

# BMJ Open Cohort profile: SUPER-Finland – the Finnish study for hereditary mechanisms of psychotic disorders

Markku Lähteenvuo <sup>1,2</sup>, Ari Ahola-Olli <sup>2,3,4</sup>, Kimmo Suokas <sup>5</sup>, Minna Holm,<sup>6</sup> Zuzanna Misiewicz,<sup>2</sup> Tuomas Jukuri,<sup>7</sup> Teemu Männynsalu,<sup>8</sup> Asko Wegelius,<sup>9,10</sup> Willehard Haaki,<sup>11,12</sup> Risto Kajanne,<sup>2</sup> Aija Kyttälä,<sup>6</sup> Annamari Tuulio-Henriksson,<sup>13</sup> Kaisla Lahdensuo,<sup>2,14</sup> Katja Häkkinen,<sup>1</sup> Jarmo Hietala,<sup>11,12</sup> Tiina Paunio,<sup>6,9,15</sup> Jussi Niemi-Pynttari,<sup>8</sup> Tuula Kieseppä,<sup>9</sup> Juha Veijola,<sup>16,17</sup> Jouko Lönnqvist,<sup>6,9</sup> Erkki Isometsä,<sup>9</sup> Olli Kampman,<sup>18,19,20,21,22</sup> Jari Tiihonen,<sup>1,23,24</sup> Steven Hyman,<sup>3,25</sup> Benjamin Neale,<sup>3,4</sup> Mark Daly,<sup>2</sup> Jaana Suvisaari,<sup>6</sup> Aarno Palotie<sup>2,3,4</sup>

**To cite:** Lähteenvuo M, Ahola-Olli A, Suokas K, *et al.* Cohort profile: SUPER-Finland – the Finnish study for hereditary mechanisms of psychotic disorders. *BMJ Open* 2023;**13**:e070710. doi:10.1136/bmjopen-2022-070710

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-070710>).

Received 05 December 2022  
Accepted 18 March 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Markku Lähteenvuo;  
markku.lahteenvuo@niiva.fi

## ABSTRACT

**Purpose** SUPER-Finland is a large Finnish collection of psychosis cases. This cohort also represents the Finnish contribution to the Stanley Global Neuropsychiatric Genetics Initiative, which seeks to diversify genetic sample collection to include Asian, Latin American and African populations in addition to known population isolates, such as Finland.

**Participants** 10 474 individuals aged 18 years or older were recruited throughout the country. The subjects have been genotyped with a genome-wide genotyping chip and exome sequenced. A subset of 897 individuals selected from known population sub-isolates were selected for whole-genome sequencing. Recruitment was done between November 2015 and December 2018.

**Findings to date** 5757 (55.2%) had a diagnosis of schizophrenia, 944 (9.1%) schizoaffective disorder, 1612 (15.5%) type I or type II bipolar disorder, 532 (5.1%) psychotic depression, 1047 (10.0%) other psychosis and for 530 (5.1%) self-reported psychosis at recruitment could not be confirmed from register data. Mean duration of schizophrenia was 22.0 years at the time of the recruitment. By the end of the year 2018, 204 of the recruited individuals had died. The most common cause of death was cardiovascular disease (n=61) followed by neoplasms (n=40). Ten subjects had psychiatric morbidity as the primary cause of death.

**Future plans** Compare the effects of common variants, rare variants and copy number variations (CNVs) on severity of psychotic illness. In addition, we aim to track longitudinal course of illness based on nation-wide register data to estimate how phenotypic and genetic differences alter it.

## INTRODUCTION

Psychotic disorders are severe mental disorders which often have profound effects on quality of life and general functioning.<sup>1,2</sup> With a lifetime prevalence of 3.1% in Finland,<sup>3</sup> they are also fairly common. The most common psychotic disorders, schizophrenia

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Limitations are generally related to the aim to collect a large number of psychosis patients for a genetic study in a relatively short time frame. This resulted in a framework where ascertainment was largely based on the accessibility of study subjects and that the phenotyping had to be based on traits that could be collected during a 1-hour visit or from health record data.
- ⇒ We might have over-representation of subjects with severe disease phenotype who need assisted housing as they were more likely to be contacted during housing unit visits.
- ⇒ Strengths include the availability of longitudinal registry data, for example on hospitalisations from 1969 onwards.
- ⇒ SUPER project has exome sequenced a large cohort of psychosis cases from a genetically isolated population of Finland with known enrichment of rare and low-frequency variants.

and bipolar disorder, are both highly heritable, with heritability estimates around 65%–85%.<sup>4,5</sup> Both are highly polygenic phenotypes.<sup>5</sup> Increasing sample sizes have enabled the identification of common and rare variants in schizophrenia and bipolar disorder, but the understanding of disease mechanics is still limited.<sup>6</sup>

Although high-impact coding variants are rare and most psychosis cases do not carry them, identification of such variants provide opportunities for functional follow-up studies and potential insights into the mechanisms of these disorders.<sup>7–10</sup> The Finnish population history provides enhanced opportunities to identify such low-frequency coding variants. Genetically, Finland is an isolated population with two recent bottlenecks, which has

caused the enrichment of a set of low-frequency alleles, including coding and loss-of-function variants in the population.<sup>11</sup> Migration of population from coastal early settlement regions to inland created internal subpopulations with distinct genetic features.<sup>12</sup> Taken together, identification of low-frequency variants contributing to disease risk might be easier in Finland than in other non-bottlenecked European populations due to enhanced statistical power.<sup>13</sup> Such rare variants predisposing to schizophrenia<sup>14–16</sup> and to other chronic diseases have already been discovered.<sup>17</sup>

