

Importance of Block Randomization When Designing Proteomics Experiments

Bram Burger,* Marc Vaudel,* and Harald Barsnes



Cite This: *J. Proteome Res.* 2021, 20, 122–128



Read Online

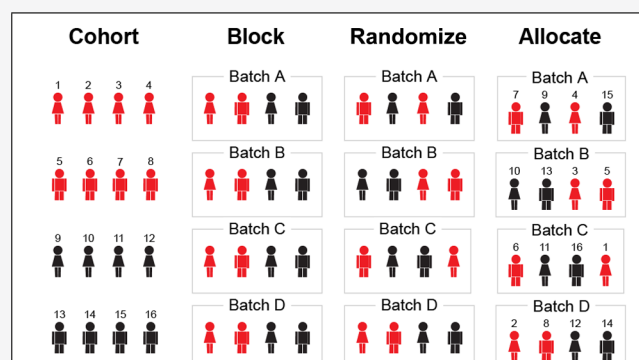
ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: Randomization is used in experimental design to reduce the prevalence of unanticipated confounders. Complete randomization can however create imbalanced designs, for example, grouping all samples of the same condition in the same batch. Block randomization is an approach that can prevent severe imbalances in sample allocation with respect to both known and unknown confounders. This feature provides the reader with an introduction to blocking and randomization, and insights into how to effectively organize samples during experimental design, with special considerations with respect to proteomics.

KEYWORDS: batches, blocking, experimental design, randomization, sample handling



INTRODUCTION

A vital part of experimental design consists of defining the order of sample processing and, if necessary, the creation of batches. The aim is then to avoid the introduction of confounders that would bias the interpretation of the data. One of the most famous examples of a confounded experiment is the observation of “water memory” by Davenas et al.,¹ i.e., the claimed ability of water to retain a memory of substances previously dissolved in it, which could not be replicated in a double-blinded experimental design,² suggesting that the initial data were the results of experimenter bias. Flaws in experimental design and unintended confounding can also be found in proteomics literature, as shown in, e.g., Sorace and Zhan,³ Hu et al.,⁴ Morris et al.,⁵ and Mertens.⁶ Knowing how to deal with the challenges of experimental design is therefore central to achieving reproducible experiments. For a general introduction to experimental design, see, for example, Box et al.,⁷ Ruxton and Colegrave,⁸ Lawson,⁹ and for proteomics specifically, see Burzykowski et al.¹⁰ and Maes et al.¹¹

While some common confounders, like sample annotation, date and order of processing, and their associated solutions are generic, others are more field-specific. In proteomics, the use of liquid-chromatography systems coupled to mass spectrometers (LC-MS) notably poses major challenges in terms of long-term performance and influence by outside factors.¹² Similarly, there can be differences introduced, notably during protein digestion,¹³ peptide fractionation and enrichment,¹⁴ and data interpretation.¹⁵ A good experimental design must therefore take into account both generic and field-specific confounders,

which can be extremely complex in larger experiments. This is especially challenging when experiments combine multiple analytical readouts, and when the allocation of patients to treatment arms and sample collection may be constrained.

Given the constraints of an experiment, the goal of defining sample order and batches is to (a) minimize the influence of anticipated confounders, (b) mitigate the risk that unanticipated confounders bias the interpretation of the results, and (c) ensure that the results remain interpretable if something does not go as planned and, for example, a batch is lost. While not inherently different from experimental design in other related fields, in our experience, block randomization is currently not a widespread concept in the proteomics community. Our aim with this feature is to introduce the proteomics audience to this topic through small and simple examples that the reader can easily extend to their own experiments.

In the following we will assume the conceptually simplest setting of label-free quantification without the use of reference samples. The concepts and considerations are however generally independent of the experimental setup. Special considerations for labeled experiments, experiments with

