

Modern biobanks in Norway

Per Magnus¹ and Jostein Holmen²

¹Norwegian Institute of Public Health, Oslo, Norway

²HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger, Norway

E-mail: per.magnus@fhi.no

This is an open access article distributed under the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Population-based studies serve many purposes. One is screening for early signs of disease. This type of secondary prevention is usually part of the health services. Another administrative purpose is health statistics and surveillance of diseases. A third is health services research, and a fourth is analytical studies with the goal of understanding the etiology of chronic diseases. These objectives can often be combined in single studies. For all these purposes, sampling of biological samples may be performed and biobanks can be established. The intention of this paper is to discuss how different types of population-based studies with biobanks have been set up during the past decades in Norway, in order to serve the abovementioned purposes. The presentation will concentrate on large cohorts, in particular the Nord-Trøndelag Health Study (HUNT) and the Norwegian Mother and Child Cohort Study (MoBa).

SCREENING FOR EARLY DETECTION OF DISEASE

In Norway, the National Health Screening Service (NHSS) was established in 1942 as the National X-Ray Screening Service. Tuberculosis was a major contributor to the burden of disease, and detection and eradication was an important method, together with vaccination, of reducing the prevalence. For the period 1942 to 1975 almost all Norwegians took part in this screening program. Height and weight was measured at the time of screening. As a consequence, data files on height and weight for about 1.9 million Norwegians, some with repeated measures, have been stored and are available for further research through the Norwegian Institute of Public Health (NIPH) (www.fhi.no). The files have been linked to the Cause of Death Registry (1). Many of the subjects have participated in other cohorts.

In Bergen, in 1950-1951 and 1963-1964, all adult citizens were invited to blood pressure measurements, including measures of height and weight. Data files from these screening activities are available for 90,000 subjects with repeated measures for about 20,000 of them.

By the early 1970s, tuberculosis in Norway was under control, while cardiovascular morbidity and mortality reached a historically high level. The experience,

competence and infrastructure developed for tuberculosis screening was then utilized in a cardiovascular screening program. More information on the prevalence of cardiovascular risk factors was needed. To provide this information and to screen for individuals at high risk, the NHSS changed its program into a cardiovascular health screening program (2).

According to the screening program, all subjects in the selected samples received a personal invitation to the health examination. They should fill in a questionnaire and at the examination, height, weight and blood pressure was measured. Also, a blood sample for analyzing lipid levels was taken. The remaining serum was stored in a biobank – the Janus biobank – financed mainly by The Norwegian Cancer Society. Health data from the 40-42-year screening program (1985-1999) was used to monitor the cardiovascular health in Norway, but were also used for research purposes. Often, research-oriented subprojects were added to the basic data collection. More than 600,000 persons have been included in these studies, which were repeated in each county every three years. Serum from the Janus biobank has been used extensively in cancer research.

In the 1990s, it became clear that positive effects of individual screening for cardiovascular risk factors were difficult to demonstrate, and this type of screening on a population level was gradually abandoned in Norway. Other population based screening activities were initiated within the health services, such as screening for early signs of breast cancer and cancer of the uterine cervix, both performed by the Cancer Registry. The regional cohort studies that were set up in many parts of the country, were conducted using the same procedures as had been used in the screening program managed by the NHSS. In fact, the basic data collection was performed by the NHSS. The data collection consisted of a) questionnaires including information about subjective health, diseases, complaints, life style and other exposure factors, b) physical examinations including height, weight and blood pressure (subsequent studies included several other tests), and c) collection of biological samples, especially blood and urine, for adolescents also buccal smears for DNA extraction. However, each research group added on a series of subprojects to this basic data set. The subprojects could involve new questions in the questionnaires, new physical measures, or extra samples of biological material.

EARLY ANALYTICAL STUDIES

In the period 1974-78 all men and women aged 35 to 49 years, living in three mostly rural counties (Finnmark, Sogn og Fjordane and Oppland) in different parts of Norway, were invited to a cardiovascular health screening (2). The participants were re-invited to similar surveys in 1977-83 (U2) and in 1985-88 (U3) in addition to the inclusion of new birth cohorts. In total, 81,687 persons were invited to U3, and 68,645 persons participated (84%). The participants have filled in questionnaires on life styles and other exposures, have been examined with respect to height, weight and blood pressure, and have been bled for analyses of blood lipids. This 3-counties cohort is an example of what was initially a screening effort, has turned into a cohort study which can be used to shed light on the determinants of chronic diseases and to describe the clinical progression of disease. Repeated measures of exposures are also important for understanding their total effects on health.