In addition to genes, major environmental and social factors shape the risks of psychotic as well as other mental disorders.<sup>5 18–21</sup> These factors play a role after the early phases of brain development, pointing out the need for comprehensive phenotype and environmental risk factor data in order to understand the causes of the disorders.<sup>22,23</sup> Besides the advantages conferred by the small founder population, rapid recent population growth and the high prevalence of schizophrenia, Finland has also the advantage of comprehensive health and population registers to facilitate the analysis of the overall health consequences of rare and common variants and gene-environment interactions in large population samples.<sup>24–26</sup>

Here, we present a new patient cohort, SUPER-Finland, which aims to benefit from the population history of Finland, national registers and high regional prevalence of schizophrenia to advance the field of psychosis research. We have collected a cohort of 10 474 patients with a history of a self-reported psychotic disorder and obtained permissions for extensive registry data extraction from the Finnish national registries. Study participants were genotyped with single nucleotide polymorphism (SNP) array and exome sequenced. A subset from Northern Finland was whole genome sequenced. Phenotypic information was collected with an interview, questionnaires and cognitive testing. The cohort collection has been funded by the Stanley Center for Psychiatric Research and forms one arm of the Stanley Global Neuropsychiatric Genomics Initiative, which aims to collect an international cross-continent psychosis cohort of over 100 000 individuals. In this paper, we describe the study protocol for SUPER-Finland.

## COHORT DESCRIPTION

The SUPER-Finland study collected 10 474 participants with a history of at least one psychotic episode. The collection was conducted during a 3-year period between November 2015 and December 2018. Subjects with a diagnosis of a schizophrenia spectrum psychotic disorder (ICD-10 codes F20, F22–F29), bipolar disorder (F31) or major depressive disorder with psychotic features (F32.3 and F33.3) were recruited from inpatient and outpatient psychiatric, general care and housing units. The study was advertised in local newspapers. Minors (<18 years) and subjects unable to give informed consent were excluded from the study. Participants recruited from healthcare

and housing units were always initially contacted by their treating units instead of project staff. Special care was taken to ensure wide coverage of known population isolates.

The study protocol has been approved by local ethical committees. Separate consents were sought for (1) study participation, including sample and data storage, and data collection from national registers, (2) additional sampling for induced pluripotent stem cell (iPS) line production, (3) permission to be re-contacted for additional studies/data set enhancement, (4) permission to contact the physician in charge if any health-related results of clinical importance were to emerge from the study. A separate biobank consent for THL Biobank was collected if participant allowed storage and use of collected data in biobank research. The cohort participants have been examined at baseline, but consent was asked for whether study personnel can recontact subjects in the future for additional samples or phenotype information. Longitudinal data can be acquired through national registers.

The study protocol included a questionnaire, an interview conducted by a study nurse, anthropometric measurements and blood/saliva sampling. In addition, subjects completed the Cambridge Neuropsychological Test Automated Battery (CANTAB) with a tablet computer. The average length of one study assessment was between 60 and 90 min.

## Questionnaire and interview

Participants were asked to fill out a questionnaire (when able, by themselves, otherwise assisted) followed by an interview with a study nurse. During the interview, data on chronic diseases, education, school difficulties, living status, anthropometric traits and current medication was collected. The questionnaire contained items concerning adverse childhood experiences, current psychosocial well-being, sleep, substance use and self-perceived cognitive problems. The questions were adopted from previous Finnish cohort studies whenever available, in order to facilitate case-control comparisons. Detailed information from the interview and questionnaire is presented in online supplemental tables 1 and 2.

## Cognitive assessment

During the interview the study subject performed two computerised CANTAB tasks.<sup>27</sup> We had to limit the number of tests due to time and resource constraints. *Reaction time* was assessed with the 5-choice Reaction Time task (RTI). This test was chosen because information processing speed is one of the central cognitive deficits in psychotic disorders.<sup>28</sup> The task provides information on motor and mental response speed, response accuracy and impulsivity. Visual learning and memory were assessed with the Paired Associates Learning (PAL) task.<sup>29</sup> This test assesses learning and memory which are relevant processes among the cognitive endophenotypes in psychotic disorders.<sup>30</sup> The instructions for PAL test have been translated into Finnish and the test has been used

previously in the Northern Finland 1966 Cohort Study.<sup>31</sup> RTI was translated specifically for the SUPER Finland study.

### Sample collection

Blood samples were collected by venipuncture for DNA extraction (2× Vacutainer EDTA K2 5/4mL, BD), serum (Vacutainer STII 10/8mL gel, BD) and plasma (Vacutainer EDTA K2 10/10mL, BD) analyses. In cases where venipuncture was not possible, a saliva sample (DNA OG-500, Oragene) was collected for DNA extraction. RNA samples were collected from the first 1500 participants (PAXgene Blood RNA Tube, Qiagen). Serum and plasma were left to settle for 30 min and then centrifuged, aliquoted in 0.5 mL fractions (Fluid X tubes) and frozen (−20°C) on-site within 60 min of sampling. Fasting before venipuncture was not required, but fasting time was documented. Also, possible infections (or fever) within 7 days prior to sampling were documented. All frozen samples were sent to the Finnish Institute for Health and Welfare (THL) within 3 months for long-term storage. If the study participant gave his/her consent, an extra blood sample (Vacutainer 10/10 mL Li-He, BD) was collected for isolation of peripheral blood mononuclear cells (PBMCs) for iPS cell production. PBMCs were isolated by Ficoll-Paque method within 30 hours after sampling and stored in LN2 vapour phase.

