Bias & Confounding

Chihaya Koriyama

August 4th, 2014

Key concepts

- **Confounding**

→ Indicative of true association. Can be controlled at the designing or analysis stage.

- **Bias**

→ Should be minimized at the designing stage.

- **Random errors**

→ Is the nature of quantitative data.

- Non-differential misclassification

→ Is the nature of (inaccurate) measurement.

ERROR VS. BIAS

Two types of errors: ---Error or bias?

- Random error
  - ⇨ is the nature of quantitative data.

- Systematic error (=bias)
  - ⇨ should be minimized at the designing stage.

Random error

Measured value (mm)
53
47
48
49
51
52
50
Mean=50

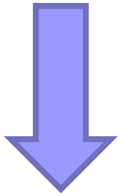
Systematic error

Measured value (mm)
48
48
48
48
48
48
48
Mean=48

God knows that the true value is 50mm.

Which is a proper comparison?

- Using accurate data
- Using inaccurate data



Can't we use our data when it is NOT accurately measured?

Is the following study acceptable?

- We want to compare the mean of blood pressure levels between two groups.
- The blood pressure checker has a problem and always gives 5mmHg-higher than true values.
- All subjects were examined by the same blood pressure checker.

Internal comparison

Proper comparison between groups :

1) Comparison using accurate data

2) Comparison using (in)accurate data

As long as the magnitude of random error and bias occur in a same manner among groups.

Two types of misclassification

- **Non-differential** misclassification
  - Systematic error may not be a critical issue as long as it occurs in all comparison groups.

- **Differential** misclassification
  - If the error occurs only in one specific group due to bias, the risk estimate deviate from null.

Two-group comparison with random errors

God knows that the true value is 50mm in both groups.

Group A(mm)	Group B(mm)
53	47
47	51
48	52
49	50
51	48
52	49
50	53

Mean?

Mean difference=0 → correct result

Systematic error occurred in both groups

God knows that the true value is 50mm in both groups.

Group A(mm)	Group B(mm)
49	48
48	49
46	49
47	48
49	46
49	47
48	49

Non-differential misclassification

Mean?

Mean difference=0 → correct result

Systematic error occurred in only group B

God knows that the true value is 50mm in both groups.

Group A(mm)	Group B(mm)
53	45
47	49
48	50
49	48
51	46
52	47
50	51

Differential misclassification

Mean?

Mean difference = 2 → wrong result

BIAS IN EPIDEMIOLOGIC STUDY

Different types of bias

- **Selection** bias:

  It occurs at sampling

- **Detection** bias:

  It occurs at diagnosis (outcome)

- **Measurement (information)** bias:

  It occurs at surveillance

  - **Recall** bias
  - Family information bias

Selection bias

- Selective differences between comparison groups that distort the relationship between exposure and outcome

- Unrepresentative nature of sample

  Usually, comparative groups *NOT* coming from the same study base and *NOT* being representative of the populations they come from

A case-control study of childhood leukemia and exposure to electromagnetic field (EMF)

- If parents of cases, living in the neighborhood of power lines, suspect the association and tend to agree on participation to the study,

  ⇨ the association may become stronger than what it should be.

A case-control study of childhood leukemia and exposure to electromagnetic field (EMF)

- All the parents of cases may be willing to participate in the study. On the other hand, the parents of healthy children may tend to participate in the study only if they live in the neighborhood of power lines since EMF exposure is strongly suspected to be related to power line.

⇨ The association may become weaker than what it should be.

Selection bias caused by low participation rate

- In a case-control study for lung cancer
- Cases were identified by cancer registry
- Controls were recruited from a population base but the **participation rate was too low**, say 20% (in general, **health-conscious** people tend to participate in this kind of study).

What happened in the association between smoking and lung cancer risk is that ….

⇨ the association become **stronger** than what it should be

Is Reserpine a cause of breast cancer?

Cases: Breast cancer patients

(no selection bias) (selection bias)

Controls: Patients at the same hospital

Horwitz RI, Feinstein AR. Exclusion bias and the false relationship of reserpine and breast cancer. Arch Intern Med. 1985;145(10):1873-5.

