Clinical uses of polygenic/genetic risk scores (PRS/GRS)

Andrew Paterson

The Hospital for Sick Children & Dalla Lana School of Public Health, University of Toronto

andrew.paterson@sickkids.ca

IGES PRS Workshop; 12 Oct 2021

Topics

- Efficacy & Safety
- Logistics: Implementation challenges
- Communication

Topics

- **Efficacy & Safety**
- Logistics: Implementation challenges
- Communication

Hippocratic Oath: *primum non nocere*

- First, do no harm
- But … compromise
- Balance harms and benefits

FRAMINGHAM RISK SCORE (FRS)

Estimation of 10-year Cardiovascular Disease (CVD) Risk

Date: _____

Patient's Name: _____

Step 1[1]

In the "points" column enter the appropriate value according to the patient's age, HDL-C, total cholesterol, systolic blood pressure, and if they smoke or have diabetes. Calculate the total points.

Risk Factor		Risk Points				Points
		Men		Women		
Age						
30-34		0		0		
35-39		2		2		
40-44		5		4		
45-49		7		5		
50-54		8		7		
55-59		10		8		
60-64		11		9		
65-69		12		10		
70-74		14		11		
75+		15		12		
HDL-C (mmol/L)						
>1.6		-2		-2		
1.3-1.6		-1		-1		
1.2-1.29		0		0		
0.9-1.19		1		1		
<0.9		2		2		
Total Cholesterol						
<4.1		0		0		
4.1-5.19		1		1		
5.2-6.19		2		3		
6.2-7.2		3		4		
>7.2		4		5		
Systolic Blood Pressure (mmHg)		Not Treated	Treated	Not Treated	Treated	
<120		-2	0	-3	-1	
120-129		0	2	0	2	
130-139		1	3	1	3	
140-149		2	4	2	5	
150-159		2	4	4	6	
160+		3	5	5	7	
Smoker	Yes	4		3		
	No	0		0		
Diabetes	Yes	statin-indicated condition				
	No	0		0		
Total Points						

1 Adapted from: D'Agostino RB et al.(). General cardiovascular risk profile for use in primary care. The Framingham Heart Study. Circ 2008;117:743-53.
2 Adapted from: Genest J et al.(). 2009 Canadian Cardiovascular Society/Canadian guidelines for the diagnosis and treatment of dyslipidemia and prevention of cardiovascular disease in the adult. Can J Cardiol. 2009;25(10):567-579.
3 Adapted from: Anderson T et al.(). 2012 Update of the Canadian Cardiovascular Society guidelines for the diagnosis and treatment of dyslipidemia for the prevention of cardiovascular disease in the adult. Can J Cardiol. 2013;29(2):151-167.

† apoB: apolipoprotein B stat, CVD: cardiovascular disease, FRS: Framingham Risk Score, HDL-C: high-density lipoprotein cholesterol, LDL-C: low-density lipoprotein cholesterol.
* Statins indicated as initial therapy
** Consider LDL-C < 1.8 mmol/L for subjects with acute coronary syndrome (ACS) within past 3 months

Provided courtesy of **Canadian Cardiovascular Society**
Leadership. Knowledge. Community.

Step 2[1]

Using the total points from Step 1, determine the 10-year CVD risk (%).*

Total Points	10-Year CVD Risk (%)*	
	Men	Women
-3 or less	<1	<1
-2	1.1	<1
-1	1.4	1.0
0	1.6	1.2
1	1.9	1.5
2	2.3	1.7
3	2.8	2.0
4	3.3	2.4
5	3.9	2.8
6	4.7	3.3
7	5.6	3.9
8	6.7	4.5
9	7.9	5.3
10	9.4	6.3
11	11.2	7.3
12	13.3	8.6
13	15.6	10.0
14	18.4	11.7
15	21.6	13.7
16	25.3	15.9
17	29.4	18.51
18	>30	21.5
19	>30	24.8
20	>30	27.5
21+	>30	>30

* Double cardiovascular disease risk percentage for individuals between the ages of 30 and 59 without diabetes if the presence of a positive history of premature cardiovascular disease is present in a first-degree relative before 55 years of age for men and before 65 years of age for women. This is known as the modified Framingham Risk Score.[1]

Step 3[1]

Using the total points from Step 1, determine heart age (in years).

Heart Age, *y*	Men	Women
<30	<0	<1
30	0	
31		1
32	1	
34	2	2
36	3	3
38	4	
39		4
40	5	
42	6	5
45	7	6
48	8	7
51	9	8
54	10	
55		9
57	11	
59		10
60	12	
64	13	11
68	14	12
72	15	
73		13
76	16	
79		14
>80	≥17	15+

Step 4[2,3]

Using 10-year CVD risk from Step 2, determine if patient is Low, Moderate or High risk.† Indicate Lipid and/or Apo B targets

Risk Level†	Initiate Treatment If:	Primary Target (LDL-C)	Alternate Target
High FRS ≥20%	• Consider treatment in all *(Strong, High)*	• ≤2 mmol/L or ≥50% decrease in LDL-C *(Strong, Moderate)*	• Apo B ≤0.8 g/L or • Non-HDL-C ≤2.6 mmol/L *(Strong, High)*

Intermediate
FRS 10-19%
	• LDL-C ≥3.5 mmol/L *(Strong, Moderate)*
	• For LDL-C <3.5 mmol/L consider if:
	  • Apo B ≥1.2 g/L
	  • OR Non-HDL-C ≥4.3 mmol/L *(Strong, Moderate)*
	• *Men ≥50 and women ≥60 with 1 risk factor: low HDL-C, impaired fasting glucose, high waist circumference, smoker, hypertension*
		• ≤2 mmol/L or ≥50% decrease in LDL-C *(Strong, Moderate)*	• Apo B ≤0.8 g/L or
			• Non-HDL-C ≤2.6 mmol/L *(Strong, Moderate)*

Low FRS <10%	• statins generally not indicated	• statins generally not indicated	• statins generally not indicated
Statin-indicated conditions**	• *Clinical atherosclerosis** • *Abdominal aortic aneurysm* • *Diabetes mellitus* *Age ≥ 40 years* *15-Year duration for age ≥ 30 years (DM1) Microvascular disease* • *Chronic kidney disease (age ≥ 50 years) eGFR <60 mL/min/1.73 m2 or ACR > 3 mg/mmol*		

Lipid targets LDL-C: _____ or Apo B: _____

PRS Example

- 1371 European subjects with type 1 diabetes
- From DCCT/EDIC study (NEJM 1993)
  - Randomized clinical trial/epidemiological follow-up
  - 27 years of observation since baseline
  - Annual exams

  - Retrospective analysis
    - DNA was collected at end of trial
    - No genetic hypotheses originally proposed (1983-9)
    - Secondary data analysis

  - Outcome: Time from baseline to Coronary Artery Disease (CAD)
  - Predictor: ~6M SNP CAD PRS (Khera et al., Nature Genetic 2018)