DNA extraction from EDTA-blood tubes was performed using PerkinElmer Janus chemagic 360i Pro Workstation with the CMG-1074 kit. Saliva samples (n=509) were incubated at +50°C on/before DNA extraction and processed using Chemagen Chemagic MSM I robot with CMG-1035-1 kit. Genotyping and sequencing were performed at the Broad Institute of MIT and Harvard, Boston, USA.

### Data handling

A secure online submission system for phenotype data was created for the study. The integrity of the data was verified at the time of transmission. Personal data and information on consents were collected manually on paper and have been stored separately from the genotype and phenotype data. The DNA samples, blood samples and phenotype data collected will ultimately be stored at the THL Biobank if consent was given.

### Linking of registry data

A unique feature for Finland and other Nordic countries is their nationwide registers kept for administrative and statistical purposes.<sup>32</sup> Finnish Prescription Drug register was initiated in 1996. The Care Register for Health Care is a continuation of the Hospital Discharge Register which was launched in 1969 and the register contains data on every hospitalisation and specialty clinic outpatient visit in the country. Psychiatric hospitalisations have information from involuntary treatments and psychosocial functioning on admission and at discharge. Unique social security number given to every resident guarantees accurate data combination across registers. These registers

are comprehensive (ie, every resident is included) and provide a ready tool for data harmonisation as the same data formats are used for every person. The accuracy of the register data has been shown to be accurate for many disease areas, including psychotic disorders and cardiometabolic diseases.<sup>26 33–35</sup> For having any psychotic disorder, both the specificity (98.3%) and the sensitivity (86.1%) of the register data is excellent.<sup>3</sup> Register-based schizophrenia diagnosis is specific, that is, there are few false positives,<sup>35</sup> but the sensitivity is lower because patients with schizophrenia tend to receive other non-affective psychosis diagnosis at the beginning of their course of illness.<sup>36</sup> The registers and the information which can be retrieved through them is described in online supplemental table 3. The ICD-codes used in case definition is described in online supplemental table 4.

The data obtained from registers will be used to verify and detect discrepancies in diagnostic, education and pharmacological treatment data collected with questionnaires and interviews from study participants. In addition, the registers provide information on environmental risk factors, social and occupational functioning and on the longitudinal course of disorder and its treatment.

Statistics Finland maintains the cause of death register and the register containing information on education and employment. We requested four age, sex and birth-place matched controls for each SUPER-Finland study participant to compare these data between population controls and psychosis cases.

### Genotyping strategy

As the genetic discovery strategy in the SUPER-Finland study is based on the population isolate concept, we genotyped all samples with SNP array for genome-wide coverage and identified rare coding variants by exome sequencing. SNP array data will be imputed using a population-specific imputation panel containing 4000 Finnish deep whole genome sequences. The panel has been constructed as part of the Sequencing Initiative Suomi project ([www.sisuproject.fi](http://www.sisuproject.fi)). Using this panel, we can reliably impute variants to SNP array data down to 0.1%–0.2% population frequency.<sup>37</sup> Sequencing is needed to identify de novo and very rare variants. An observed association between a coding variant and a disease is more likely to stimulate meaningful functional analyses compared with disease association involving a non-coding variant. If such variants exist and if they are enriched in Finland due to population history, we have enhanced statistical power to detect them compared with similar sized studies using more admixed populations.<sup>38 39</sup> Both genotyping and exome sequencing have been completed at the Broad Institute. Characteristics of individuals for whom genotyping and/or sequencing data is missing, did not pass quality control or are not available are described in online supplemental table 5. Controls will be drawn from existing Finnish genotype and exome sequence data collections.<sup>13</sup>

## EXOME AND WHOLE SEQUENCING

To identify ultra-rare variants, potentially contributing to psychotic disorder susceptibility we exome sequenced 9381 samples and whole genome sequenced 897 samples. Sequencing was performed at the Broad Institute of MIT and Harvard using the Illumina HiSeq X platform with the use of 151 base pair paired-end reads. Exome samples were enriched using the Illumina Nextera capture kit and sequenced until 80% of the target capture was covered at 20×, while whole-genome samples were sequenced at 20 or 30×. BAM files were processed using Picard sequence processing pipeline (<http://broadinstitute.github.io/picard/>), before mapped onto the human genome reference build 37 (grch37) using BWA.<sup>40</sup> This exactly followed standard best practice alignment and read processing protocols described in more detail in earlier exome analysis publications.<sup>41 42</sup> Variant calling was performed using the Genome Analysis Toolkit (GATK).<sup>43 44</sup> First, GATK (version 3.4) was used to perform local realignment around indels and recalibrate base qualities in each sample BAM. We called each sample using HaplotypeCaller, generating gVCF files containing every position of the genome with likelihoods for variants or the genomic reference. We merged samples into batches of 200 using CombineVCFs, and joint-called samples using GenotypeVCFs, all using default settings according to the best-practice pipeline. We annotated all variants using the Variant Quality Score Recalibration tool in GATK (version 3.6). The output consists of a VCF with germline single nucleotide variants (SNVs) and indels for all samples used in the core SCHEMA analysis. The variant joint calling was equivalent to the pipeline used in the generation of the gnomAD database.<sup>41</sup>

## PATIENT AND PUBLIC INVOLVEMENT

This research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy. However, we acknowledge that in general public–patient involvement is beneficial for psychiatric research and hope to include more public–patient involvement in the future.

