medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Genetic, lifestyle and environmental influences on health – a Finnish biobank recall study protocol (BioRecall)

Sillanpää E.[1,2], Föhr T.[1], Kurtti E.[3,4], Aittola K.[5,6], Mäkelä J.[7], Southerington T.[7, 8], Lakka T.A. [9,10,11], Jokela T.[12,13], Ahtiainen M.[11], Laakkonen E.K.[1,14], Rantakokko M.[1,2,14], and Ravi S.[1]

[1]Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, FINLAND

[2]Wellbeing Services County of Central Finland, Jyväskylä, FINLAND

[3]Open Science Centre, University of Jyväskylä, Jyväskylä, FINLAND

[4]Faculty of Humanities and Social Sciences, University of Jyväskylä, Jyväskylä, FINLAND

[5]Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, FINLAND

[6]Unit of Public Health, Finnish Institute for Health and Welfare, Helsinki, FINLAND

[7]Finnish Biobank Cooperative (FINBB), Turku, FINLAND

[8]Faculty of Law, University of Turku, Turku, FINLAND

[9]Institute of Biomedicine, University of Eastern Finland Kuopio, FINLAND

[10]Department of Clinical Physiology and Nuclear Medicine Kuopio University Hospital Kuopio, FINLAND

[11]Foundation for Research in Health Exercise and Nutrition Kuopio, Research Institute of Exercise Medicine, Kuopio, FINLAND

[12]Central Finland Biobank, Wellbeing Services County of Central Finland, Jyväskylä, FINLAND

[13]Department of Biological and Environmental Science, Faculty of Mathematics and Science, University of Jyväskylä, Jyväskylä, FINLAND

[14]Gerontology Research Center, University of Jyväskylä, Jyväskylä, FINLAND

Corresponding Author:

Elina Sillanpää, PhD, Associate Professor

Faculty of Sport and Health Sciences, P.O. Box 35 (VIV)

FIN-40014 University of Jyväskylä, Finland.

Tel. +358 40 1429 639, e-mail: elina.sillanpaa@jyu.fi

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Introduction

Non-communicable diseases are the leading causes of premature mortality worldwide. Both genetic predispositions and environmental exposures affect disease risk. Biobanks have increased understanding of genetic predictors of these diseases, but environmental influences are expected to be larger. Individuals also create their own environments and lifestyles based on genetically regulated preferences, leading to gene-environment interactions that require large datasets to study. Biobanks may lack sufficient lifestyle and environmental data, which limits their use for studies targeting disease prevention. We present a protocol for a biobank-recall study (BioRecall) to collect lifestyle and environmental exposures and combine these with genotypes, biological samples and clinical outcomes. Central Finland Biobank donors with a diagnosis of type 2 diabetes will be invited to participate in the pilot study.

Methods and analysis

All previously genotyped donors of the Central Finland Biobank, diagnosed with type 2 diabetes, and consenting to recalls will be invited to complete an electronic questionnaire on the MyBiobank portal. According to the preliminary feasibility assessment, there are 1,386 suitable candidates. The questionnaire includes validated questions on current and past lifestyles, anthropometrics, weight loss history, health, symptoms, work characteristics, emotional state, and residential environment. Postcode information will allow spatial environmental data to be added. Genotype and related clinical data will be provided to the study in accordance with the Finnish Biobank Act and combined with questionnaire data.

Ethics and open data

The Human Sciences Ethics Committee of the University of Jyväskylä has delivered a favourable statement about the study protocol (1671/13.00.04.00/2023). Central Finland Biobank approved the research plan (no: BB24-0333-A01). Data collected will be returned to the Central Finland Biobank for research purposes with participant's consent. Permission for data usage can then be applied through standard protocols of the Fingenious® service (www.fingenious.fi). If successful, the study will be expanded to other donors and Finnish biobanks.

Key words: biobank, lifestyle, environment, electronic survey study, genetics, non-communicable diseases, recall study, exposome

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Strengths and limitations of this study

- **Low-cost, scalable approach**: Fully electronic consent and data collection concept with the potential to expand data collection to all genotyped Finnish individuals with biobank consent.
- **Extensive dataset**: The study will form an extensive dataset in terms of lifestyle and environmental exposures to investigate the prevention of non-communicable diseases. However, in this pilot phase, the sample size is limited, and only one biobank is included.
- **Automated feedback**: As the survey progresses, respondents receive feedback on their own answers in relation to national lifestyle recommendations. Automated instant feedback may improve response rates, and the feature offers possibilities for low-cost health promotion interventions.
- **Innovative physical activity assessment**: Novel questions for strength training and balance that complement WHO's Global Physical Activity Questionnaire (GPAQ) enable the assessment of compliance with current physical activity recommendations.
- **Lack of wearable data**: The concept does not include data collection using wearables.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Non-communicable diseases (NCDs) account for over 70% of deaths worldwide (1). While the risk of these diseases increases with age at the population level, individual risk and onset times vary significantly. Many common NCDs tend to cluster within families (2,3), and recent research has made significant advances in understanding their heritability and individual risk measurement (4–6). This progress is largely due to the development of large datasets from biobanks and advancements in genome-wide genetic analysis and polygenic scoring. Genome-wide risk scores can estimate heritable disease risk, aiding in the development of personalised prevention and treatment strategies (7,8). However, for many diseases, the predictive capability of genetic scores at the individual level is low (9).

Genetic predispositions do not explain population-level changes in NCD prevalences such as the rapid increase of type 2 diabetes. In Western countries, characterized by long lifespans and high socio-demographic indices, leading health risks include unhealthy lifestyle habits (such as smoking, alcohol consumption, and suboptimal dietary patterns), markers of poor metabolic health (such as high blood pressure and body mass index), and ambient particulate matter pollution (10). These findings underscore the importance of lifestyle and environmental factors, alongside biological risk factors, in predicting individual risks for NCDs.