Polygenic Risk Scores

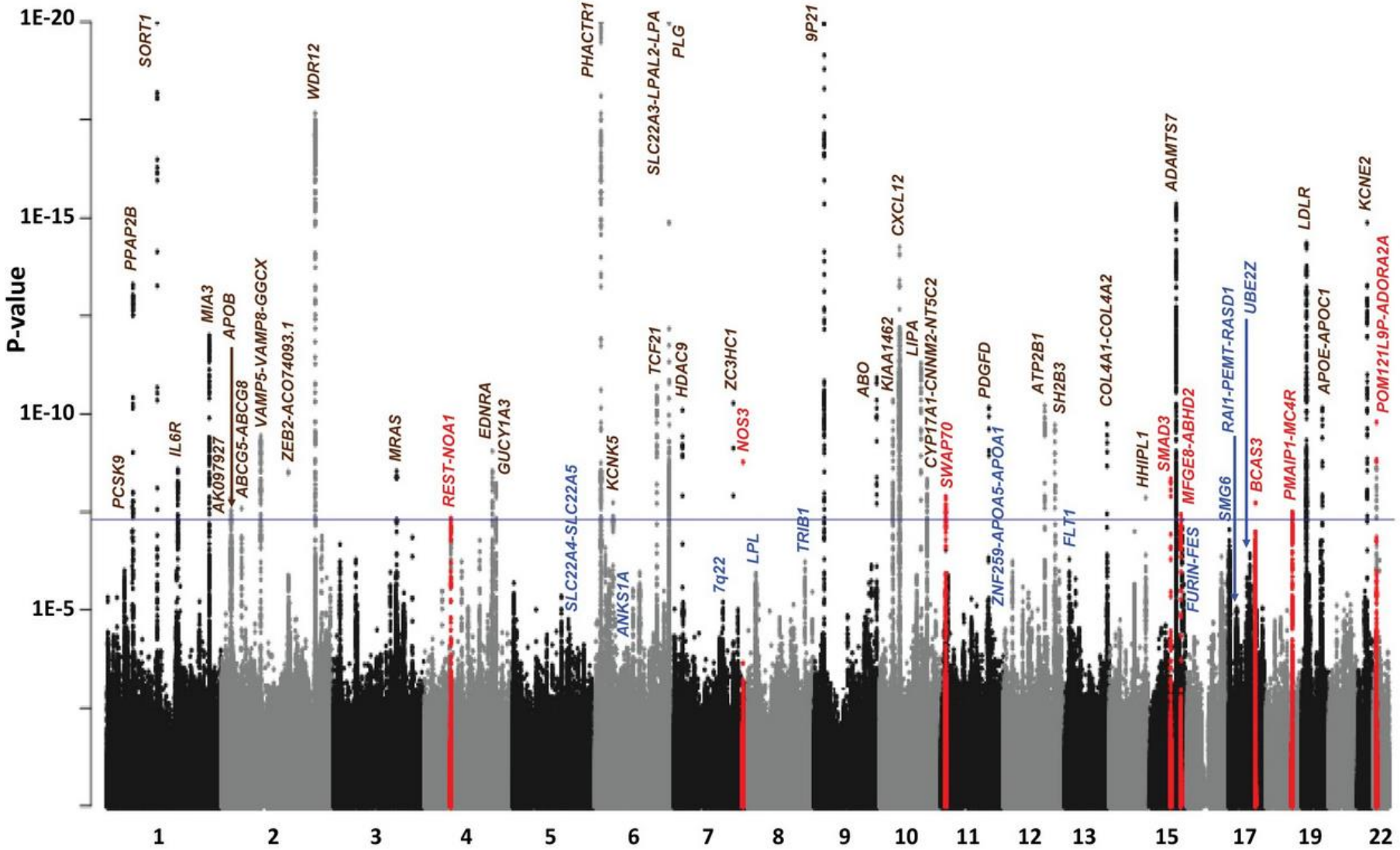
Table 1 | GPS derivation and testing for five common, complex diseases

Disease	Discovery GWAS (*n*)	Prevalence in validation dataset	Prevalence in testing dataset	Polymorphisms in GPS	Tuning parameter	AUC (95% CI) in validation dataset	AUC (95% CI) in testing dataset
CAD	60,801 cases; 123,504 controls[16]	3,963/120,280 (3.4%)	8,676/288,978 (3.0%)	6,630,150	LDPred ($\rho = 0.001$)	0.81 (0.80–0.81)	0.81 (0.81–0.81)
Atrial fibrillation	17,931 cases; 115,142 controls[30]	2,024/120,280 (1.7%)	4,576/288,978 (1.6%)	6,730,541	LDPred ($\rho = 0.003$)	0.77 (0.76–0.78)	0.77 (0.76–0.77)
Type 2 diabetes	26,676 cases; 132,532 controls[31]	2,785/120,280 (2.4%)	5,853/288,978 (2.0%)	6,917,436	LDPred ($\rho = 0.01$)	0.72 (0.72–0.73)	0.73 (0.72–0.73)
Inflammatory bowel disease	12,882 cases; 21,770 controls[32]	1,360/120,280 (1.1%)	3,102/288,978 (1.1%)	6,907,112	LDPred ($\rho = 0.1$)	0.63 (0.62–0.65)	0.63 (0.62–0.64)
Breast cancer	122,977 cases; 105,974 controls[33]	2,576/63,347 (4.1%)	6,586/157,895 (4.2%)	5,218	Pruning and thresholding ($r^2 < 0.2$; $P < 5 \times 10^{-4}$)	0.68 (0.67–0.69)	0.69 (0.68–0.69)

AUC was determined using a logistic regression model adjusted for age, sex, genotyping array, and the first four principal components of ancestry. The breast cancer analysis was restricted to female participants. For the LDPred algorithm, the tuning parameter ρ reflects the proportion of polymorphisms assumed to be causal for the disease. For the pruning and thresholding strategy, r^2 reflects the degree of independence from other variants in the linkage disequilibrium reference panel, and *P* reflects the *P* value noted for a given variant in the discovery GWAS. CI, confidence interval.