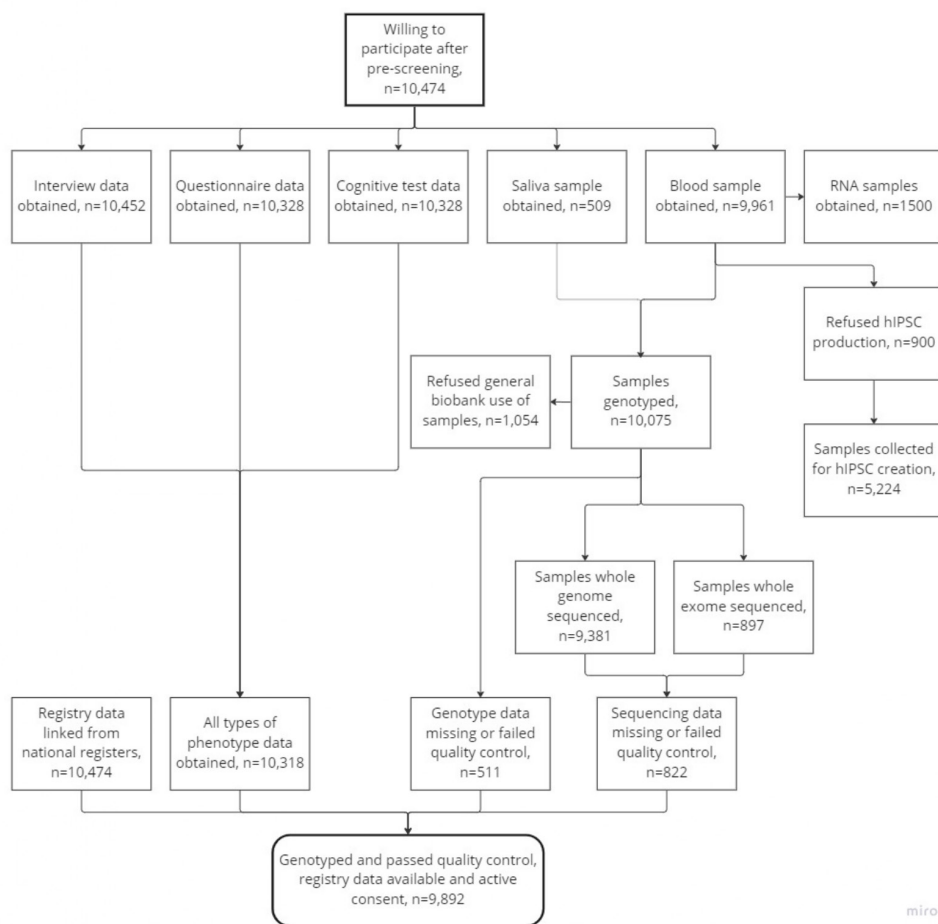
## FINDINGS TO DATE

The SUPER-Finland study finished recruitment at the end of 2018 and managed to recruit 10474 participants (mean age 46.7 (SD 14.8) years; 5184 women; 5290 men). These data are displayed in [table 1](#) and online supplemental figure 1. Blood samples for DNA analyses were collected from 9961 participants and saliva samples from 509 participants. Cells for iPS generation were collected and stored from 5224 participants. A flowchart of the study data collection and data processing is shown in [figure 1](#).

**Table 1** Participant characteristics

		n	%
Sex	Female	5184	49.5
	Male	5290	50.5
Region of residence	Helsinki	2501	23.9
	Kuopio	2282	21.8
	Oulu	2074	19.8
	Tampere	2522	24.1
Native language	Turku	1095	10.5
	Finnish/Swedish	10 128	98.3
Diagnosis	Other	174	1.7
	Schizophrenia	5757	55.2
	Schizoaffective disorder	944	9.1
	Bipolar disorder (I and II)	1612	15.5
	Psychotic depression	532	5.1
Living	Other psychosis	1047	10.0
	Other mental disorder	530	5.1
	Independently	7213	68.9
Education	Supported housing or other	3261	31.1
	Basic	4454	43.4
	Upper secondary	4519	44.0
Marital status	Higher	1293	12.6
	Married or cohabiting	2150	20.8
Employment status	Other	8162	79.2
	Full-time employment	532	5.1
	Part-time employment	320	3.1
	Student	423	4.1
	Disability pension	5234	50.6
	Old-age pension	1191	11.5
	Unemployed	578	5.6
Psychotropic medication use	Rehabilitation	1736	16.8
	Other*	338	3.3
Reported having more than one psychiatric diagnosis	Yes	9789	95.6
	No	452	4.4
Reported having more than one psychiatric diagnosis	Yes	1643	15.8
	No	8757	84.2

Persons who had only completed compulsory education and had no vocational training beyond a vocational course or on the job training and who had not taken the matriculation examination, were classified as having a basic education. Completion of vocational school was defined as secondary education regardless of basic education. Also, those who had passed the matriculation examination but who had no vocational training beyond a vocational course or on the job training were also classified into this intermediate group. Higher education comprised degrees from higher vocational institutions, polytechnics and universities. Psychotropic medications refer to ATC-class N medications. \*Taking care of own household or family members, military service or other (one of the options in the structured interview).



**Figure 1** The data collection and processing workflow chart. Additional information is available in online supplemental file 1. hiPSC, human induced pluripotent stem cell.

### Baseline characteristics

The participants had a positive attitude towards biobanking, as 90% consented to donate their samples and data to the THL Biobank for use in biobank research. 16.7% refused to be re-contacted for additional sampling or other data collection and 11.3% did not want their physician to be contacted in case of significant incidental clinical findings (table 2).

