medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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-ND 4.0 International license .

Re-contacting biobank participants: lessons from a pilot study within FinnGen

Rodosthenis S. Rodosthenous[1], Mari E.K. Niemi[1], Lila Kallio[2], Merja Perälä[2], Perttu Terho[2], Theresa Knopp[3], Eero Punkka[3], Enni M. Makkonen[4], Paula Nurmi[4], Johanna Mäkelä[5], Pauli Wihuri[5], Marco Hautalahti[5], Corianna Moffatt[6], Paolo Martini[6], Laura Germine[6], Viola A. Mäkelä[1], Oona A. Karhunen[7], Jari Lahti[7], Tero S. Hiekkalinna[8], Tero Jyrhämä[1], Huei-Yi Shen[1], FinnGen, Heiko Runz[9], Aarno Palotie[1], Markus Perola[8], Andrea Ganna[1] †

Affiliations

[1]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland.
[2]Auria Biobank, University of Turku and Turku University Hospital, Turku, Finland.
[3]Helsinki Biobank, University of Helsinki, and Hospital District of Helsinki and Uusimaa, Helsinki, Finland.
[4]Finnish Clinical Biobank Tampere, Tampere University Hospital, Tampere, Finland.
[5]FinBB, Finnish Biobank Cooperative, Helsinki, Finland.
[6]The Many Brains Project, Belmont, MA, United States.
[7]Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, 00014 Helsinki, Finland.
[8]Finnish Institute for Health and Welfare (THL), Helsinki, Finland.
[9]Translational Biology, Research and Development, Biogen Inc., Cambridge, MA, USA.

†Corresponding author

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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/10.1101/2022.04.07.22273501; this version posted April 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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-ND 4.0 International license .

Abstract

Background Biobank studies are well suited to identify novel disease risk factors. However, these studies often offer limited insights into participants' lifestyle, behavioral, or cognitive characteristics, among others. FinnGen is a nation-wide biobank project in Finland and one of the largest in the world that has successfully linked genetic and register-based information for 259,578 individuals (N=500,000 by 2023); however, it lacks cognitive, behavioral, and lifestyle measurements.

Methods We conducted a pilot study to evaluate the feasibility of obtaining cognitive, behavioral, and lifestyle information by re-contacting 5,995 participants from three FinnGen-participating Biobanks. We created an online portal to allow participants to securely log-in, provide consent, and access the study survey tools.

Results The overall participation rate was 18.6% (23.1% among individuals aged 18-69). A second reminder letter yielded an additional 9.7% participation rate in those who did not respond to the first invitation. Re-contacting participants via an online healthcare portal yielded lower participation than re-contacting via physical letter. The completion rate of questionnaire and cognitive tests was high (92% and 85%, respectively), and measurements were overall reliable among participants. For example, the correlation (*r*) between self-reported body mass index and that collected by the biobanks was 0.92.

Conclusions In summary, this pilot suggests that re-contacting FinnGen participants with the goal to collect a wide range of cognitive, behavioral and lifestyle information without additional engagement, results in a low participation rate, but with reliable data. We suggest that such information be collected at enrollment, if possible, rather than via post-hoc re-contacting.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

Introduction

Biobank studies are being set up across the world[1-9]. These studies are characterized by the possibility to link biological measurements, such as DNA, proteins, and metabolites with extensive longitudinal health information. Health outcomes are often obtained by linkage with electronic health records or national registers, as in the case of the Nordic countries. Biobank studies have allowed extensive characterization of the genetic architecture of common diseases[10-12], provided novel epidemiological insights[13 14] and identified novel disease markers[15].

The UK Biobank study[1 16], one of the largest and most widely used biobanks, has also collected lifestyle, behavioral information, and anthropometric measurements for all their participants. This was possible, at the cost of collecting a non-representative sample of the population[17-19], because all participants were enrolled via in-person visit to one of 22 the recruitment centers. However, most of the other biobank studies use a different approach collecting samples via the healthcare system or via other approaches that do not entail an extensive in-person examination at recruitment. For this reason, it is often difficult to obtain extensive behavioral and lifestyle information from biobank participants and re-contacting after recruitment is required.

The FinnGen study is a public-private partnership research project combining genotype data generated from Finnish biobank samples and digital health record data from Finnish health registers (https://www.finngen.fi/en) aiming to provide new insight in disease genetics[20]. Up to 500,000 participants will be part of FinnGen and > 350,000 have already been genotyped and linked with comprehensive health registers. Participants to the FinnGen study are recruited by several biobanks across Finland and all participants have signed a broad Biobank consent in accordance with the Finnish Biobank Act. Participants are enrolled because they are part of previous research studies or via hospitals and blood donation centers, but no extensive behavioral and lifestyle information is systematically collected for everyone at recruitment. Overall, FinnGen does not suffer from the “healthy volunteer effect” and, on the contrary, is enriched for individuals that are more likely to have been in contact with the healthcare system. Samples from consented individuals can be used across many research projects, if approved by the biobanks. Thus, individuals are not actively informed of their participation in FinnGen. Here we describe the results of a pilot study that aimed at re-contacting FinnGen participants with the goal to collect cognitive, behavioral and lifestyle information via a custom-made online platform.

