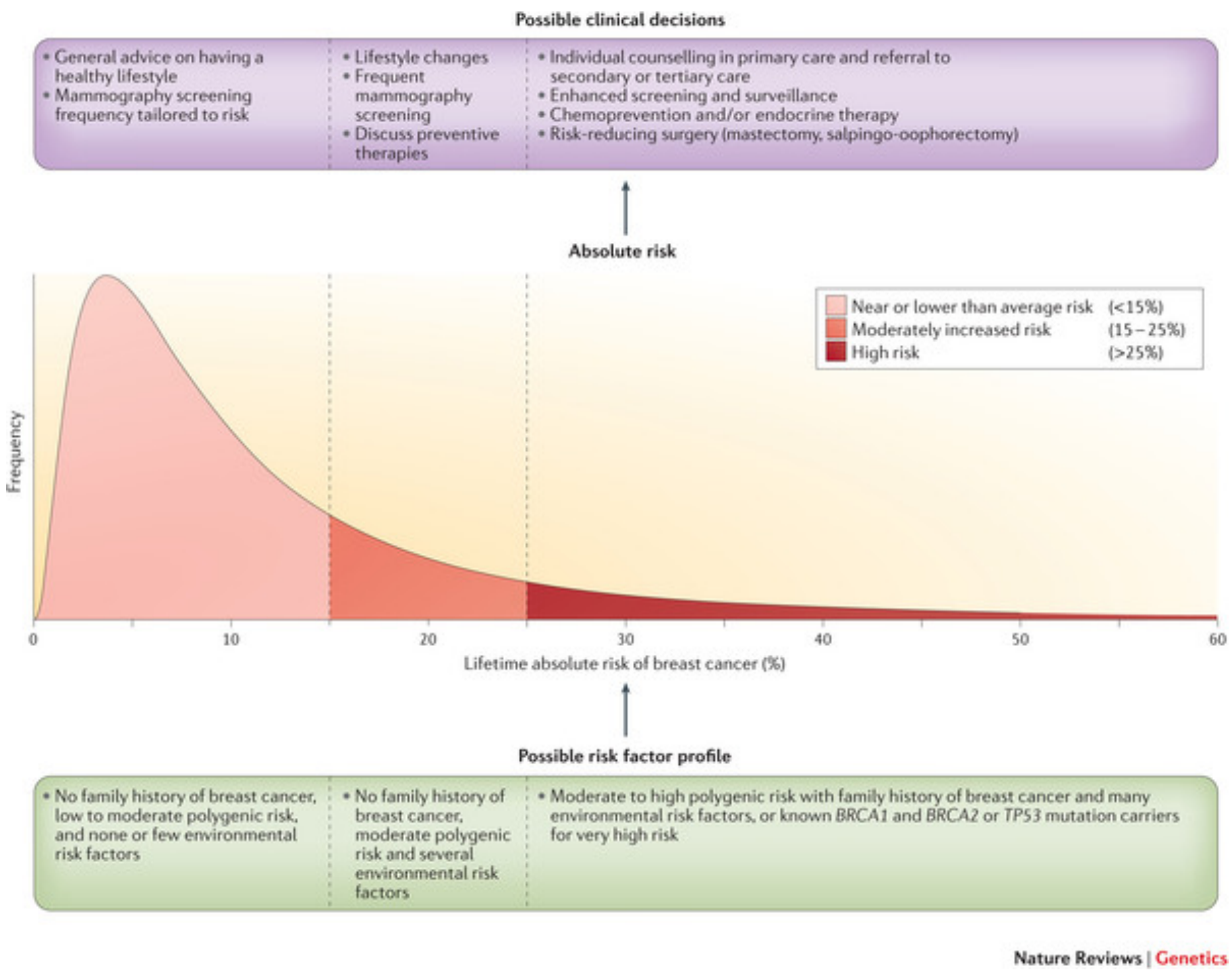
Risk prediction

1. Can we identify groups in the population that exhibit high risk?

Application: Screening

2. Can we estimate the risk for a single patient?

Application: Prevention



Chatterjee, Nat Rev Genetics 2016

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Cumulative Association of Five Genetic Variants with Prostate Cancer

No. of associated factors**						
0	144 (5.0)	174 (10.1)	NA	1.00		
1	778 (26.9)	581 (33.6)	0.48	1.62 (1.27–2.08)	1.27×10^{-4}	
2	1053 (36.4)	622 (36.0)	0.73	2.07 (1.62–2.64)	5.86×10^{-9}	
3	642 (22.2)	286 (16.6)	0.99	2.71 (2.08–3.53)	9.54×10^{-14}	
4	236 (8.2)	60 (3.5)	1.56	4.76 (3.31–6.84)	9.17×10^{-19}	
≥5	40 (1.4)	5 (0.3)	2.24	9.46 (3.62–24.72)	1.29×10^{-8}	4.78×10^{-28}

"A patent application has been filed by the Wake Forest University School of Medicine, Johns Hopkins University School of Medicine, and Dr. Henrik Grönberg at Karolinska Institutet, Stockholm, to preserve patent rights for the technology and results described in this study"

Zheng SL et al. N Engl J Med 2008

Risk score based on genome-wide significant SNPs

- Your genetic risk score (GRS) is a continuous variable.
- Two main approaches: Unweighted scores and weighted score
- Unweighted score in individual *i* for *m* SNPs: add up number of alleles for each individual

$$GRS_i = \sum_{j=1}^{m} G_{ij}$$

- Weighted score in individual *i* for *m* SNPs: multiply number of alleles for each SNP with *published* effect sizes for each individual

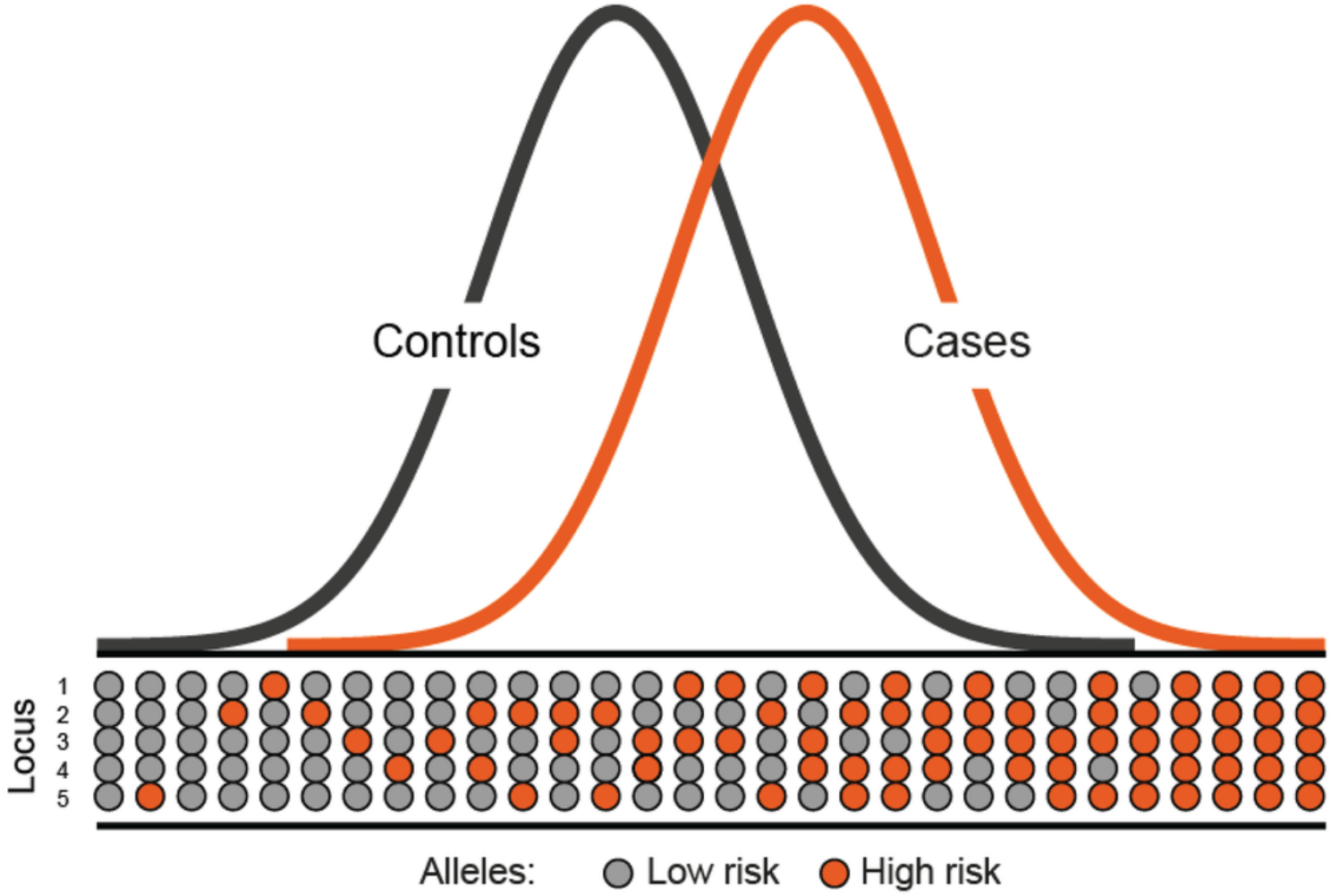
$$GRS_i = \sum_{j=1}^{m} \beta_{ij} G_{ij}$$

