METHODOLOGY
European Journal of Research Methods for the Behavioral and Social Sciences

PsychOpen GOLD

Original Article

Check for updates

A Simulation-Based Comparison of Minimization, Rerandomization, and Anticlustering for Creating Experimental Conditions

Martin Papenberg [1,2], Tim Angelike [1,3]

[1] *Department of Experimental Psychology, Heinrich Heine University, Düsseldorf, Germany.* **[2]** *Center for Digital Medicine, Heinrich Heine University, Düsseldorf, Germany.* **[3]** *Department of Psychology, Philipps-Universität, Marburg, Germany.*

Methodology, 2026, Vol. 22(2), 127–150, https://doi.org/10.5964/meth.17973

Received: 2025-05-09 • **Accepted:** 2026-03-08 • **Published (VoR):** 2026-06-30

Handling Editor: Juraj Medzihorsky, Durham University, Durham, United Kingdom

Corresponding Author: Martin Papenberg, Heinrich-Heine-Universität, Düsseldorf, Institut für Experimentelle Psychologie, Universitätsstraße 1, 40225 Düsseldorf, Germany. E-mail: martin.papenberg@hhu.de

Supplementary Materials: Code, Data, Materials [see Index of Supplementary Materials]

Abstract

Anticlustering has been used as a novel method to assign subjects to conditions in experiments. Anticlustering can be applied when covariate measurements are available at the beginning of an experiment and minimizes differences in covariates between conditions. In a simulation study implementing a two-group between-subjects design, we compared anticlustering with established methods for minimizing covariate imbalance: rerandomization and minimization. Anticlustering most strongly reduced covariate imbalance, followed by rerandomization and minimization. Lower covariate imbalance increased the precision of the effect size estimate. The average statistical power of the unadjusted analysis (independent *t*-test) was not improved when using covariate-based assignment as compared to random assignment. However, with random assignment, the statistical power of the unadjusted analysis depended on observed covariate imbalance; with covariate-based assignment, the statistical power of the unadjusted analysis was less affected by covariate imbalance because imbalance was minimized. Statistical adjustment via regression was most important to maximize statistical power.

Keywords

Anticlustering, balancing covariates, experimental design, clinical trials

The gold standard for conducting experiments is random assignment of subjects to conditions, allowing to establish the causal effect of an experimental intervention on an outcome. While preventing systematic differences between conditions, random assignment does not prevent that experimental groups randomly differ with regard to covariates that influence the outcome (Morgan & Rubin, 2012; Nguyen et al., 2017; Senn, 2013). The need to deal with nuisance variance due to measured covariates has long been recognized. However, there is an ongoing debate about how to address it: Should covariates be actively balanced among experimental conditions — "messing" with the traditional random assignment — or is statistical adjustment of covariates sufficient (Sella et al., 2021; Senn, 2022; Taves, 1974; Toorawa et al., 2009; Treasure & MacRae, 1998)? According to Hu et al. (2014), there are three main advantages of actively balancing covariates. First, balancing allows for a more precise estimate of the effect of an experimental manipulation; second, balancing increases the credibility of the analysis by increasing the comparability of the different conditions; third, with balancing, the analysis is less prone to misspecification of the statistical model (Qin et al., 2024). Consequently, there seems to be a clear intuitive impetus towards active balancing (cf. Senn, 2005). Practitioners as well as researchers tend to view model-based adjustment of covariates with skepticism if considerable imbalance is observed, and prefer to prevent imbalances altogether (Coart et al., 2023).[1] If covariate imbalance occurs as a consequence of random assignment (Nguyen et al., 2017), statistical adjustment seems to leave an "air of uncertainty about the validity of the conclusion[s]" (Treasure & MacRae, 1998, p. 362).

Many methods have been brought forward as potential enhancements of random assignment that increase covariate balance between experimental conditions (Lin et al., 2015). Moreover, the development and investigation of new covariate-balancing techniques is an active area of study (e.g., Ma et al., 2024; Qin et al., 2024; Zhang et al., 2024). When a covariate is categorical (e.g., biological sex or socioeconomic status), stratification techniques such as stratified block randomization are frequently applied in clinical studies, ensuring that combinations of categories — so called strata — are balanced between treatment and control conditions (for overviews, see Kernan et al., 1999; Lin et al., 2015). Stratification is based on assignment rules, for example through a permuted block design, that balances the occurrence of each condition within each stratum (Pocock & Simon, 1975). However, blocking rules cannot be applied when there are many strata or when covariates are numeric (Morgan & Rubin, 2012). Balancing numeric variables typically requires an automated and algorithmic assignment that employs a mathematical notion of covariate imbalance between conditions. In this paper, we investigate anticlustering, which uses an objective function to quantify covariate imbalance and uses an algorithmic assignment to optimize balance. We compared anticlustering to

1) The German Wikipedia site on Analysis of Covariance even claims that covariate adjustment should not at all be used in clinical experiments ("Kovarianzanalyse (Statistik)," 2025).

conceptually similar methods (minimization and rerandomization) that also use objective functions to quantify covariate imbalance.

Minimization is a traditional method to achieve balance in numeric covariates between conditions (Pocock & Simon, 1975; Taves, 1974; Treasure & MacRae, 1998). It can be used in sequential designs where subjects enter an experiment one after another. Minimization chooses for each new subject the experimental condition in such a way that covariate imbalance is reduced most effectively. In a sequential design, selecting the imbalance-minimizing condition will often be a deterministic choice, which may be undesirable in clinical experiments where the experimenter should not have definitive knowledge of the condition to which a subject is assigned (Pocock & Simon, 1975; Sella et al., 2021). It is therefore possible to induce a "biased coin" procedure where the balance-maximizing condition is only selected with a probability $p < 1$. Whether such a random element is necessary, however, depends on the circumstances of a particular experiment, for example on regulatory expectations.

In nonsequential experiments, covariate measurements for all subjects may be available at the beginning of an experiment. Hence, the conditions can be allocated using all covariate information, thereby potentially reducing covariate imbalance more strongly than would be possible in sequential experiments (Kapelner et al., 2021). Rerandomization is a popular method developed by Morgan and Rubin (2012) that can be used to allocate conditions in nonsequential experiments. Rerandomization first formalized sentiments about repeating a random assignment when a first attempt "went wrong" and resulted in high imbalance among conditions. Rerandomization consists of the following steps (Morgan & Rubin, 2015): (a) recruit experimental subjects and and collect all covariate measurements; (b) define a criterion of covariate imbalance that is deemed "acceptable"; (c) randomize subjects to conditions; (d) compute covariate balance; (e) if covariate balance is not acceptable, repeat randomization until it is; (f) start the experiment. Morgan and Rubin (2012) used the mahalanobis distance of covariate means between treatment and control group as criterion of covariate imbalance that should be minimized.

A critical choice for rerandomization is selecting a maximum amount of imbalance that is deemed "acceptable". The decision criterion has usually been formalized as an acceptance probability. The acceptance probability roughly corresponds to the extremeness of the observed mahalanobis distance as compared to the distribution of all mahalanobis distances, which is usually approximated using the χ^2 distribution (e.g., Zhang et al., 2024). A lower acceptance probability leads to less imbalance in covariates. However, low acceptance probabilities potentially lead to longer run times of the rerandomization method, and possibly prevent finding acceptable solutions altogether — especially when the number of covariates increases (Qin et al., 2024).

