Twin Research and Human Genetics (2025), 1–7
doi:10.1017/thg.2025.10025

Article

Data Resource Profile: The Finnish TwinRegistry Project

Stephanie Zellers[1], Sarah Niemi de Paiva[1,2], Mikko Olkkonen[1], Hannu Karhunen[3], Terhi Maczulskij[4], Annina Ropponen[5], Jaakko Kaprio[1] and Antti Latvala[6]

[1]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland, [2]Gillings School of Global Public Health, University of North Carolina Chapel Hill, Chapel Hill, North Carolina, USA, [3]Labour Institute for Economic Research LABORE, Helsinki, Finland, [4]ETLA Economic Research, Helsinki, Finland, [5]The Finnish Institute of Occupational Health, Helsinki, Finland and [6]Institute of Criminology and Legal Policy, University of Helsinki, Helsinki, Finland

---

Abstract

The Older Finnish Twin Cohort was established 50 years ago and includes twins born in Finland before 1958. Members of the cohort have responded to detailed questionnaires about their health, habits, and lifestyle up to four times, in 1975, 1981, 1991, and 2011. In 2019, the Finnish Parliament approved the Act on the Secondary Use of Health and Social Data, which enables wider use of data from national social and healthcare registers as well as various patient systems and social services. This data resource article describes the linkage of the Older Finnish Twin Cohort to numerous social and healthcare registers, alongside linked data from their families and the broader Finnish population born in 1945–1957, which serves as a reference population for generalizability and other analyses.

Keywords: Data profile; Registry data; Cohort profile

(Received 19 June 2025; revise received 26 August 2025; accepted 27 August 2025)

The Older Finnish Twin Cohort (FTC) was established 50 years ago and has been described in existing reviews, most recently in 2019 (Kaprio, 2006, 2013; Kaprio et al., 1978, 2019; Kaprio & Koskenvuo, 2002). Importantly, also in 2019, The Finnish Parliament approved the Act on the Secondary Use of Health and Social Data, which enables wider use of data from national social and healthcare registers as well as various patient systems and social services (Ministry of Social Affairs and Health, n.d.). Permits for health and social matters registry data have been centralized since via the Finnish Social and Health Data Permit Authority (FinData), and data curation and access are provided by Statistics Finland (Findata, n.d.). Finnish national register data is comprehensive and of high quality, with generally long lengths of follow-up and population-level coverage.

Across 2020–2023, we have provided Statistics Finland with the FTC questionnaire data, and we have received permission to link that questionnaire data to multiple health and social registers, in an effort we have named TwinRegistry. TwinRegistry has been approved by the research ethical board of the Finnish Institute for Health and Welfare (THL), with data access permissions granted by FinData and Statistics Finland. Statistics Finland has pseudonymized and compiled all component datasets. Access to data for analyses is via a secure remote access system, FIONA, administrated by Statistics Finland (Tilastokeskus, n.d.). In addition to register data from individuals who have completed the FTC surveys, we have additionally received permission to access the same sources of register data for a broader sample of individuals that we will describe below as the reference population.

The goal of TwinRegistry is to advance the use of register and twin questionnaire data to address complex causal research questions that require unique designs and rely on longitudinal data, familial data, and/or both self-report and objective administrative data. We will additionally describe ongoing projects and research directions.

Sample

The FTC was established in 1974 and is composed of twins born in Finland before 1958 (full range of birth years: 1880–1957). There have been four waves of questionnaires, conducted in the years 1975, 1981, 1990, and 2011. The FTC originally included 13,888 same-sex twin pairs of known zygosity in 1974 and expanded in 1996 to include approximately 8000 pairs of opposite-sex twins born between 1938–1958. Individuals were invited to participate in later waves irrespective of participation in earlier waves. Zygosity of same-sex pairs was determined by validated questionnaire in the 1975 survey, with validation via genetic markers conducted for those twins who have also participated in studies of selected twin pairs. There remain 889 same-sex twin pairs in the FTC for whom zygosity is undetermined due to inconsistent questionnaire responses. The number of pairs includes both complete and incomplete twin pairs; in the 1975 survey 83% of all pairs are complete twin pairs.

