Data Resource Profile: Linking electronic health and social records to study and lower health inequalities in cardiovascular diseases (BIG-HEART)

Laura Lõo*[a], Nikita Umov[a], Marek Oja[b], Sulev Reisberg[b,c], Anneli Uusküla[a], Raivo Kolde[b], Taavi Tillmann[a]

[a] Institute of Family Medicine and Public Health, University of Tartu, Tartu, Estonia

[b] Institute of Computer Science, University of Tartu, Tartu, Estonia

[c] STACC, Tartu, Estonia

Corresponding author: Laura Lõo, laura.loo@ut.ee

Abstract

The BIG-HEART cohort was established to study and reduce health inequalities in cardiovascular disease by linking rich, multidimensional electronic health and social data across Estonia. The dataset includes all individuals aged 36 and above residing in Estonia in 2012 (N= 770,323). Its full population coverage minimises sampling and healthy volunteer bias. Existing funding and permits will support annual health outcome follow-up through at least 2026, with possible future extensions. The dataset integrates all routinely collected individual-level primary and secondary care health data (including in- and outpatient visits, diagnoses, prescriptions issued and filled), mortality data, and extensive social data (e.g. ethnicity, education, marital status, social benefits, unemployment history, land and business ownership) from eight national registries. This enables exploration of novel social epidemiology dimensions—such as unbiased wealth measures, medication adherence, and care quality—and supports development of equity-enhancing clinical risk prediction algorithms and large language models. Health and social data are linked using pseudonymised identifiers derived from national personal identification numbers, ensuring accuracy and privacy. The data are stored in the OMOP common data model, facilitating international collaboration. Collaboration inquiries are welcome and can be directed to the BIG-HEART team at taavi.tillmann@ut.ee.

Data resource basics

The rationale for establishing BIG-HEART

Cardiovascular diseases (CVD) remain a leading global public health concern despite being largely preventable [1, 2]. 60 years of research has shown a strong association between socioeconomic factors and CVD risk [3]. However, these insights have rarely been integrated into clinical practice. Primary and secondary prevention services can exacerbate health inequalities when attendance at screening programs, preventive medication use, and success with behaviour change services follow social gradients [4–6]. We propose that integrating multidimensional socioeconomic and psychosocial data with electronic health records can first deepen our understanding of the drivers and patterns of health inequity. Second, it can accelerate implementation by enabling developers of risk prediction algorithms and care services to directly incorporate socioeconomic factors into their models and pathways. Existing CVD risk prediction models recommended in clinical practice [7–9] typically incorporate at most one socioeconomic indicator, despite emerging evidence suggesting that each additional layer may enhance both predictive performance and fairness [10–11]. This lack of integration may worsen inequalities from the earliest stage of care—risk prediction—compounded by unequal uptake and adherence to recommended services [12–15]. Estonia has among the highest cardiovascular disease mortality rates in Europe—nearly twice the EU average—and modestly above the global average, which likely reflects higher disease incidence rather than poorer outcomes after diagnosis [16–19]. This underscores both the statistical power and public health potential of conducting CVD research in Estonia. BIG-HEART integrates health and social data to bridge basic epidemiological research with applied solutions, promoting more precise and equitable risk assessments and services.

Estonian health and social data

Estonia, a country in Northeastern Europe with a population of 1.3 million, has long led globally in building secure and efficient digital infrastructures for public services. Since 1989, every resident has been assigned an 11-digit personal identification code, facilitating data linkage across government systems. In 2001, Estonia launched "X-Road"—a secure data exchange system enabling interoperability across governmental databases while maintaining strict security standards [20]. Managed by the Estonian Information System Authority, this

infrastructure supports resilience to changes in underlying database architectures or digital services.

As of 2023, approximately 95% of Estonia's population is covered by the publicly funded Estonian Health Insurance Fund (EHIF) [21], similar to the UK's National Health Service [22, 23]. Estonia's healthcare system includes roughly 400 primary care centers [24] and 20 hospitals [25], each with its own IT system. These local systems are not fully interoperable, allowing flexibility in implementation. However, providers must submit claims, prescriptions, and discharge summaries to the national system, enabling comprehensive, standardised national coverage. Estonia's national e-prescribing platform records all prescribed medications as well as dispensing status, allowing assessment of medication initiation, adherence, and persistence across population subgroups. Blockchain technology logs all clinician access to personally identifiable health data, enhancing transparency and public trust. Residents can access their records via the national Health Portal [26], similar to systems in Finland and Iceland [27].

