Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers

Nina Mars, MD, PhD[1] Jukka T. Koskela, MD, PhD [1] Pietari Ripatti, MD, [1] Tuomo T.J. Kiiskinen, MD,[1] Aki S. Havulinna, DSc (tech),[1,2] Joni V. Lindbohm, MD, PhD[3] Ari Ahola-Olli, MD, PhD[1] Mitja Kurki, PhD,[1,4,5], Juha Karjalainen, PhD,[1,6,7] Priit Palta, PhD,[1,8] FinnGen, Benjamin M. Neale, PhD,[4,7] Prof Mark Daly, PhD,[1,6] Veikko Salomaa, MD, PhD[2] Prof Aarno Palotie, MD, PhD[1,5,6] Elisabeth Widén, MD, PhD[1] Prof Samuli Ripatti, PhD[1,3,6]*

1) Institute for Molecular Medicine Finland, FIMM, HiLIFE, University of Helsinki, Helsinki, Finland 2) Finnish Institute for Health and Welfare, Helsinki, Finland 3) Department of Public Health, Clinicum, University of Helsinki, Helsinki, Finland, 4) Program in Medical and Population Genetics and Genetic Analysis Platform, Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA, 5) Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Analytic and Translational Genetics Unit, Department of Medicine, and the Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA 6) Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA 7) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA 8) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia

* **Corresponding author**

Tel: +358 40 567 0826

Email: samuli.ripatti@helsinki.fi

Address: Institute for Molecular Medicine (FIMM), Biomedicum 2U, Tukholmankatu 8, 00290, Helsinki, Finland

ABSTRACT

Polygenic risk scores (PRS) have shown promise in predicting susceptibility to common diseases.[1-3] We estimated their added value in clinical risk prediction of five common diseases, using large-scale biobank data (FinnGen; N=135,300), and the FINRISK study with clinical risk factors to test genome-wide PRSs for coronary heart disease (CHD), type 2 diabetes (T2D), atrial fibrillation (AF), and breast and prostate cancer. We evaluated the lifetime risk at different PRS levels, and the impact on disease onset and on prediction together with clinical risk scores. Compared to average PRS, having a high PRS contributed to 21% to 38% higher lifetime risk, and 4 to 9 years earlier disease onset. PRS improved model discrimination over age and sex in T2D, AF, breast cancer, and prostate cancer, and over clinical risk in T2D, breast cancer, and prostate cancer. In all diseases, PRS improved reclassification over clinical thresholds, with largest net reclassification improvements for early-onset CHD, AF, and prostate cancer. This study provides evidence for the additional value of PRS in clinical disease prediction. The practical applications of polygenic risk information for stratified screening or for guiding lifestyle and medical interventions in the clinical setting remain to be defined in further studies.

Common chronic diseases present a huge burden to societies, with an estimated one billion prevalent cases diagnosed with cardiovascular diseases, diabetes, or neoplasms worldwide.[4] Consequently, the development of strategies to prevent these diseases is critically important. To facilitate prevention, a clear understanding of individual risk is essential to determine whether an individual warrants an intervention as well as to gauge the impact of different interventions. These risk models typically incorporate clinical and laboratory-based risk factors, and can identify individuals at high risk suitable for selective prevention strategies, such as prescribing cholesterol-lowering medications for reducing coronary heart disease (CHD) risk.[5] Although clinical risk scores enable the identification of individuals who may benefit from preventive interventions, they come with some limitations. For instance, cardiovascular risk calculators fail to identify up to 40% of persons who develop the diseases, and their utility is limited among young individuals.[5,6] For breast cancer, many reproductive, hormonal, and lifestyle risk factors are common. However, they are relatively weak predictors, identifying only a small fraction of those at high long-term risk.[7] Positive family history is an important risk factor in most cardiometabolic diseases and common cancers, but its utility is limited by aspects such as the number, age, and type of relatives affected.[8]

Large-scale genetic screens comparing disease cases with controls have identified thousands of genetic loci associated with risk of complex disorders,[9] suggesting that genomic information has become a promising candidate for improving clinical risk assessment.[10,11] While, individually, the associated loci typically modify the disease risks only marginally, for many diseases the cumulative impact of risk across the genome is considerable.[12] Polygenic risk scores (PRS) measuring this cumulative genetic burden[13] have recently been shown to correlate with case status in many complex diseases including CHD, type 2 diabetes (T2D), and breast cancer.[1,2,14] However, limited information exists regarding both the performance of PRS over the life course in a prospective setting and their value when integrated with the established clinical risk factors and biomarkers.

We set out to test the utility of PRSs derived from large-scale genomic information for predicting first disease events in five diseases: CHD, T2D, atrial fibrillation or flutter (AF), and breast and prostate cancer. Specifically, we tested three hypotheses: 1) are the PRS associated with first disease events over a long follow-up and how much does PRS affect lifetime risk, 2) what is the impact of PRS on age at disease onset, and 3) what is the impact of PRS on clinical risk prediction? We tested these hypotheses within the FinnGen study cohort comprising of 135,300 individuals with genome-wide genotyping and up to 46 years of follow-up.

We first derived PRSs for the five diseases, CHD, T2D, AF, breast cancer, and prostate cancer by weighting the individual single nucleotide polymorphisms (SNP) by their effect sizes from published genome-wide association studies (GWAS) and by accounting for linkage disequilibrium (LD) between markers. We tested the association between these newly derived PRSs and disease events within the independent FinnGen study cohort (n = 135,300), which comprised 20,179 individuals with CHD, 17,519 with T2D, 12,809 with AF, 4,960 with breast cancer, and 3,617 with prostate cancer. FinnGen was comprised of 56.3% women, with mean age 59.2 (standard deviation, SD 16.6) at the end of follow-up.

For all five diseases, a higher PRS was strongly associated with a higher incidence rate (Figure 1; Supplementary Table S1). The hazard ratio (HR) per SD increment was 1.31 for CHD (95% CI 1.29-1.33, $p < 1.00 \times 10^{-300}$), for T2D 1.74 (1.72-1.77, $p < 1.00 \times 10^{-300}$), for AF 1.62 (1.59-1.65, $p < 1.00 \times 10^{-300}$), for breast cancer 1.64 (1.60-1.69, $p = 7.40 \times 10^{-268}$), and for prostate cancer 1.83 (1.78-1.90, $p = 9.32 \times 10^{-296}$). Compared to individuals with average PRS (20-80th percentile of the PRS distribution), being in the top 2.5% of the distribution translated into HRs ranging from 2.03 in CHD to 4.07 in prostate cancer (p-values 1.96×10^{-59} to 1.88×10^{-317}; Supplementary Table S1).

Similarly, when comparing the average PRS to the lowest 2.5%, the HRs ranged from 0.21 in prostate cancer to 0.61 in CHD (p-values 5.74×10^{-11} to 7.11×10^{-64}). Investigating goodness-of-fit indicated that the PRS are well calibrated (Extended Data Fig 1).

These effect sizes translated to the following increases in lifetime risk: from average PRS (20-80th percentile) to the top 2.5% of the PRS distribution, the risk increased for CHD from 37.2% (95% CI 36.9-37.5%), to 63.9% (62.3-65.5%) (all categories in Supplementary Table S1). The respective increases were for T2D from 28.3% (28.0-28.6%) to 66.7% (65.1-68.3%), for AF from 24.4% (24.1-24.7%) to 61.1% (59.5-62.7%), for breast cancer from 13.3% (13.0-13.6%) to 33.9% (31.8-36.0%), and for prostate cancer from 16.3% (15.9-16.7%) to 50.0% (47.5-52.5%).

In addition, we built estrogen receptor-negative (ER-negative) and estrogen receptor-positive (ER-positive) breast cancer PRSs (Extended Data Fig. 2). With any breast cancer as the outcome, for ER-negative PRS the HR for average PRS vs top 2.5% of the PRS distribution was 1.69 (95% CI 1.47-1.95, $p = 2.55 \times 10^{-13}$) and for ER-positive PRS, 2.72 (95% CI 2.42-3.06, $p = 2.52 \times 10^{-62}$).

The higher the PRS, the earlier was the disease onset for all five diseases (Figure 2**,** Extended Data Fig. 3, sex-specific results in Extended Data Fig. 4). Compared to individuals with average PRS, those in the top 2.5% of the distribution had a disease onset 4.35 (CHD), 8.81 (T2D), 6.64 (AF), 4.89 (breast cancer), and 5.53 (prostate cancer) years earlier. The largest difference in age at onset between the top and bottom 2.5%, 13.4 years, was seen for T2D. For CHD, the differences in age at disease onset were larger in men than in women (Extended Data Fig. 4).

In estimating clinical risk, we used the following clinical risk factors: 1) The ASCVD risk calculator[15] used for CHD includes age, sex, total cholesterol (TC), high-density lipoprotein (HDL),

systolic blood pressure (SBP), blood pressure-lowering medication, diabetes, and smoking status, 2) T2D analyses include age, sex, body mass index (BMI), history of stroke or CHD, parental history of any diabetes, SBP, diastolic blood pressure (DBP), HDL, and triglycerides, 3) the CHARGE-AF calculator[16] used for AF includes age, height, weight, SBP, DBP, smoking status, blood pressure-lowering medication, diabetes, heart failure, and history of myocardial infarction, 4) breast cancer analyses include age, family history of breast cancer, current smoking, BMI, alcohol use disorder, years of hormone replacement therapy, and having given birth one or more children, and 5) prostate cancer analyses include age, family history, and history of benign prostate hyperplasia.

