The Full Reproductive History of all Women in the Danish National Birth Cohort

This document describes the procedure used to identify and document all pregnancies of women in the Danish National Birth Cohort (DNBC).

Two data files are generated, one in a long format and one in a wide format.

- **"DNBC_Full_reproductive_history_long_20170113.dta"**
  In the long format each pregnancy has its own line. The variable "event" presents the events for each woman (lbnr)

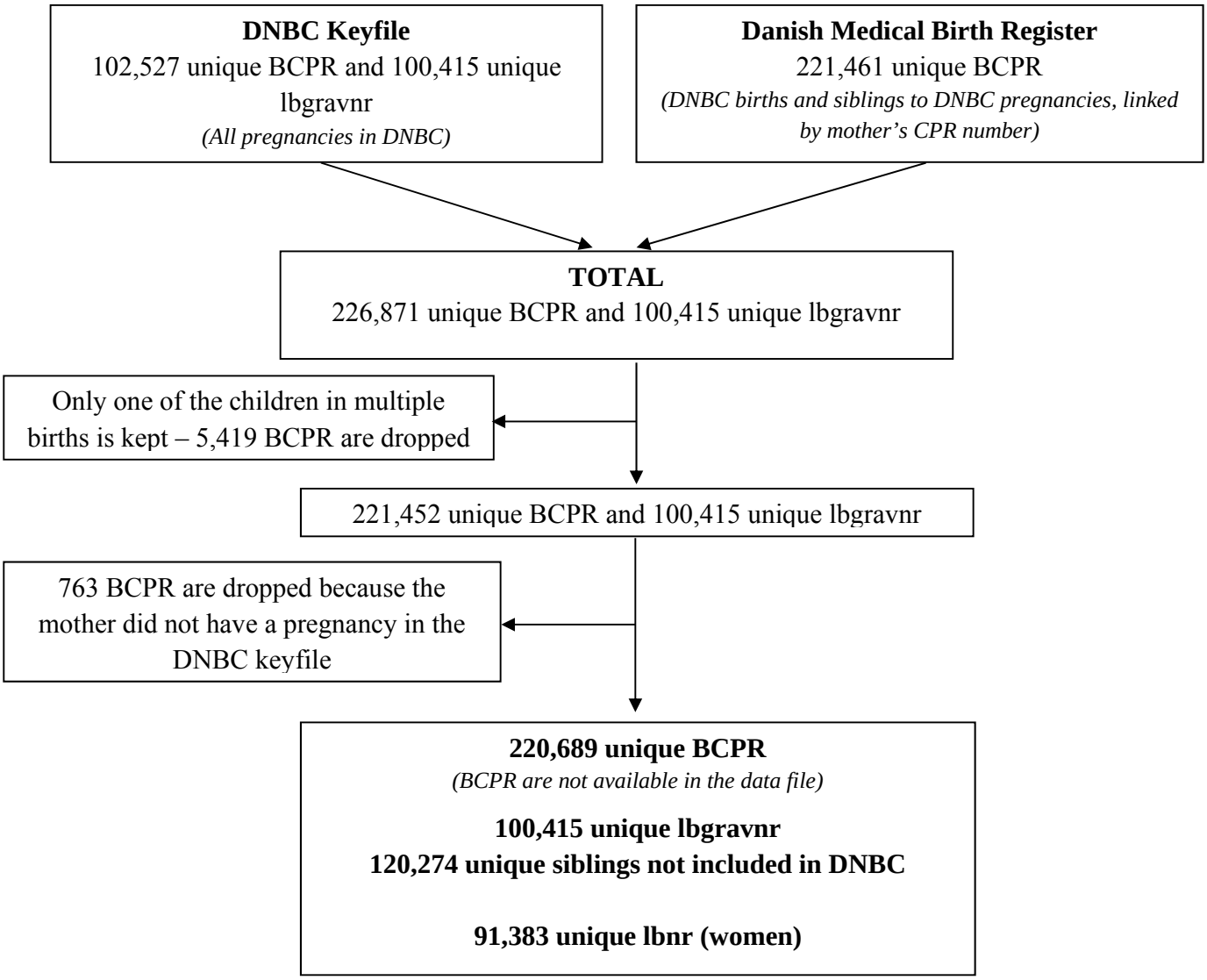
- **"DNBC_Full_reproductive_history_wide_20170113.dta"**
  The wide data set was reshaped, so each unique DNBC pregnancy (lbgravnr) has its own line. Information regarding the unique DNBC pregnancy has the index number 100, and an increasing number define the mother's subsequent pregnancies whereas a decreasing numbers define the mother's previous pregnancies

