

Quick Review of Case-cohort Design

Boyi Guo

Department of Biostatistics
University of Alabama at Birmingham

August 5th, 2021

TL;DR

TL;DR

- ▶ Case-cohort, a special case of case-control, is designed for flexibility when studying multiple endpoints or staggered entry studeis
 - ▶ Only use random sub-cohort part of data for accurate odds ratio estimation (Prentice 1986)
- ▶ Case-cohort study consists three parts:
 - ▶ Random Sub-cohort: *sub-cohort cases* and *sub-cohort non-cases*
 - ▶ *Non sub-cohort cases*
- ▶ Use sampling weight when doing analysis
 - ▶ *Denominator Weights* need to be calculated for cases
- ▶ When the case definition changes, your analytic dataset changes
 - ▶ REGARDS used stratification when random sampling the sub-cohort part. Hence, there are multiple case-cohort studies where the *survey weights* are different
 - ▶ Do not change your sub-cohort sample

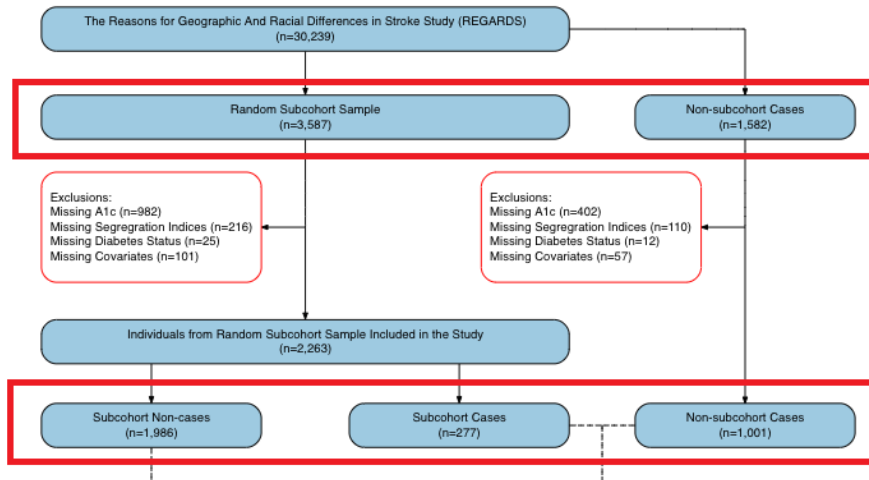
Motivation

Motivation

- ▶ The purpose for customized study designs is mostly for accurate estimation of a statistics (mean or measure of association)
 - ▶ Golden standard: cohort study
- ▶ The cohort design is not efficient for studying outcomes that are rare or has long follow-up
 - ▶ Case-control design
- ▶ Many types of case-control designs, depending on how controls are sampled
 - ▶ Case-cohort is a special case of case-control
- ▶ A big part of data collected in the case-cohort design, the random sub-cohort, can be repeatedly used when studying other definitions of cases.

Study Design

Study Design



Logistics

- ▶ At baseline
 - ▶ Collect blood sample for all participants and store in a freezer
 - ▶ Create random sub-cohort¹
- ▶ At the time of case-cohort study
 - ▶ Define cases and assemble the analytic dataset
 - ▶ Analyze the blood sample for the analytic dataset
 - ▶ Run analysis

¹This step can be also done retrospectively

Analysis

Survival Analysis

- Define *Denominator Weights* following Barlow et al. (1999)

TABLE 2. Denominator weights in the pseudolikelihood for cases ad controls by method

Outcome type and timing	Prentice [2]	Self and Prentice [7]	Barlow [8]
Case outside subcohort before failure	0	0	0
Case outside subcohort at failure	1	0	1
Case in subcohort before failure	1	1	$1/\alpha$
Case in subcohort at failure	1	1	1
Subcohort control	1	1	$1/\alpha$

- Robust variance estimations are recommended
- Software implementation are documented Barlow et al. (1999) with a dated Macro provided

Cautious

Cautious

- ▶ Different from nested case-control design
 - ▶ Risk set for controls sampling are different
- ▶ Define your cohort clearly
 - ▶ Sampling the random subcohort before/after exclusion of sample makes different to the estimation
 - ▶ Most often, REGARDS have pre-defined random sub-cohort. Don't do it on your own.

Methodology Development

Methodology Development

- ▶ Linear regression in a case-cohort study: interpretation and statistical inference
- ▶ Bootstrapping strategies in a case-cohort study and its implications on mediation analysis

Reference

Reference

- Barlow, William E., Laura Ichikawa, Dan Rosner, and Shizue Izumi. 1999. "Analysis of case-cohort designs." *Journal of Clinical Epidemiology* 52 (12): 1165–72. [https://doi.org/10.1016/S0895-4356\(99\)00102-X](https://doi.org/10.1016/S0895-4356(99)00102-X).
- Prentice, R. L. 1986. "A Case-Cohort Design for Epidemiologic Cohort Studies and Disease Prevention Trials." *Biometrika* 73 (1): 1. <https://doi.org/10.2307/2336266>.