

The Norwegian Mother and Child Cohort Study (MoBa) – new research possibilities

Per Magnus

Division of Epidemiology, Norwegian Institute of Public Health, P.O.Box 4404 Nydalen, NO-0403 Oslo, Norway

Telephone: +4723408211 Telefax: +4723408252 e-mail: per.magnus@fhi.no

ABSTRACT

The Norwegian Mother and Child Cohort Study (MoBa) is an observational, prospective study of up to 100 000 families. The study aims to find causes of diseases, with a focus on the interplay between early exposures and genetic factors. The collection of possible exposures starts in pregnancy. As of September 2007, about 80 000 families participate. A series of subprojects are ongoing, and analyses and publications are on their way. This paper presents the present situation of MoBa, and highlights new challenges and opportunities.

INTRODUCTION

Data collected through the health care system that are made available to researchers through databases such as the Medical Birth Registry of Norway (MBRN) (1) have been of great value for resolving a number of scientific questions, including inquiries into the causes of disease. However, for many conditions and diseases, more detailed data are needed. Many hypotheses are concerned with both predisposing genes and specific environmental exposures that are not routinely measured in health care systems. Consequently, researchers in many countries have set up cohort studies that collect detailed information from clinical examinations, questionnaires and interviews, and that include the collection of several types of biological materials at various points in time. In the 1980s, the importance of the intrauterine environment and conditions in early life was highlighted, partly as a consequence of observations made by Forsdahl (2) in Norway and Barker (3) in England, leading to the fetal origins hypothesis (4). In Norway, a number of researchers in perinatal epidemiology agreed in the 1990s to embark upon an ambitious pregnancy cohort, called the Norwegian Mother and Child Cohort Study (MoBa) (5).

During the last few years, the rapid development in high throughput technology in molecular medicine have increased the scope of what is possible to measure and include in epidemiological investigations. An enormous amount of genes in a single person can now be genotyped, albeit at a relatively high cost. Also, new methods in the detection and quantification of expressed genes, proteins and metabolites have paved the way for discovery research that pose statistical and interpretational challenges not previously encountered by epidemiologists.

The aim of this paper is to describe the present state of affairs as regards data collection and subprojects in MoBa and to discuss some new research challenges.

THE DATA COLLECTION

The first unit of observation in MoBa is the pregnancy. Pregnant women are approached by mail and asked to participate before they attend the routine ultrasound

screening at the local hospital, usually in week 15-18. They can also be informed about the study when they attend the ultrasound laboratory at the hospital. If they are willing to participate they have to sign a consent form. Routinely, they are asked to invite their partner, who has to sign a separate consent form if he decides to take part in the study. As of June 30, 2007, 222,484 pregnant women have been invited and 89,594 have agreed to participate, yielding a recruitment rate of 40.3%. As has been described previously (5), the response is lower for women who are invited in a second pregnancy, and differs according to their participation in the first pregnancy.

Recruitment started in 1999 in a single hospital, and has, since January 2006, included 50 out of 52 hospitals with maternity units. The long recruitment period increases the variability in intake of nutritional factors, infections and other exposures among participants. It also opens up for family studies as more pairs of siblings are included. This enhances research since family studies provide the possibility of estimating recurrence risks and correlations between relatives for many traits and diseases, and it opens for better studies of the associations between diseases and gene variants. In addition, effects of exposures that differ between pregnancies to the same woman can be estimated.

After giving consent, the pregnant woman is expected to fill in three questionnaires during pregnancy, one (Q1) at the time of recruitment, concerning general health, exposures and background data, a second (Q2) at week 22 and a third (Q3) at week 30. Q2 is a food frequency questionnaire (6), while Q3 covers health conditions during pregnancy. After birth, three questionnaires are sent: when the child is 6 months (Q4), 18 months (Q5) and 3 years (Q6). These questionnaires are concerned both with maternal and child health and with new exposures.

Blood and urine samples are donated by the participating women at the time of recruitment, and a second blood sample is given after birth. At birth, a blood sample is also taken from the umbilical cord. Details of the content and processing of these samples have been given elsewhere (7).

The partner fills in a relatively short questionnaire on history of diseases, life styles, work exposures and

other relevant background data. He is also expected to donate a blood sample, usually at the time of the ultrasound examination. The fathers constitute a cohort in themselves, just like the mothers and the children can be thought of as separate cohorts, to be followed later with respect to many outcomes, and the data provide an important source for genetic studies. In addition, the possible effects of various exposures on gametic mutations are important to study, as are all other effects on child health that can be attributed to paternal influences.