Results

In this genetically informed pilot study, we invited 5995 FinnGen participants across three Biobanks who were selected according to the inclusion criteria shown in **Figure 1**. Out of 5995 individuals invited, 1115 individuals accessed the study online portal and successfully completed the general questionnaire with an overall participation rate of 18.6% **(Table 1)**. The

medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

participation rate in the 18-69 years old age group was slightly higher at 23.1% (1018/4399). The highest overall participation rate across all ages was observed in Helsinki Biobank with 23%, followed by Tampere Biobank with 17% and Auria Biobank with 15.8%. Lower participation rate in Tampere and Auria biobanks is partially explained by the higher age range of the invited individuals. Among the 18-69 years old range, the participation rate was similar across all biobanks. The highest participation rate was consistently observed in the 40-69yo age group across all biobanks **(Table 1 and Supplementary Table 1)**. The participation rates among individuals who were ≥70yo was substantially lower (6.1%) as compared to the <70yo group (23.1%). We also tested if sending a follow-up reminder letter would significantly improve participation rate among non-responders. To assess this, we sent a reminder by post to 422 individuals from Tampere (n=240) and Auria (n=182) Biobanks who were invited initially but did not participate in the study during the first phase. The overall response rate for this second letter was 9.7% (41/422), higher in Tampere Biobank (11.5%, 28/240) than Auria Biobank (7.1%, 13/182).

Using FinnGen data, we further investigated whether any differences in basic characteristics and disease prevalence among invitees could have potentially impacted their participation in the study. No significant age difference was observed between invitees who participated (54.9±13.5, [mean±SD]) versus those who did not participate (55.6±15.6, [mean±SD]) in the study (p=0.08); however, we found that invitees who participated had a lower BMI (28.0±6.2, [mean±SD]) compared to those who did not participate (28.5±6.2, [mean±SD]) in the study (p=0.02). We also found that invitees with a university degree or higher were more likely to participate (24%, 433/1810) as compared to invitees who had a lower education (17%, 519/3099). The difference in the two proportions between the groups was statistically significant (p<0.0001). When assessing health information using FinnGen data, we found that a significantly higher (p<0.0001) proportion of invitees who did not participate were previously diagnosed with hypertension (24.9%, 1210/4854) as compared to invitees who participated (16.9%, 186/1101). No differences in other disease prevalences, such as asthma, arthrosis, depression, and immune bowel disease were observed between the two groups.

In addition, we compared the participation rate between individuals invited via physical letters and those invited electronically via the Tampere Biobank OmaTays healthcare portal. Out of 1995 invitees, 759 received their invitations via the OmaTays healthcare portal and 1236 via physical letter. The participation rate was 12.1% among those invited via the OmaTays compared to 21.4% among those who were invited via a physical letter from Tampere Biobank. Data retrieved from the OmaTays healthcare portal showed that 451/759 (59.5%) of those who were invited electronically did not open the invitation at all. Among the rest who viewed the invitation, ~30% (92/308) went on to participate in the study. Individuals invited in this pilot study were consented in the first place via the Biobank consents, in accordance with the Finnish Biobank Act. Thus, they could decide to withdraw their consent for their samples to be used in research studies. In this study, a 0.3% of the invited individuals contacted any of the three biobanks to withdraw their consent.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

Among individuals who started answering the online questionnaire, the completion rate was 92%. The completion rate for the entire battery of cognitive tests was 85%, despite the length of the tests (estimated to be 30-40 min). However, substantially fewer people started the cognitive test compared to the general questionnaire (n=699 vs n=1115). Among 143 participants who provided their experience through the feedback questionnaire, the median NPS score was 8, indicating that most individuals were happy with the current design. More specifically, 121/143 (85%) gave a score >7, which is interpreted as good/very good. On the contrary, 22/143 (15%) scored the questionnaire with a 6 or less, indicating they had a negative experience (**Figure 2**). It is worth mentioning that only 53% (73/137) had previous experience with online questionnaires, whereas for 47% (64/137) of the participants this was their first online questionnaire.

For a subset of FinnGen participants, information about body mass index (BMI) was available and extracted from electronic health records or in-person visits. Thus, for 673 individuals, we could compare the self-reported BMI data obtained from the questionnaire with those previously available in FinnGen and extracted from electronic health records or in-person measurements. We found a high correlation between the two measurements (Spearman's r=0.92) despite these being collected at different time points and with different approaches (**Figure 3**). The mean±SD of the self-reported and FinnGen collected BMI data was 28.1±6.0 and 28.0±6.2, respectively. Similarly, we could compare self-reported disease diagnoses from the questionnaire with disease information available in FinnGen and obtained from national health registers **(Supplementary Table 2)**. In the questionnaire, participants were asked whether they had been professionally diagnosed with any of the listed diseases. Overall, we noticed a higher prevalence for most diseases when self-reported compared to when extracted from the health registers. For example, asthma had a prevalence of 11.3% (124/1101) when data was extracted from the health registers compared to 18.5% (204/1101) when self-reported via the online questionnaire. We observed similar trend with vitiligo: The prevalence of vitiligo from health register data was 0.1% (1/1101) *vs* 1.7% (19/1101) when self-reported. On the other hand, in the case of primary sclerosing cholangitis, we found the same prevalence of 0.7% (8/1101) between the health register and self-reported data. For some diseases, a higher prevalence is expected because health registers do not provide good coverage if diagnosed in primary care. On the other hand, some individuals might over-report or mis-report disease diagnoses. A head-to-head comparison between self-reported and register-based diseases is challenging because different combinations of diagnostic and medication codes can be used to define the same disease from health registers.

