University of Zurich[UZH]

Epidemiology, Biostatistics
and Prevention Institute,
Biostatistics Department
www.biostat.uzh.ch

Matching

Klaus Steigmiller, BSc
PD Ulrike Held

August 31, 2018

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 cannot be carried out by hand.

In clinical research applications, the following 4 steps need to be carried out:

1. Selection of the variables used for matching
2. Choice of the matching algorithm
3. Analysis of the outcomes after matching
4. Sensitivity analysis of the results

1. Matching should be based on variables describing the individual patients as a whole, and potentially influence 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 case patient is an algorithmic task. Modern algo- rithms search for a global optimum in the sense that not single individuals are paired with their closest matching partner, but covariate balance between the case 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 mis- conception is that pairs of case-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, 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, matching will lead to the same matched pairs when repeated and can be extended easily to include new case patients, as well as additional confounders. Reproducible and global optimization of covariate balance cannot be achieved with matching by hand.

3. The aim of an analysis after matching is the estimation of an unbiased treatment effect, or between-group difference. If all confounders are balanced after matching, in the sense that the standardized mean difference (SMD) is smaller than 0.1 [Austin, 2009], then the subsequent analysis can be simple,

e.g. a paired t-test or Wilcoxon-test for continuous outcomes, or McNemar test for binary outcomes. Pairing accounts for the matched set up of case and control patients. If balancing cannot be achieved through matching, the subsequent analysis should be model-based. Typically, multiple logistic or linear mixed effects regression models are used, with treatment group and unbalanced confounders as independent variables and random intercepts for pairs of case-control patients.

4. 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.

Please be aware that algorithmic matching is a challenging task. Hand-matching for 2-3 variables is sometimes used to avoid complex techniques or simple ignorance of the effect it may have on the results. In-depth understanding of the algorithms and different aspects of matching is a pre- requisite to obtain valid study results.

In order to increase the quality of medical research, please accept that we do not analyze datasets that have been matched by hand.

References

Peter 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: 12228–1234, 2009.

G. King, R. Nielsen, C. Coberly, J. E. Pope, and A. Wells. *Comparative effectiveness of matching methods for causal inference*, 2011. URL `http://j.mp/jCpWmk,visited2018-07-12`.

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*, 110:515–527, 2015.

P. R. Rosenbaum. Hodges-lehmann point estimates of treatment effect in observational studies. *Journal of the American Statistical Association*, 80:1250–1253, 1993.

P. R. Rosenbaum. *Sensitivity Analysis in Observational Studies*, 2005. American Cancer Society.

T. J. van der Weele and P. Ding. Sensitivity analysis in observational research: Introducing the e-value. *Annals of Internal Medicine*, 167:268–274, 2017.