In 1972-1973 a randomized, controlled intervention study to prevent coronary heart disease was performed on 18,000 men (3) in the city of Oslo. In 2000, repeated measures were performed for about 7,000 of them. These data files contain information on anthropometric data, results of lipid analyses and other exposures and health outcomes.

The Tromsø Health Study, covering Tromsø City, was initiated by the University of Tromsø in 1974 (4). The initial aim was to understand the factors underlying the high incidence of cardiovascular disease. Later, many other chronic conditions have been studied. Selected samples of the population have been surveyed several times during the past forty years, and the program has expanded gradually. The last survey, Tromsø VII, started in March 2015 and includes cardiovascular health, lung function, chronic pain, bone mass, diabetes, mental health, youth health, and genetics among other areas. The data is used extensively in etiological research. Blood and urine samples are stored in a biobank.

CONOR

The idea of a large unselected cohort with an associated biobank including DNA-samples was put forward to the Research Council of Norway (RCN) in 1988 by the program committee for epidemiological research (5). The idea was to study both genetic and environmental causes of rare chronic diseases. After prolonged discussions among Norwegian epidemiologists, it was agreed that the Cohort of Norway (CONOR) should include a basic set of questions, some anthropometric measures as well as a certain amount of full blood from individuals who took part in regional health studies (6). The inclusion to CONOR started with a data collection in Tromsø in 1994-5, continued with the second data collection in the county of North-Trøndelag (HUNT II) in 1994-95, and was followed up by

examinations of people living in Hordaland (HUSK) in 1997-99, Oslo (HUBRO) in 2001-02, Hedmark and Oppland in 2001-02 and Troms and Finnmark in 2002. The DNA components of the biobanks that were set up in these studies, have been centralized in Levanger as part of the HUNT biobank. Extraction of DNA has been done from the full-blood samples. The CONOR biobank as well as the biobank of the Norwegian Mother and Child Cohort Study (MoBa) has been supported first by a Research Council of Norway (RCN) grant called BIOHEALTH NORWAY and presently by BIOBANK NORWAY (www.ntnu.edu/biobanknorway).

HUNT

The Nord-Trøndelag Health Study (The HUNT Study) was first initiated by the National Institute of Public Health in 1984-1986. All citizens in Nord-Trøndelag county were invited to a health examination. The first survey (HUNT1) aimed at collecting data on cardiovascular health, diabetes, tuberculosis and quality of life. The survey was repeated in 1995-1997 (HUNT2) but now with a far more extensive program including the major public health issues (7). Blood and urine samples were collected and stored in a biobank. Since 2001, the HUNT study has been run by the Norwegian University of Science and Technology (NTNU). A third HUNT study (HUNT3) was conducted in 2006-08, using mainly the same procedures as in HUNT2, but with larger emphasis on collecting and storing blood samples according to strict procedures (www.ntnu.no/hunt) (8).

MOBA

The Norwegian Mother and Child Cohort Study (MoBa) was initiated by the Norwegian Institute of Public Health, and has the ambitious aim of discovering new causes of serious diseases (9). Inclusion of women from all over the country started in 1999. Fathers were also invited, and by 2008 data on more than 100,000 pregnancies were included. Biological samples and questionnaire data were collected from both parents in pregnancy week 17 and stored in a biobank. In addition, umbilical cord blood samples and a second sample from the mother were sampled after birth. Many exposures, such as environmental toxicants, medications, infections, diet, working conditions and genetics were collected, and a large number of disease endpoints have been assembled (www.fhi.no/moba).

ETHICS AND PERSONAL PROTECTION

The population-based health studies mentioned above are connected to the Norwegian personal identification number, and data handling is strictly regulated by law. The data files delivered to researchers are depersonalized. The studies are based on informed consent and

are approved by the Norwegian Data Inspectorate. Since the establishment of the Regional Committees for Medical and Health Research Ethics in 1986, the studies have also been approved by one of these.

FUNDING AND NATIONAL COLLABORATION

The Norwegian population studies are mainly publicly funded through support from the government, the universities, the health authorities and county councils. Some of the sub-studies are supported by pharmaceutical and other private companies. The studies are run by different institutions, but since CONOR was established in the early 1990s, there has been a close collaboration between the institutions. CONOR data is administered by a steering group with members from the collaborating institutions. Data from other surveys is regulated by each institution, and generally research data are accessible for all researchers or research institutions in Norway or abroad. The charge for accessing data may be different for each institution.

