-based regression estimates:
  - Estimate  $\hat{\beta}$
  - Standard error  $\hat{SE}(\hat{\beta})$
  - P-value  $p_T$
  - T-statistic  $t_{obs}$
  - Confidence interval  $\hat{CI}(\hat{\beta})$
- Re-randomization based regression estimates, including:
  - Re-randomized allocation sequence vectors  $\mathbf{Z}$ , each with their associated estimate  $\hat{\beta}$

We structured the simulation around using the ”**simulator**” R package, which handles file management (saving/loading data), RNG seeds for parallel processing, and sequence of simulation steps.

## Issues

**Memory management** was poor. The simulator stored data as so:

- each **Model** object contains the simulation parameters

- each **Draw** object contains the prognostic factor and entry times (**X**, **T**) for all subjects
- the first **Output** object contains the allocation sequence vector and outcome vector (**Z**, **Y**) for all subjects
- the second **Output** object contains the model-based regression estimates ( $\hat{\beta}$ ,  $\hat{SE}(\hat{\beta})$ ,  $p_T$ ,  $t_{obs}$ ,  $\hat{CI}(\hat{\beta})$ )
- the third **Output** object contains the re-randomization based regression estimates ( $\hat{\beta}$ ,  $\hat{SE}(\hat{\beta})$ ,  $p_T$ ,  $t_{obs}$ ,  $\hat{CI}(\hat{\beta})$ )

Each 'look' at the data took lots of time, and (to this day) we have not figured out how to efficiently run the code. Granted, the task was computationally expensive: on a grid of models, for each simulation, perform 500 re-allocations and estimate the parameters, and then compute summary statistics on the re-randomized procedures?

At this point, we had already invested significantly into the "simulator" package to consider alternative approaches.

**Computing metrics** was not as straightforward as the package authors made it seem.

We computed metrics on the output, but loading the required **Output** objects into memory was too computationally costly.

We ended up running a separate R script that loaded each "Output" object individually, computed the metrics we wanted, and then saving both the output as .csv and metrics as .csv files.

**Memory management** The problem posed challenges for both memory management, and code performance as the number of simulations grew.

## Lessons

**Know when to use packages, and when to make your own code.** At first, using the "simulator" package was great because it handled the RNG generation, file management, and forced us to delineate the steps of the simulation process.

Where using the "simulator" package fell short was when issues arose and we wanted to diagnose problems in the code, as well as work on memory and performance speedups.

The file sizes were extremely large (even though they were stored as .Rdata files). The output of all functions had to be in list() format.

**Create a roadmap for programming, including which tools (packages) you're planning to use.** We originally set forth to plan each simulation separately, but due to concerns with memory management (and to a lesser extent, with handling RNGs under parallelization) we decided on using the 'simulator' package.

This proved to be costly to port the code as well as to learn the structure of a new package. Downstream, it cost more time to diagnose issues, as well as increased the difficulty for optimizing the code.

**Properly define the scope of your project.** We had to pare down the questions we were going to ask, because the scope of the questions we sought to ask was rapidly growing in complexity. We wanted to consider the impact of

- a time-dependent shift in outcome measure or prevalence (drift) by modeling entry time  $\mathbf{T}$ .
- allocation method parameters, including
  - block size  $\mathbf{B}$  for SBR and its' impact on controlling for drift,

- maximum imbalance bounds **MI** for CAA procedures and its' impact on inference metrics
- 

## 6.2 Further development

**Package for implementing rerandomization analysis** It would be useful to address the gap between analysis methods proposed in the literature prior to advances in computing that make rerandomization an attractive approach for estimating uncertainty.

**Visualizing the restrictions to allocation sequence hyperplane** Part of the difficulty in understanding the effect of covariate adaptive allocation procedures on the validity (in the philosophical sense) of results is the disconnect between the trade-off between forcing/inducing balance in key prognostic factors across treatment groups and the implications for how 'random' we can consider the treatment assignment variable.

When stepping away from complete randomization to restricted randomization approaches, one loses the ability to treat all subjects' treatment assignments as independent. The validity of the test (hypothesis test rejection proportion under the null hypothesis, and whether it is below the prespecified type I error threshold) is recovered only when the test is adjusted for the prognostic factors used in the balancing process.

A visualization of all possible randomization schemes (perhaps colored based on the likelihood of observing each sequence) would be useful to:

1. Understand the impact of restricted randomization methods on restricting the size of the reference distribution,

2. Show that certain sequences are 'impossible' under restricted randomization methods,
3. Show the properties of SBR (i.e. allocation proportion equals allocation ratio after each filled block)
4. Shows control of balance sequentially, one advantage to such procedures.

**Model-based approaches** We did not consider a class of covariate-adaptive allocation methods that posit a model for the relationship between prognostic factors, treatment assignment, and outcome measure in order to bias treatment assignment towards sequences that minimize the variance of the estimated treatment effect. As noted in the optimal ratio of prognostic factors within subgroups is unlikely to be 50:50, meaning the ratio for minimizing the variance of the adjusted treatment effect depends on the strength of the prognostic factor on the outcome (the odds ratio).