Last, access to genetic data for both participants and non-participants in this pilot study allowed us to test if our strategy of recruiting individuals in the top and bottom 3% of a polygenic score (PS) for cognitive abilities resulted in differential study participation between the two groups. We found that a higher PS for cognitive performance was associated with higher participation. In particular, 19.4% of individuals who participated in the pilot study were in the top 3% of the PS for cognitive performance compared to 16.9% of those that did not participate. The difference in the two proportions between the groups was statistically significant (p=0.01) but modest **(Figure 4)**.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

Discussion

Several biobank studies are designed to prioritize scalable and economical sample collection by using existing biological banks or hospital-based recruitment strategies. Often, extensive characterization of lifestyle, cognitive and behavioral information is done *a posteriori* by re-contacting study participants. In this pilot study we have proposed and tested a strategy to re-contact participants in one of the largest biobank studies in the world: FinnGen. To this goal, we have established a scalable re-contacting process and designed an online re-contacting platform (i.e., OmaBiopankki) for secure identification of participants. The platform can now be utilized as a benchmark for future re-contacting studies in Finland.

Despite declining response rates in population-based surveys globally[21], research studies conducted in Finland have shown a high participation rate (>50%)[22-25]. In this pilot study, we have observed a lower participation rate of 23% in the 18-69 years old age range, which may be explained by several reasons. First, all participants provided a general consent to their biobank that covers current studies such as FinnGen and a wide range of possible future research studies. Therefore, consented individuals may not be aware of their participation or be directly engaged at the time of contacting on behalf of the FinnGen study. Second, this pilot study was designed to capture a broad range of cognitive, behavioral, and lifestyle information from invited individuals and not to target any specific disease(s). The three biobanks included in this pilot recruit individuals that are hospitalized or have been directly in contact with the healthcare system. Thus, studies that target specific diseases directly relevant to each participant's own health may result in higher participation rates. Third, for the same reason, consenting participants are likely to be sicker than the general population, which may subsequently impact their participation in such studies. In fact, lower participation rates among less healthy individuals are well documented[26 27].

As part of our efforts to boost the participation rate in this study, we tested several methods. Initially, we restricted the age group of invitees to under 70 years old because we observed a significantly lower participation rate among those over 70 years old. Based on the feedback we got through our study contact email and phone helpline, we believe that this difference was mainly due to the limited access or lack of familiarity with the use of the internet and mobile banking applications required for secure authentication among older individuals. In addition, we sent a follow-up letter as a reminder to a subset of non-responders to the initial invitation and observed a significant boost in response rate (9.7%). Our finding is very comparable to Harrison's et. al. study that found a 9% increase in participation rates after sending a reminder letter to non-responders[28]. Another study by Smith et.al., also found that sending follow-up letters significantly increases the chance of response among non-responders to the initial invitation[29]. We also evaluated response rates between sending a physical or an electronic letter and found that invitees who received a physical letter were much more likely to respond to the survey. In a recent study, researchers tested the same invitation methods and found a striking difference in participation rates when sending a physical (26.8%) or an online (1.8%) invitation[28].

medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

To the best of our knowledge, this is one of the first studies that assessed participation rate in relation to the cognitive PS score of participants. To further support our finding that individuals with a higher cognitive PS score are more likely to participate in online surveys, we also compared the education level between those who participated compared to those who did not participate in the study. We found that invitees who had a university degree or higher were more likely to participate (24%, 433/1810) as compared to invitees who had a lower education (17%, 519/3099).

In conclusion, this pilot suggests that re-contacting individuals that have consented to be part of a biobank study with the goal to collect a wide range of cognitive, behavioral and lifestyle information can be challenging and may result in lower-than-expected participation. We speculate that returning some tangible incentive and/or relevant health information to participants might improve participation rates. Future studies are warranted to test this hypothesis and other strategies to improve participation in such survey research studies. Nonetheless, we suggest that cognitive, behavioral and lifestyle information be collected, whenever possible, at enrollment rather than via a post-hoc re-contacting process.

Methods

Study population

We used data from FinnGen release R6, which included 259,578 individuals with genetic information available. According to the Finnish Biobank Act (688/2012), biobanks may re-contact a person who has given such permission in his/her biobank consent. Three biobanks (Helsinki, Auria and Tampere) encompassing 100,040 FinnGen participants (Helsinki n=58,518; Auria n=29,159; Tampere n=12,363) participated in the pilot study. We included individuals that were 18 years or older at the time of the initiation of the study (Feb 2021). For Helsinki biobank, we restricted inclusion to individuals younger than 70 years of age. We excluded individuals with severe neuropsychiatric disease, or cognitive or physical disabilities, as we expected the participation rate to be lower in this group. More specifically, we excluded individuals with any of the following diagnoses as obtained from the health register data: 1) Any dementia (ICD-10: G30,F051,F00-03); 2) Neurodegenerative diseases (ICD-10: G310-312, G318-319); 3) Parkinson's disease (ICD-10: G20) 4); Multiple sclerosis (ICD-10: G35); 5) Schizophrenia, schizotypal and delusional disorders (ICD-10: F20); 6) Mental retardation (ICD-10: F70-73, F79-78); 7) Stroke (ICD-10: I60-64); 8) Transient ischemic attack (ICD-10: G45); 9) Visual impairment including blindness (binocular or monocular) (ICD-10: H54). Finally, we selected individuals in the top and bottom 3% of a polygenic score (PS) for cognitive performances[30]. This was motivated by previous work showing a strong association between a PS for cognitive performances and study participation[19 31]. Our aim was to prospectively assess if recruiting individuals based on this PS would result in a different study participation.