Generating a genetic risk score

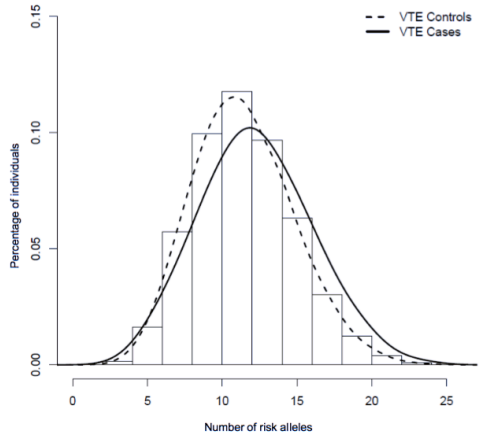
- If you are using a weighted score, do not use βs from your own data -> model overfitting

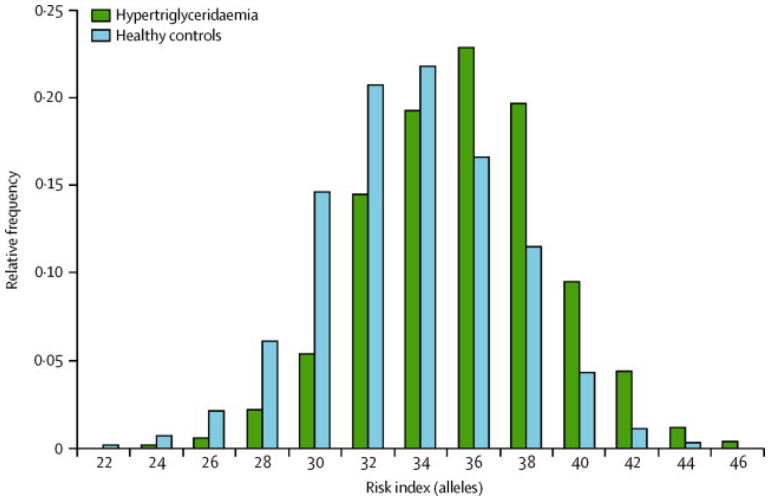
- Need to handle missing data
  - Complete case analysis (remove all samples with ≥1 SNP missing)
  - Impute
    - LD (do not always have this information, e.g. only GRS SNPs were genotyped)
    - Expected value based on allele frequency (PLINK)
    - Sampling from your data conditioned on some variables (case-control status, age)

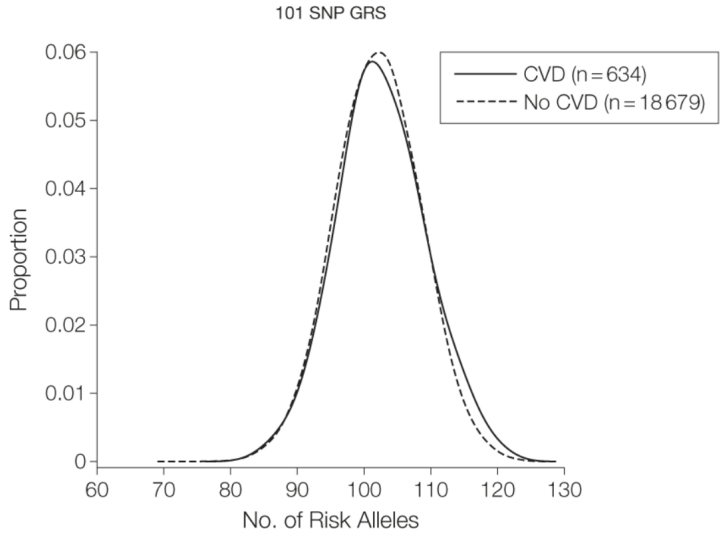
Distribution of genetic risk scores (GRS)

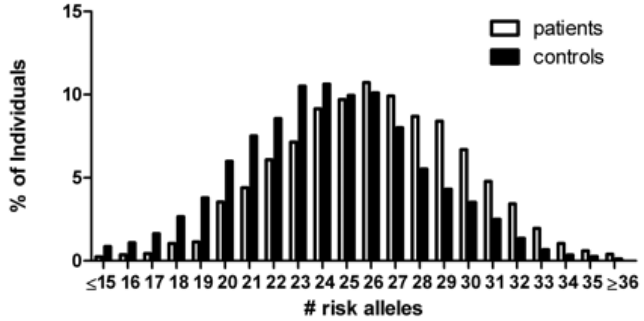


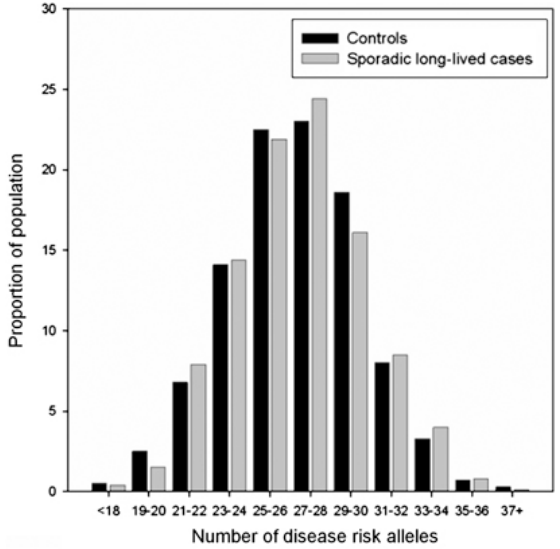
Distribution of GRS for complex diseases