Khera et al., 2018 Nature Genetics

Nikpay et al., NG 2015 CAD GWAS



* Y axis was truncated at P=1E-20

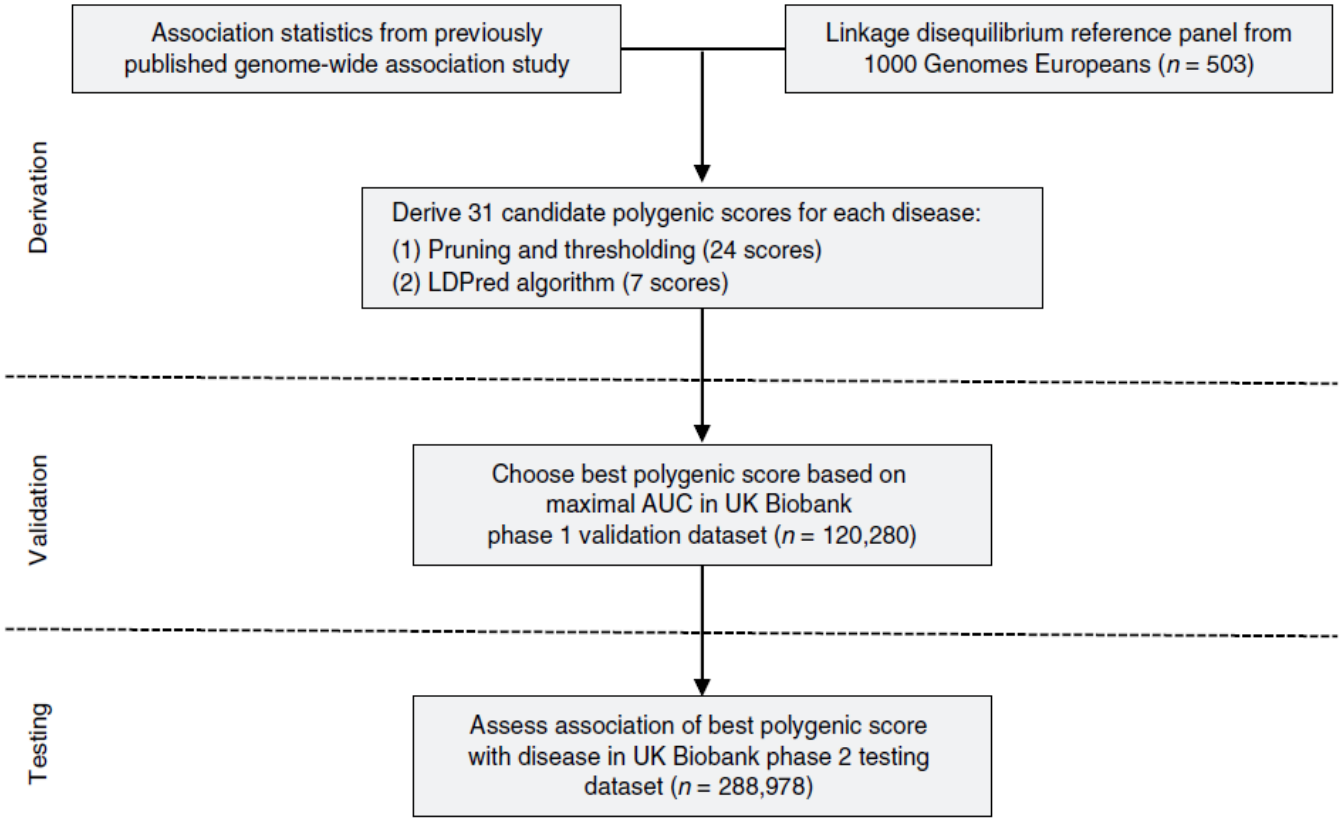


Fig. 1 | Study design and workflow. A GPS for each disease was derived by combining summary association statistics from a recent large GWAS and a linkage disequilibrium reference panel of 503 Europeans[34]. Then, 31 candidate GPSs were derived using two strategies: (1) 'pruning and thresholding' (that is, the aggregation of independent polymorphisms that exceeded a specified level of significance in the discovery GWAS); and (2) the LDPred computational algorithm[13], a Bayesian approach to calculate a posterior mean effect for all variants based on a prior (effect size in the previous GWAS) and subsequent shrinkage based on linkage disequilibrium. The seven candidate LDPred scores vary with respect to the tuning parameter ρ (that is, the proportion of variants assumed to be causal), as previously recommended[13]. The optimal GPS for each disease was chosen based on the AUC in the UK Biobank phase 1 validation dataset (n = 120,280 Europeans) and subsequently calculated in an independent UK Biobank phase 2 testing dataset (n = 288,978 Europeans).

Khera et al., 2018 Nature Genetics

Supplementary Table 1. Association of candidate polygenic scores with prevalent coronary artery disease

Derivation Strategy	Tuning Parameter	N Variants Available / N Variants in Score (%)	OR per SD (95% CI)	AUC
Genome-wide Significant	$p < 5x10^{-8}$ and $r^2 < 0.2$	74/74 (100.0%)	1.39 (1.35-1.44)	0.791
Pruning & Thresholding	$p < 5x10^{-8}$ and $r^2 < 0.4$	100/100 (100.0%)	1.39 (1.35-1.44)	0.791
Pruning & Thresholding	$p < 5x10^{-8}$ and $r^2 < 0.6$	137/137 (100.0%)	1.39 (1.35-1.44)	0.790
Pruning & Thresholding	$p < 5x10^{-8}$ and $r^2 < 0.8$	204/204 (100.0%)	1.37 (1.33-1.42)	0.789
Pruning & Thresholding	$p < 5x10^{-6}$ and $r^2 < 0.2$	192/192 (100.0%)	1.46 (1.42-1.51)	0.794
Pruning & Thresholding	$p < 5x10^{-6}$ and $r^2 < 0.4$	257/257 (100.0%)	1.47 (1.42-1.52)	0.794
Pruning & Thresholding	$p < 5x10^{-6}$ and $r^2 < 0.6$	345/345 (100.0%)	1.45 (1.41-1.50)	0.793
Pruning & Thresholding	$p < 5x10^{-6}$ and $r^2 < 0.8$	505/505 (100.0%)	1.43 (1.38-1.48)	0.792
Pruning & Thresholding	$p < 5x10^{-4}$ and $r^2 < 0.2$	1269/1273 (99.7%)	1.53 (1.48-1.58)	0.797
Pruning & Thresholding	$p < 5x10^{-4}$ and $r^2 < 0.4$	1590/1594 (99.7%)	1.56 (1.51-1.61)	0.798
Pruning & Thresholding	$p < 5x10^{-4}$ and $r^2 < 0.6$	1997/2001 (99.8%)	1.55 (1.50-1.60)	0.797
Pruning & Thresholding	$p < 5x10^{-4}$ and $r^2 < 0.8$	2706/2710 (99.9%)	1.53 (1.48-1.58)	0.797
Pruning & Thresholding	$p < 5x10^{-2}$ and $r^2 < 0.2$	56941/57276 (99.4%)	1.48 (1.44-1.53)	0.794
Pruning & Thresholding	$p < 5x10^{-2}$ and $r^2 < 0.4$	70491/70831 (99.5%)	1.54 (1.49-1.60)	0.797
Pruning & Thresholding	$p < 5x10^{-2}$ and $r^2 < 0.6$	84921/85264 (99.6%)	1.57 (1.52-1.63)	0.798
Pruning & Thresholding	$p < 5x10^{-2}$ and $r^2 < 0.8$	105595/105942 (99.7%)	1.59 (1.54-1.64)	0.799
Pruning & Thresholding	$p < 5x10^{-1}$ and $r^2 < 0.2$	413921/417670 (99.1%)	1.44 (1.39-1.49)	0.792
Pruning & Thresholding	$p < 5x10^{-1}$ and $r^2 < 0.4$	590581/594406 (99.4%)	1.48 (1.43-1.53)	0.794
Pruning & Thresholding	$p < 5x10^{-1}$ and $r^2 < 0.6$	768415/772288 (99.5%)	1.51 (1.46-1.56)	0.795
Pruning & Thresholding	$p < 5x10^{-1}$ and $r^2 < 0.8$	996630/1000544 (99.6%)	1.53 (1.48-1.58)	0.796
Pruning & Thresholding	$p < 1$ and $r^2 < 0.2$	634268/641894 (98.8%)	1.44 (1.39-1.48)	0.792
Pruning & Thresholding	$p < 1$ and $r^2 < 0.4$	973234/981023 (99.2%)	1.48 (1.43-1.52)	0.794
Pruning & Thresholding	$p < 1$ and $r^2 < 0.6$	1349381/1357303 (99.4%)	1.50 (1.46-1.55)	0.795
Pruning & Thresholding	$p < 1$ and $r^2 < 0.8$	1848045/1856048 (99.6%)	1.52 (1.47-1.57)	0.796
LDPred Algorithm	$\rho = 1$	6629369/6630150 (>99.9%)	1.52 (1.47-1.58)	0.796
LDPred Algorithm	$\rho = 0.3$	6629369/6630150 (>99.9%)	1.53 (1.48-1.58)	0.796
LDPred Algorithm	$\rho = 0.1$	6629369/6630150 (>99.9%)	1.54 (1.49-1.59)	0.796
LDPred Algorithm	$\rho = 0.03$	6629369/6630150 (>99.9%)	1.57 (1.52-1.62)	0.798
LDPred Algorithm	$\rho = 0.01$	6629369/6630150 (>99.9%)	1.62 (1.57-1.68)	0.801
LDPred Algorithm	$\rho = 0.003$	6629369/6630150 (>99.9%)	1.69 (1.63-1.75)	0.805
LDPred Algorithm	**$\rho = 0.001$**	**6629369/6630150 (>99.9%)**	**1.72 (1.67-1.78)**	**0.806**

