

Data analysis with R software

Data-analyysi R-ohjelmistolla

Tommi Härkänen

National Institute for Health and Welfare (THL), Helsinki
E-mail: tommi.harkanen@helsinki.fi

University of Helsinki, February 8, 2012

Contents

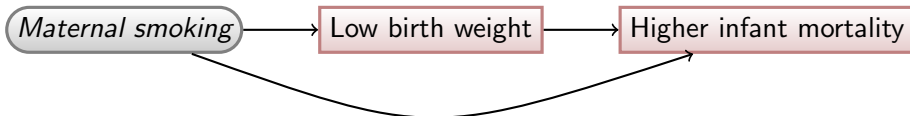
Adjustment

Causes

Causality relations are often depicted using graphs. **Nodes** are connected with **arrows**, which represent causality. Example:

Low birth weight → Higher infant mortality

Problem is that maternal smoking appeared to decrease infant mortality! Why? (e.g. Hernandez-Diaz *et al.* 2006, Wilcox 2006)



Low birth weight is often called an **intermediator**.

Confounders

- ▶ “Note that there might be common causes of smoking and infant mortality (e.g., socioeconomic factors) that would induce **confounding**.”
- ▶ “there would be common causes of LBW and mortality (e.g., birth defects, malnutrition).”

Here socioeconomic factors are background factors, which usually **precede** smoking.

Confounders

“In summary, there will be selection bias when stratifying on [or adjusting for] a variable that is

1. affected by exposure and
2. shares common causes with the outcome.

Such a variable does not need to be on a causal pathway between the exposure and the outcome (i.e., it does not need to be an intermediate variable).”

Confounders

Confounders – necessary conditions¹. The factor must:

- C1 be a **cause of the disease, or a surrogate measure of a cause**, in unexposed people; factors satisfying this condition are **called risk factors** and
- C2 **not be an intermediate step** in the causal pathway between the exposure and the disease
- C3 **not be affected** by the exposure

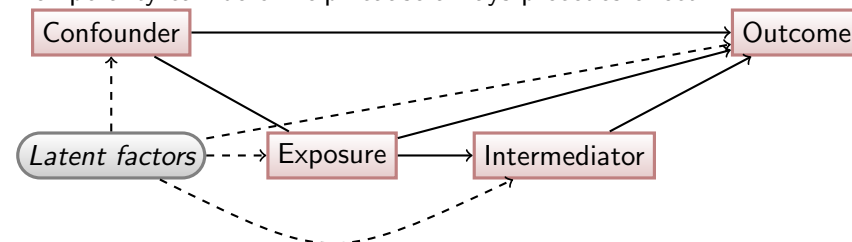
Confounders usually need to be adjusted for in statistical analyses.

¹<http://oem.bmj.com/content/60/3/227.full>

Relationships of variables: Summary

Before building a (regression) model, the relations of different variables must be assessed with care.

Temporality can be of help: cause always precedes effect.



Effects of latent factors are difficult to assess. Randomization is often the only way to remove the confounding.