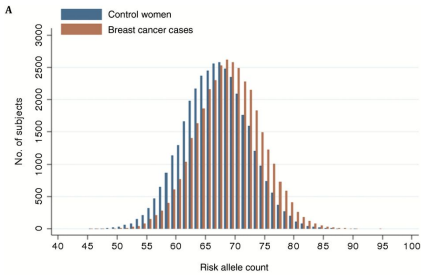


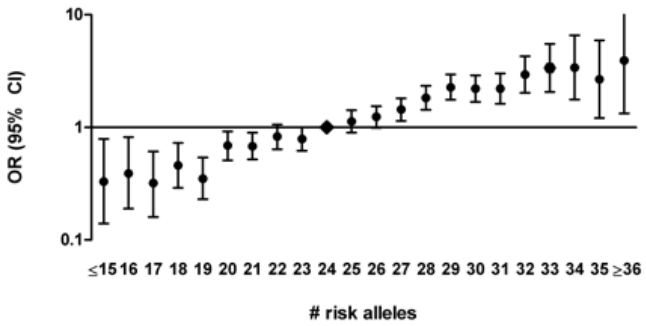


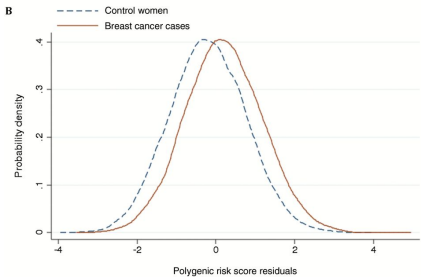




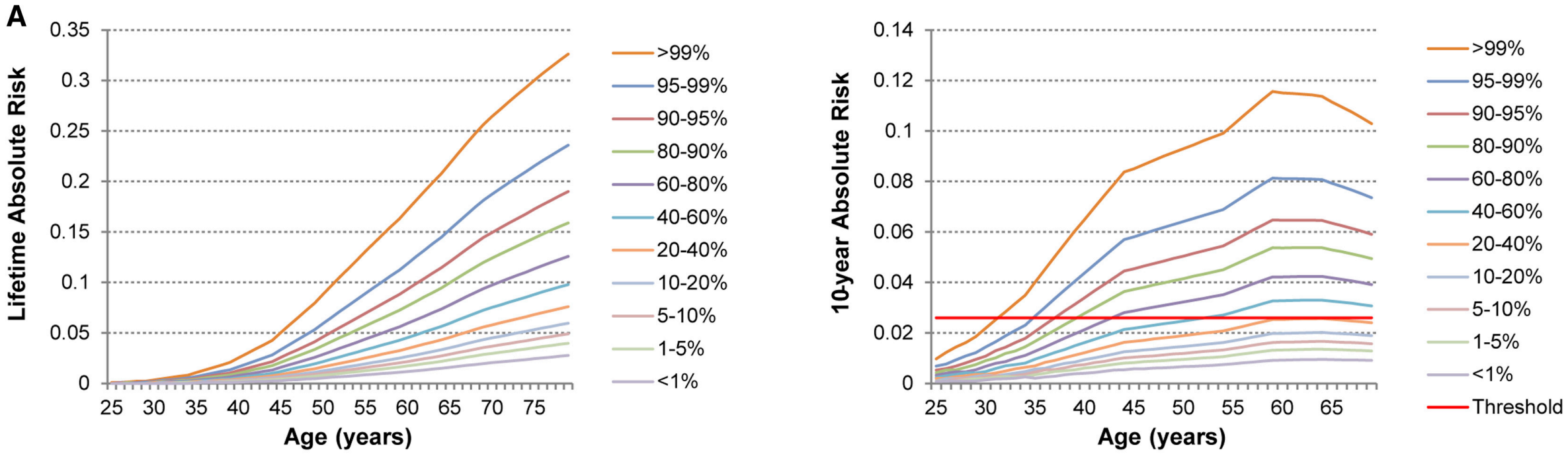








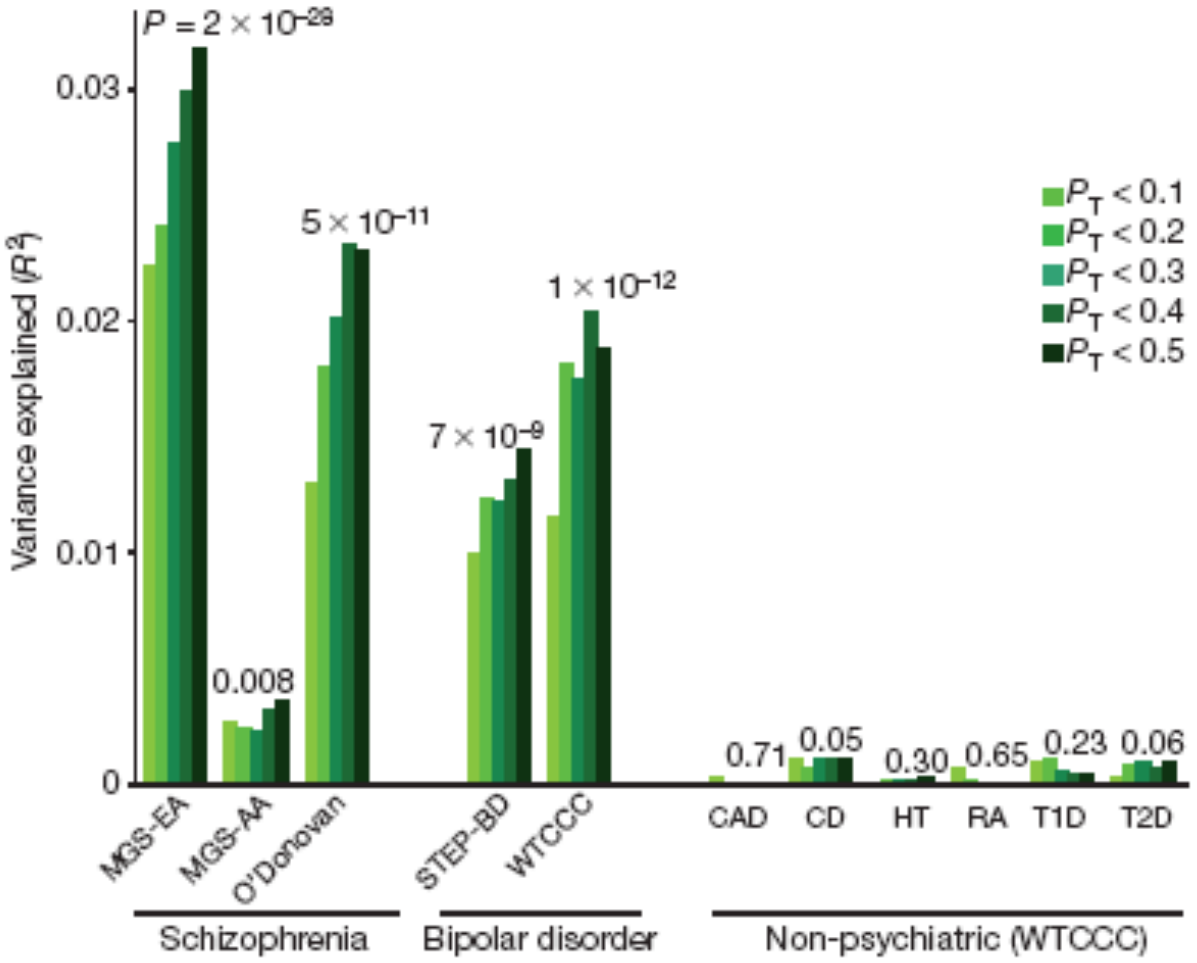
Association between a genetic risk score and breast cancer risk in women of European origin



"The average 10-year absolute risk of breast cancer for a 47-year-old woman (the age at which women become eligible to enter the UK breast cancer screening program) in the general population is 2.6%. However, the 19% of women with highest genetic risk will attain this level of risk by age 40 years"