Khera et al., 2018 Nature Genetics

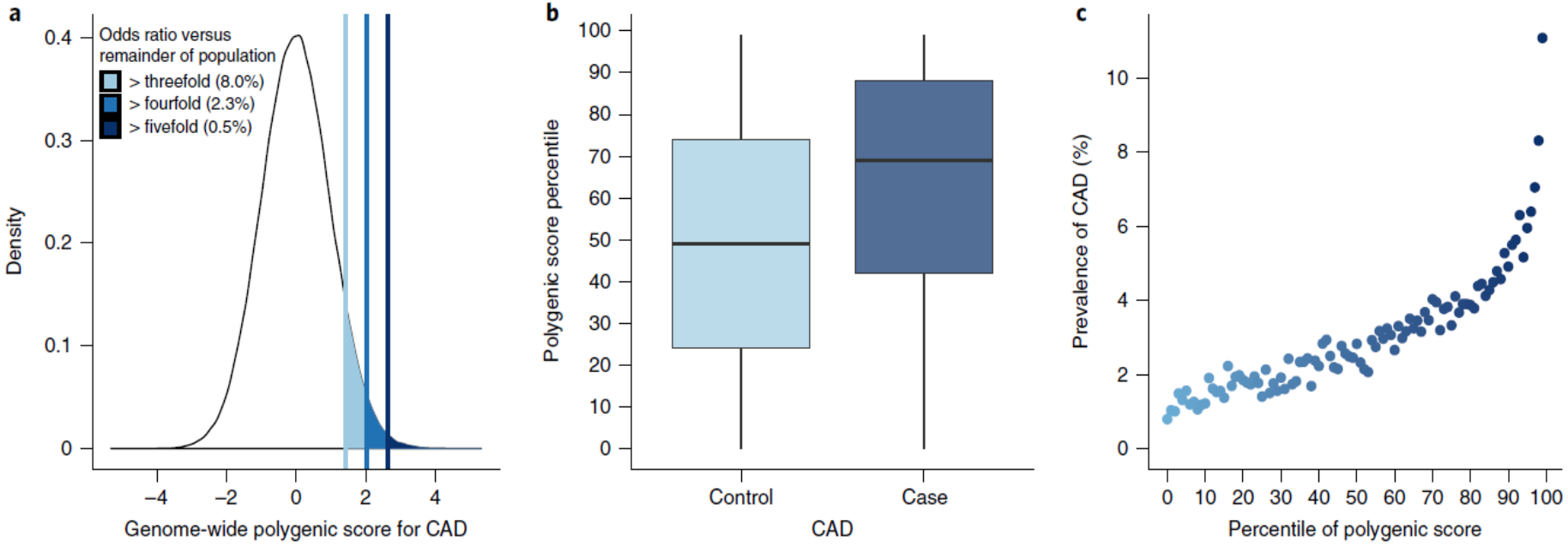


Fig. 2 | Risk for CAD according to GPS. a, Distribution of GPS_{CAD} in the UK Biobank testing dataset ($n = 288{,}978$). The *x* axis represents GPS_{CAD}, with values scaled to a mean of 0 and a standard deviation of 1 to facilitate interpretation. Shading reflects the proportion of the population with three-, four-, and fivefold increased risk versus the remainder of the population. The odds ratio was assessed in a logistic regression model adjusted for age, sex, genotyping array, and the first four principal components of ancestry. **b**, GPS_{CAD} percentile among CAD cases versus controls in the UK Biobank testing dataset. Within each boxplot, the horizontal lines reflect the median, the top and bottom of each box reflect the interquartile range, and the whiskers reflect the maximum and minimum values within each grouping. **c**, Prevalence of CAD according to 100 groups of the testing dataset binned according to the percentile of the GPS_{CAD}.

Khera et al., 2018 Nature Genetics

Managing Diabetes is as simple as ABC:

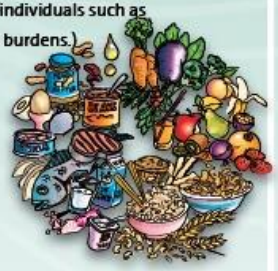
A1C Below 7%

- The A1C test reflects your blood glucose control over the past few months.
- **Work with your doctor to have your own personal A1C goal.**
- Every 1% above 6% elevates the risk for diabetes complications.
- Your blood glucose will stay lower when you exercise and restrict calorie intake.

Blood Pressure Below 140/80

- Up to 60% of people who have diabetes also have high blood pressure.
- High blood pressure can cause heart attack, stroke, and kidney disease.
- Blood pressure stays low when you reduce sodium in your diet and follow the DASH diet.

(Lower systolic targets, such as <130 may be appropriate for certain individuals such as younger patients, if they can be achieved without undue treatment burdens.)



