The Family Relations Database, **30/10/2013**

Carsten Bøcker Pedersen, Professor, DrMedSc
National Centre for Register-based Research, Aarhus University

In order to gain easier, faster, and more secure access to research data of high quality, a research database, named the Family Relations Database, has been established. It includes information on all persons registered in the Danish Civil Registration System, linked individually with data from the Medical Birth Registry, the Psychiatric Central Research Register, the National Patient Register, the In Vitro Fertilization Register, the Cancer Register, and the Register of Causes of Death. Today, The Danish Health Data Authority (DHDA) manages the Family Relations Database; and the National Centre for Register-based Research (NCRR) is data processor for DHDA. Data sources at the DHDA form the basis of information in the Family Relations Database. However, the contents of the Family Relations Database provide researchers with access to health-related data of a higher quality and data sources that are streamlined for longitudinal population-based studies. This is the case for all data sources in the Family Relations Database. E. g.:

1) The Danish Civil Registration System (CRS) contains information on mothers, fathers and siblings, if any. Beginning in 1968, the CRS established links to parents by means of families residing at the same address. Today, parental links for all persons born in Denmark 1969 or later are considered to be 100% correct. As some links to family members are based on shared address, inconsistencies exist[1]. Only parental links that are deemed valid are included in the Family Relations Database. Today, all persons born in Denmark 1960 or later have complete link to mother and father. In combination with information recorded in the local church books, it is possible to extend the completeness of information on parental links to one more generation.
2) The Danish Civil Registration System contains information on permanent residence including emigrations, immigrations, and disappearances. This is recorded in 9 different data sources with both missing and inconsistent information, where persons may be registered to reside at a specific address in Denmark while also being registered as living abroad. The residential information in the Family Relations Database contains longitudinal information on residence for all Danish residents who have been living in Denmark at least one day since 1968, accounting for these issues.
3) The National Patient Register contains an administrative database of patient-level information, a database describing diagnoses given at somatic in- and outpatient departments, and procedures performed at somatic in- and outpatients departments. This information is contained in one database per year, yielding so far 35 administrative databases, 35 diagnostic databases, and 35+ databases on procedures performed. In addition, there are some changes in information included, as well as inconsistencies due to varying variable lengths between years. In the Family Relations Database, the National Patient Register has been transformed from these 100+ datasets into one administrative database covering the whole period, one diagnostic database covering the whole period, and one database on procedures covering the whole period. In all instances, the minor inconsistencies in information between years have been corrected. Also, information from DUSAS is included in these 3 databases.

4) The Medical Birth Registry contains information on all births in Denmark since 1973, where information on, e.g., birth weight and gestational age has been recorded in different variables and using different classifications since 1973. The Medical Birth Registry, included in the Family Relations Database, corrects for inconsistent recording of birth weight, gestational age, birth length, and Apgar score.
5) The Danish Register of Causes of Death contains information on date and cause of death for all persons who died in Denmark since 1970.There are minor inconsistencies when comparing date of death in the Cause of Death Register with date of death in the Civil Registration System. The version of the Cause of Death Register contained in the Family Relations Database corrects for these inconsistencies.
6) Due to errors in assignment of CPR numbers, most prevalent in the initial operation of the Danish Civil Registration System, 110,000 persons have both a current and a historical CPR-number, and such persons may be registered in all national registers using both the current and the historical CPR number. One premise for high-quality register-based research is accurate and reliable individual longitudinal data. Therefore, a scrambled unique personal identifier replaces current and historical CPR numbers for the persons accounting for potential changes of CPR-numbers. This procedure is imposed on all registers included in the Family Relations Database. In addition, in order to increase the quality of the longitudinal data, it also enables researchers to have access to individually linked data without revealing the CPR number.

More information is provided in the detailed documentation of the Family Relations Database.

The development of the above-mentioned data administration and quality assurance procedures roughly corresponds to 5 man-years, and the implementation of these tasks corresponds to a half man-year, per year.

Reference List

(1) Pedersen CB, Gotzsche H, Moller JO, Mortensen PB. The Danish Civil Registration System. A cohort of eight million persons. *Dan Med Bull* 2006;53:441-449.

FIGURE 1: THE FAMILY RELATIONS DATABASE