The percentages of participants who fill in and return questionnaires are shown in Table 1. It should be kept in mind that the data collection is ongoing and most families have not been sent questionnaires 5 and 6. Presently, the oldest children in the cohort are reaching the age of seven. In the autumn of 2007, a four-page questionnaire on somatic diseases in the child and some exposures will be mailed to these families. Later, we plan to send questionnaires that focus on paternal health, maternal health and the child's development and mental health. New data collection initiatives will be made when the children are adolescents. Also, new approaches to data collection through web-based questionnaires and through cell phone reminding, are being implemented.

As is evident from Table 1, we do not receive biological samples from all participants. If combinations of samples are required, for instance DNA from the mother, father and child for genotyping in the case-parent design, only around 60% of families will have that combination. Furthermore, at present, one third of DNA samples are unavailable due to the reduced capacity in the biobank to process samples. Consequently, only about 20,000 triads are available for analysis today. This fact limits MoBa's flexibility to rapidly respond to important research questions.

This autumn (2007), decisions will be made on how long the cohort will recruit new participants. In the original protocol, it was stated that 100,000 women

should be recruited. This statement can be interpreted as 100,000 pregnancies or as 100,000 different women, each participating with 1 or more pregnancies. At present, about 90,000 pregnancies from about 80,000 different mothers have been included. If the decision is the latter alternative, MoBa will be the most comprehensive of the international pregnancy cohorts. However, regardless of the choice, for many research purposes, collaboration with these other cohorts will be needed to provide precise results.

SOME EARLY RESULTS

Questionnaire data are quality-controlled and entered into a large database. As they are made available, MoBa has delivered datasets to researchers after contracts have been made according to the guidelines for access to data (www.fhi.no/morogbarn). The latest delivery (version 3) was given out in April 2007. It includes, among other elements, 67,355 records with information from Q1 and 63,182 records from the MBRN.

Thus, analytic results are beginning to appear and include the observation that paternal BMI is related to subfertility (8). Others have studied the relation between temperament and weight gain in the child (9), the effect of prenatal and postnatal smoking on respiratory health of the child (10), and the effect of affectivity on breastfeeding (11). Also, studies have estimated the validity of the food frequency questionnaire, Q2 (12-14), and of methods used in the biobank (15). In addition, behaviours and disorders in the mother are being analysed (16-18). A series of other papers have been submitted or are in preparation.

MANY SUBPROJECTS

A large number of subprojects are established. The electronic project database maintained at the Department of Research Data in the Norwegian Institute of Public Health has recorded 121 MoBa subprojects as of September 2007.

MoBa has collaborated with The National Institute of Environmental Health Sciences in the U.S. for many years, and in 2007 this collaboration was extended for another 10 years. An agreement has been made that extra blood and urine should be collected from all participating women at the time of recruitment. These samples are dedicated to the study of environmental toxicants, and will be used in scientific collaborations between U.S. and Norwegian scientists.

There has also for several years been a joint project with researchers at Columbia University in New York, sponsored by the National Institute of Neurological Disorders and Stroke. The aim is to study how genes and environmental factors interact to cause autism-spectrum disorders, and to describe the early trajectories of these disorders. For these purposes a subcohort of children who are clinically examined at the age of three years is established, partly consisting of control children and partly of children where responses to the Q6 has been above a certain threshold on a scale on early signs of autism (communication, social skills,

Table 1. Numbers of responses and response rates to questionnaires and numbers and proportions of participants in MoBa who have donated biological samples as of June 30, 2007.

Data item	N	% of participants*
Q1: pregnancy week 15-20	85,176	95.1
Q2: week 22	80,543	92.7
Q3: week 30	76,255	91.4
Q4: 6 months	61,841	85.8
Q5: 18 months	40,930	74.8
Q6: 3 years	17,501	62.0
Paternal questionnaire	64,679	94.5
Maternal blood sample, week 17**	79,231	88.4
Maternal blood sample, after birth	66,381	79.3
Child, umbilical cord blood	70,748	84.5
Paternal blood sample	56,026	79.3

* The denominator (the number of subjects who have been asked to respond) differs across data items due to the continuing follow-up

** includes a urine sample for most subjects

repetitive behaviour). In addition, MoBa children can be referred to this examination from the child-psychiatric units around Norway.

A new subproject that is aimed at understanding development and causes of attention deficit/hyperactivity disorder (ADHD) is starting in the autumn of 2007, much along the same lines as the previous project.

Since 2002, MoBa has formed one half of BIOHEALTH NORWAY, which is a human biobank technology platform in collaboration between Norwegian research institutions aimed at making population-based biobanks better equipped to perform genetic studies. The support comes from The Research Council through the FUGE (functional genomics) program. The other large biobank belongs to CONOR, a large cohort study of adults (19). FUGE also supports specific research projects that utilize the MoBa biobank and database.

