Preventing premature deaths through polygenic risk scores

Melisa Chuong, Deborah Thompson, Michael Weale, Fernando Riveros-McKay, Nilesh Samani, Daniel Wells, Vincent Plagnol, Gil McVean, Euan A Ashley, Peter Donnelly, Seamus Harrison*, Jack W O'Sullivan*

Abstract

Background:
Polygenic risk scores (PRS) have demonstrated predictive validity across a range of cohorts and diseases, but quantifying their clinical utility remains a challenge. As PRS can be derived from a single biological sample and remain stable throughout life, we explore the potential of PRS to optimize existing screening programs.

Methods:
Using UK Biobank (n = 444,949), we quantified the potential clinical benefits arising from a knowledge of PRS across seven diseases with existing screening programs globally (breast cancer, prostate cancer, colorectal cancer, abdominal aortic aneurysm, type 2 diabetes, hypertension, and coronary artery disease). We identified individuals at high genetic risk (PRS OR>2) and very high genetic risk (PRS OR>3) and estimated the optimal screening ages for these genetically high risk individuals, based on the equivalent risk to population-level risk at recommended screening ages. We then leveraged published mortality estimates, with and without screening-based interventions, to assess the potential benefits of tailoring screening age based on genetic risk. We also estimated the case enrichment ratio, which is a ratio of the percentage of cases in the high PRS risk group and in the total population.

Findings:
Very high risk individuals reach the risk level associated with usual starting screening age on average 10.8 years earlier, high risk individuals 8.9 years earlier and reduced risk individuals (OR<0.5) 16.8 years later. During this time, case enrichment in the high risk group is between 1.7 and 3.0 depending on disease. Across all seven diseases, appropriate interventions following PRS-guided screening would reduce premature deaths in high risk individuals by 19.0%.

Conclusion:
Knowledge of genetic risk, measured using PRS, has the potential to deliver substantial public health benefits when aggregated across conditions, and could reduce premature mortality by tailoring existing screening programs.

Introduction

Common, chronic diseases account for over 90% of the $4.5 trillion annual healthcare spend in the US.[1] There are a range of public health screening programs aimed at reducing the burden of these diseases, either via early detection (e.g. screening for asymptomatic cancers) or prevention (e.g. screening for cardiovascular disease risk). Most programs use simple risk factors for determining initial access to screening, notably age and sometimes sex. This approach leads to overtreatment in some,[2] and undertreatment in others.[3]

Despite its widespread use, fixed age cutoffs fail to account for the considerable variability in individual risk, often missing high risk individuals who are younger than the age threshold and including lower risk individuals who are older.[4] Data on the effectiveness of age-based criteria in capturing true high risk populations are limited, and the current approach may inadvertently reduce the efficiency and equity of screening programs. Understanding the spectrum of risk across the population may allow better targeting of interventions to those most likely to benefit, and least likely to be harmed.

One promising avenue for improving risk stratification lies in genetic factors. Increasingly recognized as a key determinant of disease risk, genetics is regularly the strongest measurable risk factor.[5] Polygenic risk scores (PRS), an aggregate of the effects of common genetic variants associated with a disease into a single score, have emerged as novel biomarkers that are stable throughout life and available for a broad spectrum of diseases.[6] Unlike other biomarkers, a single genetic test allows the construction of multiple PRS that predict the risk of a broad range of diseases.

In this study, we estimate the potential clinical benefits of tailoring current screening programs to genetic risk. Specifically, we calculate the proportion of the population who are at sufficient risk of disease to warrant earlier initiation, identify the age at which this earlier screening should occur, and quantify the mortality benefits of earlier, PRS-guided screening.

Methods

Screening diseases and methods overview

We identified a list of seven common diseases for which age-related public health screening programs exist globally: female breast cancer, colorectal cancer, prostate cancer, type 2 diabetes mellitus, hypertension, coronary artery disease which we define as a combination of either fatal coronary heart disease and/or myocardial infarction, and abdominal aortic aneurysm. ‘Screening’ refers here either to screening for an existing condition, or to screening for risk in order to offer preventive intervention.

For these seven diseases, we identified individuals at high risk due to their PRS (Figure 1A). We then identified the age at which PRS-defined high risk individuals would reach the same risk as that of the average population at the conventional starting age for screening (Figure 1B). Finally, we used this information to estimate the number of mortality events that would occur prior to the conventional starting age for screening in this group, and to estimate the

potential benefit by leveraging published estimates of reductions in mortality from interventions (Figure 1C).