Finland is divided into five university hospital districts. 23.9% of the patients were recruited from the Helsinki and Uusimaa district (Southern Finland), 21.8% from the Kuopio district (Eastern Finland), 19.8% from the Oulu district (Northern Finland), 24.1% from the Tampere district (Central Finland) and 10.5% from the Turku district (Western Finland). The geographic distribution of patients is shown on the map of Finland in figure 2. The distribution of participants reflects the regional variation in the prevalence of psychotic disorders in Finland (online supplemental figure 2).

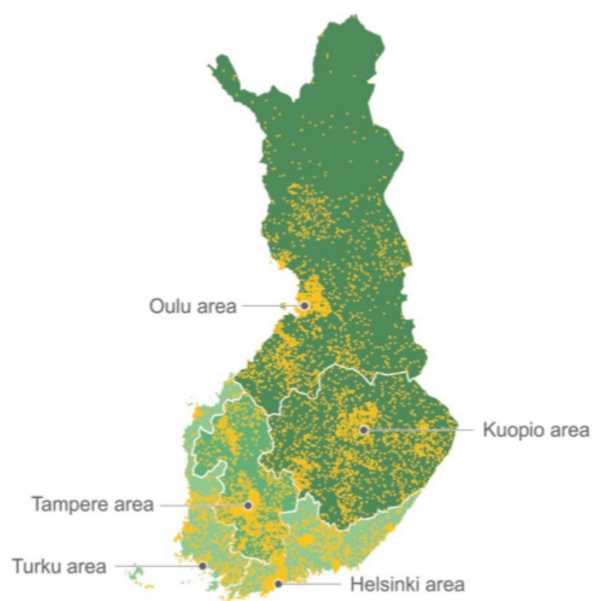
Of the total 10474 participants at the time of recruitment 5757 (55.2%) had a diagnosis of schizophrenia, 944 (9.1%) schizoaffective disorder, 1612 (15.5%) type I or type II bipolar disorder, 532 (5.1 %) psychotic depression, 1047 (10.0%) other psychosis and for 530 (5.1%) the self-reported psychosis at recruitment could not be

confirmed from register data (table 1). We are still waiting for data from the prescription drug register, which will be used to evaluate whether subjects in this subset have a psychotic disorder or not since the Care Register for Health Care only includes hospital discharge diagnoses

**Table 2** Consent information

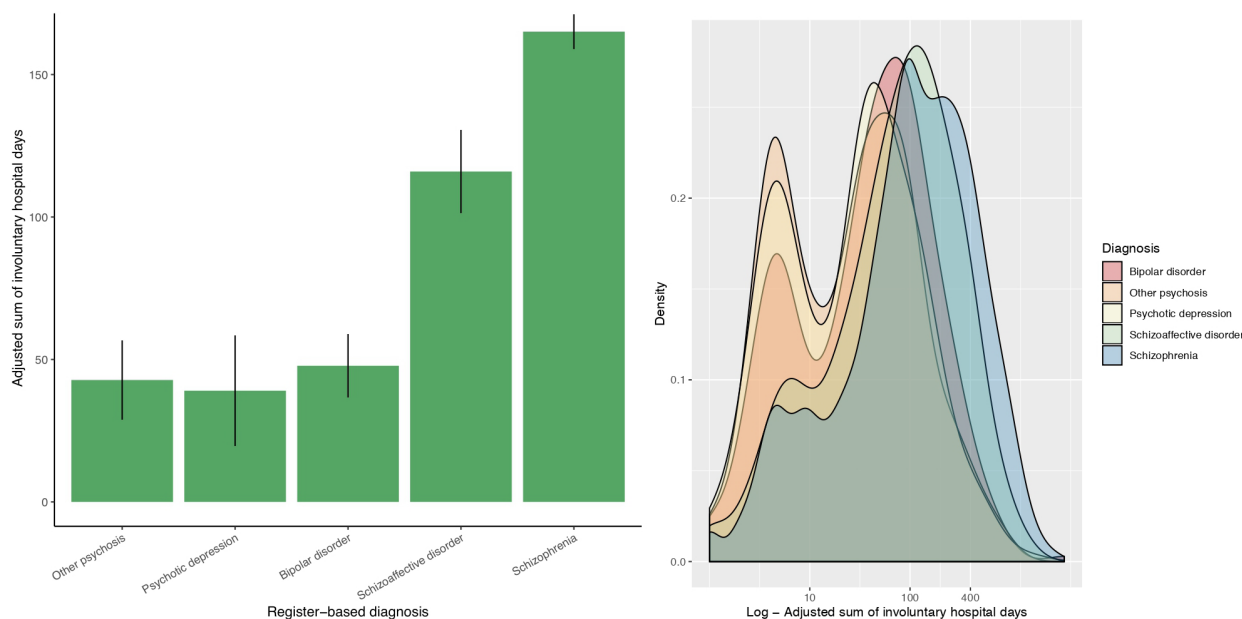
	n	%
Willing to participate in the study after prescreening	10474	99.2
Refuse iPS cell production	900	8.6
Refuse Biobank use	1054	10.1
Did not allow to contact their physician in case of incidental clinical finding	1185	11.3
Refuse recontact for additional samples or data and new studies	1746	16.7

The first row describes the number of subjects who were willing to participate after prescreening. The subjects were first prescreened for willingness to participate by healthcare unit staff after which the study personnel gave them more detailed information about the study. The participation rate is based on data collected by the study nurse.  
iPS cell, induced pluripotent stem cell.



