EpiHealth: a large population-based cohort study for investigation of gene–lifestyle interactions in the pathogenesis of common diseases

Lars Lind, Sölve Elmståhl, Ebba Bergman, Martin Englund, Eva Lindberg, Karl Michaelsson, Peter M. Nilsson & Johan Sundström

European Journal of Epidemiology
Affiliated to the European Epidemiology Federation

ISSN 0393-2990

Eur J Epidemiol
DOI 10.1007/s10654-013-9787-x

Springer

Eur J Epidemiol
DOI 10.1007/s10654-013-9787-x

NEW STUDY

EpiHealth: a large population-based cohort study for investigation of gene–lifestyle interactions in the pathogenesis of common diseases

Lars Lind · Sölve Elmståhl · Ebba Bergman · Martin Englund · Eva Lindberg · Karl Michaelsson · Peter M. Nilsson · Johan Sundström

Received: 28 September 2012 / Accepted: 16 February 2013

Abstract The most common diseases affecting middle-aged and elderly subjects in industrialized countries are multigenetic and lifestyle related. Several attempts have been made to study interactions between genes and lifestyle factors, but most such studies lack the power to examine interactions between several genes and several lifestyle components. The primary objective of the EpiHealth cohort study is to provide a resource to study interactions between several genotypes and lifestyle factors in a large cohort (the aim is 300,000 individuals) derived from the Swedish population in the age range of 45–75 years regarding development of common degenerative disorders, such as cardiovascular diseases, cancer, dementia, joint pain, obstructive lung disease, depression, and osteoporotic fractures. The study consists of three parts. First, a collection of data on lifestyle factors by self-assessment using an internet-based questionnaire. Second, a visit to a test center where blood samples are collected and physiological parameters recorded. Third, the sample is followed for occurrence of outcomes using nationwide medical registers. This overview presents the study design and some baseline characteristics from the first year of data collection in the EpiHealth study.

Keywords Epidemiology · Lifestyle · Gene · Prospective · Cohort study

Introduction

Most diseases affecting middle-aged and elderly subjects are multigenetic, but lifestyle exposures are also of major importance in the pathogenesis of those disorders, especially with increasing age. Many studies of interactions between genes and lifestyle factors are hampered by a lack of power to examine more than one interaction at a time, as recently discussed in several papers [1–8]. It has been estimated by an expert panel assembled by the US National Human Genome Research Institute in 2004 that several hundred thousand subjects are needed to study gene–environmental interactions in a proper way in prospective cohort studies [9].

There are two principally different ways of dealing with this problem. The first one is to pool individual data or meta-analyse data from existing studies. Several large-scale studies exist, such as the UK Biobank [10], deCODE in Iceland [11], Millennium in Japan [12], the Kadoorie study of Chronic Diseases in China [13] and CONOR/HUNT in Norway [14], mainly for the study of diseases in the middle-aged and elderly. LifeLines in Holland [15], the Norwegian Mother and Child Cohort study (MoBa) [16], ALSPAC (the Avon Longitudinal Study of Parents and Children) [17], the US National

Lars Lind and Sölve Elmståhl contributed equally to this work.

L. Lind (✉) · E. Bergman · E. Lindberg · J. Sundström
Department of Medical Sciences, Uppsala University,
751 85 Uppsala, Sweden
e-mail: lars.lind@medsci.uu.se

S. Elmståhl
Division of Geriatric Medicine, Malmö University Hospital,
205 02 Malmö, Sweden

M. Englund
Department of Orthopaedics, Lund University,
221 85 Lund, Sweden

K. Michaelsson
Department of Surgical Sciences, Uppsala University,
751 85 Uppsala, Sweden

P. M. Nilsson
Department of Clinical Sciences, SUS Malmö,
205 02 Malmö, Sweden

Published online: 24 February 2013

Table 1 Since the EpiHealth study intends to be used to study several of the common degenerative disorders seen in middle-age and in the elderly, no primary endpoint is identified

Prioritized area of research	Preliminary endpoints
Cause-specific mortality	Overall and cause-specific death
Cardiovascular disease	Myocardial infarction, stroke, heart failure, atrial fibrillation, aortic aneurysm, hypertension
Pulmonary disease	Chronic obstructive pulmonary disease, sleep-apnea syndrome
Osteoporosis	Bone fractures
Cancer	Incident cancer subtypes
Diabetes/obesity	Diabetes, body mass index, metabolic syndrome
Cognitive function	Dementia, cognitive function tests
Psychiatry	Depression, anxiety
Pain	Self-assessed pain, osteoarthritis, joint replacement
Functional capacity	Self-assessed by questionnaire
Quality of life	Self-assessed by questionnaire

Instead, eleven different prioritized areas of research have been identified with different specified endpoints within each area

Child Study [18], the Global Asthma and Allergen European Network (Ga2len) [19], and the Southampton Women's Survey [20] are all large studies in younger subjects that will be very valuable in the future, but will demand several decades of follow-up before they can be use for the study of diseases affecting the middle-aged and elderly.