In addition to the FTC, we obtained register data from the spouses of the FTC, siblings of the FTC, and spouses of the siblings of the FTC. The basis for the TwinRegistry data began from the FTC and the sample was expanded to facilitate various specific

Corresponding author: Antti Latvala; Email: antti.latvala@helsinki.fi

Cite this article: Zellers S, Niemi de Paiva S, Olkkonen M, Karhunen H, Maczulskij T, Ropponen A, Kaprio J, Latvala A. Data Resource Profile: The Finnish TwinRegistry Project. *Twin Research and Human Genetics* https://doi.org/10.1017/thg.2025.10025

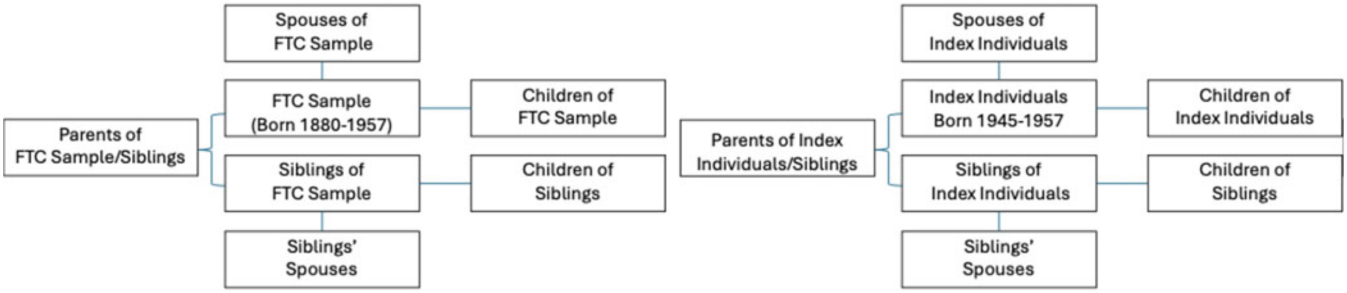

Figure 1. Figure depicting multigenerational data available from the Finnish Twin Cohort and register sources.

research designs. We additionally obtained intergenerational data by including the parents of the FTC, the children of the FTC, and the children of siblings of the FTC. The FTC questionnaires are linked with the register data via personal identity code assigned to Finnish citizens and residents.

Lastly, we created a reference population composed of all individuals born in Finland between 1945–1957. This reference population aligns with a subset of the FTC that were eligible for inclusion in all four FTC questionnaires (birth years 1945–1957 eligible for surveys in 1975, 1980, 1990, 2011). Then, for the reference population, we obtained their family members for the same relationships included for the FTC (e.g., spouses, parents, children, siblings, siblings' children). The familial linkages include both biological and adoptive family members (parents, siblings, and children); spousal relationships are recorded on a yearly basis.

Across the FTC, reference population, and all linked relatives, TwinRegistry includes approximately 4.6 million unique individuals. A single individual can have multiple familial roles. Figure 1 provides an illustration of the sample and familial relationships. From these linked familial relationships, we can create complete familial units; for example, complete parent-child trios for the FTC and reference population. Intergenerationally, there are approximately 25,000 complete parent-offspring trios from the FTC and their parents, 840,000 from the reference population and their parents, 54,000 from the FTC and their children, and 980,000 from the reference population and their children. Table 1 and Table 2 provide a more thorough overview of sample sizes for various roles (Table 1) and familial units (Table 2).

The following sections on register and questionnaire data provide a detailed narrative on a collection of comprehensive datasets available in TwinRegistry, describing aspects of social, economic, and health data. These datasets serve as resources for researchers exploring longitudinal trends in health, demographic changes, and the impact of policy interventions over time. All data use pseudonymized IDs for confidentiality.

Register Data

TwinRegistry data from the registers can be categorized into several partially overlapping categories: demographics, education, employment and socioeconomic position, health and social care, benefits and pensions, and criminal offences (see Table 3 for details). Registers have varying years of data collection and come from several different Finnish agencies. Ready-made datasets, compiled by Statistics Finland, are available for the entire Finnish population, as stated in the metadata as provided by Statistics Finland. Custom datasets, or those created specifically for the TwinRegistry project, are available for the subsample of FTC individuals, the reference population, and all linked relatives (~4.6 million individuals). Sample sizes for the registers vary based on the year of data collection and whether the dataset is ready-made or custom. The sample sizes for the number of individuals range from 91,909 (custom-made KELA pension dataset) up to 8,552,770 (ready-made FOLK children and parents' data). Similarly, some registers may contain multiple observations for a single individual, and these sample sizes also vary substantially by register type and number of years of data collection. The smallest numbers of observations range from 1,048,575 for the custom-made KELA pension data and up to 211,555,041 observations for the KELA dispensed prescription register. Any diagnostic data is coded via International Classification of Diseases codes (ICD; *ICD-10 Version: 2019*, n.d.), though the version of ICD relevant depends on the register and date, as described in available metadata from the agencies hosting each register. Prescription data for medications uses Anatomical Therapeutic Chemical (ATC) classification codes.