**Figure 2** Sample recruitment by university hospital district. The background colour in parenthesis describes the number of recruited samples relative to the number of individuals living in the area. Individuals have been grouped based on the municipality they live in and each dot has been placed within the municipality in which the subject lives to depict the geographical distribution of the sample.

and outpatient visits from specialised healthcare whereas some forms of psychotic disorders, such as substance use psychoses, may be managed at ER departments by primary care physicians.<sup>45</sup>



**Figure 3** (A) The mean number of involuntary hospital days by diagnosis group. The means are adjusted for age, sex and recruitment area. The whiskers in the bar chart represent 95% CIs. (B) Density plots of involuntary hospital days on logarithmic x-axis.

### Longitudinal views from register data

At the time of recruitment, the mean duration of schizophrenia was 22.0 (SD 12.8) years as calculated from the study visit date and date of first hospitalisation due to psychosis. Duration of schizoaffective disorder at study visit was on average 21.9 (SD 12.0) years, but duration of bipolar disorders was notably shorter, 13.7 (SD 9.0) years which was of similar magnitude with the duration of psychotic depression, 14.1 years (SD 9.1). Other psychosis onset was on average 15.3 years before study visit (SD 11.7).

To demonstrate how register data can be applied to give insights into longitudinal course of disorder, we calculated the sum of involuntary hospital days for each subject. Since the sum variable was skewed and had a heavy right tail in the distribution, we calculated the natural logarithm from the sum variable and used R's *emmeans* package to calculate back-transformed means adjusted for age, sex and recruitment area for each diagnosis category (figure 3A). Subjects with schizophrenia had the highest sum of involuntary hospital days followed by schizoaffective disorder. Involuntary hospital days between patients with bipolar disorder, psychotic depression or other psychosis did not differ from each other. As the cohort contains subjects recruited from forensic psychiatric hospitals, the numbers might not represent the overall situation of involuntary hospitalisations in Finland among psychotic patients. Figure 3B depicts the bimodal distribution of involuntary hospital days with the first peak around 7 days and second peak around 100 days. 29% of subjects with schizophrenia had no involuntary hospital days compared with 53% of subjects with bipolar disorder. Similarly, 53% of subjects with psychotic depression had zero involuntary hospital days but those

with schizoaffective disorder had a similar rate of involuntary hospitalisations as subjects with schizophrenia (27%).

We received cause of death register data from Statistics Finland. These data contain information on whether the recruited subject had died by the end of the year 2018. We grouped the deaths based on the first character of ICD-10 code. Total of 204 subjects had died between recruitment and the end of the year 2018. Most common cause of death was cardiovascular diseases (n=61, ICD-10 codes beginning with I) followed by neoplasms (n=40). Mean age at death was 59.1 years and ranged between 21 and 91 years. Thirteen subjects had died due to external causes, such as different kinds of falls. We will request updates for the cause of death register data later to collect more follow-up information and compare these data against data derived from age, sex and birthplace matched controls. To depict psychiatric comorbidity, we plotted the number of psychiatric diagnoses each subject has using the register data (online supplemental figure 3).

### Future plans

In the future, we aim to compare the effects of common variants, rare variants and copy number variations (CNVs), on severity of psychotic illness. The severity of psychotic illness and associated comorbidities and the longitudinal disease course will be analysed using the longitudinal information available from nationwide healthcare registries to estimate how phenotypic and genetic differences alter the disease course.

### Strengths and limitations

In a small country with a population of 5.5 million, we were able to recruit over 10 000 participants with a psychotic disorder, with a geographical distribution mirroring that of the prevalence of psychotic disorders and excellent participation particularly from those areas where the prevalence is high.<sup>2</sup> Another main strength of the study is the availability of longitudinal national registry data, which provides a lifetime perspective on courses of disorders and some environmental risk factors and substantially enriches the phenotype data collected during the study visit. To ensure the high quality of the data and samples, the study nurses were trained to perform the study protocol and sample treatment in a similar manner on all research sites. Performance of the study protocol was audited twice a year at all collection sites. No serious adverse events among participants were reported during the study cohort recruitment process.

Because most of the participants were recruited from those who had current treatment contact in specialised mental health services, the diagnosis distribution in the study cohort recruited does not seem to follow the general prevalence of psychotic disorders.<sup>3</sup> This is reflected in the fact that the study recruited

more schizophrenia patients while the proportion of other nonaffective psychotic disorders was smaller. In particular, the proportion of patients recruited through supported housing units was large. On the other hand, having a study sample where patients with poor functional outcomes are well represented allows us to study the genetic factors contributing to the outcome of psychotic disorders. The unique population history of Finland combined with comprehensive healthcare registries allow good opportunities for genetic research in longitudinal settings. Unfortunately, due to resource constraints, we were not able to use validated clinical symptom scales such as the Positive and Negative Symptoms Scale to assess the study participants.

### Collaboration

The SUPER-Finland website can be accessed for further information (<https://www.superfinland.fi/english>). The data from SUPER-Finland participants who gave biobank consent can be acquired from the THL Biobank when released from the original study (<https://thl.fi/en/web/thl-biobank>).