The second approach is to build new studies with sufficient power to analyze several gene–environment interactions in a proper way. In Sweden, two such studies have been initiated in recent years, the LifeGene study and the EpiHealth study. The LifeGene study aims to collect 500,000 individuals aged 0–45 years, while the EpiHealth study aims to collect 300,000 individuals aged 45–75 years. Thus, together these two studies will cover most of the lifespan except for the oldest inhabitants. The LifeGene study has recently been presented in detail [21], and the present paper aims to outline the design and some results from the first year of data collection in the EpiHealth study.

Project organization and aims of the study

The EpiHealth consortium

The EpiHealth cohort study is planned and carried out by the Lund–Uppsala University epidemiology collaboration (EpiHealth) that was named in 2009 by the Swedish Research Council a "center of excellence" regarding epidemiological research and is committed to large-scale projects in the future. The Swedish Research Council has provided funding for the study for the first 5 years. EpiHealth is governed by a Steering Committee with representatives from Lund and Uppsala universities. The Steering Committee is responsible for the performance and conduct of the study. The EpiHealth Data Access Committee (EDAC), appointed by the Steering Committee, will review applications for withdrawal of data and samples from the EpiHealth database and biobank, respectively.

The study started with two centers, in Uppsala and Malmö, but the intention is to open centers in other Swedish towns in the future to obtain a geographical spread and to collaborate with the LifeGene Test Centres.

Study objectives

Primary objectives

The primary objectives of the EpiHealth cohort study is to provide a resource to study interactions between genotypes and lifestyle factors in a large cohort (aim is 300,000 individuals) derived from the Swedish population in the age range 45–75 years regarding development of common degenerative disorders (Table 1).

Secondary objectives

Secondary objectives of EpiHealth are:

1. To study lifestyle patterns in the ages 45–75 years and differences between genders.
2. To study heritability of common degenerative disorders in the Swedish population by using data from families attending both EpiHealth and LifeGene.
3. To evaluate novel biomarkers in serum and plasma for incidence of common degenerative disorders.
4. To study associations of genotypes with development of common degenerative disorders.
5. To study associations of lifestyle factors with development of common degenerative disorders.

Study design

EpiHealth is a population-based, multicenter longitudinal cohort study in subjects aged 45–75 years in which exposures (lifestyle factors and genes) are collected at a baseline investigation aiming for 300,000 individuals.

The study consists of three parts:

1. A collection of data on lifestyle factors by self-assessment using an internet-based questionnaire.
2. A visit to a test center where blood samples are collected and physiological parameters recorded.
3. The population is followed for occurrence of future endpoints defined from eleven prioritized research areas (see Table 1).

Repeated collection of data on exposures, i.e. web-based questionnaires and test center visits, are planned for every 5th and 10th year, respectively, following the baseline investigation.

Study outline and objectives of the pilot phase

A pilot study was performed in 2011 in the first 1,000 individuals mainly in order to test some logistical issues. The objectives of the pilot phase were:

1. To test the achievability of the study
   - 1a. Participation rates in the different age strata.
   - 1b. Collection of fasting blood samples at a test center visit separate from the collection of the rest of the data using drop-in time and "short visits."
2. To evaluate the feasibility of answering the questionnaire by age.
3. To evaluate the instruments used at the test centers to measure physical variables in order to optimize throughput.

It should be pointed out that we did not expect any major alterations in the collection of data from the pilot phase to the main study, so the data collected in the pilot phase would be included in the main study. Thus, from that perspective, the subjects investigated in the pilot phase will contribute as much as subsequent participants to the results in the study.

Only very minor changes in the recruitment and collection of data were made when moving from the pilot phase to the permanent study.