We used only incident cases for comparing PRS to clinical risk calculators (comparison of effect sizes for PRS in prevalent versus incident cases in Supplementary Table S2). We first assessed the effect of adding PRS to cardiometabolic clinical risk scores for CHD, T2D, and AF using the FINRISK study (n = 21,813, mean age at baseline 48.0, 52.7% women) which has major cardiometabolic risk factors measured (Supplementary Table S3). For breast and prostate cancers, the effect of adding PRS to the clinical risk factors was assessed in FinnGen. The number of incident cases and controls was 1,209 and 18,956 for CHD, 1,346 and 19,684 for T2D, 229 and 10,332 for AF, 742 and 37,099 for breast cancer, and 1,172 and 47,679 controls for prostate cancer. Overall, the Pearson correlation between polygenic and clinical risk scores was low (r ranging from -0.01 in AF to 0.11 in T2D), and family history of CHD or T2D had only a minor effect on the association between polygenic risk and disease (Extended Data Fig. 5; Supplementary Table S4).

PRS improved model discrimination over age and sex in T2D, AF, breast cancer, and prostate cancer, and over clinical risk in T2D, breast cancer, and prostate cancer (Table 1). The improvement in C-index over clinical risk scores ranged from 1.0% in T2D to 3.9% in breast cancer. We then evaluated the reclassification of individuals across commonly used absolute risk

thresholds when adding the PRS to the clinical risk scores. These thresholds were 10-year risk ≥7.5% for CHD, 10-year risk ≥33% for T2D, 5-year risk ≥5% for AF, and a 10-year risk ≥5% for breast and prostate cancer. Adding PRS improved case reclassification with NRI ranging from 4.8% in T2D (95% CI 3.2-6.3%) to 12.9% in breast cancer (95% CI 9.9-15.9%)(Table 2, Supplementary Table S5). CHD PRS showed improvement in reclassification in early-onset cases (3.9%, 95% CI 1.6-6.2%) and in late-onset controls (1.5%, 95% CI 0.8-2.3). For T2D and breast cancer, the NRI was larger for late-onset than for early-onset disease. Lastly, we assessed how often PRS alone is elevated in early- and late-onset cases (Figure 2). In early-onset cases, this proportion was 12.6% for CHD, 17.9% for T2D, 27.9% for AF, 10.9% for breast cancer, and 29.9% for prostate cancer.

The differences in absolute risk across PRS categories in breast and prostate cancer could have an impact on screening practices. For example, breast cancer screening starts from age 50 in Finland, when its cumulative incidence reaches 2.0%. To bring the assessment of absolute risk differences across PRS categories to this screening context, using the FINRISK study we estimated the age when the cumulative incidence reached 2.0% in the different PRS groups: 44.5 years in PRS category >97.5%, 45.4 years in 80-97.5%, 50.0 years in 20-80%, and 58.5 years in PRS <20%. Similarly, a 2% cumulative incidence for prostate cancer in men was reached at age 62. When estimated across the PRS categories, a 2% cumulative incidence was reached at age 55.6 in PRS category >97.5%, at 59.4 in 80-97.5%, at 62.4 in 20-80%, and at 69.9 in <20%.

For the studied diseases, CHD, T2D, AF, breast cancer, and prostate cancer, we show that higher polygenic risk is associated with higher disease risk. This risk elevation also translated into large absolute risk differences over the lifespan as well as into large shifts towards earlier disease onset. We also show that in all five diseases, PRS has additional predictive value in clinical risk

prediction. Adding PRS to clinical risk prediction improved also reclassification over routinely used clinical thresholds.

For many diseases, particularly CHD and breast cancer, previous large studies have demonstrated strong associations between high polygenic risk and risk of disease.[1-3,14,17-19] Many of these studies have, however, used only tens or hundreds of genetic markers, looked at only prevalent cases in a cross-sectional setting, or have only a short follow-up. Importantly, a study assessing clinical risk factors with PRS has not been previously performed at this scale. We modeled the risk conferred by PRS over the life course, using a dataset with population-level disease prevalence. Moreover, we predicted new, future disease cases by studying prospectively only incident cases when comparing the PRS to clinical risk calculators, instead of using prevalent cases, which tend to produce higher effect sizes and are confounded by secondary prevention such lipid-lowering therapy prescribed in prevalent CHD. The impact of PRS was similar across all five diseases with respect to identifying subsets of the population at high risk for disease and at risk for earlier disease onset. The PRS had similar benefits also in clinical risk prediction, but the implications for clinical decision-making and the age in which utility was largest, vary between the diseases.

In CHD, adding PRS to clinical risk prediction showed two patterns with implications for clinical utility. First, for early-onset CHD, the CHD PRS identified individuals missed by clinical risk scores, comprising 13% of the early-onset cases. Most cardiovascular risk calculators have been trained with data on middle-aged individuals, and their ability to identify persons at risk for early-onset CHD is therefore limited.[5] Improved identification of these high-risk individuals could allow for targeted preventive efforts, for instance, targeting cholesterol-lowering treatments or lifestyle modification may be particularly useful in individuals with a high CHD PRS.[20-22] Second, CHD PRS improved reclassification of older non-case individuals towards lower risk. As age is an

important risk driver in most cardiovascular risk calculators and can therefore lead to false positives in older age groups, CHD PRS may potentially reduce overestimation of risk and subsequent overtreatment.

Both in early- and late-onset T2D cases, high T2D PRS was the only risk factor in approximately every sixth individual, but in the reclassification analyses, T2D PRS improved reclassification of clinical risk prediction only for late-onset disease. T2D PRS may have a role in identifying individuals for targeted screening, or in personalization of preventative options.[23] For AF, the PRS improved identification of high-risk individuals in both early- and late-onset disease, but also improved classification of older non-cases towards lower risk. In AF, identification of high-risk individuals is important for prevention of stroke, a relatively common and potentially severe consequence of AF. Potential clinical applications for AF PRS include targeted screening for timely diagnosis, or applying it as a biomarker for risk of stroke.[19,24]

For breast and prostate cancer, PRS improved identification of high-risk individuals in both early- and late-onset disease. In breast cancer, clinical risk prediction improvement was larger for late-onset disease, and in prostate cancer for early-onset disease. For prostate cancer, the older non-cases were reclassified towards lower risk, which might help prevent overdiagnosis. In many countries including Finland, breast cancer screening is initiated at age 50, by which approximately 2% of women have been diagnosed with breast cancer.[25,26] In our data, this 2% prevalence was reached at very different ages in the different PRS categories, with the difference between the tails of the PRS distribution amounting to 14 years. With similar results for prostate cancer, PRS could bring value to risk stratification to guide screening recommendations in both cancers, in line with previous studies, some of which have applied also clinical risk factors.[2,3,11,27,28] For breast cancer, the timing

and frequency of mammography screening could be stratified based on risk, and for prostate cancer, stratification based on PRS could assist in the decision-making for screening.

As the data comprised of individuals of European ancestry, PRSs need to be tested also in non-European samples. It is of utmost importance to conduct GWASs in non-Europeans to provide input for PRS in populations of non-European origin.[29] Although our analyses were performed in Finns, our results are in line with earlier reports from other samples of European origin.[1,2,18] A fraction of individuals in FinnGen were ascertained through hospital biobanks or disease-based cohorts, which may lead to some overestimation of risks. However, the effects of the PRSs in FinnGen were highly similar to those in the population-based FINRISK (Supplementary Tables S6-S7, Extended Data Fig. 6).

In conclusion, when predicting first disease events, polygenic risk scores identified individuals missed by established clinical risk prediction models, particularly those at high risk for early-onset disease. The practical applications of polygenic risk information for stratified screening or for guiding lifestyle and medical interventions in the clinical setting remain to be defined in further studies.

METHODS

Individuals

The data comprised of 135,300 Finnish individuals from FinnGen Preparatory Phase Data Freeze 3, which includes prospective epidemiological and disease-based cohorts, and hospital biobank samples (Supplementary Table S8). The data, representing roughly 3% of Finnish adult population, were linked by the unique national personal identification numbers to national hospital discharge (available from 1968), death (1969-), cancer (1953-), and medication reimbursement (1995-) registries.

A subset of FinnGen, the population-based FINRISK study with 21,813 individuals was selected for analyzing the PRSs together with clinical risk factors. The FINRISK surveys, performed in 1992, 1997, 2002, and 2007 comprised random samples of adults within five geographical areas in Finland. The baseline data covered self-reported information assessed by questionnaires, anthropometric measurements, and blood samples. Additional details on the study protocol have been previously described.[30] The Ethics Review Board of the Hospital District of Helsinki and Uusimaa approved the FinnGen study protocol (HUS/990/2017). The FINRISK analyses were conducted using the THL biobank permission for project BB2015_55.1. All participants gave written informed consent.

Disease endpoints

Using the national registries, we studied the incidence of five diseases: CHD, T2D, AF, breast cancer, and prostate cancer (diagnoses based on International Classification of Diseases, ICD-8, ICD-9, and ICD-10 in Supplementary Table S9). Follow-up ended at first-ever diagnosis of the disease of interest, death, or at the end of follow-up on December 31, 2018, whichever came first.

Clinical risk factors

The 10-year risk of hard atherosclerotic cardiovascular disease (ASCVD) was evaluated with the pooled cohort equations (PCE) according to guidelines,[15,31] comprising age, sex, self-reported ancestry, total cholesterol (TC), high-density lipoprotein (HDL), systolic blood pressure (SBP), blood pressure-lowering medication, prevalent diabetes, and smoking status. The 10-year risk for ASCVD was categorized as intermediate to high risk (≥7.5% which often leads to consideration of preventive medication) or low to borderline risk (<7.5%).[31] Analyses on family history for CHD were based on self-reported parental history of early myocardial infarction (MI). 23 individuals with missing data for the risk variables were excluded. Participants with prevalent diabetes (n = 671) were excluded from all ASCVD assessments.

For T2D, we constructed a 10-year risk score that included available risk factors listed in the American Diabetes Association (ADA) criteria for testing for diabetes or prediabetes in asymptomatic adults.[32] High risk was defined as a 10-year risk exceeding 33%.[33] The risk factors included age, sex, body mass index (BMI, kg/m^2), history of stroke or CHD, parental history of any diabetes, systolic and diastolic blood pressure (SBP, DBP), HDL, and triglycerides. History of cardiovascular disease was defined as physician-diagnosed CHD or stroke (see Supplementary Table 2 for definition of coronary heart disease; stroke was any of I61, I63, I64 except I63.6 (ICD-10) or 431, 4330A, 4331A, 4339A, 4340A, 4341A, 4349A, 436 (ICD-9) as the underlying or direct cause of death, or as the main or side diagnosis at hospital discharge. 82 individuals with missing data on BMI were excluded from these analyses involving clinical risk assessment of T2D.