## 7 *Figures*

## 7.1 Batch 1: Binary Outcome, Binary Predictors

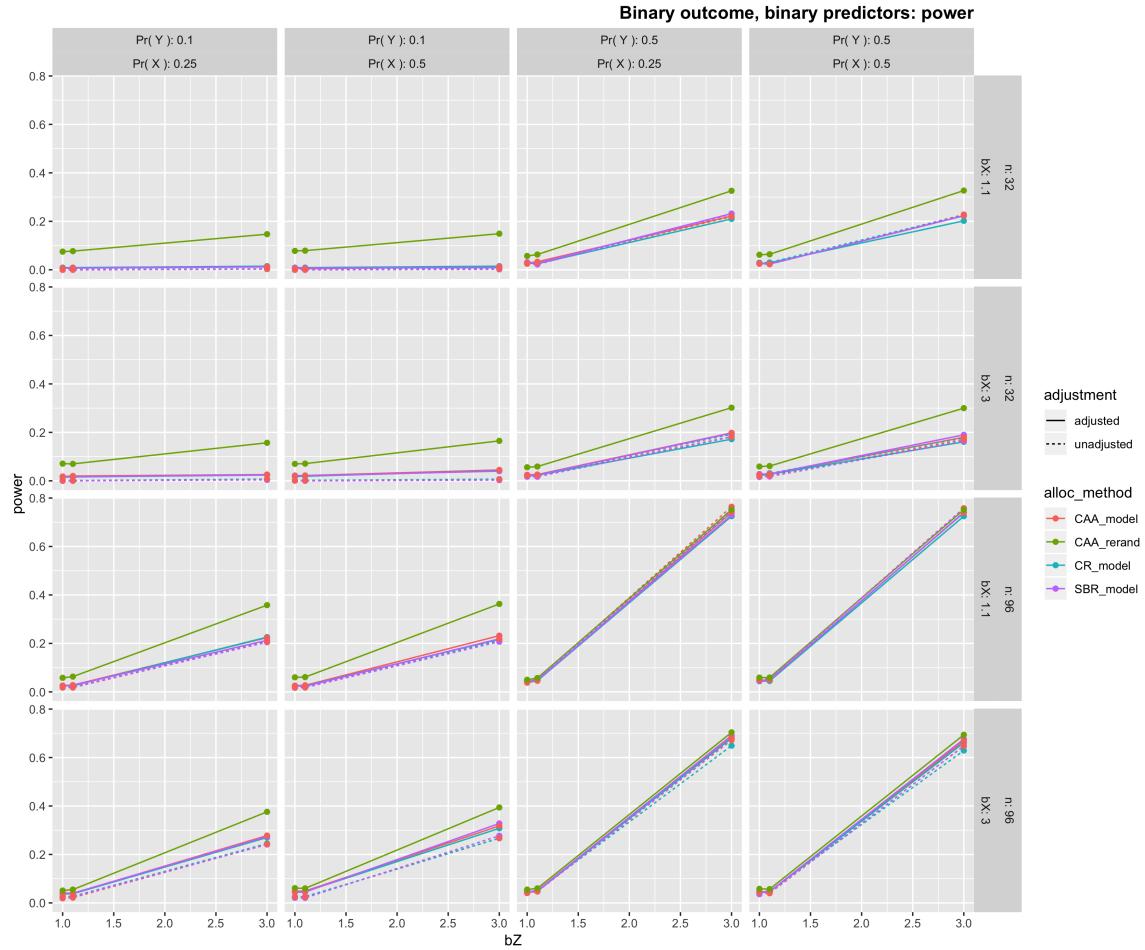


Figure 7.1: Batch 1: Power

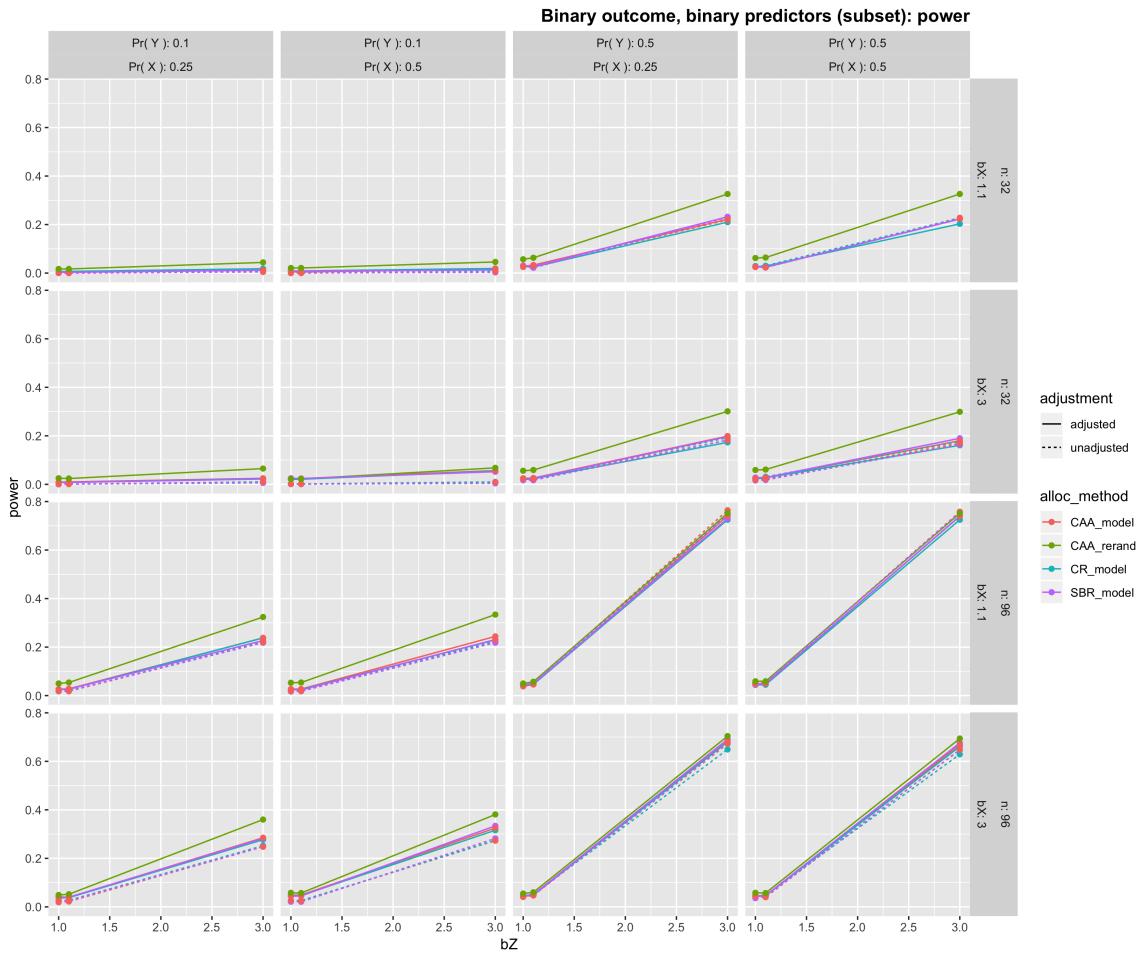


Figure 7.2: Batch 1 subset: Power

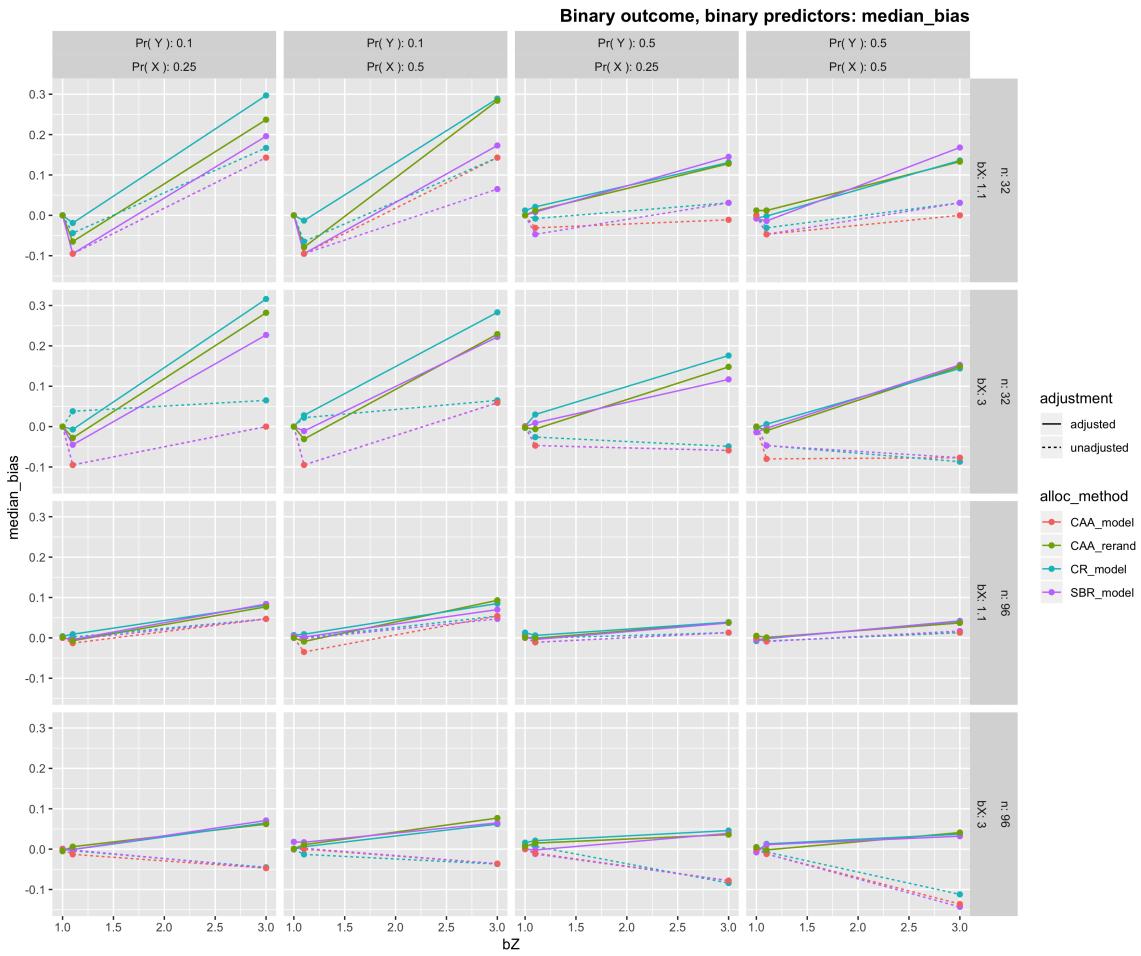