Recruitment of participants

A randomization procedure will assure that all age groups and both genders are equally represented in the main study. The sample will be included employing population-based randomization using official registries and/or commercial address databases. Inclusion criteria are age 45–75 years, a permanent address in Sweden and a Swedish civic registration number. No exclusion criteria based on prevalent diseases or medications exists.

In the pilot phase, an equal number of male and female subjects were invited to the Uppsala Test Center using the following age strata; 45–49, 50–54, 55–59, 60–64, 65–69, and 70–75 years. We have also chosen to use this procedure in the main study, although this will entail some underrepresentation in the younger males, please see Tables 4 and 5.

Invitation of participants

Potential participants in the age group 45–75 years will receive information about EpiHealth and an invitation to participate in the study by postal mail. The postal invitation will consist of:

1. Letter of invitation with activation codes and instruction on how to enroll in EpiHealth on the website www.epihealth.se or via the EpiHealth Screening Center.
2. Brochure about EpiHealth including general information about EpiHealth, the two parts of the study (questionnaire and test center visit) and also information about informed consent.
3. The Informed Consent form.

A person who wants to participate in EpiHealth will be instructed to visit www.epihealth.se and log on using the activation code supplied in the invitation letter. The website contains general information about EpiHealth and detailed information about consent and how to partially or completely withdraw consent, and this information will be publicly accessible. In addition, the questionnaire is accessible on the website before enrollment. Participants can also, if preferable, contact an EpiHealth Screening Center for help with enrollment.

Initial consent is sought on the website following activation using a check box where the participant confirms participation and that he or she has read and understood the consent agreement.

Formal written consent is thereafter sought from participants in EpiHealth during the first visit to the test center. The potential participant will be given detailed information about the formal written consent by personnel at the test center. A major purpose of EpiHealth is to enable scientists to apply for data and/or samples from the EpiHealth database and biobank, respectively. Therefore, the consent given by the participant is broad, comprising: (1) handling personal data and saving results from the questionnaires and test center visits, (2) biobank blood samples, (3)

biobank DNA, (4) using data and samples collected for health-related research in specific projects approved by an Ethical Committee, (5) merging data with official registries and national quality registries to gain the information specified in the EpiHealth information brochure, (6) merging data with medical records; and that participant has been informed about the possibility of partially or completely withdrawing his or her consent without further explanation. Consent can be partially or completely withdrawn at the request of participants at any time. Participants themselves schedule a time for the visit to the test center via a calendar function on the website.

EpiHealth questionnaire

The questionnaire is about medical history, family history, and symptoms, as well as lifestyle factors, divided into fifteen sections. The estimated time to complete the form is 40–60 min. Computers are available at the test center for participants who wish to fill the questionnaire at the test center and for those who need computer assistance. The baseline questionnaire is customized for gender and is hierarchically branched, based on the participant's answers. The groups of items included in the lifestyle questionnaire are given in Table 2. No attempt is made to check whether the questionnaire is filled in properly. In most cases, the duration of a disease, mediation, smoking, etc. is asked for.

Physical measurements at the test center

Following the signing of the formal written consent, the receptionists take a simple 1-lead ECG (lead I) for 30 s by contact between the participant's hands and ECG electrodes (Zenicor-EKG®, Zenicor, Stockholm, Sweden) to screen for atrial fibrillation and other arrhythmias [22].

Physical tests performed in the test room:

- Blood pressure and pulse rate are recorded twice in the sitting position by an automatic device (Omron, Kyoto, Japan) [21].
- Height is recorded and waist circumference is measured at the umbilical level [21].
- Weight is recorded on a scale that uses bioimpedance to calculate fat mass (Tanita, Tokyo, Japan) [21].
- A simplified resting lung function test is performed to measure FEV1 and VC (MiniSpir®, MIR Medical International Research, Waukesha, WI, USA) [21].
- The cognitive function tests Trail making Test A and B are performed at a computer [23].

Table 2 The groups of items included in the personal background and lifestyle questionnaire are the following

Family structure	Tobacco use
Social group	Physical activity
Education	Work life history
Reproduction	Pain history
History of diseases	Quality of life
Medication usage	History of injuries
Food intake	Environmental exposures
Alcohol consumption	Sleeping habits

At the test center, 100 ml of blood is taken. 90 ml is prepared into plasma, serum and whole blood (for later DNA extraction) and stored in a biobank facility for later analysis. 10 ml are used for determinations of fasting glucose, LDL- and HDL-cholesterol, and serum triglycerides at the hospital laboratory using an Architect Ci8200 analyzer (Abbott Laboratories, Abbott Park, IL, USA) [24].