Cholesterol in Check

- The LDL goal for most people is below 100
- The HDL goal for most people is above 40 for men and 50 for women
- Keep trigylcerides lower than 150

- If you have diabetes you are more prone to cholesterol abnormalities and heart disease.
- LDL or "bad" cholesterol can clog your blood vessels and cause heart attack or stroke. Your LDL target may be lower if you have cardiovascular disease; talk to your doctor
- LDL stays low when you eat less saturated fat and cholesterol from animal foods.
- HDL or "good" cholesterol helps remove cholesterol from your blood vessels.
- HDL stays high when you exercise, eat fish and soluble fiber and live a healthy lifestyle.

foodandhealth communications®

For more information, visit diabetes.org and www.nhlbi.nih.gov

Multivariate clinical predictor for Cardiovascular disease (CAD) in Type 1 Diabetes

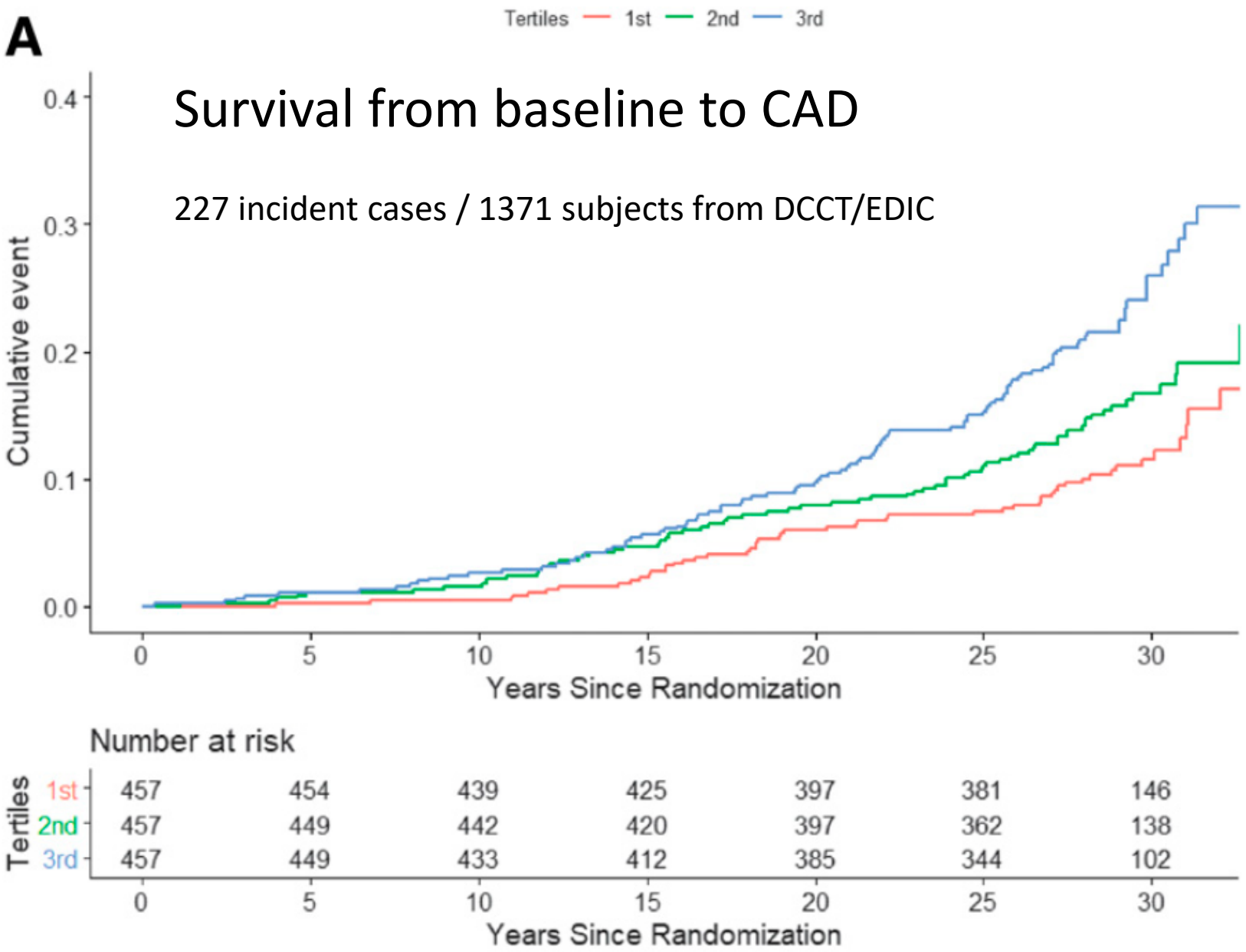
Table 4—The final multivariable Cox models for any-CVD as a function of fixed (baseline) and time-dependent covariates, the latter either the current value or mean from baseline as stated

	HR (95% CI)*	*Z*-test value	*P* value
Any-CVD model covariate			
Baseline age (per 5 years)	1.5366 (1.3641, 1.731)	7.0711	<0.001
Mean HbA_{1c} (per 1%)	1.3115 (1.1488, 1.4972)	4.0133	<0.001
Mean systolic blood pressure (per 10 mmHg)	1.3186 (1.1096, 1.567)	3.1419	0.002
Current triglycerides (log)	1.5536 (1.1688, 2.065)	3.0346	0.003
Mean pulse rate (per 10 bpm)	1.3855 (1.1051, 1.737)	2.8267	0.005
Baseline duration of diabetes (per 5 years)	1.247 (1.0514, 1.4789)	2.5364	0.02
Current use of ACE inhibitor (yes vs. no)	0.6732 (0.4777, 0.9486)	−2.2611	0.03
Baseline family history of MI (yes vs. no)	1.3866 (1.0294, 1.8678)	2.1507	0.04
Mean LDLc (per 10 mg/dL)	1.0721 (1.0037, 1.1451)	2.0697	0.04

DCCT/EDIC, Diabetes 2016; 65: 1370-1379.

Some baseline risk factors, some current

But many others are means over ~27 years prior to event



Khera et al., 2018 Nat Genet CAD PRS

Bebu et al., 2021 Diabetes Care 2021 Apr; dc202388

Association of CAD PRS and CAD risk

CAD PRS	Unadjusted		Adjust Age and HbA1c		Full adjust*	
	HR (95%CI)	p	HR (95%CI)	p	HR (95%CI)	p
Highest 33%	1.8 (1.4-2.4)	6E-6	1.8 (1.4-2.3)	1E-5	1.7 (1.3-2.3)	4E-5
Per SD increase	1.4 (1.2-1.6)	1E-7	1.4 (1.2-1.6)	9E-8	1.4 (1.2-1.6)	2E-6

* age; time-weighted HbA1c, SBP, pulse, LDLc; triglycerides, family history, duration of T1D, ACE inhibitor use

Bebu et al., 2021 Diabetes Care 2021 Apr; dc202388

Table 3—Association of clinical risk and genetic factors with subsequent risk of CVD and MACE in separate multivariable Cox proportional hazards models

Covariate	HR (95% CI)	*z* score	*P*
CVD (Akaike IC 2,909.790)			
Age	1.09 (1.06, 1.11)	7.6014	**2.9E−14**
Mean updated HbA_{1c}	1.39 (1.22, 1.58)	4.9452	**7.6E−7**
Log triglycerides	1.75 (1.37, 2.25)	4.4327	**9.3E−6**
CAD PRS	1.34 (1.18, 1.53)	4.3637	**1.3E−5**
Mean updated SBP	1.03 (1.01, 1.05)	3.3602	**7.8E−4**
Mean updated pulse	1.02 (1.00, 1.05)	2.2041	**2.8E−2**
Mean updated LDLc	1.01 (1.00, 1.01)	1.9127	5.6E−2
Duration of T1D	1.00 (1.00, 1.00)	1.7542	7.9E−2
Family history of MI	1.26 (0.96, 1.66)	1.6843	9.2E−2
Any ACE inhibitor use	0.78 (0.58, 1.05)	1.6373	1.0E−1