The data file contains 100,415 unique lbgravnr (pregnancies in DNBC). For all twins, triplets, and quadruplets only one of the children is kept in the data file.
The data file contains 91,383 unique lbnr equal to 91,383 women.
The outcomes for the 100,415 unique lbgravnr were as follow:

100 outcom_f	Freq.	Percent	Cum.
live birth, singleton	92,668	92.29	92.29
spontaneous abortion	4,734	4.71	97.00
stillbirth, singleton	303	0.30	97.30
late ind.abort.fetal disease	284	0.28	97.58
mola hydatidosa	67	0.07	97.65
extraut. pregnancy	48	0.05	97.70
early ind.abortion	118	0.12	97.82
late ind.abort, mat.disease	44	0.04	97.86
twin birth, both stillborn	3	0.00	97.86
twin birth, both live born	2,024	2.02	99.88
unknown	16	0.02	99.89
twin birth, binary outcome	21	0.02	99.92
triplets, all alive	32	0.03	99.95
emmigrated in pregnancy	49	0.05	100.00
deceased in pregnancy	3	0.00	100.00
.	1	0.00	100.00
Total	100,415	100.00	

Cohort definition:

The definition of the cohort included in this dataset is presented in the following flowchart.



Additional information:

97 singleton livebirths, 5 singleton stillbirths, 3 twin stillbirths, 3 twin livebirths and 10 twin binary outcome from DNBC were not included in the Danish Medical Birth Registry, because

- they were not born in Denmark, or
- they were born in Denmark, but the mother was not registered in the CPR-registry at the time of birth

The Danish National Birth Registry includes only birth of women registered in the CPR-registry at the time of birth. (This information was given by Steen Rasmussen from the National Board of Health)

General information:

Birth Registry:

- Danish Medical Birth Registry (MBR):
  - All live and stillbirths (1973-2003: defined as born wk ≥28+0 of gestation; 2004-present: defined as born wk ≥22+0 of gestation)

Abortion Registries:

- Legally Induced Abortion Registry (LIAR):
  - December 1974-December 1994: all legally induced abortions (primarily before the end of week 12) were reported using a paper-based survey (first years may be less valid because some doctors were unwilling to report)
  - 1995-2013: all information was obtained from the National Patient Registry (NPR)
- National Patient Registry (NPR):
  - All spontaneous abortion were reported (1977-2003: defined as deliveries before wk 28+0 of gestation; 2004-August 2013: defined as deliveries before wk 22+0 of gestation)
  - Induced abortions (until and including 1994 primarily induced abortions after week 12 due to maternal/fetal disease, and since 1995 all induced abortions)

Pregnancy outcome definitions:

Pregnancy outcomes in the data set follow the already defined coding in the DNBC:

1 live birth, singleton
2 spontaneous abortion
4 stillbirth, singleton
5 late ind. abort. fetal disease (after week 12+6)
6 molar pregnancy
7 extrauterine pregnancy
11 early ind. abortion (before week 13+0)
12 late ind. abort, mat. disease (after wk 12+6)
15 twin birth, both stillborn
18 twin birth, both live born
20 unknown
21 twin birth, binary outcome
22 triplets, all alive
23 triplets, mixed outcome
24 quadruplets, all alive
25 emigrated in pregnancy
26 deceased in pregnancy

Strategy used with inconsistency in the registers:

Within the NPR, one reproductive event may have been registered multiple times. This was determined because of inconsistencies in gestational dates. For entries with the same outcome, the entry with the later event date was kept. For entries with different outcomes, a hierarchy was used to decide which entry to keep.