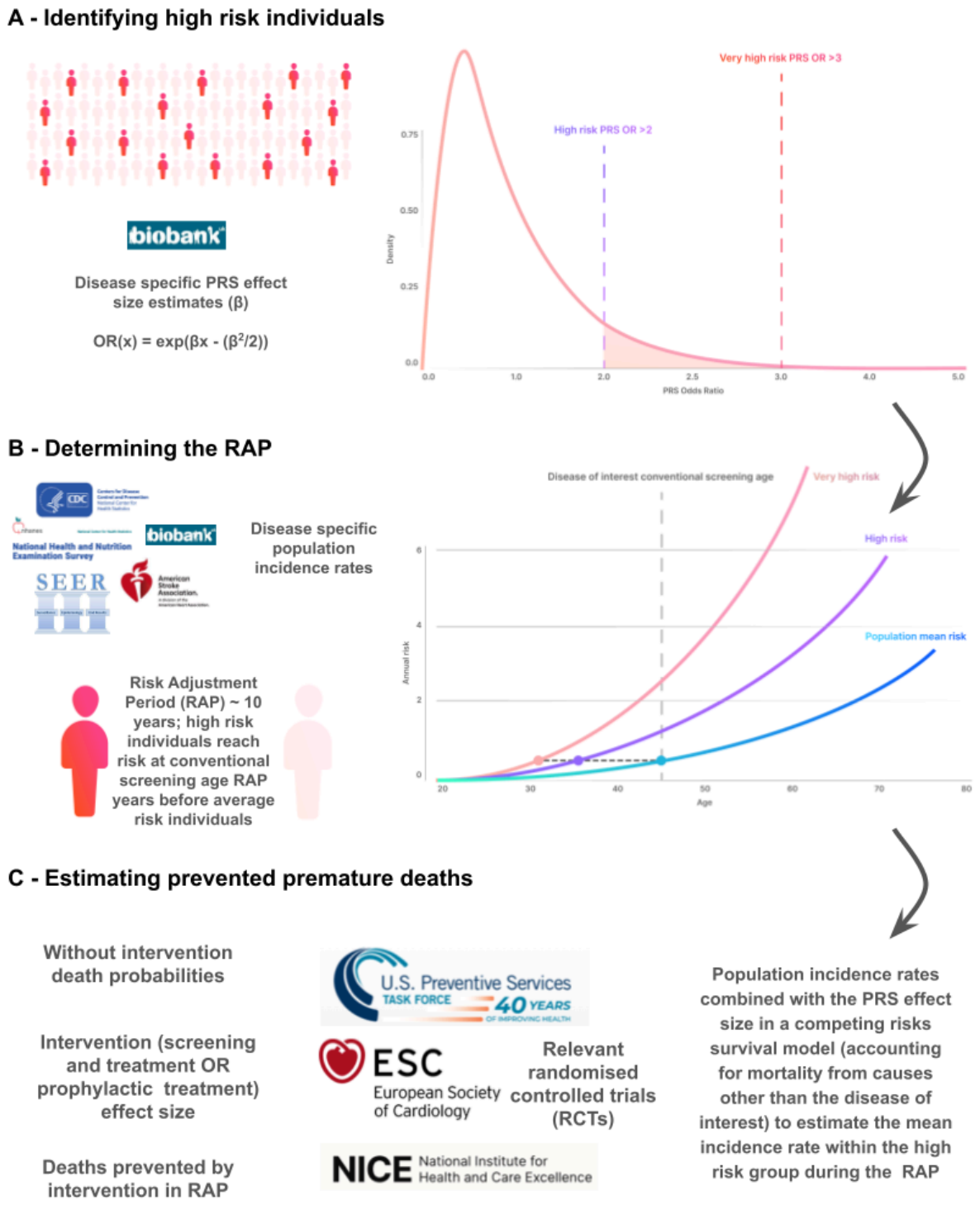

Figure 1. Overview of study methods.

PRS and PRS effect sizes

We previously constructed and released the UK Biobank PRS Release v2 as a freely-available resource for UK Biobank researchers.[7] The UK Biobank (UKB) is a UK based prospective cohort of ~500,000 individuals, aged 40-69 when recruited between 2006 and 2010.[8,9] We used the Standard PRS set, which was constructed from non-UKB training data only, as our source of high-performance PRS, along with their corresponding trait definitions. Briefly, the PRS were generated using meta-analyzed external genome-wide association study summary statistics. Per-individual PRS values were calculated as the genome-wide sum of the per-variant posterior effect size multiplied by allele dosage. All PRS were centered and standardized based on external reference data (1000 Genomes Project) to ensure means of approximately zero and standard deviations of approximately one within major genetically inferred ancestry groups. The same pipeline was used to construct a new PRS for abdominal aortic aneurysm.

PRS scores were available for all 454,949 individuals in the UKB cohort, which was then split into study training and test sets. The UKB training set consisted of 122,285 individuals. Using previously described methods applied to their genetic data,[7] 2867 were inferred with East Asian (EAS) ancestry, 100,000 with European (EUR), 9543 with South Asian (SAS), 9486 with Sub-Saharan African (AFR), and 389 with Native/Indigenous American (AMR) ancestry. The UKB training set was used to estimate the odds ratio per standard deviation (OR per SD) of each PRS in each ancestry group. The remaining individuals outside of the UKB training set were of EUR ancestry only. Following removal of individuals who were related (up to third degree) with individuals in the UKB training set, 332,664 individuals were placed in an independent UKB testing set. This testing set was used for empirical sensitivity analyses outlined below, and to calculate percentages of individuals at high risk for multiple diseases.

Following previous practice,[5,10–12] we defined individuals to be at ‘high risk’ of a disease if their PRS value was associated with an odds ratio (OR) >2 compared to the population mean risk. For ‘very high risk’, we used a PRS value associated with OR>3, and for ‘reduced risk’ we used a PRS value associated with OR<0.5 (Figure 1A). A detailed description of the risk group calculation is available in the Supplementary Materials.

Risk advancement periods (RAP)

For each of the seven diseases, we searched international guideline registers including: United States Preventive Services Task Force (USPSTF), United Kingdom National Institute of Clinical Excellence (NICE), European society guidelines (e.g. European Society of Cardiology) and national specialty guidelines (e.g. American Heart Association, American Society of Clinical Oncology). From these guidelines we extracted recommendations on screening and screening start age.

For each disease, we then used published age, ethnicity, race and sex-specific incidences for the US population to calculate a set of baseline one-year incidence rates from age 0 to 80 years, representing the average population rate with age. Population incidence rates were obtained from a number of sources, as described in the Supplementary Materials. To estimate the disease rate in each disease-specific high risk group, baseline incidences were

multiplied by the mean relative risk within the high risk group compared to baseline, which was estimated from the mean OR per SD in the relevant upper portion of the PRS distribution (using the equivalence between relative risk and OR that applies when baseline incidences are low, and adjusting for bias in the transformation from PRS to OR).[13,14] We then compared the incidence curves in order to find the age at which individuals in the high risk group reached the same level of risk as that found in the whole population at the conventional screening start age. We refer to this starting age as the PRS-guided screening age. The difference between the conventional screening age and the PRS-guided screening age defines the risk advancement period (RAP) (Figure 1B).[15] Our primary analyses used PRS-adjusted screening ages that were calculated using baseline rates averaged over males and females, where applicable (sex-specific ages are reported in the Supplementary Materials). Sensitivity analyses using different conventional screening start ages were conducted for breast cancer, and colorectal cancer, in line with alternative guidelines. We also estimated the 'case enrichment ratio', which is a ratio of the expected case incidence in the high PRS risk group to the expected case incidence in the total population (equivalently, by rearrangement, it is the ratio of the percentage of all cases who are in the high PRS risk group to the percentage of the total population who are in the high risk group). For example, for breast cancer, 7.08% of the cases are in the high risk group, and 2.98% of the population is in the high risk group, thus the case enrichment ratio is 2.37.