Invitation procedures

Invitation was carried out via postal letter for 5995 FinnGen participants by Helsinki (n=2000), Auria (n=2000) and Tampere (n=1236) biobanks. For an additional 759 individuals from

medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

Tampere biobank, an invitation was sent electronically via the OmaTays healthcare portal, which has been used already for medical communication and online booking appointments by the Tampere biobank. A random subset of 422 participants aged 18-69 from Tampere (n=240) and Auria (n=182) biobank that did not reply to the first invitation were re-contacted with a second invitation letter.

Portal for enrollment and collecting survey information

We collaborated with the Finnish biobank cooperative (FINBB) to create a portal that allows study participants to securely log-in, identify themselves, provide consent, and access the study survey tools. The portal is hosted at https://omabiopankki.fingenious.fi/ and secure authentication is guaranteed by a bank ID identification system which is available to most Finnish residents. Once the participants log into the system, the study invitation and consent can be viewed online. After the participants consent to participate in the study, they can view the two information collection tools: a general questionnaire and a battery of the cognitive tests. By selecting each tool, participants firstly view a short description prior to their completion. Both the general questionnaire and cognitive tests are developed on third-party platforms. Once the participants have completed the online questionnaire or cognitive test battery, they are redirected back to the Omabiopankki portal where they can complete the remaining tasks.

General study questionnaire

The general study questionnaire included 18 categories of questions which cover a broad range of health-related topics that are not obtainable or are difficult to obtain through hospital and register data. All questions were accompanied by the option “don’t know” and “prefer not to answer” to allow the participants to voluntarily skip any of the questions. The following categories were included, with the scientific justification: 1) General questions (sex at birth, height, weight); 2) Women's health; 3) Smoking; 4) Family medical history; 5) Disease diagnoses (including diseases not well captured by hospital and register data); 6) Medication history (including over-the-counter medications); 7) Alcohol consumption; 8) Early onset neurodevelopmental disorders and psychiatric disorders; 9) Mental health and mood; 10) Education levels; 11) Physical activity; 12) Multi-site pain; 13) Risk taking; 14) Flu and viral infections (respiratory); 15) Sleep; 16) Oral health; 17) Diet; 18) Sauna habits.

Cognitive tests

We collected data from a test battery designed to capture different aspects of cognitive abilities. This test battery was provided by *TestMyBrain* and translated into Finnish. *TestMyBrain* has previously validated all their cognitive test batteries [https://psyarxiv.com/dcszr/]. Eight different tests were performed: Digit Symbol Matching, Flicker Change Detection, Visual Paired Associates, Multiracial Emotion Identification, Gradual Onset Continuous Performance, Matrix Reasoning, and Vocabulary test. Because the vocabulary test could not be directly translated from English, we created a Finnish-specific version of the test by selecting 19 out of 30 candidate words that provided with the highest correlation with the well-established Wechsler Adult Intelligence Scale -Revised (WAIS-R) Vocabulary task[32] in a sample of N=24 Finns (79.2% women) with an average age of 24.3 years

medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

(SD: 3.0). Our new online vocabulary test had a Spearman's correlation of 0.86 with the Vocabulary task from the WAIS-R. The analysis of collected cognitive test data is out of the scope of this paper and therefore not shown here. Participants could see a summary of their results and how they scored in each test compared to the average scores after completion of all the tests.

Feedback questionnaire

To get a sense of the participants' experience upon recontacting, we introduced a feedback questionnaire directed to the individuals logging into the Omabiopankki portal. The main question to participants was to rate their overall experience in completing the online questionnaire and cognitive tests using a Net Promoter Score (NPS) scale from 0 to 10. In addition, we asked if they have participated in an online questionnaire study before, if they experienced any technical issues while completing the survey, and to share with us what they enjoyed or did not enjoy while completing the survey.

Ethics

Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, older research cohorts, collected prior the start of FinnGen (in August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea, the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) approved the FinnGen study protocol Nr HUS/990/2017. The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019, THL/1524/5.05.00/2020, and THL/2364/14.02/2020), Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020 and Statistics Finland (permit numbers: TK-53-1041-17 and TK-53-90-20). The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 6 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, Auria Biobank AB17-5154, Biobank Borealis of Northern Finland_2017_1013, Biobank of Eastern Finland 1186/2018, Finnish Clinical Biobank Tampere MH0004, Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

Acknowledgements

We want to acknowledge the participants and investigators of FinnGen study. The following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx), Finnish Clinical Biobank Tampere (www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta) and Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/). All Finnish Biobanks are members of BBMRI.fi infrastructure (www.bbmri.fi). The Finnish Biobank Cooperative - FINBB (https://finbb.fi/) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services (https://site.fingenious.fi/en/) managed by FINBB.

Funding

The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme Corp, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis AG, and Boehringer Ingelheim.

Competing interests

M.E.K.N. is a full-time employee at Novartis AG and H.R. is a full-time employee at Biogen Inc. All other authors declared no competing interests.

Data availability statement

Data are available on reasonable request.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

Author contributions

All authors met the criteria of authorship as outlined in the ICMJE and approved this manuscript for submission. Specific contributions were as follows: conceptualisation: AG, MEKN, RSR, MP, AP, JL; methodology: AG, MEKN, RSR, PW, MH, CM, PM, LG, VAM, OAK; project administration and investigation: RSR, LK, MP, PT, TK, EP, EMM, PN, JM, MP, HYS; data curation: MEKN, RSR, PT, PN, TK, RSR, TSH, TJ; formal analysis: RSR, MEKN, AG; writing original draft: RSR, AG; writing, review and editing: JL, HR, AP, MP, AG.

