WP08_02 Register-based research in the Nordic countries

Eero Kajantie, Signe Opdahl

Transcript

Slide 1

This is how much of health research has historically been done. You as a researcher have an idea. To realise an idea, you usually need to plan a study to collect new data – for example, a health examination. If you plan the data collection well, you can get the type of data you want. However, this takes a lot of time and money.

In many countries, in particular high-income countries, the authorities collect lots of data in individuals in administrative registers. These data can be very valuable for resarch

However, you have to remember that they have been collected for some other purpose than research.

Here we give a brief overview on how linkage data from different registers can be used to study long-term outcomes of preterm birth. In some cases register data can also be used to enrich data collected from individuals in a cohort study. In this presentation we, however, restrict to studies that are only based on register data.

Slide 2

The words register and registry are both used, sometimes interchangeably. When you look up in a dictionary, a register refers to the list or database itself, whereas registry has a flavour of an office or authority that maintains registers. The interchangeable use is illustrated by that the official English names of registers are different in different Nordic countries.

Nonwithstanding this, the word originates in Latin, and for our purposes it means indivudial-level information on a specific group of individuals, aimed to comprise the whole target population and linked to an identifier which allows regular updating.

When we use these data in research, we in general talk about secondary use of administrative data.

Slide 3

Here are some examples from Norway illustrating different registries.

Slide 4

The essential ingredient to combine individual data from different registries, is the national identity number. In the 1940s to 1960s, all the Nordic countries assigned a unique identification number to all their residents, and to all children born after this time point, as well as all immigrants. As an example, the Norwegian identity number consist of a person's date of birth, followed by 5 digits chosen after a specific pattern, 11 digits in total. This number follow all of us in all our contact with authorities and other situations where definitive identification is needed, including the registries. This allows for 100 % reliable linkage of individual level information from different registries for the entire population, but requires a high level of trust between the public and the authorities to be acceptable from an ethical perspective.

Slide 5

In RECAP, research using data from the Nordic registries is organised in a separate work package. To the left, we show the main collaborating institutions, and to the right, we have listed the size of the total population and the typical size of the birth cohorts in each country. You can also see when the Medical birth registries, which form the basis for the research cohorts in WP8, were established.

Slide 6

Some of the most obvious strenghts of registry data stem from the fact that the registries are nationwide and most have mandatory or automated reporting. This means that the entire population is included, which gives us large number of participants with practically no selection or loss to follow-up.

Other advantages are that the data are collected regardless of our particular research project and mostly as part of clinical routine, so participants are not subjected to any extra tests or procedures as a result of our research. Although there are costs involved in the linkage procedures and the data management is complicated and time consuming, registry-based research have low costs compared to clinical studies.

The registration of family relations also allows for inventive designs taking genetics and other family factors into account.

Slide 7

If you are thinking that this sounds too good to be true, we can reassure you that there are also a number of limitations to be aware of when using registry data for health research.

It is important to remember that the data are collected as part of clinical or administrative routine, and not specifically for your research project. Typically, data are recorded for diseases and medical procedures, but there are generally few data describing function, behaviour or lifestyle. You might not find all the data on your wish-list and important confounders might be unavailable.

Most of our key health registries have existed for decades, and although this allows for trend studies and long-term follow-up, it may also mean that registration practices have changed over time and it certainly means that a large number of health workers have performed the registration.

Slide 8

Here we illustrate by some examples how the quality of different variables may differ, and even for some variables it may differ based on definition and time.

Death registers and cancer registers are in general very reliable, use multiple sources to obtain information and have resources to verify individual-level data should questions arise. Birth weight, recorded in a medical birth register, is also usually very complete and reliable. The same applies to a number of severe and common medical disorders usually diagnosed or treated in a hospital. For such disorders there often are published studies in which the diagnoses have been validated against medical records.

Asthma is an example of a disease which can be indicated by several types of register data. You can use diagnoses in a specialty inpatient or oupatient clinic. However, at least for adults, only people with more severe symptoms are likely to visit these clinics, while most are diagnosed in primary care. Primary care data have traditionally not been available for research, althought this is changing. But currently we need to restrict us to specialty clinic diagnoses, which generally are quite reliable, but capture only the more severe cases; in other words, they are specific but not very sensitive. The same applies to something we call medication special reimbursement, which is unique to Finland; if you fulfil defined lung function criteria, by a clinician's statement approved by a physician in the national social insurance institution, and need medication most of the year, you qualify. Again, specific but not very sensitive. By contrast, if you take all who have purchased a prescibed asthma medication, you will identify more people, but a proportion of them probably have medication for symptoms and may not qualify for the lung function criteria.

Gestational diabetes is an example of a disease which varies by country and over time. This example is from Finland. Since 2004, the Medical Birth Register has included information on whather an oral glucose tolerance test has been abnormal and whether a mother has used insulin in pregnancy. This is fairly sensitive and specific. IN 2003 and before, this information come through ICD diagnosis codes recorded in Medical Birth Register or Hospital Register. The ability to distinguish gestational from other types of diabetes, mainly type 1, has also improved with the newer ICD versions.

Slide 9

Because of privacy restrictions, we have not been able to make individual-leven register data directly available through the RECAP platform. However, the key datasets have been harmonised and metadata are available on the platform. How can you access these data?

I want to emphasise that now, in 2021, the system is very much in transition. Decision-makers are more and more understanding the utility of register data, and technical developments have allowed the creation of secure remote access systems, where datasets specified in data permissions are located in a centralised server, which typically operate standard statistical software (but may not accommodate special software), but typically do not allow export of individual-level data. These systems can be very helpful in allowing data access from different locations, but have restrictions, in particular when using data from different countries simultaneously.

Regardless of a possible secure remote access system, the use of individual-level data is always based on a research plan and a permission from the register authority. This is given for variables specified in the application and restricted to individuals listed in that application. In Norway and Sweden, you in addition need an Ethics approval. Generally, if you have a sensible research plan, you should get a permission, but sometimes this can take very long time; the time can be difficult to predict.

Traditionally there have been restrictions in accessing the data from abroad, although, with some exceptions, there should not be legal constraints within the European Economic Area.

All this means that this is a long-term process, and many researchers choose to collaborate with a Nordic group. In the RECAP register work package, we welcome collaborations.

Slide 10

You can find information on the Nordic register core dataset harmonised for RECAP on the RECAP platform. By clicking on read more…

Slide 11

…and scrolling the page down you will find a link to a ”crash course” document in which you will find more detail. You are also welcome to contact one of our team leaders in these e-mail addresses.