### Author affiliations

- <sup>1</sup>Department of Forensic Psychiatry, University of Eastern Finland, Niuvanniemi Hospital, Kuopio, Finland
- <sup>2</sup>Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
- <sup>3</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
- <sup>4</sup>Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA
- <sup>5</sup>Faculty of Social Sciences, Tampere University, Tampere, Finland
- <sup>6</sup>Finnish Institute for Health and Welfare, Helsinki, Finland
- <sup>7</sup>Department of Psychiatry, Oulu University Hospital, Oulu, Finland
- <sup>8</sup>Psychiatric and Substance Abuse Services, City of Helsinki, Helsinki, Finland
- <sup>9</sup>Department of Psychiatry, Helsinki University Central Hospital, Helsinki, Finland
- <sup>10</sup>Department of Mental Health and Substance Abuse Services, Finnish Institute for Health and Welfare, Helsinki, Finland
- <sup>11</sup>Department of Psychiatry, University of Turku, Turku, Finland
- <sup>12</sup>Department of Psychiatry, TYKS Turku University Hospital, Turku, Finland
- <sup>13</sup>Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland
- <sup>14</sup>Mehiläinen Oy, Helsinki, Finland
- <sup>15</sup>SleepWell Research Program, Faculty of Medicine, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
- <sup>16</sup>Research Unit of Clinical Neuroscience, Department of Psychiatry, University of Oulu, Oulu, Finland
- <sup>17</sup>Medical Research Centre Oulu, University Hospital of Oulu and University of Oulu, Oulu, Finland
- <sup>18</sup>Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
- <sup>19</sup>Department of Psychiatry, The Pirkanmaa Wellbeing Services County, Tampere, Finland
- <sup>20</sup>Department of Clinical Sciences, Psychiatry, Umeå University, Umeå, Sweden
- <sup>21</sup>Department of Clinical Medicine (Psychiatry), Faculty of Medicine, University of Turku, Turku, Finland
- <sup>22</sup>Department of Psychiatry, The Wellbeing Services County of Ostrobothnia, Vaasa, Finland
- <sup>23</sup>Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
- <sup>24</sup>Center for Psychiatry Research, Stockholm City Council, Stockholm, Sweden
- <sup>25</sup>Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA

**Acknowledgements** We want to thank SUPER-Finland study participants. In addition, we want to thank the study nurses who worked on the field during the

collection phase and the THL data collection and sample processing team whose work allowed for almost real-time surveillance of the collected material. Especially, we would like to thank Hannu Turunen for data management and Auli Toivola and Noora Ristiluoma for research coordination. We want to thank the outpatient polyclinics and other healthcare units for their cooperation during the recruitment. We also wish to thank Anders Kämpe and Lea Urpa for their help with the final revisions of the manuscript.

**Contributors** ML, AA-O, KS, JS and AP drafted the manuscript. MH, ZM, TJ, TM, AW, WH, RK, AK, AT-H, KL, KH, JH, TP, JN-P, TK, JV, JL, EI, OK, JT, SH, BN and MD substantially contributed to the conception and design of the work and interpretation of the data and gave critical revision for important intellectual content. AP is responsible for the overall content as guarantor.

**Funding** This work was supported by the Stanley Center for Psychiatric Research at Broad Institute (award/grant number is not applicable).

#### Disclaimer

**Map disclaimer** The inclusion of any map (including the depiction of any boundaries therein), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

**Competing interests** None declared.

**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.

**Patient consent for publication** Not applicable.

**Ethics approval** This study involves human participants and was approved by the coordinating ethical board of the Helsinki and Uusimaa hospital district: 36/13/03/00/2016. Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Deidentified participant data is available upon reasonable request to the authors. Registry data may be obtained from the pertinent registry holders. The SUPER-Finland website can be accessed for further information (<https://www.superfinland.fi/english>). The data from SUPER-Finland participants who gave biobank consent can be acquired from the THL Biobank when released from the original study (<https://thl.fi/en/web/thl-biobank>).

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Markku Lähteenvuo <http://orcid.org/0000-0002-7244-145X>