Received: July 13, 2020

Published: September 21, 2020



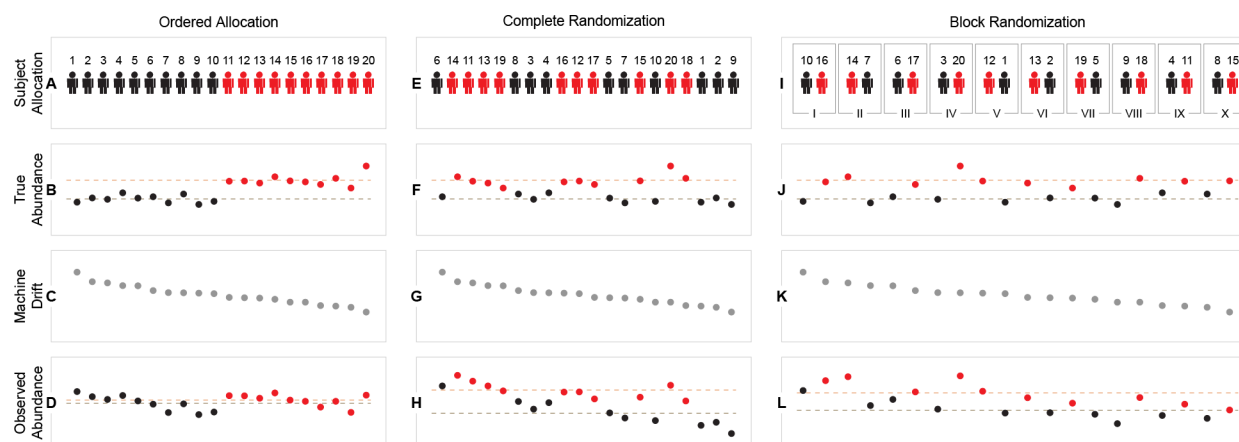


Figure 1. Randomization to account for machine drift. (A, E, I) Sample allocation order for 10 subjects receiving placebo (black figures, subjects 1–10) and treatment (red figures, subjects 11–20); ordered allocation (A), complete randomization (E), and block randomization (I). (B, F, J) True abundance for the subjects if there would be no machine drift; ordered allocation (B), complete randomization (F), and block randomization (J), (identical except for the sample order). (C, G, K) Simulated machine drift, identical for all three settings. (D, H, L) Observed abundances = true abundance + machine drift; ordered allocation (D), complete randomization (H), and block randomization (L). Dashed lines indicate the group means for each setting.

reference samples, and experiments with repeated measures are provided at the end.

SAMPLE RANDOMIZATION

Deciphering the association between sample characteristics, e.g., tumor types, and the proteome holds the key for improved diagnostic and treatment of diseases. In this setting, protein abundances are the responses, or outcome variables, of the experiment, and the other variables in the model are the explanatory variables.^{11,16,17} Often, one is interested in studying the association of one or more explanatory variable(s) with the response(s), while having to control for other explanatory variables that are not of primary interest. The explanatory variables of interest are also referred to as treatment variables, e.g., treatment, disease status, or tumor type.

The variables included in the model because they are expected to potentially influence the outcome, although not being of primary interest, are referred to as control variables, and include properties such as enzyme batch, column, or day of acquisition. Note that some variables may belong in either category, depending on the goals of the experiment, e.g., age, sex, and patient ancestry. If there is an association between treatment and control variables, this can impair the ability to estimate the effect of the treatment variable. In an extreme case, if *Controls* are handled first, and *Patients* last, to what extent are the observed differences between patients and controls genuine and not artifacts introduced during sample handling?

In addition to the monitored variables included in the model, other variables may also affect the results, such as machine drift or environmental changes during the analysis. Obviously one cannot control for all variables that may have an influence on the response variable. By their nature, unobserved variables cannot be included in the model as they are not observed, and including too many variables in the model reduces the power of the experiment. A high number of samples combined with randomization provides a safeguard, in the long run, against undue influence of unobserved variables.^{7,18,19} For more details on the importance of

randomization in proteomics, please see Morris et al.⁵ and Mertens.⁶

To illustrate the effects of randomized versus ordered allocation, let us assume that we have ten patients receiving a given *Treatment* and ten a *Placebo*, with the treatment resulting in a minor mean increase in the analytical readout—e.g., the abundance of a given protein, see Figure 1. Figure 1A shows the experimental setting for the ordered allocation with *Placebo* subjects processed first, while Figure 1E shows the same subjects in a complete randomized allocation. The “true” protein abundance for each patient is plotted in Figure 1B and F for the ordered and complete randomized allocations, respectively. Note that, apart from the order of the samples, these figures are exactly the same.

Now let us introduce a machine drift (Figure 1C,G) that causes the mass spectrometer to detect slightly less of the protein over time. For the ordered allocation, the observed protein abundances show almost no difference between the two group means (Figure 1D). Conversely, the difference in group means for the randomized allocation is nearly the same as the “true” difference (Figure 1H), and only has added variance caused by the machine drift (Figure 2). Note that if the groups were reversed in the ordered sample allocation scheme, the group mean difference would have been exaggerated instead.

