

SMP: Genetic Epidemiological Methods (GEM)**Project: popgen****Michael Krawczak - Christian Albrechts University, Kiel - krawczak@medinfo.uni-kiel.de****Introduction**

For most complex human diseases, the recruitment of samples for gene finding studies is retrospective in nature and limited to cases of a particularly pronounced phenotype, or with a strong family history. This strategy of “extreme (retrospective) sampling” is thought to maximise the chances for the initial detection of a disease-associated genetic variant]. However, the clinical impact of a disease gene discovery is usually only made through the implementation of diagnostic and therapeutic algorithms in an unselected (i.e. “normal”) population. For such purposes, covariate-adjusted absolute and relative genotypic disease risks are the key parameters of interest and these can only be estimated from representative population samples which must often be established anew. Even within a typical retrospective case-control study, evaluating the relative risk of a genetic variant requires knowledge of its background frequency in the study population and large samples of unrelated controls must be collected. Also, in order to be able to examine the impact of genetic factors on not only acute but also chronic disease, it is worthwhile to extend the retrospective ascertainment of phenotypes to the prospective follow-up of a subset of cases, defined for example by an incidence cohort.

Unbiased sampling of patients at the population level is best achieved by defining a confined geographical catchment area in which all clinically overt cases with the disease in question are recruited and from which no systematic or uncontrolled “escape” of patients occurs. In this respect, Northern Schleswig-Holstein (approx. 1.1 million people) in Germany (Figure 1) is ideal because its tight geographical borders (Denmark, North Sea, Baltic Sea, Kiel Canal with limited crossing points) and low density of treatment facilities (the only tertiary referral centre is the University Hospital Schleswig-Holstein Campus Kiel) set stringent limits for patients seeking treatment either inside or outside its confines. This is the target region of “popgen”. The project was initiated by clinical and non-clinical partners at the Christian-Albrechts-University Kiel in 2003 to provide disease-orientated projects from the NGFN with a unique interdisciplinary platform for the identification and cross-sectional recruitment of all locally prevalent cases with the disease in question.

Confinement of patient and proband sampling activities to the popgen target region ensures that

- all individuals classified as being affected by a disease can be identified using the resources of the German public health care system
- diagnostic accuracy and ascertainment efficiency can be monitored in a standardised fashion
- patients can easily be enrolled in a follow-up scheme
- a DNA bank can be established and made available to partners from within and outside the NGFN
- processing of phenotype data is centralised and consistent
- standard operating procedures and quality control measures can be implemented at all stages of the recruitment process.

The choice of target diseases for popgen is based upon

- an active interest expressed by one of the NGFN networks
- the (likely) availability of a susceptibility genotype for the disease in question
- a prevalence that matches the popgen design (i.e. high enough to allow reliable risk assessment, but below the limits of suitability for a cohort study)

- a phenotype that can be assessed from existing files without additional examinations.

Diseases covered by popgen, or intended to be covered in the near future, include Autism, Essential Tremor, Seizure, Parkinson Disease, Bipolar Disorder, Arteriosclerosis/ Coronary Heart Disease, Dilated Cardiomyopathy, Atopic Eczema/ Asthma, Inflammatory Bowel Disease, Sarcoidosis, Periodontitis, Colonic Adenocarcinoma, Gallstones, and Juvenile Pneumonia.

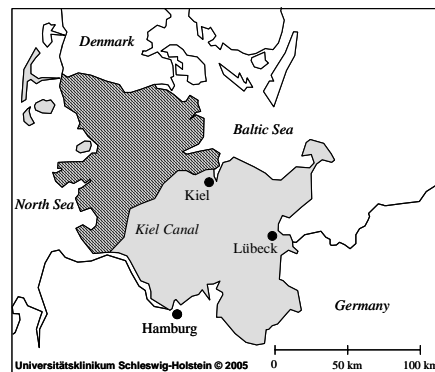


Fig 1: Map of Schleswig-Holstein (grey), the northernmost state of Germany. The popgen target area lies north of the Kiel Canal but excludes the North-Frisian Islands.

Patients are identified and contacted through their health care providers or insurers. Interested individuals respond directly to popgen and give permission for popgen personnel to obtain complete health care records. Diagnoses are verified on the basis of the available documentation using pre-established criteria as defined by the clinical partners from the respective NGFN disease-orientated network. After phenotype assessment, a representative proportion of patients are asked to participate in a follow-up scheme. Healthy control individuals are identified through official population registries and contacted by mail. The control group will ultimately comprise 7200 individuals, with 2400 people in each of three age groups (18-30 years, 30-50 years, 50-80 years). The declarations of written informed consent that are used for both patients and controls fully comply with current ICH standards for the conduct of clinical research, with some biobank-specific items added. Participants are granted at least 24-hours to withdraw from the study prior to the pseudonymization or anonymization of their data. All recruitment and data management procedures have been approved by the ethics committee of the Kiel Medical Faculty and by the data protection officer of the University Hospital Schleswig-Holstein.

All participants donate 30 ml of EDTA blood (yielding 600-1000 microgram of DNA – sufficient for a large number of genetic tests). The resulting DNA bank is accessible to all NGFN members for exploring/verifying genotype-phenotype relationships, provided that promising disease-associated genetic variants have been detected for the condition in question. Access to the DNA bank must be approved by the local ethics committee and is free of charge to NGFN members.

Results/Project Status

For several diseases, identification and/or contacting of patients has already been completed. Recruitment for

additional diseases, involving both popgen staff and representatives of the respective NGFN disease-orientated networks, are underway.