The personnel at the test centers have been trained together to minimize measurement variation. We use the same equipment at all test centers. We use the same biochemical laboratory for analysis of glucose and lipids.

If consented to, follow-up personal information and data from the test center will be available on the participant's personal EpiHealth page at www.epihealth.se within 3 weeks. The feedback will include the actual figures, reference values, and recommendations on actions to be taken, if any, in most cases a visit to the GP's office or some other personal physician.

Cut-off values used for recommendations:

1. Fasting plasma glucose >7.0 mmol/l. A new measurement at GP's office recommended within a couple of weeks.
2. LDL >3.5 mmol/l. Decision at GP's office about dietary treatment/lipid lowering therapy needed within 2 months.
3. Serum triglycerides >2.0 mmol/l. Decision at GP's office about dietary treatment/lipid lowering therapy needed within 2 months.
4. Blood pressure >140/90 mmHg. A new measurement at GP's office recommended within a couple of weeks.
5. An FEV1 <70 %. A visit to the GP's office recommended within a couple of weeks if a diagnosis of asthma or COPD is not previously known.

For blood pressures >180/110, fasting plasma glucose >10.0 mmol/l, or a previously unknown atrial fibrillation, the participant is contacted within a few days with a recommendation of an urgent check-up by their GP. If the rate of atrial fibrillation is >100 beats/min, the participant is referred to the nearest emergency department.

Participants will not receive any feedback information regarding genetic analysis or analysis of blood samples stored in the biobank and later analyzed.

Table 3 The EpiHealth database will be merged with several official national Swedish registers

1	*Swedish Population Registry*; regarding address, place of birth, parents place of birth and marriage
2	*Swedish Censuses (1960–1990)*; regarding social group, income, living conditions, education
3	*Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknads-studier (LISA) (from 1990)*; regarding employments and sick leave
4	*Swedish National Insurance Agency*; regarding sick leave
5	*Utbildningsregistret (UREG)*; regarding education
6	*Swedish Multi-Generation Register;* regarding family members
7	*Medicinska födelseregistret (MFR)*; regarding pregnancies and deliveries, birth-weight
8	*Swedish Patient Registry;* regarding in-hospital care
9	*Swedish Prescription Registry;* regarding use of medication
10	*Swedish Cause-of Death Registry;* regarding vital status and cause of death
11	*Swedish Cancer Registry;* regarding cancer
12	*Outpatient Registries;* regarding care in primary care facilities (includes *Day Surgery Registry*)
13	*Swedish military service conscription register*; regarding blood pressure, exercise capacity, body height and weight
14	*Swedish Information System on Occupational Accidents and Work-related Diseases*; regarding occupational accidents and work-related diseases

Assessment of outcomes

In most cases, endpoints will be collected by merging the database with the official national registers specified in Table 3. Some of the registers in Table 3 will also be used to obtain data on exposures. The EpiHealth database will annually be merged with the registers providing data on outcomes. Hospital or primary care records might also be needed for the validation of the correctness of some diagnoses in the official registers. EpiHealth data might also be merged with data from the Swedish Quality of Care registers, usually run by research groups.

Statistical evaluation

The statistical methods to be used in the evaluation of the data in the study will be dependent on the research question and the nature of the data. In the case when incident cases of a certain disease after a certain follow-up period are analyzed as an outcome and several life-style factors together with several genotypes are used as exposures, Cox proportional hazard analysis or Poisson regression analysis will typically be performed. If cross-sectional analysis of a dichotomous outcome variable is analyzed, typically logistic regression analysis will be used. If cross-sectional analysis of a continuous outcome variable is analyzed, typically linear regression analysis will be used. Also non-parametric analysis will be performed if the data do not permit parametric tests.

Ethics policy

The study has been approved by the Ethics Committee of Uppsala University at the start of the pilot phase of the study and before the continuation of the study (Log. No. 2010/402), and by the Swedish Data Inspection Board (No. 307-2011). Approval by the Ethics Committee will also be sought each time a statistical analysis or sample withdrawal from the biobank is to be performed according to the decision of the EpiHealth Data Access Committee. The EpiHealth study has a written Ethics Policy document that is publically displayed on the homepage.

