

SMP: Genetic Epidemiological Methods (GEM)**Project: Internet-based Access to KORA as Source for Population Controls and Population Genetics**

H.-Erich Wichmann, Christian Gieger, Thomas Illig - GSF – Institute of Epidemiology, Neuherberg and IBE – Chair of Epidemiology, LMU München - wichmann@gsf.de

Introduction

KORA-gen is a resource for genetic epidemiological research, based on the KORA platform (Cooperative Health Research in the Region of Augsburg). Biosamples and phenotypic characteristics as well as environmental parameters of 18,000 adults from Augsburg and the surrounding counties are available. KORA-gen can be used by the projects of the National Genome Research Network (NGFN). Interested parties can inform themselves interactively via internet about the available data and the rules of access (www.gsf.de/KORA-gen).

Basic tools to perform genetic epidemiology are detailed information about the disease phenotype, and biological samples for molecular research. However, these minimum requirements might be not sufficient for successful results. Genetic epidemiology mainly deals with complex diseases. On the one hand, most of them are relatively frequent in the population. On the other hand, they are polygenic in origin, with several genes involved, in addition to environmental influences like lifestyle and contact to toxic or carcinogenic substances. Typically, the contribution of single genes as well as of single external risk factors is small to moderate. Therefore, in addition to large sample numbers a high quality of data as well as sound epidemiological methodology are prerequisites of successful research.

To study the genetic basis of complex diseases, typically thousands rather than hundreds of patients with the disease of interest are needed. Therefore, the most common way is to recruit them via hospitals or doctor's offices. However, the quality of phenotypic characterization might be variable. It is relatively easy to recruit large numbers if the standards for diagnostic criteria are low, but if well defined patients are needed, the available numbers are much smaller. To find a strong genetic effect, a more crude phenotypic characterization might be sufficient, but for weak effects, thorough phenotypic characterization is important.

The simplest way of phenotyping for the disease of interest is to take doctors' diagnoses from the records of the patients. However, it is quite clear for many diseases that under the umbrella of one diagnosis several distinct pathophysiological entities are summed up, which might have a different genetic origin. Therefore it is crucial for genetic epidemiology to be based on well defined, standardized phenotyping, which then allows us to define subgroups of patients that have more specific clinical or sub-clinical properties. This makes it necessary to use specific laboratory parameters as well as refined diagnostic tools.

Another strong argument for the importance of sound phenotyping is the matter of intermediate phenotypes. Many parameters show an early reaction of the body, and they are associated with the disease after disease manifestation. They do, however, not yet represent a status of disease. These parameters like specific IgE, elevated cholesterol, elevated C-reactive protein or obesity are often called intermediate phenotypes. The chance of success might be higher if we try to identify genes that influence intermediate phenotypes instead of trying to identify the genetic influence on an endpoint which may be influenced by dozens of intermediate phenotypes and therefore hundreds of genes.

As long as genetic epidemiology is "only" interested in the identification of genes, most environmental factors may be ignored. However, the situation is different if one is interested in gene environment interaction. In complex diseases, it is likely that a combination of genes predisposing for the

disease and environmental factors exacerbating the impact of these genes are jointly responsible for disease development in populations. In addition, environmental factors which seem to have only a moderate impact at the population level might have larger relative risks in subpopulations with certain genetic predispositions. Classical epidemiology has always been dealing with these "environmental" risk factors, but only today we are able to combine knowledge on the genetic background with classical epidemiological research, and we have tools to investigate the interaction of genes and the environment whose applications help to understand diseases [1].

In Germany promising developments took place recently. In the past, the ethical rules of genetic epidemiological studies were strongly dependent on the local ethics committees and could be quite restrictive. Meanwhile, the situation is better since common rules have been agreed upon ([2], www2.gsf.de:6666/gem/ethik.htm). The situation has improved further in 2004, when the German National Ethics Council has published its Opinion on Biobanks for Research. In this Opinion new and research-friendly proposals for ethical regulations are made. It is suggested that in the future it should be possible to perform research without informed consent if the samples and data are anonymized or pseudonymized. It should be possible that the donor gives generalized consent for medical research, including unlimited storage. The use of old collections of biosamples should be possible under specific conditions ([3], www.ethikrat.org).