Figure 7.3: Batch 1: Median bias

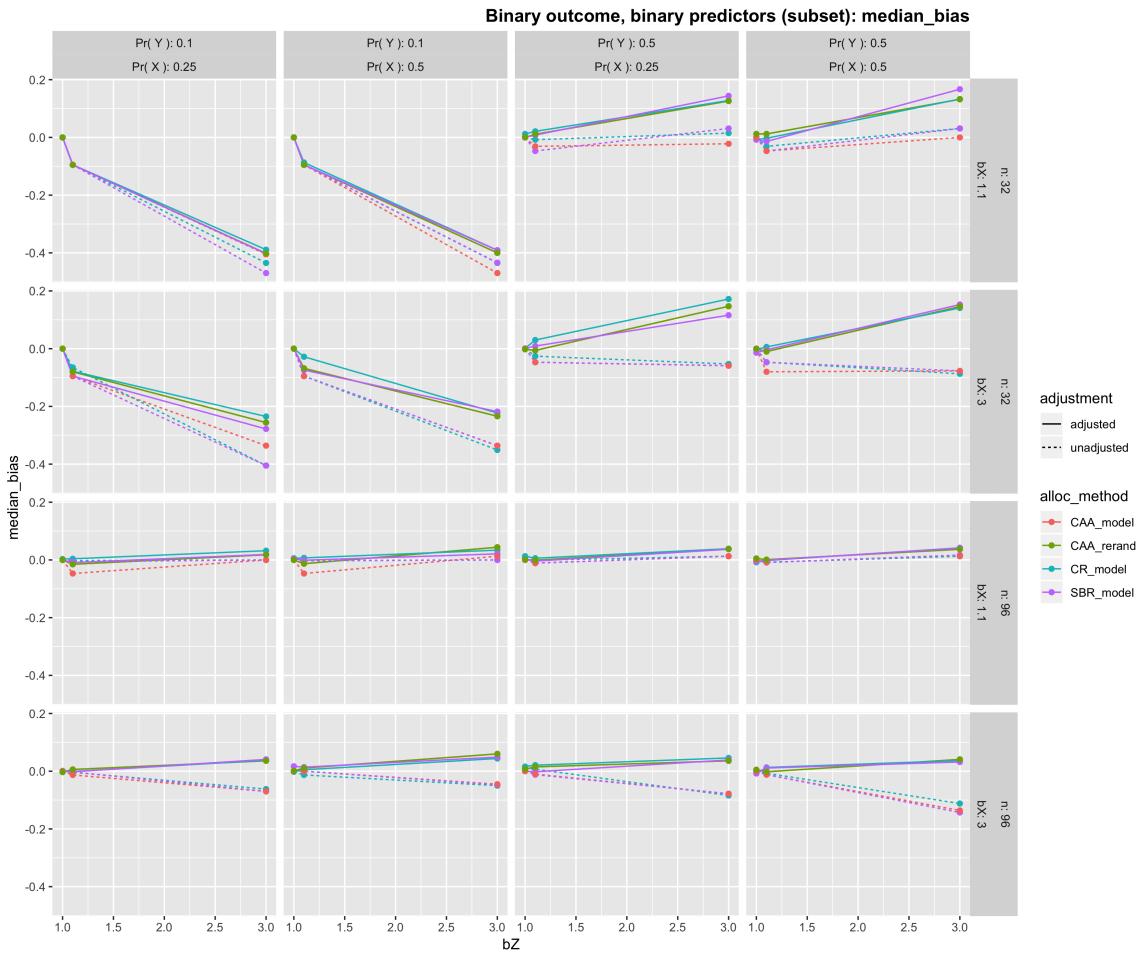


Figure 7.4: Batch 1 subset: Median bias

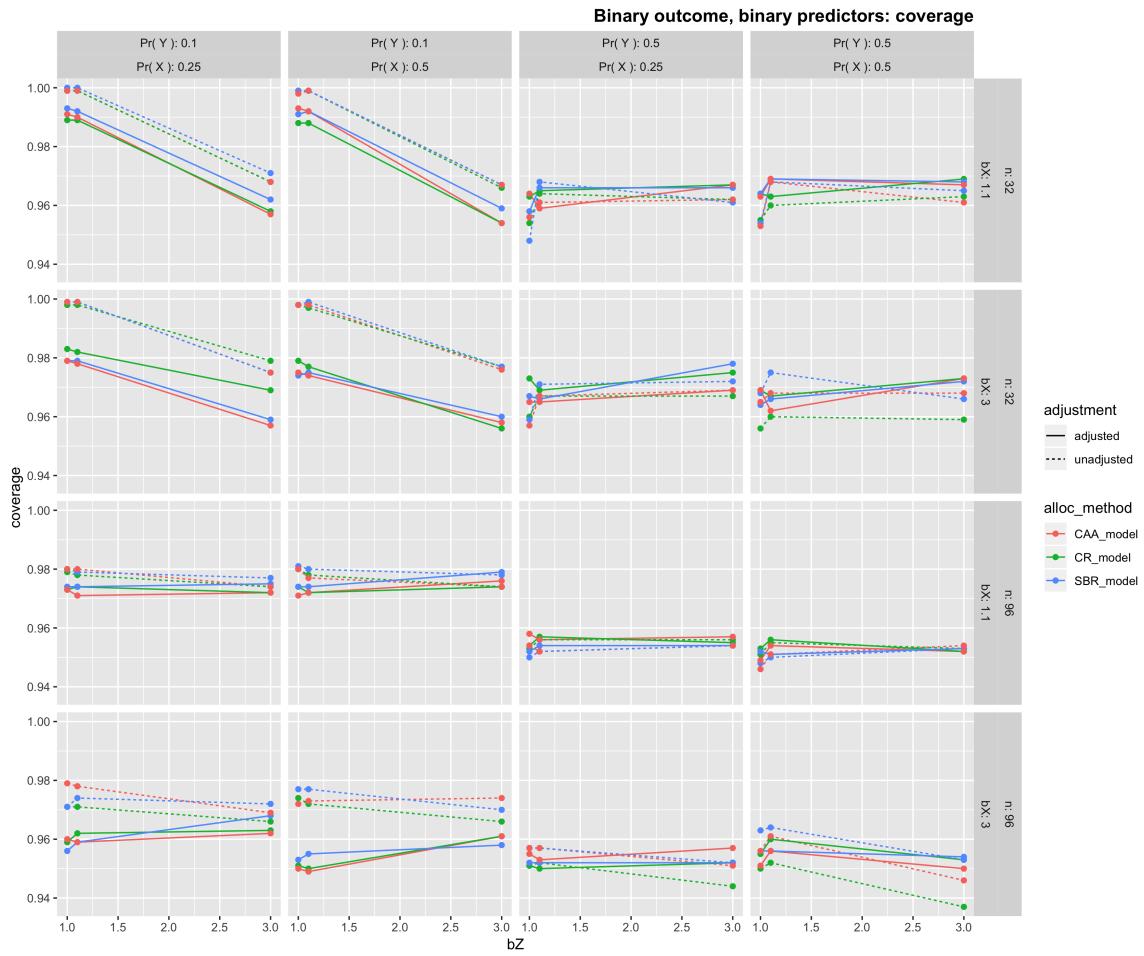


Figure 7.5: Batch 1: Coverage probability