Secondary use of Estonian health data for research

As the collection and coding of electronic health data differ across countries, there is growing interest in harmonising cleaned, research-ready datasets rather than raw data. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [28] has become a widely adopted standard for enabling cross-country comparisons and large-scale analyses. Estonia contributed early to this effort, first mapping a 10% random sample of national health data to OMOP [29], followed by a 30% sample that supported several international COVID-19 studies. These analyses provided valuable insights—for example, they confirmed that post-acute COVID-19 symptoms defined by the WHO were frequently reported, though with variation across healthcare settings and countries [30]. They also demonstrated that vaccination, particularly the first dose, significantly reduced the risk of long COVID [31], and showed that serious adverse events like thromboembolism and myocarditis were more common after SARS-CoV-2 infection than in historical cohorts [32].

BIG-HEART: Advancing health and social data integration

Building on Estonia's electronic health record infrastructure, BIG-HEART further integrates data by enriching health records with socioeconomic data. BIG-HEART was established in 2024, covering all 770,323 Estonian residents aged 36 and older who as of 01.01.2012 were resident in Estonia according to the Population Register. The cohort is tracked annually for health and mortality outcomes through 2026. Embedded in the OMOP CDM, the cohort supports harmonised, large-scale analyses and international collaborations on health equity.

Data collected

Data sources

BIG-HEART is both a longitudinal cohort and a data resource, enabling analysis of health and social exposures over time. It is based on routinely collected, secondary administrative data, structured as a register-based cohort with repeated annual updates. The cohort was constructed using the Population Register, which defined the study population. Table 1 shows baseline demographic characteristics. Data are grouped into (1) health and (2) social domains. Eight national datasets were linked at the individual level via project-specific pseudonyms. Figure 1 provides an overview of the linked data sources, coverage, and age distribution.

Table 1. Baseline demographic characteristics of the BIG-HEART cohort (N=770,323) as of 01.01.2012, from the Population Register.

Characteristic	Number of people	%
Age		
Mean (SD, range)	57.4	14, 36–108
Median (IQR)	56	46–68
Male sex	334,151	43.4
Citizenship		
Estonian	608,990	79.0
Russian	79,847	10.4
Ukrainian	4723	0.6
Finnish	4183	0.5
Latvian	1607	0.2
Lithuanian	1124	0.1
Other	69,758	9.1
Missing	91	0.01
Self-reported nationality		
Estonian	499,090	64.8
Russian	204,487	26.5
Ukrainian	20,563	2.7
Finnish	8055	1.0
Latvian	1679	0.2
Lithuanian	1468	0.2
Other	28,282	3.7
Missing	6699	0.9
Education level[a]		
Less than primary	3389	0.4
Primary (ca age 11)	34,198	4.4
Lower secondary (ca age 15)	94,820	12.3
Lower secondary, vocational	15,564	2.0
Upper secondary (ca age 18)	192,709	25.0
Upper secondary, vocational	73,164	9.5
Upper secondary, specialised	141,479	18.4
Bachelor's degree	37,645	4.9
Master's degree	126,368	16.4
Doctorate	4,904	0.6
Missing	46,083	6.0
Marital status		
Married	379,490	49.3
Divorced	155,869	20.2
Widow/widower	100,471	13.0
Single	111,964	14.5
Missing	22,529	2.9

[a] Education levels and associated ages are aligned with the Soviet-era school structure, as the youngest individuals in the dataset were born prior to Estonia's post-independence educational reforms. Certain categories (such as upper secondary specialised) have since been discontinued in the modern system.