Two major German biobanking activities are ongoing. In Northern Germany POPGEN (www.popgen.de) has been established. The concept of POPGEN is to recruit patients for 8 selected diseases. It has been started to recruit 15,000 patients and a random sample of 10,000 controls. In Southern Germany, KORA ([4], www.gsf.de/KORA) has been used for collaborative genetic epidemiological research, mostly within the National Genome Research Network (NGFN, www.ngfn.de/ngfn) since 2001. KORA has been used for more than 30 studies on cardiovascular diseases, obesity, type 2 diabetes, allergies, asthma, neurologic disorders, different forms of cancer, and rare mendelian diseases, as well as for projects dealing with population genetics. Here, KORA S4 has served to look for population stratification within Germany and for linkage disequilibrium patterns among European populations. A list of publications resulting from these collaborations can be found on the KORA-gen website (www.gsf.de/KORA-gen).

Results/Project Status

In the framework of MONICA (Monitoring of trends and determinants in cardiovascular diseases) and KORA (Cooperative Health Research in the Region of Augsburg), four large population-based cross-sectional studies have been carried out, and a biological specimen bank was established in order to enable the KORA researchers to perform epidemiologic research with respect to molecular and genetic factors. The KORA study center conducts regular follow-up investigations and has collected a wealth of information on sociodemography, general medical history, environmental factors, smoking, nutrition, alcohol consumption, and various laboratory parameters. This unique resource will be increased further by follow-up studies of the cohort.

In 2004 this collection has been opened also for external researchers, under the name KORA-gen [5]. The objective of KORA-gen is to provide access to information about available population controls for genetic studies as well as provision of DNA samples/data and phenotype data. The KORA-gen infrastructure is instrumental in questions of study design, sampling, and matching, of DNA handling and determination of genetic markers, and of data structures and formats. This is supported by an internet based information resource and by providing competent individual counselling and assistance. A web-portal for genetic control populations has been made available. Partners can choose genetic controls based on age, sex and basic phenotype information. This automatic pre-filtering allows a more informed choice of controls that can further be detailed through individual and person-based counselling.

The biological samples can be genotyped directly at the GSF facilities. But it is also possible to perform the genotyping at other genotyping centers of the NGFN. There is an amount-dependent fee per sample. The underlying KORA-gen database will be fed with all genotypes and linked with the KORA database of the GSF. This gradual accumulation will add significant value to the overall dataset. In keeping and administrating the central KORA-gen database, the GSF acts as a trustee. For all genetic and phenotype data items ownership is defined to those scientists who provided these data. Access to these data for scientific analyses is only granted with permission of the data owners. Rules for data ownership and data access have been formulated and documented as Standard Operating Procedures (SOPs). The genetic database will adhere to existing standards for data communication with other partners.

KORA-gen provides data and biosamples for about 18,000 adults from the general population. It is based on 4 surveys of 4,000-5,000 participants each, the individuals being 25 to 74 years old, performed in the city of Augsburg and the two neighbour counties, with a population of 600,000 inhabitants. The number of participants is shown in Table 1.

	S1 (84/85)	S2 (89/90)	S3 (94/95)	S4 (99/01)
Men	2023	2482	2405	2090
Women	1999	2458	2451	2171

Tab 1: Sample size of KORA-gen: $n=18,079$ participants of the MONICA/KORA surveys S1 to S4 in Augsburg. The age range was 25 to 74 years at recruitment and is 30 to 90 years in 2005.

