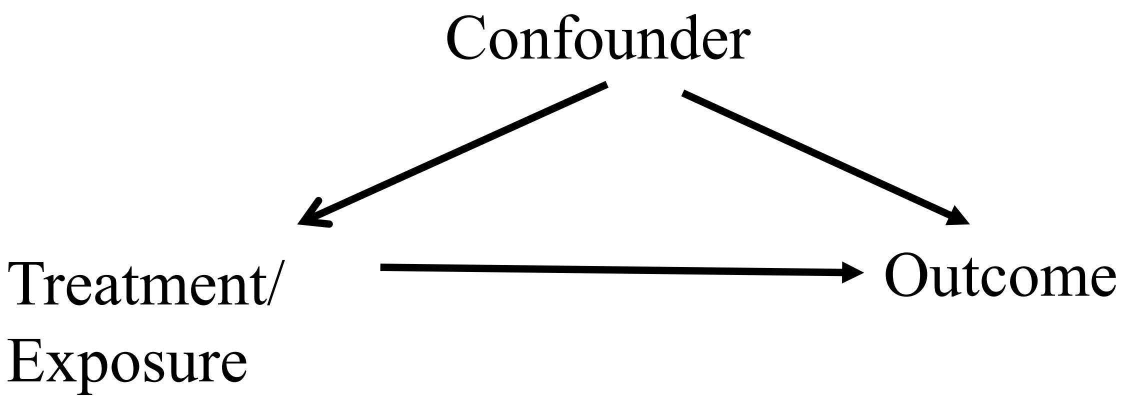
2.2 Role of matching in cohort and case-control studies

Confounding

confounding is due to “imbalance”, so idea is to

“balance” the design

achieved by randomization in experimental studies

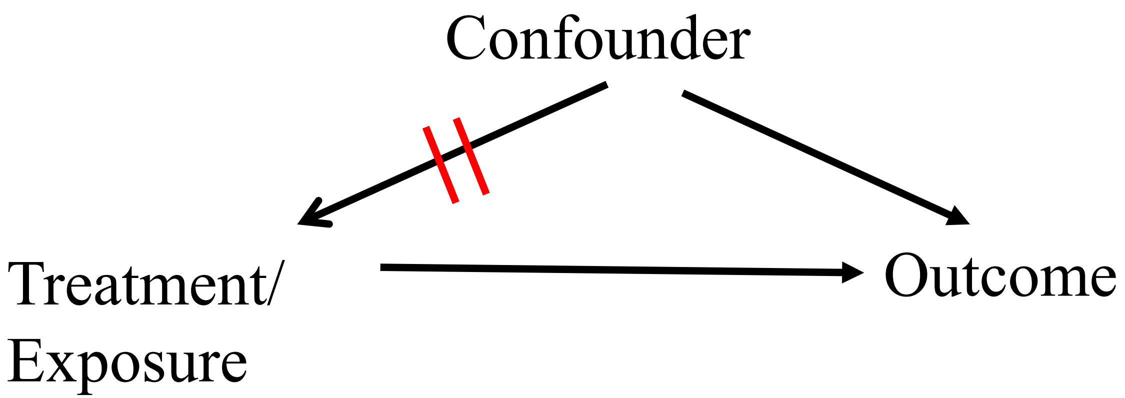


Confounding

confounding is due to “imbalance”, so idea is to

“balance” the design

achieved by randomization in experimental studies



In observational studies: matching

Matching

Used in both cohort and case-control studies, traditionally more common in case-control studies

Purpose:

- Overcome/reduce confounding due to imbalance
- Enable adjustment for confounders that are difficult/impossible to measure (e.g. neighborhood, family environment,….)

How to:

Groups to be compared (exposed/unexposed, or case/control) chosen to be similar on one or more potential confounders

Matching may be done
on a group basis ***(frequency matching)***
or an individual basis (***pair-matching***).

Example of balance by matching

Extreme positive confounding (Breslow & Day case-control studies p102)

102 BRESLOW & DAY

	Factor C+ Exposure E +	Factor C+ Exposure E −	Factor C− Exposure E +	Factor C− Exposure E −	Pooled levels of C Exposure E +	Pooled levels of C Exposure E −
Case	90	10	1	9	91	19
Control	9	1	10	90	19	91
Odds ratio	1		1		22.9	

→ Frequency matching does not eliminate confounding , but reduces it

	Factor C+ Exposure E +	Factor C+ Exposure E −	Factor C− Exposure E +	Factor C− Exposure E −	Pooled levels of C Exposure E +	Pooled levels of C Exposure E −
Case	50	50	90	10	140	60
Control	10	90	50	50	60	140
Odds ratio	9		9		5.44	

improved precision, and:

pooled OR in same direction as common OR but diluted (closer to 1)