BLOCK RANDOMIZATION

Complete randomization can produce severely imbalanced sample allocations, e.g., randomly assigning all subjects receiving treatment to one batch and all subjects receiving placebo to another batch. Then, batch and treatment are completely confounded, and it is impossible to perform an analysis with regards to the treatment. In such a situation it is not uncommon to resort to adjusting batches and sample ordering manually, or simply “randomize until it looks good”. Both of these procedures are poorly reproducible, and potentially introduce unintended biases. A structured way to solve this problem is to rather rely on block randomization.²⁰

As a simple example, let us revisit the experimental setting from Figure 1. To ensure that both treatments are equally represented throughout the run, we make ten blocks of two

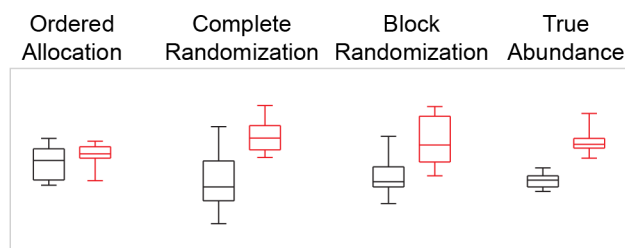


Figure 2. Boxplots for the true and observed protein abundances simulated in Figure 1. Black boxes (left) indicate *Placebo*, red boxes (right) indicate *Treatment*. The “true abundance” is what would have been measured without the machine drift (Figure 1, second row). The “ordered allocation”, “complete randomization”, and “block randomization” is as depicted in Figure 1D, H, and L, respectively.

subjects: one *Treatment* and the other *Placebo*. These are the smallest blocks we can make where each treatment is proportionally represented. The order of the treatments within the blocks (*Treatment* first or *Placebo* first) is chosen randomly for each block. Finally the subjects are randomly assigned to the blocks (see the right panels of Figure 1). Thus, we group small, representative subsets of the experiment together (the blocks), but within the blocks the order of the treatments is random, and within these constraints the assignment of subjects to blocks is also random. Consequently, while in complete randomization all sample sequences are possible, block randomization returns a sample sequence from a subset of all possible sequences where, by design, biases introduced by sequential processing are distributed as evenly as possible over the treatment groups.

In practice, it is not always possible or preferable to have the same number of subjects in all groups of the treatment variable(s).^{18,19} The block randomization procedure with groups of different sizes is only slightly different. As before, first, blocks of samples where each group is proportionally represented are created. For example, if one group is twice as large as the other group, each block would consist of three subjects, one of the smaller group and two of the larger group (Figure 3A). When the groups do not have a small common divisor, one can create blocks of different sizes. For example, in a nine vs ten setting, one would make eight blocks consisting of

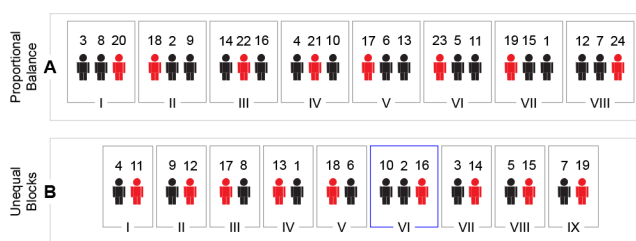


Figure 3. Examples of block randomization. (A) 16 subjects receiving placebo (black, subjects 1–16) and eight treatment (red, subjects 17–24), results in eight blocks of three subjects, each containing two *Placebo* and one *Treatment*. Subjects are randomly assigned to a block and the order within each block is randomized. (B) 10 subjects receiving placebo (black, subjects 1–10) and nine treatment (red, subjects 11–19), results in one block of three subjects, containing two *Placebo* and one *Treatment*, and eight blocks of two subjects, containing one *Placebo* and one *Treatment*. The block containing three subjects is randomly placed among the other blocks. Subjects are randomly assigned to a block and the order of the subjects within each block is randomized.

one subject of each group, and one block with the remaining three subjects (Figure 3B). In an experiment with multiple treatment levels, e.g., *Placebo*, *Treatment 1*, and *Treatment 2*, the blocks would consist of subjects from all treatments. As previously, the blocks are put in random order, the order of the treatments within the blocks is chosen randomly for each block, and subjects are finally randomly allocated according to their characteristics.

■ ACCOUNTING FOR CONTROL VARIABLES

The previous examples included two groups of subjects where the treatment was assumed to be the only difference, and where all samples could be processed at the same time. Most experiments have to account for control variables when estimating the treatment effects. Some common control variables are technical in nature, such as protease batches, freezer locations, and biobanks, but sample characteristics such as sex, age, and patient ancestry also commonly fall into this category. As the complexity of the design increases, it is common that not all samples can be processed at the same time in the same way at the same location. The sets of samples created by this process are referred to as batches, and this becomes yet another control variable to account for.