The available data and biosamples are described in table 2. However, not all parameters are available for all participants. For the surveys S1 to S3 two follow-up interviews with self-administered questionnaires and mortality follow-up have been performed in 1997/98 and 2002/03. Since 2004, the KORA study center conducts regular follow-up investigations of the original survey population. For details see [4]. Certain conditions have to be fulfilled when using KORA-gen. The rules set by the responsible ethics committee and the office for privacy/data protection have to be followed. Quality standards have to be met with respect to scientifically sound research questions, study designs have to be based on realistic sample size calculation, quality of lab tests and genotyping has to be fulfilled according to internationally accepted standards. Furthermore the rights and scientific interests of the KORA researchers have to be taken into account in a fair manner, since they performed the field work and invested a lot of costs and energy to accumulate the data and biosamples.

Outlook

KORA-gen will be further developed with respect to its organizational as well as its technical structure. The web-portal in its first version is a static information resource. In future we will turn it into a dynamic system based on our phenotype and genotype databases. This will allow our

project partners to better plan their projects and easier access the steadily growing genotype database.

In collaboration with GSF - Institute of Human Genetics and several external partners the KORA S3/F3 cohort will be used for a genome wide association study on the basis of the Affymetrix 500k chip technology.

The Public Population Project in Genomics (P3G, www.p3gconsortium.org) is a consortium that will foster collaboration between researchers in the field of population genomics. The aim of P3G is the establishment of standards, nomenclatures, communication tools and sharing of technological know-how. This will allow efficient sharing of data between projects and with the international human genetics community. KORA-gen is a member of P3G since mid 2005 and will actively contribute to the project in the next years.

Currently five publicly funded, population-based regional epidemiological cohorts exist in Germany, which would be able to provide biosamples for genetic epidemiological projects: KORA Augsburg, SHIP Greifswald, EPIC Heidelberg, EPIC Potsdam, and Heinz Nixdorf Recall Essen. They in total enclose more than 75.000 participants from the general population. As a first step it has been agreed that a minimum data set will be defined for all cohorts. With these, an aggregated database will be established on a common internet page to give an overview over the data of the single projects.

Interview	Medical examinations
<ul style="list-style-type: none"> ○ Socio-demographic variables ○ Utilization of the medical health care system ○ Smoking ○ Nutrition ○ Physical activity ○ Medication use ○ Family history ○ Women specific variables ○ Self-reported health status ○ Psychosocial variables 	<ul style="list-style-type: none"> ○ Systolic and diastolic blood pressure ○ Anthropometrical measurements (body height and weight, waist and hip circumference) ○ Bioelectrical Impedance Analysis (BIA) ○ Echocardiography ○ Electrocardiogram ○ Pulse wave analysis
Laboratory examinations	Biosamples
<ul style="list-style-type: none"> ○ Cholesterol (total, HDL, LDL) ○ Uric acid ○ Glucose (partly fasting, partly OGTT) ○ Triglycerides (partly fasting) ○ Creatinine ○ Blood cell counts ○ HbA1c ○ Haemostasiological and inflammatory parameters 	<ul style="list-style-type: none"> ○ Serum ○ Plasma ○ DNA ○ Immortalized lymphocytes ○ Urine

Tab 2: Data and biosamples of the MONICA/KORA surveys which can be used in KORA-gen (some variables are only available for subgroups).

Lit.: 1. Wichmann HE. Genetic epidemiology – from biobanking to genetic statistics. *Methods of Information in Medicine* 2005 (in press). 2. Wichmann, H.E., Jäger, L., Taupitz, J., Doppelfeld, E.: *Ethik und genetische Epidemiologie*. Deutsches Ärzteblatt 2002; 99: 2-4. 3. German National Ethics Council (Nationaler Ethikrat): *Biobanks for research – opinion* (Biobanken für die Forschung) Saladruck, Berlin 2004. 4. Holle R, Happich M, Löwel H, Wichmann HE, for the KORA Study Group. KORA – A research platform for population based health research. *Gesundheitswesen* 2005, 67 Suppl.1, 19-25. 5. Wichmann, H.E., Gieger, C., Illig, T. for the KORA Study Group: KORA-gen - Resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen*, 2005; 67 Suppl. 1, 26-30.