References

1. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562(7726):203-09. doi: 10.1038/s41586-018-0579-z [published Online First: 2018/10/12]
2. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* 2016;70:214-23. doi: 10.1016/j.jclinepi.2015.09.016 [published Online First: 2015/10/07]
3. Karlson EW, Boutin NT, Hoffnagle AG, et al. Building the Partners HealthCare Biobank at Partners Personalized Medicine: Informed Consent, Return of Research Results, Recruitment Lessons and Operational Considerations. *J Pers Med* 2016;6(1) doi: 10.3390/jpm6010002 [published Online First: 2016/01/20]
4. Nagai A, Hirata M, Kamatani Y, et al. Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol* 2017;27(3S):S2-S8. doi: 10.1016/j.je.2016.12.005 [published Online First: 2017/02/13]
5. Al Kuwari H, Al Thani A, Al Marri A, et al. The Qatar Biobank: background and methods. *BMC Public Health* 2015;15:1208. doi: 10.1186/s12889-015-2522-7 [published Online First: 2015/12/05]
6. Kuriyama S, Yaegashi N, Nagami F, et al. The Tohoku Medical Megabank Project: Design and Mission. *J Epidemiol* 2016;26(9):493-511. doi: 10.2188/jea.JE20150268 [published Online First: 2016/07/05]
7. Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* 2011;40(6):1652-66. doi: 10.1093/ije/dyr120 [published Online First: 2011/12/14]
8. Wei CY, Yang JH, Yeh EC, et al. Genetic profiles of 103,106 individuals in the Taiwan Biobank provide insights into the health and history of Han Chinese. *NPJ Genom Med* 2021;6(1):10. doi: 10.1038/s41525-021-00178-9 [published Online First: 2021/02/13]
9. Finer S, Martin HC, Khan A, et al. Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *Int J Epidemiol* 2020;49(1):20-21i. doi: 10.1093/ije/dyz174 [published Online First: 2019/09/11]
10. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet* 2018;50(11):1593-99. doi: 10.1038/s41588-018-0248-z [published Online First: 2018/10/24]
11. Backman JD, Li AH, Marcketta A, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* 2021;599(7886):628-34. doi: 10.1038/s41586-021-04103-z [published Online First: 2021/10/19]

medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

12. Kanai M, Akiyama M, Takahashi A, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* 2018;50(3):390-400. doi: 10.1038/s41588-018-0047-6 [published Online First: 2018/02/07]
13. Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. *Lancet* 2015;386(9993):533-40. doi: 10.1016/S0140-6736(15)60175-1 [published Online First: 2015/06/08]
14. Richmond RC, Anderson EL, Dashti HS, et al. Investigating causal relations between sleep traits and risk of breast cancer in women: mendelian randomisation study. *BMJ* 2019;365:l2327. doi: 10.1136/bmj.l2327 [published Online First: 2019/06/28]
15. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large population-based studies. *Nat Rev Genet* 2021;22(1):19-37. doi: 10.1038/s41576-020-0268-2 [published Online First: 2020/08/30]
16. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12(3):e1001779. doi: 10.1371/journal.pmed.1001779 [published Online First: 2015/04/01]
17. Swanson JM. The UK Biobank and selection bias. *Lancet* 2012;380(9837):110. doi: 10.1016/S0140-6736(12)61179-9 [published Online First: 2012/07/17]
18. Munafo MR, Tilling K, Taylor AE, et al. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* 2018;47(1):226-35. doi: 10.1093/ije/dyx206 [published Online First: 2017/10/19]
19. Pirastu N, Cordioli M, Nandakumar P, et al. Genetic analyses identify widespread sex-differential participation bias. *Nat Genet* 2021;53(5):663-71. doi: 10.1038/s41588-021-00846-7 [published Online First: 2021/04/24]
20. Kurki MI, Karjalainen J, Palta P, et al. FinnGen: Unique genetic insights from combining isolated population and national health register data. *medRxiv* 2022:2022.03.03.22271360. doi: 10.1101/2022.03.03.22271360
21. Boyle J, Berman L, Dayton J, et al. Physical measures and biomarker collection in health surveys: Propensity to participate. *Res Social Adm Pharm* 2021;17(5):921-29. doi: 10.1016/j.sapharm.2020.07.025 [published Online First: 2020/08/18]
22. Pyykkonen AJ, Isomaa B, Pesonen AK, et al. Sleep duration and insulin resistance in individuals without type 2 diabetes: the PPP-Botnia study. *Ann Med* 2014;46(5):324-9. doi: 10.3109/07853890.2014.902226 [published Online First: 2014/05/13]
23. Borodulin K, Tolonen H, Jousilahti P, et al. Cohort Profile: The National FINRISK Study. *Int J Epidemiol* 2018;47(3):696-96i. doi: 10.1093/ije/dyx239 [published Online First: 2017/11/23]
24. Ruokolainen O, Ollila H, Karjalainen K. Determinants of electronic cigarette use among Finnish adults: Results from a population-based survey. *Nordisk Alkohol Nark* 2017;34(6):471-80. doi: 10.1177/1455072517736618 [published Online First: 2017/12/01]
25. Lobier M, Niittymaki P, Nikiforow N, et al. FinDonor 10 000 study: a cohort to identify iron depletion and factors affecting it in Finnish blood donors. *Vox Sang* 2020;115(1):36-46. doi: 10.1111/vox.12856 [published Online First: 2019/10/28]
26. Lindsted KD, Fraser GE, Steinkohl M, et al. Healthy volunteer effect in a cohort study: temporal resolution in the Adventist Health Study. *J Clin Epidemiol* 1996;49(7):783-90. doi: 10.1016/0895-4356(96)00009-1 [published Online First: 1996/07/01]
27. Pinsky PF, Miller A, Kramer BS, et al. Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. *Am J Epidemiol* 2007;165(8):874-81. doi: 10.1093/aje/kwk075 [published Online First: 2007/01/25]