Matching in cohort vs. case-control studies

Common purpose: balance within confounder strata to avoid sparse data (that could result from random sampling)

Different consequences for analysis

Cohort study: balance exposed and unexposed

(matching only affects independent variables)

Case-control study: balance cases and controls in

(matching depends on outcome)

Examples of matched cohorts

Karolinska Institutet

Early examples: Injury/accident research

Common examples:

- matching on neighbourhood (control socio-economic/ life-style factors)
- matching on family (control genetic factors and family environment)

Recent examples: (many matched patient cohorts)

- surgical intervention in diabetes vs non-diabetes patients
- Medication effects in new vs. non-new users
- Outcomes (e.g. MI & stroke) following Covid-19

Often additional matching on age, sex,..

Choice of matching variable(s) for matched cohort

Important to avoid overmatching

(i.e. making exposed and unexposed more similar for the outcome)

Inappropriate to match on a variable that is:

- A mediator (intermediary between exposure and outcome)

  consequence: bias

- strongly associated with outcome but little/ no association with exposure (i.e. is not a confounder)

  consequence: loss in efficiency

General Principle:

match on just a few well-established confounders

Matching in case-control studies

Purpose: to balance the number of cases and controls within confounder strata

Overmatching
(making **cases** and **controls** more similar for the **exposure**)

Inappropriate to match on a variable that is:
A mediator (intermediary between exposure and outcome)

strongly associated with **exposure** but little/ no association with **outcome** (i.e. is not a confounder)