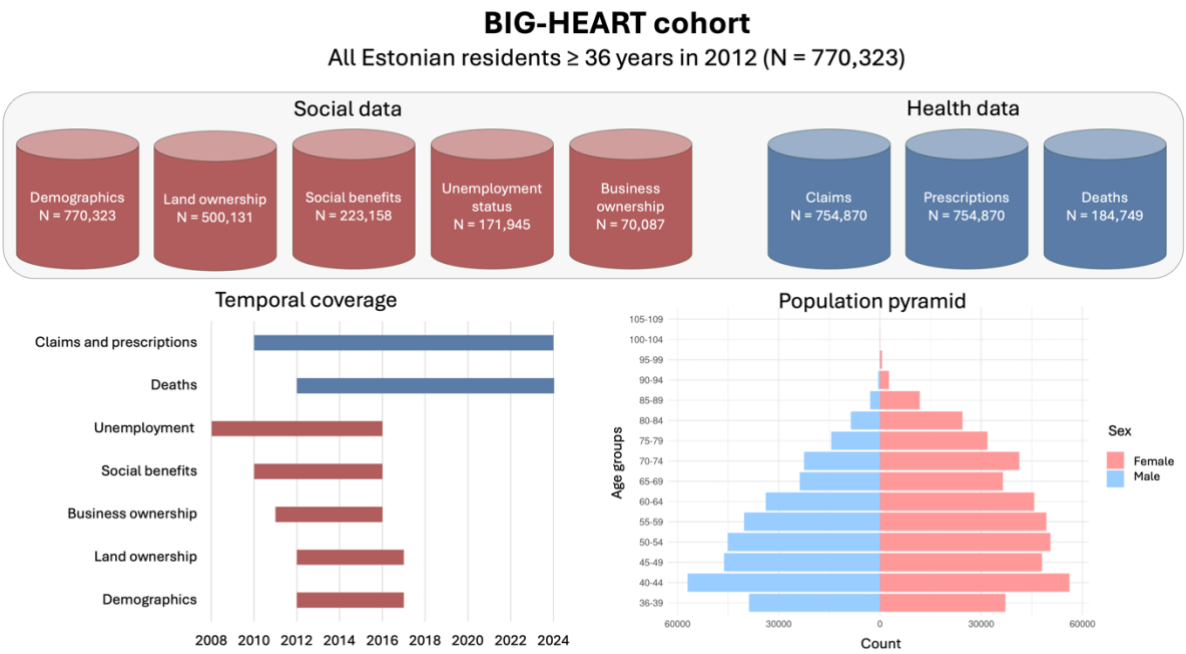

Figure 1. Data sources included in data linkage, data timeliness and population pyramid of the BIG-HEART cohort.

Health data are primarily derived from a national billing database managed by the EHIF [33]. This comprehensive dataset includes all reimbursed insurance claims submitted by primary and secondary healthcare providers, covering both in- and outpatient services. Each claim records the procedures performed and diagnoses—both incident and pre-existing—coded according to the International Statistical Classification of Diseases, 10th Revision (ICD-10), and the NOMESCO Classification of Surgical Procedures. Table 2 highlights key diseases of public health relevance, primarily based on the top 30 conditions ranked by Disability Adjusted Life Years (DALY), along with other commonly reported cases from the BIG-HEART database.

Table 2. Number of cases of selected diseases with significant public health impact, based on DALY rankings and commonly reported conditions, in the BIG-HEART cohort (N=770,323). Data includes individuals with prevalent disease at baseline or incident disease during follow-up.

Diagnosis	ICD-10 code	Number of people	%
Hypertensive heart disease	I10–15	487,320	63.3
Dorsalgia	M54	413,701	53.7
Dyslipidaemia	E78	327,025	42.5
Falls	W00–19	314,609	40.8
COVID-19	U07	270,460	35.1
Virus identified	U07.1	176,358	22.9
Virus not identified	U07.2	139,547	18.1
Heart failure	I50	188,465	24.5
Ischemic heart disease	I21–25	136,746	17.8
Type 2 diabetes	E11	128,924	16.7
Cerebrovascular disease	I60–69	126,331	16.4
Depression	F32–33	124,628	16.2
Anxiety	F41	99,288	12.9
Asthma	J45	78,040	10.1
Chronic obstructive pulmonary disease	J44	68,660	8.9
Chronic kidney disease	N18	63,627	8.3
Age-related hearing loss	H91	58,593	7.6
Prostate cancer	C61	22,693	2.9
Migraine	G43	22,334	2.9
Colorectal cancer	C18–21	18,119	2.4
Breast cancer	C50	17,320	2.2
Psychotic disorders	F20–25, F28–29	14,568	1.9
Type 1 diabetes	E10	13,621	1.8
Lung cancer	C34	13,294	1.7
Alcohol addiction or use of specialist services	Z71.4, Z72.1, F10	12,233	1.6
Tobacco addiction or use of cessation services	Z71.6, Z72.0, F17	10,519	1.4
Other COVID outcomes	U09	6204	0.8
Pancreatic cancer	C25	4692	0.6
Use of psychoactive substances	F11–16, F18–19	4234	0.5
Cervical cancer	C53	3696	0.5
Alzheimer's disease	G30	3141	0.4
Manic episode	F31	2189	0.3
COVID-19 vaccine causing adverse effects	U12	803	0.1
Multisystem infammatory syndrome associated with COVID-19	U10	349	0.05
Bipolar affective disorder	F30	317	0.04

