

Exercise for day 1 (Monday before lunch):

Background for defining research questions

Number 1 – Diabetes and mortality

You want to find out if people with type I and type II diabetes are at risk of dying from other causes than people without diabetes. Define exposure and outcomes and formulate one research question which can be answered by yes or no, and describe how register data can be used to design a study to examine this research question?

Number 2 – Air pollution and acute cardiovascular events

You wonder if air pollution affects the risk of acute cardiovascular events. Exposing subjects to pollution would be unethical, so conducting a randomized study is out of the question. Define exposure and outcomes and formulate a research question which can be answered by yes or no, and explain how you intend to answer that question using register data. Could other data sources be used to improve your study?

Number 3 – Vitamin D and colorectal cancer risk

You want to find out if low serum vitamin D levels affect the risk of colorectal cancers. A randomized trial is not feasible, and you decide to use data that are already collected. Formulate a research question that can be answered by yes or no, and explain how you intend to answer this using register data in combination with other data.

Number 4 – Asthma medication and ADHD

Montelukast is a leukotriene receptor antagonist drug used in the treatment of asthma. Side effects of the drug include mood changes and sleeping problems and you want to find out whether the drug increases the risk of Attention Deficit Hyperactivity Disorder (ADHD). A randomized study is difficult because the risk of ADHD is low and large sample sized would be needed in a RCT study. Formulate a research question that can be answered by yes or no, and design a register-based study to answer this question.

Number 5 – Change of partner and risk of preeclampsia

Preeclampsia is a condition that affects about 2-8% of pregnant women and causes a life-threatening rise in blood pressure. The only known cure is to end the pregnancy, preferably by giving birth. The risk of preeclampsia is lower among women who have given birth before, and you want to find out if this is also true for women who change partners between their first and second pregnancy. Formulate a suitable research question that can be answered with yes or no, and design a register-based study that answers the question.

Number 6 – Childhood cancer and socioeconomic status in adulthood

You want to investigate the long-term socioeconomic consequences of surviving childhood cancer. A randomized study is difficult to perform. Explain how you would use registry data to conduct this study. What are your outcomes?

Number 7 – Use of painkillers and risk of violent death

You suspect that use of prescription painkillers is associated with behaviors that increase the risk of dying a violent death (accident, murder or suicide). An ethical and feasible randomized trial seems very difficult. Formulate a research question that can be answered with yes or no, and design a register-based study that answers this questions.

Number 8 – Death in close family and depression or anxiety

You want to study the impact of death in close family on the risk of depression or anxiety. Formulate a research question that can be answered with yes or no, and design a register-based study that answers this answers.

Number 9 – Pierre Robin syndrome and school grades

Pierre Robin syndrome is a birth defect affecting about one in ten thousand children that is characterized an unusually small mandible (jawbone). This syndrome may cause delayed speech development, and you wonder if this may also lead to worse school grades. Formulate a research question that can be answered with yes or no, and design a register-based study that answers the question. Pay particular attention to power issues.

Number 10 – Long-term use of benzodiazepines and risk of heart attack

Benzodiazepines are widely used to treat anxiety and insomnia and some persons have long-term use. There has been raised concern of a detrimental effect on the heart including heart attacks by long-term use. Formulate a research question that can be answered with yes or no, and design a register-based study that answers the question.

Exercise day 2 (Tuesday before lunch)

Study design

On Monday you discussed how to answer research questions based on registers. This exercise focuses on the same research questions you discussed yesterday:

1.

Which epidemiological study design do you think is most appropriate for your research question?

2.

What are the characteristics of your study population and analysis data set?

Please consider the following dimensions:

- Unit of observation
- Open or close population
- Any sampling of persons included in the study
- Any inclusion or exclusion criteria

3.

What are the main strengths and limitations of using this study design? Please include your responses to question 2.

4.

What alternative study designs could be used? And what is the reason for not choosing this or these designs?