For AF, the clinical score for the 5-year absolute risk for individuals above age 45 was carried out with the CHARGE-AF score, comprising age, height, weight, SBP, DBP, smoking status, blood

pressure-lowering treatment, prevalent diabetes, heart failure, and history of MI.[16] CHARGE-AF was revised and the risk was categorized as ≤5% or >5%.[16] When taking all calculator components from the original study, the original CHARGE-AF score showed poor calibration with a mean 5-year risk 0.02% in individuals aged ≥45. To improve calibration, we obtained the mean component from FINRISK individuals aged 45 to 74, which resulted in a 5-year mean risk of 4.3% (standard deviation 4.6%). We did not revise the baseline hazard, as the original baseline hazard ≈0.972 was similar to ours (≈0.977). 85 individuals with missing data for the risk variables were excluded.

For CHD, T2D, and AF, the clinical risk factors were available in FINRISK. For breast and prostate cancers, the clinical risk factor comparisons were done in FinnGen. For breast cancer, we modeled the 10-year risk, with a high-risk definition of ≥5%.[34] Due to the lack of absolute risk calculators for prostate cancer in the general population, we applied these thresholds also for prostate cancer.

In the cancer analyses, follow-up was restricted to start from Jan 1st, 2000, leading to 742 incident breast cancer cases with 37,099 controls (mean 10-year risk of 2.0%, SD 2.1%), and 1,172 incident prostate cancer cases with 47,679 controls (mean 10-year risk 2.6% with SD 4.3%). For breast cancer, we modeled the 10-year risk using available risk factors: age, family history of breast cancer, current smoking, BMI, alcohol use disorder, years of hormone replacement therapy (estrogen-only or estrogen-progestagen preparations), and having given birth to one or more children. Prostate cancer analyses comprised age, family history, and history of benign prostate hyperplasia. The detailed definitions are provided next.

The cancer risk factors were modeled from registry data using available risk factors, limiting analyses to individuals born before or in 1975. Breast cancer analyses include age (on Jan 1st 2000), family history of breast cancer (ICD-10 code Z80.3), current smoking, body mass index, years of

hormone replacement therapy before start of follow-up (individual years with purchases with ATC codes G03CA, G03FA, G03FB, or G03CA), having given birth one or more children before start of follow-up (ICD-10 codes O80, O81, O82, O84, O85 and ICD-9 codes 650, 651, 6695, 6696, 6697), and alcohol use disorder (following diagnoses defined in Kiiskinen et al,[35] ICD-10 codes F10, G31.2, G41.51, G62.1, I42.6, O35.4, K29.3, K70, K85.2, K86.0, P04.3 X45, Z71.4, E24.4, T51.1-9; ICD-9 codes 291, 303, 305A, 3575A, 4255, 5353A, 5710-3, 5770D-F, 5771C-D, 7607A, 9801-9; ICD-8 codes 291, 303, 5710, 9801-9; medication purchases with ATC codes N07BB01, N07BB02, N07BB04). Prostate cancer analyses include age (on Jan 1st 2000), family history (Z80), and history of benign prostate hyperplasia before start of follow-up.

For comparing early- and late-onset cases, the definition for early-onset cases was age below 55 for CHD, 45 for T2D, 60 for AF, 45 for breast cancer, and 55 for prostate cancer.

Genotyping and imputation in FinnGen

FinnGen samples were genotyped with Illumina and Affymetrix arrays (Illumina Inc., San Diego, and Thermo Fisher Scientific, Santa Clara, CA, USA) and put through the same rigorous QC steps as described below. Genotype imputation was carried out by using the population-specific SISu v3 imputation reference panel with Beagle 4.1 (version 08Jun17.d8b, https://faculty.washington.edu/browning/beagle/b4_1.html) as described in the following protocol: dx.doi.org/10.17504/protocols.io.nmndc5e. Post-imputation QC involved excluding variants with imputation INFO <0.7.

Genotyping and imputation in FINRISK

26,404 FINRISK samples were genotyped using several arrays: the HumanCoreExome BeadChip, the Human610-Quad BeadChip, the Affymetrix6.0, and the Infinium HumanOmniExpress

(Illumina Inc., San Diego and Affymetrix, Inc., Santa Clara, CA, USA). Genotype calls were generated together with other available data sets using zCall at the Institute for Molecular Medicine Finland (FIMM). After sample-wise quality control (exclude samples with ambiguous gender, missingness (>5%), excess heterozygosity (+-4SD), non-European ancestry) and variant-wise quality control (exclude SNPs with high missingness (>2%), low HWE P-value (<1e-6), minor allele count (MAC) <3 (in case Zcall'ed chip data) or MAC <10 (chip data called using Illumina GenCall) steps, the samples were pre-phased using Eagle2 (version 2.3). Genotype imputation was carried out by using a Finnish population-specific reference panel consisting of 2,690 high-coverage WGS and 5,092 WES samples with IMPUTE2 (version 2.3.2) that allows the usage of two panels at the same time (the 'merge_ref_panels' option). Post-imputation quality control involved excluding variants imputed with imputation INFO < 0.7. Chromosome X variants were also excluded from the downstream analyses. We excluded one individual of each sample-pair with kinship >0.125, and calculated principal components for the unrelated individuals. The 26,404 samples contained the 2012 FINRISK cohort; this study used only FINRISK cohorts from 1992, 1997, 2002, and 2007, comprising 21,813 individuals.

Polygenic risk scores

In all five diseases, we used external GWAS for building the PRS. The summary association statistics came from recent GWAS (Supplementary Table S10, Extended Data Fig. 7).[17,36-39] LDpred was used to account for linkage disequilibrium among loci,[40] with whole-genome sequencing data on 2,690 Finns serving as the LD reference panel (using only autosomes). After performing quality control, the final scores were generated with PLINK2[24] by calculating the weighted sum of risk allele dosages for each SNP. The final PRSs comprised 6,412,950 variants for CHD PRS, 6,437,380 for T2D PRS, 6,171,733 for AF PRS, 6,390,808 for breast cancer PRS, and 6,606,785 for prostate

cancer PRS (candidate LDpred scores concerning the tuning parameter in Supplementary Table S11).

Due to the high LD in the isolated Finnish population, we selected an LD-radius approximately twice the radius recommended, which is M/3,000, where M is the total number of single nucleotide polymorphisms used in the analysis. Variants with minor allele frequency less than 1% are excluded by the software.

We calculated the polygenic risk scores by summing the dosage of each risk allele carried by an individual (ranging from 0 to 2 for each variant, dosage used for incorporating imputation uncertainty), weighting each variant by its natural logarithm of the relative risk extracted from the genome-wide association study. For each individual i, this results in a single value on a continuous scale:

$$\mathrm{PRS}_i = \sum_{j=1}^{M} \hat{\beta}_j \times dosage_{ij}$$

where $\hat{\beta}_j$ is the weight for variant j obtained from GWAS summary statistics.

Statistical analysis

Cox proportional hazards model was used to estimate survival curves and hazard ratios (HRs) and 95% confidence intervals (CI). Schoenfeld residuals and log-log inspection showed that proportional assumption criteria applied in our models. Unless otherwise stated, we adjusted for FINRISK survey collection year (in FINRISK), genotyping array/batch, the first ten principal components of ancestry, and stratified the models by sex. Breast cancer was studied only in women and prostate cancer only in men. Lifetime risk by age 80 was estimated from the adjusted survival curves with confidence intervals for lifetime risks obtained by normal approximation. Performance

metrics were calculated with risk measures and PRS on the continuous scale. Model discrimination was assessed with the concordance index (C-index; confidence intervals obtained by normal approximation), which is an extension of the area under the receiver operating characteristic curve to survival analysis. Goodness-of-fit for the Cox proportional hazards model was assessed with R package *survMisc*, following methodology proposed by May & Hosmer.[41]

We then evaluated improvements in clinical risk stratification when adding PRS to the clinical risk score. We report the number of individuals reclassified across following clinical thresholds: 10-year risk ≥7.5% for CHD, 10-year risk ≥33% for T2D, 5-year risk ≥5% for AF, and a 10-year risk ≥5% for breast and prostate cancer. Reclassification was also assessed with net reclassification improvement (NRI). In C-index and NRI comparisons, the PRS was added to the linear predictor alongside the original regression coefficients for the risk factors in the clinical risk scores (ASCVD for CHD and CHARGE-AF for AF), or as an independent risk factor in the regression model (breast and prostate cancer; T2D using the ADA high-risk definition variables on their original scale). For the linear predictor, the effect sizes for CHD PRS ($\beta = 0.222$ in women and $\beta = 0.271$ in men, per SD increase) and AF PRS ($\beta = 0.539$ per SD increase) were obtained from models fitted in FINRISK when adjusting for the risk factors included the original risk calculators. Reclassification was assessed also separately for early- and late-onset disease by dividing individuals into groups according to age at baseline (55 for CHD, 45 for T2D, 60 for AF, 45 for breast cancer, and 55 for prostate cancer).

We chose 2.5% as our top tail and divided the PRS into bins of <2.5%, 2.5-20%, 20-80%, 80-97.5%, and >97.5%. The 20-80% bin was used as the reference, to display the results with respect to a large group of individuals with average risk. The definition of the top and bottom 2.5% also

follows the principle, where a reference range of a laboratory test is often defined by selecting values within which 95 percent of the population fall.