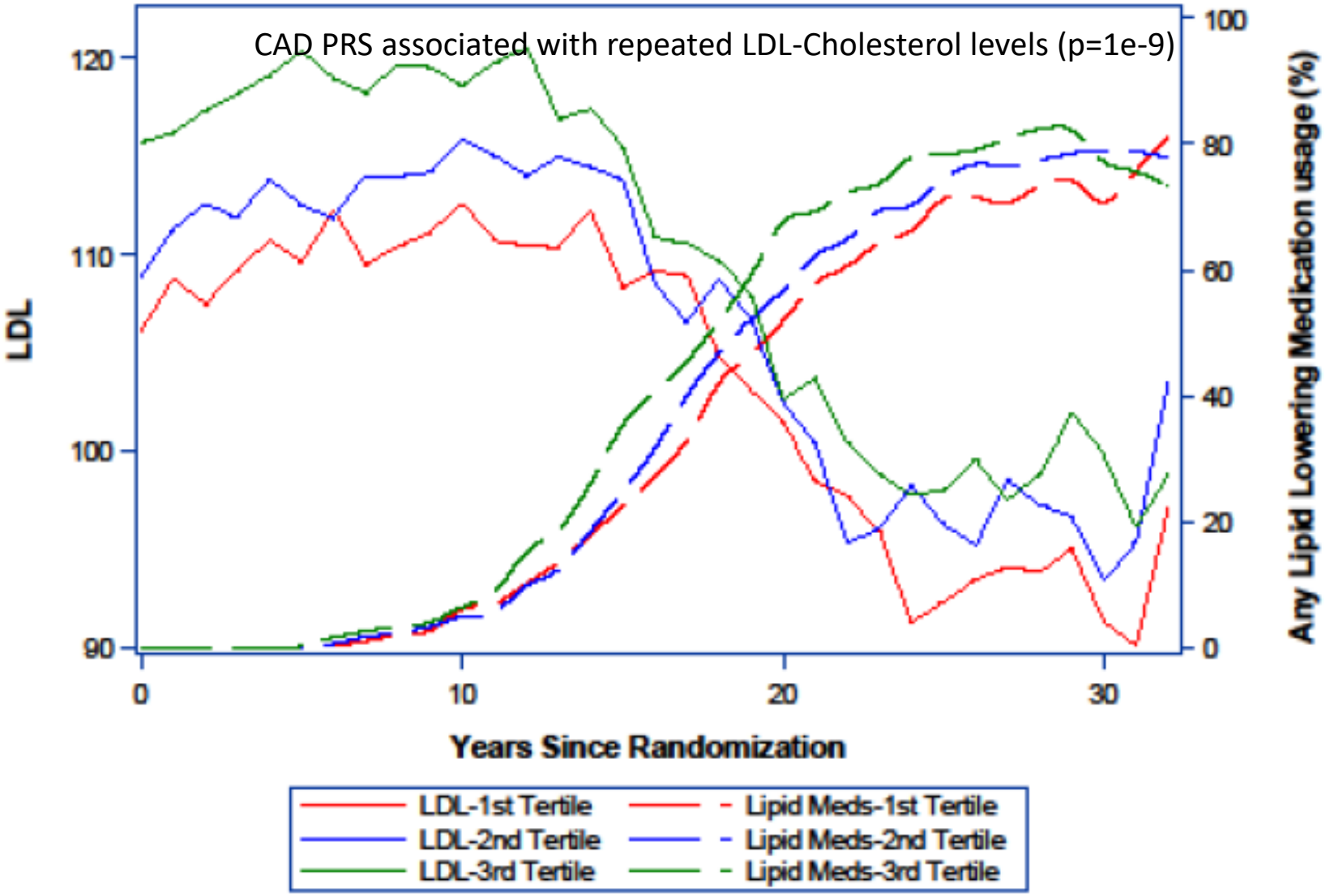
…

AUC: clinical model =0.698
AUC: clinical model + CAD PRS = 0.702

But do we typically have repeated HbA1c, BP, pulse, LDL-C over 27 years?

Bebu et al., 2021 Diabetes Care 2021 Apr; dc202388

LDLc levels (solid) and % of statin treatment (dash) by CAD PRS tertiles



Some of CAD PRS effect is mediated through effect on LDL-C (or factors correlated with it)

Bebu et al., 2021 Diabetes Care 2021 Apr; dc202388


A B C D E F
G H I J K L
M N O P Q R
S T U V W
X Y Z

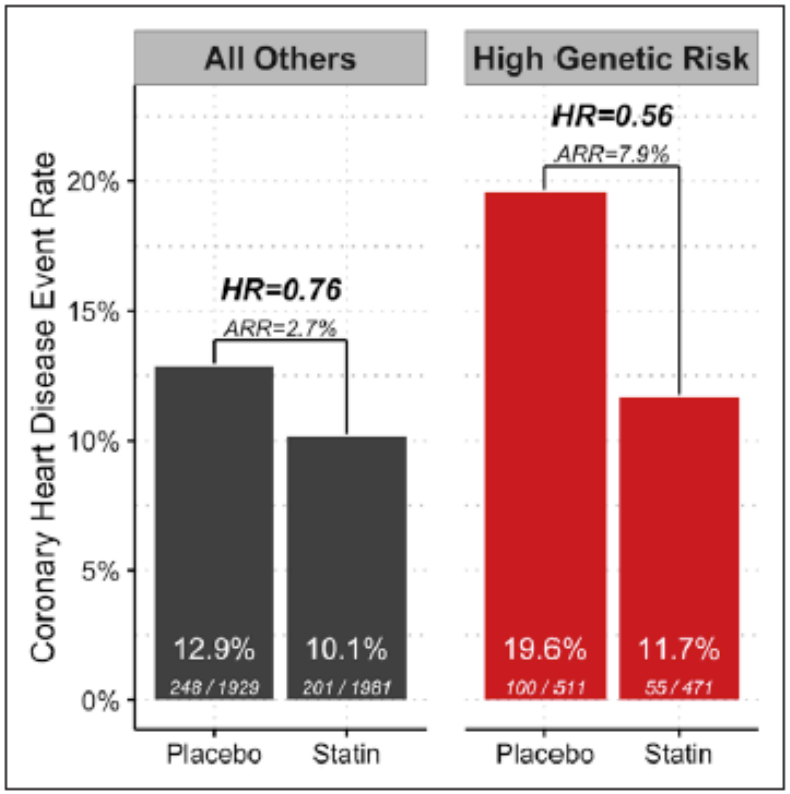


Figure 1. Incident coronary heart disease events by statin therapy and genetic risk group in WOSCOPS (West of Scotland Coronary Prevention Study).
Nonfatal myocardial infarction or death resulting from coronary heart disease rate is shown by randomized treatment group and polygenic risk group in WOSCOPS. Absolute events (and percentage) per individual in each group are shown at the bottom of the bars. This represents 604 events over 64031 total patient-years of follow-up. The follow-up period within the trial was 4.8 years (SD, 0.7 years) for both the placebo and statin groups and out of the trial was 8.1 years (SD, 3.4 years) for the placebo group and 8.4 years (SD, 3.0 years) for the statin-treated group. ARR indicates adjusted relative risk; and HR, hazard ratio.

