Epidemiological cohort studies are typical implemented in investigating associations between disease outcomes and certain risk factors. The major effort and cost of conducting a cohort study mostly arise from the assembling of covariate measurements. The information on assessing the association mostly comes from cases, subjects experiencing the disease. But when the disease of interest is rare, most subjects in the study cohort do not experience the disease event by the end of the study period. In this case, conducting a full-scale cohort study might be too costly and might not be feasible for this purpose. To reduce the cost in such studies and achieve the same goals as a cohort study, several cohort-sampling designs have been proposed.

The case-cohort study design proposed by Prentice (1986) is the most widely used one, especially useful when the disease rate is low. The main idea is to sample a subset disproportionately within the study cohort focusing on cases: a random subset of the cohort (subcohort) and remaining cases in the cohort. Note that this case-cohort sample is not a random sample and a valid estimation procedure needs to take this account.

We will begin with this design feature, and discuss its variations and estimation procedures.

- Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, *73*(1), 1-11.

References

1. For theoretical justifications of (modified) Prentice(1986)'s estimator

- Self, S. G., & Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *The Annals of Statistics*, 64-81.

2. Variations of the design

- Kulich, M., & Lin, D. Y. (2004). Improving the efficiency of relative-risk estimation in case-cohort studies. *Journal of the American Statistical Association*, *99*(467), 832-844.
- Cai, J., & Zeng, D. (2007). Power calculation for case–cohort studies with nonrare events. *Biometrics*, *63*(4), 1288-1295.

3. Extension to multivariate failure time data

- Kim, S., Cai, J., & Lu, W. (2013). More efficient estimators for case-cohort studies. *Biometrika*, *100*(3), 695-708.

4. More efficient estimators

-Barlow, W. E., Ichikawa, L., Rosner, D., & Izumi, S. (1999). Analysis of case-cohort designs. *Journal of clinical epidemiology*, *52*(12), 1165-1172.
- Breslow, N. E., Lumley, T., Ballantyne, C. M., Chambless, L. E., & Kulich, M. (2009). Improved Horvitz–Thompson estimation of model parameters from two-phase stratified samples: applications in epidemiology. *Statistics in biosciences*, *1*(1), 32-49.