Comparing high clinical and high polygenic risk separately for early and late-onset cases, we calculated the proportion of individuals exceeding the absolute risk thresholds applied in the clinical risk calculators, which was 10.4% (CHD, with 10-year risk ≥7.5%), 38.4% (T2D, with 10-year risk ≥33%), 10.7% (AF, with 5-year risk ≥5%), 4.8% (breast cancer with 10-year risk ≥5%), and 13.9% (prostate cancer with 10-year risk ≥5%). Based on these, we defined elevated PRS as a polygenic risk score above the 90th percentile.

In FINRISK analyses, the association between PRS and the disease was tested for incident cases only. The number of prevalent cases excluded in FINRISK was 954 for coronary heart disease, 671 for T2D, 351 for atrial fibrillation (AF), 164 for breast cancer, and 59 for prostate cancer. The FINRISK had 1,805 individuals overlapping with the AF GWAS, and we excluded these individuals from the AF analyses. Age at disease onset and the differences between PRS categories were estimated with restricted mean survival time (RMST).[42] RMST (age 85 as the upper limit) were estimated by fitting flexible parametric survival models, which generated very similar effect sizes as the Cox proportional hazards models. Adjusted survival curves were plotted with the R package *survminer*, using the calculation parameter "conditional", which after rebalancing averages for the polygenic risk score categories. For statistical analyses, we used R 3.5.2, and Stata 14.2 (College Station, TX, USA).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Acknowledgements We would like to thank Sari Kivikko, Huei-Yi Shen, and Ulla Tuomainen for management assistance. The FINRISK analyses were conducted using the THL biobank permission for project BB2015_55.1. The FINRISK data used for the research were obtained from THL Biobank. For the Finnish Institute of Health and Welfare (THL) driven FinnGen preparatory project (here called FinnGen), all patients and control subjects had provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, older cohorts were based on study-specific consents and later transferred to the THL Biobank after approval by Valvira, the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Valvira. The Ethics Review Board of the Hospital District of Helsinki and Uusimaa approved the FinnGen study protocol Nr HUS/990/2017. The FinnGen preparatory project is approved by THL, approval numbers THL/2031/6.02.00/2017, amendments THL/341/6.02.00/2018, THL/2222/6.02.00/2018 and THL/283/6.02.00/2019. Following biobanks are acknowledged for collecting the FinnGen project samples: Auria Biobank (https://www.auria.fi/biopankki/en), THL Biobank (https://thl.fi/fi/web/thl-biopankki), Helsinki Biobank (https://www.terveyskyla.fi/helsinginbiopankki/en), Northern Finland Biobank Borealis (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki), Finnish Clinical Biobank Tampere (https://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (https://ita-suomenbiopankki.fi/), Central Finland Biobank (https://www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (https://www.bloodservice.fi/Research%20Projects/biobanking). We thank all study participants for their generous participation in FINRISK and FinnGen. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was supported by the Finnish Foundation for Cardiovascular Research [to S.R., V.S., and A.P.]; Sigrid Jusélius Foundation [to S.R. and A.P.]; University of Helsinki HiLIFE Fellow grants 2017-2020 [to S.R.]; Academy of Finland Center of Excellence in Complex Disease Genetics [grant number 312062 to S.R., 312074 to A.P., 312075 to M.D]; Academy of Finland [grant number 285380 to S.R, 128650 to A.P.]; The Finnish Innovation Fund Tekes [grant number 2273/31/2017 to E.W.]; Foundation and the Horizon 2020 Research and Innovation Programme [grant number 667301 (COSYN) to A.P]; Ida Montin Foundation [to P.R.]; Doctoral Programme in Population Health, University of Helsinki [to P.R.]; and Emil Aaltonen Foundation

[to P.R.]. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and nine industry partners (AbbVie, AstraZeneca, Biogen, Celgene, Genentech, GSK, MSD, Pfizer and Sanofi). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

S.R and N.M. conceived and designed the study. N.M., and P.R. carried out the statistical and computational analyses with advice from S.R., J.T.K., E.W., J.V.L., A.A.-O., M.D., V.S., B.M.N., and A.P. Quality control of the data was carried out by N.M., A.S.H., T.T.J.K., M.K., J.K., P.P. The manuscript was written and revised by all the co-authors. All co-authors have approved of the final version of the manuscript.

Conflicts of interests A.P. is a member of the Pfizer Genetics Scientific Advisory Panel. V.S. has participated in a conference trip sponsored by Novo Nordisk and received an honorarium for participating in an advisory board meeting (unrelated to the present study). V.S. also has research collaboration with Bayer Ltd (unrelated to the present study). B.M.N. is a member of the scientific advisory board at Deep Genomics and consultant for Camp4 Therapeutics, Takeda Pharmaceutical and Biogen.

Data availability The FinnGen data may be accessed through Finnish Biobanks' FinnBB portal (www.finbb.fi) and THL Biobank data through THL Biobank (https://thl.fi/en/web/thl-biobank).

Code availability The full genotyping and imputation protocol for FinnGen is described at dx.doi.org/10.17504/protocols.io.nmndc5e

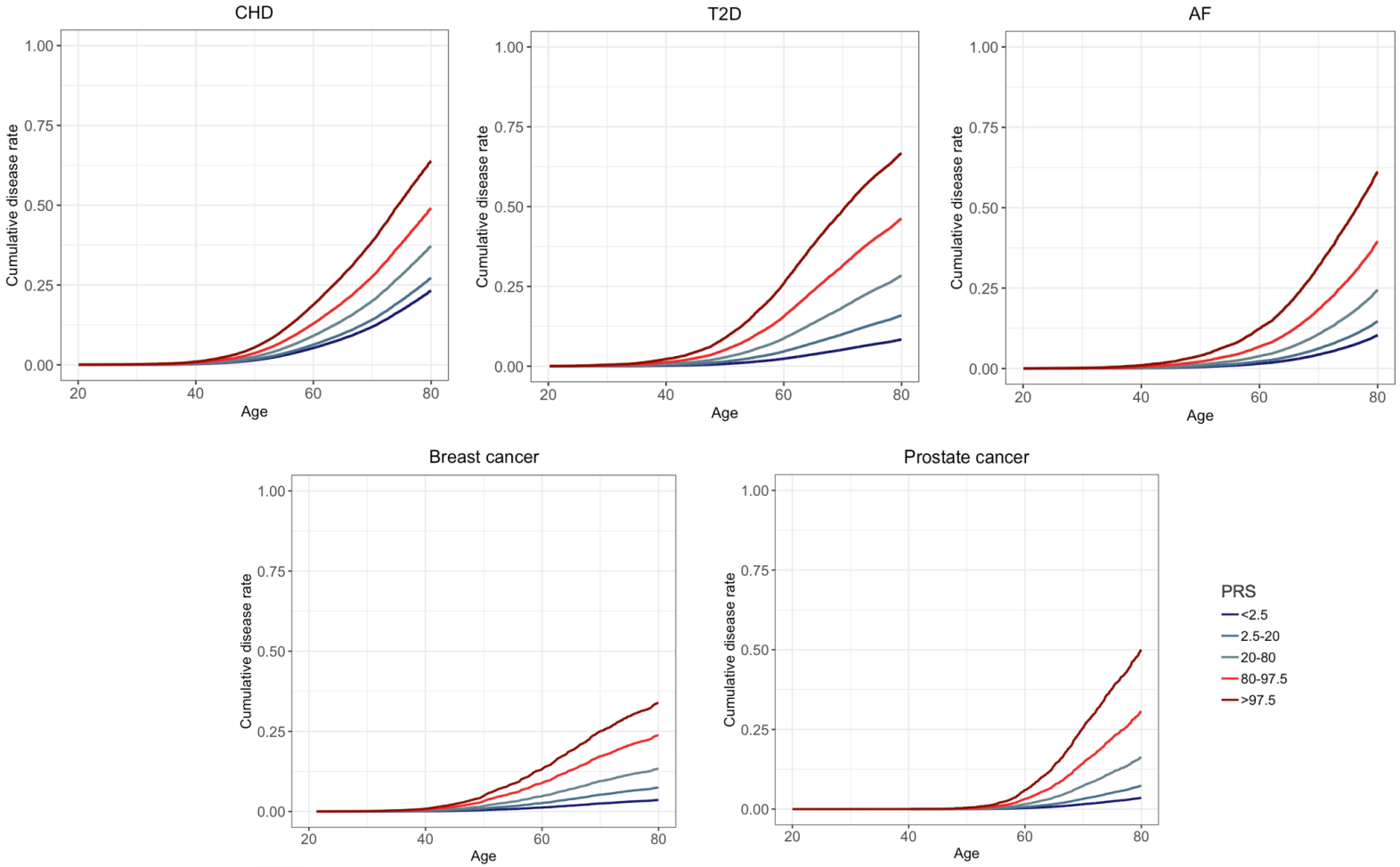
REFERENCES

1. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018;50:1219-24.
2. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet 2019;104:21-34.
3. Seibert TM, Fan CC, Wang Y, et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. BMJ 2018;360:j5757.
4. Global Burden of Disease. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1789-858.
5. Piepoli MF, Hoes AW, Agewall S, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts)Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J 2016;37:2315-81.
6. Lindbohm JV, Sipila PN, Mars NJ, et al. 5-year versus risk-category-specific screening intervals for cardiovascular disease prevention: a cohort study. Lancet Public Health 2019;4:e189-e99.
7. Brentnall AR, Cuzick J, Buist DSM, Bowles EJA. Long-term Accuracy of Breast Cancer Risk Assessment Combining Classic Risk Factors and Breast Density. JAMA Oncol 2018;4:e180174.
8. Wilson BJ, Qureshi N, Santaguida P, et al. Systematic review: family history in risk assessment for common diseases. Ann Intern Med 2009;151:878-85.
9. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 2017;101:5-22.
10. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet 2018.
11. Choudhury PP, Wilcox AN, Brook MN, et al. Comparative validation of breast cancer risk prediction models and projections for future risk stratification. J Natl Cancer Inst 2019.
12. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. Nat Rev Genet 2018;19:110-24.
13. Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet 2016;17:392-406.
14. Abraham G, Havulinna AS, Bhalala OG, et al. Genomic prediction of coronary heart disease. Eur Heart J 2016;37:3267-78.
15. Goff DC, Jr., Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014;129:S49-73.
16. Alonso A, Krijthe BP, Aspelund T, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. J Am Heart Assoc 2013;2:e000102.
17. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 2018;50:1505-13.