Mavaddat et al. AJHG 2019

Going beyond genome-wide significant SNPs

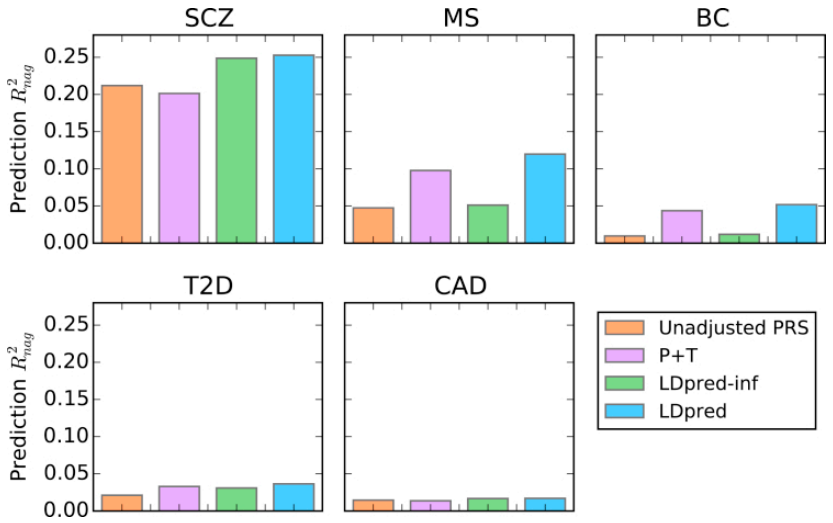


Purcell, Nature 2009

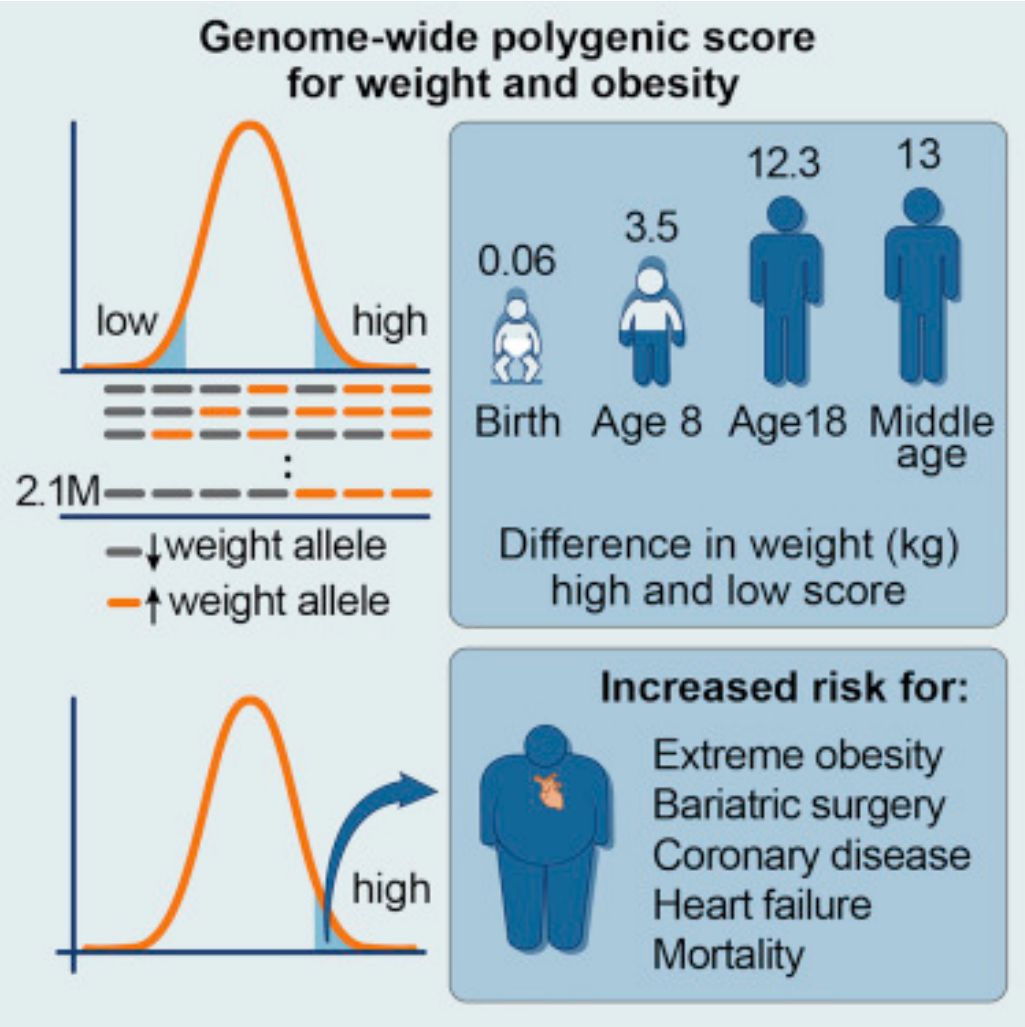
Genome-wide polygenic risk scores

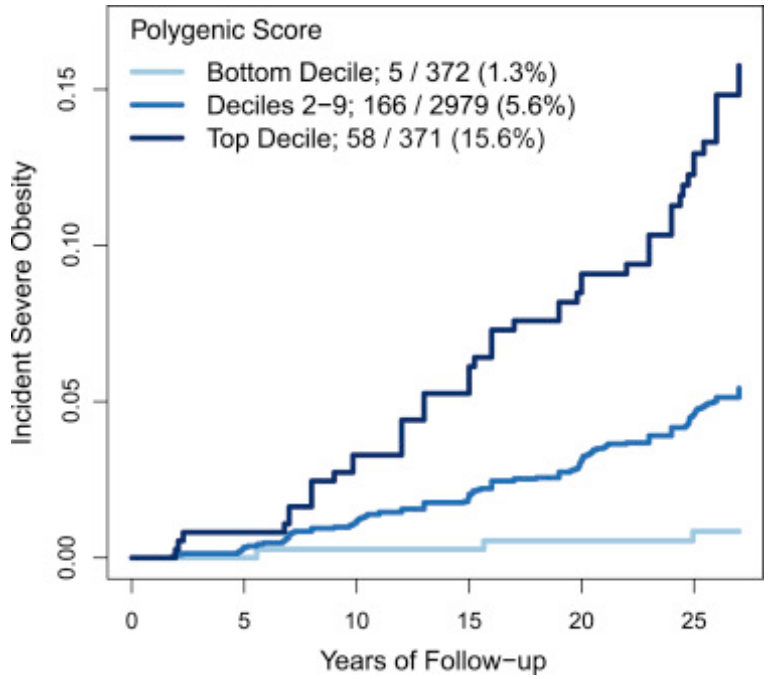
- Common approach: Prune SNPs based on LD:
  1. Rank SNPs based on p-value
  2. Going down the list of SNPs, Remove any SNPs that are in LD with a previous SNP and has higher p-value
  - Will result in loss of power

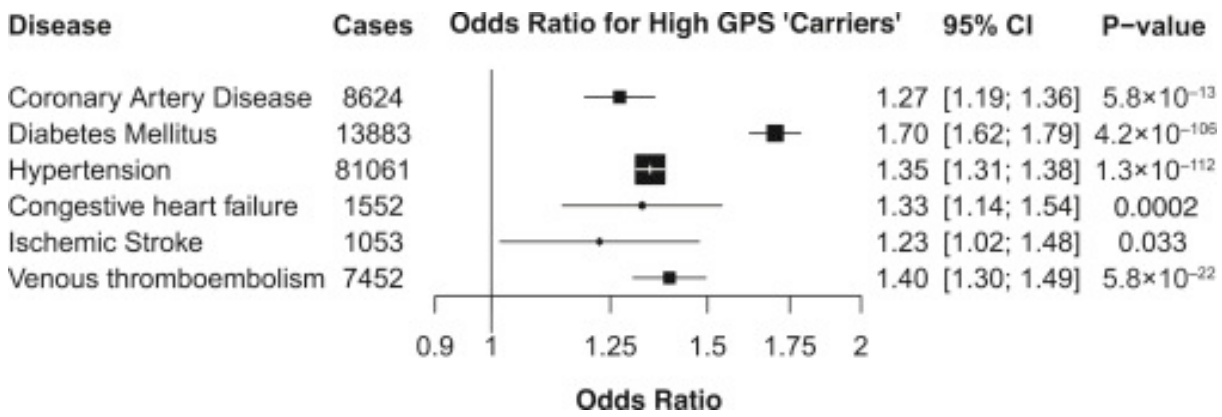
- Newer methods (e.g. Vilhjalmsson, AJHG 2015) incorporates LD in the score leading to improved accuracy



Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood

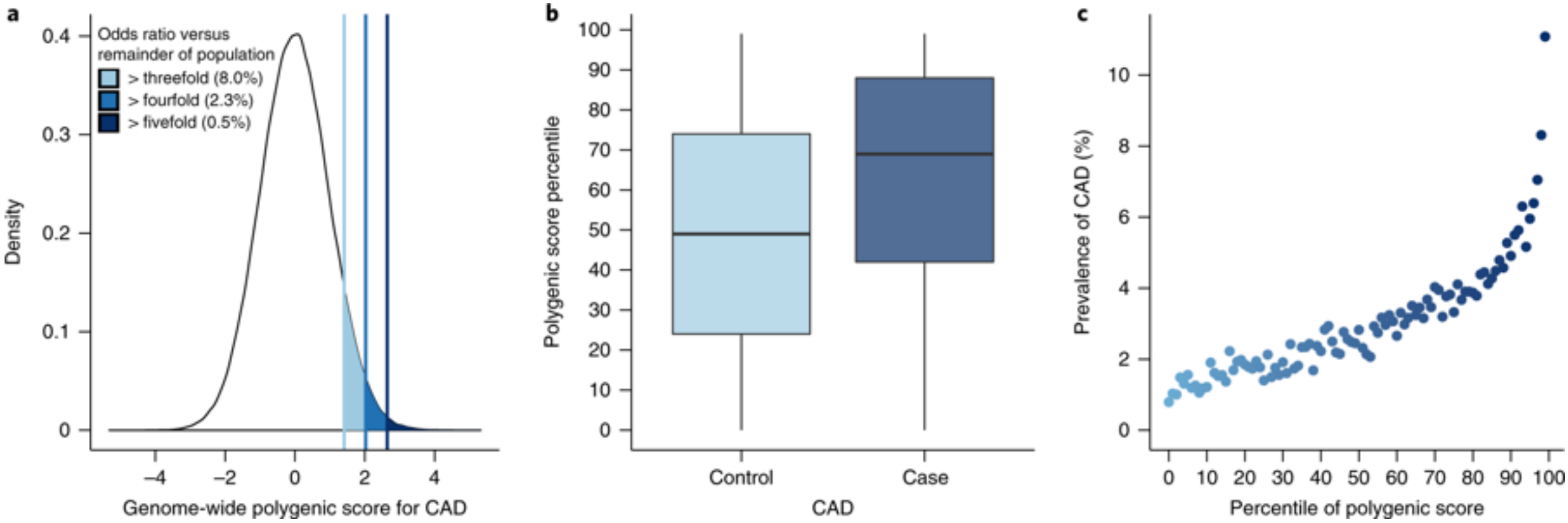






Khera, Cell 2019

Risk for coronary artery disease according to genome-wide polygenic score.



(**a**) Distribution of genome-wide polygenic score for CAD (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 standard deviation of 1 to facilitate interpretation. Shading reflects proportion of population with 3, 4, and 5-fold increased risk versus remainder of the population. Odds ratio assessed in a logistic regression model adjusted for age, sex, genotyping array, and the first four PCs; (**b**) GPS_{CAD} percentile among CAD cases versus controls in the UK biobank validation cohort. Within each boxplot, the horizontal lines reflect the median, the top and bottom of the box reflects the interquartile range, and the whiskers reflect the maximum and minimum value within each grouping; (**c**) prevalence of CAD according to 100 groups of the validation cohort binned according to percentile of the GPS_{CAD}.

Khera, Nat Genet 2018

US 20190017119A1

(19) **United States**

(12) **Patent Application Publication**

Khera et al.

(10) **Pub. No.: US 2019/0017119 A1**

(43) **Pub. Date: Jan. 17, 2019**

(54) **GENETIC RISK PREDICTOR**

(71) Applicant: **THE GENERAL HOSPITAL CORPORATION**, Boston, MA (US)

(72) Inventors: **Amit Khera**, Boston, MA (US); **Derek Klarin**, Boston, MA (US); **Sekar Kathiresan**, Boston, MA (US)

(21) Appl. No.: **16/034,260**

(22) Filed: **Jul. 12, 2018**

Related U.S. Application Data

(60) Provisional application No. 62/531,762, filed on Jul. 12, 2017, provisional application No. 62/583,997, filed on Nov. 9, 2017, provisional application No. 62/585,378, filed on Nov. 13, 2017.