Though not developed for this purpose, anticlustering has recently been used as a new method to achieve covariate balance in nonsequential experiments (e.g., Stadelmann

et al., 2025). Anticlustering can be used to partition a pool of experimental subjects into disjunct groups with the aim of maximizing similarity between groups, based on the available covariates (Späth, 1986). Due to the availability of the free and open source software package `anticlust` (Papenberg & Klau, 2021), anticlustering has been applied increasingly in a wide range of research fields, including psychology (Nagel et al., 2024; Schaper et al., 2023), chemistry (Rahu et al., 2024), machine learning (Mauri et al., 2023), and cell biology (Papenberg, Wang, et al., 2025). Even though the original introduction of `anticlust` did not discuss using it for allocating experimental conditions, it has already been used for this purpose (Even et al., 2023; Stadelmann et al., 2025; Tuti et al., 2022).

Anticlustering is characterized by employing objectives of between-group balance that are known from cluster analysis. In the context of anticlustering, subjects are assigned to groups in such a way that these objectives are maximized; in cluster analysis, the same objectives would be minimized (Brusco et al., 2020). As already recognized by Späth (1986), when maximizing the *variance* — i.e., the objective function used in *k*-means clustering — it is possible to create groups that are similar to each other. In particular, *k*-means anticlustering equates the mean values of the covariates between groups. Thereby, *k*-means anticlustering is conceptually similar to rerandomization, which also minimizes differences in covariate means. However, rerandomization uses the mahalanobis distance between covariate means, thereby controlling for the covariance between covariates: Covariates that are correlated with other covariates obtain a lower weight than covariates that provide unique information. *K*-means anticlustering uses the same weighting for all covariates regardless of their covariances.

Using the mahalanobis distance is an appealing feature of rerandomization that so far has not been incorporated into anticlustering. However, anticlustering offers several advantages over rerandomization, which may make its use beneficial in certain circumstances. First, anticlustering is naturally defined for an arbitrary number of groups and not only two. While Morgan and Rubin (2012) discuss that criteria used in MANOVA testing can be adapted for rerandomization for more than two groups, we are not aware of actual implementations for three or more groups. The extension to three and more groups would also require investigating suitable values for the acceptance probability, which are probably different from suitable values in the context of just two groups. The acceptance probability is a necessary parameter in rerandomization, but not required in anticlustering. Moreover, anticlustering can also employ other objective functions than just the *k*-means criterion, which only minimizes differences in covariate means. For example, the *k*-plus objective incorporates higher order moments such as the variance, skewness and kurtosis, in addition to means (Papenberg, 2024). This feature may particularly relevant in the case of nonlinear covariate effects (see Ma et al., 2024).

To summarize, when measurements are available at the beginning of the experiment, anticlustering can be applied in a wide range of settings: It can be used to assign

subjects to an arbitrary number of conditions; it handles an arbitrary number of covariates; it scales from small to large experiments (Papenberg & Klau, 2021); it handles categorical as well as numeric covariates (Papenberg, Wang, et al., 2025); and it can be used to balance different properties of the data such as means and variances of covariates (Papenberg, 2024). However, the benefits of using anticlustering for allocating experimental conditions have not been previously investigated. Because anticlustering has however already been applied for this purpose, such an evaluation is much-needed. We close the gap in the current paper using a simulation study that allowed to establish a ground-truth for the presence of covariate effects on the outcome to be predicted. We compared anticlustering with standard random assignment, minimization and rerandomization, which are well-known methods for minimizing covariate imbalance (Coart et al., 2023; Morgan & Rubin, 2012; Sella et al., 2021; Treasure & MacRae, 1998). In the simulation, we implemented the most simple design — but arguably one of the most important designs as well — that can be used to adjust for covariates in experimental research: It implemented one experimental factor that varied in two levels, one numeric covariate,[2] and one numeric dependent variable. Such a design can be analysed using a simple *t*-test when the covariate is not adjusted, and ANCOVA or regression when the covariate is adjusted. As previous research has amplified the need to carefully select prognostic covariates for analysis (Raab et al., 2000; Wysocki et al., 2022), we included an additional condition where five "noise variables" were generated that were unrelated to the outcome. We implemented (a) random assignment, (b) assignment via anticlustering based on the covariate(s), (c) assignment via minimization based on the covariate(s) (Sella et al., 2021; Treasure & MacRae, 1998), and (d) assignment via rerandomization based on the covariate(s) (Morgan & Rubin, 2012). We applied both an unadjusted analysis (i.e., the *t*-test) and a covariate-adjusted analysis (i.e., regression) for all four assignment methods. As dependent variables, we investigated the degree of covariate imbalance between conditions.[3]

Method

We implemented a simulation study using the R programming language (Version 4.5.1, R Core Team, 2022). An accompanying Open Science Framework (OSF) repository contains all code and data required to fully reproduce the simulation and its analysis (see

2) We focused on numeric covariates in our study, but note that anticlustering can also effectively be applied to balance categorical covariates. Categorical covariates could also be dealt with by using stratification-based methods, which were not the focus of our study. Previous results, however, indicate that anticlustering performs better than stratification-based methods with regard to balancing multiple categorical variables (Papenberg, Wang, et al., 2025).

3) Note that the imbalance of covariates between experimental conditions is not the same as the size of the covariate effect, which was an independent variable in our simulation.

Papenberg & Angelike, 2026). The simulation was conducted to compare four different methods for assigning subjects to experimental conditions: standard random assignment, anticlustering, rerandomization, and minimization. For anticlustering, we used the implementation provided in the `anticlust` package (Version 0.8.10-1, Papenberg & Klau, 2021). We applied *k*-means anticlustering, which minimizes the mean values of covariates between conditions, via the default exchange algorithm. The exchange algorithm is a non-deterministic optimization method based on pairwise interchanges. It starts by randomly initializing the groups under the restriction of equal-sized conditions. It then iterates through all subjects and attempts to improve balance by swapping each subject with a subject that is currently assigned to a different condition. A subject is not swapped if no exchange would improve balance according to the objective function. The process stops after all possible exchanges have been evaluated for each subject. Balance could potentially be further improved by not stopping after a single iteration through all subjects; instead it is possible to repeat the iteration process until no single exchange is able to further improve the anticlustering objective, i.e., until a local maximum is found (Weitz & Lakshminarayanan, 1998). From the viewpoint of minimizing covariate imbalance, the local maximum method is unambiguously better. However, in the current study, we opted for using the non-locally optimal exchange method, which may be preferable in the context of designing experiments: Even when explicitly minimizing covariate imbalance, a remaining degree of randomness is desirable (Kapelner et al., 2021), and the non-locally optimal algorithm maintains a higher similarity to the initial random assignment. In any case, it is not possible to predict the final group allocation without knowledge about the random initialization. Therefore, anticlustering would arguably be in line with potential regulatory expectations demanding unpredictability of treatment assignment.