Questionnaire Data

The FTC questionnaires offer longitudinal insights through surveys conducted in 1975, 1981, 1990, and 2011. In 1974, researchers identified potential twins from the Central Population Registry of Finland, selecting pairs of individuals of the same sex with the same birth date, same surname at birth, and same local community of birth (Kaprio et al., 2019). The initial survey in 1975 was sent to these identified individuals and inquired whether the individuals were twins, as well as zygosity information for twins. A total of 36,922 individuals were invited in 1975, of which 31,145 replied (response rate of 84.6%). This number of responses includes $N = 1175$ individuals who indicated they were not a biological twin. A total of 29,311 individuals were invited in 1981 with a total of 24,684 responses (84.2% response rate), including 2451 individuals who had not responded in 1975. In terms of retention, 22,122 twins participated in both 1975 and 1981, with 4308 twins who participated in 1975 but not 1981 (83.7% retention from 1975 to 1981).

In 1991 and 2011, only individuals in certain birth years were invited to participate. In 1991, 16,179 twins born between 1930–1957 who had also participated in either 1975 or 1981 were invited and the response rate was 77.3% ($N = 12{,}502$). In 2011, all twins born between 1945–1957, regardless of earlier participation, were invited to participate ($N = 11{,}738$ invited, $N = 8410$ responses, participation rate 71.6%).

These surveys cover aspects of health, employment, family dynamics, personality, stressful life events, and housing conditions, reflecting changes in respondents' situations over the life course. As the FTC questionnaires have been described extensively elsewhere (Kaprio, 2006, 2013; Kaprio et al., 2019; Kaprio et al.,

Table 1. Sample sizes by role

Twins and their families		Population reference individuals and their families	
Role	**Sample size**	**Role**	**Sample size**
Twins	46,497	Non-twin population reference individuals	1,254,124
Mothers of twins and their siblings	13,281 Biological 90 Adoptive	Mothers of reference individuals and their siblings	491,100 biological 5802 adoptive
Fathers of Twins and their Siblings	11,624 Biological 80 Adoptive	Fathers of reference individuals and their siblings	434,488 biological 6416 adoptive
Siblings of twins	18,957 Biological Full 10 Biological Half 4376 Biological of Unknown Relatedness	Siblings of reference individuals	745,470 biological full 5990 biological half 151,354 biological of unknown relatedness
Children of twin mothers	38,343 Biological 140 Adoptive	Children of reference individual mothers	1,077,367 bBiological 4516 adoptive
Children of twin fathers	39,244 Biological 230 Adoptive	Children of reference individual fathers	1,055,323 biological 6090 adoptive
Children of sisters of twins	23,027 Biological 77 Adoptive	Children of sisters of reference individuals	342,864 biological 1291 adoptive
Children of brothers of twins	25,468 Biological 129 Adoptive	Children of brothers of reference individuals	353,479 biological 1859 adoptive

Note: An individual may have multiple roles. For example, a sibling of a twin could themselves be part of the reference population; the biological child of a twin mother may also be a biological child of a reference individual father.

Table 2. Sample sizes for various family units

Family unit	Both parents and a child	One parent and a child
Twins and their parents	23,068	4578
Siblings of twins and their parents	23,856	4373
Twins and their children	54,894	22,293
Siblings of twins and their children	36,209	12,092
Parents of twins – twins – children of twins	16,987	2997
Parents of twins' siblings – twins' siblings – children of twins' siblings	17,904	2925
Population reference individuals and their parents	838,051	144,222
Siblings of reference individuals and their parents	354,566	68,688
Population reference individuals and their children	981,309	350,391
Siblings of reference individuals and their children	261,417	352,696
Parents of reference individuals – reference individuals – children of reference individuals	644,344	104,185
Parents of reference individuals' siblings – reference individuals' siblings – children of reference individuals' siblings	250,870	44,391

Note: Biological and adoptive families are combined.