Estimation of events and deaths without PRS-guided screening

Sex-specific population incidence rates (derived as described above) were combined with the PRS effect size in a competing mortality survival model,[16] using US-specific estimates of mortality from other causes.[17] This model was used to estimate the disease incidence rate that would occur without screening-based intervention within each disease-specific high risk group during the RAP time interval between the earlier PRS-guided screening age and the conventional screening age. These disease events represent opportunities for early detection or preventive intervention (Figure 1C). We note that this is a conservative estimate of the true number of cases for some diseases, because it is based on current reported population incidence rates (cases that were diagnosed at these ages in the absence of screening, presumably symptomatically), and so it does not include additional early-stage prevalent cases that would be diagnosed once screening begins.

The predicted event rate in each high risk group over the RAP was multiplied by the proportion of the population estimated to be in the high risk group and scaled to give the number of cases predicted to occur in the high risk group out of a hypothetical total population of 100,000 males and 100,000 females. We describe these as the '*without-intervention events*'. The without-intervention event rates were calculated separately for males and females. Confidence intervals (CI) were derived from the 95% CI of the log(OR) per SD of the PRS, assuming that the PRS-guided screening age and the proportion of the population in the high risk groups are known without error, and ignoring uncertainty in the population baseline rates.

To estimate *without-intervention mortality* rates, we further identified, from relevant randomized controlled trials, the control group probabilities of disease-specific death among diagnosed individuals over approximately 10 years following diagnosis. The

without-intervention events were multiplied by these disease-specific mortality rates to obtain an estimate of the without-intervention disease-specific deaths.

Quantification of clinical benefits of PRS-based screening

We consider two kinds of clinical benefit: an ‘early detection’ benefit that arises from earlier screening, and thus, earlier treatment for current disease, and an ‘early prevention’ benefit that arises from earlier preventive intervention that is administered to individuals at high risk due to their PRS. A literature search was conducted to identify randomized controlled trials reporting benefit effect sizes of each intervention over sufficiently long follow-up time periods.

Analyses were repeated for disease-specific very high risk groups (OR>3) and for reduced risk groups (OR<0.5). For reduced risk groups, the PRS-adjusted screening ages were later than the conventional screening ages, reflecting a negative RAP (i.e. a risk delay period). For some diseases, the reduced risk group never reaches the average risk at the conventional screening age. In this situation, calculations were based on taking the PRS-adjusted screening age to be 80 years, although we note that this does not reflect a realistic use-case.

When quantifying the combined benefit over all diseases, to avoid double-counting, we only included the benefit of ‘early detection’ and excluded the benefit of ‘early prevention’ for breast cancer. We did not estimate clinical benefits of delaying screening in the reduced risk groups as there are no randomized controlled trials or effect size estimates for delayed screening.

As a sensitivity analysis, the without-intervention events and disease-specific deaths were also estimated empirically in the UK Biobank testing set. Disease rates within the screening groups of interest (for the period of time defined by the RAP), were defined from incident rates within UK Biobank, conditional on being disease-free at the starting age.

Results

Individuals at risk for multiple diseases

PRS effect size estimates (OR per SD) were determined in the UKB training set (Table S1) and used to define the disease-specific high risk and very high risk individuals in the UKB testing set. Of the 332,664 individuals in the UKB testing set, there were 83,950 individuals at high risk (PRS OR>2) and 18,956 individuals at very high risk (PRS OR>3) for at least one of the seven diseases for which current screening programs exist (Table 1), corresponding to approximately 25% and 6% of the cohort, respectively.

Number of Diseases	At High Risk	At Very High Risk

0	248714 (74.8%)	313703 (94.3%)
1	73119 (22.0%)	18520 (5.6%)
2	9905 (3.0%	436 (0.1%
3	868 (0.3%)	5 (0%)
4	55 (0.0%)	0 (0%)
5	3 (0.0%)	0 (0%)
At least one	83950 (25.2%)	18956 (5.7%)

Table 1. Count (and percentage) of individuals in the UKB testing set at high (PRS OR>2) and very high risk (PRS OR>3) for the seven diseases).

High PRS individuals reach equivalent risk earlier - Risk Advancement Period (RAP)

For each of the seven diseases, we define the target incidence rate as the mean population incidence at the age at which the screening program begins. The main analyses used White self-reported ethnicity population incidence rates and PRS effect size estimates for genetically-inferred European-ancestry individuals in UKB. Multiplying population age-specific incidence rates by the average relative risk within each disease-specific high risk PRS group (PRS OR>2), we estimate that these high risk individuals would reach the target incidence rate between 4 and 22 years (mean 8.9 years) before the conventional screening age, depending on disease (Figure 2, Table S2). The diseases with the largest RAP for high risk individuals were hypertension and breast cancer (Figure 2, Table S2). A similar pattern was seen for the very high risk groups (OR>3), with RAP ranging from 5 to 22 years (Table S2) (mean 10.8 years). Individuals with a reduced PRS risk (OR <0.5) reached population level risk 6 to 40 years later (mean 16.8 years) than the conventional screening ages; for prostate cancer (based on a screening start age 55) and hypertension, we predict that the reduced risk groups will never reach the target incidence (Figure 2, Table S2).

Alternative conventional start ages for breast cancer, colorectal cancer and coronary artery disease were utilized for sensitivity analyses, with results summarized in Figure S1 and Table S2. The analyses were also repeated, this time replacing the population incidence rates with those for Black self-reported ethnicity and using a PRS effect size estimated in the genetically-inferred African-ancestry UKB training set (Figure S2, Figure S3, Table S1). Here the PRS effect sizes are smaller, meaning that fewer individuals have a PRS OR high enough to be in a high risk group. However, the slightly lower average OR within each high risk group leads to only a relatively small reduction in the length of the RAP for these diseases.