However, current literature likely oversimplifies or underestimates the influence of various important factors due to challenges in measuring exposures and their cumulative impact on NCDs. The reductionist approach—studying exposures mainly as isolated factors or small groups of exposures—may not be optimal for studying population health. In reality, individuals are simultaneously exposed to numerous behavioral and environmental factors, some increasing risk and others providing resistance to common NCDs. The concept of the exposome, which encompasses the totality of environmental exposures from conception onwards, has gained considerable interest among researchers (11). The exposome is classified into three categories: 1) the general external environment (climate, urbanization, socio-economic factors), 2) the individual environment (lifestyle, environmental pollutants, family relations, work), and 3) the internal environment (metabolic factors, inflammation, aging) (12). Unlike genetic risk scores, environmental risk scores vary with age and over time, depending on exposure accumulation.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Another challenge in predicting individual disease risks lies in gene-environment interactions. Genetic inheritance, which influences health behaviors, can directly affect disease risk regardless of the adopted behavior (13), suggesting genetic pleiotropy—where the same genetic variation affects both health-related behavior and disease incidence (13,14). Inherited disease risk may also be nonlinearly related to environmental factors. For example, individuals tend to select their environments based on their genetic preferences and the role of heredity in obesity development is greater in obesogenic environments (15). Better identification of lifestyle and environmental risks and a deeper understanding of gene-environment interactions could help prioritize risk factors at individual and population levels and help in developing predictive strategies for public health promotion. However, large datasets that include a broad spectrum of different levels of data are rare. UK biobank includes broad phenotype data

Here, we present a protocol for the biobank recall study concept (BioRecall), which allows the collection of comprehensive data on traditional and emerging lifestyle and environmental risk factors through electronic surveys and open-source data. This data will be combined with genome data, biological samples, and register-based disease outcomes from the biobank registry. The BioRecall concept, developed in collaboration with the Finnish Biobank Cooperative (FINBB), aims to navigate complex administrative processes not yet customized for clinical or public health recall activities in Finland. It also presents practical solutions for crossing interfaces in re-contacting, consenting, and contracting activities between multiple parties.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Methods and analysis

Study design and setting

BioRecall is a cross-sectional pilot study that will be conducted in the Central Finland Biobank, Finland, and at the University of Jyväskylä, Finland, in collaboration with the FINBB and the University of Eastern Finland. This study pilots a recall study procedure (Figure 1) where lifestyle and environmental data is aimed to be collected and combined with biological and clinical data stored at the biobank registry. Lifestyle and environmental data will be collected by an electronic survey and from national open-source registers through postal code linkage.

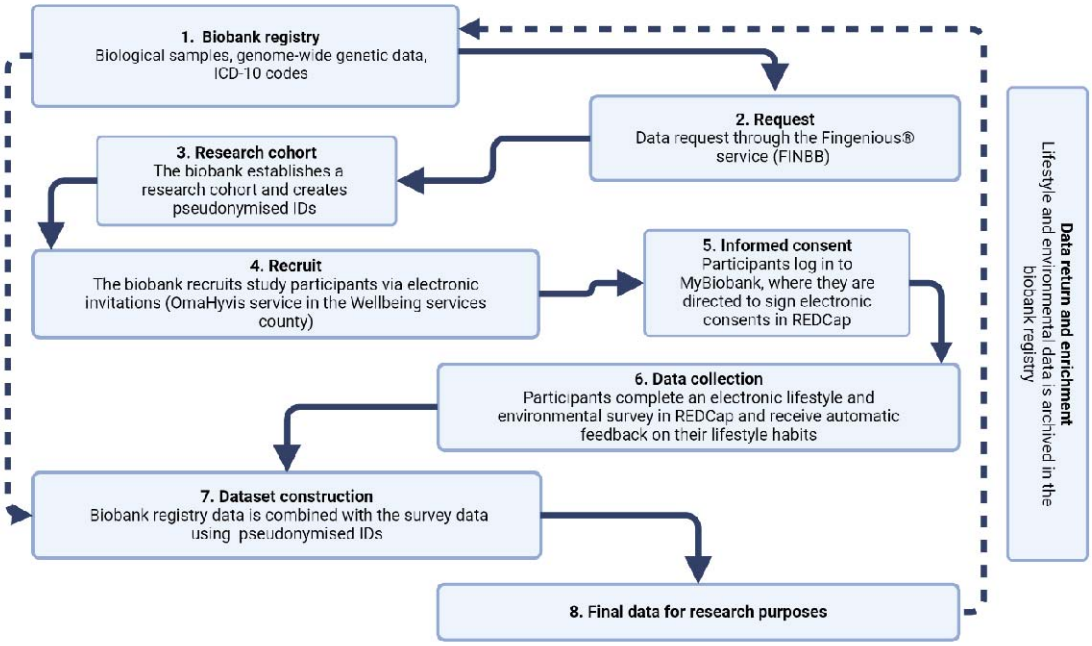


Figure 1. Key steps in the process of the biobank recall study. FINBB, Finnish Biobank Cooperative.

Eligibility criteria

In this pilot study, the target group consists of individuals with a diagnosis of type 2 diabetes (T2D). T2D is one of the most common chronic diseases worldwide and is strongly linked to lifestyle factors. Inclusion criteria for the invited participants are as follows: have a diagnosis of T2D (ICD-10 codes E11, E11.0, E11.1, E11.2, E11.3, E11.4, E11.5, E11.6, E11.7, E11.8, or E11.9), have available genome-wide DNA polymorphic data, and have given consent to be

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

recontacted for research purposes. All individuals in the Central Finland Biobank are adults (>18 years old), hence we do not apply any age restrictions. According to a feasibility assessment conducted by the Central Finland Biobank, 1,386 participants are eligible.