Hierarchy of which outcomes to keep:

0. Birth (both live and still)
1. Molar
2. Extrauterine
3. Spontaneous
4. Induced (DNBC outcome type 11)
5. Induced (DNBC outcome type 12)

6. Induced (with fetus disease) (DNBC outcome type 5 (ICD-10 O05.3+O05.4))
7. Other (unknown, deceased during pregnancy, emigrated)

This hierarchy is partly obtained from the paper by Kjaersgaard MI (Kjaersgaard MI1, Parner ET, Vestergaard M, Sørensen MJ, Olsen J, Christensen J, Bech BH, Pedersen LH, Prenatal antidepressant exposure and risk of spontaneous abortion - a population-based study. PLoS One. 2013 Aug 28;8(8):e72095) and further discussed with and accepted by Professor Ellen Aagaard Nøhr.

Although this was the hierarchy followed in the data cleaning process of the register data, we disregarded the hierarchy when combining the register data with the DNBC pregnancies data. Researchers had already defined the outcomes of the pregnancies included in the DNBC, cross checking registers, medical records, and personal responses to questionnaires. Therefore, we were confident that the DNBC data was more accurate than conflicting data from the registers.

Procedure for data management:

The Legally Induced Abortion Registry:

We used data from 1974 to 1995, because the abortions after 1995 were obtained from the NPR.

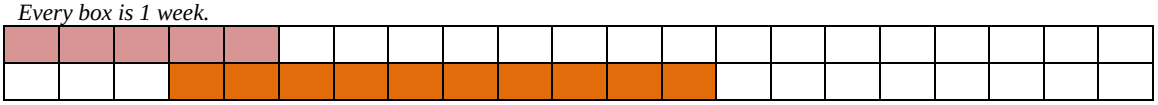
1. Type of outcome was categorized based on the variable "INDIK1" indication for abortion1 or if missing "INDIK2" indication for abortion 2. INDIK tells us which paragraph of the abortion legislation was used
    a. There were 15 entries with missing both indication numbers, so outcome was based on gestational age (outcome=12 if gestational age was more than 12 weeks, outcome=11 if gestational age was 12 weeks or less)
    b. For the entries missing both gestational age and indication number, they were coded as outcome 11 (induced before week 12) because this is the most common outcome of induced abortions
    c. 5 entries found with conflicting gestational age and indication number; gestational length was set to missing for these abortions
2. We had to check gestational length for INDIK1 and INDIK2: 52, 62, 63, 70, 71 to decide if they should be coded 12 (induced after week 12) or 11 (induced before the end of week 12)
    a. 52 is abortions before the end of week 12
    b. 62 includes both; outcomes were determined based on gestational length
    c. 63 is abortions after the end of week 12
    d. 70 not present in the data set
    e. 71 is abortions before the end of week 12
3. Date of abortion was defined based on "inddto" (indlæggelsesdato) and if missing then "ambdto" (dato for ambulant ingreb). If both of these were missing then "uddto" (udskrivningsdato) was used
4. Abortions with missing event date were dropped (16 entries)
5. Gestational age in LIAR was recorded as the week in which the abortion was conducted (not the number of full weeks). We calculated the gestational length for each pregnancy

based on the following: if gestational age was reported as 11, we knew that the abortion was conducted after the 10 full weeks but we didn't know how many days after that. We added 4 days to the full number of weeks to get the gestational age (e.g. (10*7)+4=74 days)

a. For women with missing gestational age either because variable "svlangde" was coded as unknown or there was conflicting gestational age/outcome (13 entries total), we imputed gestational age to a new variable "imputed_ge_days" using the median length from other abortions with the same outcome
    i. Induced after the end of wk 12, fetus disease: 123 days (17 weeks)
    ii. Induced before the end of wk 12: 60 days (8 weeks)
    iii. Induced after the end of wk 12, other reasons: 109 days (15 weeks)

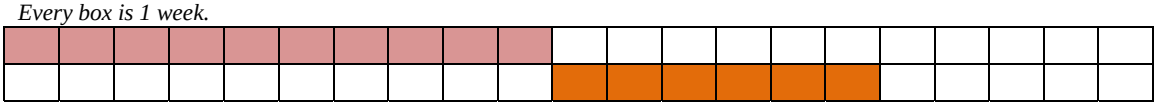
6. The first day of the last menstrual period (LMP) was calculated by subtracting the gestational age from the date of the abortion
7. For abortions that were recorded as occurring within 28 days (4 weeks), the abortion with the last abortion date was kept (see chart below for two examples)

Every box is 1 week.



