



Matching

Prof. Dr. Ulrike Held
Klaus Steigmiller, MSc
Priska Heinz, MSc
Lea Bühner, MSc

June 13, 2025

Matching on treatment is a technique to reduce bias due to confounding. The aim of matching is to imitate a randomized study design and thus be able to infer causal treatment effects from the matched samples. Many observational studies use this technique nowadays. In this summary, we address difficulties and advantages of matching and we explain why matching is an algorithmic task that is part of the statistical analysis.

In clinical research applications, the following five steps need to be carried out:

1. Selection of the variables used for matching (i.e., confounders).
2. Choice of the matching algorithm.
3. Evaluation of the quality of the matching by assessing the balance between treatment groups.
4. Analysis of the outcomes after matching (i.e., estimation of the treatment effect).
5. Sensitivity analysis of the results (e.g., by varying the matching algorithm).

1. Matching should be based on variables describing the individual patients as a whole, and potentially influencing the treatment outcome. These variables are called confounders. Confounders may be demographic, clinical, as well as imaging or laboratory variables. Confounders must be measured at baseline; variables collected after baseline may not be used. Matching on age and gender alone is not sufficient, as other unbalanced confounders are likely to affect the results.
2. Finding a suitable control patient for each treated patient, or stratification into subgroups of treated and control patients, is an algorithmic task. Modern algorithms search for a global optimum in the sense that not single individuals are paired or grouped with their closest matching partner, but covariate balance between the treated and control patients as a whole is maximized [King et al., 2011]. The choice of an appropriate matching algorithm is a crucial step and a poor choice can compromise the results of the study. A common misconception is that pairs of treated and control patients need to be of the same age, gender, etc. This is not necessary. Analogous to a randomized study, the distribution of confounders needs only to be balanced across treatment groups. Specific algorithms may be used to perform propensity score matching, exact matching, optimal matching, genetic matching, or to address specific features of the dataset, such as a large number of categorical

confounders [Pimentel et al., 2015]. The algorithmic task is reproducible in the sense that, for a given data set and seed value, matching will lead to the same matched pairs or groups when repeated. It can be extended easily to include new patients, as well as additional confounders.

3. The quality of the matching is evaluated by assessing the balance of covariates between control and treated patients. Often this is done by using the standardized mean difference (SMD). It can be useful to try out several matching algorithms and then choose the one leading to the most balanced covariates. If all confounders are balanced after matching, in the sense that the absolute SMD is smaller than 0.1 [Austin, 2009], the analysis of the outcomes, i.e. estimation of the treatment effect can be performed.
4. The aim of an analysis after matching is the estimation of an unbiased treatment effect. It is advisable to use double-adjustment to reduce residual confounding by adjusting for the matching variables in the analysis of the outcomes, as discussed by Nguyen et al. [2017]. Typically, multiple logistic, Cox, or linear regression models are used, with treatment group and confounders as independent variables and accounting for substratification through matching. Details on estimands and marginal as well as conditional estimates can be found in Heinz et al. [2022].
5. A common critique of matched observational studies is that the results may still have been affected by unmeasured confounders, leading to a biased estimate of the treatment effect. Therefore, sensitivity analysis after matching should be used to assess the robustness of the results in case unmeasured confounders were present. For that, different methods, including Rosenbaum bounds for p-values [Rosenbaum, 2005], Hodges-Lehman point estimates [Rosenbaum, 1993], as well as the E-value for binary outcomes [van der Weele and Ding, 2017], have been proposed.

Algorithmic matching is a challenging task, and in-depth understanding of the algorithms and different aspects of matching is a prerequisite to obtain valid study results. The matching process is part of the outcome analysis, and it needs to be described in the statistical analysis plan. Pairing of observations from treatment and control group based on single variables like age and sex based on clinical knowledge cannot be considered valid for outcome analysis.

References

- P. C. Austin. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Communications in statistics-simulation and computation*, 38(6):1228–1234, 2009.
- P. Heinz, P. D. Wendel-Garcia, and U. Held. Impact of the matching algorithm on the treatment effect estimate: A neutral comparison study. *Biom. J.*, April 2022.
- G. King, R. Nielsen, C. Coberly, J. E. Pope, and A. Wells. Comparative effectiveness of matching methods for causal inference. *Unpublished manuscript, Institute for Quantitative Social Science, Harvard University, Cambridge, MA*, 2011.
- T.-L. Nguyen, G. S. Collins, J. Spence, J.-P. Daurès, P. J. Devereaux, P. Landais, and Y. Le Manach. Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. *BMC Med. Res. Methodol.*, 17(1), December 2017.

- S. D. Pimentel, R. R. Kelz, J. H. Silber, and P. R. Rosenbaum. Large, sparse optimal matching with refined covariate balance in an observational study of the health outcomes produced by new surgeons. *Journal of the American Statistical Association*, 110(510):515–527, 2015.
- P. R. Rosenbaum. Hodges-lehmann point estimates of treatment effect in observational studies. *Journal of the American Statistical Association*, 88(424):1250–1253, 1993.
- P. R. Rosenbaum. Sensitivity analysis in observational studies. *Encyclopedia of statistics in behavioral science*, 2005.
- T. J. van der Weele and P. Ding. Sensitivity analysis in observational research: introducing the E-value. *Annals of internal medicine*, 167(4):268–274, 2017.