Recruitment

Researchers submit a data request through the Fingenious® service for genotyped participants with T2D included in the Central Finland biobank registry. Fingenious® service is a digital platform that provides researchers with access to Finnish biobank samples, data, and participant recruitment. The Fingenious® service is provided by the FINBB.

Based on the request, the Central Finland Biobank will establish a research cohort for this recall study. Before the initiation of the study, the biobank will create pseudonymized IDs for all invited participants. With the help of pseudonymized IDs, the Central Finland Biobank can link the collected survey data to the biobank's registry data, ensuring that directly identifiable information, such as personal identification numbers or contact information, does not need to be collected on the survey form. Hence, the data is pseudonymized from the beginning and participants retain their anonymity to the researchers.

The Central Finland biobank registry contains the names and postal addresses of individuals. Before sending the invitations, the biobank will verify that the invitees are alive and have consented to be recalled. The invitations are primarily sent electronically to those who have consented to electronic communication. This will be done via the online health care portal (OmaHyvis) owned by the Wellbeing services county of Central Finland, which is co-owner of the biobank. In case of missing consent for electronic contact in OmaHyvis, the invitation will be sent by mail. At this stage, the researcher will not know the identities of the individuals to whom the information is sent. The invitation will include the study information sheet and the biobank opt-out form. After the invitations are sent, the responsibility for the study lies with the researchers; for example, all participant inquiries related to the study will be directed to the researcher. We plan to send two electronic reminders within a month and, if the response rate is low, a paper questionnaire and a pre-paid return envelope.

Survey implementation

In the information sheet, participants are instructed to log in to MyBiobank (omabiopankki.fingenious.fi). MyBiobank is a portal for Finnish biobanks operated by FINBB. The portal allows study participants to securely log in, identify themselves, provide

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

consent, and access the study survey tools (16). Authentication is guaranteed through Suomi.fi, which is a secured identification system available to Finnish residents. MyBiobank links a participant to the electronic survey study, which is implemented using the REDCap platform. The REDCap is a secure web-based tool for collecting sensitive data as the collected data is stored on the University of Jyväskylä's own secure server. The questionnaire takes approximately 30–40 minutes to complete. As the respondent progresses through the survey, he/she will receive immediate feedback on his/her lifestyle habits in relation to existing Finnish national recommendations.

Informed consents

Before accessing the survey, electronic consent for participation in the study will be collected using the REDCap software. Participants will proceed with the survey only after accepting the consent. Additionally, participants will consent to the archiving of the collected survey data in the Central Finland biobank. This allows the data to be made available for future research approved by the biobank.

Participant coverage analysis

Central Finland Biobank will get the information on which individuals were invited but did not participate in the study. From these individuals, the biobank will form a pseudonymized patient cohort, which allows for examining how the groups of participants and non-participants differ, for example, in terms of age, gender, and place of residence.

Measures

Physical activity

The Global Physical Activity Questionnaire (GPAQ) (17) will be used to assess physical activity participation at work, traveling, and recreational activities along with sedentary behavior. The GPAQ was developed by the World Health Organization (WHO), and it has been shown to be a valid measure of assessing moderate-to-vigorous physical activity (18).

In addition to the GPAQ, we complemented physical activity questions with additional questions to assess the respondents' participation in muscle-strengthening and balance exercises. This allows assessing the implementation of the current physical activity recommendations (19). Complementary questions were developed by an expert group, which included sports science academics. Participants will report whether they engage in muscle-

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

strengthening activities, such as gym or home workouts, kettlebell exercises, Pilates, BodyPump, or CrossFit. In case of a positive response, the frequency and type of these activities are inquired. Furthermore, to determine whether participants meet the strength training recommendations set by the American College of Sports Medicine (ACSM) (20), their engagement in strength training exercises targeting the lower, middle, and upper body at least twice a week, using a resistance that allows them to perform each movement a maximum of 12 times per set, is assessed. Additionally, participants will report if they engage in balance exercises, such as dancing, yoga, Tai Chi, skating, ball games, cross-country skiing, or stand-up paddleboarding. If so, they will indicate how often they practice these activities each week.

Finally, the life-long physical activity history will be assessed by asking the participants to describe their physical activity level at all applicable ages from the following options: 7–12, 13-16, 17–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, and 80 or older. The response options are: 1) no exercise, 2) regular independent leisure-time physical activity, 3) regular other supervised physical activity in a sport club etc., and 4) regular competitive sport and training related to that sport. The definitions of these response options are reported elsewhere (21).

Diet and alcohol consumption

Diet quality will be measured using the Healthy Diet Index (HDI) (22), a dietary screener based on the D2D Food Intake Questionnaire (23) with slight updates. This questionnaire evaluates key food groups on both daily and weekly basis, including meals and snacks, fast food, dishes made with vegetables and legumes, fish, poultry, meat and sausages, fat and cream quality used in cooking, spreads, and salad dressings, fruits and vegetables, nuts and seeds, dairy products, grain products, processed meat products, sweets, pastries, and desserts, and sugary drinks and alcoholic beverages.