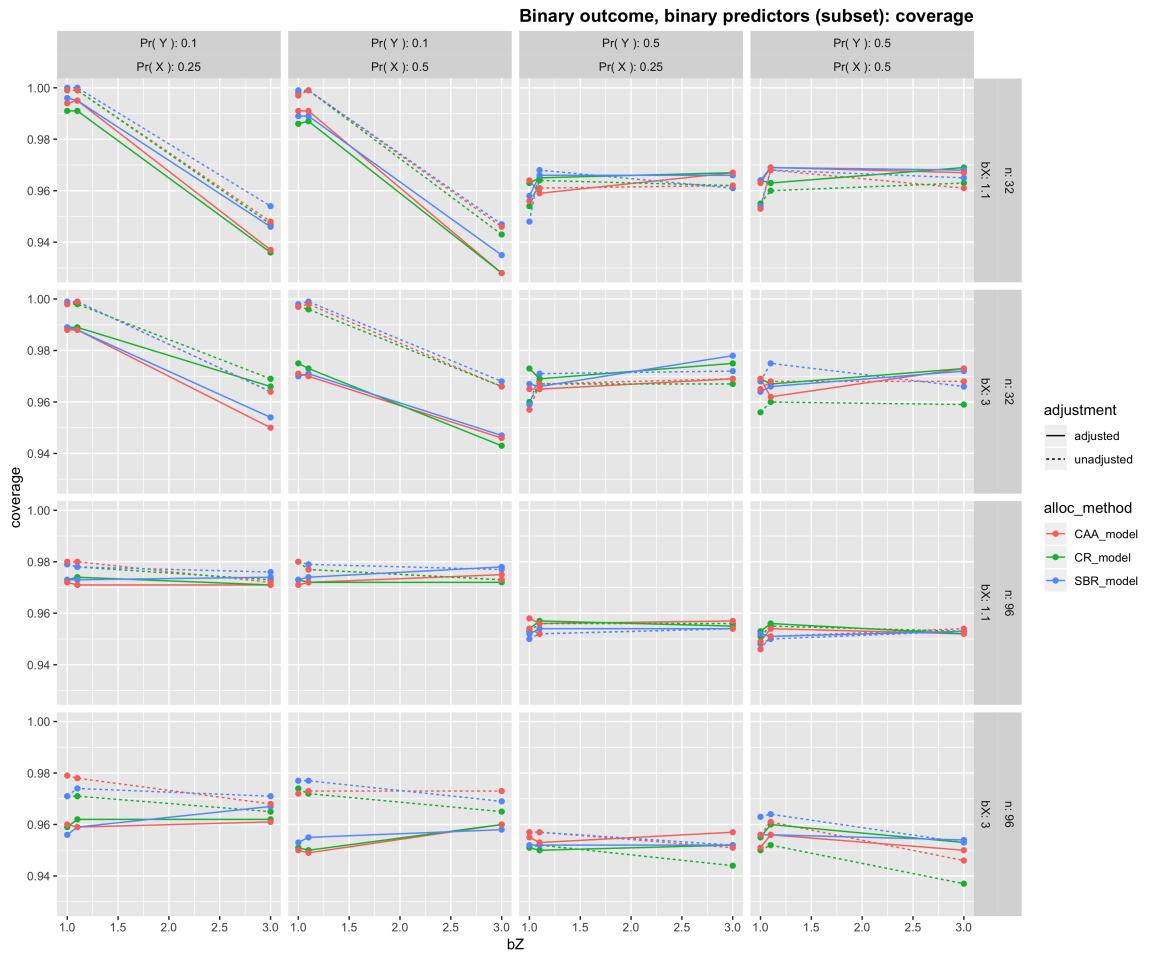


Figure 7.6: Batch 1 subset: Coverage probability

## 7.2 Batch 2: Binary Outcome, Continuous Predictors

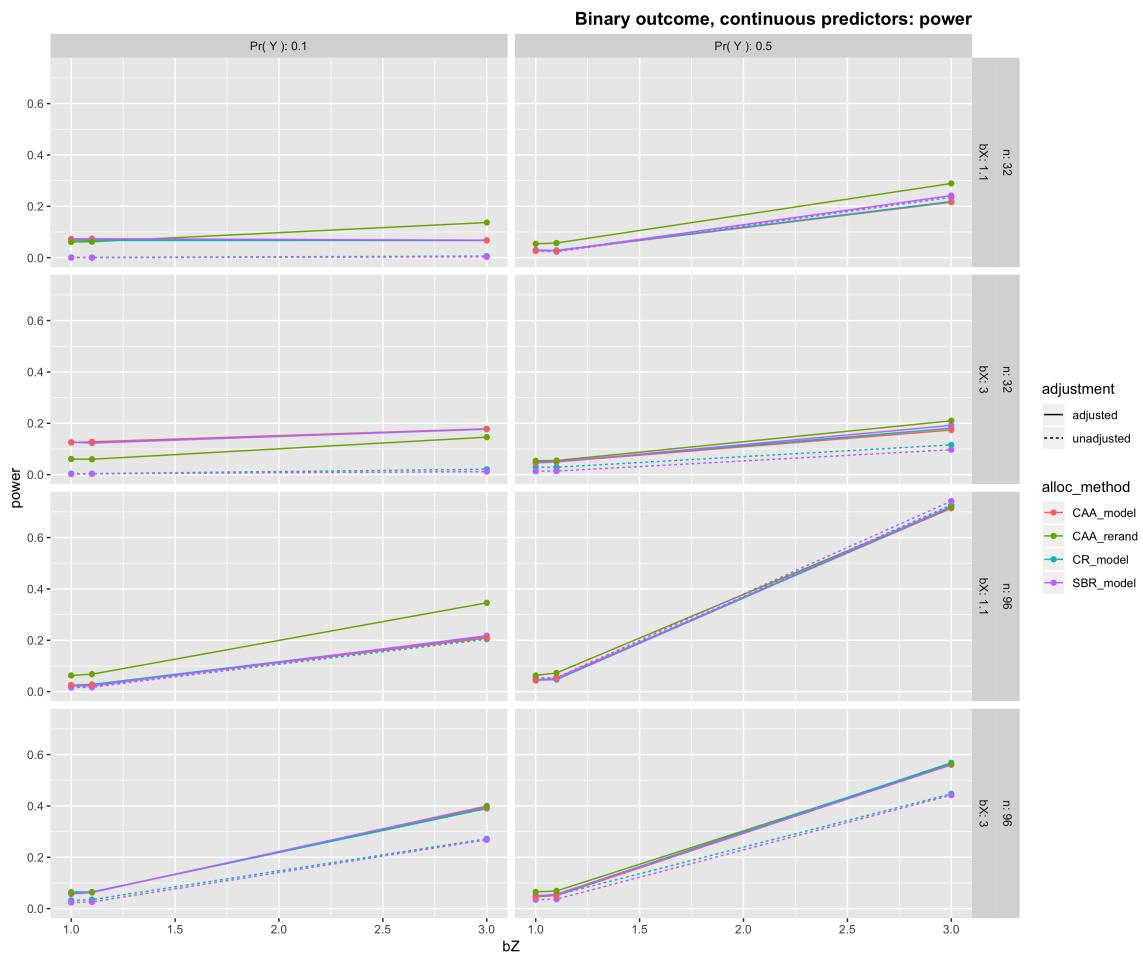


Figure 7.7: Batch 2: Power

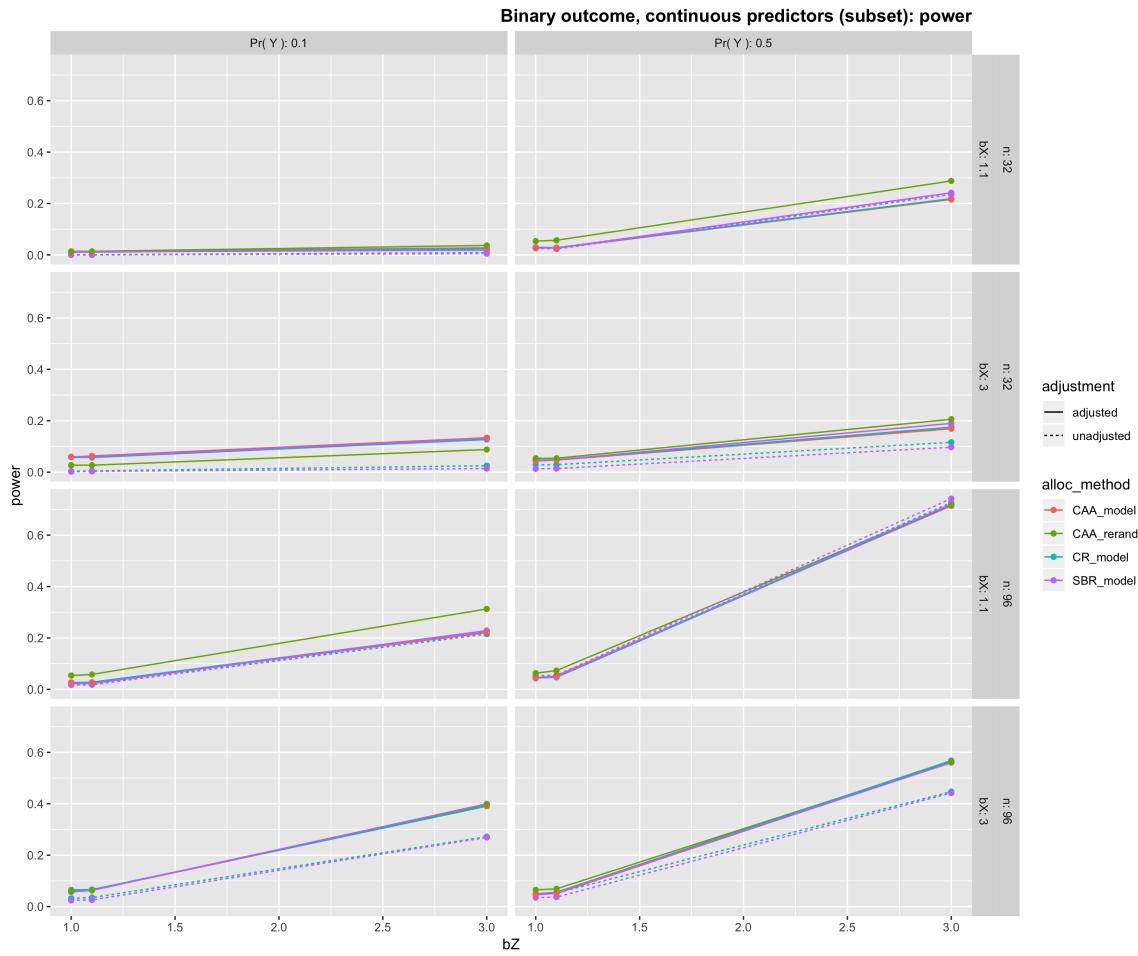