Publication Classification

(51) **Int. Cl.**

C12Q 1/6883 (2006.01)

C12Q 1/6886 (2006.01)

(52) **U.S. Cl.**

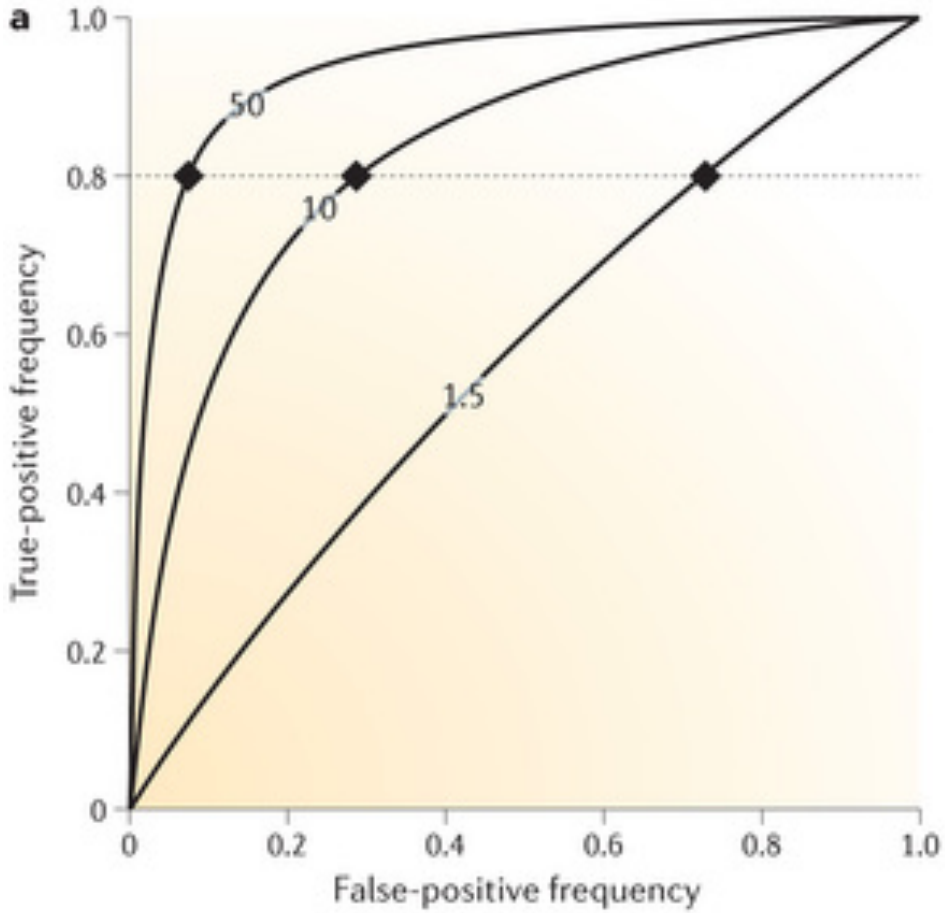
CPC ***C12Q 1/6883*** (2013.01); *C12Q 2600/156* (2013.01); ***C12Q 1/6886*** (2013.01)

(57) **ABSTRACT**

The present disclosure relates to a method of determining a risk of developing coronary artery disease in a subject, the method comprising identifying whether at least 95 single nucleotide polymorphisms (SNPs) from Table D is present in a biological sample from the subject, wherein the presence of a risk allele of a SNP from Table D indicates that the subject has an increased risk of coronary artery disease, and wherein the presence of an alternative allele indicates that the subject has a decreased risk of coronary artery disease.

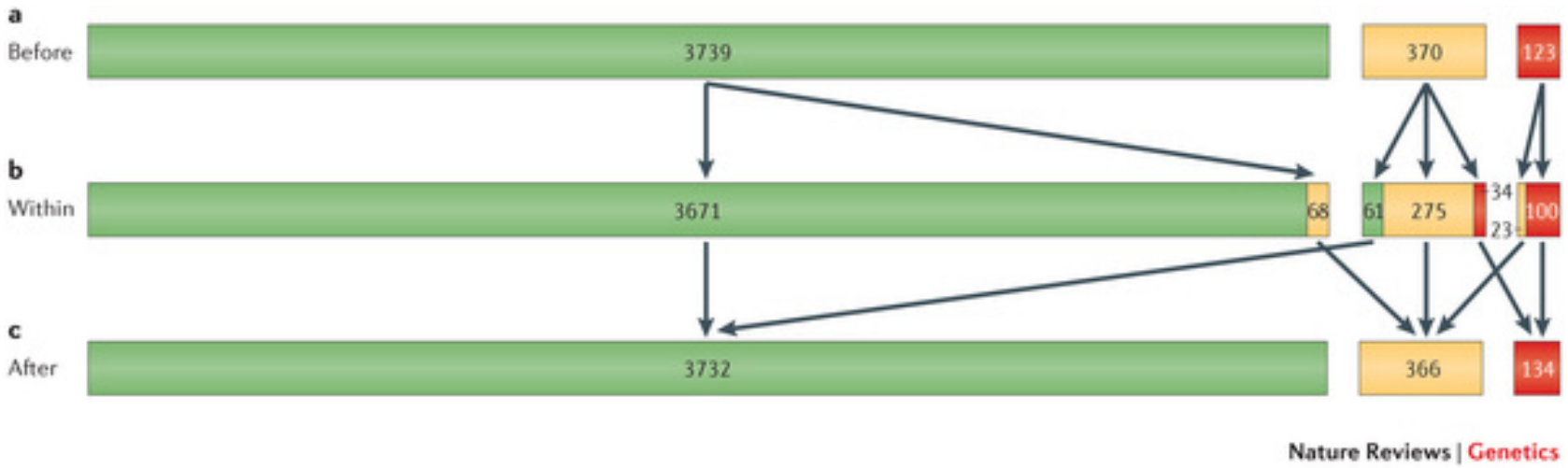
Measures of risk prediction performance (i)

- Area under the receiver operator characteristic (ROC) curve
  - The ROC curve plots the true-positive fraction (sensitivity) against the false-positive fractions (1-specificity)
  - Ranges from 0.5 (no discrimination between cases and controls) to 1.0 (perfect discrimination)



Measures of risk prediction performance (ii)

- Reclassification based on genetic risk scores



A cohort of 4,232 people was classified into low (<10%; green), medium (>10–<20%; yellow) and high (>20%; red) 10-year risk of cardiovascular disease before and after applying genotype risk score.

a | Before incorporating genotype score (standard risk factors)

b | Reclassification based on genotypes

c | After incorporating genotype score

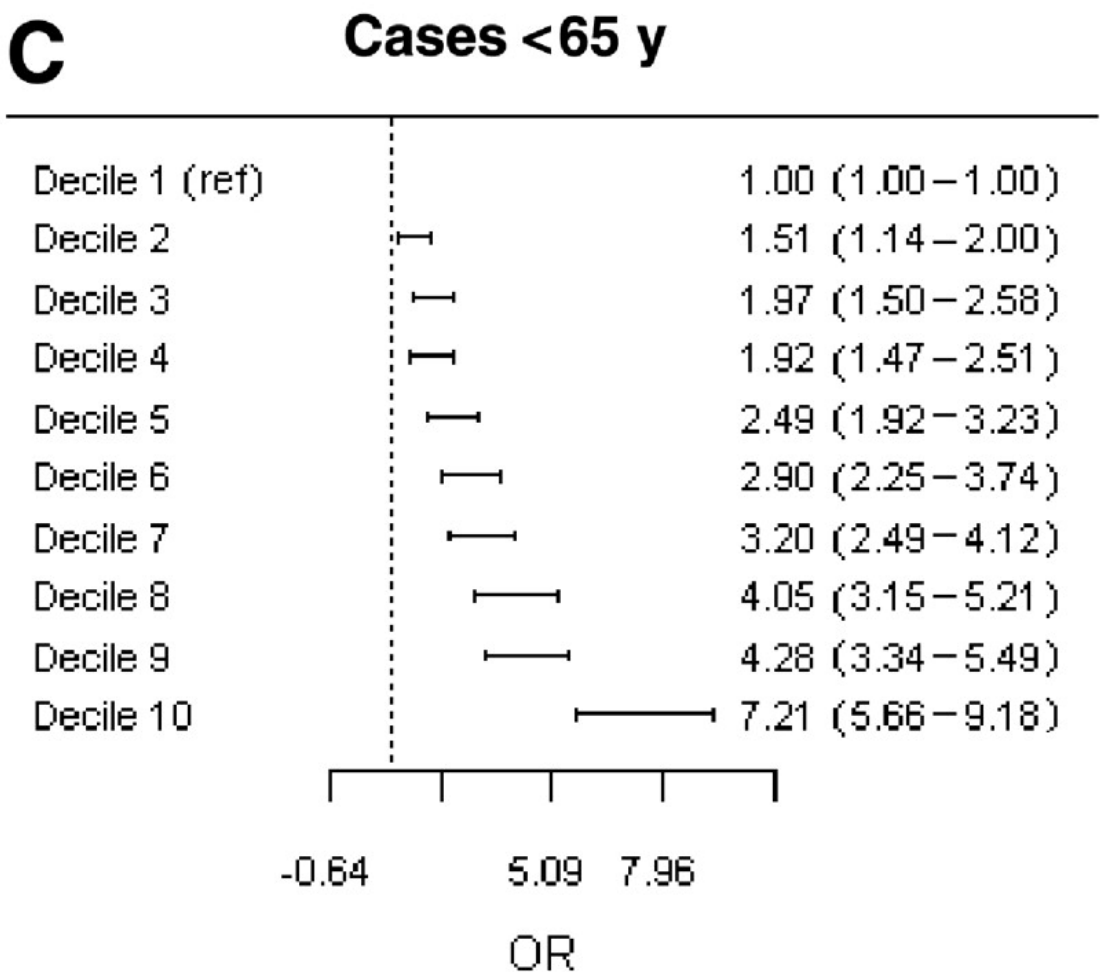
Reclassification statistics and outcome data show improvement in classification