18. Inouye M, Abraham G, Nelson CP, et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol 2018;72:1883-93.
19. Lubitz SA, Yin X, Lin HJ, et al. Genetic Risk Prediction of Atrial Fibrillation. Circulation 2017;135:1311-20.
20. Khera AV, Emdin CA, Kathiresan S. Genetic Risk, Lifestyle, and Coronary Artery Disease. N Engl J Med 2017;376:1194-5.
21. Natarajan P, Young R, Stitziel NO, et al. Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. Circulation 2017;135:2091-101.
22. Mega JL, Stitziel NO, Smith JG, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. Lancet 2015;385:2264-71.
23. McCarthy MI. Painting a new picture of personalised medicine for diabetes. Diabetologia 2017;60:793-9.
24. Bapat A, Anderson CD, Ellinor PT, Lubitz SA. Genomic basis of atrial fibrillation. Heart 2018;104:201-6.
25. Average Number of New Cases Per Year and Age-Specific Incidence Rates per 100,000 Females, UK. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer Accessed 4 April, 2019.
26. NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 8.2 (26.03.2019). http://www-dep.iarc.fr/NORDCAN/ Accessed 4 April, 2019.
27. Maas P, Barrdahl M, Joshi AD, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. JAMA Oncology 2016;2:1295-302.
28. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med 2019;21:1708-18.
29. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 2019;51:584-91.
30. Borodulin K, Tolonen H, Jousilahti P, et al. Cohort Profile: The National FINRISK Study. Int J Epidemiol 2017.
31. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2018.
32. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2019. Diabetes Care 2019;42:S13-S28.
33. Lindstrom J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care 2003;26:725-31.
34. Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of breast cancer risk: a validation study. Lancet Oncol 2019;20:504-17.
35. Kiiskinen T, Mars NJ, Palviainen T, et al. Genomic prediction of alcohol-related morbidity and mortality. Translational Psychiatry 2020;10:23.
36. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 2018;50:1335-41.
37. Schumacher FR, Al Olama AA, Berndt SI, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet 2018;50:928-36.
38. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551:92-4.

39. Nielsen JB, Thorolfsdottir RB, Fritsche LG, et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. Nat Genet 2018;50:1234-9.
40. Vilhjalmsson BJ, Yang J, Finucane HK, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. Am J Hum Genet 2015;97:576-92.
41. May S, Hosmer DW. A simplified method of calculating an overall goodness-of-fit test for the Cox proportional hazards model. Lifetime Data Anal 1998;4:109-20.
42. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol 2013;13:152.

Figure 1. Adjusted survival curves from Cox proportional hazards models, showing the cumulative risk of disease by polygenic risk score (PRS) categories in FinnGen (n = 135,300 individuals).



CHD = coronary heart disease, AF = atrial fibrillation or flutter, T2D = type 2 diabetes. P-values for trend: CHD $p = 2.80 \times 10^{-256}$; T2D $p < 1.00 \times 10^{-300}$; AF $p < 1.00 \times 10^{-300}$; breast cancer $p = 3.07 \times 10^{-183}$; prostate cancer $p = 2.41 \times 10^{-243}$. Incident and prevalent cases included. All tests were two-tailed.

Table 1. C-index for model discrimination assessed for combinations of age, sex, disease-specific polygenic risk score (PRS) and clinical risk scores.

	N total	N cases	Age + sex*	Age + sex* + PRS	Clinical	Clinical + PRS
CHD	20,165	1,209	0.830 (0.825-0.834)	0.832 (0.828-0.836)	0.823 (0.819-0.827)	0.820 (0.816-0.824)
T2D	21,030	1,346	0.728 (0.723-0.733)	0.763 (0.758-0.767)	0.835 (0.831-0.839)	0.845 (0.841-0.849)
AF	10,561	229	0.709 (0.702-0.716)	0.751 (0.744-0.757)	0.725 (0.719-0.732)	0.734 (0.728-0.741)
Breast cancer	37,841	742	0.693 (0.689-0.696)	0.737 (0.733-0.741)	0.711 (0.707-0.714)	0.750 (0.746-0.753)
Prostate cancer	48,851	1,172	0.827 (0.824-0.829)	0.857 (0.855-0.859)	0.840 (0.837-0.842)	0.866 (0.863-0.868)

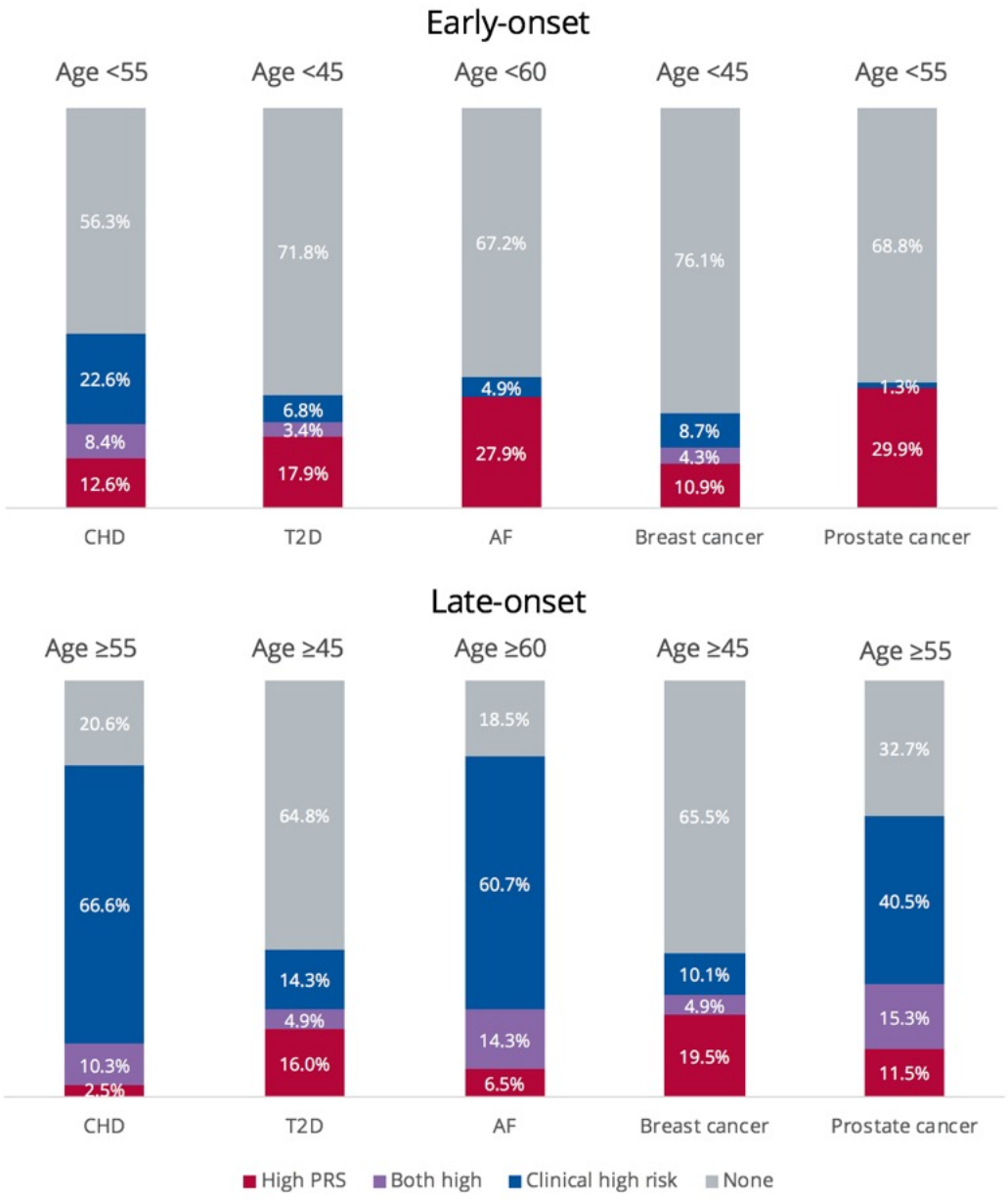
*Sex not included for breast and prostate cancer. CHD = coronary heart disease, AF = atrial fibrillation or flutter, T2D = type 2 diabetes. Only incident cases included. Time horizons: 10-year risk for CHD, T2D, breast and prostate cancer; 5-year risk for AF. Clinical risk factors: The ASCVD risk calculator used for CHD includes age, sex, total cholesterol, high-density lipoprotein, systolic blood pressure, blood pressure-lowering medication, diabetes, and smoking status; T2D analyses include age, sex, body mass index, history of stroke or CHD, parental history of diabetes, systolic and diastolic blood pressure, high-density lipoprotein, and triglycerides; the CHARGE-AF calculator used for AF includes age, height, weight, systolic and diastolic blood pressure, smoking status, blood pressure-lowering medication, diabetes, heart failure, and history of myocardial infarction; breast cancer analyses include age, family history of breast cancer, current smoking, body mass index, alcohol use disorder, years of hormone replacement therapy, and having given birth one or more children; prostate cancer analyses include age, family history, and history of benign prostate hyperplasia. PRS was added to the linear predictor of existing continuous calculators (ASCVD[15] for CHD, CHARGE-AF[16] for AF), or in the case of individual risk factors as an independent covariate in the regression model (T2D, breast cancer, prostate cancer).

Table 2. Net reclassification improvement (NRI) with addition of polygenic risk score (PRS) to clinical risk scores.