The EHIF database contains comprehensive information on sick leave episodes and health insurance coverage, including the start and end dates of coverage and the reasons behind changes in status. Our data suggest that 6–8% of individuals may experience fluctuations in their insurance coverage, implying that stable coverage is closer to 91%, rather than the usually reported 95%. These fluctuations may be due to changes in employment or registration status and should be taken into account when interpreting analyses based on insurance data. In addition to coverage and sick leave data, the EHIF also maintains an extensive prescription records system. This includes detailed information on all dispensed medications—classified using the Anatomical Therapeutic Chemical system—along with the prescribed dosage, prescribing provider, and both the date and location of dispensing. Metadata from packaging codes further enrich these records by identifying the drug manufacturer, strength, and pharmaceutical form (e.g., tablet, cream).

The Causes of Death Register [34] collects comprehensive mortality data. In the BIG-HEART cohort, we have full details on the cause of death when the underlying cause is circulatory (ICD-10 codes I00–I99) or sudden death of unknown cause (R96.0, R96.1). For all other causes, the date of death is recorded, which supports time-to-event analyses and the handling of competing non-cardiovascular outcomes. Should future research require detailed cause-of-death data beyond the circulatory domain, extended access can be requested to support scientifically justified hypotheses.

Social data are currently available from five registries. Table 3 presents the socioeconomic and health-related characteristics of the BIG-HEART cohort across the study period. While the primary purpose of collecting social data was to characterise baseline exposures—not to assess the incidence of social events—future extensions may incorporate variables like occupation and income if supported by research hypotheses. Consequently, social data have shorter time coverage than health data. The Population Register [35] includes demographic and migration variables such as sex, birth year, marital status, citizenship, self-reported nationality, county of residence, emigration date, and death date, as well as the most recently recorded educational attainment. Since education data were provided as a 2023 snapshot rather than time-stamped records, we cannot determine participants' education status at baseline and assume no changes during follow-up.

Table 3. Socioeconomic and health characteristics of the BIG-HEART cohort (N=770,323) over the full study period, from linked databases.

Data source	Characteristic	Number of people (%)
Social Insurance Records Database	Receiving any social benefit at least once between 2010–2016	223,158 (29.0)
	Benefit for having a disability	164,940 (21.4)
	Lifelong pension for having no capacity to work	105,951 (13.8)
	Victim of Soviet political repression during 1945–1991	11,874 (1.5)
	Occupational injury/disease	2363 (0.3)
	Forced to work during the Chernobyl clean up in 1986–1989	1464 (0.2)
E-Business Register	*If analysing people as the unit of analysis*	
	Number of companies owned by one participant, at baseline	
	0	730,726 (94.9)
	1	32,631 (4.2)
	2	5057 (0.7)
	3	1250 (0.2)
	>3	659 (0.1)
	If analysing companies as the unit of analysis, at baseline	
	Number of companies based on their size	
	Micro enterprises (0–9 employees)	40,753
	Small enterprises (10–49 employees)	3142
	Medium and larger enterprises (≥ 50 employees)	512
	Missing	34
	Average company's staff size	
	Median	1
	Mean	5.6
	Median profit (in euros, €)	742
	Median accumulated profit prior to baseline (€)	5894
	Dividends paid (€)	
	Median	5112
	Mean	35,389
	Number of loss-making companies	13,975
	Number of companies with accumulated losses	7374
E-Land Register	Property ownership	500,131 (64.9)
	Owners of only non-residential property	96,826 (19.4)
	Owners of some residential property	403,305 (52.4)
	Owners of residential property in a major city[a]	232,355 (30.2)
	Owners of residential property in the capital city[b]	153,156 (19.9)
Unemployment Insurance Fund	*If analysing people as the unit of analysis*	
	Unemployed at least once between 2008–2016	171,945 (22.3)
	Unemployed ≥ 2 times between 2008–2016	72,891 (9.5)
	If analysing all unemployment events (N= 319,338) as the unit of analysis	
	Median unemployment period in days (IQR) per episode	181 (76–337)
	Median time in days from finding a job until becoming unemployed again (IQR)	149 (62–297)

