SISCER Module 12
Lecture 4: Matched Case-Control Studies

Ting Ye & Qingyuan Zhao

University of Washington & University of Cambridge

July 2023

Plan

Motivating examples

Matched case-control studies

Key references for this lecture

- Rosenbaum (2002, Chapter 3.3)

Motivating example 1: DES and vaginal cancer

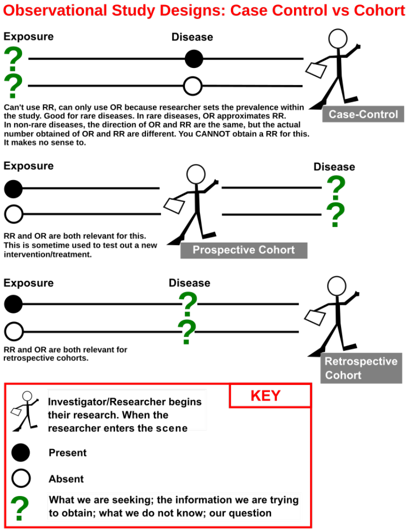
- Herbst et al. (1971) were interested in the possibility that a drug, diethylstilbestrol or DES, given to pregnant women, might be a cause of vaginal cancer in their daughters.
- Each of the eight cases was matched to four controls (referents), that is, to four women without vaginal cancer who were born within five days of the birth of the case at the same hospital, and on the same type of service.
- There were 8 cases of vaginal cancer and 32 referents, and the study compared the use of DES by their mothers.

	DES	No DES
case (vaginal cancer)	7	1
referent (no vaginal cancer)	0	32

- If a conventional test designed for use in a randomized experiment (e.g., Fisher exact test) is used, it would be highly significant.
- How far would the observational study have to depart from a randomized experiment to produce such a relationship if DES were harmless?

Motivating example 2: Firearms in the home and suicide risks

- Wiebe (2003) tested the hypothesis that having a gun in the home is a risk factor for adults to commit suicide
- 1,959 suicide case subjects were drawn from the 1993 National Mortality Followback Survey (NMFS). 13,535 living referent subjects were drawn from the 1994 National Health Interview Survey (NHIS), matched to the case subject by sex, race, and age.
- Subjects were considered exposed if they lived in a home where one or more firearms were reported present.
- Logistic regression was used to calculated conditional odds ratio, adjusting for other potential confounders such as marital status and education.
- Compared with persons living in a home with no firearms, the conditional odds ratio for suicide was 3.44 (95% CI 3.06 to 3.86) for persons living in a home with firearms.



In a case-referent study, cases are deliberately over-represented and referents are under-represented.

Advantages of case-referent studies

- Enable studying rare outcomes
- Easy to look at multiple risk factors

Common pitfalls of case-referent studies

- Confounding bias (as before)
- Recall bias
- Selection bias

(More details at the end...)

https://en.wikipedia.org/wiki/Case-control_study

Interpreting case-referent studies based on Silber et al. (2001)

- Population: all Medicare patients in Pennsylvania
- Cases: patients died after surgery
- Referents: patients did not die after surgery
- Exposure: certain risk factor
- Assumption: NUCA
- Key insight: The odds ratio in Tables 2-3 are the same, both equal to

$$\frac{\mu_1(x)/(1-\mu_1(x))}{\mu_0(x)/(1-\mu_0(x))}$$

So the odds ratio in case-control studies is a measure of the treatment effect.

Table: Data *before* selecting cases and referents

$X_i = x$	Exposed	Unexposed
Case	$\sum A_i Y_i(1)$	$\sum (1-A_i) Y_i(0)$
Referent	$\sum A_i(1-Y_i(1))$	$\sum (1-A_i)(1-Y_i(0))$

Table: Expected counts *before* selecting cases and referents; $\pi(x) = P(A=1 \mid X=x)$, $\mu_a(x) = E(Y(a) \mid X=x)$

$X_i = x$	Exposed	Unexposed
Case	$n_x \pi(x) \mu_1(x)$	$n_x(1-\pi(x))\mu_0(x)$
Referent	$n_x \pi(x)(1-\mu_1(x))$	$n_x(1-\pi(x))(1-\mu_0(x))$

Table: Expected counts *after* selecting cases and referents

$X_i = x$	Exposed	Unexposed
Case	$k_{1x} n_x \pi(x) \mu_1(x)$	$k_{1x} n_x(1-\pi(x))\mu_0(x)$
Referent	$k_{0x} n_x \pi(x)(1-\mu_1(x))$	$k_{0x} n_x(1-\pi(x))(1-\mu_0(x))$

Inference about treatment effect

- Mantel–Haenszel test
- Logistic regression adjusting for treatment, stratum indicators (and potentially other covariates)

Rosenbaum sensitivity analysis for matched case-referent studies (Rosenbaum, 1991)