Ari Ahola-Olli <http://orcid.org/0000-0002-4451-3487>

Kimmo Suokas <http://orcid.org/0000-0001-6296-6343>

## REFERENCES

- Saarni SI, Viertiö S, Perälä J, *et al*. Quality of life of people with schizophrenia, bipolar disorder and other psychotic disorders. *Br J Psychiatry* 2010;197:386–94.
- Perälä J, Saarni SI, Ostamo A, *et al*. Geographic variation and sociodemographic characteristics of psychotic disorders in Finland. *Schizophr Res* 2008;106:337–47.
- Perälä J, Suvisaari J, Saarni SI, *et al*. Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Arch Gen Psychiatry* 2007;64:19–28.
- Pettersson E, Lichtenstein P, Larsson H, *et al*. Genetic influences on eight psychiatric disorders based on family data of 4 408 646 full and half-siblings, and genetic data of 333 748 cases and controls. *Psychol Med* 2019;49:1166–73.
- Owen MJ, Sawa A, Mortensen PB. Schizophrenia. *Lancet* 2016;388:86–97.
- Ripke S, Neale BM, Corvin A. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014;511:421–7.
- Marshall CR, Howrigan DP, Merico D, *et al*. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* 2017;49:27–35.
- Purcell SM, Moran JL, Fromer M, *et al*. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014;506:185–90.
- Singh T, Kurki MI, Curtis D, *et al*. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci* 2016;19:571–7.
- Mukai J, Cannavò E, Crabtree GW, *et al*. Recapitulation and reversal of schizophrenia-related phenotypes in setd1a-deficient mice. *Neuron* 2019;104:S0896–6273(19)30787–1:471–487..
- Chheda H, Palta P, Pirinen M, *et al*. Whole-Genome view of the consequences of a population bottleneck using 2926 genome sequences from Finland and United Kingdom. *Eur J Hum Genet* 2017;25:477–84.
- Martin AR, Karczewski KJ, Kermine S, *et al*. Haplotype sharing provides insights into fine-scale population history and disease in Finland. *Am J Hum Genet* 2018;102:760–75.
- Lim ET, Würtz P, Havulinna AS, *et al*. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet* 2014;10:e1004494.
- Stoll G, Pietiläinen OPH, Linder B, *et al*. Deletion of TOP3β, a component of FMRP-containing mRNPs, contributes to neurodevelopmental disorders. *Nat Neurosci* 2013;16:1228–37.
- Steinberg S, Gudmundsdottir S, Sveinbjornsson G, *et al*. Truncating mutations in RBM12 are associated with psychosis. *Nat Genet* 2017;49:1251–4.
- Peltola MA, Kuja-Panula J, Liuhanen J, *et al*. AMIGO-kv2.1 potassium channel complex is associated with schizophrenia-related phenotypes. *Schizophr Bull* 2016;42:191–201.
- Locke AE, Steinberg KM, Chiang CWK, *et al*. Exome sequencing of Finnish isolates enhances rare-variant association power. *Nature* 2019;572:323–8.
- Murray RM, Bhavsar V, Tripoli G, *et al*. 30 years on: how the neurodevelopmental hypothesis of schizophrenia morphed into the developmental risk factor model of psychosis. *Schizophr Bull* 2017;43:1190–6.
- World Health Organization, Calouste Gulbenkian Foundation. World health organization and calouste gulbenkian foundation. In: *Social determinants of mental health*. Geneva: World Health Organization, 2014.
- Radua J, Ramella-Cravaro V, Ioannidis JPA, *et al*. What causes psychosis? an umbrella review of risk and protective factors. *World Psychiatry* 2018;17:49–66.
- Uher R, Zwickler A. Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness. *World Psychiatry* 2017;16:121–9.
- McGrath JJ, Mortensen PB, Visscher PM, *et al*. Where GWAS and epidemiology meet: opportunities for the simultaneous study of genetic and environmental risk factors in schizophrenia. *Schizophr Bull* 2013;39:955–9.
- Owen MJ, O'Donovan MC. Schizophrenia and the neurodevelopmental continuum: evidence from genomics. *World Psychiatry* 2017;16:227–35.
- Furu K, Wettermark B, Andersen M, *et al*. The Nordic countries as a cohort for pharmacoepidemiological research. *Basic Clin Pharmacol Toxicol* 2010;106:86–94.
- Sund R. Quality of the Finnish hospital discharge register: a systematic review. *Scand J Public Health* 2012;40:505–15.
- Pihlajamaa J, Suvisaari J, Henriksson M, *et al*. The validity of schizophrenia diagnosis in the Finnish hospital discharge register: findings from a 10-year birth cohort sample. *Nord J Psychiatry* 2008;62:198–203.
- CANTAB®. Cognitive assessment software. 2019.
- Dickinson D, Iannone VN, Wilk CM, *et al*. General and specific cognitive deficits in schizophrenia. *Biol Psychiatry* 2004;55:826–33.
- Barnett JH, Blackwell AD, Sahakian BJ, *et al*. The paired associates learning (pal) test: 30 years of CANTAB translational neuroscience from laboratory to bedside in dementia research. *Curr Top Behav Neurosci* 2016;28:449–74.
- Greenwood TA, Lazzeroni LC, Maihofer AX, *et al*. Genome-Wide association of endophenotypes for schizophrenia from the

- Consortium on the genetics of schizophrenia (COGS) study. *JAMA Psychiatry* 2019;76:1274–84.
- 31 Taivalantti M, Barnett JH, Halt A-H, *et al.* Depressive symptoms as predictors of visual memory deficits in middle-age. *J Affect Disord* 2020;264:29–34.
- 5 Haukka J. n.d. Finnish health and social welfare registers in epidemiological research. *Nor J Epidemiol*;14.
- 33 Haukka J, Suvisaari J, Tuulio-Henriksson A, *et al.* High concordance between self-reported medication and official prescription database information. *Eur J Clin Pharmacol* 2007;63:1069–74.
- 34 Leppälä JM, Virtamo J, Heinonen OP. Validation of stroke diagnosis in the national hospital discharge register and the register of causes of death in Finland. *European Journal of Epidemiology* 1999;15:155–60.
- 35 Pajunen P, Koukkunen H, Ketonen M, *et al.* The validity of the Finnish hospital discharge register and causes of death register data on coronary heart disease. *Eur J Cardiovasc Prev Rehabil* 2005;12:132–7.
- 36 Kiviniemi M, Suvisaari J, Pirkola S, *et al.* Regional differences in five-year mortality after a first episode of schizophrenia in Finland. *Psychiatr Serv* 2010;61:272–9.
- 37 Mitt M, Kals M, Pärn K, *et al.* Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet* 2017;25:869–76.
- 38 Dwivedi OP, Lehtovirta M, Hastoy B, *et al.* Loss of znt8 function protects against diabetes by enhanced insulin secretion. *Nat Genet* 2019;51:1596–606.
- 39 Flannick J, Thorleifsson G, Beer NL, *et al.* Loss-Of-Function mutations in Slc30a8 protect against type 2 diabetes. *Nat Genet* 2014;46:357–63.
- 40 Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 2009;25:1754–60.
- 41 Karczewski KJ, Francioli LC, Tiao G, *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434–43.
- 42 Lek M, Karczewski KJ, Minikel EV, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91.
- 43 Van der Auwera GA, Carneiro MO, Hartl C, *et al.* From fastq data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013;43:11.
- 44 McKenna A, Hanna M, Banks E, *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
- 45 Suvisaari J, Perälä J, Saarni S, *et al.* The epidemiology and descriptive and predictive validity of *dsm-IV* delusional disorder and subtypes of schizophrenia. *Clinical Schizophrenia & Related Psychoses* 2009;2:289–97.