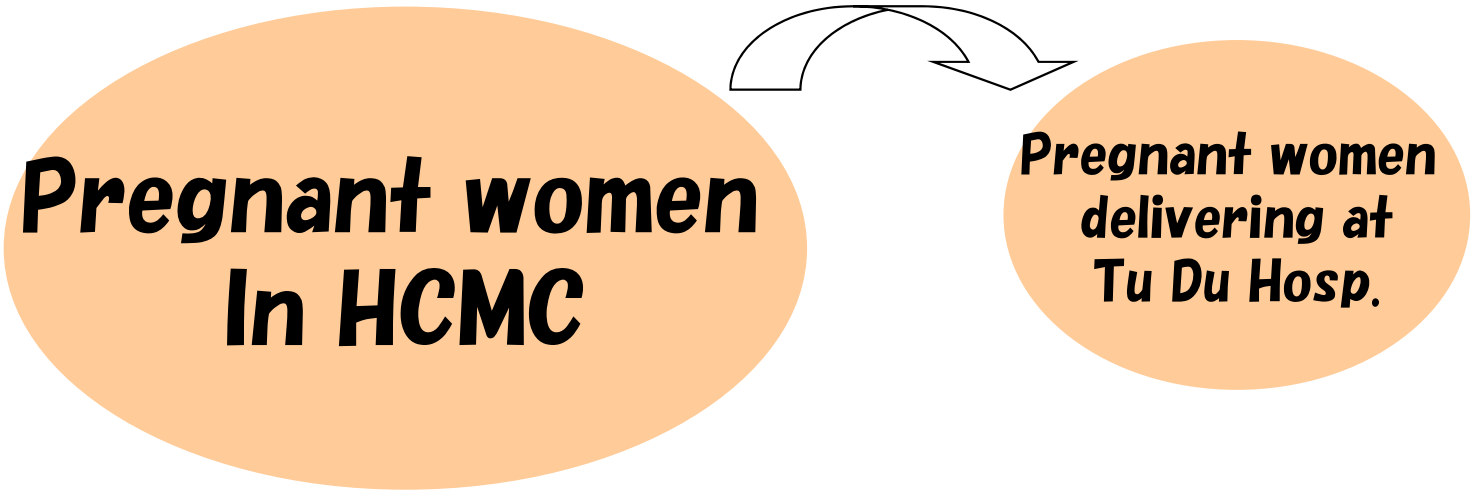
(Except who have cardiovascular diseases to which Reserpine is likely to be prescribed.)

Selection bias influences internal validity of the obtained results.

NOTE for advance learners:

Sampling is a different issue from selection bias.

Prevalence of postpartum depression at Tu Du
= Prevalence in HCMC?



Sampling influences **generalizability (external validity)** of the obtained results.

Detection bias

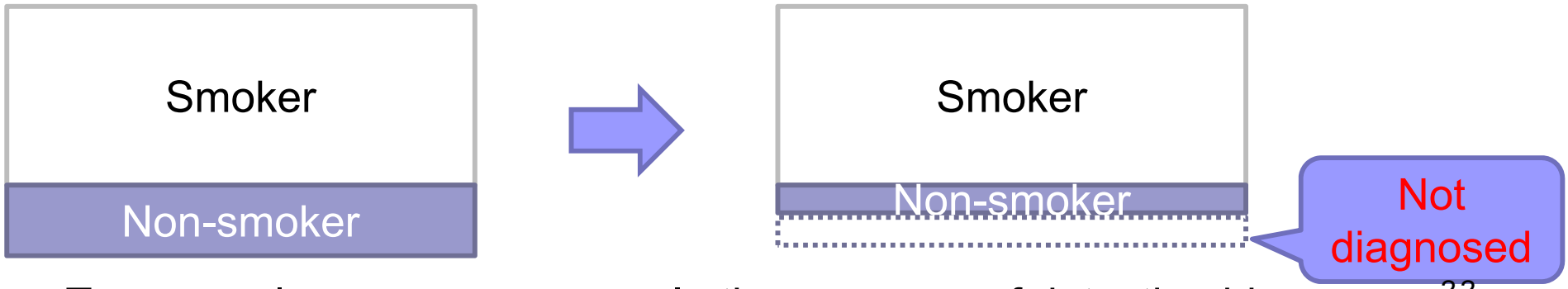
- Typically, this is the situation where the exposure of interest makes asymptomatic case to symptomatic.

- It is a special situation where case ascertainment depends on exposure.

Diagnosis of lung cancer among smokers

- A doctor may examine the patient's X-ray more carefully if he knew the patient was a heavy smoker but not for the non-smoking patient.

  ⇒ the association may become **stronger** than what it should be.



True prevalence

In the presence of detection bias

A case-control study of acoustic neuroma and mobile phone use

- This brain tumor is asymptomatic and is occasionally noticed by hearing difficulty or hearing loss. In other words, those who use mobile phone may have a higher chance of noticing unilateral hearing difficulty and visiting hospitals, where the acoustic neuromas are detected.

⇨ the association may become **stronger** than what it should be.