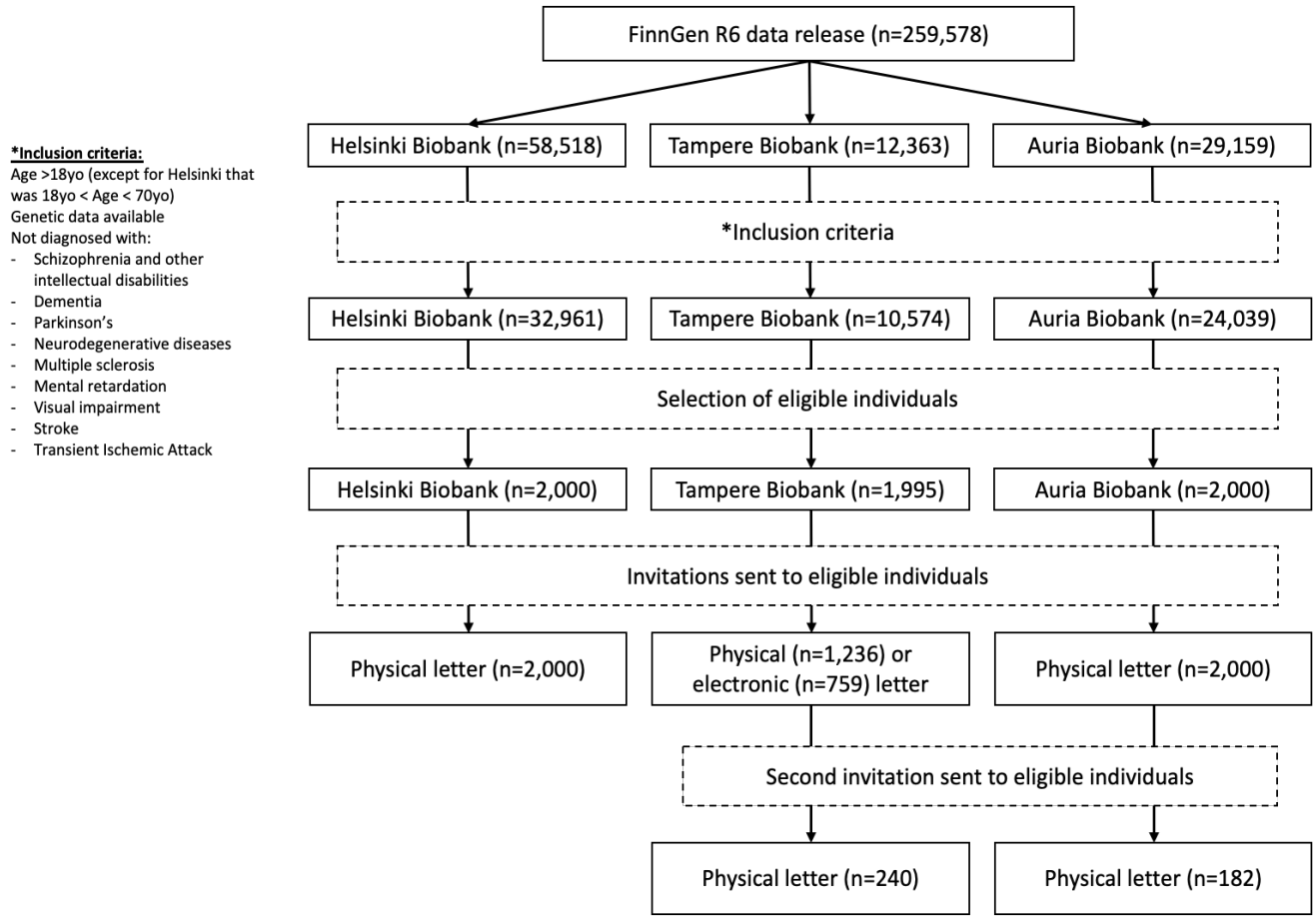
medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

28. Harrison S, Henderson J, Alderdice F, et al. Methods to increase response rates to a population-based maternity survey: a comparison of two pilot studies. *BMC Med Res Methodol* 2019;19(1):65. doi: 10.1186/s12874-019-0702-3 [published Online First: 2019/03/22]
29. Smith MG, Witte M, Rocha S, et al. Effectiveness of incentives and follow-up on increasing survey response rates and participation in field studies. *BMC Med Res Methodol* 2019;19(1):230. doi: 10.1186/s12874-019-0868-8 [published Online First: 2019/12/07]
30. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 2018;50(8):1112-21. doi: 10.1038/s41588-018-0147-3 [published Online First: 2018/07/25]
31. Tyrrell J, Zheng J, Beaumont R, et al. Genetic predictors of participation in optional components of UK Biobank. *Nat Commun* 2021;12(1):886. doi: 10.1038/s41467-021-21073-y [published Online First: 2021/02/11]
32. Wechsler D. WAIS-R: Wechsler adult intelligence scale-revised. Rev. ed ed. New York: New York, NY : Harcourt Brace Jovanovich [for] Psychological Corp., ©1981. 1981:1896-1981.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

Figures

Figure 1. Flow diagram depicting the selection of study participants.



medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

Figure 2. Net Promoter Scores (NPS) from the feedback survey asking for the experience of participants completing the online questionnaires (n=143).