The HDI assesses overall diet quality on a scale from 0 to 100 points, with higher scores indicating better adherence to Finnish nutrition recommendations (22,24). It also evaluates seven specific components of the diet, weighted by their relative importance to an overall healthy diet: meal pattern: 10 points, grains: 20 points, fruits, and vegetables: 20 points, fats: 15 points, dairy: 10 points, fish and meat: 10 points, and snacks and treats including beverages: 15 points.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Additionally, special diets, food allergies, and the use of vitamin D and calcium supplements, which may influence diet quality and overall nutrient adequacy, are assessed. Finally, to complement the weekly alcohol consumption data collected as part of the HDI, respondents are asked to report the frequency of their alcohol consumption using the following response options: daily, several times a week, once a week, a couple of times per month, approximately once a month or more rarely, or not at all.

Anthropometrics and weight loss history

Participants will report their current weight, or their pre-pregnancy weight if currently pregnant or recently given birth, in kilograms and height in centimeters. They will also measure their waist circumference following the provided instructions. In addition, participants will report how many times they have lost at least 5 kilograms of weight and whether a healthcare professional has recommended that they lose weight. In case of a positive response to the latter question, follow-up questions will examine the settings, as well as the quality and quantity of the support provided by healthcare professionals.

Smoking and substance use

Participants will report their smoking habits using questions adapted from the Finnish Twin Cohort study (25). Additionally, information will be collected on the use of snus, nicotine pouches, cigars, cigarillos, pipes, cannabis, and other intoxicating substances.

Medical history and current health status

Participants will be presented with a list of medical diagnoses and asked to indicate any that have been confirmed by a doctor. Additionally, they will report their current medications. Finally, participants will evaluate their subjective experience of their health, physical fitness, and functional performance compared to similar-aged people.

Questions for women

Questions regarding menstrual history, parity, and the use of hormonal contraceptives or hormone replacement therapy will be directed to those who report being women. The age at menarche and whether secondary amenorrhea (defined as the absence of menses for at least three months due to reasons other than pregnancy or hormonal contraceptive use (26) has ever occurred will be self-reported by the participants. They will also provide information on their menopausal age (if applicable), specifying whether menopause occurred spontaneously

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

or as a result of medical intervention (e.g., oophorectomy, radiation therapy, or chemotherapy). Additionally, participants will report the number and timing of their pregnancies. Finally, details about the current use of hormonal contraceptives or hormone replacement therapy are inquired.

Work characteristics

Participants who have ever been employed or are currently employed will answer work-related questions. These questions are instructed to be answered based on the most recent employment. In addition to their job title, participants will be asked several questions related to their work characteristics. The number of hours per week engaged in paid employment will be inquired. Information on the work schedule will be obtained by assessing the participant's work type: fixed days, fixed nights, rotating shift with a two-shift pattern without night shift, rotating shift with a two-shift pattern with night shift, rotating shift with a three-shift pattern (25).

Occupational physical loading will be assessed using a single item: "What kind of work do you do?" The response alternatives will be: (1) mainly sedentary work, which requires very little physical activity; (2) work which involves standing and walking, but no other physical activity; (3) work which in addition to standing and walking requires lifting and carrying; and (4) heavy physical work. Participants will be able to select only one category to describe his/her work (25).

Self-rated work ability will be measured using a single item: "Let's assume that your work ability at its all-time best would be given 10 points, and 0 points would indicate that you are completely unable to work. How would you score your current work ability?". This single question measure is based on the use of the first item of the Work Ability Index's (WAI) seven-item questionnaire (27), which is also referred to as the Work Ability Score (WAS). WAS has been shown to have a similar validity but it is a simpler alternative to WAI (28).

Work stress will be assessed using effort-reward imbalance (ERI). ERI was assessed by four questions from the 10-item scale by Siegrist (29–31). Participants will respond to a 5-point Likert scale (1 = very little, 5 = very much) to one question concerning effort and three questions concerning reward. ERI will be calculated by dividing the effort score by the mean score of reward variables.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Work engagement will be assessed using three items from the Utrecht Work Engagement Scale (UWES) (32): “At my work, I feel full of energy” (vigor), “I am enthusiastic about my work” (dedication), and “Time flies when I am working” (absorption). Participants will respond to a 5-point Likert scale (1 = always, 5 = never).

Emotional states

Positive and negative affectivity will be assessed using the Internationally Reliable Short form of the Positive and Negative Affect Schedule (PANAS; with 10 adjectives: five for positive affect and five for negative affect): “Thinking about yourself and how you normally feel. How do you usually feel? Choose the most appropriate option.” Participants will respond to each item on a 5-point Likert scale (1 = does not describe me at all, 5 = describes me very well) (33).

Sleep characteristics

The Jenkins Sleep Scale (34) will be utilised to assess sleep quality. This scale consists of four items that evaluate sleep problems during the last month: trouble falling asleep, waking up several times during the night, trouble staying asleep, and waking up feeling tired. Response options for each item are on a 6-point scale from “not at all” (0 points) to “22-31 days” (5 points) (34). Information on the duration of nightly sleep and potential naps will also be inquired.

Residential environment

Factors related to the local environment will be assessed using questions modified from national Health 2011 survey (35,36). The questions examine the availability of walking and cycling paths near the home, along with access to certain daily services, such as grocery stores, supermarkets, libraries, post offices, sports facilities, and green spaces. In addition, the residential area will be defined based on the postal code reported in the questionnaire. Environmental exposures on the postal code area level will be curated from different sources using both open-source statistical and geospatial data and data received through collaboration in exposome networks. Annual average air quality and pollution exposures of the closest observation station will be retrieved from the Finnish Meteorological Institute. The National Land Survey of Finland (NLS) offers information on natural features, such as forest and agricultural areas, rivers, lakes, and built environments (roads, buildings). Access to green and blue spaces and their respective quality can be estimated by e.g. tree coverage,

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

vegetation, and biodiversity indices. The Finnish Environment Institute (SYKE) offers information on land cover, land use, and urban zones. Additionally, data are openly available for Central Finland from the European Environmental Agency (EEA; Europe-wide Urban Atlas) and USGS Landsat or Copernicus Open Access Hub (Sentinel-2 Data). Personal lifestyle etc. exposures will be collected using the questionnaire. Socioeconomic characteristics e.g. demographics, education, income, employment, housing, crime and social inequality, political orientation and voting behaviour in the postal code area will be obtained from Statistics Finland.