For minimization, we used the implementation provided by Sella et al. (2021), which uses the variance in covariate means as the objective to be minimized (Pocock & Simon, 1975). It can be used as a deterministic procedure or with a "biased coin" rule. We used the deterministic version that always selected the best possible condition with regard to covariate balance. While a random component may be preferable in some practical applications, using the deterministic procedure allowed us to investigate the best possible performance of minimization. Employing a random component would decrease observed balance, but offers no statistical advantages that would be useful in a simulation study. Thus, using the deterministic implementation of minimization yielded a more fair comparison to anticlustering and rerandomization that had the inherent advantage of being nonsequential methods. Even though all observations are available simultaneously in the simulation, the minimization method is based on sequentially assigning subjects to conditions. Hence, the *i*th subject is assigned to a condition using the information on covariate imbalance computed via the previous $i - 1$ subjects. For rerandomization, we used an implementation of the standard procedure of Morgan and Rubin (2012), provided

by Zhang et al. (2024). Rerandomization requires to define an acceptance probability used as the criterion for accepting an assignment based on covariate imbalance. Lower probabilities imply stricter requirements regarding observed balance between conditions. We set a strict acceptance probability of $p = .001$ following Morgan and Rubin (2015).

Data Generating Process

The following regression equation was assumed as the process to generate $4 \times N$ observations for each run of the simulation study:

$$y_{ij} = \beta_C C_i + \beta_T T_{ij} + \epsilon_i$$

Here, y_{ij} is the value on the dependent variable of the ith person ($i = 1, ..., N$) for the jth assignment method ($j = 1$: random assignment; $j = 2$: minimization; $j = 3$: anticlustering; $j = 4$: rerandomization). The values y_{ij} were generated as a linear combination of the individual manifestation on the numeric covariate (C_i), the treatment indicator variable ($T_{ij} \in \{0, 1\}$) and an error term (ϵ_i). For each simulation run, C_i and ϵ_i were drawn as random variates from a normal distribution, respectively ($M = 0$, $SD = 1$). The regression coefficient of the covariate β_C varied between 0 and 1 depending on the simulation condition (see following section). It can be interpreted as the strength of association between the covariate and the dependent variable. The regression coefficient β_T varied between 0 and 1 depending on the simulation condition; it can be interpreted as the strength of the effect of the treatment. The values T_{ij} were generated in dependence of the assignment method: via random assignment, via minimization, via anticlustering, or via rerandomization. Minimization, anticlustering and rerandomization took into account the individual covariate values C_i and sought to balance them between conditions. All four assignment methods always created equal-sized experimental conditions.

Design

The entirety of all simulation conditions was the Cartesian product of the following factors: The regression coefficient of the treatment β_T varied in 6 levels [0, 0.2, 0.4, 0.6, 0.8, 1]; the regression coefficient of the covariate β_C varied in 6 levels [0, 0.2, 0.4, 0.6, 0.8, 1]; the sample size N varied in 15 levels [20, 40, 60, ..., 280, 300]; the noise condition had two levels [noise, no noise]. When noise was introduced, we generated an additional 5 covariates that were used for the covariate-based assignment methods and the covariate-adjusted analyses, which were however not predictive of the outcome. In the "no noise" condition, these additional variables were not considered for assignment or adjustment. For each of the $6 \times 6 \times 15 \times 2$ conditions, we generated 1000 data sets. For each data set, 8 methods of data analysis were applied, which resulted from crossing the four assignment methods [random, minimization, rerandomization, anticlustering] with the method of analysis [adjusted, unadjusted]. For the unadjusted method, we com-

puted an independent *t*-test. For the adjusted method, we computed a linear regression with y_i as criterion and the covariate and treatment variable as additive predictors. For each test, the *p*-value associated with the treatment factor was recorded. As dependent variable, we determined whether each *p*-value was significant using the conventional $\alpha = .05$ level. In total, this resulted in $6 \times 6 \times 15 \times 2 \times 1000 = 1{,}080{,}000$ simulation runs and $6 \times 6 \times 15 \times 2 \times 1000 \times 8 = 8{,}640{,}000$ data points used in the analysis of the simulation. As additional criteria for evaluation, we computed the standardized effect size *d* (Cohen, 1988) to quantify the imbalance in covariate distribution between conditions (d_{cov}), and to compute the effect of the treatment (d_{treat}), respectively.

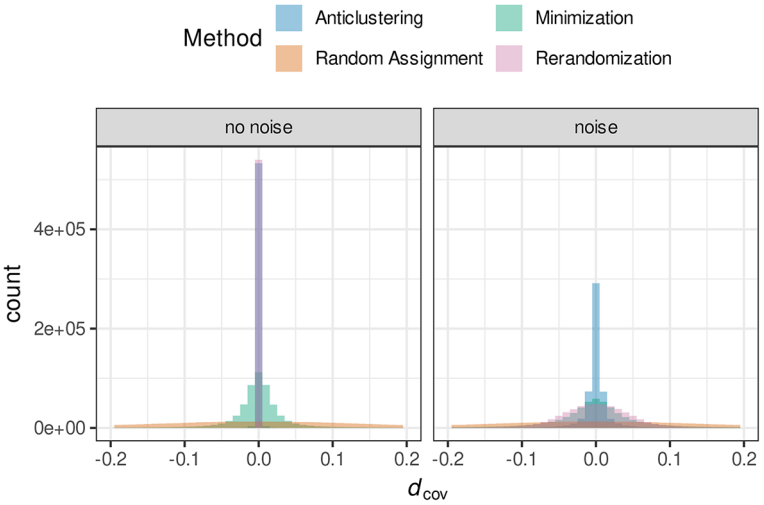
Results

Covariate Imbalance

Figure 1 shows the distribution of covariate imbalance by assignment technique (random, minimization, anticlustering, rerandomization), depending on whether additional noise variables were included or not. As expected, minimization reduced covariate imbalance as compared to random assignment. Rerandomisation and anticlustering exhibited an even more pronounced reduction in covariate imbalance, as shown by two highly peaked distributions of d_{cov} in the left panel of Figure 1. Without noise variables, the standard deviation of the distribution of d_{cov} was $SD = 0.001$ for anticlustering, $SD = 0.000$ for rerandomization, $SD = 0.053$ for minimization, $SD = 0.217$ for random assignment. When noise variables were included, the covariate-based assignments had more difficulties in balancing the prognostic covariate. This result was expected because achieving balance on multiple variables simultaneously is more difficult than achieving balance on only one variable alone (Papenberg, 2024). In this condition, anticlustering exhibited the most pronounced reduction in covariate imbalance, as shown by the single peaked distribution of d_{cov} in the right panel of Figure 1. With noise variables, the standard deviation of the distribution of d_{cov} was $SD = 0.032$ for anticlustering, $SD = 0.053$ for rerandomization, $SD = 0.105$ for minimization, and $SD = 0.216$ for random assignment.

Figure 1

Distribution of Covariate Imbalance for the Four Assignment Techniques When Noise Variables Were Included or Not



Note. Covariate imbalance was computed using the standardized effect size *d* (Cohen, 1988) for the standard normally distributed covariate. Note that the histograms were cut at $d_{cov} = -0.2$ and $d_{cov} = 0.2$ to properly portray the largest portion of the distribution. Only 7.6% of all values were outside of this interval, but there were outliers of up to $d_{cov} = -2.85$ and $d_{cov} = 2.42$.