Circulation. 2017;135:2091–2101. DOI: 10.1161/CIRCULATIONAHA.116.024436

Primary prevention in subjects with hypercholesterolemia

57 SNP CHD weighted GRS

Top GRS quintile vs the rest

Natarajan et al., Circ 2017

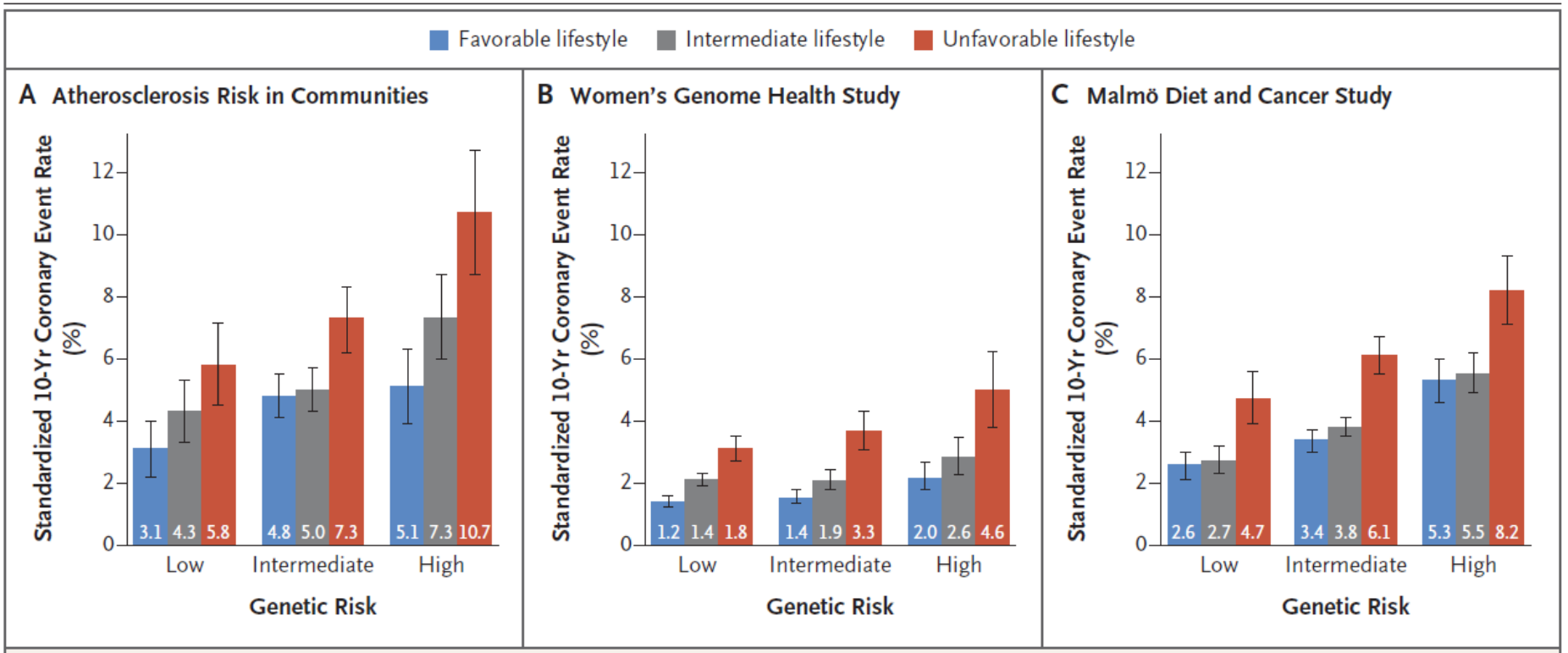


Figure 3. 10-Year Coronary Event Rates, According to Lifestyle and Genetic Risk in the Prospective Cohorts.

Shown are standardized 10-year cumulative incidence rates for coronary events in the three prospective cohorts, according to lifestyle and genetic risk. Standardization was performed to cohort-specific population averages for each covariate. The I bars represent 95% confidence intervals.

'Lifestyle' factors : smoking, obesity, physical activity, diet; simple sum

Khera et al., NEJM 2016

Does PRS make any change to clinical recommendations?

- Any difference from the standard of care
- Need trials/studies to examine GRS-specific efficacy/safety

Topics

- Efficacy & Safety
- **Logistics: Implementation challenges**
- Communication

Implementation challenges

- Most clinical molecular genetics labs not trained / set-up to report GWAS array results
  - Single rare coding variants
- Next Generation Sequencing (NGS) data processing/format
  - Single-sample VCF
    - Variants that differ from a specific reference sequence
      - Do not tell us genotype at all variants in the genome
        - Cannot differentiate between homozygous ref and missing
- Don’t impute
  - Imputation servers may not meet clinical criteria
    - Errors
    - downtime
- Don’t have GWAS expertise
  - Which study to use?
    - Which GRS (P value, LD threshold, method, etc.)?
- Genetic counsellors overwhelmed and not trained in PRS

Topics

- Efficacy & Safety
- Logistics: Implementation challenges
- **Communication**

Communication science

Best evidence for conveying complex risk information to general population

The New York Times

Yes, Many of Us Are Stress-Eating and Gaining Weight in the Pandemic

A global study confirms that during the pandemic, many of us ate more junk food, exercised less, were more anxious and got less sleep.