Figure 7.8: Batch 2 subset: Power

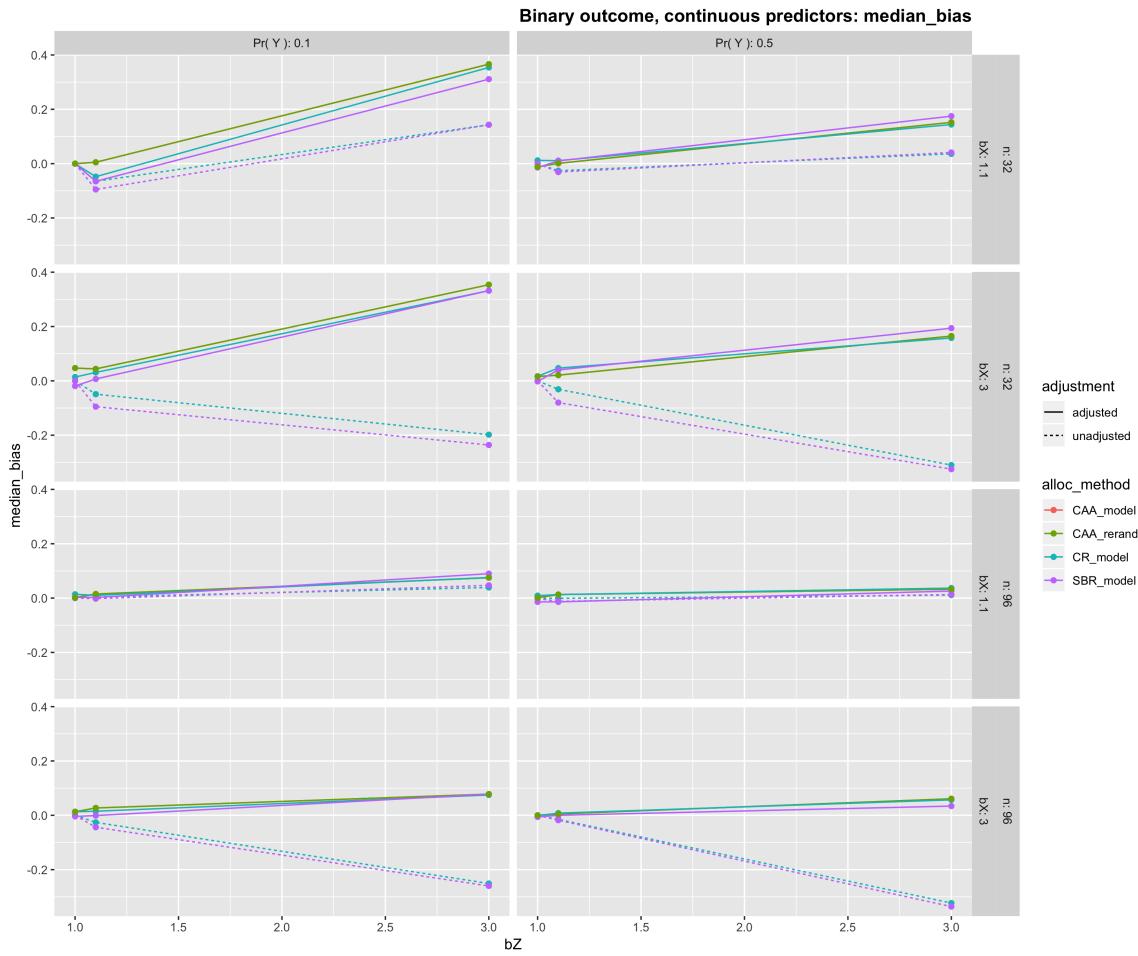


Figure 7.9: Batch 2: Median bias

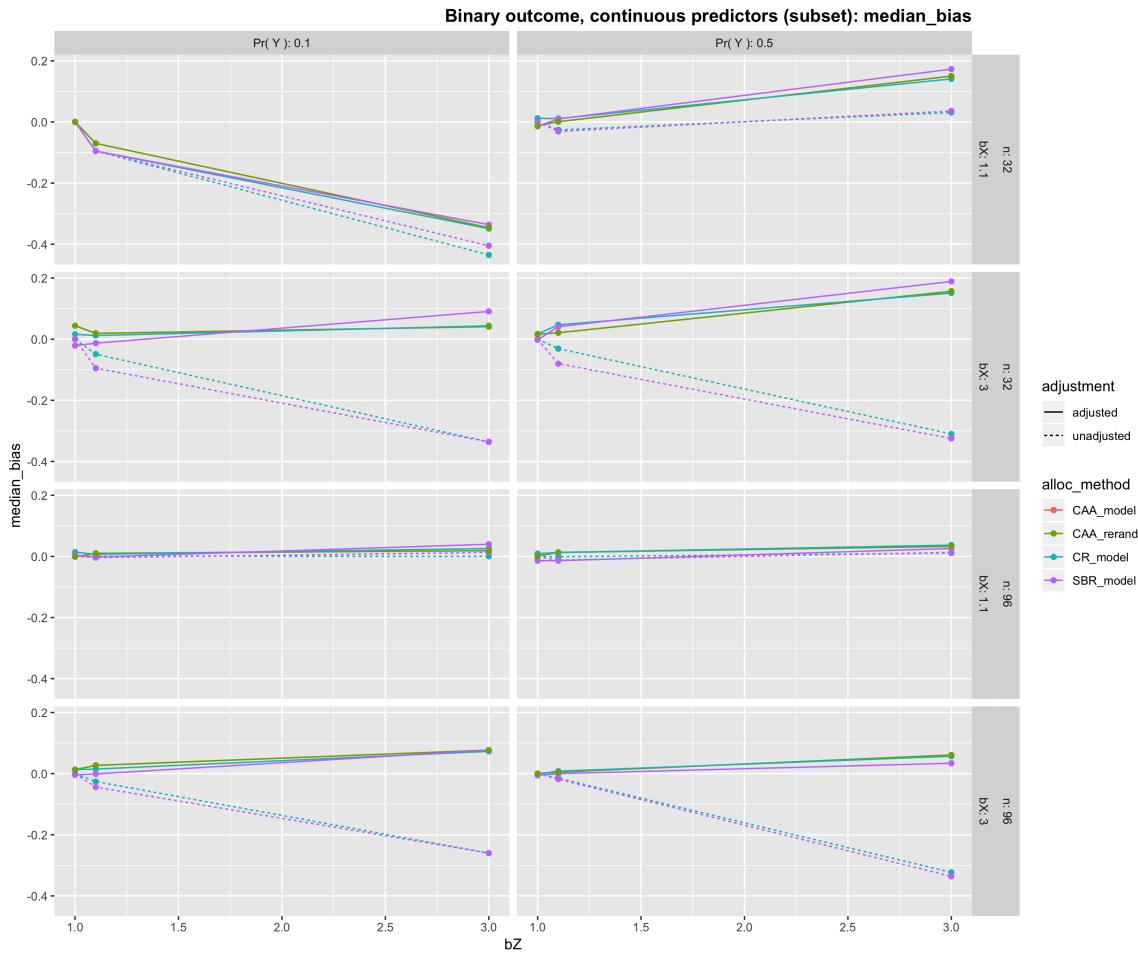


Figure 7.10: Batch 2 subset: Median bias

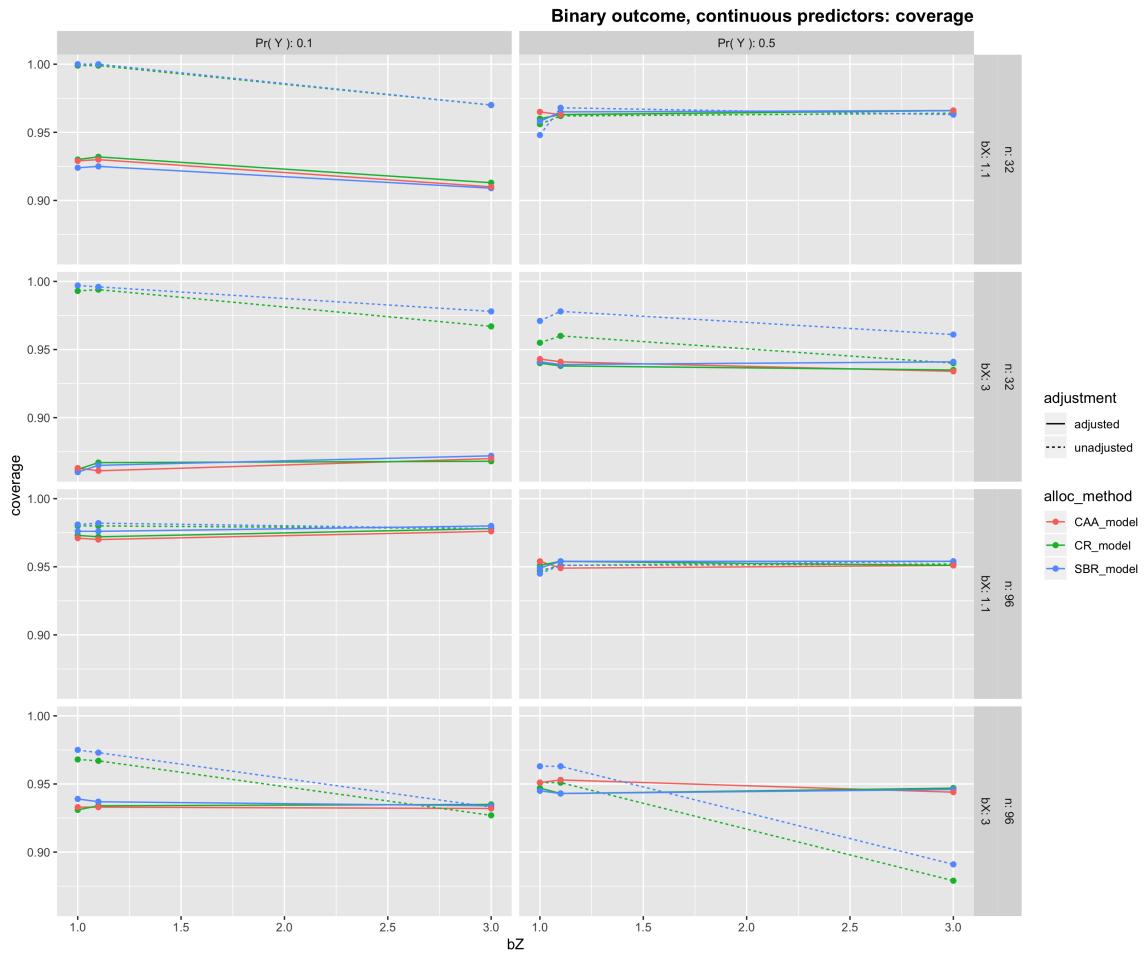