	Reason for registering as "unemployed", top 3 reasons	
	Redundancy	41,781 (13.1)
	End of fixed-term contract	35,059 (11.0)
	Employee's initiative	20,403 (6.4)
	Reason for terminating the "unemployed" status, top 3 reasons	
	New employment	189,006 (59.2)
	Breach of unemployment obligations	80,085 (25.1)
	Requested by unemployed person	39,454 (12.4)
	People covered by the national health insurance system	
Estonian Health Insurance Fund	Insured at least once between 2012–2023	754,870 (98.0)
	Insured at baseline (01.01.2012)	703,433 (91.3)
	Individuals with ≥ 90 days of sick leave	29,798 (3.9)
	All-cause deaths, between 2012–2023	184,749 (24.0)
Death Register	CVD deaths (I00–99), between 2012–2023	95,019 (12.3)
	Atherosclerotic death (I21–25 + I60–69), between 2012–2023	45,771 (5.9)

[a] Tallinn, Tartu, Narva, or Pärnu
[b] Tallinn

The Social Insurance Records Database [36] covers social benefits – such as type, amount, and payment period – for various reasons, including having a disability, being a victim of severe violence, political repression during the Soviet Union from 1945 to 1991, being forced to work during the Chernobyl clean-up operation from 1986 to 1989, living alone as a pensioner, suffering health damage from an occupational disease or work accident, or receiving an incapacity pension. This data is available from 2010 to 2016.

The Unemployment Insurance Fund data source [37] provides longitudinal data on unemployment episodes, prior occupations, and reasons for entering and exiting unemployment. For the ~25% of participants with at least one unemployment spell, it is possible to reconstruct pre-unemployment income, enabling nuanced analyses of economic vulnerability. Occupations are coded using the International Standard Classification of Occupations, allowing longitudinal tracking of employment stability and career disruption. Evidence suggests that the frequency of labour market transitions may be more predictive of disease incidence than unemployment duration [38]. Among the 10% of participants with multiple unemployment episodes, employment and unemployment often alternate in six-month cycles. The BIG-HEART cohort enables investigation of the contextual and personal factors influencing unemployment resolution—whether through re-employment, retirement, or other paths. Currently, the cohort includes 8 years of employment data, which could be extended to 16 years in future updates if needed.

Objective wealth indicators are available from two registries. The E-Land Register [39] records all landholdings—residential or commercial—owned by participants, down to the

neighborhood level (excluding street names). As shown in Table 2, about 65% of the BIG-HEART cohort owns some form of land, enabling rare analyses of how real estate wealth relates to life expectancy, healthcare use, and outcomes—supporting fairness assessments across socioeconomic strata.

The E-Business Register [40] provides data on approximately 40,000 individuals who co-own or operate a legal entity. It offers six years of detailed longitudinal information, including turnover, profitability, tax arrears, employee count, and sector classification using NACE codes. Among the broader population, a smaller share—about 5%—owns business assets, typically through micro-enterprises. This supports research on how financial stability or instability in business ownership affects physical and mental health, healthcare use, and outcomes. Wealth across property and business domains offers a multidimensional view of financial security, precarious employment, and health—especially relevant during external shocks like the COVID-19 pandemic. The dataset could be extended to examine whether business disruptions led to differences in healthcare access, diagnoses, or medication use.

Secure data processing

To protect privacy, the Population Register assigned a unique, project-specific pseudonym (data key) to each person, replacing national personal identification numbers. These pseudonyms were shared with data providers to enable secure record linkage, but not with the University of Tartu, which never had access to identifiable information. Each data provider used the pseudonym to extract relevant records and returned pseudonymised datasets to a secure computing environment, ensuring that researchers only accessed de-identified data.