Data access procedure

EpiHealth is a resource open to all scientists. EpiHealth welcomes research applications from scientists other than those actively working with the collection of data. The application process consists of four stages, which must be completed before samples and/or data are provided:

1. The researcher submits an application (research proposal) to the EpiHealth Data Access Committee, which reviews the application and determines the availability of samples and/or data and the eligibility of the proposed study for access (the process is shortened when no samples involved).
2. After possible modifications to the research proposal suggested by the EpiHealth Data Access Committee, the researcher submits an application to and obtains permission from the regional Ethical Committee to perform the research specified in the research proposal.
3. The researcher contributes to displaying a short description of the research proposal on EpiHealth's homepage for 1 month, so that participants in EpiHealth can withdraw consent if they are not comfortable with the research proposal.
4. The researcher agrees to the conditions of access and signs the material transfer agreement/data transfer agreement (MTA/DTA).

Discussion

This overview describes the design and first results of the EpiHealth cohort study in Sweden, which aims to collect data on 300,000 individuals. The date of including the first patient in the study was April 26th 2011 in Uppsala, and in January 9th 2012 the test center in Malmö started to include subjects. Thus, today inclusion of subjects is performed in two different parts of Sweden. It is evident, however, that if the EpiHealth study is to reach its goal of 300,000 screened individuals, other test centers will need to open in the two largest cities in Sweden, Stockholm and Gothenburg. Furthermore, in order to be a nation-wide study it is necessary to also have a test center in the Northern part of Sweden, although that part of the country is more sparsely populated.

As previously stated, the EpiHealth database and biobank is intended for use not only by investigators involved in the collection of data. Thus, in the Spring of 2013 we plan to circulate a call to researchers at the Swedish universities to submit research proposals for use of the EpiHealth database and biobank samples.

The EpiHealth screening has thus far experienced a moderate participation rate. In January 1st 2013, >7,000 subjects had been included in the study, and the participation rate was around 25 %. It is therefore obvious hat EpiHealth cannot serve as a database to study true prevalence rates of diseases in Sweden. In contrast, the study will well serve the postulated main aim, to investigate gene–lifestyle interactions in the causation of common diseases in the middle-aged and elderly, since this only requires sufficiently large ranges in the distributions of lifestyle factors and genetic variants. The large sample size of EpiHealth will ensure sufficient diversity in the distributions of lifestyle factors and genetic variants, even though the sample might not be truly representative of the Swedish population. It should be noted that the latter limitation is not unique to the EpiHealth cohort. The healthy participant effect is an issue in most population-based research.

Although we cannot investigate the true prevalences of diseases with this low participation rate, it will be possible to perform other kinds of traditional epidemiology, such as investigating the associations between non-genetic exposures and different outcomes, since the large number of observations collected in the study will also ensure that exposure levels will be represented in the cohort that would not have been included in a small study with a low participation rate. Thus, the EpiHealth cohort study will also serve as a basis for investigating many research questions apart from gene–environmental interactions.

It has been estimated that several hundreds of thousands of subjects are needed in order to study several gene–environmental interactions in a proper way in prospective cohort studies [9]. The power in the present study will vary across the different outcomes collected in the present study. Since we expect fewer incident cases of aortic aneurysms than cases of myocardial infarction, for example, the power will naturally be higher for outcomes with higher incidence rates. Furthermore, we have outcomes in the form of continuous data, like quality of life or pain, with data collected from everyone in the questionnaire at baseline, which will entail better investigative power than when incident cases have to be collected. Thus, the power to investigate gene–environmental interactions will vary considerable according to the chosen outcome. This power will also be contingent on the number of interactions between genes and environmental factors that are to be studied.

Assuming an incidence rate of 10 % over 10 years for a binary outcome, the study has power to detect interactions up to the order of 11 for 11 different exposures.

The EpiHealth study collaborates closely with the LifeGene study [21] in that we share the same IT system. Further, more than 75 % of all questionnaire items are identical between the studies, and we also use almost identical clinical examinations and blood sample collection procedures. Therefore, we can easily use a wealth of data from both studies to investigate changes in several lifestyle-related and physiological variables throughout the lifespan, except in the oldest segment of the age distribution. The reason for restricting the upper age limit to 75 years is both that there is a long induction time period for many late onset diseases and that knowledge of how to use a computer is fundamental for the data collection in EpiHealth. The proportion of the Swedish inhabitants that can use a computer drops dramatically above the age of 75. We have a limited capacity to help elderly persons with computer problems at the test centers.