Background characteristics

In addition to age and sex, the questionnaire will cover questions about the economic well-being of the household, education, occupational group, and employment status.

Automatic feedback

We have developed an automatic feedback function for REDCap software that provides participants with feedback on their lifestyle habits in relation to national lifestyle recommendations for health promotion. This was done to make the questionnaire more engaging for participants and to motivate invited individuals to complete the questionnaire; thus, increasing the response rate. Feedback will be provided from the following parts of the questionnaire: physical activity, dietary habits, waist circumference, work stress, and sleep.

The feedback on physical activity summarizes the participant's weekly amount of exercise, categorized into moderate and vigorous activity as well as strength and balance training. Additionally, it provides information on recommended levels of physical activity based on current guidelines (19).

The feedback on dietary habits uses HDI scoring and provides automated feedback based on nutrition recommendations to support health-promoting diet (24,37). This feedback system has been refined through prior research projects (38,39).

In the feedback, participants are informed about healthy waist circumference and its risk thresholds in relation to their reported waist circumference. Alongside their reported sleep duration, they are informed about the recommended sleep duration and the potential risks associated with sleep deprivation.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The feedback on work stress is based on ERI (29–31). The participant receives an individual-level ERI ratio score (the ratio of the self-assessed effort score and the mean of the self-assessed reward scores) and an interpretation of the score. The ratio scores of ERI range from 0.2 to 5.0. An imbalance between work efforts and rewards is present when the ratio score of ERI $\neq 1$. The ratio score of ERI <1 indicates an imbalance in favour of rewards and ratio score >1 indicates an imbalance in favour of effort (40).

Schedule

Data collection will start in February 2025. Data will be used to publish a manuscript about lifestyles and weight management among Type 2 Diabetic patients in the Central Finland Wellbeing Services County area. This will be drafted during the year 2025. Data will also be used in several M.Sc. thesis, which preparation will start in the fall of 2025. These theses will be published in the Jyväskylä University Digital Repository JYX. If the protocol is successful, data collection will be extended to all donors of the Central Finland biobank in autumn 2026.

Sample size calculation and statistical analysis plan

Since the primary focus of the project is on piloting the operational model, no formal sample size calculation has been performed, and no detailed statistical analysis plan has been developed.

Ethics and dissemination

The Human Sciences Ethics Committee of the University of Jyväskylä has given favorable consent to the study protocol (1671/13.00.04.00/2023). Central Finland Biobank has approved the study (BB24-0333-A01).

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Discussion

One of the greatest challenges in public health promotion is to recognize individuals at risk for developing diseases. This would help to direct personalised preventative actions before disease manifestation. This ambitious and important goal is not possible to achieve without proper datasets. Large-scale biobank datasets include data from hundreds of thousands of genotyped individuals along with register-based disease endpoints and biological samples. Complementing biobank data with lifestyle and environmental risk factors using the biobank recalls would offer possibilities to study the complex interactions between genetic predispositions, lifestyle choices, and environmental exposures. Customized clinical recall studies for Finnish biobank donors have been piloted recently as part of a large public partner project the FinnGen (16,41,42). These pilot studies, which have been focused on neurocognitive diseases (41,42), have shown that the recall concept is feasible but also highlighted the need for developing a more centralized recall procedure for Finnish biobanks (16) and testing recall concepts with different subgroups. The BioRecall study concept attempts to address this need.

The BioRecall concept presented in this study offers a cost-effective and fully electronic approach to data collection, which is crucial for large-scale epidemiological research. This pilot is conducted in the Central Finland Biobank with the genotyped individuals having type 2 diabetics, but the concept has the potential to expand data collection to all genotyped Finnish individuals with biobank consent (10% of the population). Another primary strength of the BioRecall concept is its versatility in collecting comprehensive data on genetic information, lifestyles, environmental exposures, and clinical outcomes. This holistic exposome approach addresses the current challenges in understanding the multifactorial nature of NCDs. It is expected that larger datasets with novel machine-learning tools will help to identify key risk factors and protective factors at both individual and population levels.

The biobank recall concept presented here may allow longitudinal monitoring of participants in a cost-effective manner. This capability is essential for tracking changes in health behaviours and outcomes over time, providing valuable data for the prevention and management of NCDs. Collaboration with the Wellbeing services counties can facilitate the further development of this concept, ensuring that the data collected is used to improve public

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

health strategies and outcomes. This cooperation can lead to the development of more targeted and effective health policies and interventions. The potential for health promotion interventions through feedback activities further enhances the utility of this approach, offering opportunities for personalised health recommendations and interventions. The feedback tool is particularly advantageous in addressing the common issue of low response rates in lifestyle surveys (16). Automated personalised feedback mechanisms are expected to improve participant engagement and response rates, which is a significant advancement over traditional survey methods. Returning personal health risk information may also be used to promote lifestyle changes.