Statistical Power

We computed the average number of significant results while excluding all simulation runs where the effect of the experimental condition was zero ($\beta_T = 0$) to obtain estimates of statistical power. Table 1 shows the global average statistical power for all methods, aggregated across all simulation conditions. Using covariate-based assignment alone did not improve power as compared to random assignment. Statistical power was however substantially increased by statistically adjusting for the covariate via regression. Using covariate-based assignment in addition to statistically adjusting for the covariate increased power marginally by less than 1 percentage point.

Table 1

Global Simulation Results

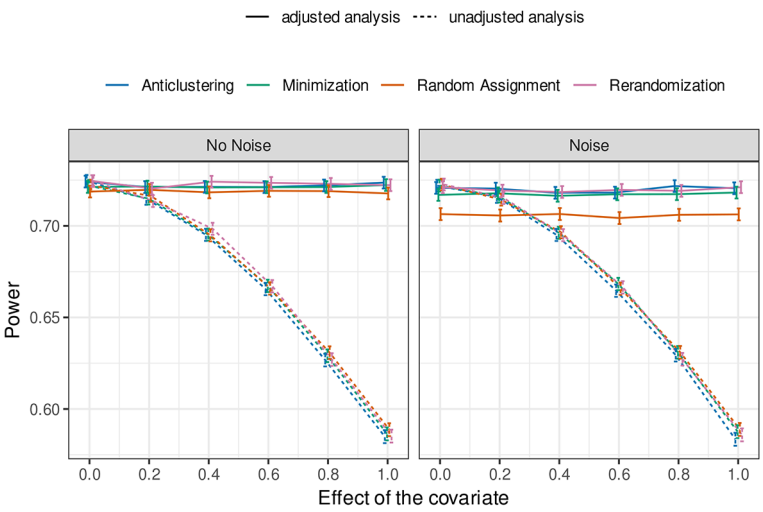
Analysis	Power	95% CI	n
Anticlustering + unadjusted analysis	.669	[.668,.669]	900000
Rerandomization + unadjusted analysis	.669	[.668,.670]	900000
Minimization + unadjusted analysis	.670	[.669,.671]	900000
Random Assignment + unadjusted analysis	.670	[.669,.671]	900000
Random Assignment + adjusted analysis	.712	[.711,.713]	900000
Minimization + adjusted analysis	.719	[.718,.720]	900000
Anticlustering + adjusted analysis	.721	[.720,.722]	900000
Rerandomization + adjusted analysis	.721	[.720,.722]	900000

Note. Statistical power of all methods, aggregated across all simulation conditions when there was an effect of the condition ($\beta_T > 0$); *n* refers to the numer of simulation runs in which $\beta_T > 0$.

Figure 2 shows that statistical power of the unadjusted analyses decreased when the effect of the prognostic covariate on the dependent variable increased. Statistically adjusting for the covariate mitigated the reduction in power completely regardless of how experimental groups were assigned. When noise variables were included in the analysis, statistical power was reduced slightly for the adjusted analysis. Interestingly, using covariate-based assignment techniques prevented this reduction in power due to the presence of noise variables. Average power for the covariate-adjusted analyses with noise variables was .720 [.719, .721] with assignment via anticlustering, .720 [.718, .721] with assignment via rerandomization, .717 [.716, .719] with assignment via minimization, and .706 [.705, .707] with random assignment. When no noise variables were present, there was basically no difference in power between the assignment techniques: Average power for the covariate-adjusted analyses was .722 [.721, .724] with assignment via anticlustering, .723 [.722, .724] with assignment via rerandomization, .721 [.720, .723] with assignment via minimization, and .719 [.717, .720] with random assignment. Hence, the overall increase in power due to balancing shown in Table 1 was due to the presence of additional non-prognostic noise variables.

Figure 2

Statistical Power by Covariate Effect and Noise Condition



Note. Error bars indicate the 95% confidence interval.

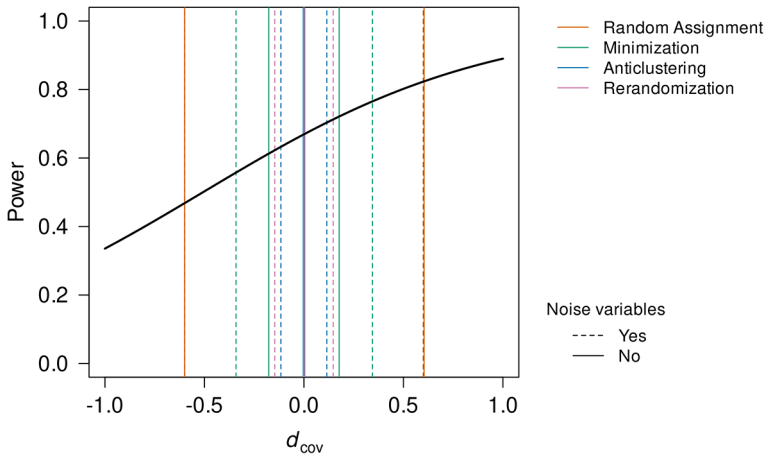
Table 1 showed that balancing covariates by assignment did not improve statistical power, even though nuisance variance was better balanced between conditions than when random assignment was applied (see Figure 1). This result probably conflicts with the intuition of many researchers who might assume that balancing covariates increases statistical efficiency. Why do our results conflict with this intuition? In fact, we must differentiate, (a) how the long-term properties (i.e., statistical power) of statistical tests are influenced by balancing, and (b) how the properties of an individual study are affected by balancing. As our results indicate, the long-term frequency of significant results is not influenced by actively balancing. However, the probability that an individual study finds a significant result is affected by covariate imbalance. To corroborate this sentiment, we predicted the probability that a *t*-test found a significant result. In particular, we computed a logistic regression predicting statistical significance (1 = significant; 0 = not significant) using covariate imbalance d_{cov} as predictor. Again, we only used the conditions when there was an effect of the treatment ($\beta_T > 0$).

The logistic regression coefficient of covariate imbalance was significant, $b = 1.39$, 95% CI [1.34, 1.44], $z = 56.80$, $p < .001$, showing that covariate imbalance was related to the probability of an individual study to find a treatment effect – if no statistical adjustment was applied. Figure 3 illustrates the relationship between covariate imbalance and statistical power, using the coefficients obtained from the logistic regression. It depicts a quite strong relationship between covariate imbalance and the probability to

find an effect. Depending on the direction of the covariate imbalance, power can be increased or decreased. That is, if the treatment has a positive effect on the outcome and the covariate also has a positive effect, finding a significant effect is facilitated if there are higher average values of the covariate in the treatment condition. However, finding a significant effect is aggravated if there are lower average values of the covariate in the treatment condition. This result explains why the long-term power is not negatively affected by covariate imbalance; imbalance can also facilitate finding an effect if the imbalance "favors" the treatment condition. Note that covariate imbalance did not predict statistical power of the regression analysis that adjusted for the covariate, as shown by another logistic regression, $b = -0.02$, 95% CI $[-0.07, 0.03]$, $z = -0.80$, $p = .427$.