Measurement (information) bias

- Once the subjects to be compared have been identified, the information to be compared must be obtained.
- Information bias can occur whenever there are errors in the measurement of subjects, but the consequence of the errors are different, depending on whether distribution of errors for one variable (e.g., exposure or disease) depends on the actual values of other variables.
- For discrete variables, measurements error is called classification error or misclassification.

"Modern Epidemiology", Rothman, Greenland, and Lash

Rheumatism history in parents was asked to rheumatic patients and their siblings

	the history of rheumatism in parents	
	Yes	No
Rheumatic Patients	73%	27%
Siblings	50%	50%

Recall bias

"systematic error due to differences in accuracy or completeness of recall to memory of past events or experiences.

(JM Last)

Case-control study on relationship of prenatal infections and congenital malformations

Cases (mothers of babies with defect)

Controls (mothers of healthy babies)

Cases' mother <u>recall better</u> about prenatal episode of infections since they tend to think about possible causes of their babies illness.

Obtained results: Relationship between baby's defect and prenatal infection will be exaggerated.

Method to minimize this recall bias
→Consider using a hospital control.

Controlling for misclassification

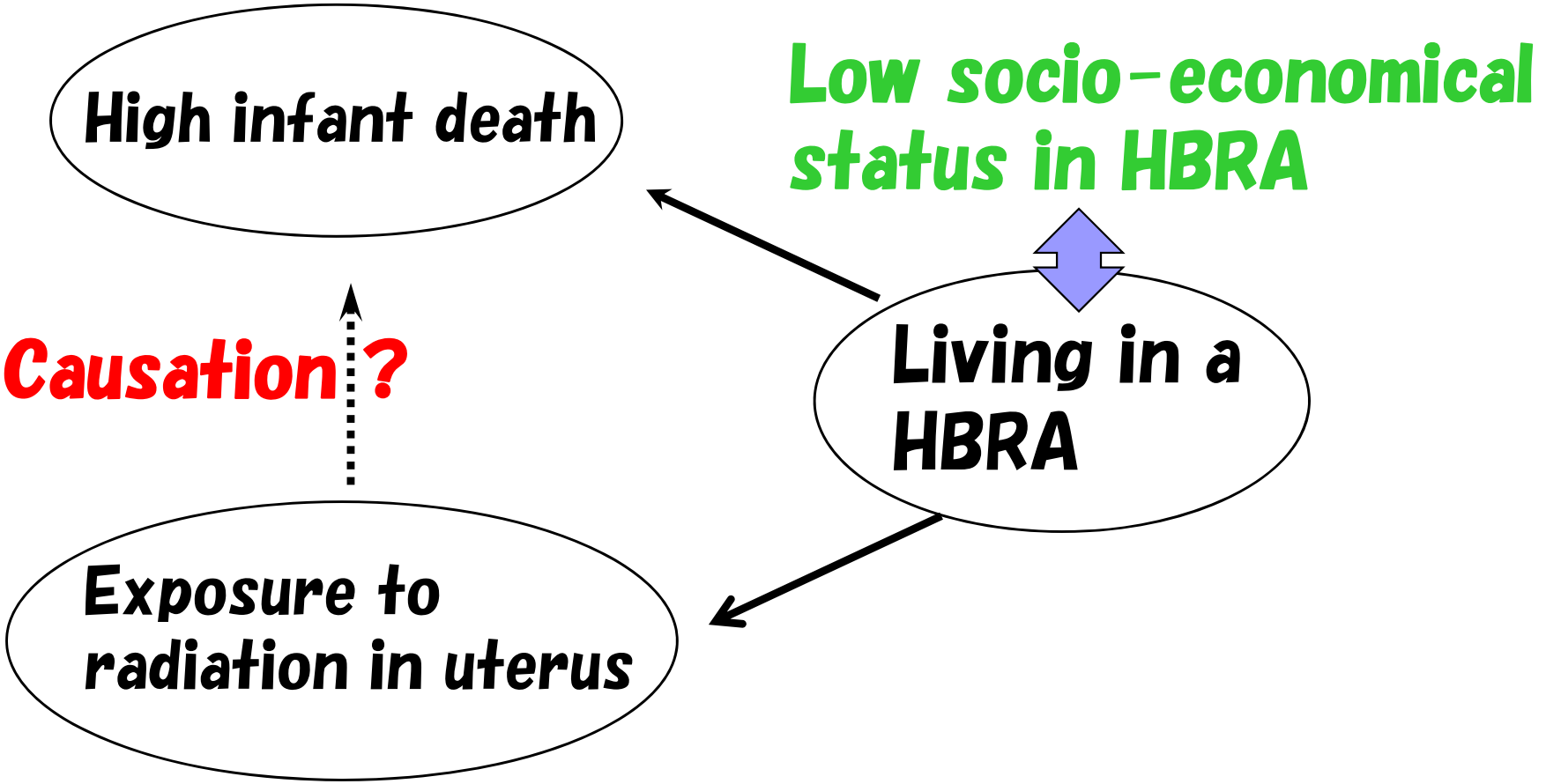
- - **Blinding**
  - prevents investigators and interviewers from knowing case/control or exposed/non-exposed status of a given participant
- - **Form of survey**
  - mail may impose less "white coat tension" than a phone or face-to-face interview
- - **Questionnaire**
  - use multiple questions that ask same information
- - **Accuracy**
  - Multiple checks in medical records & gathering diagnosis data from multiple sources