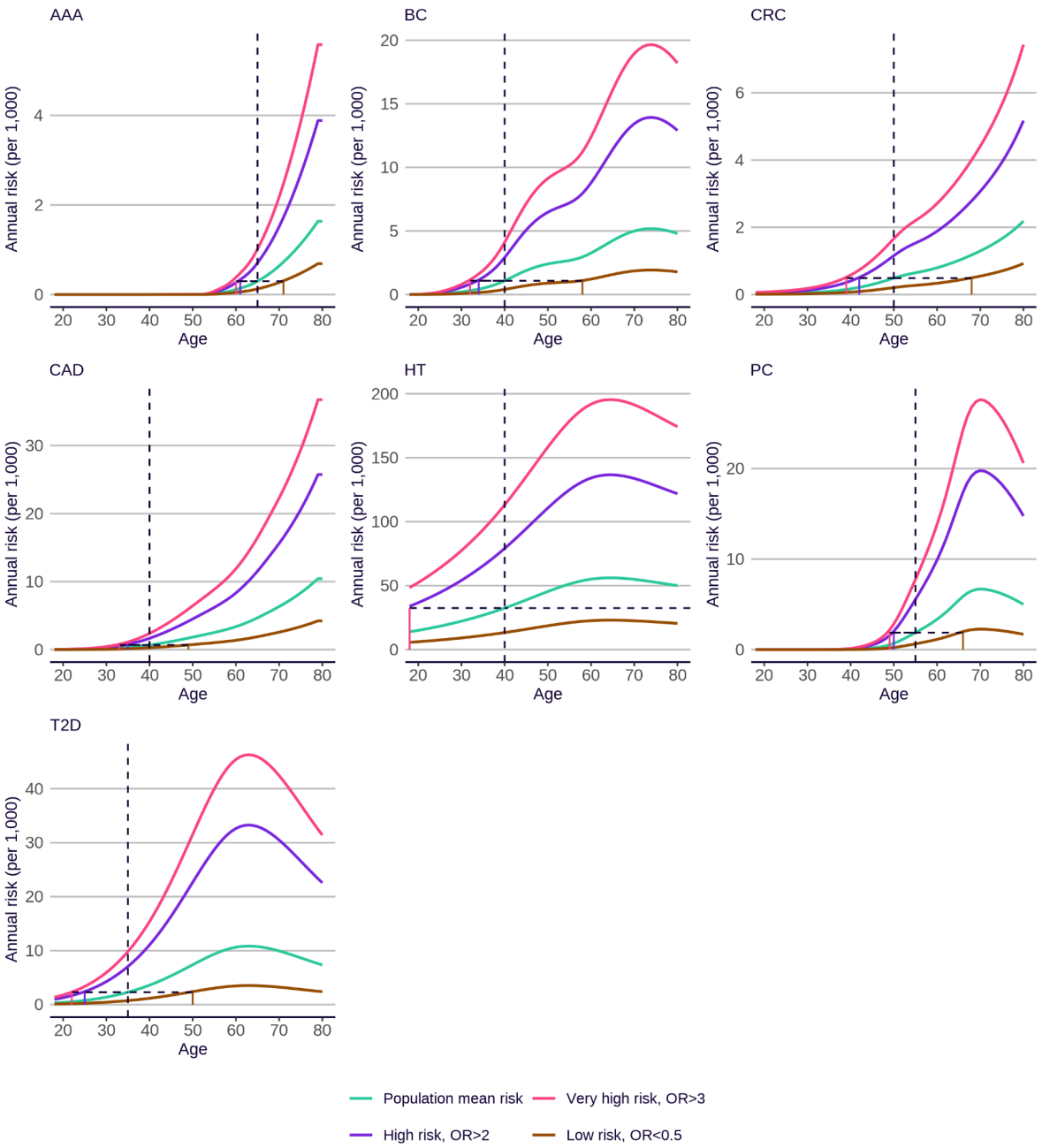

Figure 2. Annual incidence with age for individuals at high (PRS OR>2), very high (PRS OR>3), and reduced risk (PRS OR<0.5), compared to the population mean. Incidence rates are for the White self-identified ethnicity, and PRS effect sizes are estimated in UK Biobank training set individuals with estimated European genetic ancestry (see Figure S3 for an equivalent figure using Black self-identified ethnicity population rates and PRS effect sizes estimated in the African genetic ancestry subset of the UK Biobank training set). Incidences are sex-averaged, apart from prostate cancer (male) and breast cancer (female). Vertical dotted lines indicate the conventional screening start age. Horizontal dotted lines indicate the RAPs for different PRS-based risk groups. AAA = abdominal aortic aneurysm; BC = breast cancer; CRC = colorectal cancer; CAD = coronary artery disease; HT = hypertension; PC = prostate cancer; T2D = type 2 diabetes mellitus.

The clinical benefits of PRS-guided screening and treatment

For the seven diseases with existing screening programs, we determined the number of disease-specific deaths that could be prevented by offering earlier PRS-guided screening to those with high or very high PRS. For each disease, we identified the probability of disease associated death once diagnosed (over 10 years) within control groups of relevant randomized controlled trials; the appropriate intervention based on guidelines; and the corresponding intervention effect size based on randomized controlled trials (Table 2). For breast cancer, we identified effect sizes for both screening (mammography) and for risk-lowering preventive intervention (e.g. prophylactic treatment with tamoxifen, raloxifene, or aromatase inhibitors[18]).

Benefit	Disease	Probability of disease associated death once diagnosed (over 10 years)	Intervention	Intervention effect size	Measured outcome
Detection	Abdominal aortic aneurysm	0.009 (0.009 - 0.10) [19]	Abdominal Ultrasound [19]	0.65 (0.57 - 0.74) [19]	AAA mortality
Detection	Breast cancer	0.23 (0.22 - 0.25) [20]	Mammography [21]	0.75 (0.58 - 0.97) [20]	BC mortality in women screened 40-50
Detection	Colorectal cancer	0.3 (0.27 - 0.33) [22]	Colonoscopy [22]	0.74 (0.68 - 0.80) [22]	CRC mortality
Detection	Prostate cancer	0.08 (0.076 - 0.088) [23]	Prostate specific antigen (PSA) [24]	0.84 (0.73 - 0.96) [24,25]	PC mortality
Prevention	Type 2 diabetes mellitus	0.25 (0.24 - 0.27) [26]	Metformin [26]	0.64 (0.45 - 0.91) [26]	All-cause mortality
Detection	Hyper-tension	0.030 (0.02 - 0.042) [27]	Office blood pressure measurement [27]	0.91 (0.86 - 0.97) [27]	MI hospitalization
Prevention	Coronary artery disease	0.005 (0.04 - 0.06) [28]	HMG-CoA reductase inhibitors (Statins) [28]	0.75 (0.72 - 0.78) [29]	MI or coronary death
Prevention	Breast cancer	0.23 (0.22-0.25) [20]	Tamoxifen [18]	0.69 (0.59 - 0.84) [18]	Rate of invasive breast cancer in high risk women

Table 2. Probability of disease associated death (with 95% CI), guideline-based intervention, and intervention effect sizes (with 95% CI) for the seven screening diseases under early detection or early prevention benefit. References for the estimate or guideline are provided in square brackets. MI = myocardial infarction; for other disease abbreviations see Figure 2.