Figure 3

Relationship Between Statistical Power and Covariate Imbalance for the Unadjusted t-Test



Note. The coloured vertical lines indicate the 1^{st} and 99^{th} percentile of observed covariate imbalance, depending on assignment method and the presence of noise variables. For random assignment, the dashed and solid lines overlap because covariate imbalance is not affected by the presence of noise variables if the assignment is entirely random. When no noise variables were included, anticlustering and rerandomization practically negated covariate imbalance. When noise variables were included, anticlustering led to the most pronounced reduction in covariate imbalance.

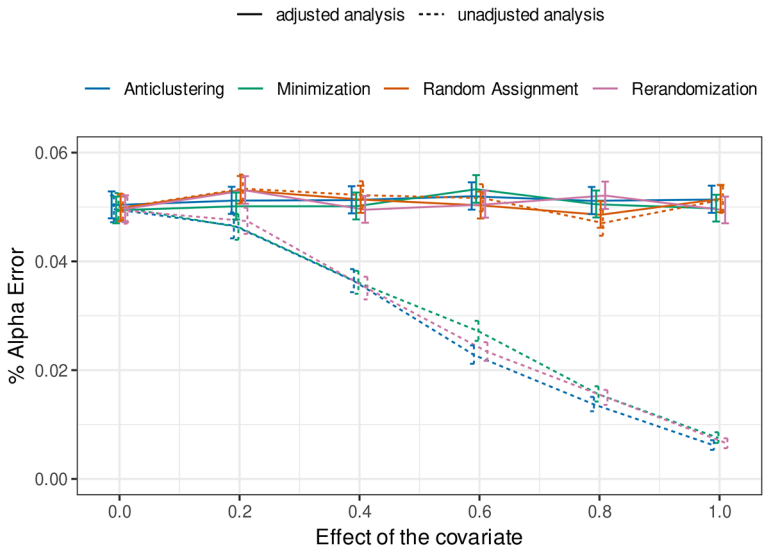
Alpha Error Probability

The following analysis only considered simulation conditions in which $\beta_T = 0$ to assess the occurrence of false positive results, i.e., alpha errors. Table 2 illustrates alpha error probability for each method. Interestingly, only using covariate-assignment without applying statistical adjustment decreased the probability of committing an alpha error. The other methods did not depart from the nominal alpha level of.05. Figure 4 illustrates the alpha error probability in dependence of the covariate effect: When variation due to the

covariate increased, using covariate-based assignment decreased the alpha error probability. Alpha error probability was most attenuated when the covariate effect was maximal ($\beta_C = 1$). While lower alpha errors may at first seem beneficial, they are indicative of a generally more conservative behaviour of the unadjusted test based on covariate-based condition assignment (e.g., Shao et al., 2010). In fact, the conservativeness obscured true statistical power, which was only revealed when using covariate-adjustment during analysis (see Table 1 and Figure 2).

Figure 4

Probability of Alpha Errors by Covariate Effect



Note. Error bars indicate the 95% confidence interval.

Table 2

Global Simulation Results

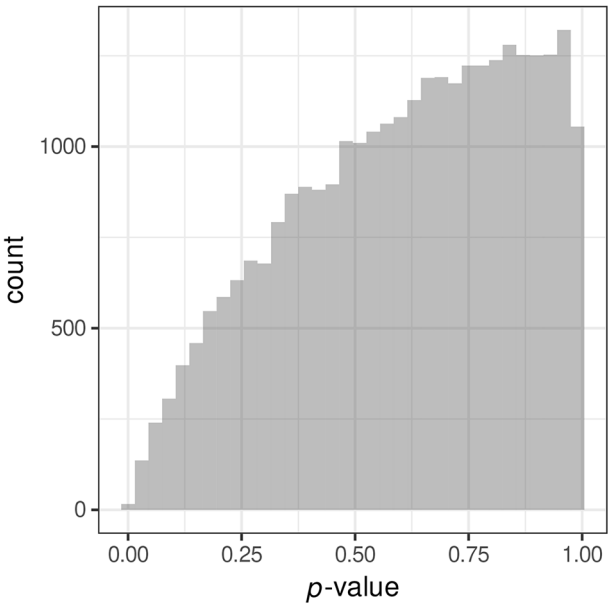
Analysis	% Alpha Error	95% CI	n
Anticlustering + adjusted analysis	.051	[.050,.052]	180000
Anticlustering + unadjusted analysis	.029	[.028,.030]	180000
Minimization + adjusted analysis	.051	[.050,.052]	180000
Minimization + unadjusted analysis	.030	[.030,.031]	180000
Random Assignment + adjusted analysis	.051	[.050,.052]	180000
Random Assignment + unadjusted analysis	.051	[.050,.052]	180000
Rerandomization + adjusted analysis	.051	[.050,.052]	180000
Rerandomization + unadjusted analysis	.029	[.029,.030]	180000

Note. Alpha error probability of all methods, aggregated across all simulation conditions when there was no effect of the condition ($\beta_T = 0$); *n* refers to the number of simulation runs in which $\beta_T = 0$.

Figure 5 illustrates the distribution of p-values in the condition with $\beta_C = 1$ (i.e., the largest effect of the covariate) and $\beta_T = 0$ (i.e., no condition effect) for the unadjusted analysis when using anticlustering assignment. Interestingly, the distribution of p-values was highly skewed, favoring larger p-values, thus explaining the low occurrence of alpha errors. With standard random assignment, p-values are uniformly distributed; that p-values diverge from the uniform distribution when using covariate-based assignment is quite striking and requires an explanation. In our simulation, when there was no treatment effect, variation in the dependent variable was generated based on the influence of the covariate (β_C) and random error (ϵ_i). Because the covariate was actively balanced between conditions, all of the systematic portion of variance in the outcome was evenly distributed between conditions. For this reason, there was less difference in the outcome between conditions than would be expected under random assignment, skewing the distribution of p-values, and decreasing the probability of committing an alpha error. Though the resulting conservativeness of the unadjusted test has also previously been proven theoretically (e.g., Ma et al., 2015, 2024; Shao et al., 2010), we believe that it is of educational value to reproduce the consequence of such proofs in an empirical simulation.

Figure 5

Distribution of p-Values of the Unadjusted t-Test



Note. Illustrates the distribution of all p-values when using anticlustering-based assignment, and a large covariate effect ($\beta_C = 1$) and no condition effect ($\beta_T = 0$) were present.

Effect Size of Treatment

Table 3 illustrates the average standardized effect size of the treatment d_{treat} and its standard deviation in dependence of the assignment method, aggregated across all simulation runs. Average effect size was not affected by the assignment method. However, the standard deviation of the effect size was larger for standard random assignment than for the covariate-based assignments. As shown by a Morgan-Pitman test for dependent variances (Wilcox, 2015), the differences in standard deviations (random assignment vs. anticlustering; random assignment vs. minimization; random assignment vs. rerandomization) were significant ($p < .001$). Note that even the minuscule difference in standard deviation between minimization versus anticlustering and rerandomization was significant ($p < .001$). Hence, balancing covariates reduced the variance of the standardized effect size estimate without introducing bias. This effect was strongest for anticlustering and rerandomization.