However, there are challenges to consider. Based on prior evidence among Finnish participants postal contact may yield better response rates than contacting via a healthcare portal and reminders may be needed (16). However, a similar concept with automatic feedback and among the current patient group has not been tested before. Comprehensive data can be obtained using a broad questionnaire. On the other hand, the broad scope of the questionnaire may result in low adherence due to the time required to complete it. The response rates in the piloting phase will show how balancing with these challenges succeeded. An important limitation of this study is that in this pilot phase, the sample size is limited, and only one regional biobank is included. However, all other software, except OmaHyvis, which is utilised for the first contact, is expected to be available at the national level and all wellbeing services counties have a system allowing electronic contact for their residents.

In conclusion, the biobank recall concept represents a promising and innovative approach to epidemiological research. Its ability to integrate versatile data types, respond to current survey challenges, and provide cost-effective longitudinal monitoring positions it as a valuable tool for advancing our understanding of NCDs and improving public health.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Acknowledgements

We thank all biobank donors for their valuable contributions to science.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Funding

This project is supported by the JYU.Well strategic funding received from the University of Jyväskylä (to ES) and the research infrastructure funding received from the Research Council of Finland (grant number: 358894 to ES) as part of the BBMRI.fi consortium. ES is supported by the Research Council of Finland (grant numbers: 341750, 346509, and 361981), Juho Vainio Foundation and Päivikki and Sakari Sohlberg Foundation. SR and TF are supported by Juho Vainio Foundation. MR is supported by the Research Council of Finland (grant number: 330185). KA is partially supported by the European Regional Development Fund through the Regional Council of Northern Savo (No. A80407) for the development of FOODNUTRI infrastructure and by the Joint Action on Cardiovascular Diseases and Diabetes (JACARDI), funded by the EU4Health Programme (Grant Agreement No. 101126953).

Data availability

All Finnish biobanks are members of the BBMRI.fi infrastructure (https://www.bbmri.fi). The FINBB (https://finbb.fi/) is the coordinator of BBMRI.fi and the Finnish node of the Pan-European BBMRI-ERIC research infrastructure. The Finnish biobank data can be accessed through the Fingenious services (https://site.fingenious.fi/en/) managed by FINBB. Novel physical activity questions developed in the project as well as their REDCap implementation are available from the corresponding authors upon a reasonable request. The HDI questionnaire is available and administered by the Finnish Institute for Health and Welfare (THL). Developed feedback based on the HDI can be obtained upon a reasonable request from the FOODNUTRI project (KA).

Competing interests: None.

Contributions:

Original idea ES, SR, TF, MR, EL, JA

Funding ES

Resources ES, TJ, MA, EK, JA, JM, TS, JM and TS on behalf of the FINBB

Concept development ES, SR, TF, MR, EL, KA, EK, TAL, JA

Writing the original draft SR, TF, ES, KA

Revising the intellectual content MR, EL, KA, EK, TAL, TJ, MA, JA, JM, TS

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Data management plan ES, SR, TF, MA, TJ

Data collection ES, SR, TF, MA, TJ

Administration and study coordination SR, TF, JM and TS on behalf of the FINBB

Project management and supervision ES

Legal advice TS

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Noncommunicable diseases [Internet]. [cited 2024 Dec 17]. Available from: https://www.who.int/health-topics/noncommunicable-diseases

2. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. Journal of Internal Medicine. 2002;252(3):247–54.

3. Marenberg ME, Risch N, Berkman LF, Floderus B, Faire U de. Genetic Susceptibility to Death from Coronary Heart Disease in a Study of Twins. New England Journal of Medicine. 1994 Apr 14;330(15):1041–6.

4. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. NatGenet. 2018;50(9):1219–24.

5. King A, Wu L, Deng HW, Shen H, Wu C. Polygenic risk score improves the accuracy of a clinical risk score for coronary artery disease. BMC Med. 2022 Nov 7;20(1):385.

6. Mars N, Koskela JT, Ripatti P, Kiiskinen TTJ, Havulinna AS, Lindbohm JV, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. NatMed. 2020;26(4):549–57.

7. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. NatRevGenet. 2018;19(9):581–90.

8. Lankinen MA, Nuotio P, Kauppinen S, Koivu N, Tolonen U, Malkki-Keinänen K, et al. Effects of Genetic Risk on Incident Type 2 Diabetes and Glycemia: The T2D-GENE Lifestyle Intervention Trial. The Journal of Clinical Endocrinology & Metabolism. 2025 Jan 1;110(1):130–8.

9. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Med. 2020 May 18;12(1):44.

10. GBD 2016 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017 Sep 16;390(10100):1345–422.

11. Vermeulen R, Schymanski EL, Barabási AL, Miller GW. The exposome and health: Where chemistry meets biology. Science. 2020 Jan 24;367(6476):392–6.

12. Siroux V, Agier L, Slama R. The exposome concept: a challenge and a potential driver for environmental health research. European Respiratory Review. 2016 Jun 1;25(140):124–9.

13. Tynkkynen NP, Törmäkangas T, Palviainen T, Hyvärinen M, Klevjer M, Joensuu L, et al. Associations of polygenic inheritance of physical activity with aerobic fitness, cardiometabolic risk factors and diseases: the HUNT study. Eur J Epidemiol. 2023 Sep 1;38(9):995–1008.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

14. Sillanpää E, Palviainen T, Ripatti S, Kujala UM, Kaprio J. Polygenic Score for Physical Activity Is Associated with Multiple Common Diseases. Medicine & Science in Sports & Exercise. 2022 Feb;54(2):280.

15. Jackson SE, Llewellyn CH, Smith L. The obesity epidemic – Nature via nurture: A narrative review of high-income countries. SAGE Open Med. 2020 Apr 28;8:2050312120918265.

16. Rodosthenous RS, Niemi MEK, Kallio L, Perala M, Terho P, Knopp T, et al. Recontacting biobank participants to collect lifestyle, behavioural and cognitive information via online questionnaires: lessons from a pilot study within FinnGen. BMJ Open. 2022 Oct 5;12(10):e064695.