It is important to distribute, as best one can, the different levels of the treatment variables equally over the different levels of the control variables. In case of substantial confounding, as when most of the subjects receiving *Placebo* are *Female* and most of those receiving the *Treatment* are *Male*, it can become impossible to estimate the treatment effect. Similarly, when all *Treatment* subjects are in one batch and all *Placebo* in the other, batch and treatment are confounded and without a batch-independent reference it is not possible to distinguish the treatment effect from a possible batch effect. Thus, it is important to account for control variables in the experimental design and in the sample organization, and this can also be achieved using block randomization.

As a simple example, let us reuse the experimental setting from Figure 1, but this time split the experiment into two batches, e.g., due to different days of processing. Given that both treatment groups have ten subjects each, it is possible to divide both treatment groups into two subgroups of equal size. The first batch, *Batch A*, consists of five randomly chosen subjects from the *Treatment* group plus five randomly chosen subjects from the *Placebo* group, while the second batch, *Batch B*, consists of the remaining subjects (Figure 4A). The randomization scheme is then exactly as before, except that half of the randomized blocks are now assigned to *Batch A*, and the other half to *Batch B*.

Assume that for some unknown reason, something is different for the second batch, resulting in increased protein abundance measurements (Figure 4C). Each batch can now be seen as a separate experiment: the difference between treatment and placebo can be calculated in *Batch A*, and similarly in *Batch B* (Figure 4D). In both batches the treatment effect is close to the “true” difference (Figure 4B), even though the measured abundances are different.

When the experiment consists of multiple batches, they will undergo the same experimental protocol at different points in time and/or space. Every step of the protocol may then introduce variation that is specific for each batch. If samples are moved across batches between processing steps, each processing step has its own specific sample-to-batch allocation. Then, each processing step will have its own batch effect, and

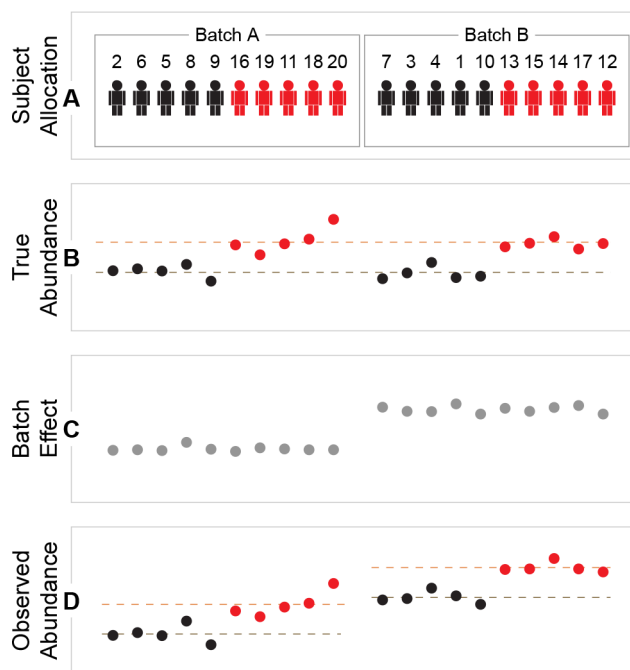


Figure 4. Randomized batch allocation. (A) Sample allocation order for subjects receiving placebo (black, subjects 1–10) and treatment (red, subjects 11–20). (B) True abundance if no batch effect. Dashed lines indicate group means for all placebo subjects and all treatment subjects, considered to be the true group means. (C) Simulated batch effect. (D) Observed abundances = true abundance + batch effect. Dashed lines indicate group means within the batches.

each of these will have to be estimated, unnecessarily increasing the complexity of the model. It is instead recommended to keep the same batches throughout the experiment, so that possible batch effects from different processing steps are combined into one overall batch effect.

If different processing steps of the protocol have different size constraints, i.e., one step requires more batches than another, being able to combine the smaller batches into larger batches without having to split the smaller batches is ideal. For example, when one experimental step can process 12 samples at once, while another step can process 24 samples at once, two batches from the first step can be combined for the second processing step. When this is not possible, it makes most sense to set up the batches according to the smallest constraints, and keep these batches throughout.

As a general advice, to reduce the complexity of experimental design, execution, and downstream analysis, it is recommended to keep the distribution of control variables as simple as possible. Note that this also applies to sample characteristics such as sex, age, and patient ancestry. For example, if one has to use two batches and is not interested in the association of the response with the sex of the patients, it is perfectly fine, and even recommended, to process *Male* and *Female* samples separately, hence confounding batch allocation with sex, thus reducing the complexity of the model.