MoBa participates in two integrated projects in EU's 6th framework program for research. EARNEST (www.metabolic-programming.org) is a large consortium that studies early nutrition in relation to later diseases, and both the other two large pregnancy cohorts in Europe, ALSPAC (20) in England and the Danish Birth Cohort (21) are partners. The other integrated project is NewGeneris (www.newgeneris.org) which investigates the role of prenatal and early-life exposure to genotoxic chemicals present in food and the environment in the development of childhood cancer and immune disorders.

NEW CHALLENGES

There are no exclusion criteria for participants in MoBa and the cohort is set up to answer all possible research questions that deal with causality, given the limitations in the data set. One limitation is that cell lines from the participants have not been provided. Also, the questionnaires are of limited size, and many compromises were made when the content was determined. Another limitation is linked to the living conditions and exposures prevailing in the population of Norway in this particular period of time.

There are several possibilities that have not been exploited yet. One is the creative use of population registries, such as the Central Population Registry, The Medical Birth Registry and the National Twin Registry. From these registries informative family constellations with supplementary data can be constructed and linked to the MoBa database, opening new avenues in life course and intergenerational research. There are several exposure registries that can be used for linkage, such as the water-work registry, education registry and other databases on socioeconomic factors.

Also, no linkage has yet been made to disease registries, such as the Cancer Registry, the Cause of Death Registry or the Registry for Type 1 Diabetes. Linkage to the Prescription Registry will allow for studies of the influence of medications as exposures, but will also, for some disorders, serve as a source of diagnostic information. In February 2007, Norwegian authorities decided to launch a person-identifiable National Pa-

tient Registry with diagnoses for both outpatient visits and hospitalizations, and this will be of large importance for the follow-up of parents and children.

The ideology of MoBa is to divide research interests according to very narrow and specific research questions. In this way many researchers can study the same exposure or the same disorder, but the link of a specific exposure to a specific disorder is given to only one researcher or research group for a certain time period as specified in the guidelines for access to data. This approach can be thought of as hypothesis-driven research.

We are now facing the challenge of genomewide association studies, which can be described as pure discovery or agnostic research (22). This may be an advantage. Previously, genetic studies were based on candidate genes, which were genes that reasonably could be assumed to be involved in causal pathways. It is interesting that many new associations have been with genes that were not suspected of being involved with diseases, and associations have also been discovered to chromosomal regions where no genes have been described. During the summer of 2007, several case-control studies utilizing SNP-chip technology with more than 500,000 single-nucleotide polymorphisms (SNPs) have been published. This gives an enormous amount of false positive results, depending on the cut-off for what is considered to be statistically significant. It is generally agreed that replication studies are needed in order to have any faith in the reported associations. For instance, a recent publication on a new genetic association to rheumatic arthritis (RA) (23) used what they call a combined or a multi-stage case-control design, where SNPs that had an association with RA with a p-value below 1×10^{-8} in the first stage with 1522 cases and 1850 controls were retyped in an independent set of 997 cases and 1777 controls. Thus, large case-control samples will be needed with high-quality diagnostic information.

In addition to SNP-variability, DNA and RNA will be analysed for epigenetic phenomena, transcript profiling as well as the detection of insertions, deletions and variation in the number of gene copies. For autism, for instance, de novo mutations that are each very rare has added to the complexity. With its inbuilt family design, MoBa also lends itself to studies of the effects of maternal genes and their interaction with fetal genes.

Another challenge to MoBa is the use of scarce plasma resources for discoveries of biomarkers. Biomarkers can be exposures, but also early signs of disease development. There is, for instance, interest in early signs of pre-eclampsia and of autism. Discovery of new biomarkers may be very useful for understanding early mechanisms of disease development and help diagnose and manage these and other diseases at an early stage. New methods in proteomics and metabolomics give rise to a multitude of comparisons between cases and controls, and a general strategy for use of this resource is needed. In the autumn of 2007, a panel of experts in proteomics will be assembled to develop a strategy for MoBa.

In conclusion, the greatest challenge to MoBa right

now is to what extent we should use high quality datasets and biological samples to take part in pure discovery research. It is maybe wise to save most of the scarce biological material at the present time. In contrast to many large case-control samples that are

usually based on patients in clinical settings, MoBa has a wealth of longitudinally collected exposure data, and a major contribution is likely to be the study of environmental effects contingent on the presence or absence of candidate genes discovered by others.