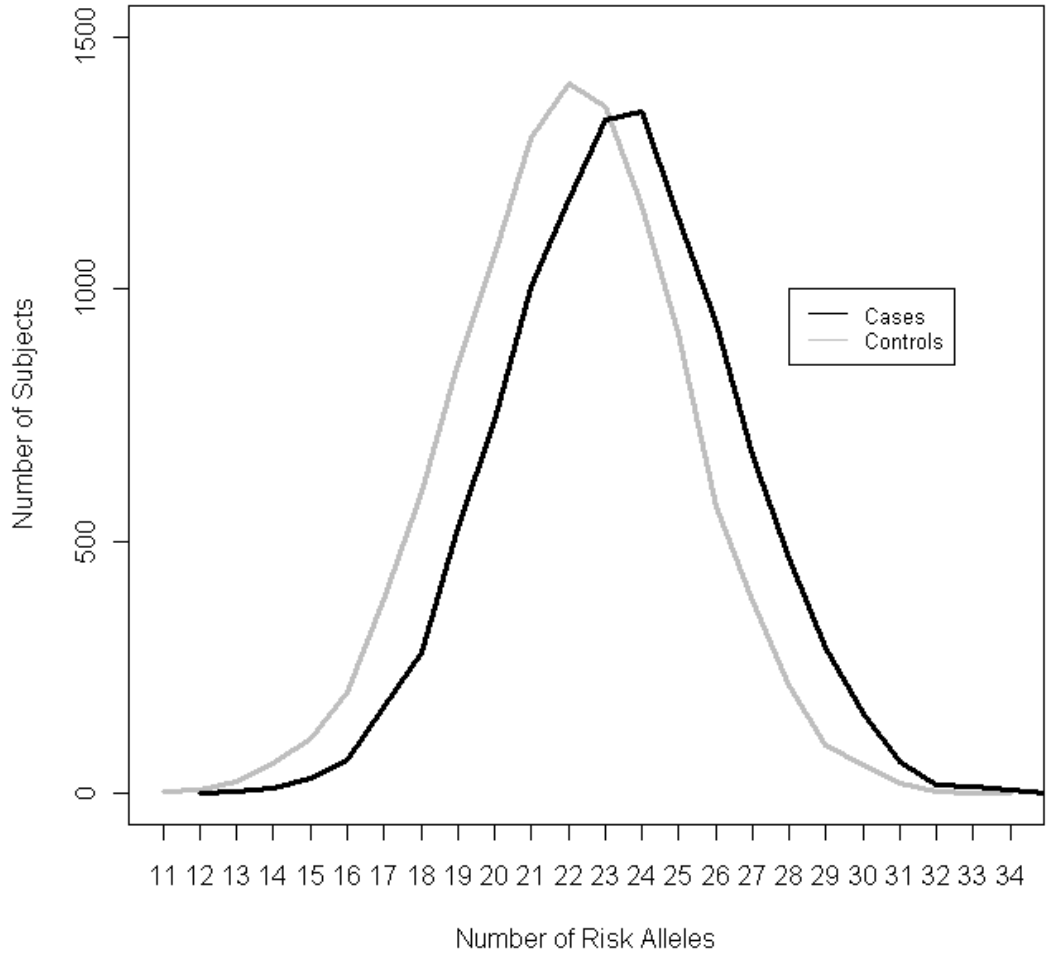
Manolio, Nat Rev Genetics 2013

Two empirical examples

Prostate Cancer	Pancreatic Cancer
Common	Rare
Few known environmental risk factors	Many known environmental risk factors
Often a long natural history with disease that does not progress	Often detected too late and with poor prognosis.
Many common genetic variants identified	Few common genetic variants identified
7,509 cases and 7,652 controls of European Ancestry	3,349 cases and 3,654 controls of European Ancestry
We generated risk models using family history and 25 SNPS	We generated risk models using Smoking, Heavy alcohol use, Body Mass Index, Diabetes, Family history and 4 genetic variants

Prostate cancer - Risk model performance





Lindström et al, CEBP 2012

Does performance vary with age?

Age	Model 1: Family History	Model 2: Genetics	Model 3: Genetics + Family History
-60	0.55 (0.53-0.56)	0.66 (0.64-0.69)	0.68 (0.65-0.71)
61-65	0.53 (0.52-0.54)	0.65 (0.63-0.67)	0.65 (0.63-0.67)
66-70	0.53 (0.52-0.54)	0.63 (0.62-0.65)	0.65 (0.63-0.66)
71-75	0.52 (0.51-0.53)	0.63 (0.61-0.65)	0.64 (0.62-0.66)
75+	0.51 (0.49-0.52)	0.60 (0.57-0.63)	0.60 (0.57-0.63)

Lindström et al, CEBP 2012

Absolute risks of prostate cancer as a function of family history and genetic risk

Table 3. Age-specific mortality-adjusted 10-year absolute risks of prostate cancer among white U.S. men as a function of family history of prostate cancer and genetic risk (as estimated by model 2)

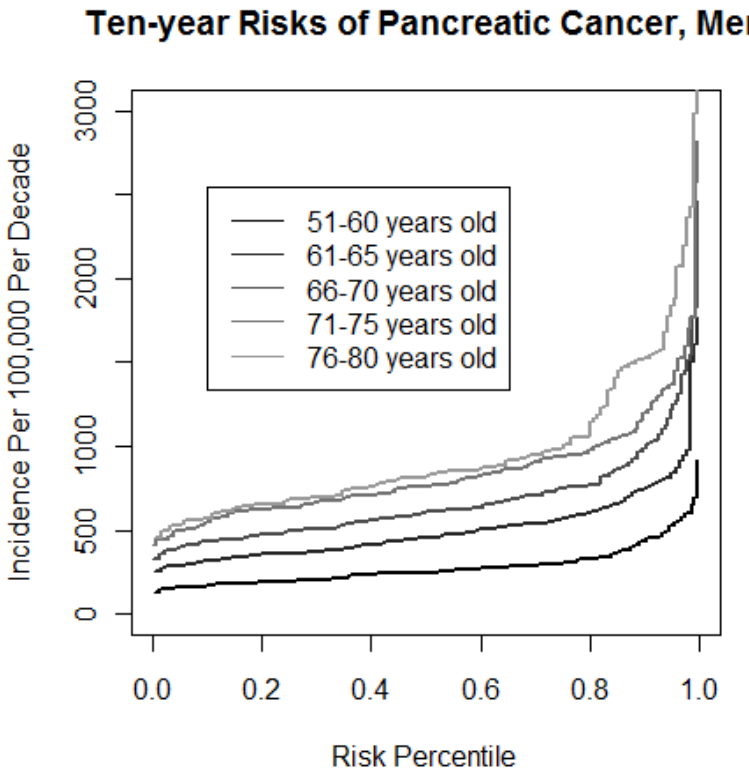
Age	Family history	No information on genetics	10th percentile	30th percentile	50th percentile	70th percentile	90th percentile
50	Negative FH	0.020	0.008	0.012	0.017	0.023	0.034
	Positive FH	0.042	0.016	0.027	0.038	0.049	0.067
60	Negative FH	0.064	0.029	0.043	0.056	0.075	0.109
	Positive FH	0.134	0.057	0.088	0.122	0.154	0.231
70	Negative FH	0.089	0.046	0.065	0.081	0.102	0.139
	Positive FH	0.183	0.104	0.137	0.175	0.209	0.271
80	Negative FH	0.063	0.039	0.049	0.060	0.071	0.089
	Positive FH	0.131	0.085	0.114	0.132	0.143	0.181

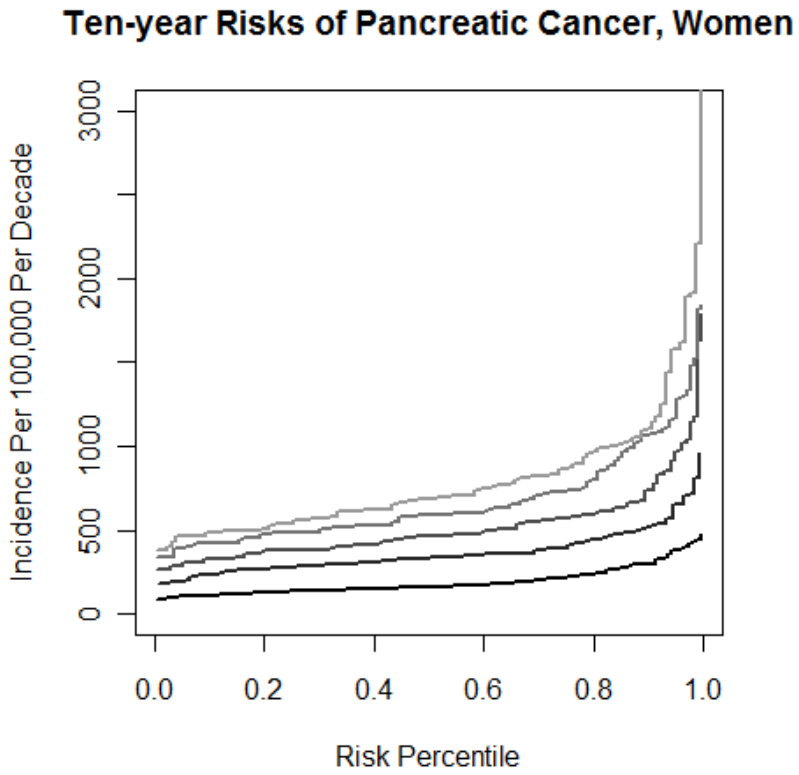
NOTE: Quintiles of genetic risk were based on the distribution in controls. All calculations are based on regression parameters estimated in the imputed data set. Incidence rates are based on SEER data.
Abbreviation: FH, family history.