Figure 7.11: Batch 2: Coverage probability

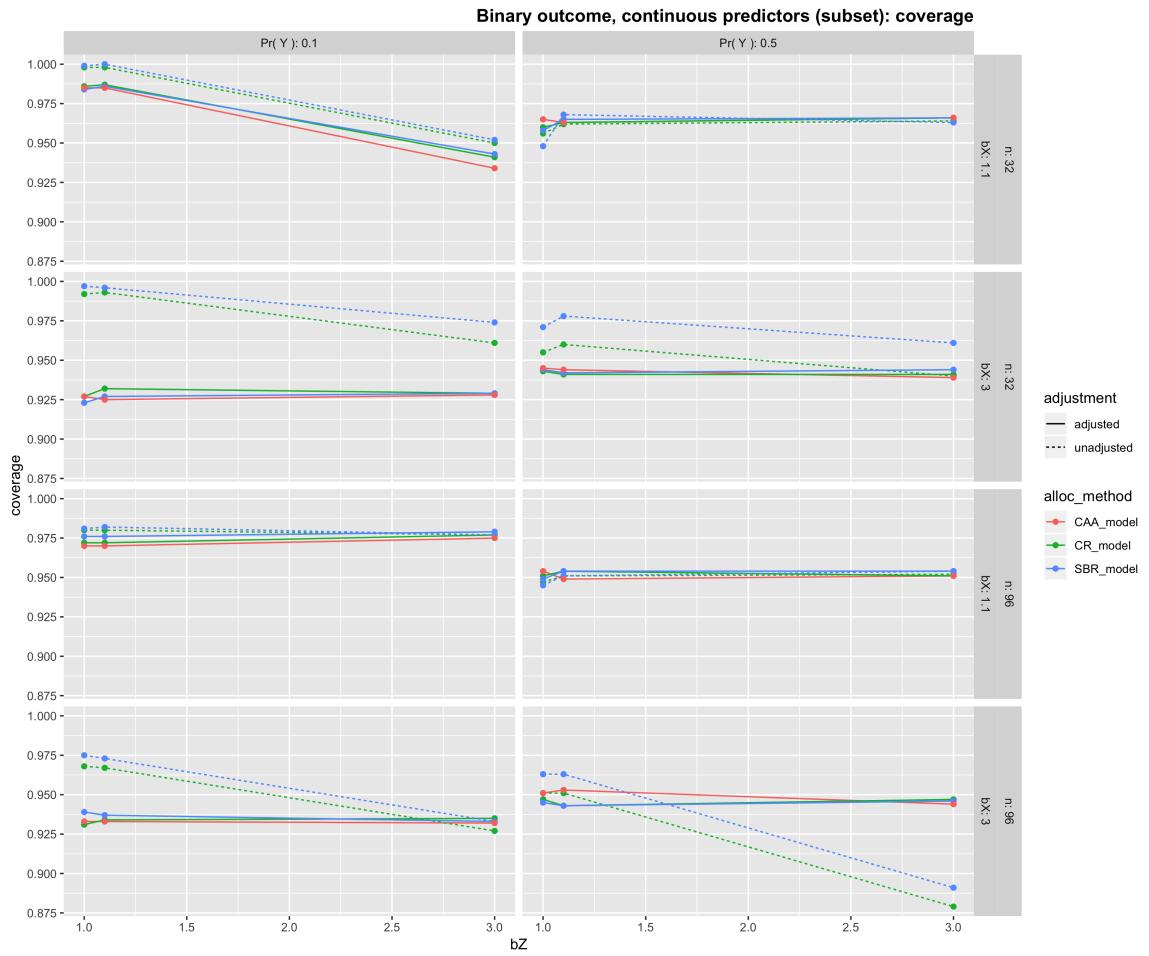


Figure 7.12: Batch 2 subset: Coverage probability

## 7.3 Batch 3: Continuous Outcome, Binary Predictors

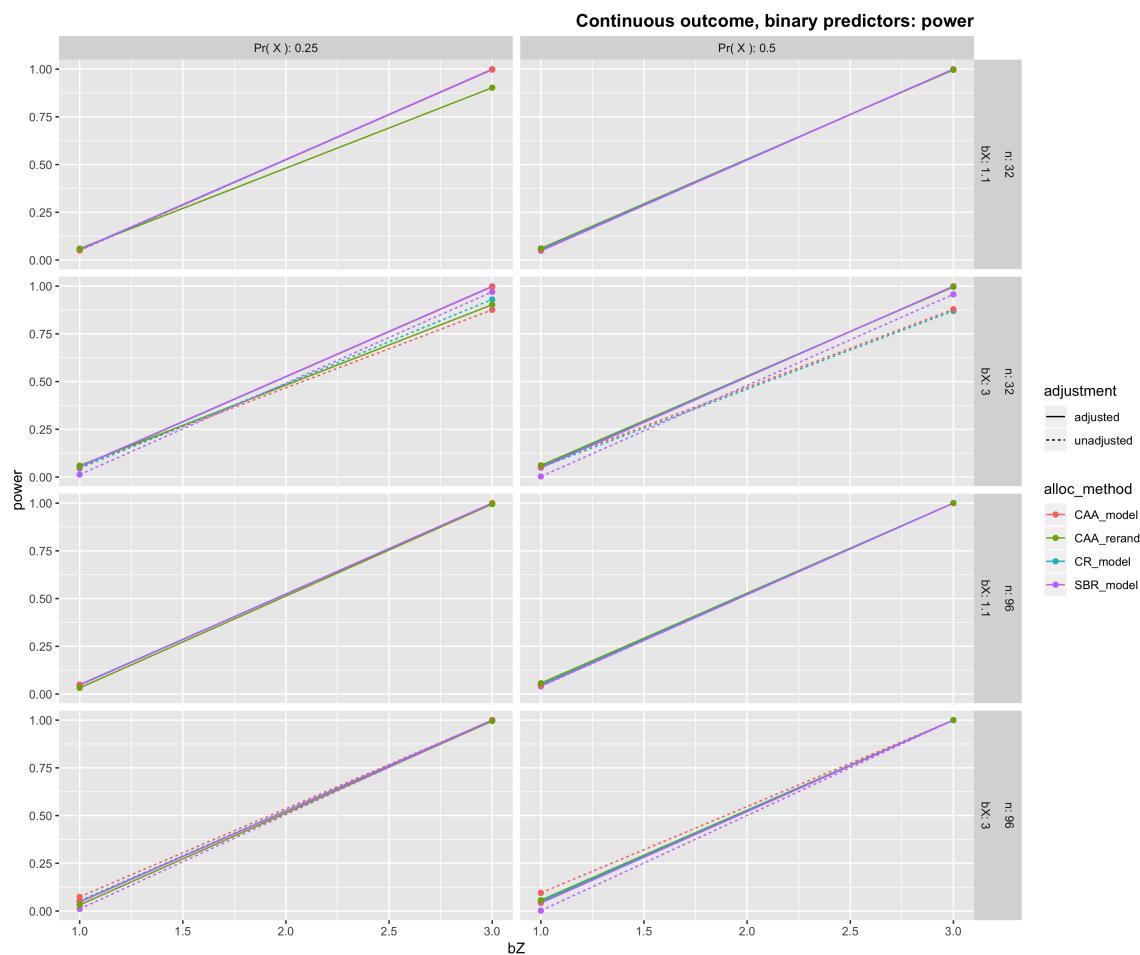


