

Birth data

Secondary data from non-centralized data source

Table of contents

1	Control population	3
2	Data documentation	5
2.1	foedselsdata.sas7bdat	5

Go to Data documentation

Data from Danish midwife journals have been collected for all individuals enrolled in DD2 per 22 February 2022, *i.e.*, a total of $n=9,549$ individuals. Data include information on birth weight, birth length, born at term (yes/no, and if no then also the number of weeks), and twin (yes/no).

The afleverings-/afslutningsrapport from Rigsarkivet (the Danish National Archive) can be downloaded here (Danish):

There is similar description in English in the electronic supplemental material from Hansen et al. (2023):

CPR-numbers for a total of $n=9,549$ individuals were sent to Rigsarkivet and $n=8,896$ fulfilled the criteria for potentially having available birth information (born in the period from ~1920 to 1988, born in Denmark, and identifiable through the biological mother's name). During the retrieval, data were uniformized (*e.g.*, to include birth weight in grams and not pounds) and proofread.

Data include $n=9,549$ CPR-numbers (one CPR-number is included twice), $n=8,346$ have non-missing birth weight, and a total of $n=8,364$ had a non-missing value in at least one of the five birth variables. A total of $8,869-8,364=505$ were not included despite fulfilling the criteria (not registered¹).

¹It is not marked in the data whether missing was due to not being eligible or due to other reasons

Chapter 1

Control population

In addition to the individuals from DD2, a control population was recorded. For every individual in DD2, approximately two random individuals were selected based on the births from the same midwife information sheet (with birth information on ~6–8 different births). Controls from the same midwife record thus served as a match on date of birth, midwife, and geographical location. A total of $n=18,210$ individuals are recorded in the control population and data include information on year and month of birth, birth weight, birth length, sex, born at term, twin, and geographical location (lægekreds/fødselsamt). No CPR-numbers are available for the control population (data permissions and ethical and practical reasons) and neither are links between the controls and the individuals in DD2. Data on controls are currently not available on the servers¹.

Selected pages from the opgavebeskrivelse (del 1 and del 2) can be downloaded here (Danish):

¹The current data permission includes only individuals in DD2 and therefore not the matched controls. As CPR-numbers are not available and controls ultimately cannot be linked to their respective DD2 individual there is no reason to upload data to the servers because all analyses can be performed locally.

Chapter 2

Data documentation

2.1 foedselsdata.sas7bdat

Format (var x obs)	Id variables	Unique key	Important dates
Wide (9 x 9,544)	CPR	CPR	-

The dataset include birth data for N=9,544 individuals identified via CPR. There are no dates in the dataset, however, information relates to birth date.

Table 2.2: Illustration of the overall data structure. The dataset is in wide format (9 variables x 9,544 observations), with CPR as the unique key.

Row	CPR	foedselsvaegt	foedselslaengde	fuldblest	baaren	huger_for_tid	tillyngfoedsel
1	CPR1	num.	num.	Ja			0
2	CPR2	num.	num.	Nej	2 til 3		0
3	CPR3	num.	num.	Ja			0
...
9,544	CPR9544	num.	num.	Ja			1

Hansen AL, Thomsen RW, Brøns C, Svane HML, Jensen RT, Andersen MK, et al. Birthweight is associated with clinical characteristics in people with recently diagnosed type 2 diabetes. Diabetologia. 2023;66(9):1680–92.