THE BIOBANKS

The data and the biological samples from a study are managed by the responsible institution, but the biological samples are centralized in three major biobanks: The Janus Serum bank in Oslo, the HUNT Biobank in Levanger, and the MoBa Biobank in Oslo.

The Janus Serum Bank

The Janus Serum Bank is a unique population-based, nationwide, research biobank containing prediagnostic biospecimens from 317,000 individuals, of whom 64,000 have developed cancer (Table 1). The serum bank is ideal for studies of novel biomarkers potentially relevant for etiology, early diagnosis, prognosis and choice of therapy. Today the Janus samples are stored at a commercial freezer department, with a constant temperature at -25°C . All sample retrievals are done manually, which is very time consuming, and the working conditions are sub-optimal. There is an urgent need to establish a new storage facility to increase biobank security, improve logistics, and ease the workflow.

The HUNT biobank

The present biobank facility was established in conjunction with the HUNT 3-survey in 2006 as a state-of-the-art biobank of $> 2000\text{ m}^2$, well equipped with storage room and laboratory facilities. Blood sampling has followed a strict quality protocol, collecting serum, plasma, buffy coat, immortalized cells for cell line production, specialized tubes for trace metal/elements-analysis, RNA-tubes and urine. Fully automated procedures have been established for i) blood fractionation and distribution into 2-D bar coded cryo-tubes, ii) DNA extraction, QC and normalization and iii) storage and retrieval of 1 million DNA samples operating at

Table 1. Serum samples in the Janus Biobank (<http://www.krefregisteret.no/no/Forskning/Janus-serumbank/Janus-materialet/>).

Sources	Men	Women	Total
Health surveys	150 779	138 310	289 089
Red Cross blood donors	17 438	14 581	32 019
The Radium Hospital	1 494	1 822	3 316
Donors in several categories	4 321	3 152	7 473
All donors	165 390	151 561	316 951

-20°C . A liquid nitrogen storage facility is also in place with a storage capacity of 1.5-3 mill cryo tubes. This is organized as archive storage with the longest perspective for preserving sample integrity and quality. Samples are also stored at -80°C in manual ultra-freezers, but funding has been provided for automated storage solutions also at -80°C with an expected capacity of 10-12 mill 0.5 ml tubes, to be installed in 2015. Additionally, from the Young-HUNT3 Study (2006-08) 7638 buccal smears are stored on FTA paper for genetic analyses.

HUNT biobank has a dedicated and highly specialized core staff of personnel preparing biological material for a large number of researchers world-wide. The biobank is using the Nautilus LIMS for cold chain logistics and management and as a sample inventory system. The LIMS is integrated with the fully automated sample fractionation and aliquoting system operating at 4°C . In total, samples from approximately 220,000 donors, collected over almost three decades, are stored as more than four million aliquots in HUNT biobank.

The HUNT biobank is also a national biobank, primarily based on samples from all the larger regional health surveys in Norway, organized as CONOR. From 2008, a complete collection of blood samples (EDTA full blood/DNA) from $\sim 250,000$ participants of the CONOR cohorts (Table 2) has been stored as a national CONOR biobank at the HUNT biobank in Levanger. DNA has been extracted and stored from more than 140,000 of these samples. All the CONOR-cohorts and the regional health trusts (clinical biobanks) have since 2011 been part of the national biobank infrastructure, Biobank Norway or BBMRI.no, which is part of the European biobank network, BBMRI.ERIC (<http://bbmri.eu/web/guest/scientific-publications>). Biobank Norway is coordinated by HUNT/NTNU and lead by professor Kristian Hveem, also serving as the leader of HUNT Biobank. Biobank Norway is supported by the Research Council of Norway.

The HUNT Research Center, the databank and biobank are all ISO-certified (NS-EN ISO 9001:2008) since April 2011 for the following services: population-based health surveys, data collection, sample handling and storage of biological material, quality assurance/quality control, documentation, analysis and delivery of biological material and analytical results for research,

Table 2. Number of participants in the major CONOR cohorts.