The early detection and early prevention benefits are displayed respectively in Tables 3 and 4, using a PRS threshold of OR>2 to define disease-specific high risk groups (sensitivity

analyses using alternative conventional screening ages are shown in Table S3). Across the seven diseases, between 2.9% (colorectal cancer) and 10.0% (type 2 diabetes) of the population are identified as being at high risk due to their PRS. Within these high risk groups, there is a case enrichment within the RAP of between 1.7 (hypertension) and 3.0 (type 2 diabetes) compared to the general population, resulting from their increased genetic risk (Figure 3). If relevant preventive measures were applied within these high risk groups, we estimate that between 9% (hypertension) and 35% (abdominal aortic aneurysm) of disease-specific deaths within these high risk groups could be prevented via early detection (Table 3), and between 24.4% (coronary artery disease) and 35.5% (type 2 diabetes) could be prevented via early treatment (Table 4). Summing across both types of benefit (but counting only the early detection benefit for breast cancer to avoid double counting), we estimate that, in a population of 100,000 individuals of each sex, 6386 disease cases would occur within the RAP in individuals with a high PRS risk for the relevant disease. These cases, which fall outside of current screening ages, would be expected to lead to 499 deaths. If a PRS-guided early detection and/or early prevention strategy was adopted, 95 of these deaths (19.0% of disease-specific deaths in the respective high risk groups) would be saved.

For example, the 7.2% of women in the high risk group for breast cancer could begin the intervention i.e. screening for early detection at age 34, instead of at the conventional age of 40 (Table 3). During this 6-year RAP, we would expect to see 75 cases of breast cancer in this group. Assuming that 23% of cases would lead to death (Table 2),[20] this would result in 17 breast cancer-specific deaths. The UK Age randomized control trial reported a 25% reduction in 10-year breast cancer mortality associated with mammographic screening between ages 40-49 (as opposed to the UK standard screening age of 50 years).[20] We hence estimate that 4.3 deaths could be prevented by PRS-guided screening in a population of this size.

Restricting interventions to individuals with a very high risk PRS (OR>3) would result in longer risk advancement periods, but a more modest overall impact, because the earlier interventions would be offered to fewer people (e.g. 1.6% of women at very high risk of breast cancer, as opposed to 7.2% at high risk). Over all seven diseases, we estimate that 70 disease-specific deaths in the very high risk groups could be prevented in a population of 100,000 individuals of each sex, combining early detection and early prevention benefits (Tables S4 and S5).

In addition to predicting numbers of cases based on US population incidence rates, we also looked empirically at the number of cases during the RAP in the high risk PRS groups in the held out testing set UK Biobank and found similar results (Tables S6 and S7).

Disease	AAA	AAA	BC	CRC	CRC	HT	HT	PC	Total
Benefit	detection	detection	detection	detection	detection	detection	detection	detection	-
Sex	female	male	female	female	male	female	male	male	-
Conventional screening age	65	65	40	50	50	40	40	55	-
PRS-guided screening age	61	61	34	42	42	18	18	50	-
Risk adjustment period in years (RAP)	4	4	6	8	8	22	22	5	-
% population in high risk group	2.98	2.98	7.22	2.89	2.89	3.96	3.96	9.4	-
Cases in whole population	13.89	134.59	390.29	234.35	276.59	31036.96	43031.61	554.79	75673
% cases in high risk group	7.08	7.07	19.32	6.83	6.84	7.53	6.78	27.59	-
Case enrichment	2.37	2.37	2.68	2.36	2.36	1.90	1.71	2.93	-
Cases in high risk group without intervention	0.98 (0.81 - 1.19)	9.52 (7.83 - 11.47)	75.39 (72.01 - 78.85)	16.00 (14.75 - 17.33)	18.92 (17.45 - 20.49)	2335.89 (2311.24 - 2360.26)	2916.87 (2897.02 - 2936.36)	153.05 (145.87 - 160.39)	5526
Expected cause-specific deaths without intervention	0.01 (0.01 - 0.01)	0.09 (0.07 - 0.10)	17.34 (16.56 - 18.14)	4.80 (4.43 - 5.20)	5.68 (5.23 - 6.15)	70.08 (69.34 - 70.81)	87.51 (86.91 - 88.09)	12.24 (11.67 - 12.83)	197
Expected cause-specific deaths prevented by PRS-guided intervention	0.0031 (0.0025 - 0.0037)	0.030 (0.025 - 0.036)	4.33 (4.14 - 4.53)	1.25 (1.15 - 1.35)	1.48 (1.36 - 1.60)	6.31 (6.24 - 6.37)	7.88 (7.82 - 7.93)	1.96 (1.87 - 2.05)	23

Table 3. Estimates of clinical benefits of PRS-guided screening leading to early detection benefit. ‘Cases in high risk group’ reports the number of cases which would have occurred in each disease-specific high PRS risk group during the RAP period (per 100k population of each sex). ‘Expected cause-specific deaths without intervention’ reports the number of deaths within these cases which would have occurred without screening to detect them at an earlier stage (per 100k population of each sex). Numbers in brackets are the 95% confidence intervals. For disease abbreviations see Figure 2.