The BIG-HEART cohort is stored within one of Estonia's highest-security computing infrastructures [41], featuring 13,000 CPU cores and 64 GPUs. Approved researchers access the environment remotely via a Virtual Private Network, using a browser-based virtual desktop interface. The system is fully isolated from the public internet. To ensure data protection, copy-pasting is disabled, and all user activity—including video recordings of analytic sessions—is logged for auditing. Data imports and exports are controlled through secure S3 gateways, with every export requiring manual review and approval by a designated principal investigator.

Data resource use

The BIG-HEART dataset became available to us in the middle of 2024, making it a comparatively new and highly promising resource for studying socioeconomic dimensions of health and healthcare. Although no studies have been published to date, several have been initiated by our comparatively small team to date. Among healthy participants at baseline, we are investigating whether registry-based measures of real estate and business wealth, unemployment episodes, and social benefits predict CVD incidence, and whether such predictors add value to CVD risk prediction algorithms. We are also curious to see if machine learning, artificial intelligence and large language models will create risk prediction models that outperform conventional parametric regression models, both in terms of overall accuracy and algorithm fairness, which the BIG-HEART cohort is exceptionally well placed to evaluate with its rich array of socioeconomic layers and complete population representation. Much of this work is used to inform the design of future interventions, such as identifying persons who are amenable to preventative care (with statins, hypertensives or behaviour change services) and evaluating the effectiveness of novel interventions.

For persons with established disease at baseline, we are describing sociodemographic variations in the uptake and impact of healthcare services, including attendance at screening programmes and medication adherence. We are extending this to also predict which individuals have greater risk of not adhering to medications in the future, and considering pilots to feed this back to frontline clinicians to increase adherence. Beyond these initial opening studies, BIG-HEART is well positioned to address many public health questions that lie at the intersection between computer science, epidemiology, social sciences and the more applied development of novel software as medical devices and related healthcare services. We are keen to collaborate on proposals that either a) analyse the existing data to answer novel questions (which can take as little as 3–6 months to generate results), or b) to add novel data dimensions and layers to the existing dataset via Estonian regulators (which usually takes around 12 months).

Strengths and weaknesses

BIG-HEART has three core strengths. First, it includes a complete population sample: all Estonian residents aged 36 and above were included via the Population Register. This eliminates sampling bias and sets BIG-HEART apart from biobanks and survey-based cohorts, where participation is voluntary and often limited (5–50% response rates). Second, it offers complete healthcare service capture, due to Estonia's national electronic infrastructure and OMOP-standardised data. These features support accurate estimation of healthcare use, and of conditions that usually result in clinical contact. Third, BIG-HEART integrates rich social data at the population level. These data, though not originally collected for research, are nearly complete and enable rigorous analysis of social determinants. Social epidemiologists can evaluate how socioeconomic factors affect disease, while data scientists can validate prediction models and assess fairness across subgroups.

Despite these strengths, BIG-HEART has limitations. Some valuable data layers remain unlinked, including the Health Information System (with discharge summaries, clinical notes, and lab results), Estonian Tax Board registries (with declared income and occupation for those never unemployed), and disease-specific registries for myocardial infarction and cancer. Presented health needs may also underestimate true needs. Chronic conditions (e.g. hypertension, obesity, depression) often precede care-seeking and may go undocumented, particularly among disadvantaged groups. This could bias risk estimates and fairness evaluations toward the null. Although official coverage is 95%, baseline data show that only 91% were insured at linkage, and 6–8% cycle in and out of coverage. If uninsured individuals have higher health needs, prevalence and fairness measures may be overly optimistic. Finally, some individuals may have emigrated before baseline or during follow-up without updating their residency. These misclassified participants may be incorrectly considered at risk, biasing estimates toward the null. While we have not yet quantified this, Estonia's strong incentives for residents to update their status (e.g. for tax, transport, childcare, or mortgage benefits) likely reduce the problem. Future work may identify likely emigrants based on complete disengagement from health and social systems.