		All individuals				Early-onset				Late-onset			
		Individuals reclassified		NRI		Individuals reclassified		NRI		Individuals reclassified		NRI	
		Up (%)	Down (%)	Value (%)	95% CI	Up (%)	Down (%)	Value (%)	95% CI	Up (%)	Down (%)	Value	95% CI
CHD N = 20,165	**Cases**	2.5	1.6	0.9	-0.2, 2.0	4.4	0.6	3.9	1.6, 6.2	1.7	2.0	-0.4	-1.6, 0.9
	Non-cases	2.1	2.3	0.2	-0.1, 0.5	1.7	1.3	-0.4	-0.7, -0.1	3.1	4.6	1.5	0.8, 2.3
	All	-	-	1.1	-0.1, 2.2	-	-	3.5	1.2, 5.8	-	-	1.2	-0.3, 2.7
T2D N = 21,030	**Cases**	6.6	1.9	4.8	3.2, 6.3	1.4	1.0	0.5	-1.6, 2.6	7.6	2.0	5.5	3.7, 7.3
	Non-cases	0.8	0.5	-0.1	-0.2, -0.1	0.1	<0.1	-0.3	-0.2, 0.0	1.3	1.0	-0.3	-0.6, -0.1
	All	-	-	4.5	3.0, 6.1	-	-	0.4	-1.7, 2.5	-	-	5.2	3.4, 7.0
AF N = 10,561	**Cases**	16.2	3.5	12.7	6.9, 18.4	24.7	4.3	20.4	9.5, 31.4	10.3	2.9	7.4	1.2, 13.5
	Non-cases	7.7	5.0	-2.7	-3.4, -2.0	7.3	1.6	-5.7	-6.4, -5.0	8.5	11.5	3.0	1.6, 4.5
	All	-	-	10.0	4.2, 15.7	-	-	14.8	3.8, 25.7	-	-	10.4	4.1, 16.7
Breast cancer N = 37,841	**Cases**	15.2	2.3	12.9	9.9, 15.9	3.2	0.6	2.6	-0.5, 5.6	18.4	2.7	15.7	12.0, 19.4
	Non-cases	4.2	2.0	-2.2	-2.5, -2.0	0.4	<0.1	-0.3	-0.4, -0.3	7.8	3.9	-3.9	-4.4, -3.4
	All	-	-	10.7	7.7, 13.7	-	-	2.2	-0.8, 5.3	-	-	11.8	8.0, 15.5
Prostate cancer N = 48,851	**Cases**	14.3	6.4	7.9	5.3, 10.5	18.3	2.7	15.6	10.8, 20.5	12.7	7.9	4.8	1.7, 7.9
	Non-cases	3.3	4.2	0.8	0.6, 1.1	1.8	0.8	-0.9	-1.1, -0.8	8.1	14.1	6.0	5.2, 6.9
	All	-	-	8.8	6.1, 11.4	-	-	14.7	9.8, 19.6	-	-	10.8	7.6, 14.0

CHD = coronary heart disease, AF = atrial fibrillation or flutter, T2D = type 2 diabetes. Only incident cases included. Reclassification proportion of cases calculated with all cases in the denominator, with a similar approach for controls. Reclassification thresholds: 10-year risk ≥7.5% for CHD, 10-year risk ≥33% for T2D, 5-year risk ≥5% for AF, and a 10-year risk ≥5% for breast and prostate cancer. Clinical risk scores based on following risk factors: The ASCVD risk calculator used for CHD includes age, sex, total cholesterol, high-density lipoprotein, systolic blood pressure, blood pressure-lowering medication, diabetes, and smoking status; T2D analyses include age, sex, body mass index, history of stroke or CHD, parental history of diabetes, systolic and diastolic blood pressure, high-density lipoprotein, and triglycerides; the CHARGE-AF calculator used for AF includes age, height, weight, systolic and diastolic blood pressure, smoking status, blood pressure-lowering medication, diabetes, heart failure, and history of myocardial infarction; breast cancer analyses include age, family history of breast cancer, current smoking, body mass index, alcohol use disorder, years of hormone replacement therapy, and having given birth one or more children; prostate cancer analyses include age, family history, and history of benign prostate hyperplasia. Early- and late-onset assessment performed by dividing individuals into groups according to age at baseline (55 for CHD, 45 for T2D, 60 for AF, 45 for breast cancer, and 55 for prostate cancer).

Figure 2. The proportion early- and late-onset cases with high clinical risk, high polygenic risk, or neither.



CHD = coronary heart disease, AF = atrial fibrillation or flutter, T2D = type 2 diabetes. High PRS = in top decile of the distribution. The number of early- and late-onset cases for CHD was 190 and 1,019, for T2D 117 and 1,229, for AF 61 and 168, for breast cancer 46 and 696, and for prostate cancer 77 and 1,095. CHD, T2D, and AF cases from FINRISK and breast and prostate cancer from FinnGen, all incident cases. Clinical high risk definitions were following: For CHD, the 10-year risk calculator for hard atherosclerotic cardiovascular disease (ASCVD) ≥7.5%, according to the Pooled Cohort Equations by ACC/AHA (2013); for T2D, a 10-year risk ≥33% when constructing a calculator with risk factors listed in the American Diabetes Association (ADA) criteria for testing for diabetes or prediabetes in asymptomatic adults; for AF, the 5-year risk of AF >5%, by a revised version of the CHARGE-AF; for breast and prostate cancer a 10-year risk ≥5% with clinical risk factors.

Table S1. Hazard ratios (HR) and 95% confidence intervals (CI) for polygenic risk score (PRS) bins in FinnGen.

	HR (95% CI)	Lifetime risk, % (95% CI)	p	Cases / Controls
CHD PRS				
<2.5	0.61 (0.55-0.69)	23.2 (21.8-24.6)	3.04×10^{-18}	329 / 3,054
2.5-20	0.72 (0.69-0.75)	27.2 (26.6-27.8)	2.41×10^{-53}	2,580 / 21,097
20-80	1 (reference)	37.2 (36.9-37.5)	-	11,832 / 69,348
80-97.5	1.38 (1.34-1.43)	49.1 (48.5-49.7)	1.30×10^{-76}	4,532 / 19,145
>97.5	2.03 (1.90-2.18)	63.9 (62.3-65.5)	2.87×10^{-93}	906 / 2,477
T2D PRS				
<2.5	0.22 (0.19-0.27)	8.3 (7.4-9.3)	7.11×10^{-64}	131 / 3,195
2.5-20	0.48 (0.46-0.51)	15.9 (15.4-16.4)	1.90×10^{-157}	1,653 / 21,548
20-80	1 (reference)	28.3 (28.0-28.6)	-	9,869 / 68,998
80-97.5	1.92 (1.85-1.99)	46.2 (45.6-46.8)	7.04×10^{-287}	4,776 / 17,970
>97.5	3.45 (3.24-3.67)	66.7 (65.1-68.3)	1.88×10^{-317}	1,090 / 2,072
AF PRS				
<2.5	0.39 (0.33-0.46)	10.3 (9.3-11.3)	1.34×10^{-27}	136 / 3,247
2.5-20	0.56 (0.53-0.59)	15.9 (15.4-16.4)	3.92×10^{-81}	1,281 / 22,396
20-80	1 (reference)	24.4 (24.1-24.7)	-	7,159 / 74,021
80-97.5	1.81 (1.73-1.88)	39.5 (38.9-40.1)	3.57×10^{-177}	3,417 / 20,260
>97.5	3.50 (3.26-3.77)	61.1 (59.5-62.7)	2.11×10^{-250}	816 / 2,567
Breast cancer PRS				
<2.5	0.28 (0.20-0.40)	3.6 (2.8-4.4)	5.24×10^{-13}	33 / 1,873
2.5-20	0.54 (0.49-0.60)	7.5 (7.1-7.9)	4.61×10^{-32}	421 / 12,918
20-80	1 (reference)	13.3 (13.0-13.6)	-	2,703 / 43,030
80-97.5	1.88 (1.77-2.01)	23.8 (23.1-24.5)	5.32×10^{-85}	1,483 / 11,856
>97.5	2.87 (2.56-3.23)	33.9 (31.8-36.0)	1.10×10^{-70}	320 / 1,586
Prostate cancer PRS				
<2.5	0.21 (0.13-0.34)	3.5 (2.6-4.4)	5.74×10^{-11}	18 / 1,459
2.5-20	0.43 (0.37-0.49)	7.4 (6.9-7.9)	7.47×10^{-36}	242 / 10,095
20-80	1 (reference)	16.3 (15.9-16.7)	-	1,934 / 33,507
80-97.5	2.08 (1.93-2.24)	30.7 (29.8-31.6)	4.86×10^{-84}	1,118 / 9,219
>97.5	4.07 (3.61-4.60)	50.0 (47.5-52.5)	8.09×10^{-114}	305 / 1,172

CHD = coronary heart disease, AF = atrial fibrillation or flutter, T2D = type 2 diabetes. Incident and prevalent cases included. The estimates were obtained from Cox proportional hazards models described in detail in the Methods. All tests were two-tailed.

Table S2. Odds ratios and 95% confidence intervals per standard deviation increase in the polygenic risk scores (PRS) in FINRISK (n = 21,813 with detailed information in Supplementary Table S3), demonstrating how in most diseases using prevalent cases yields higher effect sizes compared to analyses with only incident cases. We therefore used only incident cases for analyses comparing PRS to clinical risk assessment.

	Prevalent only	**Incident only**	**Incident and prevalent**
CHD	1.58 (1.46-1.70)	1.31 (1.25-1.38)	1.38 (1.32-1.44)
T2D	2.04 (1.85-2.25)	1.80 (1.71-1.89)	1.83 (1.75-1.92)
AF	1.88 (1.68-2.10)	1.71 (1.61-1.81)	1.73 (1.64-1.82)
Breast cancer	1.73 (1.48-2.03)	1.77 (1.60-1.96)	1.77 (1.62-1.93)
Prostate cancer	2.29 (1.75-3.00)	2.00 (1.81-2.22)	2.04 (1.85-2.25)

CHD = coronary heart disease, AF = atrial fibrillation or flutter, T2D = type 2 diabetes. All tests were two-tailed.

Table S3. Baseline characteristics for FINRISK.