Disease	BC	CAD	CAD	T2D	T2D	Total
Benefit	prevention	prevention	prevention	prevention	prevention	
Sex	female	female	male	female	male	-
Conventional screening age	40	40	40	35	35	-
PRS-guided screening age	34	35	35	25	25	-
Risk adjustment period in years (RAP)	6	5	5	10	10	-
% population in high risk group	7.22	4.43	4.43	10	10	-
Cases in whole population	390.29	117.32	324.87	1513.07	1192.27	3537
% Cases in high risk group	19.32	10.94	10.88	29.95	30.2	-
Case enrichment	2.68	2.47	2.45	2.99	3.02	-
Cases in high risk group without intervention	75.39 (72.01 - 78.85)	12.83 (12.23 - 13.45)	35.33 (33.69 - 37.03)	453.19 (440.20 - 466.32)	360.03 (349.61 - 370.56)	936
Expected cause-specific deaths without intervention	17.34 (16.56 - 18.14)	0.064 (0.061 - 0.067)	0.18 (0.17 - 0.19)	113.30 (110.05 - 116.58)	90.01 (87.40 - 92.64)	220
Expected cause-specific deaths prevented by PRS-guided intervention	5.38 (5.14 - 5.62)	0.016 (0.015 - 0.017)	0.044 (0.042 - 0.046)	40.77 (39.62 - 41.99)	32.40 (31.46 - 33.35)	78

Table 4. Estimates of clinical benefits of PRS-guided screening leading to preventive treatment. ‘Cases in high risk group without intervention’ reports the number of cases that would have occurred without preventive treatment in each disease-specific high risk group during the RAP period (per 100k population of each sex). ‘Expected cause-specific deaths without intervention’ reports the number of deaths within these cases which would have occurred without preventive treatment (per 100k population). Numbers in brackets are the 95% confidence intervals. For disease abbreviations see Figure 2.

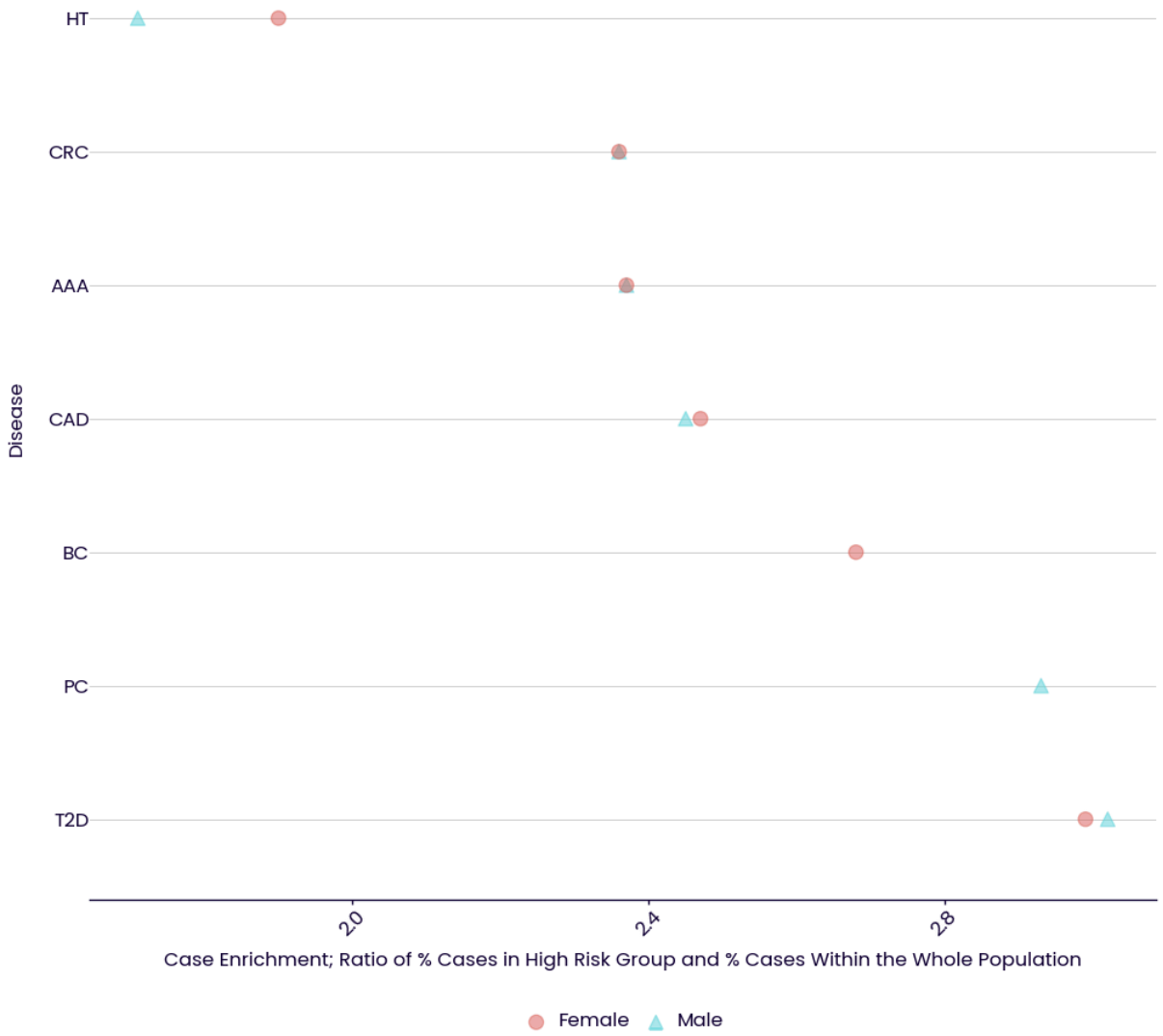

Figure 3. Case enrichment (ratio of the percentage of cases in high risk (PRS OR>2) individuals compared to that in the whole population), in men and women for the seven screening diseases.

Discussion

Healthcare systems continue to invest large amounts of resources into screening for many common diseases. In order to optimize cost-effectiveness and minimize harms from overtreatment, screening for early detection or prevention is targeted to groups of the population in whom the benefits are deemed to be greatest. This targeting is achieved using a very minimal number of predictors, typically age and sex. As these are relatively crude predictors of risk, a reliance upon these factors can lead to overtreatment in some groups, and undertreatment in others. Here we consider using an additional immutable risk factor, PRS, to support stratification of screening and screening-based intervention opportunities.

Specifically, we focus upon the potential clinical benefit of earlier intervention based upon genetic stratification across a range of important and common diseases.

Our paper shows three main results: 1) a single biological sample obtained once can predict risk of many important common diseases, and can identify a substantial proportion of the population as high risk; 2) PRS can be used to propose rational starting ages for early screening in high risk people; and 3) if PRS-guided interventions were implemented, premature mortality could be substantially reduced.

First, we show that PRS can predict the risk of seven common diseases (breast cancer, colorectal cancer, prostate cancer, type 2 diabetes, hypertension, coronary artery disease and abdominal aortic aneurysm), and that a quarter of the population is at high risk of at least one of these diseases. PRS has the advantageous property that data from a single test can be obtained once at an early age, in principle from birth, and can capture risk simultaneously across many diseases.