Two reported pregnancies for the same woman. Gestational length is defined as the period from the first day of last menstrual period. In this example, we assumed that this was the same pregnancy and same abortion, so we kept the abortion according to the hierarchy or the last entry if outcome were the same.

Every box is 1 week.



Two reported pregnancies for the same woman. The abortion date could theoretically be the first day of the menstrual period with the next pregnancy starting right after. However, the earliest abortion date for the next pregnancy must be at least 4 weeks after last abortion date because the woman needs at least 4 weeks to observe absence of menstruation before getting another abortion. Since the second abortion here occurred more than 4 week after the first, we would assume that there were two different pregnancies, so we would keep both entries.

The National Patient Registry:

Data on all abortions was obtained from January 1977 to August 2013.

The type of outcome was defined based on the ICD-8 or ICD-10 codes.

- 2: Spontaneous abortion
    - ICD-8: 643x-644x, 645.1x, 645.4x, 645.7x, 645.9x
    - ICD-10: O02x, O03x
- 5: Induced abortions after week 12 of gestation, fetus disease
    - ICD-8: 641.3x, 641.4x
    - ICD-10: O053x, O054x
- 6: Molar
    - ICD-8: 634.29, 634.60, 634.69, 645.0x

    - ICD-10:O01x
- 7: Extrauterine pregnancy
    - ICD-8: 631x
    - ICD-10: O00x
- 11: Induced abortions before end of week 12 of gestation
    - ICD-8: 640.x
    - ICD-10: O04x, O060x, O062x, O064x, O066x, (O068x: no gestational length)
- 12: Induced abortions after week 12 of gestation, other reasons
    - ICD-8: 641.0x, 641.1x, 641.2x, 641.5x, 641.6x, 641.7x, 641.9x
    - ICD-8: 642x, 645.5x, 645.6x, 645.8x → no length
    - ICD-10: O050x, O051x, O052x, O055x, O056x, O057x, O058x, O059x, O061x, O063x, O065x, O067x
- Other:
    - O06 and O069: if gestational length was greater than or equal to 12 weeks, then it was coded as outcome=12, if no gestational length recorded, it was coded as outcome=11
    - 645.3x and 645.2x: women with this as an A diagnosis were deleted because both of these diagnoses are side effects from abortions, so they must have another entry with the actual abortion

1. We defined the outcome based on the A-diagnose type, or B if no A was available (see above for ICD-8/ICD-10 codes)
2. When pregnancy outcome was different outcome for the same hospital admission number (recnum), we used the hierarchy (see above) to decide which one to keep
3. Abortion date was defined based on hospital admission date
4. Gestational length was available in the register
    a. If different gestational lengths and outcomes were reported for the same recnum, we followed the hierarchy (see above) to decide which one to keep
    b. If the outcomes were the same for the same recnum, we kept the abortion with the later abortion date
    c. There were 2 entries in which gestational length and outcome did not correspond, so gestational length was set to missing and imputed as we did for the other women missing gestational length
5. When gestational length was missing, we imputed the median value of the available gestational lengths for the specific outcome type
    a. Induced abortion after the end of week 12, fetus disease: median length = 124 days (17 weeks)
    b. Induced abortion before the end of week 12: median length = 56 days (8 weeks)
    c. Induced abortion after the end of week 12, other reasons = 105 days (15 weeks)
    d. Molar = 82 days (11 weeks)
    e. Spontaneous = 64 days (9 weeks)
    f. Extrauterine = 49 days (7 weeks)
6. The first day of the LMP was defined based on abortion date and gestational length (if weeks were known and days were unknown we added 4 extra days, as done in the LIAR)

7. If there was overlap in the pregnancy periods or dates were less than 28 days (4 weeks) between abortion dates and outcome was the same, then we the kept the entry with the last abortion date (same procedure as for LIAR)
8. If there was overlap in the pregnancy periods or the dates were less than 28 days (4 weeks) between abortion dates and outcome was NOT the same, the hierarchy (see above) was used to decide which one to keep (same procedure as LIAR)

Combining data from LIAR, NPR and the existing data set of birth from MBR (here called DNBC/MBR):