Table 3

Mean Treatment Effects and Their Standard Deviations for Each Assignment Method Aggregated Across All Simulation Conditions

Method	$M(d_{treat})$	$SD(d_{treat})$	n
Anticlustering	0.441	0.365	1080000
Minimization	0.441	0.367	1080000
Random Assignment	0.442	0.380	1080000
Rerandomization	0.441	0.366	1080000

Note. n refers to the total number of simulation runs.

Discussion

Our study set out to investigate anticlustering as a novel method to balance covariates among experimental conditions, which researchers have recently started doing (Even et al., 2023; Stadelmann et al., 2025; Tuti et al., 2022). Using a simulation study, we compared it with established methods for allocating experimental subjects to conditions: rerandomization and minimization, which are alternative methods to achieve balance in covariates (Morgan & Rubin, 2012; Treasure & MacRae, 1998), and standard random assignment. In doing so, we were able to locate the unfamiliar method of anticlustering in respect to these well-known and established methods. Moreover, we also provide some insight regarding the general question on whether actively balancing covariates or statistical adjustment of covariates is more important in the analysis of experiments.

In our simulation, we implemented the arguably most basic design used in experimental research: the independent two-group design, which can be analysed using the ubiquitous *t*-test or a regression (or ANCOVA) when a measured covariate is controlled

for. Our results can be summarized as follows: When a covariate is prognostic of the outcome variable, statistical adjustment is crucial for maximizing statistical power (see Figure 2). Balancing covariates by itself does not increase power, and using it on top of statistical adjustment only marginally improves power if the analysis includes covariates that are not prognostic of the outcome. These results for example confirm an investigation by Coart et al. (2023). Regarding statistical power alone, our results therefore attest to the importance of statistical adjustment and not necessarily to the usefulness of balancing covariates. Hence, our results confirm the points repeatedly made by Senn (1989, 2005, 2013, 2022).

However, there are advantages to actively balancing covariates that could be shown by our study. Our results confirm that balancing via minimization, rerandomization and especially via anticlustering leads to less observed imbalance in covariates (see Figures 1 and 3). If researchers worry that observed covariate imbalance will decrease the credibility of their findings (cf. Senn, 2005; Treasure & MacRae, 1998), balancing does provide a potent counter-measure. Moreover, we find that balancing decreases the risk of having "bad luck" when no statistical adjustment is employed: There is less variance in statistical power if covariates have been balanced (see Figure 3). Imbalance in covariates can facilitate as well as aggravate detecting existing effects of the treatment — depending on whether the direction of the imbalance is consistent with the direction of the treatment effect. However, researchers may prefer to exert control on the level of imbalance instead of hoping for good luck that the true effect of interest is not cancelled out by the effect of the covariate. Moreover, we find that the standardized effect size d of the treatment effect has less variability when covariates were balanced (see Table 3). Observed estimates of d will tend to be closer to the true value when balancing was applied. This finding confirms previous theoretical results according to which balancing covariates increases the precision of the treatment effect (Hu et al., 2014; Morgan & Rubin, 2012).

Previously, researchers have highlighted that actively balancing covariates may invalidate the usual interpretation of statistical tests: These tests assume that true random sampling is used to allocate experimental subjects to experimental conditions (e.g., Moulton, 2004; Senn, 2013). When covariate-based assignment is used and the covariate is correlated with the dependent variable, the classical statistical tests are more conservative than required (Morgan & Rubin, 2012; Qin et al., 2024). Our study confirms these previous theoretical results. For the unadjusted analysis, we find that the actual alpha error departs from the nominal alpha error when covariates have been balanced. Balancing decreased the probability of alpha errors, which is arguably less problematic than an increase would be. However, we do not recommend regarding the reduced alpha error rate as beneficial, because it was due to a more conservative test behaviour that came at the cost of statistical power. Instead, we recommend using statistical adjustment of the covariate in combination with covariate-based assignment, which maximizes power

and maintains the nominal alpha error level. An alternative to statistical adjustment via regression would be to implement a randomization test that incorporates the logic of the covariate-based assignment process into the significance test (Bugni et al., 2018; Ma et al., 2015; Morgan & Rubin, 2012; Shao et al., 2010). These tests can increase statistical power and maintain the nominal alpha level without employing statistical adjustment. In practice, however, it seems that randomization tests are rarely used (Lin et al., 2015) and they are not readily available in standard software (Ohashi, 1990). In particular, randomization tests must incorporate the specific procedure that was used for condition assignment, which cannot be captured by software defaults. In contrast, statistical control of a covariate is always possible and — as confirmed by our results — highly useful (e.g., Raab et al., 2000; also see Senn, 1989). Moreover, for anticlustering, a randomization test must still be developed, which is certainly an interesting venue for future research.

In general, all balancing techniques that we investigated (minimization, rerandomization and anticlustering) provided rather comparable results across the dependent variables. Anticlustering and rerandomization were more successful at minimizing observed covariate imbalance than minimization. This is not surprising, given that anticlustering and rerandomization use all information simultaneously to maximize balance. Minimization is a sequential method that assigns each subject one-by-one after another. It operates using less information. Rerandomization and anticlustering obtained highly similar results across all dependent variables. Again, this is not surprising, given that their objective functions are basically equivalent for uncorrelated covariates — at least when using the *k*-means objective for anticlustering, which was applied in the simulation study. When there were several covariates (in the noise condition) anticlustering achieved increased balance as compared to rerandomization. This is because anticlustering uses a more potent optimization algorithm than rerandomization, which relies on repeated resampling without systematically searching for improved balance.

With regard to statistical power and the probability of alpha errors, minimization, rerandomization and anticlustering provided very similar results (see Table 1 and Table 2). Therefore, we expect that the question whether to apply anticlustering, rerandomization or minimization will not depend on general superiority of either method; instead, it probably depends on the conditions of the data inquiry: If information on all subjects is available at the start, researchers have the opportunity to apply anticlustering or rerandomization. If subjects are recruited sequentially, however, anticlustering or rerandomization cannot be applied, but minimization can.

With regards to the relative merits of anticlustering versus rerandomization, the present study establishes practical equivalence between the two methods. Again, which method should be applied may depend on the circumstances. The anticlustering method is readily available for researchers via the open source and free software package `anticlust` (Papenberg & Klau, 2021) that has been maintained for several years and

includes much documentation and convenience functionality. Anticlustering also has the particular advantage that it can be applied for more than two experimental conditions, which is so far lacking for rerandomization. Moreover, anticlustering can easily be adapted for more advanced use cases such as balancing higher order moments instead of only covariate means (Papenberg, 2024), or including constraints on group assignments (Papenberg, Breuer, et al., 2025; Papenberg, Wang, et al., 2025). Rerandomization on the other hand has received more scrutiny on a theoretical level for the purpose of balancing covariates in experiments (Johansson et al., 2021; Morgan & Rubin, 2012); the origin of anticlustering has not been in the realm of statistical inference (Brusco et al., 2020; Späth, 1986). We therefore call for further investigation of rerandomization and anticlustering in different contexts to establish additional evidence on the relative usefulness of the two methods.