Second, regarding screening for these seven diseases, we show how PRS information could be used to propose rational earlier screening start times for people at high risk. We show that current screening misses opportunities to intervene in 6386 disease-specific cases per 100,000 individuals of each sex. These individuals have no current way (other than PRS) of being identified before the start of current screening programs. These missed opportunities for intervention inevitably lead to preventable, premature mortality in some. Compared to when conventional screening begins, we predict that high risk individuals (PRS OR>2) would reach the equivalent population risk around a decade earlier (ranging from 4 to 22 years across diseases). Hypertension and breast cancer were the diseases with the longest risk advancement periods. Conversely, reduced risk individuals (OR < 0.5) reached the equivalent population risk on average more than a decade after conventional screening begins. Together, this demonstrates that using age alone is a relatively crude way to stratify population health programs.

Third, we show that premature, disease specific deaths that occur before conventional screening occurs are potentially preventable with PRS-guided interventions. Across our included diseases, our results suggest that 499 disease-specific deaths (per 100,000 individuals of each sex) will occur in the corresponding high risk groups before conventional screening occurs, and 19% of the disease-specific deaths within these high risk groups would be preventable with PRS-guided interventions. We estimate that the highest proportions of preventable deaths during the risk advancement period are for type 2 diabetes, abdominal aortic aneurysm, and colorectal cancer, with respectively an estimated 36%, 35% and 26% of disease-specific deaths preventable with PRS-guided interventions. In absolute terms, we found that hypertension, type 2 diabetes and prostate cancer had the potential for the largest number of deaths saved by this approach.

We also note that there are likely to be additional benefits to PRS-guided screening, beyond mortality reduction. In this analysis, to provide uniformity across the diseases, our benefit analyses focused on mortality. However, identifying high risk individuals or those with diseases earlier also has the potential to reduce non-fatal events such as non-fatal myocardial infarction and stroke. Indeed, in absolute terms, the reduction in non-fatal events is likely to be greater.

The existing literature regarding PRS largely examines prediction of lifetime disease susceptibility,[10,31] and integration of PRS into existing clinical risk scores.[10,32,33] However, there are only a few studies concerning the identification of earlier conventional screening ages with PRS[31,34], or predicting cumulative disease burden across a number of diseases using PRS.[35] Chen et al estimated the personalized screening age of men and women in the highest PRS decile for colorectal cancer.[34] Our results were congruent for men, but not for women (we predicted women would reach the same risk at 34 where they predicted 48). Reasons for this difference may include the use of different conventional screening ages (50 versus 55 years), and also our use of national sources of incidence rates to support our estimates. Jermy et al examined the ability of PRS to estimate PRS-guided screening ages for breast cancer and type 2 diabetes.[31] This paper used a different methodology, comparing PRS percentiles, and used the old USPSTF breast cancer screening age (50), and found that women at the top 5% PRS risk would be eligible for breast cancer screening at 43 years. In contrast, we used the new USPSTF breast cancer screening age of 40 and found a PRS-guided screening age of 34 years. For type 2 diabetes, Jermy et al found people at the top 5% PRS risk for type 2 diabetes would be eligible for screening at 29 years, whereas our results suggest 25 years of age. However, Jermy et al did not use USPSTF guidelines to define conventional screening ages, and did not use national US incidence rates to support estimates. Widen et al examined the ability of PRS to identify people at high risk across a number of diseases.[35] These authors constructed polygenic health indexes, and estimated the ability of PRS to predict lifetime disease burden when combining PRS across a number of diseases. Our study quantifies the number of people at high and very high risk for a number of diseases and then estimates associated mortality estimates.

Collectively, our results show the power of PRS across multi-disease risk prediction and precision-guided screening, with the potential to save premature, preventable deaths in younger adults. When considered in combination with prior work showing the feasibility of integrating PRS into primary care,[30] our results show a promising clinical use case for PRS. Our results have implications for patients, clinicians, researchers and policy makers. For clinicians, PRS are increasingly available both via direct to consumer tests and within healthcare systems themselves.[36] We anticipate our data may provide some guidance - by providing some data to support conversations regarding risk and screening for patients presenting to clinics with PRS results. For patients, we hope our results show that PRS should be considered a risk factor that may help guide health behaviors. For policy makers, our results highlight the use of PRS as a risk factor that can improve screening. PRS can enhance accuracy and potentially identify people who are being missed by conventional screening programs. We have focussed on the benefit that people at high risk can obtain from earlier screening, but people at very low polygenic risk may also benefit from screening commencing at a later than conventional age, or less frequently, especially if screening is intrusive or carries other risks. We outline future research priorities for PRS in the section below..

Limitations

Our study has several limitations. Our analyses in UK Biobank are limited by the predominance of individuals of European ancestry, although we also modelled African ancestries and found similar results. We further attempted to mitigate any inequity in PRS

performance by using a validated approach to translate and improve PRS across ancestral groups.[32] Estimates of benefit effect sizes come from interventional trials, and real world adherence was not modelled. We estimated ancestry-specific PRS effect sizes in UK Biobank and applied these to models using predominantly US-based incidence and mortality rates - this assumes that ancestry-specific PRS effect sizes are consistent between these two countries. We did not consider pathogenic rare variants, which have an important impact on risk for carriers (e.g. cancer risk for *BRCA1/2* carriers, coronary artery disease risk for *APOB* carriers), but which have a low population prevalence (<0.5%).[5,12,37] A cost-benefit economic analysis was considered beyond the scope of this paper, but is necessary for future work. Our estimates of the absolute number of deaths prevented is small, but identifying and preventing mortality in younger adults is likely to lead to large improvements in quality adjusted life years, however this was not formally assessed. We focused on coronary artery disease, rather than on cardiovascular disease including ischemic stroke. We excluded ischemic stroke due to concerns around phenotype heterogeneity and imprecise phenotyping in biobanks,[38,39] which could confound our results. Lastly, guidelines are not unified in their recommendations for the seven diseases in which we quantified screening benefits, and alternative intervention pathways are also possible.

Conclusions

Genetically-guided public health screening can identify high risk individuals earlier and potentially prevent premature mortality in the general population.