	FINRISK 1992 N = 4,745	FINRISK 1997 N = 6,733	FINRISK 2002 N = 5,427	FINRISK 2007 N = 4,908
Follow-up in years, mean (SD)	22.3 (4.3)	17.5 (3.7)	13.3 (2.0)	8.7 (1.0)
Age, mean (SD)	44.3 (11.4)	48.2 (13.4)	48.3 (13.1)	51.1 (13.9)
Age ≤50, %	65.9	54.9	52.4	45.2
Women, %	53.8	51.0	53.4	53.3
Current smokers, %	28.0	23.6	26.3	19.9
TC, mean (SD)	5.6 (1.1)	5.5 (1.1)	5.6 (1.1)	5.3 (1.0)
LDL, mean (SD)	3.5 (1.0)	3.5 (0.9)	3.4 (1.0)	3.2 (0.9)
HDL, mean (SD)	1.4 (0.3)	1.4 (0.4)	1.5 (0.4)	1.4 (0.4)
TG, mean (SD)	1.5 (1.1)	1.5 (1.0)	1.4 (1.0)	1.4 (0.9)
SBP, mean (SD)	135.3 (19.3)	136.0 (19.9)	135.2 (20.0)	136.2 (20.3)
BMI, mean (SD)	26.1(4.4)	26.6 (4.5)	26.9 (4.7)	27.2 (4.9)
WHR, mean (SD)	0.8 (0.1)	0.9 (0.1)	0.9 (0.1)	0.9 (0.1)
Blood pressure-lowering treatment, %	9.0	13.0	14.3	21.3
Lipid-lowering treatment, %	1.5	3.2	7.1	14.6
Positive family history for any diabetes, %	N/A	25.8	26.4	28.7
Positive family history for early MI, %	23.6	25.5	25.6	15.5
ASCVD risk, mean (SD)	4.4 (5.5)	6.0 (7.9)	5.7 (7.3)	6.6 (7.9)
CHARGE-AF, mean (SD)	2.7 (2.3)	4.4 (4.8)	4.3 (4.4)	5.7 (5.6)
T2D risk, mean (SD)	4.6 (7.3)	6.3 (9.5)	8.0 (11.2)	8.9 (11.3)
Prevalent CHD, %	3.3	5.3	5.0	6.4
Prevalent MI, %	0.8	1.3	1.0	1.7
Prevalent AF, %	0.9	1.8	1.7	2.9
Prevalent T2D, %	0.4	2.8	3.8	4.5
Prevalent breast cancer in women, %	1.0	1.2	1.3	2.3
Prevalent prostate cancer in men, %	0.0	0.6	0.5	1.2
Incident CHD, %	14.2	13.8	8.8	5.3
Incident MI, %	5.6	5.7	3.7	2.0
Incident AF, %	9.5	9.1	6.0	4.0
Incident T2D, %	15.7	13.2	10.1	7.2
Incident breast cancer in women, %	5.3	4.1	3.0	1.5
Incident prostate cancer in men, %	5.5	5.6	3.6	2.0

TC = total cholesterol, LDL = low-density lipoprotein (using the Friedewald equation), HDL = high-density lipoprotein, TG = triglycerides, SBP = systolic blood pressure, BMI = body mass index, WHR = waist-hip ratio, CHD = coronary heart disease, MI = myocardial infarction. T2D = type 2 diabetes, AF = atrial fibrillation or flutter. Units: lipid measurements mmol/l, SBP mmHg, BMI kg/m2. ASCVD = the 10-year risk calculator for hard atherosclerotic cardiovascular disease according to the Pooled Cohort Equations by ACC/AHA (2013). CHARGE AF = AF risk calculator. ASCVD, CHARGE-AF and T2D risk contains only incident cases and controls.

Table S4. Impact of family history on polygenic risk score (PRS) effect size estimates (per standard deviation increment) in FINRISK (total n = 21,813; detailed information in Supplementary Table S3), obtained from Cox proportional hazards models.

	HR (95% CI)	p
CHD PRS	1.27 (1.22-1.32)	4.77×10^{-28}
Family history of early MI	1.49 (1.36-1.63)	7.57×10^{-18}
CHD PRS + family history of early MI		
CHD PRS	1.26 (1.21-1.31)	5.24×10^{-26}
Family history of early MI	1.45 (1.32-1.59)	1.21×10^{-15}
T2D PRS	1.57 (1.50-1.66)	1.24×10^{-68}
Family history of any diabetes	1.62 (1.47-1.78)	1.67×10^{-22}
T2D PRS + family history of any diabetes		
T2D PRS	1.54 (1.46-1.62)	2.24×10^{-62}
Family history of any diabetes	1.49 (1.35-1.64)	8.27×10^{-16}

CHD = coronary heart disease, T2D = type 2 diabetes. T2D models adjusted for BMI. Only incident cases included. All tests were two-tailed.

Table S5. Number and proportion of individuals reclassified with addition of polygenic risk score (PRS) to clinical risk assessment.

		All		Early-onset		Late-onset	
		Individuals reclassified		Individuals reclassified		Individuals reclassified	
		Up (%)	Down (%)	Up (%)	Down (%)	Up (%)	Down (%)
CHD	Cases	30 (2.5)	19 (1.6)	16 (4.4)	2 (0.6)	14 (1.7)	17 (2.0)
	Non-cases	397 (2.1)	428 (2.3)	234 (1.7)	183 (1.3)	163 (3.1)	245 (4.6)
T2D	Cases	89 (6.6)	25 (1.9)	3 (1.4)	2 (1.0)	86 (7.6)	23 (2.0)
	Non-cases	152 (0.8)	107 (0.5)	12 (0.1)	3 (<0.1)	140 (1.3)	104 (1.0)
AF	Cases	37 (16.2)	8 (3.5)	23 (24.7)	4 (4.3)	14 (10.3)	4 (2.9)
	Non-cases	796 (7.7)	517 (5.0)	496 (7.3)	110 (1.6)	300 (8.5)	407 (11.5)
Breast cancer	Cases	113 (15.2)	17 (2.3)	5 (3.2)	1 (0.6)	108 (18.4)	16 (2.7)
	Non-cases	1,576 (4.2)	755 (2.0)	68 (0.4)	6 (<0.1)	1,508 (7.8)	749 (3.9)
Prostate cancer	Cases	168 (14.3)	75 (6.4)	62 (18.3)	9 (2.7)	106 (12.7)	66 (7.9)
	Non-cases	1,593 (3.3)	1,986 (4.2)	625 (1.8)	294 (0.8)	968 (8.1)	1,692 (14.1)

CHD = coronary heart disease, AF = atrial fibrillation or flutter, T2D = type 2 diabetes. Reclassification proportion in cases calculated with all cases in the denominator, with a similar approach for controls. Reclassification thresholds: 10-year risk ≥7.5% for CHD, 10-year risk ≥33% for T2D, 5-year risk ≥5% for AF, and a 10-year risk ≥5% for breast and prostate cancer. Clinical risk assessment based on following risk factors: The ASCVD risk calculator used for CHD includes age, sex, total cholesterol, high-density lipoprotein, systolic blood pressure, blood pressure-lowering medication, diabetes, and smoking status; T2D analyses include age, sex, body mass index, history of stroke or CHD, parental history of diabetes, systolic and diastolic blood pressure, high-density lipoprotein, and triglycerides; the CHARGE-AF calculator used for AF includes age, height, weight, systolic and diastolic blood pressure, smoking status, blood pressure-lowering medication, diabetes, heart failure, and history of myocardial infarction; breast cancer analyses include age, family history of breast cancer, current smoking, body mass index, alcohol use disorder, years of hormone replacement therapy, and having conceived one or more children; prostate cancer analyses include age, family history, and history of benign prostate hyperplasia. Early- and late-onset assessment performed by dividing individuals into groups according to age at baseline (55 for CHD, 45 for T2D, 60 for AF, 45 for breast cancer, and 55 for prostate cancer).

Table S6. Hazard ratios (HR) and 95% confidence intervals (CI) per standard deviation increment in FINRISK, obtained from Cox proportional hazards models.

	N total	N cases	HR (95% CI)	p
Incident coronary heart disease	20,188	2,197	1.25 (1.18-1.32)	1.74×10^{-14}
Incident type 2 diabetes	21,030	1,346	1.70 (1.63-1.78)	8.82×10^{-133}
Incident atrial fibrillation or flutter	19,691	1,431	1.62 (1.54-1.70)	8.85×10^{-78}
Incident breast cancer	11,332	404	1.75 (1.59-1.92)	2.61×10^{-30}
Incident prostate cancer	10,258	444	1.88 (1.71-2.06)	4.74×10^{-41}

All tests were two-tailed.

Table S7. Hazard ratios (HR) and 95% confidence intervals (CI) for polygenic risk score (PRS) bins in FINRISK, obtained from Cox proportional hazards models.