17. Global physical activity questionnaire (GPAQ) [Internet]. [cited 2024 Dec 17]. Available from: https://www.who.int/publications/m/item/global-physical-activity-questionnaire

18. Cleland CL, Hunter RF, Kee F, Cupples ME, Sallis JF, Tully MA. Validity of the Global Physical Activity Questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behaviour. BMC Public Health. 2014 Dec 10;14(1):1255.

19. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med. 2020 Dec;54(24):1451–62.

20. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. MedSciSports Exerc. 2007;39(8):1423–34.

21. Kovanen V, Aukee P, Kokko K, Finni T, Tarkka IM, Tammelin T, et al. Design and protocol of Estrogenic Regulation of Muscle Apoptosis (ERMA) study with 47 to 55-year-old women's cohort: novel results show menopause-related differences in blood count. Menopause. 2018;25(9):1020–32.

22. Lindström J, Aittola K, Pölönen A, Hemiö K, Ahonen K, Karhunen L, et al. Formation and Validation of the Healthy Diet Index (HDI) for Evaluation of Diet Quality in Healthcare. International Journal of Environmental Research and Public Health. 2021 Jan;18(5):2362.

23. Hemiö K, Pölönen A, Ahonen K, Kosola M, Viitasalo K, Lindström J. A Simple Tool for Diet Evaluation in Primary Health Care: Validation of a 16-Item Food Intake Questionnaire. International Journal of Environmental Research and Public Health. 2014 Mar;11(3):2683–97.

24. ravitsemussuositukset_2014_fi_web_versio_5.pdf [Internet]. [cited 2024 Dec 17]. Available from: https://www.ruokavirasto.fi/globalassets/teemat/terveytta-edistava-ruokavalio/kuluttaja-ja-ammattilaismateriaali/julkaisut/ravitsemussuositukset_2014_fi_web_versio_5.pdf

25. Kaprio J. The Finnish Twin Cohort Study: An Update. Twin Research and Human Genetics. 2013;16(1):157–62.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

26. Melin A, Tornberg AB, Skouby S, Faber J, Ritz C, Sjödin A, et al. The LEAF questionnaire: a screening tool for the identification of female athletes at risk for the female athlete triad. Br J Sports Med. 2014 Apr;48(7):540–5.

27. Ilmarinen J, Tuomi K, Klockars M. Changes in the work ability of active employees over an 11-year period. Scandinavian Journal of Work, Environment & Health. 1997;23:49–57.

28. El Fassi M, Bocquet V, Majery N, Lair ML, Couffignal S, Mairiaux P. Work ability assessment in a worker population: comparison and determinants of Work Ability Index and Work Ability score. BMC Public Health. 2013 Apr 8;13(1):305.

29. Siegrist J. Adverse health effects of high-effort/low-reward conditions. Journal of Occupational Health Psychology. 1996;1(1):27–41.

30. Siegrist J, Wege N, Pühlhofer F, Wahrendorf M. A short generic measure of work stress in the era of globalization: effort–reward imbalance. Int Arch Occup Environ Health. 2009 Aug 1;82(8):1005–13.

31. Juvani A, Oksanen T, Salo P, Virtanen M, Kivimäki M, Pentti J, et al. Effort-reward imbalance as a risk factor for disability pension: the Finnish Public Sector Study. Scandinavian Journal of Work, Environment & Health. 2014;40(3):266–77.

32. The Measurement of Work Engagement With a Short Questionnaire: A Cross-National Study - Wilmar B. Schaufeli, Arnold B. Bakker, Marisa Salanova, 2006 [Internet]. [cited 2024 Dec 17]. Available from: https://journals.sagepub.com/doi/abs/10.1177/0013164405282471?casa_token=6pg1wE2a3w4AAAAA:aXNSrEkpzFECl8kse60C0fi0ZhtZawYoiWqiWfEaV2wGwNb22dRmKVgGaWjTu07KncYWaPX3DBFFTQ&casa_token=5TAXvWOXQNQAAAAA:O6M2Jcl-EvPvOqIkGqiEKwmVOX81UflfmeQWMixcGsWPtJWKWcwtLK91T1HvaaMyguOFInaVoyvkbA

33. Thompson ER. Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). Journal of Cross-Cultural Psychology. 2007 Mar 1;38(2):227–42.

34. Jenkins CD, Stanton BA, Niemcryk SJ, Rose RM. A scale for the estimation of sleep problems in clinical research. Journal of Clinical Epidemiology. 1988 Jan 1;41(4):313–21.

35. Health 2011 Survey - Methods [Internet]. THL; 2016 [cited 2024 Dec 18]. Available from: https://www.julkari.fi/handle/10024/130780

36. Finnish Institute for Health and Welfare (THL), Finland [Internet]. 2024 [cited 2024 Dec 18]. Health 2011 -forms - THL. Available from: https://thl.fi/en/research-and-development/research-and-projects/health-2000-2011/forms/health-2011-forms

37. Rune B, Rikke A, Kristoffer A Erik, Juel C Jacob, Hanna E, Maijaliisa E, et al. Nordic Nutrition Recommendations 2023: Integrating Environmental Aspects. Nordic Council of Ministers; 2023. 388 p.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.02.25319875; this version posted January 2, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