Year	Study	No of participants	Response rate (%)	Age groups	Web link
1984-2008	HUNT	107 000 ¹	54-89	13+	http://www.hunt.no
1975-2008	Tromsø	40 051 ²	63-74	25+	The Tromsø study⁴
1997-1999	HUSK	25 634	66	40-72	http://www.uib.no/isf/husk/
1995-1997	HUBRO	19 000	47	30-76	http://www.fhi.no
2001-2002	OPPHED	12 400	56	30-75	http://www.fhi.no
2001-2003	TROFINN	15 000	56	30-77	http://www.fhi.no
		219 085 ³			

¹27 000 participated in all 3 surveys, 50 000 in 2 of them

²15 557 have participated more in than 3 surveys

³All with available DNA/EDTA whole blood/buffy coat, and for several of them serum/plasma are also available

⁴http://uit.no/ansatte/organisasjon/artikkel?p_menu=42374&p_document_id=80172&p_dimension_id=88111

Table 3. Numbers of samples from mothers, fathers and children in MoBa as of August 2014 (as shown in more detail in reference 14).

	No of DNA samples	No of plasma samples	No of RNA samples	No of urine samples
Mothers during pregnancy	96 450	98 207		78 130
Mothers after birth	80 123	86 426		
Fathers	71 171	72 484		
Children	89 475	91 473	45 446	

documentation and storage of health data and delivery of health data for research.

DNA samples from the HUNT Biobank have been successfully used in a number of studies published in high-impact international journals (10-13). This also underlines the good quality of both the samples, accompanying data and related phenotypes, as well as the management of the sample collection.

The MoBa biobank

In contrast to the regional studies, MoBa collects data and biological samples from the whole of Norway (Table 3). By 2005, 50 out of 52 hospitals with maternity wards participated in the data collection. Women were recruited at the time of the routine ultrasound examination at about 17 weeks of gestation. Blood was sampled in EDTA tubes and sent by ordinary mail to the biobank in Oslo. Plasma was pipetted off one of the tubes at the hospital. DNA was extracted from fresh blood at arrival. In addition, urine samples were collected from the mothers. After birth, a second blood sample was taken from the mother. Also, a blood sample was taken from the umbilical cord. For the last 50,000 children, one EDTA tube and a Tempus tube for RNA processing were collected. All samples are aliquoted on to microtitre plates, deep well plates or single matrix tubes (14). A dedicated Laboratory Information Management System (LIMS) has been developed specifically for MoBa. The biobank has recently been ISO-certified. DNA is available from more than 58,000 mother-father-child trios. This is a unique resource for genetic studies. It allows studies of imprinting effects and the investigation of the importance of de novo mutations in disease etiology. As part of a

recent grant from the RCN, DNA from 11,000 of these trios have been genotyped, using a chip that includes about 250,000 single nucleotide polymorphisms (SNPs) from the whole genome and about the same number from the exome.

RESEARCH RESOURCES

A key element for research in Norway is the personal identification number, enabling linkage of data for each individual. Another key element is the opportunity to use national health registers, such as the Cause of Death Register, the Cancer Register, the Prescription Register, the Medical Birth Register and others. Additionally, in some regions there are established registers for specific diagnoses, like myocardial infarction, stroke, hip and forearm fractures, multiple sclerosis, dementia and other mental illnesses.

The combination of self-reported health data and data from physical examinations and biological samples from the cohorts linked to national and regional registers has been a main tool for epidemiological studies during the last decades. The research effort has been prominent, resulting in a large number of scientific papers and PhD theses covering a wide range of topics.

INTERNATIONAL COLLABORATION

Even though the Norwegian population studies include several hundred thousand people, they do not always provide sufficient statistical power. This is generally the case for genetic studies. Consequently, data are included in consortiums, where a number of population

based studies compile data to increase statistical power. In other situations Norwegian research groups have been collaborating with research groups in other countries on a systematic and long term basis.

FUTURE DATA COLLECTIONS

Two large data collections are presently being set up. One is already mentioned: The Tromsø VII study that will invite more than 33,000 persons in the city of Tromsø. The other is the fourth wave of HUNT that will commence in 2017. They will invite all 95,000 citizens aged 13 years and older in Nord-Trøndelag county. These studies will add to the CONOR biobank. MoBa has an ongoing data collection. In 2015, the plan is to send a new questionnaire to the fathers and to send the first questionnaire to be filled in by the children themselves, at the age of 13.

