RECAP summer school
WP8
Anna Pulakka

Case study: adding register data to a cohort study

Abstract
A practical overview of adding register data to a cohort study using the ADHD study as an example.

Learning aims

- To identify register data as a possibility for expanding analyses in certain study questions.
- To understand what needs to be taken into consideration when adding register data to a cohort study.

Script for ADHD case study

Title slide
Hello and welcome! My name is Anna Pulakka and I work as a senior researcher at the Finnish Institute for Health and Welfare, THL. I am mainly working with the register data in the RECAP project, work package 8.

In this section I will give you a short real life example on how to add some register data analyses to an analysis of cohort data. I aim to give this example in a practical level, and at the same time to demonstrate some aspects of working with register data.

Practicalities of adding register analyses
How did we do this in practice? In another section of this course, Rachel Robinson from the University of Helsinki told about their study in WP9, where they did an IPD meta-analysis to find out if those who were born preterm have higher risk of ADHD symptoms in adulthood. In their study they used data from 8 different cohorts. In the cohort data, preterms did not differ in the total number of ADHD symptoms compared to term borns.

Rachel and others from University of Helsinki contacted us to find out if it's possible to add some register data analyses to see how the cohort results compare to the data from registers.

We were very much interested in this collaboration, and after the contact, we agreed on a detailed analysis plan and the variables used with the team in University of Helsinki. In this case, we wanted to do the analyses as similarly as possible in both the cohort and register studies.

I did the analyses at THL, and exported the results to them. In this case it was more practical for me to do the analyses rather than us acquiring the permissions for someone at the University of Helsinki team to do the analyses. In reality there of course were several rounds or discussions and more analyses to run as the plans for the sensitivity analyses got more detailed over time

Variables

Next I'm going to describe the variables that we used for you to see an example which variables are possible from the register data.

Not surprisingly, gestational age was our main exposure variable. This was derived from the medical birth register. Earlier studies have found satisfactory validity of gestational age in the medical birth register compared to medical records (Gissler 1995). The method of assessing gestational age has obviously changed over time. At the beginning of the register, timing of the last menstrual period was used for assessing gestational age, but measurement by ultrasound scan became common early on and is now the main method.

There is no information on ADHD symptoms in the registers, however, from the hospital discharge register we can see the **diagnoses for ADHD,** by code F90 using International Classification of Diseases (ICD) codes, version 10. We can also get the admission dates for the visits when this diagnosis was received. It is important to note, that the Finnish hospital discharge register only includes visits to a secondary/speciality care. Thus, we are probably missing some people who have received the diagnosis in primary care.

In this case, we were only interested in ADHD in adulthood, thus we only included diagnoses after 18 years of age. This means that we had to limit the data to those born year 1998 or earlier. This also means that we can focus on ICD version 10 as the first diagnoses are from the year 2005. Before year 1996 the diagnoses were coded with ICD version 9.

We used quite a simple set of covariates. Sex is available at the medical birth register. Maternal education is derived from Statistics Finland. Age at the end of the follow-up was calculated using birth date from the MBR, date of death and date of emigration from the population register.

In this study we also used diagnoses for intraventricular hemorrhage (IVH) and bronchopulmonary dysplasia (BPD). These diagnoses were derived from the hospital discharge register similarly as the diagnosis for ADHD, but for these we needed to use ICD codes, version 9, too.

We can see that even with a simple set of variables, the data comes from many different registers. The key is being able to link these data together. This is possible with personal ID codes, which is an unique feature of the Nordic societies. The same personal ID code is used in all of these registers.

Conclusion

As mentioned earlier, Rachel is presenting the results of the ADHD study elsewhere at this course. I have focused on the practical aspects of adding register data to a cohort study.

There are several advantages in adding register data to a cohort study. We have register data from up to millions of people, which of course increases power of the study. This also allows studying subgroups, for example extremely preterm born people or those with small gestational weight.

However, register data has a limited set of outcome variables, mainly diagnoses received in secondary health care. Registers also do not include information of behavioural factors, such as physical activity or

food intake, or this information is limited. Thus there needs to be a clear understanding if the research question can be answered with register data.

In regards to the practical aspects of adding register data to cohort studies, simple analyses are quite feasible to run, given that there is a clear and detailed analyses plan and the variables are well defined and agreed on. Thus collaboration like the one have I described is well doable in many cases.

We do welcome you to contact us if you have research ideas that could be studied with register data. In other sections of this course, the team leaders of RECAP Nordic partners will tell shortly about the specifics of the register data in their countries. We are happy to collaborate and look forward to hearing from you.