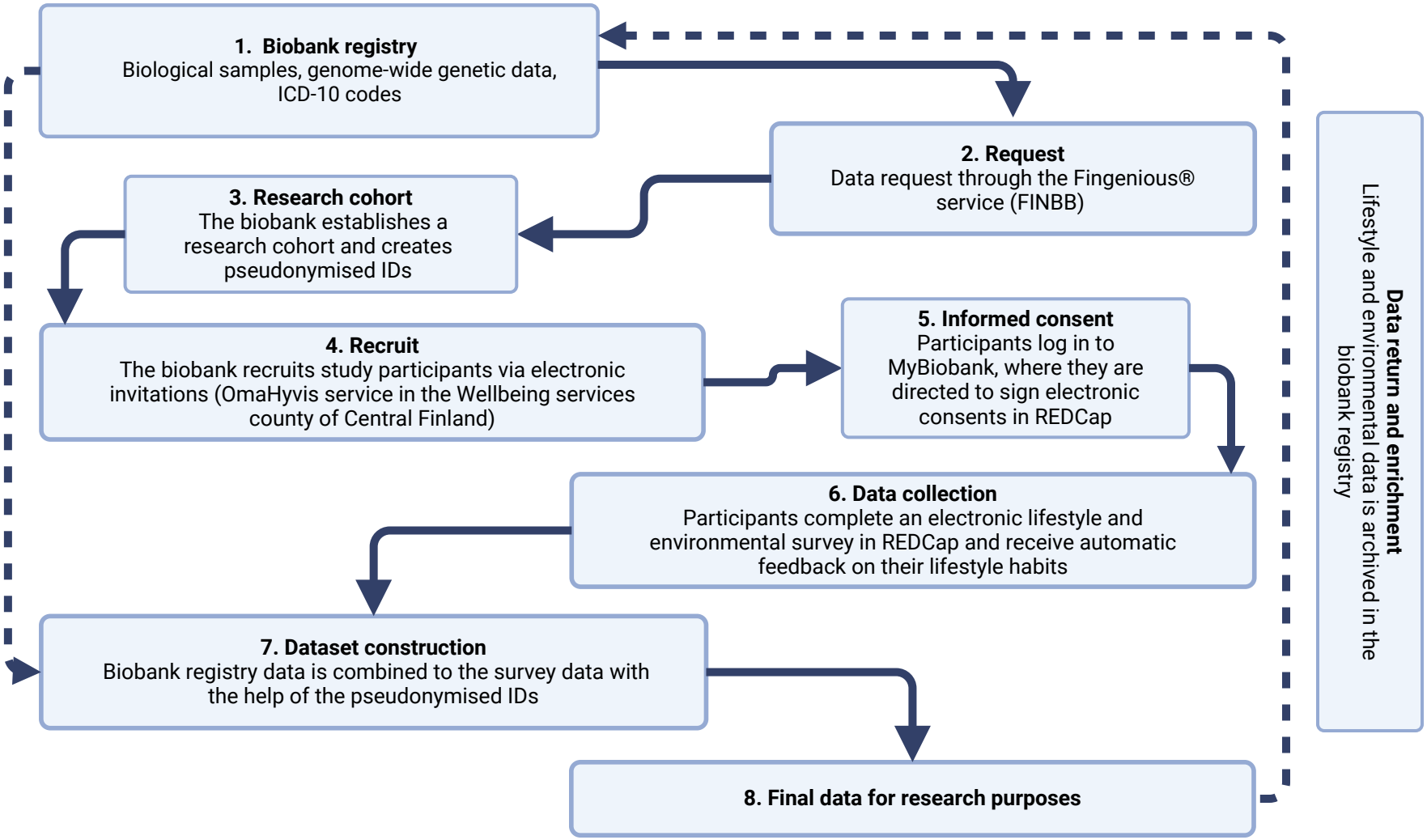
38. Nutrition, Lifestyle Factors, Health and Genes Study [Internet]. Itä-Suomen biopankki. [cited 2024 Dec 18]. Available from: https://ita-suomenbiopankki.fi/tutkimus/nutrition-lifestyle-factors-health-and-genes-study/

39. Kuopiossa kehitetään ravitsemustieteen tutkimusinfrastuktuuria | Itä-Suomen yliopisto [Internet]. [cited 2024 Dec 18]. Available from: https://www.uef.fi/fi/artikkeli/kuopiossa-kehitetaan-ravitsemustieteen-tutkimusinfrastuktuuria

40. Stanhope J. Effort–Reward Imbalance Questionnaire. Occupational Medicine. 2017 Jun 1;67(4):314–5.

41. Julkunen V, Schwarz C, Kalapudas J, Hallikainen M, Piironen AK, Mannermaa A, et al. A FinnGen pilot clinical recall study for Alzheimer's disease. Sci Rep. 2023 Aug 3;13(1):12641.

42. Vuoksimaa E, Saari TT, Aaltonen A, Aaltonen S, Herukka SK, Iso-Markku P, et al. TWINGEN: protocol for an observational clinical biobank recall and biomarker cohort study to identify Finnish individuals with high risk of Alzheimer's disease. BMJ Open. 2024 Jun 1;14(6):e081947.


1. Biobank registry
Biological samples, genome-wide genetic data, ICD-10 codes
2. Request
Data request through the Fingenious® service (FINBB)
3. Research cohort
The biobank establishes a research cohort and creates pseudonymised IDs
4. Recruit
The biobank recruits study participants via electronic invitations (OmaHyvis service in the Wellbeing services county of Central Finland)
5. Informed consent
Participants log in to MyBiobank, where they are directed to sign electronic consents in REDCap
6. Data collection
Participants complete an electronic lifestyle and environmental survey in REDCap and receive automatic feedback on their lifestyle habits
7. Dataset construction
Biobank registry data is combined to the survey data with the help of the pseudonymised IDs
8. Final data for research purposes
Data return and enrichment
Lifestyle and environmental data is archived in the biobank registry