1978; Kaprio & Koskenvuo, 2002), here we describe the broad data domains (see Table 4 and the supplementary table) and highlight the evolution of the surveys in content and methodology, reflecting shifts in research focus and public health priorities.

Broadly speaking, the domains included in the questionnaires remained stable over time, but the degree of detail varied with topical research interests and the sample age. For example, items on weight and body measurements became more detailed over time, reflecting a growing concern with longitudinal weight trends and waist circumference as a measure of abdominal obesity. Types of diseases, medications, and chronic health conditions included over time also became more inclusive and detailed, reflecting new knowledge about disease as well as the aging cohort. Health-related impacts of sleep, substance use, pain, and psychological health also became more expansive over time. While many questionnaire domains became more detailed over time, many items were maintained over time (e.g., the alcohol and tobacco use items; see the supplementary table), allowing for consistency in longitudinal analyses.

Data Resource Use

The linkage of FTC questionnaires with register data expands opportunities that utilize classical twin designs and discordant twin designs, by providing both novel exposures and outcomes, and a follow-up time of several decades. The additional availability of the reference population from the same birth cohort as a subset of the FTC permits a thorough study of generalizability from twins to the broader population, as well as the use of other genetically informative family-based research designs such as sibling and cousin analyses in practically whole-population data. The combination of twin designs with individual level analyses will permit for excellent statistical power and controls for confounding factors. Additionally, multigenerational studies are also feasible here, including the children of twins design (McAdams et al., 2018) and other extended twin family designs. These extended familial designs permit analyses of intergenerational transmission, better estimate nonadditive genetic effects, and estimate the effects of assortative mating and parental environmental effects (Keller et al., 2010).

Table 3. Domains of register data, specific registers included, years of coverage, and example data

Domain	Register name	Years of coverage	Description of contents
Demographics	FOLK census data	1970–1985	Population censuses from 1970, 1975, 1980, 1985
	FOLK basic data	1987–2021	Brief information on residence, marital and socioeconomic status, education, family size, pensions, and income
	FOLK parents and children	1970–2022	Relationships between parents and offspring, both biological and adoptive
	FOLK migration	1983–2022	Immigration and emigration, inter- and intra-municipality migration within Finland, citizenship, native language, birth country
Education	Degree register	2005–2022	Degrees awarded, institution details, and type of education
	Upper secondary applications	1985–2022	Application area, programs and institution choices, primary and lower secondary education qualifications
	Matriculation examinations	1967–2022	Exam subject, section, dates, and grades; taken at the end of upper secondary school to progress to higher education
	Higher education applications	1992–2020	Applications to polytechnic and university institutions, including institution details and field of education
Employment and income	FOLK placement and labor market training	1991–2022	Dates of placement (1991-) and labor market training (2005-) including type and reason for ending placement
	FOLK job seekers	1991–2022	Duration of job search, employment at the start of job search, and reason for ending job search
	FOLK employment data	1987–2021	Linked employer-employee data with job types, dates, duration of employment, sector, industry, and plant codes
	FOLK unemployment	1991–2022	Dates and duration of unemployment, reason for ending unemployment
	FOLK income	1987–2021	Income from wage earnings, entrepreneurial activities, property, welfare, pension, and grants; taxes and debts
Health	Causes of death	1971–2020	Date and location of death; method for determining cause of death; underlying, immediate, and contributing causes of death
	Medication reimbursements	1972–2020	Entitlement to medication reimbursement based on diagnosis; duration of entitlement
	Prescription medications	1995–2020	Medication ATC code and quantity dispensed for purchased prescriptions
	Sickness allowances	1987–2003	Relevant diagnosis, duration of allowance, and amount paid
	Hospital discharges and Healthcare Register	1972–2020	Impatient diagnoses and care (1972–) and outpatient diagnoses and care (1994–)
	Medical births	1987–2020	Care and diagnoses for mothers, live births, and stillbirths in pre- and perinatal periods
Social welfare and pensions	Income support	1985–2020	Type of income support and amount paid; participation in rehabilitation work activities
	Care Register for Social Welfare	1995–2020	Social welfare institutional care services, type of institution, reason for entering treatment, care needs, duration of care, and discharge plans
	Pensions	1972–2021	Pension type, duration, diagnoses, reason for incapacity for work. Information comes from FOLK data, KELA, and National Pension Center
Crime	Fixed fine data	1995–2020	Crime date, verdicts, fine amounts, settlement status of fines
	Criminal convictions	1977–2020	Crime date, trial date, conviction, types of punishment, imprisonment length, criminal charge, and categorization of criminal charges