Cardiovascular diseases

The Arteriosclerosis/Coronary Heart Disease (CHD) project is conducted on behalf of the Cardiovascular Disease Network ("CardioNet") of the NGFN. All patients included are below 65 years of age and show significant CHD, as validated by cardiac catheterisation. To achieve maximum recruitment efficiency, a close collaboration has been initiated between popgen and the six centres performing coronary angiography in the catchment area. All cardiac catheterisations performed at these units between January 1997 and June 2005 were scrutinised (45,000 in total) and 6000 patients identified as matching the popgen inclusion criteria. These patients have been contacted, and almost 3100 have agreed to participate in the study. Given a response rate of over 50% upon single contact, past experience from NGFN collections established under similar conditions suggests that the final recruitment rate will exceed 80% once additional approaches by mail and telephone have been made. From the treatment records, a detailed disease history is ascertained, including information on myocardial infarctions, surgery, heart and kidney function, glucose and fat metabolism etc. A standardised questionnaire is sent out to the participating patients in order to obtain additional demographic, phenotypic and environmental information. Under the same recruitment scheme as used for CHD (i.e. age ≤ 55 years), some 400 patients with Dilated Cardiomyopathy, but without CHD, could be identified. These patients have been contacted for inclusion in a separate project.

Neuropsychiatric diseases

Parkinson Disease (PD) is a key phenotype in the Neuronal Disease Network ("NeuroNet") of the NGFN and has a prevalence of approximately 100 to 200 per 100,000, with an age-related increase in all populations. As yet, however, genetic risk factors for PD have been identified mostly in early-onset patients. From the popgen target population, between 1100 and 2500 PD patients are expected to be identifiable, irrespective of age. From these candidates, PD patients will be selected and recruited in a two-tiered fashion. First, all neurologists and psychiatrists in the popgen catchment area will be contacted (approximately 70, including four hospitals) and patients identified by searching the registers of board-certified physicians. More than 500 patients have been contacted and 150 have agreed to participate. Second, a cross-sectional sample of 40,000 people of at least 60 years of age will be taken from a defined subregion of the popgen catchment area. This strategy will target PD patients who are only treated by general physicians and who would therefore escape a specialist-based recruitment scheme. A popgen project on Bipolar affective disorder (BPAD) has also started. Since BPAD affects up to 5% of the general population, an assumed response rate of 50% will result in the inclusion of some 3000-5000 individuals in the popgen BPAD sample. Some 1700 patients have so far been identified.

Environmental diseases

A close collaboration has been established between popgen and the Environmental Disease Network ("EnviroNet") of the NGFN. popgen-based patient sampling is in progress for four chronic diseases: Bronchial Asthma, Juvenile Periodontitis, Inflammatory Bowel Diseases (IBD), and Sarcoidosis. The first three are of great importance to public health due to their high prevalence (5-10% for Asthma, 1% for Juvenile Periodontitis, 0.5% for IBD). Current aetiological concepts maintain that environmental factors are important for these diseases to develop but that individuals have to be genetically predisposed in order for an external stimulus to trigger disease onset. Multiple predisposing genes have

already been identified for some of the diseases in question using highly selected patient samples, however their impact at the population level and the degree of interaction with environmental factors are still unknown.

Controls

Recruitment of popgen control individuals is almost complete for the City of Kiel. In total, 4300 probands have agreed to participate and some 3600 DNA samples are currently available. Recruitment of a second, equally sized set of controls from rural areas surrounding the City of Kiel began in Autumn 2005. Data from control individuals are totally anonymised before inclusion in the popgen database.

Ascertainment efficiency and follow-up

The response rate is constantly monitored in each phase of the popgen recruitment process and, if necessary, recruitment is promoted by repeated mail or telephone contacts. Cumulative prevalence figures per age group and geographical region have been provided by health insurers, thus allowing determination of the ascertainment efficiency. Previous experience suggests that the local set-up in Schleswig-Holstein leads to an extremely high response rate. This is exemplified by the envisaged population-wide recruitment of over 80% of diagnosed cases with CHD. Long term clinical research performed at the Department of Internal Medicine, Kiel, on similar instances of chronic illness, further indicates that more than 50% of patients can be subjected to long-term follow-up. Patients and their doctors will be contacted once a year to obtain information about the development of the patient's disease. This scientifically important follow-up scheme will nevertheless be confined to a set of key features and is primarily intended to help in characterising the natural course of disease. It will not resolve any details related to, for example, pharmacological response.

Outlook

popgen was founded in May 2003 through the Optimisation and Networking Fund of the NGFN and is currently funded through a two-year grant from the second round of NGFN funding. Since then, popgen has successfully established itself as a large-scale genetic epidemiological project of international recognition managed by a dedicated team at the University Clinic in Kiel. Standard operating procedures and stringent data security measures have been established and an independent medical director guarantees that professional standards are adhered to in all disease- and patient-related matters. In addition three local university institutions (Department of General Internal Medicine, Institute of Clinical Molecular Biology, Institute of Medical Informatics and Statistics) have assumed joint responsibility for popgen's supervision. In view of its central importance to all patient-related research in the NGFN, the NGFN project committee inaugurated an external advisory board for popgen in January 2004. This board involves the speakers of the disease-orientated networks and of the central methodological platforms (genotyping, genetic epidemiology) within the NGFN, with bylaws regulating the mutual responsibilities and interactions of the advisory board and the popgen management. popgen now provides a unique population-based resource, focused upon a specific set of diseases, that is indispensable for the future development of genetic medicine. popgen will thus be of pivotal importance for patient-based studies within the confines of the NGFN, but at the same time is open to external collaborators with an active interest in disease-orientated genomic research. You may read more about popgen at www.popgen.de and in a forthcoming manuscript [1].

Lit.: 1. Krawczak M et al. popgen: Population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Comm. Genet. (In Press).