It is important to underline that the control variables have to be part of the statistical model when analyzing the results, for example by fitting a linear model. If a control variable is not included in the model, one is in essence assuming that there is no difference between the batches, potentially leading to increased variance in the estimates for the variables of interest. On the other hand, every variable added to the model incurs a

cost. With two batches, estimating the batch effect costs one degree of freedom, with each additional batch costing yet another degree of freedom. Each degree of freedom used in estimating variables is essentially adding noise, and thus it is important to not make too many and/or small batches.⁷

MULTIPLE TREATMENT VARIABLES

When the effect of an explanatory variable, e.g., sex, on the response is of interest, it must be considered as one of the treatment variables. Similar considerations then hold as for the control variables, i.e., to be able to estimate the effect of the treatment variables, they should not be confounded with each other, nor with the control variables.

In a setting with multiple treatment variables, one can also be interested in estimating interactions between the treatment variables. These interactions then become another treatment variable, and it is important to make sure that these are also not confounded with any of the other variables. Hence, with sex and treatment as the treatment variables, one has to make sure that the effects of both separately and their interaction can be estimated. For this, all combinations of sex and treatment should be present in the experiment. Additionally the comparison between these combinations should not be confounded by a control variable. Here again, the latter can be achieved using block randomization.

For example, given 24 patients equally divided into two treatment levels, *Placebo* and *Treatment*, and two sexes, *Female* and *Male*, there are now four groups of six subjects representing each treatment–sex combination (1: *Placebo Female*, 2: *Placebo Male*, 3: *Treatment Female*, 4: *Treatment Male*). To make sure that the groups are balanced over the experimental design, we make six blocks of four subjects, one of each group, and within each block we randomly order the groups. Subsequently, we randomly allocate subjects to the blocks based on their group.

In case all samples cannot be processed together, one first creates the batches, and subsequently performs block randomization within each batch. For example, say we have to process the above samples in two batches. Given that we are interested in all comparisons between the treatment–sex combinations, we make two batches of 12 samples, each batch containing three subjects of each group (Figure 5). This way, the batches are balanced in terms of the subject-characteristics that we use in the analytical model.

In contrast to the case where sex is a control variable, putting all *Males* in one batch and all *Females* in the other should now clearly be avoided: sex and batch would be complete confounders, rendering impossible the comparison between *Males* and *Females*. Similarly, the interaction between sex and treatment would be lost.

With an increasing number of variables, the model becomes increasingly constrained. Given that one always has only a limited number of samples, variables thus need to be prioritized. The general advice from Box et al.⁷ “Block what you can and randomize what you can’t”, implies that there is only so much one can control for. Especially in proteomics studies, with generally large heterogeneity between subjects, one has to be careful not to block so heavily, and to include so many variables, that some combinations of treatment and control variables occur in only a very limited subset of subjects.

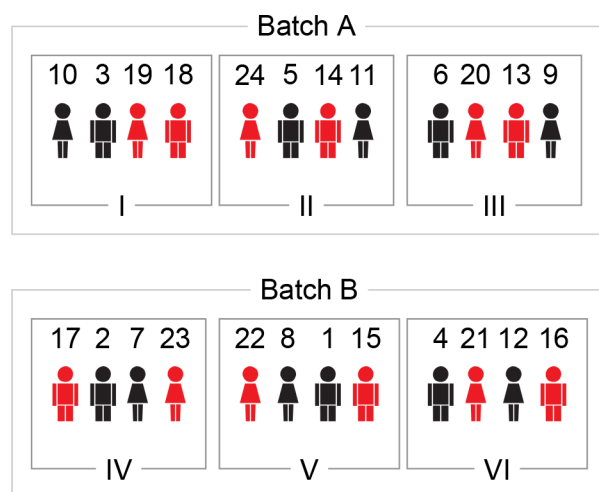


Figure 5. Example of block randomization with two variables and two batches. Treatment variables are treatment and sex, with treatment having two levels: *Placebo* (black) and *Treatment* (red). All combinations of treatment variables are represented by one subject in each block. Within these constraints subjects are randomly assigned to a block and the order of the subjects in each block is randomized.

FURTHER CONSIDERATIONS

The strategies outlined for the theoretical settings above can easily be extended to more elaborate situations. However, as for all methods, implementing block randomization can quickly become challenging in real-world situations. In this section, special considerations are introduced for situations where the reality of the experiment poses challenges in experimental design.