- Suppose the i-th case is matched to $J-1$ referents, indexed by $i1, i2, \dots, iJ$.
- Each matched set may have different number of exposed subjects, denoted by m_i.
- The Mantel-Haenszel test statistic is the sum of Fisher's exact test statistics in multiple 2×2 tables
- Consider sensitivity analysis for Fisher's exact test in one 2×2 table

		Outcome Y 0	1	Total
Treatment A	0	N_{i00}	N_{i01}	$J-m_i$
	1	N_{i10}	N_{i11}	m_i
	Total	$J-1$	1	J

Consider the null hypothesis of no treatment effect $H_0: Y_{ij}(0) = Y_{ij}(1)$, for all i, j. The test statistic is $T_i = N_{i11}$ (the number of treated units with an outcome of 1).

Rosenbaum sensitivity analysis for matched case-referent studies (Rosenbaum, 1991)

- Recall the sensitivity model in Lecture 3: for $\Gamma \geq 1$ and $\pi_{ij} = \mathbb{P}(A_{ij} = 1 | X_{ij}, U_{ij})$,

$$1/\Gamma \leq \mathrm{OR}(\pi_{i\ell}, \pi_{ik}) = \frac{\pi_{i\ell}(1-\pi_{ik})}{\pi_{ik}(1-\pi_{i\ell})} \leq \Gamma, \text{ with } X_{i\ell} = X_{ik}$$

  Or equivalently the following logistic model

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = g(X_{ij}) + \gamma U_{ij},\ 0 \leq \gamma \leq \log \Gamma, 0 \leq U_{ij} \leq 1.$$

- Under NUCA,

$$P(A_{ij} = 1 \mid \boldsymbol{X}_{i\cdot} \boldsymbol{U}_{i\cdot}, \sum_{j=1}^{J} A_{ij} = m_i) = \frac{m_i}{J}$$

  where $\boldsymbol{X}_{i\cdot} = (X_{i1}, \ldots, X_{iJ})$ and $\boldsymbol{U}_{i\cdot} = (U_{i1}, \ldots, U_{iJ})$.

Rosenbaum sensitivity analysis for matched case-referent studies (Rosenbaum, 1991)

- A consequence of Rosenbaum's sensitivity model is that

$$\frac{m_i}{m_i + (J - m_i)\Gamma} \leq P(A_{i1} = 1 \mid \sum_{j=1}^{J} A_{ij} = m_i, Y_{i1} = 1, \sum_{j=1}^{J} Y_{ij} = 1, \boldsymbol{X}_{i\cdot}, \boldsymbol{U}_{i\cdot}) \leq \frac{m_i\Gamma}{m_i\Gamma + J - m_i}$$

- Again, a fair coin toss is replace by a biased coin toss
- The Mantel Haenszel is $T = \sum_i T_i = \sum_i N_{i11}$, where the Fisher's exact test statistics T_i's are mutually independent
- So we reject H_0 if T is large (compared to a independent sum of an extended hypergeometric distribution)

Γ	1.0	2	5	**7**	10
Worst-case p-value	1.28e-05	0.0005	0.016	**0.042**	0.050

- Small's R package SensitivityCaseControl has a function sens.analysis.mh implements the sensitivity analysis for Mantel-Haenszel test.

Common pitfalls of case-referent studies

- Confounding bias (as before)
- Recall bias: the accuracy of exposure information may be different for cases and referents, especially if the exposure information comes from interviews
  - Cases may improving memory, thus enhancing sensitivity among cases
  - Cases may provoke more false memory of exposure, thus reducing specificity among cases
  - Cases (e.g., disease) may itself cloud memory and thus reduce sensitivity
- Selection bias: nonrandom selection of cases and referents may distort the odds ratio
  - E.g., if cases of lung cancer at a hospital were compared to referents who were patients with cardiac disease in the same hospital, because smoking causes both lung cancer and cardiac disease, the calculated odds ratio would be too small

Herbst, A. L., Ulfelder, H., and Poskanzer, D. C. (1971). Adenocarcinoma of the vagina: Association of maternal stilbestrol therapy with tumor appearance in young women. *New England Journal of Medicine*, 284(15):878–881.

Rosenbaum, P. R. (1991). Sensitivity analysis for matched case-control studies. *Biometrics*, pages 87–100.

Rosenbaum, P. R. (2002). *Observational studies*. Springer.

Silber, J. H., Rosenbaum, P. R., Trudeau, M. E., Even-Shoshan, O., Chen, W., Zhang, X., and Mosher, R. E. (2001). Multivariate matching and bias reduction in the surgical outcomes study. *Medical care*, pages 1048–1064.

Wiebe, D. J. (2003). Homicide and suicide risks associated with firearms in the home: a national case-control study. *Annals of emergency medicine*, 41(6):771–782.