Lindström et al, CEBP 2012

Pancreatic cancer - Risk model performance

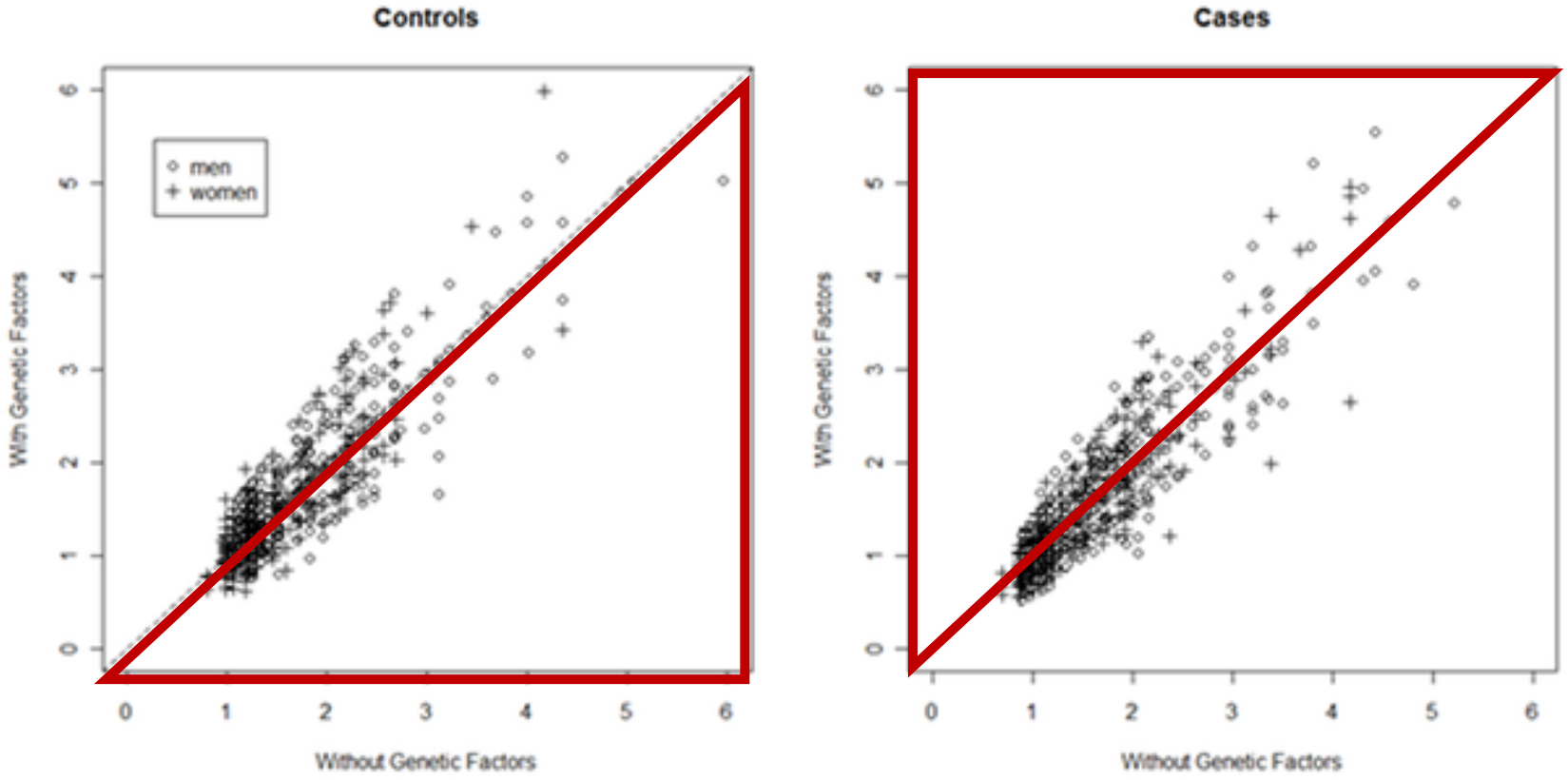
Model 1: Non-genetic risk factors	Model 2: Genetic risk factors	Model 3: Non-genetic and genetic risk factors
AUC=0.57 (0.55-0.59)	AUC=0.58 (0.56-0.60)	AUC=0.61 (0.58-0.63)





Klein et al, PLoS One 2013

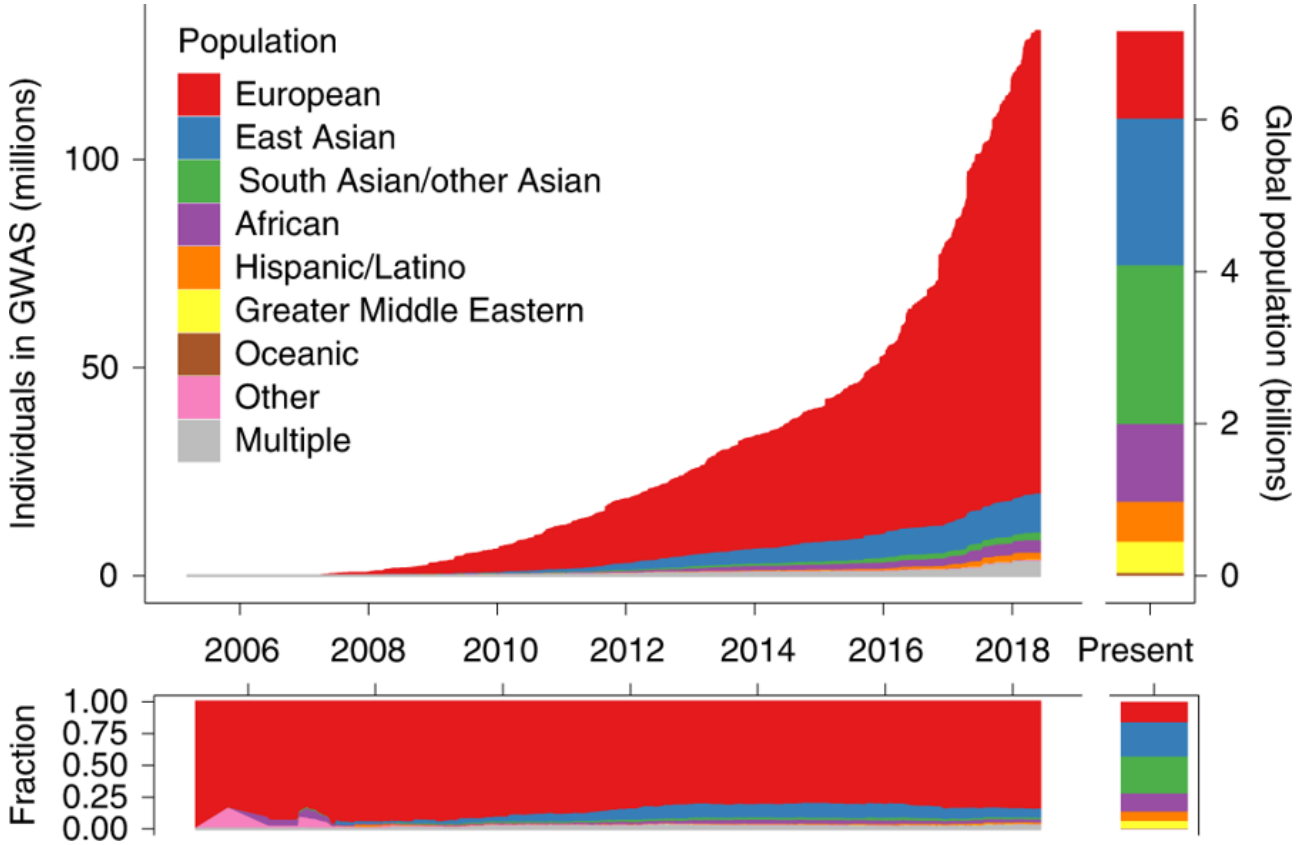
Reclassification of lifetime risk after adding genetic factors to the risk model