Continuous Variables and Fuzzy Categories

Most of the variables inspected so far have divided the samples into nonoverlapping categories, this will however not always be the case. Often, variables are continuous or categories are overlapping, such as age or disease state, respectively. In both cases categorization is commonplace, but can be problematic. Categories that span a large number of values can lead to relatively large differences between subjects within a category, while the differences between subjects at the edges of neighboring categories will be small. On the other hand, in a large enough study the randomization should mitigate this problem. Additionally, a substantial number of subjects per category is a requirement to still be able to randomize subjects. If each subject ends up being its own category, randomization will no longer be possible.

In cases where categorization is problematic, one can perform matching for the purposes of blocking, i.e., subjects that are similar according to a given variable, e.g., age or disease state, but belong to different groups with respect to all other variables, are treated as a single group for the purposes of block randomization (see for example Sekhon²¹). Note that the categorized version is solely created for block allocation, and the final analysis still uses the original variable.

Repeated Measures

In experiments with repeated measures, such as longitudinal experiments or dose–response curves, an extra dimension has to be taken into account when creating batches. For experiments where each subject has multiple samples, taken under different conditions (e.g., treatments or time points), the main goal is to detect differences within subjects, e.g., how a

disease progresses or the influence of different treatments. One should then treat all the samples from a subject as similarly as possible, and (where possible) process them in one batch. When the samples are exactly the same, e.g., with replicate samples at a certain dose/dilution, the opposite applies, as the main goal is then to detect differences between the doses. Differences between samples at the same dose are then considered as noise. In this case, one should therefore spread the samples from a particular dose over as many batches/replicates available.

Batch Size Being Smaller than the Number of Groups

When the number of variables increases, the batch size may become smaller than the number of groups. It then becomes important to distribute samples across batches in a way that makes it possible to calculate the differences of interest. For example, comparing four cell types, but having a batch size of three, it is impossible to put all cell types together in one batch. Then one would need to create a balanced set of batches, each including three of the four cell types, yielding four different batches (Figure 6). To create a balanced setting, one needs

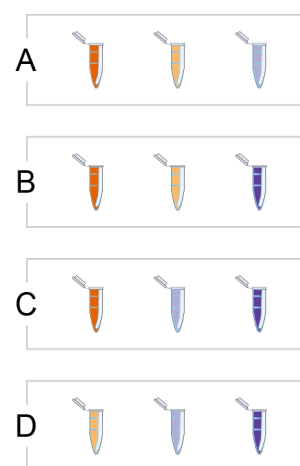


Figure 6. Four possible (ordered) batch compositions with four groups and a batch size of three. Each color indicates a cell type. Each cell type occurs in a batch alongside each other cell type exactly twice.

three samples of each cell type. Ideally, if every cell type has the same number of samples and the number of samples of each cell type is a multiple of three, one can create the same number of each of the batches (e.g., five replicates of batch A, five replicates of batch B, etc.). Hence all pairs of cell types will occur together in a batch the same number of times.⁷

As the number of groups to compare grows and/or the block size becomes smaller, one will need an increasing number of replicates to create balanced settings. This generally means that one has to prioritize some group-comparisons over others. The more often a pair of groups occurs together in a batch, the better one can estimate the given difference. On the other hand, when a pair of groups never occurs together in a batch, but share a batch with a common other group, the difference can be calculated through this common other group. For example, with three cell types (A, B, and C) and a batch size of two, and the main interest is to compare one cell type against the other two, it could be possible to mainly have batches with A vs B and A vs C. The comparison between B and C can still be made, although the power for that comparison will be much lower. For general strategies on how to construct batches in a

way that the differences of interest are estimable, see, e.g., Lawson⁹ or Box et al.⁷

Mitigation of Batch Effects Using Reference Samples

An approach that maintains comparability between samples is to introduce a common reference. For example, in targeted proteomic analyses, one spikes a known amount of the heavy version of a peptide/protein, for use as a reference to infer the concentration of the peptide/protein present in each sample, and hence compare the abundance of the specific peptide/protein across the samples.^{11,22}

A similar approach can also be used in untargeted settings, both with and without labels,^{12,23,24} where one sample is used as a standard throughout the experiment. Having a common reference makes samples more easily comparable across the different settings (e.g., batches, days of analysis, and instruments) by providing a common baseline. However, this only works for processing steps that the reference sample shares with the other samples in the relevant batch, and poses challenges in terms of missing value and dynamic range that are beyond the scope of this article.