In Tables 4 and 5 we present the first descriptive data on the major physiological and biochemistry variables collected at the test centers. These data will just provide some flavor of the data collected in the study, and may enable comparisons with other cohorts and target populations in order to judge external validity of results from the EpiHealth study.

Using the current ECG device we record 30 s. This is in fact a longer time period than is usually recorded using a 12-lead resting ECG, and since the p-wave is usually detectable in this single lead, the occurrence of a sinus rhythm is easy to detect in the vast majority of cases.

Using this device at the test center only, we cannot detect paroxysmal arrhythmias. In order to do so, we would need to perform Holter monitoring for at least 24 h, and the present high-throughput screening study cannot perform such a huge task in many thousands of subjects.

Table 4 Means and SD for men regarding some measured physiological and laboratory variables in different age groups

Variable	45–55 years		56–65 years		66–75 years	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Systolic blood pressure (mmHg)	421	132 (16)	547	138 (17)	724	143 (17)
Diastolic blood pressure (mmHg)	421	84.9 (9.9)	547	86.2 (10.0)	724	84.7 (9.5)
Height (cm)	421	179 (7.0)	548	178 (6.7)	724	177 (6.4)
Weight (kg)	421	86.1 (12.3)	548	85.8 (13.0)	723	83.6 (11.8)
Waist circumference (cm)	421	94.9 (10.0)	548	97.7 (10.3)	724	98.1 (9.5)
Hip circumference (cm)	421	102.7 (6.1)	548	103.2 (6.9)	724	103.0 (6.5)
BMI (kg/m^2)	421	26.6 (3.5)	548	26.8 (3.6)	723	26.6 (3.4)
Waist/hip ratio	421	0.92 (0.06)	548	0.95 (0.06)	724	0.95 (0.06)
Total body fat (%)	419	22.9 (5.5)	543	24.7 (5.3)	705	25.7 (5.4)
Forced expiratory volume at 1 s (l)	237	4.0 (0.7)	259	3.6 (0.8)	342	3.2 (0.9)
Forced vital capacity (l)	237	5.2 (1.0)	259	4.9 (1.0)	342	4.3 (1.0)
Trail making test A (sec)	416	31.8 (10.5)	545	37.9 (11.2)	711	45.7 (15.7)
Trail making test B (sec)	416	47.0 (19.0)	545	55.9 (22.8)	711	71.2 (33.4)
Fasting glucose (mmol/l)	410	5.9 (0.9)	525	6.2 (1.4)	708	6.3 (1.4)
Serum cholesterol (mmol/l)	410	5.90 (1.05)	526	5.92 (1.11)	708	5.66 (1.15)
LDL-cholesterol (mmol/l)	410	3.99 (0.95)	526	3.96 (0.98)	708	3.75 (1.04)
HDL-cholesterol (mmol/l)	410	1.32 (0.34)	526	1.35 (0.33)	708	1.34 (0.32)
Serum triglycerides (mmol/l)	410	1.41 (0.78)	526	1.45 (0.86)	708	1.34 (0.63)

The resting lung function test was not included in the first investigations performed. Trail making tests A and B are cognitive function tests

Table 5 Means and SD for women regarding some measured physiological and laboratory variables in different age groups

Variable	45–55 years		56–65 years		66–75 years	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Systolic blood pressure (mmHg)	720	124 (15)	711	132 (18)	720	141 (18)
Diastolic blood pressure (mmHg)	720	80.3 (9.5)	711	82.7 (9.4)	720	82.8 (10.0)
Height (cm)	720	166 (6.6)	711	165 (5.8)	720	163 (5.9)
Weight (kg)	720	70.5 (12.1)	710	70.3 (12.1)	718	69.7 (11.5)
Waist circumference (cm)	720	84.5 (11.4)	711	87.3 (11.8)	720	87.8 (11.8)
Hip circumference (cm)	720	100.9 (9.1)	711	102.1 (8.8)	720	102.7 (8.9)
BMI (kg/m^2)	720	25.4 (4.1)	710	25.7 (4.2)	718	25.9 (4.1)
Waist/hip ratio	720	0.84 (0.06)	711	0.85 (0.06)	720	0.85 (0.07)
Total body fat (%)	716	33.7 (6.5)	704	35.7 (6.2)	712	36.5 (6.2)
Forced expiratory volume at 1 s (l)	405	2.9 (0.5)	355	2.5 (0.5)	331	2.2 (0.5)
Forced vital capacity (l)	405	3.7 (0.8)	355	3.3 (0.7)	331	2.9 (0.7)
Trail making test A (sec)	718	32.7 (11.0)	702	37.9 (10.9)	715	44.9 (15.1)
Trail making test B (sec)	718	46.5 (17.9)	702	54.4 (22.2)	714	70.0 (30.5)
Fasting glucose (mmol/l)	681	5.6 (0.8)	683	5.9 (1.2)	699	5.9 (0.8)
Serum cholesterol (mmol/l)	681	5.72 (1.05)	683	6.29 (1.02)	699	6.37 (1.06)
LDL-cholesterol (mmol/l)	681	3.65 (0.95)	683	4.1 (0.91)	699	4.16 (0.96)
HDL-cholesterol (mmol/l)	681	1.64 (0.34)	683	1.71 (0.4)	699	1.69 (0.37)
Serum triglycerides (mmol/l)	681	1.01 (0.52)	683	1.15 (0.56)	699	1.25 (0.57)

