SDS

Secondary data from centralized routine registry data available on remote research server

Table of contents

Chapter 1

Sundhedsdatastyrelsen (SDS)/The Danish Health Data Authority

Sundhedsdatastyrelsen - FSEID-00006014.

Download the “udtræksbeskrivelse” here:
Dataspecifikation
Dataoversigt

Chapter 2

Population

The DD2 data are uploaded to SDS, and they check the validity of the CPR-numbers before encrypting them. The dataset `DS_EXT_DD2_POP` includes all individuals with a valid CPR-number. Individuals with a non-valid CPR-number are still included in the uploaded data, e.g. in `DS_EXT_DD2CORE`, but they will have a missing encrypted CPR-number (`CPR_ENCRYPTED`) and can thus not be linked across data sources. Previously (when we had local data), we would notify DD2 (Odense) about non-valid CPR-numbers, but this is no longer possible as we are not allowed to identify the individuals based on individual-level data on the servers.

The data files uploaded by us have the prefix `DS_EXT_`. Data are available in 2 versions; the initial upload version (without suffix) and the updated versions (with suffix _230601, _231026, etc., corresponding to the date of the update on SDS). The registry data (views) will always be based on the newest population, i.e., individuals in the newest version of `DS_EXT_DD2_POP`.

Chapter 3

Rawdata available on SDS

Data on SDS are based on data views, i.e., we have access to data updated (almost) on a day-to-day basis. The update file (SAS `IN06014.OpdateringsOversigt`) contains the time of the current update for each dataset in the library. The metadata file (SAS `metadata.metadata_allcolumns`) contains information about variables in the data. For research projects, in order to be able to reproduce results, remember to save a copy of the research data in a suitable project folder on SDS.

> **Note**
>
> Using SAS, if you extract data from views you will likely save a lot of time by writing code that is proper sql, rather than using SAS-specific syntax

```
Example:
1)

proc sql;
  create table x as
  select *
  from data.lab_lab_dm_forsker(where = (analysiscode in: ([list of npucodes])));
quit; * takes ~40 minutes to run

2)
proc sql;
  create table x as
  select *
  from data.lab_lab_dm_forsker
```

```
  where analysiscode eqt [npucode1] or ... or analysiscode eqt [npucodeN];
quit; * takes ~40 minutes to run


3)
proc sql;
  create table x as
  select *
  from data.lab_lab_dm_forsker
  where analysiscode like [npucode1%] or ... or analysiscode like [npucodeN%];
quit; * takes ~10-20 seconds to run
```

Further documentation on the main data sources can be found here:

- CPR-registret (CPR): documentation
  References: Schmidt et al. (2014)
- Landspatientregistret (LPR2 and LPR3): documentation, LPR3, LPR3 vejledning
  References: Schmidt et al. (2015)
- Lægemiddelstatistikregistret (LRS/LMS): documentation
  References: Kildemoes et al. (2011), Johannesdottir et al. (2012), Pottegård et al. (2017), Ehrenstein et al. (2010)
- Laboratoriedatabasen (LAB-F/RLRR): documentation
  References: Grann et al. (2011), Arendt et al. (2020)

3.1 Access to data

All DD2-researchers with access to FSEID-00006014 can (in principle) work with all the datasets. As these datasets are considered rawdata, they are cannot be edited/changed/moved/renamed/etc. To access data, please do the following:

- **SAS**: The data are already available in the libraries `in06014` and `metadata`
- **STATA**: Run the (relevant parts of the) DO file located at `F:\Projekter\FSEID00006014\DB Connections FSED00006014.do`
- **R**: Run the (relevant parts of the) R file located at `F:\Projekter\FSEID00006014\DB Connections FSED00006014.R`

In each research project, it is the individual researchers' responsibility to ensure data usage is allowed.

Chapter 4

Data uploads

The first data were uploaded in 2022 when the server access was initially established. Since then, the data have been updated multiple times.