The inclusion of a common reference sample is especially important when doing longitudinal/time-series experiments where samples are extracted from patients at multiple time-points. One could consider processing all of the samples at the same time, i.e., at the end of the study, but this unfortunately results in storage time being completely confounded with the time between the samples, thus making it impossible to distinguish the treatment effect from the storage effect (see, for example, refs 6, 25). Processing the samples as they become available can however introduce other confounders, e.g., machine performance or batches of chemicals. The solution is to process the samples at set time-points, and apply the same common reference sample across the entire experiment. The common reference sample then has to be present in every batch and its placement in the batch ought to be randomized.

The use of common reference samples can alleviate many challenges concerning batch effects. However, it is not always possible or desirable to include a common reference. For example, there may be constraints with regards to the available resources, e.g., the acquisition of heavy peptides for absolute quantification in targeted studies can be very expensive, and similarly, for large studies that run for an extended period of time, it is often not feasible to create a reference sample with a comparable composition relative to the experimental samples and that is large enough to last through the entire study.

CLOSING REMARKS


Proteomics has many aspects that ought to be taken into account when designing and planning experiments. The complexity of the samples, the proteome, and the analytical techniques employed make proteomics experiments particularly challenging. Especially in larger studies, the labor intensive sample preparation often means that the experiment has to be split into multiple batches. It is therefore important to design the experiment in such a way that variables and batches are not confounding. Each batch should as much as possible be its own small experiment. When the size of the batches is insufficient to achieve this, it is essential to make sure that the utilized experimental design can give answers to the scientific questions asked. To ensure this, and minimize the bias due to unobserved variables as much as possible, one should therefore block variables and randomize subjects, in the


different batches, and use this restricted random allocation to process samples in the lab and on the mass spectrometer.

It is important to underline that the challenges posed by the handling of multiple variables can only be answered if the scientific project is rigorously defined, with response, treatment, and control variables clearly identified before the samples are collected. Given the multidimensional and multidisciplinary nature of modern omics projects, it is essential that experts with the necessary expertise are involved early in the experimental design, to prevent confounding effects. Finally, while considerations of power are beyond the scope of this article, it cannot be stressed enough that an adequate number of samples is paramount, both for correct experimental design and to ensure that the research questions can be answered.

AUTHOR INFORMATION

Corresponding Authors

Bram Burger — Computational Biology Unit (CBU), Department of Informatics and Proteomics Unit (PROBE), Department of Biomedicine, University of Bergen, 5007 Bergen, Norway;  orcid.org/0000-0003-1093-1290; Phone: +47 55 58 00 00; Email: bram.burger@uib.no

Marc Vaudel — Department of Clinical Sciences, University of Bergen, 5007 Bergen, Norway;  orcid.org/0000-0003-1179-9578; Phone: +47 55 58 00 00; Email: marc.vaudel@uib.no

Author

Harald Barsnes — Proteomics Unit (PROBE), Department of Biomedicine and Computational Biology Unit (CBU), Department of Informatics, University of Bergen, 5007 Bergen, Norway

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jproteome.0c00536>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

BB and HB are supported by the Bergen Research Foundation. HB is also supported by the Research Council of Norway.

REFERENCES

- (1) Davenas, E.; Beauvais, F.; Amara, J.; Oberbaum, M.; Robinzon, B.; Miadonnai, A.; Tedeschi, A.; Pomeranz, B.; Fortner, P.; Belon, P.; Sainte-Laudy, J.; Poitevin, B.; Benveniste, J. Human basophil degranulation triggered by very dilute antiserum against IgE. *Nature* **1988**, *333*, 816–818.
- (2) Maddox, J.; Randi, J.; Stewart, W. W. “High-dilution” experiments a delusion. *Nature* **1988**, *334*, 287–290.
- (3) Sorace, J. M.; Zhan, M. A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC Bioinf.* **2003**, *4*, 24.
- (4) Hu, J.; Coombes, K. R.; Morris, J. S.; Baggerly, K. A. The importance of experimental design in proteomic mass spectrometry experiments: Some cautionary tales. *Briefings Funct. Genomics Proteomics* **2005**, *3*, 322–331.
- (5) Morris, J. S.; Baggerly, K. A.; Gutstein, H. B.; Coombes, K. R. Statistical Contributions to Proteomic Research. *Methods Mol. Biol.* **2010**, *641*, 143–166.
- (6) Mertens, B. J. A. Transformation, Normalization, and Batch Effect in the Analysis of Mass Spectrometry Data for Omics Studies. In *Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry*; Datta, S., Mertens, B. J. A., Eds.; Frontiers in

Probability and the Statistical Sciences; Springer: Cham, Switzerland, 2017; Chapter 1, pp 1–21.