Based on our findings, we provide the following guidance on when anticlustering is likely to be most beneficial in practice. First, our findings suggest that balancing covariates in general is useful and should be done if possible. In nonsequential designs, when covariate measurements are available at the beginning of an experiments, anticlustering is a suitable method to achieve balance: It increases the precision of the effect size estimation; by increasing balance between conditions it potentially increases interpretability and trustworthiness of the results; when no statistical adjustment is feasible, it reduces variability of statistical power of an unadjusted statistical analyses. However, we generally recommend using covariate-adjustment during analysis whenever possible, while adhering to good practice in covariate selection (e.g., Raab et al., 2000; Wysocki et al., 2022). Balancing via anticlustering is not a substitute for statistical adjustment. Using both covariate-based assignment and statistical adjustment of covariates provides the best of two worlds. Our results are in line with official guidelines for medical research (e.g., European Medicines Agency, 2015) that explicitly recommend: (1) assignment methods that ensure covariate balance, and (2) covariate-adjusted analyses.

Limitations

Several limitations should be considered with regard to interpreting the results of our study. The most important limitation is concerned with generalizability: Only a rather simple design was considered in our simulation. Even though we probably implemented one of the most important designs in the social sciences, in a strict sense our conclusions only pertain to the conditions that we realized in this study, which is a general drawback of simulation studies for method comparisons. In particular, our data generating process assumed that the usual assumption of normality of residuals was true, which however is often not the case in real data. Future research should investigate how non-normality due to skewed, bounded, or multimodal data may affect the relative usefulness of covariate-based assignment and covariate-adjustment. It should also be noted that we only modelled linear covariate effects. For this reason, minimizing differences in covariate

means — which was done by all three competing covariate-based assignment methods — was sufficient to obtain increased precision of the effect size estimate. However, results may differ for nonlinear covariate effects. When covariate effects are assumed to be nonlinear, it may be preferable to choose a quantification of between-group balance that also considers differences with regard to higher order moments such as the variance (see Ma et al., 2024). For anticlustering, such objectives exist and have been implemented in software (Papenberg, 2024). Rerandomization instead uses the mahalanobis distance between covariate means and does not address differences in higher order moments.

While our study did not identify advantages of balancing with regard to statistical power, this pattern may not hold for all kinds of study designs. In particular, there may be statistical advantages of balancing in group-randomized designs, where the danger of strong imbalance among conditions is more severe, given that fewer subjects are assigned to treatment groups (Moulton, 2004).

Conclusion

We investigated anticlustering as an alternative method for allocating subjects to conditions in experiments in order to minimize discrepancy in important covariates between conditions. We find that it has similar effects as the well-known methods of minimization and rerandomization, which have been considered to be useful improvements over standard random assignment (Morgan & Rubin, 2012; Taves, 1974; Treasure & MacRae, 1998). We find it worthwhile to investigate the applicability of anticlustering across other settings of experimental assignment in future research.

Funding: The authors have no funding to report.

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: The authors have declared that no competing interests exist.

Data Availability: The data and R code required to reproduce all analyses presented in this paper are openly available from Papenberg and Angelike (2026).

Supplementary Materials

Type of supplementary material	Availability/Access
Data	
Simulation/results_revision.csv	Papenberg and Angelike (2026)
Code	
Simulation/Simulation.R	Papenberg and Angelike (2026)
paper.Rmd	Papenberg and Angelike (2026)
Simulation/Minimization_Sella_et_al/VarMin.R	Papenberg and Angelike (2026)
Material	
Simulation/Minimization_Sella_et_al/VM_template.csv	Papenberg and Angelike (2026)
Simulation/Minimization_Sella_et_al/TutorialVM_R.docx	Papenberg and Angelike (2026)
Study/Analysis preregistration	
Study was not preregistered	—
Other	
Simulation/Minimization_Sella_et_al/COPYRIGHT	Papenberg and Angelike (2026)
README.md	Papenberg and Angelike (2026)
lit.bib	Papenberg and Angelike (2026)
Simulation/Randomization_Zhang_et_al/test_rer.R	Papenberg and Angelike (2026)

References

Brusco, M. J., Cradit, J. D., & Steinley, D. (2020). Combining diversity and dispersion criteria for anticlustering: A bicriterion approach. *British Journal of Mathematical and Statistical Psychology, 73*(3), 375–396. https://doi.org/10.1111/bmsp.12186

Bugni, F. A., Canay, I. A., & Shaikh, A. M. (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association, 113*(524), 1784–1796.

Coart, E., Bamps, P., Quinaux, E., Sturbois, G., Saad, E. D., Burzykowski, T., & Buyse, M. (2023). Minimization in randomized clinical trials. *Statistics in Medicine, 42*(28), 5285–5311. https://doi.org/10.1002/sim.9916

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

European Medicines Agency. (2015). *Guideline on adjustment for baseline covariates in clinical trials.* https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-adjustment-baseline-covariates-clinical-trials_en.pdf

Even, G., Mouray, A., Vandenabeele, N., Martel, S., Merlin, S., Lebrun-Ruer, S., Chabé, M., & Audebert, C. (2023). Bact-to-batch: A microbiota-based tool to determine optimal animal allocation in experimental designs. *International Journal of Molecular Sciences, 24*(9), Article 7912. https://doi.org/10.3390/ijms24097912

Hu, F., Hu, Y., Ma, Z., & Rosenberger, W. F. (2014). Adaptive randomization for balancing over covariates. *Wiley Interdisciplinary Reviews: Computational Statistics, 6*(4), 288–303.

Johansson, P., Rubin, D. B., & Schultzberg, M. (2021). On optimal rerandomization designs. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 83*(2), 395–403. https://doi.org/10.1111/rssb.12417

Kapelner, A., Krieger, A. M., Sklar, M., Shalit, U., & Azriel, D. (2021). Harmonizing optimized designs with classic randomization in experiments. *American Statistician, 75*(2), 195–206. https://doi.org/10.1080/00031305.2020.1717619

Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M., & Horwitz, R. I. (1999). Stratified randomization for clinical trials. *Journal of Clinical Epidemiology, 52*(1), 19–26. https://doi.org/10.1016/S0895-4356(98)00138-3

Kovarianzanalyse (Statistik). (2025). In *Wikipedia*. https://de.wikipedia.org/w/index.php?title=Kovarianzanalyse_(Statistik)&oldid=251698759

Lin, Y., Zhu, M., & Su, Z. (2015). The pursuit of balance: An overview of covariate-adaptive randomization techniques in clinical trials. *Contemporary Clinical Trials, 45*, 21–25. https://doi.org/10.1016/j.cct.2015.07.011

Ma, W., Hu, F., & Zhang, L. (2015). Testing hypotheses of covariate-adaptive randomized clinical trials. *Journal of the American Statistical Association, 110*(510), 669–680.

Ma, W., Li, P., Zhang, L.-X., & Hu, F. (2024). A new and unified family of covariate adaptive randomization procedures and their properties. *Journal of the American Statistical Association, 119*(545), 151–162.

Mauri, L., Apolloni, B., & Damiani, E. (2023). Robust ML model ensembles via risk-driven anti-clustering of training data. *Information Sciences, 633*, 122–140. https://doi.org/10.1016/j.ins.2023.03.085

Morgan, K. L., & Rubin, D. B. (2012). Rerandomization to improve covariate balance in experiments. *Annals of Statistics, 40*, 1263–1282. https://doi.org/10.1214/12-AOS1008

Morgan, K. L., & Rubin, D. B. (2015). Rerandomization to balance tiers of covariates. *Journal of the American Statistical Association, 110*(512), 1412–1421.