REFERENCES

1. Irgens LM. The Medical Birth Registry of Norway. Epidemiological research and surveillance throughout 30 years. *Acta Obstet Gynecol Scand* 2000; **79**: 435-9.
2. Forsdahl A. Observations throwing light on the high mortality in the county of Finnmark. Is the high mortality today a late effect of very poor living conditions in childhood and adolescence? *Int J Epidemiol* 2002; **31**: 302-8.
3. Barker DJ, Osmond C. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet* 1986; **i**: 1077-81.
4. Barker DJ. Fetal origins of coronary heart disease. *BMJ* 1995; **311**: 171-4.
5. Magnus P, Irgens LM, Haug K, Nystad W, Skjærven R, Stoltenberg C and the MoBa study group. Cohort profile: The Norwegian Mother and Child Cohort Study. *Int J Epidemiol* 2006; **35**: 1146-50.
6. Meltzer HM, Brantsæter AL, Ydersbond TA, Alexander J, Haugen M and the MoBa Dietary Support Group. Methodological challenges when monitoring the diet of pregnant women in a large study; experiences from the Norwegian Mother and Child Cohort Study. *Mat Child Nutr* (In press).
7. Rønningen KS, Paltiel L, Meltzer HM, Nordhagen R, Lie KK, Hovengen R, Haugen M, Nystad W, Magnus P, Hoppin JA. The biobank of the Norwegian Mother and Child Cohort Study. *Eur J Epidemiol* 2006; **21**: 619-25.
8. Nguyen RHN, Wilcox AJ, Skjærven R, Baird DD. Men's body mass index and reduced fertility. *Hum Reprod* 2007; **22**: 2488-93.
9. Niegel S, Ystrom E, Vollrath ME. Is difficult temperament related to overweight and rapid early weight gain in infants? A prospective cohort study. *J Develop Behav Pediatr* (In press).
10. Håberg SE, Stigum H, Nystad W, Nafstad P. Effects of pre- and postnatal exposure to parental smoking on early childhood respiratory health. *Am J Epidemiol* 2007; **166**: 679-86.
11. Ystrom E, Niegel S, Klepp K-I, Vollrath ME. The impact of maternal negative affectivity and self-efficacy on breastfeeding: The Norwegian Mother and Child Cohort Study. *J Pediatr* (In press).
12. Brantsæter AL, Haugen M, Alexander J, Meltzer HM. Validity of a new food frequency questionnaire for pregnant women in the Norwegian Mother and Child Cohort Study (MoBa). *Mat Child Nutr* (In press).
13. Brantsæter AL, Haugen M, Hagve T-A, Aksnes L, Rasmussen SE, Alexander J, Meltzer HM. Self-reported dietary supplement use in pregnancy corresponds with biological markers – A validation study in the Norwegian Mother and Child Cohort Study. *Ann Nutr Metabol* 2007; **51**: 146-54.
14. Brantsæter AL, Haugen M, Rasmussen SE, Alexander J, Samuelson SO, Meltzer HM. Urine flavonoids and plasma carotenoids in the validation of fruit, vegetable and tea intake during pregnancy in the Norwegian Mother and Child Cohort Study (MoBa). *Publ Health Nutr* 2007; **11**: 274-83.
15. Hoppin JA, Ulmer R, Calafat AM, Barr DB, Baker SV, Meltzer HM, Rønningen KS. Impact of urine preservation methods and duration of storage on measured levels of environmental contaminants. *J Exp Sci Environ Epidemiol* 2006; **16**: 39-48.
16. Bø K, Owe KM, Nystad W. Which women do pelvic floor muscle exercises six months postpartum? *Am J Obstet Gynecol* 2007; **197**: 49.e1-49.e5.
17. Wesnes SL, Rørtveit G, Bø K, Hunskaar S. Urinary incontinence during pregnancy. *Obstet Gynecol* 2007; **109**: 922-8.
18. Bulik CM, Von Holle A, Hamer R, Berg CK, Torgersen L, Magnus P, Stoltenberg C, Siega-Riz AM, Sullivan P, Reichborn-Kjennerud T. Patterns of remission, continuation, and of broadly defined eating disorders in the Norwegian Mother and Child Cohort Study (MoBa). *Psychol Med* 2007; **10**: 1-10.
19. Magnus P, Arnesen E, Holmen J, Stoltenberg C, Søgaard AJ, Tell GS. CONOR – COhort NORway: historie, formål og potensiale. *Nor Epidemiol* 2003; **13**: 79-82.
20. Golding J, Pembrey M, Jones R and the ALSPAC study team. ALSPAC – The Avon longitudinal study of parents and children. I. Study methodology. *Paediatr Perinatal Epidemiol* 2001; **15**: 74-87.
21. Olsen J, Melbye M, Olsen SF, Sørensen TIA, Aaby P, Andersen AMN, et al. The Danish National Birth Cohort – its background, structure and aim. *Scand J Publ Health* 2001; **29**: 300-7.
22. Hunter DJ, Kraft P. Drinking from the fire hose – statistical issues in genomewide association studies. *N Engl J Med* 2007; **357**: 436-9.
23. Plenge RM, Seielstad M, Padyokov L, Lee AT, Remmers EF, Ding B, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis – a genomewide study. *N Engl J Med* 2007; **10.1056/NEJMoa073491**.