
Scania Metadatabase for Epidemiology

The aim of the Scania Metadatabase for Epidemiology (SME) was to invent and create a catalogue that describes epidemiological studies collected within Lund University and Region Skåne. The project was initiated by Kristina Jakobsson (Division of Occupational and Environmental Medicine, Region Skåne, Lund), Jonas Björk (assistant coordinator of the national, interdisciplinary research project SIMSAM Lund) and Peter Nilsson (coordinator of SFO EpiHealth). The study inventory have been conducted by the R&D-centre in Scania and BBMRI.se in close cooperation with the Swedish National Data Service (SND). SME was initiated in the autumn of 2012 and it currently comprises information about 50 studies, including a total of just over 680,000 study participants as well as biological samples from more than 217 000 individuals from 21 of the studies.

Malmö-Lund, May 2015



Division of Occupational and Environmental Medicine, Lund

R&D-centre Scania

Swedish Initiative for Research on **Micro**data in the **Social And Medical** Sciences



SND Svensk Nationell Datatjänst



The recent decades numerous epidemiological studies have been implemented in the Southern healthcare region and these have resulted in databases containing a wealth of information, such as background variables that cannot be obtained from registries. These databases are a valuable resource for research. A significant part of the data that has been collected reasonably has a wide application beyond its original purpose and could therefore be used in new research projects.

The Scania Metadatabase for Epidemiology (SME) was initiated in the autumn of 2012 with the aim to create a catalogue that describes the epidemiological studies collected within Lund University and Region Skåne. Metadata means, purely linguistic, data about data or information about data and is used to describe the content and the structure for a given data collection. SME does not contain primary data or personal information, but only descriptions of the various studies and related data sets. The inclusion criteria were as follows: the research study should be conducted at Lund University or Region Skåne; it should contain collected data, through questionnaires or clinical examinations (and not only data from registers) and data should be available in electronic form and on the individual level.

In Sweden, as well as in other Nordic countries, we have a long experience of large-scale biobank- and register based research. The unique opportunity to follow up individuals in different studies using the access to population based health data registries is possible due to our personal identification numbers (PINs) that are given to all individuals in Sweden at birth or at immigration. The PIN can be used to link individuals to a variety of registries containing information about e.g. different diseases, date of birth and death and medical prescriptions. This possibility makes the information and samples, already collected in existing studies, much more valuable as more data, such as a specific cancer diagnosis, can be added to the individuals in the studies and new research questions may be answered.

The study inventory has been conducted in close cooperation with the Swedish National Data Service (SND). SND has from the Swedish Research Council been asked to be a national resource for the coordination of existing and new research databases within medicine and health, social sciences and humanities. This assignment includes the coordination of existing and new research databases and thus facilitate and develop the access to data for researchers. SND can be resembled as a shop window that displays the work and projects that are ongoing and by this create new collaborations between researchers.

The work with SME was conducted in two parts. Initially researchers involved in large population-based studies were identified. Through interviews with several senior scientists additional researchers who might have relevant data was identified, and a first list of cohorts potentially relevant for inclusion were established. The principal investigator were contacted. SND assisted with expertise and infrastructure relating to the registration of metadata. If the study followed the inclusion criteria, and the primary researcher gave their approval, metadata was registered in SND's web form. Finally, the existing database on SND's webpage was ransacked and studies that met the inclusion criteria as above were included in SME. Information that is documented include e.g. title, purpose and a general description of the study. The names of the primary researchers are listed, with e-mail information for the contact person.

At present SME comprises information about 50 studies with a total of over 680,000 study participants. Twenty-one of these studies also have sample collections, with samples from more than 217,000 of the study participants. Samples included are e.g. blood, urine and tumor tissue. Since all studies that are included in SME are tagged with the searchable keywords, EpiHealth and EpiHealth_Skåne, they are easy to find in the SND catalogue: <http://snd.gu.se/sv/catalogue>.

There are many different kinds of studies included in SME from regional health examinations and public health cohorts to occupational studies. The largest category is population-based cohorts (Table 1). The 26 population-based cohorts covers several areas; e.g. the health of the elderly population, how various environmental factors such as air pollution affects us and drug habits of the students in Scania. Several of the cohorts contain information that could contribute to increased knowledge about cancer and other common diseases such as cardiovascular diseases. Many of the cohorts are part of ongoing research projects and are therefore regularly linked to various registers.

Table 1: Different studies that are represented in the Scania Metadatabase for Epidemiology (SME)

	Number
Population-based cohorts	28
Patient-based studies	15
Occupational-based studies	4
Other studies	3
Studies with sample collections	21

Fifteen of the studies, most of them case-control studies, are focusing on patients with various diseases (Table 1). Among these are several studies investigating risk factor for various cancers such as breast cancer, lymphoma and melanoma. There are also studies following diabetes patients.

Four of the studies are occupational based studies (Table 1). Among them is a study that examines the fertility of hairdressers. There is also a study of women working within the health care or in schools. The aim of this study was to clarify the importance of physical and psychosocial work environmental factors for disorders of the muscles and joints to enable development of an intervention of ergonomic conditions that would reduce the risk. The remaining three studies fall under the category “other studies”.

The study with the most participants are “Cancer incidence and mortality in a population based investigation in the southern health care region” with information about medical prescriptions, medical expenses and tumor types from 235,000 individuals (Table 2). Thereafter follows the two diabetes studies, TEDDY and DiPiS with information about each 48,134 and 35,683 individuals.

Table 2: The five largest studies in the Scania Metadatabase for Epidemiology

Title	Number of Individuals	Sample collection
Cancer incidence and mortality in a population based investigation in the southern health care region	235,000	No
The Environmental Determinants of Diabetes in the Young – TEDDY	48,134	Yes
Diabetes Prediction in Scania – DiPiS	35,683	Yes
Malmö Preventive Medicine – MFM	33,346	Yes
Melanoma in Southern Sweden – the MISS cohort	29,508	Yes

The expectation is that SME will be used as an inspiration for new research and open up new research partnerships by visualizing the research groups and their publications. Much of the data that already have been collected can be used to answer new questions, after approval from the responsible person/ principal investigator and the usual application procedure to the ethics committee. For the scientific community the metadatabase implicates a kind of ”pension insurance” where knowledge of the studies are preserved for the future.

The aim of SME is that the database continuously will be updated with information about studies within the region. We encourage researchers to make their work visible by adding their studies to the database on SND’s website. Information on how to do this can be found on SND’s webpage under “Deposit data”. SND can be contacted on snd@gu.se for further help and advice.

SME will be the first regional of its kind in Sweden and the expectation is that it will inspire researchers in other regions to establish similar metadatabases.