Moulton, L. H. (2004). Covariate-based constrained randomization of group-randomized trials. *Clinical Trials, 1*(3), 297–305. https://doi.org/10.1191/1740774504cn024oa

Nagel, J., Morgan, D. P., Gürsoy, N. Ç., Sander, S., Kern, S., & Feld, G. B. (2024). Memory for rewards guides retrieval. *Communications Psychology, 2*(1), Article 31. https://doi.org/10.1038/s44271-024-00074-9

Nguyen, T.-L., Collins, G. S., Lamy, A., Devereaux, P. J., Daurès, J.-P., Landais, P., & Le Manach, Y. (2017). Simple randomization did not protect against bias in smaller trials. *Journal of Clinical Epidemiology, 84*, 105–113. https://doi.org/10.1016/j.jclinepi.2017.02.010

Ohashi, Y. (1990). Randomization in cancer clinical trials: Permutation test and development of a computer program. *Environmental Health Perspectives, 87*, 13–17. https://doi.org/10.1289/ehp.908713

Papenberg, M. (2024). K-plus anticlustering: An improved k-means criterion for maximizing between-group similarity. *British Journal of Mathematical and Statistical Psychology, 77*(1), 80–102. https://doi.org/10.1111/bmsp.12315

Papenberg, M., & Angelike, T. (2026). *Balancing covariates with anticlustering* [OSF project page containing study code and study data]. Open Science Framework. https://osf.io/zryf5/overview

Papenberg, M., Breuer, M., Diekhoff, M., Tran, N. K., & Klau, G. W. (2025). Extending the bicriterion approach for anticlustering: Exact and hybrid approaches. *Psychometrika, 90*(5), 1789–1808. https://doi.org/10.1017/psy.2025.10052

Papenberg, M., & Klau, G. W. (2021). Using anticlustering to partition data sets into equivalent parts. *Psychological Methods, 26*(2), 161–174. https://doi.org/10.1037/met0000301

Papenberg, M., Wang, C., Diop, M., Bukhari, S. H., Oskotsky, B., Davidson, B. R., Vo, K. C., Liu, B., Irwin, J. C., Combes, A., Gaudilliere, B., Li, J., Stevenson, D. K., Klau, G. W., Giudice, L. C., Sirota, M., & Oskotsky, T. T. (2025). Anticlustering for sample allocation to minimize batch effects. *Cell Reports Methods, 5*(8), Article 101137. https://doi.org/10.1016/j.crmeth.2025.101137

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

Qin, Y., Li, Y., Ma, W., Yang, H., & Hu, F. (2024). Adaptive randomization via mahalanobis distance. *Statistica Sinica, 34*, 353–375. https://doi.org/10.5705/ss.202020.0440

R Core Team. (2022). *R: A language and environment for statistical computing*. R Project for Statistical Computing. https://www.R-project.org/

Raab, G. M., Day, S., & Sales, J. (2000). How to select covariates to include in the analysis of a clinical trial. *Controlled Clinical Trials, 21*(4), 330–342. https://doi.org/10.1016/S0197-2456(00)00061-1

Rahu, I., Kull, M., & Kruve, A. (2024). Predicting the activity of unidentified chemicals in complementary bioassays from the HRMS data to pinpoint potential endocrine disruptors. *Journal of Chemical Information and Modeling, 64*(8), 3093–3104. https://doi.org/10.1021/acs.jcim.3c02050

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2023). Metacognitive differentiation of item memory and source memory in schema-based source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 49*(5), 743–765. https://doi.org/10.1037/xlm0001207

Sella, F., Raz, G., & Cohen Kadosh, R. (2021). When randomisation is not good enough: Matching groups in intervention studies. *Psychonomic Bulletin & Review, 28*, 2085–2093. https://doi.org/10.3758/s13423-021-01970-5

Senn, S. (1989). Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine, 8*(4), 467–475. https://doi.org/10.1002/sim.4780080410

Senn, S. (2005). An unreasonable prejudice against modelling? *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry, 4*(2), 87–89. https://doi.org/10.1002/pst.169

Senn, S. (2013). Seven myths of randomisation in clinical trials. *Statistics in Medicine, 32*(9), 1439–1450. https://doi.org/10.1002/sim.5713

Senn, S. (2022). Empirical studies of balance do not justify a requirement for 1,000 patients per trial. *Journal of Clinical Epidemiology, 148*, 184–188. https://doi.org/10.1016/j.jclinepi.2022.02.010

Shao, J., Yu, X., & Zhong, B. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika, 97*(2), 347–360.

Späth, H. (1986). Anticlustering: Maximizing the variance criterion. *Control and Cybernetics, 15*(2), 213–218.

Stadelmann, V. A., Gerossier, E., Kettenberger, U., & Pioletti, D. P. (2025). Combining systemic and local osteoporosis treatments: A longitudinal in vivo microCT study in ovariectomized rats. *Bone, 192*, Article 117373. https://doi.org/10.1016/j.bone.2024.117373

Taves, D. R. (1974). Minimization: A new method of assigning patients to treatment and control groups. *Clinical Pharmacology & Therapeutics, 15*(5), 443–453. https://doi.org/10.1002/cpt1974155443

Toorawa, R., Adena, M., Donovan, M., Jones, S., & Conlon, J. (2009). Use of simulation to compare the performance of minimization with stratified blocked randomization. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry, 8*(4), 264–278. https://doi.org/10.1002/pst.346

Treasure, T., & MacRae, K. D. (1998). Minimisation: The platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does. *BMJ, 317*, 362. https://doi.org/10.1136/bmj.317.7155.362

Tuti, T., Aluvaala, J., Malla, L., Irimu, G., Mbevi, G., Wainaina, J., Mumelo, L., Wairoto, K., Mochache, D., Hagel, C., Maina, M., & English, M.. (2022). Evaluation of an audit and feedback intervention to reduce gentamicin prescription errors in newborn treatment (ReGENT) in neonatal inpatient care in Kenya: A controlled interrupted time series study protocol. *Implementation Science, 17*, Article 32. https://doi.org/10.1186/s13012-022-01203-w

Weitz, R., & Lakshminarayanan, S. (1998). An empirical comparison of heuristic methods for creating maximally diverse groups. *Journal of the Operational Research Society, 49*(6), 635–646. https://doi.org/10.1057/palgrave.jors.2600510

Wilcox, R. (2015). Comparing the variances of two dependent variables. *Journal of Statistical Distributions and Applications, 2*, Article 7. https://doi.org/10.1186/s40488-015-0030-z

Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical control requires causal justification. *Advances in Methods and Practices in Psychological Science, 5*(2). https://doi.org/10.1177/25152459221095823

Zhang, H., Yin, G., & Rubin, D. B. (2024). PCA rerandomization. *Canadian Journal of Statistics, 52*(1), 5–25. https://doi.org/10.1002/cjs.11765

Methodology (METH) is the official journal of the European Association of Methodology (EAM).

PsychOpen GOLD is a publishing service provided by the Leibniz Institute for Psychology (ZPID), Germany.

Methodology
2026, Vol. 22(2), 127–150
https://doi.org/10.5964/meth.17973