CONFOUNDING

Confounding

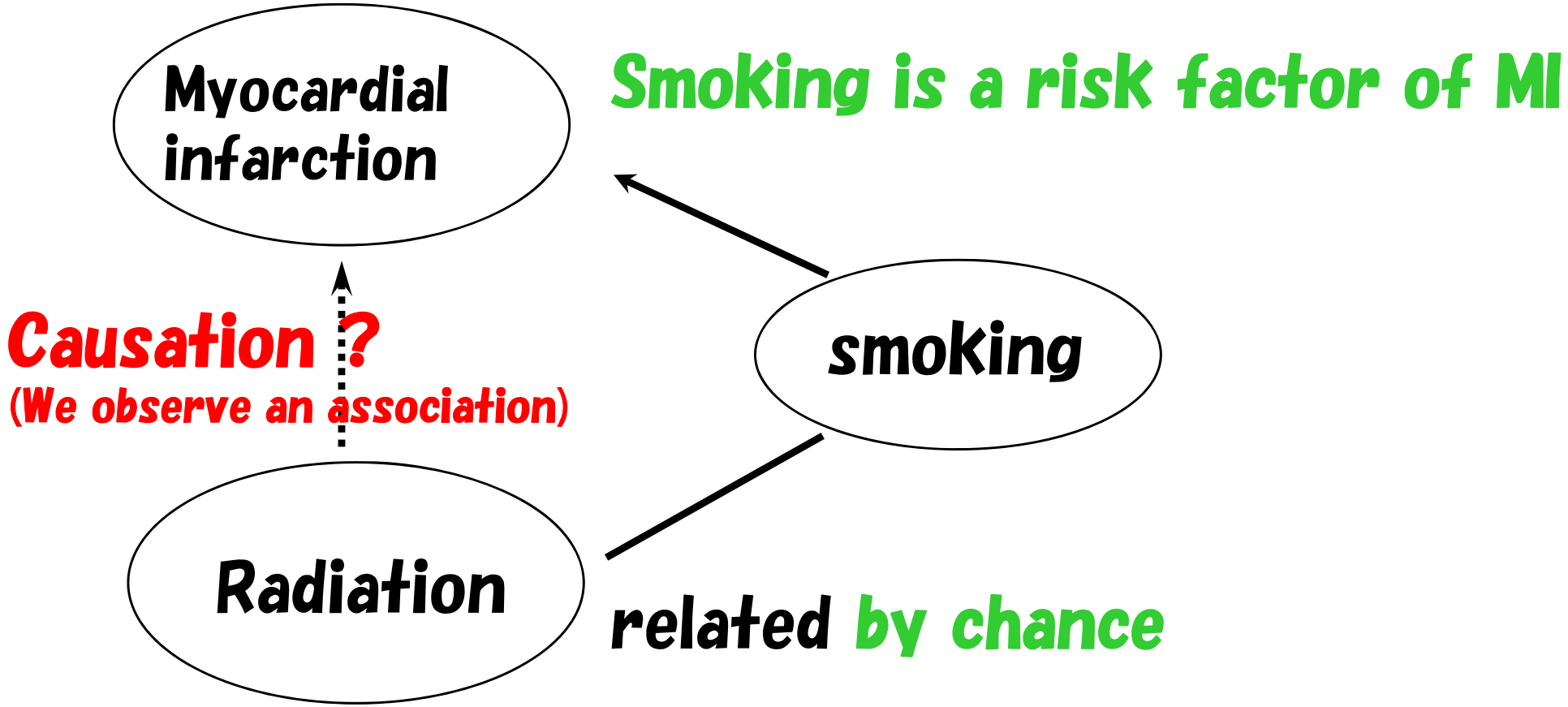
- Confounders are **risk factors** for the outcome.
- Confounders are **related to exposure** of your interest.
- Confounders are **NOT on the causal pathway** between the exposure and the outcome of your interest.