Data resource access

Access to the BIG-HEART data can be obtained with approval from the Research Ethics Committee of the University of Tartu. Researchers can contact the principal investigator Taavi Tillmann through taavi.tillmann@ut.ee.

Ethics approval

The BIG-HEART project has received approval from the Research Ethics Committee of the University of Tartu [no. 384/T-8, 20.11.2023].

Data availability

See Data resource access above.

Author contributions

T.T. conceived the original research idea and supervised the project. Data analysis was conducted by L.L. and N.U. Data processing and curation were carried out by L.L., N.U., M.O., S.R., and R.K. The original draft of the manuscript was written by L.L., and all authors reviewed, edited, and approved the final version. Funding acquisition was handled by T.T. and R.K.

Funding

This work was supported by the Estonian Research Council grants PSG809, PRG1844, and PRG2218. This research was co-funded by the European Union through the European Regional Development Fund (Project No. 2021-2027.1.01.24-0444) and through Estonian Ministry of Education and Research (project TEM-TA72; and grant TK218 establishing the Estonian Centre of Excellence for Well-Being Sciences (EstWell)).

Acknowledgements

We acknowledge the contributions of individuals and teams who established the cohorts and data resources that inspired the foundation of BIG-HEART. These include the bespoke CALIBER study of electronic health records (Spiros Denaxas and Harry Hemingway), the

Whitehall II and HAPIEE social epidemiological cohorts (Michael Marmot, Mika Kivimäki, and Martin Bobak), and FinRegistry (Andrea Ganna).

Conflict of interest

None declared.

References

1. Mensah GA, Fuster V, Murray CJL, Roth GA. Global burden of cardiovascular diseases and risks, 1990–2022. *J Am Coll Cardiol* 2023;82:2350–73.
2. The Global Cardiovascular Risk Consortium. Global effect of modifiable risk factors on cardiovascular disease and mortality. *N Engl J Med* 2023;389:1273–85.
3. Hinkle LE, Whitney LH, Lehman EW, et al. Occupation, education, and coronary heart disease. Risk is influenced more by education and background than by occupational experiencess, in the Bell System. *Science* 1968;161:238–46.
4. Murphy A, Palafox B, O'Donnell O, et al. Inequalities in the use of secondary prevention of cardiovascular disease by socioeconomic status: evidence from the PURE observational study. *Lancet Glob Health* 2018;6:e292–301.
5. Lang S-J, Abel GA, Mant J, et al. Impact of socioeconomic deprivation on screening for cardiovascular disease risk in primary prevention population: a cross-sectional study. *BMJ Open* 2016;6:e009984.
6. Ohm J, Skoglund PH, Häbel H, et al. Association of socioeconomic status with risk factor target achievements and use of secondary prevention after myocardial infarction. *JAMA Netw Open* 2021;4:e211129.
7. SCORE2 working group, ESC cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J* 2021;42:2439–54.
8. Hippisley-Cox J, Coupland CAC, Bafadhel M, et al. Development and validation of a new algorithm for improved cardiovascular risk prediction. *Nat Med* 2024;30:1440–7.
9. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: a report of the American College of

Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* 2019;140:e596–646.

10. Tillmann T, Vaucher J, Okbay A, et al. Education and coronary heart disease: a Mendelian randomization study. *BMJ* 2017;358:j3542.
11. Tillmann T, Läll K, Dukes O, et al. Development and validation of two SCORE-based cardiovascular risk prediction models for Eastern Europe: a multicohort study. *Eur Heart J* 2020;41:3325–33.
12. Kimenai DM, Pirondini L, Gregson J, et al. Socioeconomic deprivation: an important, largely unrecognized risk factor in primary prevention of cardiovascular disease. *Circulation* 2022;146:240–8.
13. Dalton JE, Perzynski AT, Zidar DA, et al. Accuracy of cardiovascular risk prediction varies by neighborhood socioeconomic position: A retrospective cohort study. *Ann Intern Med* 2017;167:456–64.
14. Lang S-J, Abel GA, Mant J, Mullis R. Impact of socioeconomic deprivation on screening for cardiovascular disease risk in a primary prevention population: a cross-sectional study. *BMJ Open* 2016;6:e009984.
15. Rangachari P, Thapa A. Impact of hospital and health system initiatives to address Social Determinants of Health (SDOH) in the United States: a scoping review of the peer-reviewed literature. *BMC Health Serv Res* 2025;25:342.
16. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020;396:1223–49.
17. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, et al. Myocardial infarction and coronary deaths in the WHO MONICA Project: registration procedures, event rates, and case fatality in 38 populations. *Circulation* 1994;90:583–612.
18. Tunstall-Pedoe H, Kuulasmaa K, Mähönen M, et al. Contribution of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA populations. *Lancet* 1999;353:1547–57.
19. Tillmann T, Pikhart H, Peasey A, et al. Psychosocial and socioeconomic determinants of cardiovascular mortality in Eastern Europe: a multicentre prospective cohort study. *PLoS Med* 2017;14:e1002459.