TwinRegistry is currently developed by researchers at the Institute of Molecular Medicine Finland (FIMM) and the Institute of Criminology and Legal Policy (Krimo) at the University of Helsinki. FIMM notably also hosts the FinnGen and FinRegistry projects described elsewhere (Kurki et al., 2023; Viippola et al., 2023). These three similar initiatives each have unique samples, strengths, and capabilities, but they do all share register data as a basis, which allows for coordination and infrastructure sharing across projects. For example, the FinRegistry project has compiled *Risteys*, a publicly available web portal that enables exploration of the over 3100 clinical endpoints available in FinRegistry (*Risteys*, n.d.). As the underlying register data in TwinRegistry and FinRegistry are from many of the same sources, many of the FinRegistry endpoints can also be generated in TwinRegistry.

As an example of causal analysis based on the register data, our recent work evaluates the influence of job loss on all-cause and specific-cause mortality, utilizing the economic job displacement natural experiment at the individual-level (Zellers et al., 2025). Finland suffered a severe recession in the early 1990s, with numerous mass layoffs and plant closures resulting in high unemployment. These job losses due to plant closure and mass layoff can be identified objectively via register data, highlighting one example of the utility that this administrative data provides. Future work will integrate the behavioral data available in the twin

Table 4. Domains of Finnish Twin Cohort questionnaire data

Domain	Subtopic	Relevant items and details
Family	Relationship with co-twin	Cohabitation status, age at separation, frequency of contact, twin birth characteristics and similarity to determine zygosity
	Marital status	Asked at all questionnaires
	Family	Added in 1990 to assess number of children and number of children at or below school age
Anthropometrics	Handedness	Asked at intake only
	Height	Asked at all questionnaires
	Weight	Asked at all questionnaires, including historical weight information and intentional weight loss efforts
Physical health	Diseases	Self-reported experience of symptoms or diagnosis of certain diseases, including experiences of pain and stiffness
	Medication use	Medications range from fortifying medications like vitamins, prescription medications for specific conditions, hormone replacement therapy, and other common over the counter medications for minor symptoms
Habits and lifestyle	Sleep	Sleep quality and duration and more detailed items on sleep disturbances, sleep medication use, chronotype, and sleep inertia
	Diet	Dietary questions include consumption of tea and coffee as well as consumption of foods from various food groups
	Exercise and physical activity	Frequency, duration, and intensity of leisure time physical activity as well as physical activity during commute
	Employment and Education	Highest level of education and employment status, as well as details about nature of work, retirement
Substance use	Alcohol use	Frequency and quantity of alcohol consumption across categories of beer, wine, spirits, as well as consequences of alcohol use
	Tobacco use	Frequency and quantity of tobacco use, nicotine dependence, and other exposure (nicotine replacement, secondhand smoke)
Mental health	Stress	Various items to gauge psychological stress and physical indicators of stress like teeth clenching
	Emotions	Trait anger, contentment, life satisfaction, depression, and loneliness
	Personality	Bortner scales, sociability and neuroticism
	Life events	Stressful life events, traumatic life events, major changes to residence and employment

questionnaires to evaluate lifestyle factors as possible mechanisms from job loss to mortality, as well as leverage ‘children-of-twins’ designs to investigate how parental job loss influences offspring mental health.

Generalizability

The comprehensive linkage of FTC data to registers and availability of register data in the reference population born 1945–1957 permits for thorough analyses of how appropriately twins participating in studies represent their birth cohort. Future projects will directly evaluate representativeness as is relevant to the projects’ specific research questions, but here we present a starting set of analyses to demonstrate the representativeness of twins to their birth cohort. To assess representativeness, we compared the subset of FTC twins born 1945–1957 that participated in the 1975 survey ($N = 23{,}650$) to the singletons in the reference population born the same years ($N = 1{,}254{,}094$) using register variables anticipated to be useful in a variety of projects. As the full reference population does include those FTC members born 1945–1957, all comparisons between the FTC sample and singleton members of the reference population are conducted after removal of the FTC from the reference population.