The resting lung function test was not included in the firstinvestigations performed. Trail making tests A and B are cognitive function tests

Population-based epidemiological studies are both costly and may entail some ethical problems that need to be handled in a careful way. The funding of EpiHealth thus far remains in the academic world and is based on the support

to EpiHealth from the Swedish Government, as administered by the Swedish Research Council, but also other financial support has to be sought.

Ethical discussions are much needed regarding population-based studies, but as EpiHealth relies on the full informed consent of study participants and has obtained ethical approval for collection of data, these aspects are now covered. Furthermore, future research projects based on data and biobank samples from the EpiHealth cohort require additional ethical approval, to be sought from regional ethics committees. Thus, all research projects emerging from the EpiHealth study have been subjected to two different ethics committee approvals.

One important goal, as stated in the aims, is to better describe and understand disease processes based on gene–lifestyle interactions. An important consequence would be to find new targets for treatment, both for new drugs and for specific lifestyle interventions tailored to the profile of individuals. This will enhance the possibility of developing what has been called personalized medicine.

Population-based studies including genetic testing are not without problems [25]. However, there is no other way to elucidate early developments of disease conditions, as this could not be done either in patients with established disease, where medical treatment will often interfere with the studied mechanisms, or in laboratory animals, or in silico, based on computer-based modeling. Large groups of subjects representing a wide range of phenotypes have to be studied. In the future international collaboration is also needed, especially within European populations. Such projects already exist to a certain degree, for example the European Prospective Investigation into Cancer and nutrition (EPIC) [26], which is investigating risk factors for cancer, diabetes, and cardiovascular disease. EPIC has collected information from a number of European cohorts mostly from the 1990s, thus reflecting the phenotyping standards of that time. New cohorts could expand in methodology to enable the better and more advanced phenotyping that is the third pillar of the modern design of epidemiological population-based studies, the two other pillars being modern genotyping (and inclusion of omics) and updated national register data information. In many countries the first two requirements can be fulfilled, but only in a few countries, Sweden being one of them, can the third requirement (national registers based on personal ID with high coverage) be added [27].

In conclusion, the EpiHealth cohort study, with the aim of collecting data on 300,000 individuals, has been successfully initiated and during its first year has established two test centers in Sweden, which thus far have examined more than 7,000 individuals. The study, given continued funding, will be a major foundation for the study of gene–lifestyle interactions in the pathogenesis of common diseases in the middle-aged and elderly in the future.

Acknowledgments We thank the Swedish Research Council for supporting the strategic research network Epidemiology for Health (EpiHealth) and thereby also the EpiHealth screening cohort.