20. Estonian Information System Authority. Project “X-Road” history. Available from: https://x-road.global/xroad-history (accessed May 2025).
21. Estonian Health Insurance Fund. Estonian Health Insurance Fund Annual Report of Financial Year 2023. 2024.
22. National Health Service. NHS Health Check. Available from: https://www.nhs.uk/conditions/nhs-health-check/ (accessed May 2025).
23. Ferris TG, Dixon J, Blumenthal D. Lessons from England’s National Health Service. *N Engl J Med* 2024;391:1852–9.
24. Estonian Health Insurance Fund. All family physicians working list Estonia. Available from: https://www.tervisekassa.ee/en/people/health-care-services/primary-health-care/all-family-physicians-working-list-estonia(accessed May 2025).
25. Ministry of Social Affairs. Haiglavõrk ja haiglavõrgu arengukavad. Available from: https://www.sm.ee/haiglavork-ja-haiglavorgu-arengukavad (accessed May 2025).
26. Health Portal. Available from: https://www.terviseportaal.ee/en/ (accessed May 2025).
27. Laugesen K, Ludvigsson JF, Schmidt M, et al. Nordic health registry-based research: a review of health care systems and key registries. *Clin Epidemiol* 2021;13:533–54.
28. OHDSI. Standardized data: the OMOP Common Data Model. Available from: https://www.ohdsi.org/data-standardization/ (accessed May 2025).
29. Oja M, Tamm S, Mooses K, et al. Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned. *JAMIA Open* 2023;6:ooad100.
30. Xie J, Lopez-Guell K, Dedman D, et al. Incidence of post-acute COVID-19 symptoms across healthcare settings in seven countries: an international retrospective cohort study using routinely collected data. *EClinicalMedicine* 2024;77:102903.
31. Catala M, Mercade-Besora N, Kolde R, et al. The effectiveness of COVID-19 vaccines to prevent long COVID symptoms: staggered cohort study of data from the UK, Spain, and Estonia. *EClinicalMedicine* 2024;12:225–36.
32. Voss EA, Shoaibi A, Lai LYH, et al. Contextualising adverse events of special interest to characterise the baseline incidence rates in 24 million patients with COVID-19 across 26 databases: a multinational retrospective cohort study. *EClinicalMedicine* 2023;58:101932.

33. Estonian Health Insurance Fund. Available from: https://www.tervisekassa.ee/en (accessed May 2025).
34. The National Institute for Health Development. Causes of Death Register. Available from: https://www.tai.ee/en/statistika-ja-registrid/causes-death-register (accessed May 2025).
35. E-Population Register. Available from: https://www.rahvastikuregister.ee/ (accessed May 2025).
36. Social Insurance Board. Social Insurance Records Database. Available from: https://iseteenindus.sotsiaalkindlustusamet.ee/dashboard/portal/login?url=%2Fdashboard%2Fportal (accessed May 2025).
37. Estonian Unemployment Insurance Fund. Available from: https://www.tootukassa.ee/en (accessed May 2025).
38. Dupre ME, George LK, Liu G, et al. The cumulative effect of unemployment on risks for acute myocardial infarction. *Arch Intern Med* 2012;172:1731–7.
39. E-Land Register Portal. Available from: https://www.rik.ee/en/e-land-register/e-land-register-portal (accessed May 2025).
40. E-Business Register. Available from: https://ariregister.rik.ee/eng (accessed May 2025).
41. HPC Public Documentation. SAPU. Available from: https://docs.hpc.ut.ee/public/services/SAPU/ (accessed May 2025).