Here we visualize representativeness of the FTC on education, mortality, employment, residence, and marital status. We draw data from the 1987 FOLK basic data (first year of availability) and the 2022 FOLK basic data (last year of availability). Results are presented in Table 5 (1987) and Table 6 (2022). We present relative proportions rather than statistical comparisons, as the sample sizes are well powered to detect small but statistically significant effects that may not be meaningful to interpret. Generally speaking, the twin sample is not different from their respective birth cohort, but this will be more thoroughly evaluated in later papers, with respect to the key outcomes and ages of interest, including statistical comparisons to quantify generalizability and statistical significance. See Niemi de Paiva et al. (2025) for an example of ongoing generalizability analyses. We note there are $N = 406{,}891$singletons (32.4% of the full reference population) who are without any siblings. This is a special group that necessarily can only occur in the singleton population and may influence differences, though we do not evaluate this here.

Strengths and Limitations

A major strength of the TwinRegistry is the ability to combine register data with self-report data, as each data source has its strengths and limitations. The register data is well powered and contains a wealth of objectively reported clinical phenotypes but lacks information on behaviors and lifestyle that may be relevant for less severe outcomes. For example, register data would contain information on diagnoses of alcohol use disorders, which are given to patients whose symptoms are severe enough to warrant medical

Table 5. Generalizable descriptives based on 1987 FOLK basic data, comparing twins born 1945–1957 and population reference individuals born same years (age range 30–42)

	Twin	Reference
% of base sample with valid data in 1987 FOLK basic data	91.7%	84.9%
% Deceased	1.65%	1.64%
Native language		
% Finnish	95.8%	94.2%
% Swedish	4.1%	5.2%
% Other	0.01%	0.0%
Marital status		
% Unmarried	27.0%	21.4%
% Married	64.9%	69.5%
% Divorced	7.4%	8.5%
% Widowed	0.06%	0.07%
Residence		
% Urban	54.9%	57.1%
% Semi-urban	13.8%	13.4%
% Rural	31.3%	29.5%
% Having taken matriculation exam	83.6%	80.0%
Highest qualification		
% Primary	35.8%	34.6%
% Upper secondary	42.3%	38.9%
% Short-cycle tertiary	12.3%	14.1%
% BA/BS	4.7%	6.3%
% MA/MS	4.4%	5.7%
% PhD	0.3%	0.4%
Main activity		
% Employed	85.0%	86.2%
% Unemployed	4.8%	4.5%
% Student	1.1%	1.1%
% Pensioner	3.8%	3.0%
% Conscript	<0.1%	<0.1%
% Outside labor force	5.3%	5.2%

Note: Individuals may be excluded from the FOLK basic dataset for a given year if they were not living in Finland on the last day of that year due to emigration or death (e.g., 31 December 1987).

Table 6. Generalizability descriptives based on 2022 FOLK basic data, comparing twins born 1945–1957 and population reference individuals born same years (age range 65–77)

	Twin	Reference
% of base sample with valid data in 2022 FOLK basic data	74.1%	69.0%
% Deceased	19.4%	19.1%
Marital Status		
% Unmarried	17.1%	13.0%
% Married	53.9%	56.7%
% Divorced	19.6%	20.0%
% Widowed	9.4%	10.3%
Residence		
% Urban	64.0%	65.8%
% Semi-urban	18.5%	17.6%
% Rural	17.5%	16.7%
Main Activity		
% Employed	4.9%	4.6%
% Student	1.0%	1.0%
% Pensioner	94.9%	95.0%
% Outside labor force	0.2%	0.3%

Note: Individuals may be excluded from the FOLK basic dataset for a given year if they were not living in Finland on the last day of that year due to emigration or death (e.g., 31 December 2022).

diagnosis and intervention, but individuals with subclinical alcohol use problems would not be well characterized in register data. On the other hand, in the FTC questionnaires, we have information on alcohol consumption, which is useful for investigating normative use, but since the FTC is a community sample, severe and heavy use are rarer and therefore challenging to investigate using only questionnaire data. The combination of register and questionnaire data provides complimentary information across the spectrum of severity. Questionnaire data additionally provides an advantage in that they contain information on risk factors, exposures, and symptoms not available in register data.