Karolinska Institutet

Example of matched case-control study

Primary breast cancer patients in Sweden (1958-2001)
N=164,228

NCC sampling

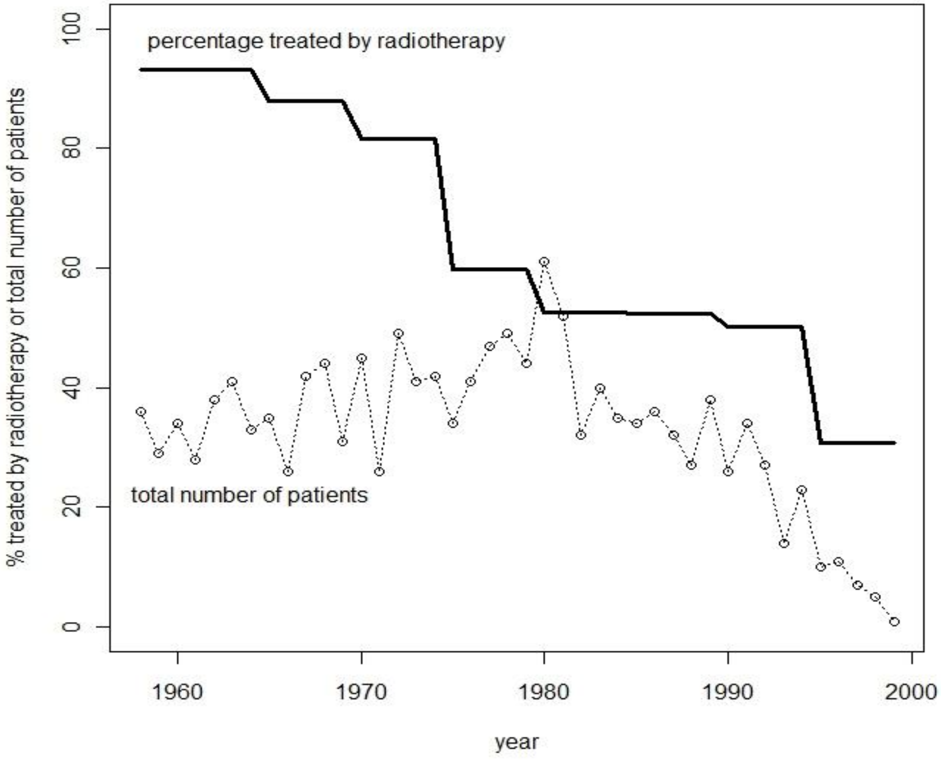
Cases of lung cancer
n=730

Controls matched on:
decade of BC diagnosis, n=726
age ±5 years, region

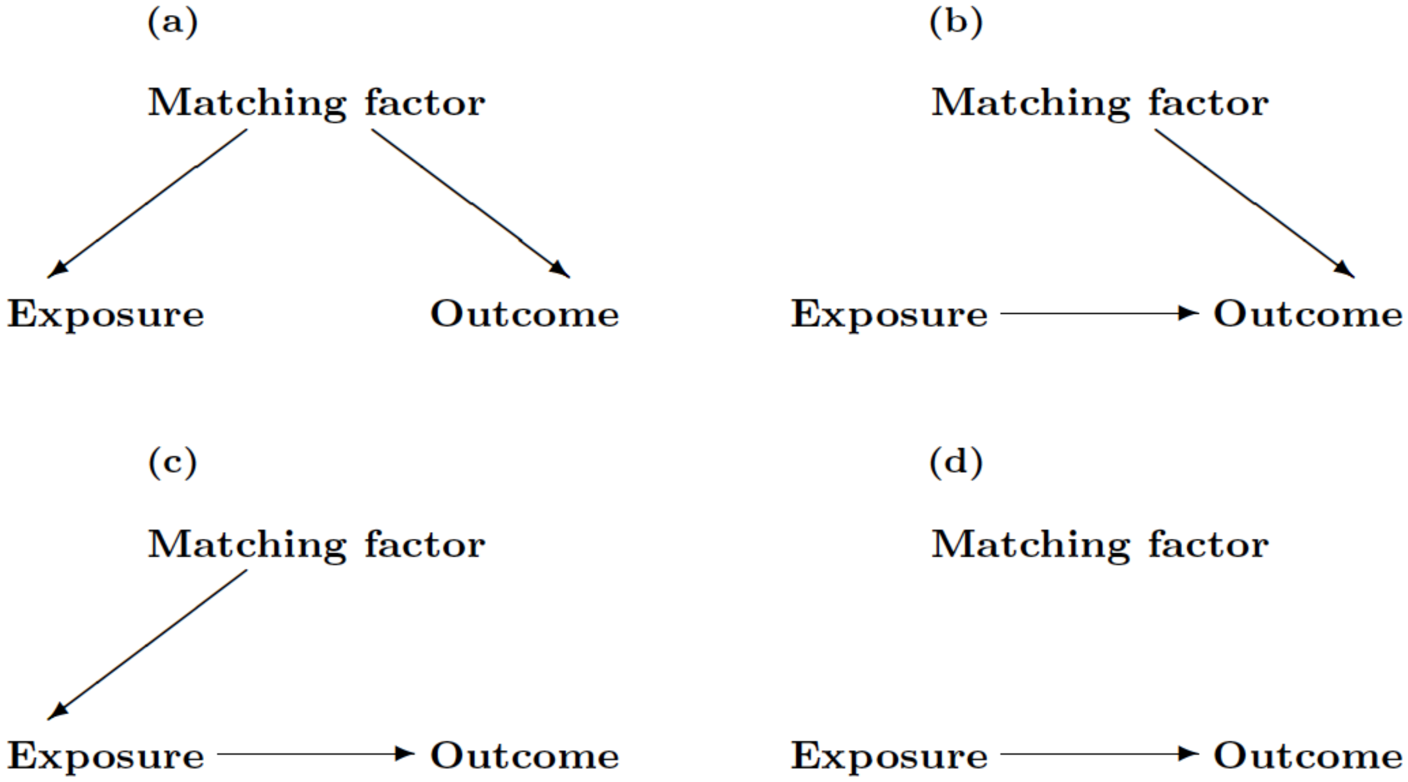
Conditional logistic regression – no significant effect of radiation

Investigators felt matching was the problem!

Radiotherapy use by calendar year: overmatched?



Choice of matching variables (Quiz)



Fine matching

Sometimes matched strata have very few observations, e.g. matched pairs often used: twins, paired organs,…

Example: matched case-control pairs:

	Control Exposed	
	Yes	No
Case Exposed	n_{11}	n_{10}
Case Unexposed	n_{01}	n_{00}

Test of association from paired data

Matched case-control pairs:

	Control Exposed Yes	Control Exposed No
Case Exposed	n_{11}	n_{10}
Case Unexposed	n_{01}	n_{00}

Test of association uses only discordant pairs:

(McNemar's Chi-Square) $$\chi^2 = \frac{(n_{10} - n_{01})^2}{(n_{10} + n_{01})}$$

i.e. where pair is discordant, does it tend to be the case that is exposed

Paired OR

Assuming ***a common OR*** in different strata/pairs, we can compute the **Mantel-Haenszel** estimator: can be done very simply as all tables have just 1's and 0's in the 4 cells (only 4 possibilities)!