Exercise day 3 (Wednesday before lunch)

Truncation bias and immortal time bias

Read Brøndum-Jacobsen et al. Skin cancer as a marker of sun exposure associates with myocardial infarction, hip fracture and death from any cause. *International Journal of Epidemiology* 2013;42:1486–1496 and answer the following questions:

1. What is the aim of the study?

In the following questions, only look at the association between exposure and death of any cause (overall mortality)

2. How are exposure and outcome defined?
3. When does the risk time start and end for both exposed and unexposed? In the alternative design (matched case-control design), how are the cases and controls sampled?
4. Please explain the statistical model used in the main study.
5. What is the main finding of the main study?
6. And what is the main finding of matched study?
7. Please try to explain how immortal time bias may have influenced the results?
8. Please consider whether truncation bias may have influenced the results?

Exercise day 4 (Thursday)

Datamangement - Exercises in R, STATA and SAS

Load birth registry data (Should take less than a minute).

1- Consider dta.mfr. This is a data set that was made to look like data from the

Medical Birth Registry. The variables are:

lopenr: ID number. Unique for all subjects

fdato: Date of birth for subject

ffdato: Date of birth for the father of subject

kjonn: Sex of subjects

vekt: Birth weight of subjects

We want to see if there are sex differences in birth weight,

and if paternal age at birth affects birth weight

i. Clean up the data

a -How many records are there?

b -How many unique "lopenr" are there?

c -Why is there a difference between your findings in a) and b)? Remove redundant records.

d -Make a histogram of fdato. Does everything look OK? Explain.

e -How many birth dates are missing?

f -How do we handle the missing data?

g -How many paternal birth dates are missing?

h -Make a histogram of paternal birth date. Does everything look OK? Explain.

i -Create a variable, agedad, which is paternal age at birth, and make a histogram.

Does everything look OK?

j -Drop records where paternal age at birth is unrealistic. Which cutoff(s) do you use?

k -Re-draw the histogram from h). Comment on the differences.

l -What are the minimum and maximum birth weights? Do they look realistic?

m -Make a histogram of birth weight. Does everything look OK?

ii. Run analyses

a -Are there sex differences in birth weight? If yes, how big?

b -Does paternal age affect birth weight? If yes, how much?

c -Repeat a) and b) on the original data set. Comment on the results.

2- Consider edu.dta. This is a data set that was made to look like data from the

Educational Database. The variables are...

lopenr: ID number. Same as in the first data set.

faar: Birth year for subject.

ffaar: Birth year for father of subject

utdaar: Education year

utd: Education in education year

0- No elementary school (barneskolen)

1- Elementary school (barneskolen)

2- Lower secondary school (ungdomsskolen)

3- Upper secondary school, first two years (VGS, grunnutdanning)

4- Upper secondary school, third year (VGS, avsluttende utdanning)

5- Upper secondary school, additional year (VGS, påbygging)

6- Lower level university (e.g., bachelor)

7- Upper level university (e.g., master)

8- PHD

9- Not given

We want to see if paternal age at birth affects education

i. Prepare data

- a -How many records are there?
- b -How many unique "lopenr" are there?
- c -Is the large number of rows a problem? Why (not)?
- d -Remove rows with same lopenr AND utdaar. Keep the bottom one (highest).
- e -Tabulate the education variable. What do the numbers in the table mean?
- f -Convert from long to wide format, using education each year as a time-varying variable
- g -How many records are there now?
- h -How many have missing values on education in 1967?
- i -How many have missing values on education in 2016?
- j -Tabulate the values in your new data.frame and compare with e).
- k -How do we handle missing data in the education variable?
- l -Tabulate education in 2016. Why are there so many zeros?

ii. Merge with former data set

- a -Merge the two data sets. How many records do not match?
- b -Handle non-matching records. What did you do?
- c -Does paternal age affect education?

Explain how you performed the analyses, and what the results were.

Exercise day 5 (Friday afternoon)

Choose one of the research questions you discussed on Monday and consider whether it is possible to conduct this study in a cross-Nordic setting. Please consider the following questions:

- What is the main benefits of conducting a cross-country study compared to a single country study?
- And the main limitations and obstacles.
- You have chosen a specific study design to answer your question. Do you have to consider an alternative study design if you need to conduct a cross-Nordic study?
- Will it be possible to obtain individual level data from all countries and will it be possible to combine all individual level data in one common database?
- Which legal considerations do you have to consider when designing the study?

Please choose another of the research questions from Monday, where you decided another study design and answer the same questions.