Figure 7.13: Batch 3: Power

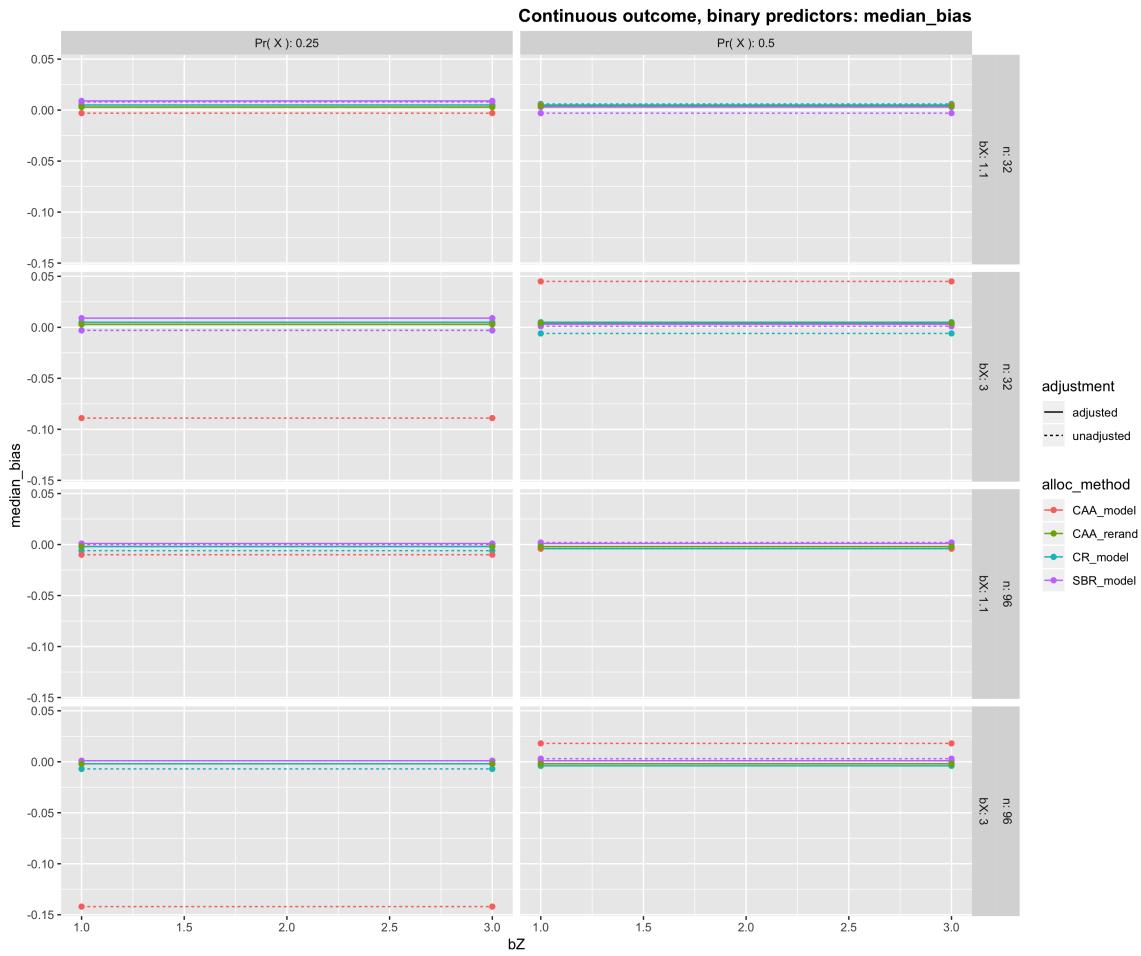


Figure 7.14: Batch 3: Median bias

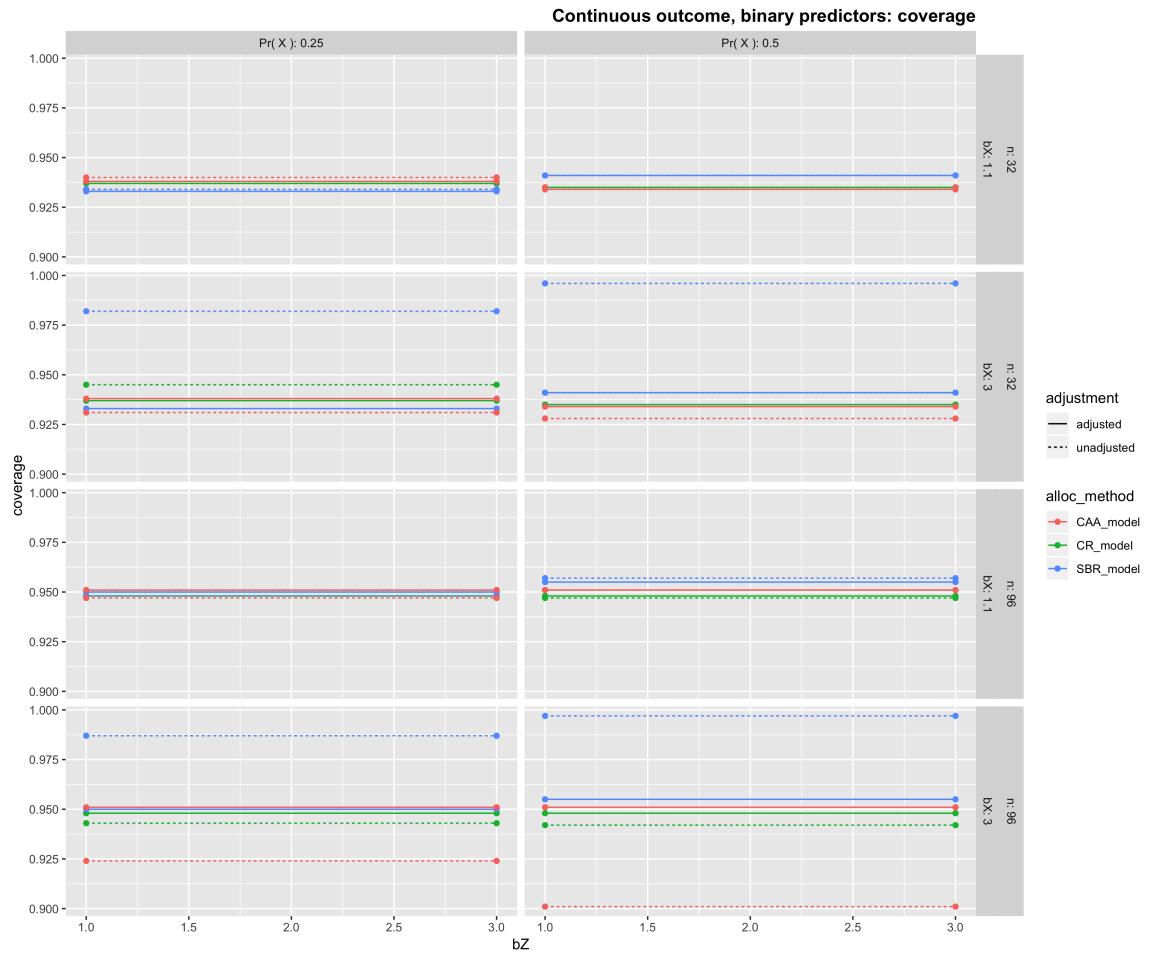


Figure 7.15: Batch 3: Coverage probability

## 7.4 Batch 4: Continuous Outcome, Continuous