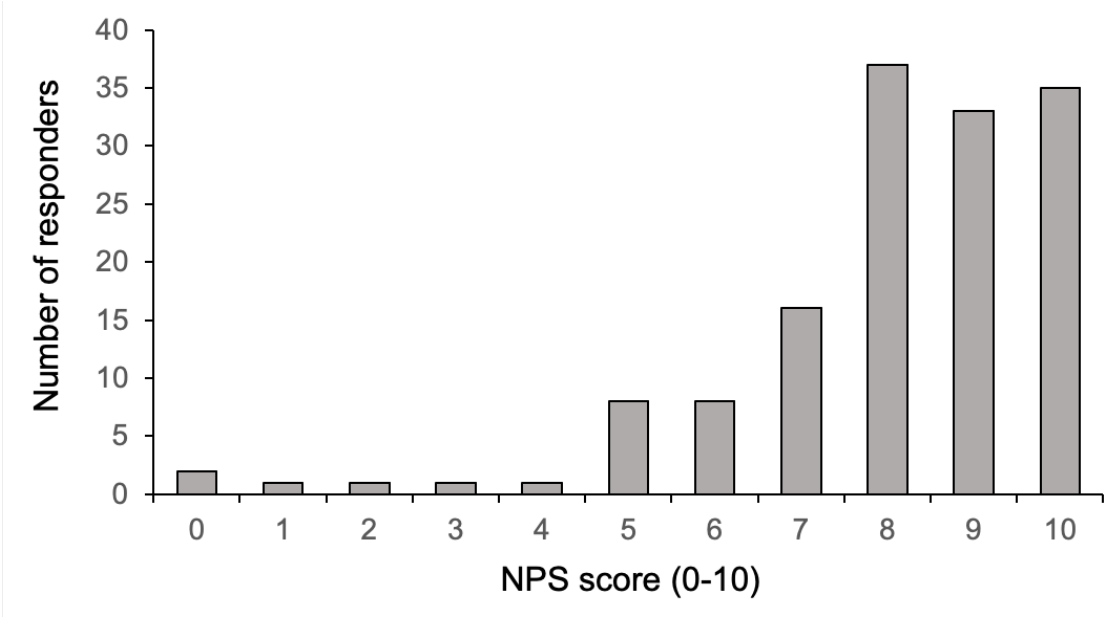
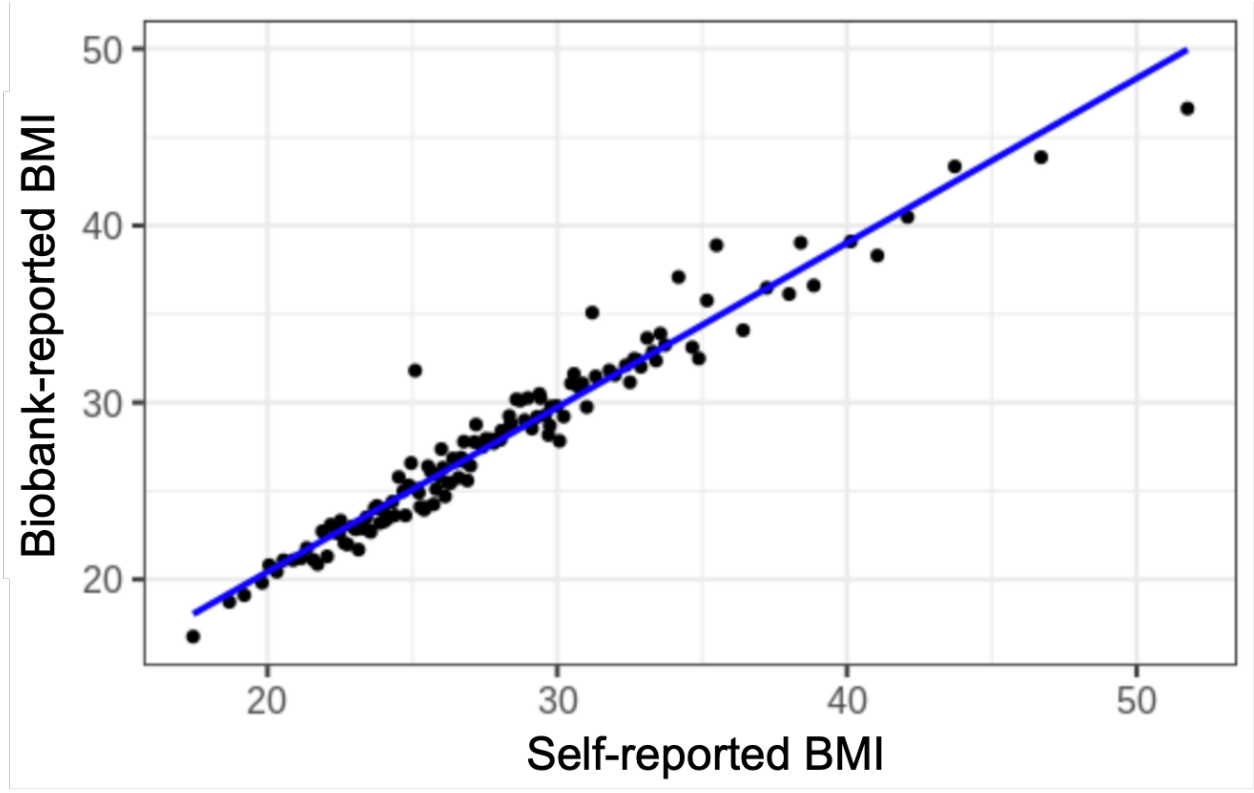
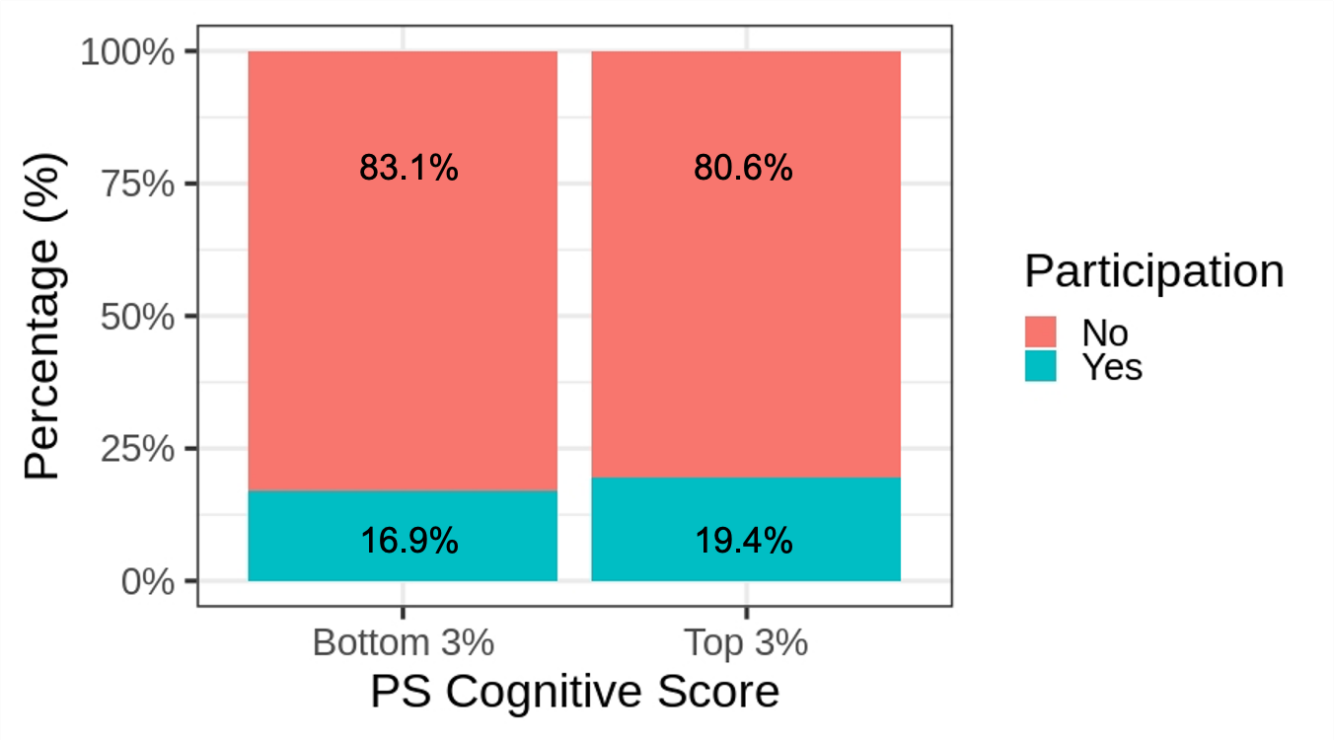