References

1. Centers for Disease Control and Prevention. Fast Facts: Health and Economic Costs of Chronic Conditions. *Chronic Disease* https://www.cdc.gov/chronic-disease/data-research/facts-stats/index.html (2024).
2. Welch, H. G., Prorok, P. C., O'Malley, A. J. & Kramer, B. S. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. *N. Engl. J. Med.* **375**, 1438–1447 (2016).
3. Saw, S. T., Leong, B. D. K. & Aziz, D. A. A. Early Detection of Undiagnosed Abdominal Aortic Aneurysm and Sub-Aneurysmal Aortic Dilatations in Patients with High-Risk Coronary Artery Disease: The Value of Targetted Screening Programme. *Vasc. Health Risk Manag.* **16**, 215–229 (2020).
4. Mukama, T. *et al.* Risk-Adapted Starting Age of Screening for Relatives of Patients With Breast Cancer. *JAMA Oncol.* **6**, 68–74 (2020).
5. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
6. Lambert, S. A. *et al.* The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat. Genet.* **53**, 420–425 (2021).
7. Thompson, D. J. *et al.* A systematic evaluation of the performance and properties of the UK Biobank Polygenic Risk Score (PRS) Release. *PLOS ONE* **19**, e0307270 (2024).
8. Sudlow, C. *et al.* UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
9. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
10. Mars, N. *et al.* Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat. Med.* **26**, 549–557

(2020).
11. Hao, L. *et al.* Development of a clinical polygenic risk score assay and reporting workflow. *Nat. Med.* **28**, 1006–1013 (2022).
12. Fahed, A. C. *et al.* Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat. Commun.* **11**, 3635 (2020).
13. Zhang, J. & Yu, K. F. What's the Relative Risk? A Method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes. *JAMA* **280**, 1690–1691 (1998).
14. Pharoah, P. D. P. *et al.* Polygenic susceptibility to breast cancer and implications for prevention. *Nat. Genet.* **31**, 33–36 (2002).
15. Brenner, H., Gefeller, O. & Greenland, S. Risk and Rate Advancement Periods as Measures of Exposure Impact on the Occurrence of Chronic Diseases. *Epidemiology* **4**, 229 (1993).
16. Gail, M. H. *et al.* Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J. Natl. Cancer Inst.* **81**, 1879–1886 (1989).
17. *CDC WONDER Multiple Cause of Death*. https://wonder.cdc.gov/mcd-icd10.html.
18. US Preventive Services Task Force. Medication Use to Reduce Risk of Breast Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA* **322**, 857–867 (2019).
19. Guirguis-Blake, J. M., Beil, T. L., Senger, C. A. & Coppola, E. L. Primary Care Screening for Abdominal Aortic Aneurysm: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* **322**, 2219–2238 (2019).
20. Duffy, S. W. *et al.* Effect of mammographic screening from age 40 years on breast cancer mortality (UK Age trial): final results of a randomised, controlled trial. *Lancet Oncol.* **21**, 1165–1172 (2020).
21. Siu, A. L. & on behalf of the U.S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann. Intern. Med.* **164**, 279–296 (2016).
22. US Preventive Services Task Force. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA* **325**, 1965–1977 (2021).
23. Thompson, I. M. *et al.* Long-Term Survival of Participants in the Prostate Cancer Prevention Trial. *N. Engl. J. Med.* **369**, 603–610 (2013).
24. US Preventive Services Task Force *et al.* Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA* **319**, 1901 (2018).
25. Tsodikov, A. *et al.* Reconciling the Effects of Screening on Prostate Cancer Mortality in the ERSPC and PLCO Trials. *Ann. Intern. Med.* **167**, 449–455 (2017).
26. Jonas, D. E. *et al.* Screening for Prediabetes and Type 2 Diabetes: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* **326**, 744–760 (2021).
27. US Preventive Services Task Force. Screening for Hypertension in Adults: US Preventive Services Task Force Reaffirmation Recommendation Statement. *JAMA* **325**, 1650–1656 (2021).
28. Chou, R. *et al.* Statin Use for the Primary Prevention of Cardiovascular Disease in Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* **328**, 754–771 (2022).
29. Baigent, C. *et al.* Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet Lond. Engl.* **366**, 1267–78 (2005).
30. Fuat, A. *et al.* A polygenic risk score added to a QRISK®2 cardiovascular disease risk calculator demonstrated robust clinical acceptance and clinical utility in the primary care setting. *Eur. J. Prev. Cardiol.* **31**, 716–722 (2024).
31. Jermy, B. *et al.* A unified framework for estimating country-specific cumulative incidence for 18 diseases stratified by polygenic risk. *Nat. Commun.* **15**, 5007 (2024).
32. Weale, M. E. *et al.* Validation of an Integrated Risk Tool, Including Polygenic Risk Score, for Atherosclerotic Cardiovascular Disease in Multiple Ethnicities and Ancestries.

Am. J. Cardiol. **148**, 157–164 (2021).
33. Riveros-Mckay, F. *et al.* Integrated Polygenic Tool Substantially Enhances Coronary Artery Disease Prediction. *Circ. Genomic Precis. Med.* **14**, (2021).
34. Chen, X., Heisser, T., Cardoso, R., Hoffmeister, M. & Brenner, H. Personalized Initial Screening Age for Colorectal Cancer in Individuals at Average Risk. *JAMA Netw. Open* **6**, e2339670 (2023).
35. Widen, E., Lello, L., Raben, T. G., Tellier, L. C. A. M. & Hsu, S. D. H. Polygenic Health Index, General Health, and Pleiotropy: Sibling Analysis and Disease Risk Reduction. *Sci. Rep.* **12**, 18173 (2022).
36. Park, J. K. & Lu, C. Y. Polygenic Scores in the Direct-to-Consumer Setting: Challenges and Opportunities for a New Era in Consumer Genetic Testing. *J. Pers. Med.* **13**, 573 (2023).
37. Beheshti, S. O., Madsen, C. M., Varbo, A. & Nordestgaard, B. G. Worldwide Prevalence of Familial Hypercholesterolemia: Meta-Analyses of 11 Million Subjects. *J. Am. Coll. Cardiol.* **75**, 2553–2566 (2020).
38. Abraham, G. *et al.* Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nat. Commun.* **10**, 5819 (2019).
39. O'Sullivan, J. W. *et al.* Combining Clinical and Polygenic Risk Improves Stroke Prediction Among Individuals With Atrial Fibrillation. *Circ. Genomic Precis. Med.* **14**, e003168 (2021).