References

1. Marras C, Goldman SM. Genetics meets environment: evaluating gene–environment interactions in neurologic diseases. Semin Neurol. 2011;31(5):553–61.
2. Renz H, Autenrieth IB, Brandtzaeg P, et al. Gene–environment interaction in chronic disease: a European Science Foundation Forward Look. J Allergy Clin Immunol. 2011;128(6 Suppl): S27–49.
3. Heim C, Binder EB. Current research trends in early life stress and depression: review of human studies on sensitive periods, gene–environment interactions, and epigenetics. Exp Neurol. 2012;233(1):102–11.
4. de Magalhaes JP, Wuttke D, Wood SH, Plank M, Vora C. Genome–environment interactions that modulate aging: powerful targets for drug discovery. Pharmacol Rev. 2012;64(1):88–101.
5. Rigoli L, Briuglia S, Caimmi S, et al. Gene–environment interaction in childhood asthma. Int J Immunopathol Pharmacol. 2011;24(4 Suppl):41–7.
6. Dunn EC, Uddin M, Subramanian SV, Smoller JW, Galea S, Koenen KC. Research review: gene–environment interaction research in youth depression: a systematic review with recommendations for future research. J Child Psychol Psychiatry. 2011;52(12):1223–38.
7. Cornelis MC, Hu FB. Gene–environment interactions in the development of type 2 diabetes: recent progress and continuing challenges. Annu Rev Nutr. 2012;32:245–59.
8. Lee YC, Lai CQ, Ordovas JM, Parnell LD. A database of gene–environment interactions pertaining to blood lipid traits, cardiovascular disease and type 2 diabetes. J Data Mining Genomics Proteomics. 2011;2(1):106.
9. Collins FS. The case for a US prospective cohort study of genes and environment. Nature. 2004;429(6990):475–7.
10. Ollier W, Sprosen T, Peakman T. UK Biobank: from concept to reality. Pharmacogenomics. 2005;6(6):639–46.
11. Gulcher J, Stefansson K. Population genomics: laying the groundwork for genetic disease modeling and targeting. Clin Chem Lab Med. 1998;36(8):523–7.
12. Nakamura Y. The biobank Japan project. Clin Adv Hematol Oncol. 2007;5(9):696–7.
13. Chen Z, Lee L, Chen J, et al. Cohort profile: the Kadoorie Study of chronic disease in China (KSCDC). Int J Epidemiol. 2005; 34(6):1243–9.
14. Naess O, Sogaard AJ, Arnesen E, et al. Cohort profile: cohort of Norway (CONOR). Int J Epidemiol. 2008;37(3):481–5.
15. Stolk RP, Rosmalen JG, Postma DS, et al. Universal risk factors for multifactorial diseases: lifelines: a three-generation population-based study. Eur J Epidemiol. 2008;23(1):67–74.
16. Nilsen RM, Vollset SE, Gjessing HK, et al. Self-selection and bias in a large prospective pregnancy cohort in Norway. Paediatr Perinat Epidemiol. 2009;23(6):597–608.
17. Golding J, Pembrey M, Jones R. ALSPAC: the Avon longitudinal study of parents and children. I. Study methodology. Paediatr Perinat Epidemiol. 2001;15(1):74–87.
18. Landrigan PJ, Trasande L, Thorpe LE, et al. The national children's study: a 21-year prospective study of 100,000 American children. Pediatrics. 2006;118(5):2173–86.
19. Keil T, Kulig M, Simpson A, et al. European birth cohort studies on asthma and atopic diseases: II. Comparison of outcomes and exposures: a GA2LEN initiative. Allergy. 2006;61(9):1104–11.

20. Inskip HM, Godfrey KM, Robinson SM, Law CM, Barker DJ, Cooper C. Cohort profile: the Southampton women's survey. Int J Epidemiol. 2006;35(1):42–8.
21. Almqvist C, Adami HO, Franks PW, et al. Lifegene: a large prospective population-based study of global relevance. Eur J Epidemiol. 2011;26(1):67–77.
22. Friberg L, Engdahl J, Frykman V, Svennberg E, Levin LA, Rosenqvist M. Population screening of 75- and 76-year-old men and women for silent atrial fibrillation (STROKESTOP). Europace. 2013;15(1):135–40.
23. Wiberg B, Lind L, Kilander L, Zethelius B, Sundelöf JE, Sundström J. Cognitive function and risk of stroke in elderly men. Neurology. 2010;74(5):379–85.
24. Carlsson L, Lind L, Larsson A. Reference values for 27 clinical chemistry tests in 70-year-old males and females. Gerontology. 2010;56(3):259–65.
25. Davey Smith G, Ebrahim S, Lewis S, Hansell AL, Palmer LJ, Burton PR. Genetic epidemiology and public health: hope, hype, and future prospects. Lancet. 2005;366:1484–98.
26. Riboli E, Kaaks R. The EPIC project: rationale and study design. European prospective investigation into cancer and nutrition. Int J Epidemiol. 1997;26(Suppl 1):S6–14.
27. Calltorp J, Adami HO, Aström H, Fryklund L, Rossner S, Trolle Y, Giesecke J. Country profile: Sweden. Lancet. 1996;347:587–94.