	HR (95% CI)	p	N cases / N contols
CHD PRS			
<2.5	0.65 (0.47-0.89)	0.008	39 / 466
2.5-20	0.81 (0.72-0.92)	0.001	313 / 3,220
20-80	1 (reference)	-	1,274 / 10,838
80-97.5	1.35 (1.22-1.50)	2.45×10^{-8}	471 / 3,062
>97.5	2.42 (1.97-2.97)	2.21×10^{-17}	100 / 405
T2D PRS			
<2.5	0.23 (0.14-0.38)	7.14×10^{-9}	16 / 513
2.5-20	0.50 (0.43-0.58)	2.80×10^{-21}	224 / 3,476
20-80	1 (reference)	-	1,406 / 11,278
80-97.5	1.90 (1.73-2.08)	4.44×10^{-43}	719 / 2,981
>97.5	2.99 (2.52-3.54)	1.44×10^{-36}	151 / 378
AF PRS			
<2.5	0.44 (0.26-0.73)	0.002	15 / 478
2.5-20	0.60 (0.50-0.72)	1.69×10^{-8}	143 / 3,303
20-80	1 (reference)	-	779 / 11,035
80-97.5	1.94 (1.72-2.19)	4.28×10^{-27}	406 / 3,039
>97.5	3.19 (2.56-3.98)	8.67×10^{-25}	88 / 405
Breast cancer PRS			
<2.5	0.43 (0.16-1.16)	0.09	4 / 280
2.5-20	0.44 (0.30-0.65)	3.58×10^{-5}	29 / 1,954
20-80	1 (reference)	-	221 / 6,577
80-97.5	1.94 (1.56-2.42)	3.90×10^{-9}	123 / 1,860
>97.5	3.05 (2.04-4.55)	5.34×10^{-8}	27 / 257
Prostate cancer PRS			
<2.5	0.10 (0.01-0.73)	0.02	1 / 256
2.5-20	0.38 (0.25-0.58)	4.90×10^{-6}	25 / 1,770
20-80	1 (reference)	-	233 / 5,921
80-97.5	2.14 (1.74-2.64)	6.58×10^{-13}	146 / 1,649
>97.5	3.93 (2.79-5.53)	3.90×10^{-15}	39 / 218

CHD = coronary heart disease, AF = atrial fibrillation or flutter, T2D = type 2 diabetes, PRS = polygenic risk score. Only incident cases included. All tests were two-tailed.

Table S8. The prospective epidemiological and disease-based cohorts, and hospital biobank samples in FinnGen Data Freeze 3.

Cohort	N
Auria biobank*	9,967
Blood Service biobank	13,222
Borealis biobank*	1,368
Botnia Family	1,216
Botnia New	6
Botnia PPP	4,856
Botnia Sib-Helsinki	431
Corogene	4,495
Eastern Finland biobank*	1,965
FinHealth 2017	5,783
FINRISK 1992-2012	29,550
GeneRISK	6,960
Health 2000	6,602
Health 2011	711
Helsinki biobank*	21,014
Kuusamo 2011	145
Migraine	7,732
SUPER	4,402
Tampere biobank*	1,973
THL Diabetes	6,983
Twins	5,919
Sum	**135,300**

*Hospital-based biobanks

Table S9. Disease endpoint definitions.

	Additional definitions	Only main diagnosis accepted	ICD-10	ICD-9	ICD-8	ICD-10 exclusions	Cause of death ICD-10	Cause of death ICD-9	Cause of death ICD-8	Cause of death ICD-10 exclu-sions	Cause of death ICD-9 exclu-sions	Topo-graphical codes*
Coronary heart disease	Myocardial infarction|Myocardial infarction, strict|Complications following myocardial infarction|Prior myocardial infactrion|Angina pectoris|Other coronary atherosclerosis|Coronary artery bypass graft**|Coronary angioplasty**											
Major coronary heart disease event	Myocardial infarction|Coronary artery bypass graft|Coronary angioplasty	Yes	I20.0 | I21 | I22	410 | 4110	410 | 411.0		I2[1-5] | I46 | R96 | R98	41[0-4] | 798	41[0-4] | 798		7980A	
Myocardial infarction, strict		Yes	I21 | I22	410	410		I21 | I22	410	410			
Myocardial infarction			I21 | I22	410	410		I21 | I22	410	410			
Complications following myocardial infarction			I23	-	-		I23	-	-			
Old myocardial infarction			I25.2	412	412		I25.3	412	412			
Angina pectoris			I20	413 | 411[0-1]	413		I20	413 | 411[0-1]	413			
Other coronary atheroclerosis			I25 | I24 | Z95.1 | T82.2	414 | 9960A	414	I25.3	I25 | I24 | Z95.1 | T82.2	414 | 9960A	414	I25.3		
Atrial fibrillation and flutter	Eligibility for special reimbursement for apixaban, dabigatran, edoxaban, rivaroxaban or dronedarone for ICD-10 I48		I48	4273	427.92		I48	4273	427.92			

Malignant neoplasm of breast / breast cancer	Reimbursement for medications used for treating breast cancer							C50
Malignant neoplasm of prostate / prostate cancer	Reimbursement for medications used for treating prostate cancer							C61
Type 2 diabetes* **	Any type 2 diabetes diagnosis defined below | Medication purchases for ATC A10B, Blood glucose lowering drugs, excluding insulins.				E10[0-9]			
Type 2 diabetes with coma		E11.0	2502A	-		E11.0	2502A	
Type 2 diabetes with ketoacidosis		E11.1	2501A	-		E11.1	2501A	
Type 2 diabetes with renal complications		E11.2	2503A	-		E11.2	2503A	
Type 2 diabetes with ophthalmic complications		E11.3	2504A	-		E11.3	2504A	
Type 2 diabetes with neurological complications		E11.4	2505A	-		E11.4	2505A	
Type 2 diabetes with peripheral circulatory complications		E11.5	2506A	-		E11.5	2506A	
Type 2 diabetes with other specified/ multiple/unspecified complications	Eligibility for medication reimbursement with ICD-10 E11	E11[6-8]	2507A | 2508A	-		E11[6-8]	2507A | 2508A	
Type 2 diabetes without complications		E11.9	2500A	-		E11.9	2500A	

* The International Classification of Diseases for Oncology, Third Edition (ICD-O-3). **Procedure code identified at hospital discharge or from the nationwide register of invasive cardiac procedures. *** In FinnGen analyses, individuals with type 1 diabetes were excluded from cases (ICD-10 E10[0-9], ICD-9 250[0-8]B as a hospital discharge diagnosis or cause of death, or E10 for medication reimbursement)

Table S10. Genome-wide association studies used for constructing the polygenic risk scores and the number of variants in the final scores.

	GWAS summary statistics source	Article link	Data download link	Most recent access to data download	SNPs in discovery GWAS	SNPs in PRS calculation	LD radius	Additional information
Coronary heart disease	UKBB SAIGE	https://www.nature.com/articles/s41588-018-0184-y	https://www.dropbox.com/sh/wuj4y8wsqjz78om/AAACfAJK54KtvnzSTAoaZTLma?dl=0	Nov 2, 2018	28 345 446	6 412 950	4 000	PheCode 411 Ischemic heart disease
Type 2 diabetes	Mahajan et al 2018	https://www.nature.com/articles/s41588-018-0241-6	http://www.diagram-consortium.org/downloads.html	Dec 21, 2018	23 465 133	6 437 380	4 000	Not adjusted for BMI
Atrial fibrillation and flutter	Nielsen et al 2018	https://www.nature.com/articles/s41588-018-0171-3	http://csg.sph.umich.edu/willer/public/afib2018/	Dec 21, 2018	34 740 187	6 171 733	4 000	
Breast cancer	Michailidou et al 2017	https://www.nature.com/articles/nature24284	http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/gwas-icogs-and-oncoarray-summary-results/	Dec 21, 2018	11 792 358	6 390 808	4 000	
Breast cancer, estrogen receptor-positive	Michailidou et al 2017	https://www.nature.com/articles/nature24284	http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/gwas-icogs-and-oncoarray-summary-results/	Dec 21, 2018	11 784 434	6 390 799	4 000	
Breast cancer, estrogen receptor-negative	Michailidou et al 2017	https://www.nature.com/articles/nature24284	http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/gwas-icogs-and-oncoarray-summary-results/	Dec 21, 2018	11 784 725	6 390 805	4 000	
Prostate cancer	Schumacher et al 2018	https://www.nature.com/articles/s41588-018-0142-8	http://practical.icr.ac.uk/blog/?page_id=8088	Dec 21, 2018	20 734 509	6 606 785	4 000	

Table S11. The LDpred algorithm uses a tuning parameter p for denoting the fraction of variants assumed to be causal for the disease. The PRS with the highest C-index (bolded) in FINRISK (total n = 21,813), was chosen for the subsequent analyses.

	Fraction of causal markers	C-index
Coronary heart disease	0.0001*	0.8163
	0.0003*	0.8162
	0.001*	0.8163
	0.003	**0.8203**
	0.01	0.8195
	0.03	0.8188
	0.1	0.8184
	0.3	0.8183
	1	0.8183
	inf	0.8183
Type 2 diabetes	0.0001*	0.7022
	0.0003*	0.7043
	0.001*	0.7033
	0.003*	0.7033
	0.01*	0.7089
	0.03*	0.7091
	0.1	0.7374
	0.3	**0.7417**
	1	0.7402
	inf	0.7398
Atrial fibrillation or flutter	0.0001*	0.7912
	0.0003*	0.7915
	0.001*	0.7921
	0.003*	0.7916
	0.01*	0.7942
	0.03	**0.8135**
	0.1	0.8107
	0.3	0.8082
	1	0.8057
	inf	0.8055
Prostate cancer	0.0001*	0.8076
	0.0003*	0.8077
	0.001*	0.8096
	0.003	0.8140
	0.01	**0.8416**
	0.03	0.8341
	0.1	0.8270
	0.3	0.8237
	1	0.8224
	inf	0.8223
Breast cancer	0.0001*	0.6426
	0.0003*	0.6403
	0.001*	0.6454
	0.003*	0.6490
	0.01*	0.6422

	0.03	**0.7042**
	0.1	0.6955
	0.3	0.6892
	1	0.6852
	inf	0.6853
Breast cancer, estrogen receptor-positive	0.0001*	0.6404
	0.0003*	0.6421
	0.001*	0.6434
	0.003*	0.6450
	0.01*	0.6479
	0.03	**0.6990**
	0.1	0.6912
	0.3	0.6868
	1	0.6834
	inf	0.6833
Breast cancer, estrogen receptor-negative	0.0001*	0.6403
	0.0003*	0.6410
	0.001*	0.6405
	0.003	**0.6511**
	0.01	0.6472
	0.03	0.6449
	0.1	0.6438
	0.3	0.6435
	1	0.6434
	inf	0.6432

*One or multiple chromosomes failed to converge. C-index from Cox proportional hazards model (follow-up as time scale), stratified by sex and adjusting for age, ten first principal components of ancestry, FINRISK survey collection year, and genotyping array. Only incident cases included.