(7) Box, G.; Hunter, J.; Hunter, W. *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed.; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, 2005.

(8) Ruxton, G.; Colegrave, N. *Experimental Design for the Life Sciences*, 2nd ed.; Oxford University Press: Oxford, U.K., 2006.

(9) Lawson, J. Design and analysis of experiments with R; *Texts in Statistical Science*; CRC: Boca Raton, FL, 2015.

(10) Burzykowski, T.; Claesen, J.; Valkenburg, D. In *Mass Spectrometry of Proteins*; Evans, C. A., Wright, P. C., Noirel, J., Eds.; Methods in Molecular Biology; Humana: New York, NY, 2019; Vol. 1977, Chapter 12, pp 181–197.

(11) Maes, E.; Kelchtermans, P.; Bittremieux, W.; Grave, K. D.; Degroove, S.; Hooyberghs, J.; Mertens, I.; Baggerman, G.; Ramon, J.; Laukens, K.; Martens, L.; Valkenburg, D. Designing biomedical proteomics experiments: state-of-the-art and future perspectives. *Expert Rev. Proteomics* **2016**, *13*, 495–511.

(12) Zhang, T.; Gaffrey, M. J.; Monroe, M. E.; Thomas, D. G.; Weitz, K. K.; Piehowski, P. D.; Petyuk, V. A.; Moore, R. J.; Thrall, B. D.; Qian, W.-J. Block Design with Common Reference Samples Enables Robust Large-Scale Label-Free Quantitative Proteome Profiling. *J. Proteome Res.* **2020**, *19*, 2863–2872.

(13) Burkhart, J. M.; Schumbrutzki, C.; Wortelkamp, S.; Sickmann, A.; Zahedi, R. P. Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J. Proteomics* **2012**, *75*, 1454–1462.

(14) Solari, F. A.; Dell'Aica, M.; Sickmann, A.; Zahedi, R. P. Why phosphoproteomics is still a challenge. *Mol. BioSyst.* **2015**, *11*, 1487–1493.

(15) Zhu, Y.; Orre, L. M.; Tran, Y. Z.; Mermelekas, G.; Johansson, H. J.; Malyutina, A.; Anders, S.; Lehtiö, J. DEqMS: A Method for Accurate Variance Estimation in Differential Protein Expression Analysis. *Mol. Cell. Proteomics* **2020**, *19*, 1047–1057.

(16) Riter, L. S.; Vitek, O.; Gooding, K. M.; Hodge, B. D.; Julian, R. K. Statistical design of experiments as a tool in mass spectrometry. *J. Mass Spectrom.* **2005**, *40*, 565–579.

(17) Oberg, A. L.; Vitek, O. Statistical Design of Quantitative Mass Spectrometry-Based Proteomic Experiments. *J. Proteome Res.* **2009**, *8*, 2144–2156.

(18) Senn, S. J. Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine* **1989**, *8*, 467–475.

(19) Senn, S. Seven myths of randomisation in clinical trials. *Statistics in Medicine* **2013**, *32*, 1439–1450.

(20) Rosenberger, W.; Lachin, J. *Randomization in Clinical Trials: Theory and Practice*, 2nd ed.; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, 2016.

(21) Sekhon, J. S. Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R. *J. Stat. Softw.* **2011**, DOI: 10.18637/jss.v042.i07.

(22) Kuzyk, M. A.; Smith, D.; Yang, J.; Cross, T. J.; Jackson, A. M.; Hardie, D. B.; Anderson, N. L.; Borchers, C. H. Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol. Cell. Proteomics* **2009**, *8*, 1860–77.

(23) Geiger, T.; Cox, J.; Ostasiewicz, P.; Wisniewski, J. R.; Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat. Methods* **2010**, *7*, 383–385.

(24) Brenes, A.; Hukelmann, J.; Bensaddek, D.; Lamond, A. I. Multibatch TMT Reveals False Positives, Batch Effects and Missing Values. *Mol. Cell. Proteomics* **2019**, *18*, 1967–1980.

(25) Gast, M.-C. W.; van Gils, C. H.; Wessels, L. F. A.; Harris, N.; Bonfrer, J. M. G.; Rutgers, E. J. Th.; Schellens, J. H. M.; Beijnen, J. H. Influence of sample storage duration on serum protein profiles assessed by surface-enhanced laser desorption/ionisation time-of-flight mass spectrometry (SELDI-TOF MS). *Clin. Chem. Lab. Med.* **2009**, *47*, 694–705.