DISCUSSION

Compared to many countries Norway has several advantages in collecting epidemiological data:

- The Norwegian population is generally highly educated with no illiteracy, making self-reported data more reliable than in some other countries.
- The population based studies take advantage of long democratic traditions in Norway, with a high degree of trust between the population and the authorities. This is probably one explanation of the relatively high attendance rate.
- In Norway every citizen has a unique personal identification number, following the individual through the whole life, enabling linkage of data and follow-up even after death.
- There is a good public health service taking care of individuals with medical needs.
- There are several institutions with long experience in conducting large population surveys, ensuring high quality data.

- Due to the long traditions with high quality population based health studies, this is an arena where Norway can compete internationally. This may be one reason why these studies are supported by politicians from all parties.

The profile of the studies has changed through the years, from screening and intervention in the 1970s and early 1980s to research in the last decades. The analyses are concentrated around the main public health issues, like cardiovascular diseases, diabetes, cancer, mental health, osteoporosis, lung diseases and musculoskeletal complaints, and are therefore relevant to disease prevention and health promotion. However, even if the intention is not to detect unknown diseases, the procedures will nevertheless identify some individuals who need more thorough examination. This provokes medical and ethical dilemmas, like the definition of who needs follow-up and what kind of information should the participants have. These questions have been discussed for decades, and will be discussed also in the future.

A main challenge is the declining attendance rate in recent surveys (15). If the Norwegian data- and biobanks shall be relevant even in the future, every effort should be made to maintain the high attendance rate, which combined with high quality data is the main advantage of Norway data- and biobanks compared to population based data in other countries.

Most of what is known about the etiology of complex, chronic diseases comes from the study of large prospective population studies with biobanks. Maintaining the high quality of the cohorts and the biobanks is costly. To secure funding is a constant struggle. The cohorts form a basis for international cooperation and funding. They should be developed further by being open to new ideas and avenues in modern epidemiological research, and by seeking new collaborations. Let us hope for a safe future for these valuable resources.

REFERENCES

1. Waaler HT. Height, weight and mortality. The Norwegian experience. *Acta Med Scand Suppl* 1984; **679**: 1-56.
2. Bjartveit K, Foss OP, Gjervig T, Lund-Larsen PG. The cardiovascular disease study in Norwegian counties. Background and organization. *Acta Med Scand Suppl* 1979; **634**: 1-70.
3. Hjermmann I, Velve BK, Holme I, Leren P. Effect of diet and smoking intervention on the incidence of coronary heart disease. Report from the Oslo Study Group of a randomised trial in healthy men. *Lancet* 1981; **2** (8259): 1303-10.
4. Arnesen E. Statens helseundersøkelser og Tromsø. *Norsk Epidemiologi* 2003; **13**: 33-44.
5. Committee report to the Council for Medical Research. The research need for the establishment of a DNA-bank in Norway. Oslo: Research Council of Norway, 1988.
6. Magnus P, Arnesen E, Holmen J, Stoltenberg, Søgaard A, Tell G. CONOR – COhort NORway: Historie, formål, potensiale. *Norsk Epidemiologi* 2003; **13**: 79-82.
7. Holmen J, Midthjell K, Krüger O, Langhammer A, Holmen T, Bratberg G, et al. The Nord-Trøndelag Health Study 1995-97 (HUNT 2): Objectives, contents, methods and participation. *Norsk Epidemiologi* 2003; **13**: 19-32.
8. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, et al. Cohort Profile: the HUNT Study, Norway. *Int J Epidemiol* 2013; **42**: 968-77.

9. Magnus P, Stoltenberg C. The Norwegian Mother and Child Cohort Study (MoBa) – past, present and future. *Norsk Epidemiologi* 2014; **24**: 3-6.
10. Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, Zhou W, et al. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat Genet* 2014; **46**: 345-51.
11. Strawbridge RJ, Dupuis J, Prokopenko I, Barker A, Ahlqvist E, Rybin D, et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* 2011; **60**: 2624-34.
12. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42**: 937-48.
13. Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, Palmer CD, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet* 2011; **7**: e1001324.
14. Paltiel L, Haugan A, Skjerden T, Harbak K, Bækken S, Stensrud N, et al. The biobank of the Norwegian Mother and Child Cohort Study – present status. *Norsk Epidemiologi* 2014; **24**: 29-35.
15. Langhammer A, Krokstad S, Romundstad P, Heggland J, Holmen J. The HUNT study: participation is associated with survival and depends on socioeconomic status, diseases and symptoms. *BMC Med Res Methodol* 2012; **12**: 143.