1. The data sets from LIAR and NPR were combined and checked for same abortion registered multiple times (check by lbnr and date); if there were multiple entries the following procedure were used:
    a. If the same abortion was reported as induced in both NPR and LIAR: we kept information from LIAR because abortions were reported in this register after they were conducted and the LIAR was the main register for reporting induced abortions during this time period
    b. If the same abortion was reported as induced in LIAR and spontaneous in NPR: we followed the predefined hierarchy (see above)
2. Gestational age was imputed as a new variable in the DNBC/MBR data set if missing based on the same process as for the LIAR/NPR:
    a. Induced abortion after the end of week 12, fetus disease: median length = 123 days (17 weeks)
    b. Induced abortion before the end of week 12: median length = 56 days (8 weeks)
    c. Induced abortion after the end of week 12, other reasons = 105 days (15 weeks)
    d. Molar = 82 days (11 weeks)
    e. Spontaneous = 64 days (9 weeks)
    f. Extrauterine = 49 days (7 weeks)
    g. Singleton = 280 days (40 weeks)
    h. Twins = 261 days (37 weeks)
    i. Triplets = 238 days (34 weeks)
3. The first day of the LMP for all births and DNBC abortions was generated by subtracting the date of the birth or abortion by the gestational age
4. When combining the LIAR/NPR abortion data set with the previously made data set including all DNBC pregnancies and all births before and after the DNBC pregnancies from the MBR (DNBC/MBR) some pregnancies were reported in both LIAR/NPR and DNBC/MBR data sets and the entries were kept based on the hierarchy if they came from the same dataset, but we maintained the outcomes from the DNBC pregnancies regardless of the established hierarchy (see described above)
5. Then sequential numbers were given in the combined data set for:
    a. All reproductive events
    b. Gravidity, not including molar/extrauterine pregnancies (total pregnancies)
        i. Women who had only molar/extrauterine pregnancies (n=7) are labeled as gravidity=0
    c. Parity (total births, both live and stillborn; not including the 3 “other” outcomes)

   d. Spontaneous + induced abortions
   e. Spontaneous abortions
   f. Induced abortions
   g. Molar pregnancies
   h. Extrauterine pregnancies
   i. Other outcomes (unknown, emigrated during pregnancy, deceased in pregnancy)
6. The total number of all of the above listed event types for each woman was estimated
   a. The emigrated outcome was not counted in parity or abortion numbers, but it was included in gravidity because we were not sure the outcome of the pregnancy if the woman emigrated

Description of the variables in the data sets of the women´s full reproductive history:

Variable Name	Description
lbnr	DNBC ID (woman id number)
eventda	date of event (abortion or birth)
pregdat	date of conception
outcom_f	outcome of pregnancy (see all possible outcomes above)
temp_outcom_f	hierarchy of outcomes
ge_days	gestational length (in days)
event	specific reproductive event number
gravidity	specific pregnancy number (excluding molar/extrauterine pregnancies)
parity	number of births
stillbirth	number of stillbirths
abortions	specific abortion number (spontaneous and induced)
s_abort	specific spontaneous abortion number
i_abort	specific induced abortion number
m_preg	specific molar pregnancy number
e_preg	specific extrauterine pregnancy number
other	specific "other" outcome number (emigrated, deceased, unknown)
lbgravnr	DNBC pregnancy identification number
lbgravf	
lbgravff	
eventdamis	Eventdate missing in keyfile put pullout from another form (gravtabinfo). Valuelabels indicate which date that was used
deathage	age of child death (in days)
c_section	c-section: yes or no
length	length of child (in cm)
sex	sex of child
birthweight	weight of child (in grams)
smoking	maternal smoking habits
m_height	maternal height (in cm), available from mid-2004
m_weight	maternal weight (in kg), available from mid-2004
bmi	maternal BMI, available from mid-2004
m_birthday	maternal birth date
m_age	maternal age at date of event
gest_wk	full weeks gestation
imputed_pregdat	imputed date of conception
imputed_ge_days	gestational length (in days, imputed), indicates which lengths were imputed
eventnr	total number of reproductive events
graviditynr	total number of pregnancies
paritynr	total number of births
stillbirthnr	Total number of stillbirths
abortnr	total number of abortions
s_abortnr	total number of spontaneous abortions
i_abortnr	total number of induced abortions
m_pregnr	total number of molar pregnancies
e_pregnr	total number of extrauterine pregnancies
othernr	total number of "other" pregnancies (unknown, emigrated, deceased)