Figure 3. Spearman's correlation between biobank-reported and self-reported body mass index (BMI) measures acquired from the same participants (n=647).



medRxiv preprint doi: https://doi.org/10.1101/2022.04.07.22273501; this version posted April 25, 2022. 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-ND 4.0 International license .

Figure 4. Stacked bar graph showing the participation proportion of study participants by PS cognitive performances group.



Tables

Table 1. Basic questionnaire participation rates among all invitees by age group and biobank.

Age	Helsinki (n=2000)			Tampere (n=1995)			Auria (n=2000)			All (n=5995)		
	Invited	Participated	Participation Rate (%)	Invited	Participated	Participation Rate (%)	Invited	Participated	Participation Rate (%)	Invited	Participated	Participation Rate (%)
20-29	149	22	14,8	116	18	15,5	41	8	19,5	306	48	15,7
30-39	426	70	16,4	273	51	18,7	115	23	20,0	814	144	17,7
40-49	379	93	24,5	231	44	19,0	159	49	30,8	769	186	24,2
50-59	484	133	27,5	303	58	19,1	218	64	29,4	1005	255	25,4
60-69	562	142	25,3	410	93	22,7	533	150	28,1	1505	385	25,6
70-79	0	0	0,0	503	62	12,3	686	21	3,1	1189	83	7,0
80-89	0	0	0,0	152	13	8,6	226	1	0,4	378	14	3,7
90-99	0	0	0,0	7	0	0,0	22	0	0,0	29	0	0,0
Total	**2000**	**460**	**23**	**1995**	**339**	**17**	**2000**	**316**	**15,8**	**5995**	**1115**	**18,6**

Supplementary Material

Supplementary Table 1. Basic questionnaire participation rates among all invitees under 70 years old by age group and biobank.

Supplementary Table 2. Prevalence of the medical history attainment for the same participants (n=1101) between the Re-contacting pilot and FinnGen studies.