

Requirements for description of data extraction

”Forskerservice Klassisk”

To ensure that you receive the data you need, and to facilitate administrative procedures as well as the work involved in extracting data, enclose a separate data extraction description with your application.

We would like to draw your attention to the fact that data on individuals’ health is personally sensitive information. This means that for each research project, we only deliver data that are necessary to carry out the project.

Unless you justify/motivate the need for e.g. all variables or all rows from a register, we normally do not deliver them all. We usually want you to restrict the data extraction you ask for.

Below, you will see a number of items to be included in the description of a data extraction. If in doubt, please contact the research support unit at SDS.

1. Information on population / cohort

If the data extraction concerns a well-defined population, you need to specify whether you already set up the population or if the research support unit should set it up for you.

In case you set up the population, you need to mail a list of CPR numbers [personal IDs] after you submitted your application. These CPR numbers should be sent by registered post on a CD / USB key labelled with the SDS file number (FSEID). The list format: SAS file or text file (e.g. csv) with one line per CPR number, preferably numbers without hyphens. When we receive the CPR numbers, we check them for duplicates and validity. Data are extracted for valid CPR numbers, only.

We also ask you to note from where you got the population’s CPR numbers.

If you want the research support unit to set up the population, you have to specify exact delimitation criteria. They could include

- Registers
- Diagnoses
- Procedures (surgical, examination, treatment)
- Causes of death
- Age intervals
- Gender
- Time periods (dates)

2. Type of data

State if you need the data as individual data with personal ID [CPR number], individual data with scrambled/de-identified personal ID or without personal IDs.

In general, we only deliver individual data with scrambled/de-identified personal IDs. If you apply for individual data with true personal IDs you have to substantiate the need for this information in your research.

3. Time period

Specify the time period for which you need the data extraction.

If you apply for data from the National Patient Registry we encourage you to consider in advance if the time period is to be defined based on admission date or discharge date for each contact to a hospital.

A time period could be defined as: *2000-2012, based on dates for hospital discharges* (= hospital contacts with discharge dates in the period 2000-2012).

4. Registers, tables and data

If you request individual data, you need to specify exactly which variables you want from which registers. A few of the registers are composed of several tables. You need to specify these, as well.

An example might be:

From the National Patient Register we request:

Table Variable

T_adm K_recnum

V_cpr

D_inddto

D_uddto

T_diag V_recnum

C_diag

C_diagtype

C_tildiag

T_sksopr V_recnum

C_opr

C_oprart

C_tilopr

If you want name and address (from the Danish Civil Registration System, CPR) for a certain population it is, however, sufficient to write 'name and address'.

On the homepage of the SDS research support unit, you may access an overview with links to pages on each of the registers. On these pages you may find links to register documentation and overviews of tables and variables. If in doubt of anything, please contact the person responsible for the register or the research support unit. [Information all in Danish]

5. Justification for registers, tables and variables applied for

We only deliver data that are strictly necessary to your project. Therefore, you are asked to describe how the data, you apply for, will be used to answer the project's research questions.

If you ask for names and addresses for persons you wish to contact in connection with recruitment or questionnaire surveys, you need to state how many persons you want to contact and how many respondents you expect.

Based on document at <http://sundhedsdatastyrelsen.dk/da/forskertjeneste/forskertjeneste-klasse/adgang-fsk/ansoegningsproces-fsk> (300316)

Translation by Hanne Jørgensen