Example of confounder
– living in a HBRA is a confounder –



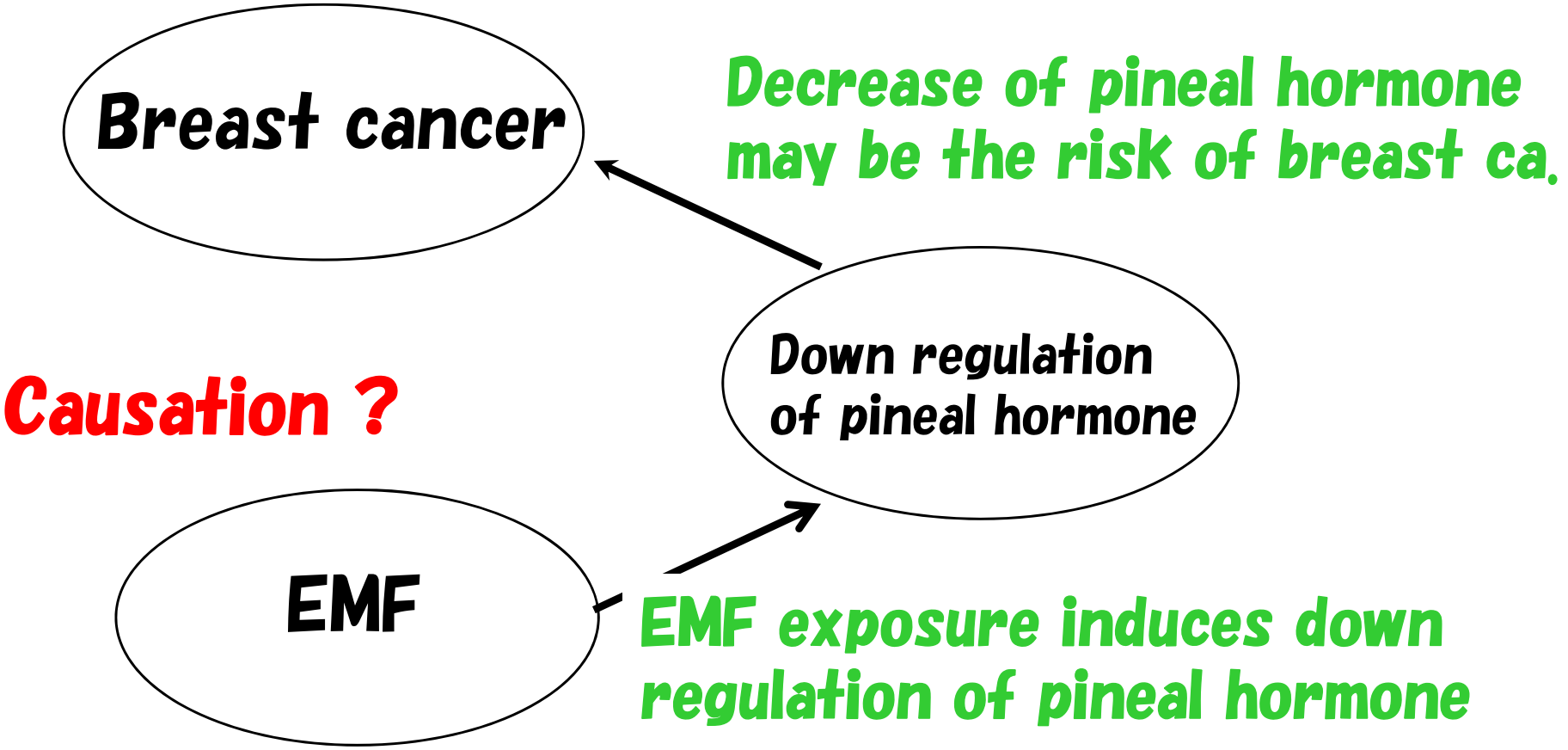
HBRA: high background radiation area

Example of confounder
– smoking is a confounder –



Example of "not" confounder

- pineal hormone is not a confounder -



EMF: electro-magnetic field

Why do we have to consider confounding?

- We want to know the "true" causal association but a distorted relationship remains if you do not adjust for the effects of confounding factors.

How can we solve the problem of confounding?

"Prevention" at study design

- ✓Limitation
- ✓Randomization in an intervention study
- ✓Matching in a cohort study

Notice: Matching does not always prevent the confounding effect in a case-control study.

How can we solve the problem of confounding?

"Treatment " at statistical analysis

- ✓ **Stratification** by a confounder
- ✓ **Multivariate** analysis