Short description	Detailed description	Date	Affected datasets
Original upload	Upload of datasets for the initial server access	Summer 2022	DD2CORE, DD2_POP, BIOBANK, DDDA, FOED-SELSDATA, IDNC, IL6_TNF

Short description	Detailed description	Date	Affected datasets
General update	Population update and inclusion of data source FODSTATUS. The biobank was updated with additional measurements of CRP, c-peptide, glucose, and HOMA (from April 2022), and additional "plade-biomarkører" were added to IL6_TNF. DDDA data were updated (June 2022) before the registry closed. IDNC was trimmed to the current population.	June 2023 (_230621)	DD2CORE, DD2_POP, BIOBANK, DDDA, FOED-SELSDATA, IDNC, IL6_TNF, FODSTATUS
Upload of biobank data	C-peptide and glucose (and therefore also HOMA) were changed, and the remaining "plade-biomarkører" were added to IL6_TNF, which is now complete	Fall 2023 (_231026)	BIOBANK, IL6_TNF

Short description	Detailed description	Date	Affected datasets
Upload of biobank data (genetics)	Genetic variables (polygenic risk scores) were added to IL6_TNF (long format). All new variables are labelled "Gen1"	April 2024 (_240429)	IL6_TNF
Upload of biobank data (genetics) and DICTA variable	Additional genetic variables were added to IL6_TNF. In DD2CORE the DICTA variable was added.	July 2024 (_240711)	DD2CORE, IL6_TNF

Short description	Detailed description	Date	Affected datasets
Population update and upload of genetics	The population (and registries) was updated (November 2023) and additional genetic variables were included. We made the SDS update in the same round where the DST data were updated (genindstilling via DDV), and decided to rename IL6_TNF to BIOMARK including only the "plade-biomarkører", and including the genetic variables in a separate dataset called GENETIK. The dataset INFODATA including e.g., DICTA was also included.	December 2024 (upload not available as we are waiting for SDS updates)	DD2CORE, DD2_POP, BIOMARK, GENETIK, INFODATA
December 2024 upload	The upload from December 2024 was made available and we added CAR and DAR from the "omlægning af registre" (still waiting for LAB).	June 2025	DD2CORE, DD2_POP, BIOMARK, GENETIK, INFODATA (and CAR+DAR from registries)

NB: we try to distinguish between “update” and “upload”. An “update” means that the population, and therefore also the registry data, has changed, and that the entire DD2 cohort should be seen as a new version. By “upload” we mean additional data uploaded to the current population. This could for example be adding analyses from the biobank or adding a new data source, i.e., something that doesn't affect the population itself. The plan is to make one annual update (November/end of the year) and two to three additional uploads during per year.

Arendt JFH, Hansen AT, Ladefoged SA, Sørensen HT, Pedersen L, Adelborg K. Existing data sources in clinical epidemiology: Laboratory information system databases in denmark. Clin Epidemiol. 2020;12:469–75.

Ehrenstein V, Antonsen S, Pedersen L. Existing data sources for clinical epidemiology: Aarhus university prescription database. Clin Epidemiol. 2010;2:273–9.

Grann AF, Erichsen R, Nielsen AG, Froslev T, Thomsen RW. Existing data sources for clinical epidemiology: The clinical laboratory information system (LABKA) research database at aarhus university, denmark. Clin Epidemiol. 2011;3:133–8.

Johannesdottir SA, Horvath-Puho E, Ehrenstein V, Schmidt M, Pedersen L, Sorensen HT. Existing data sources for clinical epidemiology: The danish national database of reimbursed prescriptions. Clin Epidemiol. 2012;4:303–13.

Kildemoes HW, Sørensen HT, Hallas J. The danish national prescription registry. Scand J Public Health. 2011;39(7 Suppl):38–41.

Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data resource profile: The danish national prescription registry. Int J Epidemiol. 2017;46(3):798–798f.

Schmidt M, Pedersen L, Sørensen HT. The danish civil registration system as a tool in epidemiology. Eur J Epidemiol. 2014;29(8):541–9.

Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The danish national patient registry: A review of content, data quality, and research potential. Clin Epidemiol. 2015;7:449–90.