Lorenzo Gritti

By Anahad O'Connor

Published Dec. 4, 2020 Updated Dec. 10, 2020

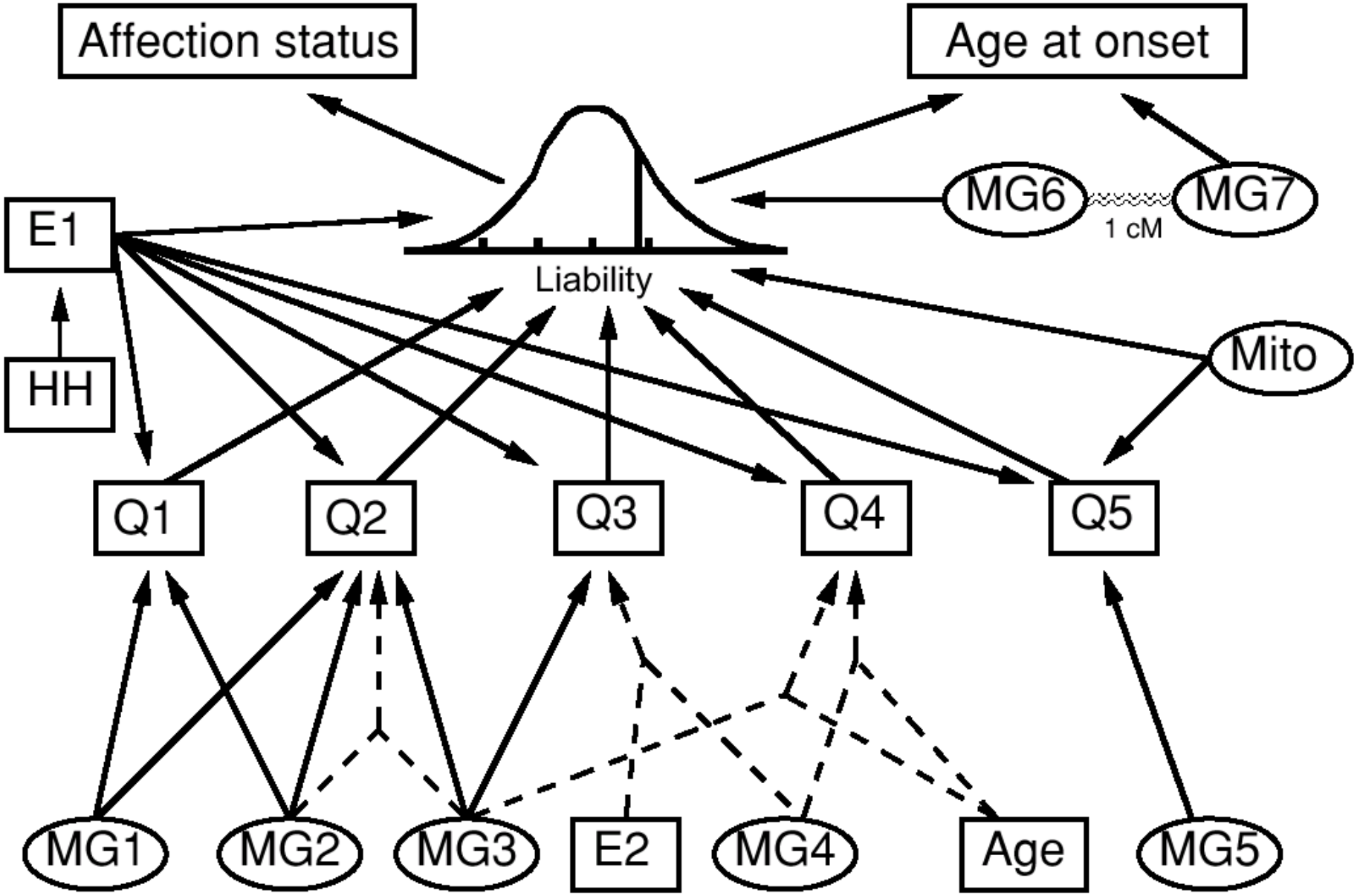
Genetic Weight

Your genes influence not just your weight, but also the impact of different healthy habits.

Overview | Scientific Details

Andrew, your genes predispose you to weigh about 3% less than average.

This predisposition doesn't mean you will definitely weigh less than average. Keep in mind that your lifestyle and environment have a big impact on your weight.



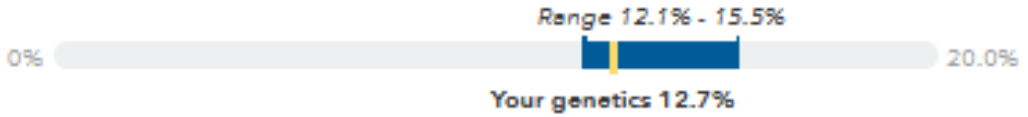
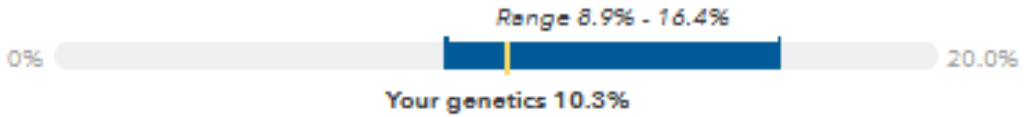
From Genetic Analysis Workshop 12
Almasy et al., Genetic Epi 2001; 21S; S332

Genetics and Lifestyle Associations at 23andMe

Your genetics can actually influence how much lifestyle impacts your weight, which is called "gene-environment interaction."

We looked for these kinds of interactions by comparing the BMIs of 23andMe research participants with different genetics and different daily habits. In general, we saw the biggest weight differences between people who practiced these habits most often compared to those who rarely or never did. Each lifestyle choice seemed to have a slightly different effect on weight, depending on genetics. This table shows the average effect associated with your genetic weight predisposition as well as the range of effect seen in people with other predispositions.

Uncovering the connections between genetics, lifestyle, and weight is an active area of science, and our research efforts are ongoing.

Lifestyle factor	Range of observed results (% weight difference), from lowest to highest genetic weight predisposition
Limiting red meat	
Avoiding fast food	
Sleeping a healthy amount	
Exercising	

25-page white paper

White Paper 23-17

The science behind 23andMe's Genetic Weight report

Estimating BMI and associated phenotypes with polygenic risk models

Authors: Michael L. Multhaup, Alisa P. Lehman, Bertram L. Koelsch, Alison Chubb, Robin P. Smith, Shirley Wu and Nicholas A. Furlotte

Created: February 2017

1. Introduction

More than a million 23andMe customers have consented to participate in research. Their contributions not only have led to more than 60 scientific publications but have also allowed our scientists to develop unique and innovative products for the 23andMe® Personal Genetic Service. We previously published a white paper describing our general approach in creating predictive models for categorical traits with a limited number of discrete outcomes such as hair color and cheek dimples[1]. Here, we extend this approach to model body mass index (BMI), a quantitative trait with a continuous numeric outcome, and detail appropriate metrics for the validation of models predicting quantitative outcomes. We also extend these methods of model creation to non-European populations. The genetic models described in this study use more than 300 single nucleotide polymorphisms (SNPs) to predict BMI based on data from more than 600,000 research participants. The out of sample variance in BMI explained by the purely genetic models ranged from 1.8% to 4.3%, depending on the ethnicity of the cohort. Finally, we present our analysis of the interaction between the BMI genetic risk scores and various lifestyle phenotypes. These sets of information are translated for use in the 23andMe Genetic Weight report.

Table 1: Model training cohorts and definitions

Name	Training Cohort	Linear regression model formula	Equation
GRS	Ancestry-specific	$BMI \sim \sum_{i=0}^{n}(\beta_i * SNP\ allelic\ dosage_i) + \beta_{intercept}$	Equation 1
Population Stratification	Ancestry-specific	$BMI \sim \sum_{i=0}^{4}(\beta_i * PC_i) + \beta_5 * GRS + \beta_6 * age + \beta_7 * sex + \beta_{intercept}$	Equation 2
Main result	Ancestry- and sex-specific	$BMI \sim \beta_0 * GRS + \beta_1 * age + \beta_2 * age^2 + \beta_{intercept}$	Equation 3
Phenotype GxE (discovery)	European	$BMI \sim \beta_0 * GRS + \beta_1 * phenotype + \beta_2 * phenotype : GRS + \beta_{intercept}$	Equation 4
Phenotype GxE (prediction)	European	$BMI \sim \beta_0 * GRS + \beta_1 * phenotype + beta_2 * phenotype^2 + \beta_3 * phenotype : GRS + \beta_{intercept}$	Equation 5

Topics

- Efficacy & Safety
- Logistics: Implementation challenges
- Communication

Discussion period