Predictors

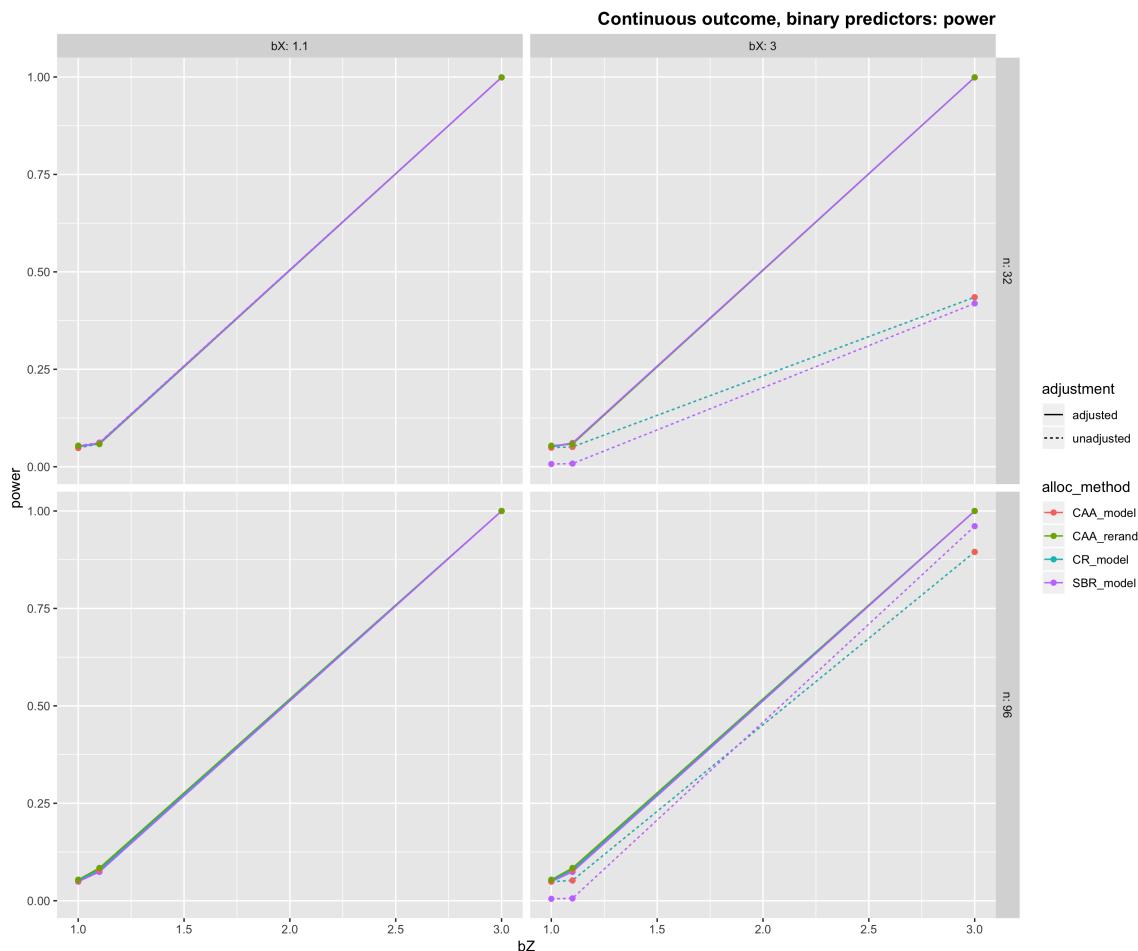


Figure 7.16: Batch 4: Power

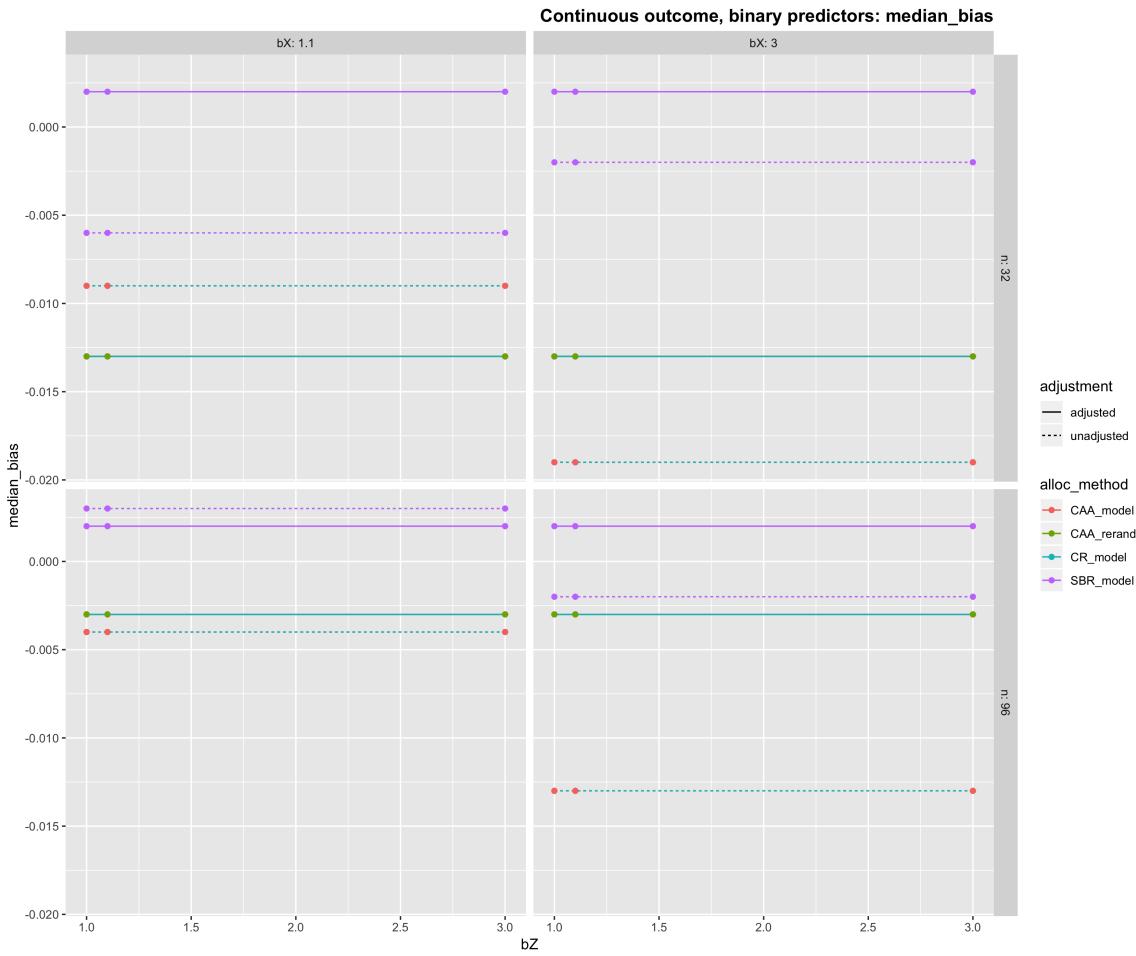


Figure 7.17: Batch 4: Median bias

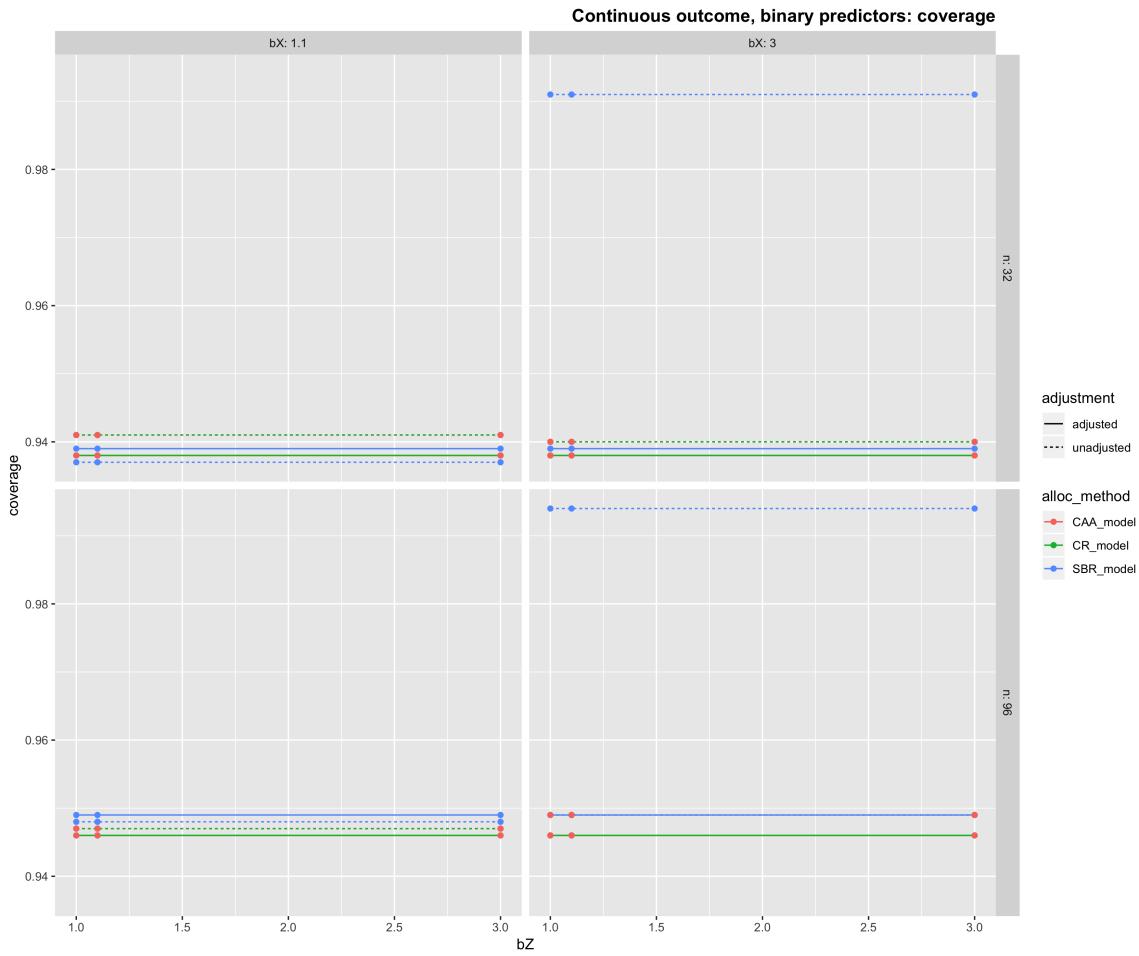


Figure 7.18: Batch 4: Coverage probability