	Concordant pairs					**Discordant pairs**			
	Y=1	Y=0		Y=1	Y=0	Y=1	Y=0	Y=1	Y=0
Exposed	1	1		0	0	1	0	0	1
Unexposed	0	0		1	1	0	1	1	0
No. pairs	n_{11}			n_{00}		n_{10}		n_{01}	

Numerator for MH: 0*n_{11} + 0 *n_{00} + $\frac{1}{2}$ *n_{10} +0 *n_{01}

Denominator: 0*n_{11} + 0 *n_{00} + 0*n_{10} + $\frac{1}{2}$ *n_{01}

$$OR_{MH} = \frac{n_{10}}{n_{01}}$$

The ratio of discordant pairs

Benefits of matching

Reduces bias due to:

- spurious imbalance
- selection bias where there is no population register
  (e.g. geographic region, hospital)

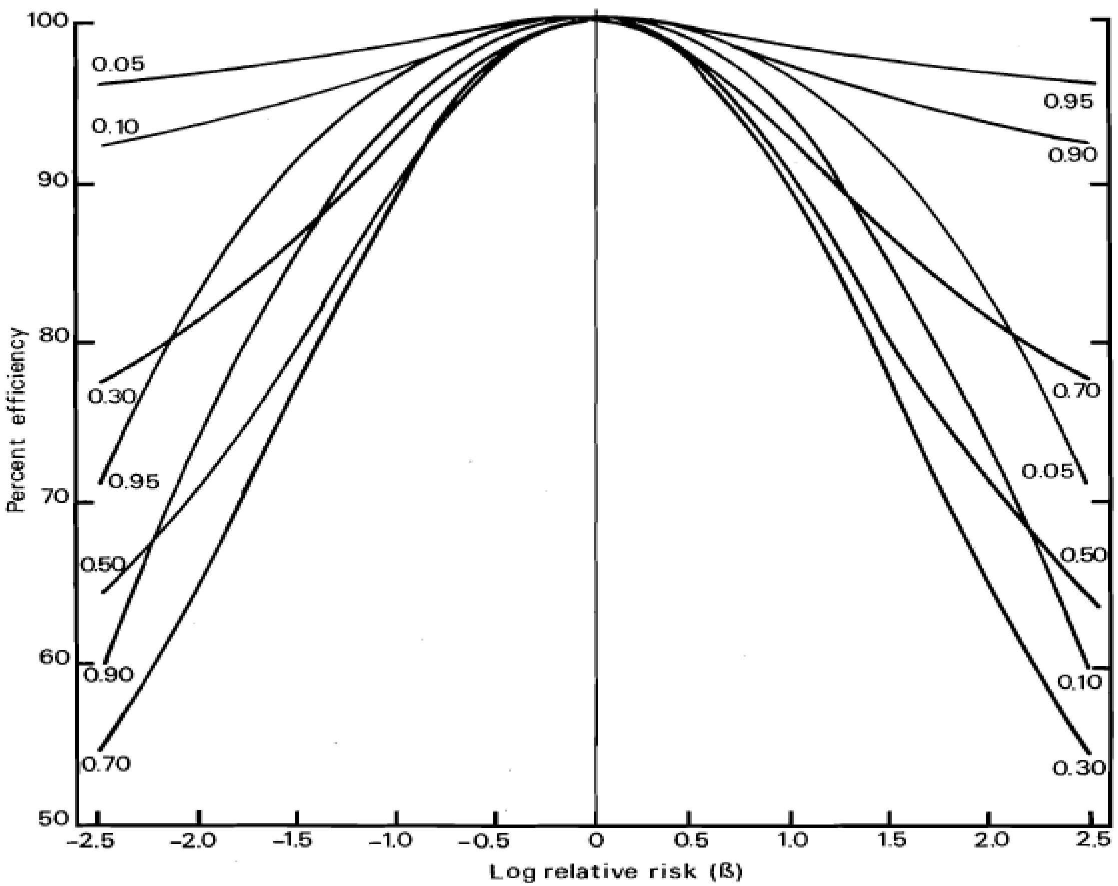
allows adjustment for “unmeasurable” confounders
(e.g. genetics, neighbourhood effects,….)

allows cases and controls matched for “exposure window”

Improves efficiency (provided matching factor IS a confounder)

Eaxmple: loss of efficiency from unnecessary matching

Fig 7.1 Breslow & Day (random pairs)



<u>Loss:</u>

- zero if OR=1
- depends on prevalence of exposure
- only large for extreme ORs

Limitations of matching

1. Potential for overmatching

2. Cost (time and effort) to ”find a match”

3. Study planning: e.g. if enrolling cases and “concurrent” controls, cannot know in advance the numbers needed

4. Choice of matching categories:
   - Too wide, insufficient adjustment for confounding
   - Too fine, loss of concordant sets

5. Unnecessary matching (Loss of efficiency)

In selecting study subjects, the objective is to avoid:

Bias (”fair”)

Confounding (”balanced”)

Chance (”large enough”)