Also, longitudinal studies requiring active participation are limited by potentially losing some of the most important subpopulations during follow-up, such as people who develop alcohol use problems, whereas register data has no such issues with attrition. Using both data sources simultaneously permits a nuanced set of analyses on normative, subclinical, and clinically relevant outcomes. An additional major strength is that register data is free from recall bias, has no loss to follow-up, and is sufficiently detailed for long-term studies.

A limitation with respect to data utility concerns biological data. As part of the Finnish Biobanking Legislation (Law 688/2012), all biological samples associated with the FTC were deposited in 2018 with the nationwide Biobank of the National Institute for Health and Welfare (Kaprio et al., 2019). Biobank data can be linked to FTC questionnaires or to register data, but so far, it has not been possible to combine biological samples with FTC questionnaires and register data simultaneously. Therefore, projects can link two of the three data sources (register data, biobank data, questionnaires), but not all three, to protect participant privacy, and thus the TwinRegistry does not currently contain biological data. An additional limitation is that the full reference population is only available for birth years 1945–1957, therefore we are only able to conduct full-population generalizability analyses for the birth years 1940–1957. Generalizability of the twins born 1880–1944 to their full birth cohort cannot be comprehensively assessed. That said, we have no reason to believe that generalizability of twins to singletons varies by birth year on the outcomes described here.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/thg.2025.10025.

Data Availability. Data for this project can be accessed by approved researchers only via the secure remote computing environment FIONA, provided by Statistics Finland. Broadly speaking, Finnish register data can be obtained by submitting a data permit application for individual-level data to FinData and/or to Statistics Finland (see Viippola et al., 2023, for details). Biobank data can be applied for via THL Biobank, detailed here: https://thl.fi/en/research-and-development/thl-biobank/for-researchers/sample-collections/twin-study. Access to individual-level FTC questionnaire data is more limited, in light of European data protection legislation and the consent forms originally provided to participants of the FTC questionnaires (Kaprio et al., 2019). We are very interested in collaboration and seeing the TwinRegistry resource used to its fullest potential. Please reach out to Antti Latvala (antti.latvala@helsinki.fi) for collaboration inquiries.

Acknowledgments. We acknowledge Statistics Finland's Tutkijapalvelut service for assistance with documentation and computing, as well as the various agencies who have permitted contributions of their registry resources. We also acknowledge the excellent data resource examples provided by FinRegistry and FinnGen, the twins who have participated in decades of research, and to the residents of Finland whose data are included in the registers.

Author contributions. Stephanie Zellers: visualization, writing original draft, project administration, data curation, supervision; Sarah Niemi de Paiva: writing original draft, data curation; Mikko Olkkonen: data curation; Terhi Maczulskij: conceptualization; Hannu Karhunen: conceptualization; Annina Ropponen: conceptualization; Jaakko Kaprio: funding acquisition, supervision, conceptualization; Antti Latvala: supervision, conceptualization. All authors: review and editing of final manuscript.

Funding. Initial permits were funded from grants to AL (grant 77208202 from the Faculty of Social Sciences, University of Helsinki) and to JK (Academy of Finland Center of Excellence in Complex Disease Genetics Grants 336823 & 352792). Additionally, as a supporting activity to ongoing research projects, funding for documentation of the TwinRegistry data resource was provided by the Broad Trauma Initiative at the Broad Institute of MIT and Harvard.

Competing of interest. The authors declare they have no conflicts of interest.

Ethics approval. TwinRegistry is a collaboration project of the Finnish Institute for Health and Welfare (THL) and the University of Helsinki. The TwinRegistry project has received approvals for data access from THL (THL/6165/14.02.00/2020, THL/3987/5.05.00/2021) and Statistics Finland (TK/177/07.03.00/2020-1). The TwinRegistry project has received IRB approval from THL (THL 220/6.02.04/2021). The acceptable level of data granularity and conditions for access were determined in the process of approval from THL and Statistics Finland. Data processing is justified under the basis of public interest.

References

Findata. (n.d.). Findata. https://findata.fi/

ICD-10 Version:2019. (n.d.). https://icd.who.int/browse10/2019/en

Kaprio, J. (2006). Twin Studies in Finland 2006. *Twin Research and Human Genetics*, *9*, 772–777. https://doi.org/10.1375/twin.9.6.772

Kaprio, J. (2013). The Finnish Twin Cohort Study: An update. *Twin Research and Human Genetics*, *16*, 157–162. https://doi.org/10.1017/thg.2012.142

Kaprio, J., Bollepalli, S., Buchwald, J., Iso-Markku, P., Korhonen, T., Kovanen, V., Kujala, U., Laakkonen, E. K., Latvala, A., Leskinen, T., Lindgren, N., Ollikainen, M., Piirtola, M., Rantanen, T., Rinne, J., Rose, R. J., Sillanpää, E., Silventoinen, K., Sipilä, S., … Waller, K. (2019). The Older Finnish Twin Cohort — 45 years of follow-up. *Twin Research and Human Genetics*, *22*, 240–254. https://doi.org/10.1017/thg.2019.54

Kaprio, J., & Koskenvuo, M. (2002). Genetic and environmental factors in complex diseases: The Older Finnish Twin Cohort. *Twin Research and Human Genetics*, *5*, 358–365. https://doi.org/10.1375/twin.5.5.358

Kaprio, J., Sarna, S., Koskenvuo, M., & Rantasalo, I. (1978). The Finnish Twin Registry: Formation and compilation, questionnaire study, zygosity determination procedures, and research program. *Progress in Clinical and Biological Research*, *24 Pt B*, 179–184.

Keller, M. C., Medland, S. E., & Duncan, L. E. (2010). Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. *Behavior Genetics*, *40*, 377–393. https://doi.org/10.1007/s10519-009-9320-x

Kurki, M. I., Karjalainen, J., Palta, P., Sipilä, T. P., Kristiansson, K., Donner, K. M., Reeve, M. P., Laivuori, H., Aavikko, M., Kaunisto, M. A., Loukola, A., Lahtela, E., Mattsson, H., Laiho, P., Della Briotta Parolo, P., Lehisto, A. A., Kanai, M., Mars, N., Rämö, J., … Palotie, A. (2023). FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*, *613*, 508–518. https://doi.org/10.1038/s41586-022-05473-8

McAdams, T. A., Hannigan, L. J., Eilertsen, E. M., Gjerde, L. C., Ystrom, E., & Rijsdijk, F. V. (2018). Revisiting the Children-of-Twins Design: Improving Existing Models for the Exploration of Intergenerational Associations. *Behavior Genetics*, *48*, 397–412. https://doi.org/10.1007/s10519-018-9912-4

Ministry of Social Affairs and Health. (n.d.). *Secondary use of health and social data*. Ministry of Social Affairs and Health. https://stm.fi/en/secondary-use-of-health-and-social-data

Niemi de Paiva, S., Hukkanen, M., Kaprio, J., & Zellers, S. (2025). *Investigating fertility rates across generations in monozygotic twins, dizygotic twins, and non-twin controls using the Finnish Twin Cohort*. OSF. https://doi.org/10.17605/OSF.IO/QBWV3

Risteys. (n.d.). https://risteys.finregistry.fi/

Tilastokeskus. (n.d.). *FIONA remote access system*. Statistics Finland. https://stat.fi/tup/tutkijapalvelut/fiona-etakayttojarjestelma_en.html

Viippola, E., Kuitunen, S., Rodosthenous, R. S., Vabalas, A., Hartonen, T., Vartiainen, P., Demmler, J., Vuorinen, A.-L., Liu, A., Havulinna, A. S., Llorens, V., Detrois, K. E., Wang, F., Ferro, M., Karvanen, A., German, J., Jukarainen, S., Gracia-Tabuenca, J., Hiekkalinna, T., … Ganna, A. (2023). Data Resource Profile: Nationwide registry data for high-throughput epidemiology and machine learning (FinRegistry). *International Journal of Epidemiology*, *52*, e195–e200. https://doi.org/10.1093/ije/dyad091

Zellers, S., Azzi, E., Latvala, A., Kaprio, J., & Maczulskij, T. (2025). Causally-informative analyses of the effect of job displacement on all-cause and specific-cause mortality from the 1990s Finnish recession until 2020: A population registry study of private sector employees. *Social Science & Medicine*, *370*, 117867. https://doi.org/10.1016/j.socscimed.2025.117867