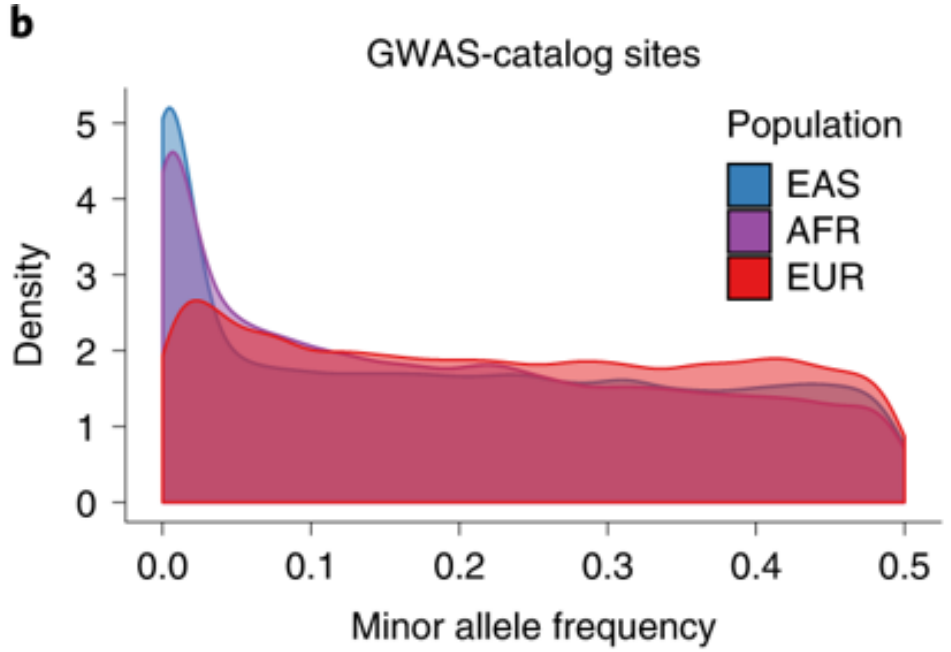


Fewer than 0.3% individuals had more than a 5% average lifetime risk.
No individual had an estimated lifetime risk above 7.5%.

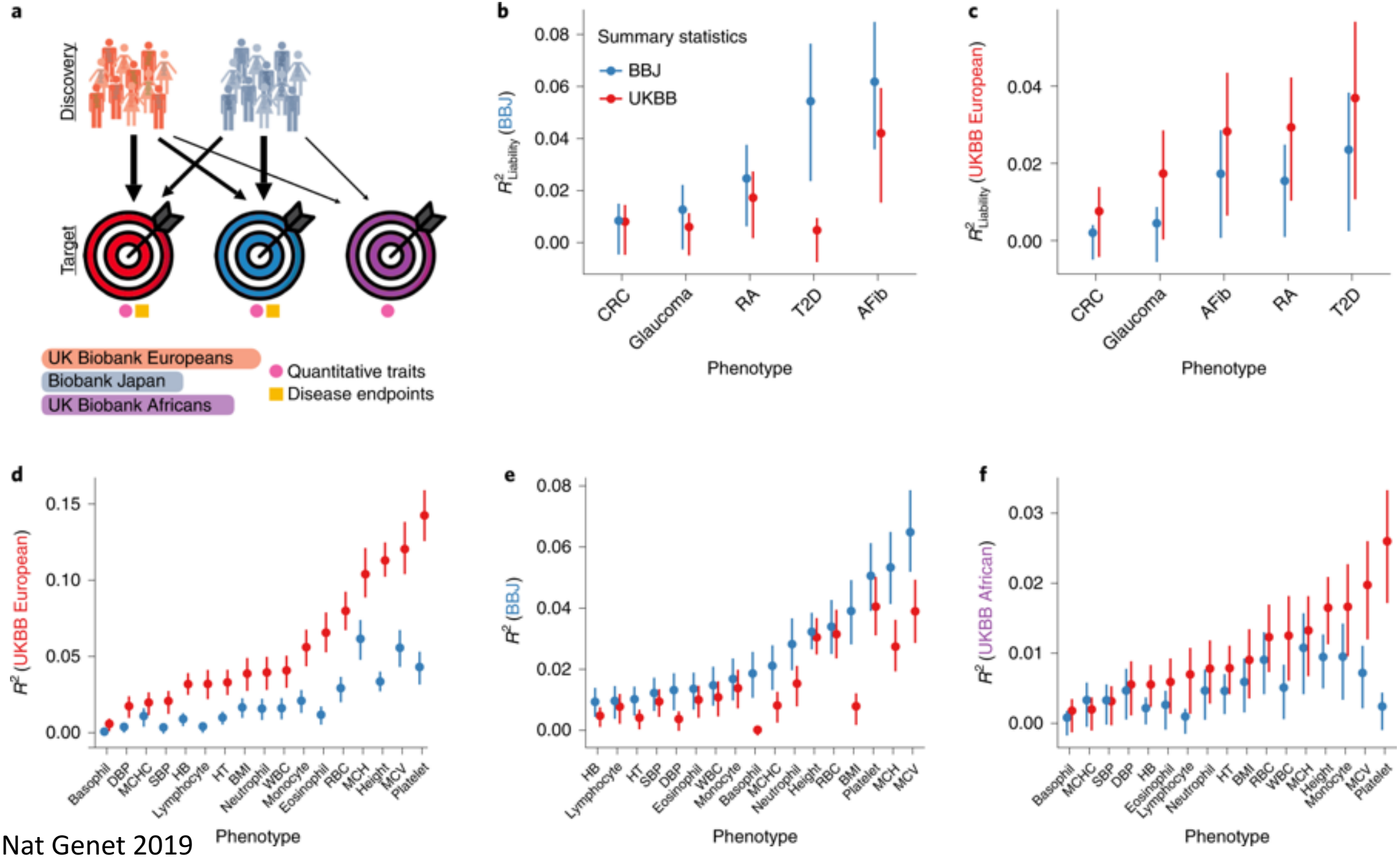
Klein et al, PLoS One 2013

Generalizing PRS across ethnicities





Martin et al, Nat Genet 2019



Martin et al, Nat Genet 2019

PERSPECTIVE
https://doi.org/10.1038/s41588-019-0379-x
nature genetics

Clinical use of current polygenic risk scores may exacerbate health disparities

Alicia R. Martin[1,2,3*], Masahiro Kanai[1,2,3,4,5], Yoichiro Kamatani[5,6], Yukinori Okada[5,7,8], Benjamin M. Neale[1,2,3] and Mark J. Daly[1,2,3,9]

"The single most important step toward parity in PRS accuracy is vastly increasing the diversity of participants included and analyzed in genetic studies, which would improve utility for all groups, most rapidly for underrepresented groups."

Martin et al, Nat Genet 2019

Can genetic information change your behavior (and long-term health)?

Changes in behavior after testing for genetic cancer risk n=762 (23andMe and Pathway Genomics)

Table 3. Unadjusted Associations Between Genetic Risk Scores and Participants' Diet at 6 Months, Overall and Separately by Behavior at Baseline

	Overall			Not Meeting CDC Recommendations for Fruit and Vegetables at Baseline			Meeting CDC Recommendations for Fruit and Vegetables at Baseline		
PGT Cancer Risk	No.	Changed Diet, %	*P*	No.	Changed Diet, %	*P*	No.	Changed Diet, %	*P*
Breast cancer risk			.50			.82			.30
Not elevated	375	34.7		180	30.6		195	38.5	
Elevated	44	29.5		27	33.3		17	23.5	
Colorectal cancer risk			.73			.90			.56
Not elevated	524	30.3		294	27.9		230	33.5	
Elevated	166	28.9		97	28.9		69	29.0	
Prostate cancer risk			.70			.24			.23
Not elevated	207	24.2		137	23.4		70	25.7	
Elevated	64	26.6		46	32.6		18	11.1	

Abbreviations: CDC, Centers for Disease Control and Prevention; PGT, personal genomic testing.

Table 4. Unadjusted Associations Between Genetic Risk Scores and Participants' Exercise Behavior at 6 Months, Overall and Separately by Behavior at Baseline

	Overall			Not Meeting CDC Recommendations for Exercise at Baseline			Meeting CDC Recommendations for Exercise at Baseline		
PGT Cancer Risk	No.	Changed Exercise, %	*P*	No.	Changed Exercise, %	*P*	No.	Changed Exercise, %	*P*
Breast cancer risk			.57			.83			.53
Not elevated	375	27.7		254	28.3		135	26.4	
Elevated	44	31.8		30	30.0		14	35.7	
Colorectal cancer risk			.27			.24			.87
Not elevated	524	24.0		346	23.7		178	24.7	
Elevated	166	28.3		104	29.8		62	25.8	
Prostate cancer risk			.052			.12			.25
Not elevated	207	18.4		120	16.7		87	20.7	
Elevated	64	29.7		43	27.9		21	33.3	

Abbreviations: CDC, Centers for Disease Control and Prevention; PGT, personal genomic testing.

Gray, JCO 2017

Changes in behavior after testing for genetic cancer risk n=762 (23andMe and Pathway Genomics)

Table 5. Unadjusted Associations Between Genetic Risk Scores and Participants' Use of Vitamins or Herbal Supplements at 6 Months, Overall and Separately by Use at Baseline

	Overall			Not Using Vitamins or Herbal Supplements at Baseline			Using Vitamins or Herbal Supplements at Baseline		
PGT Cancer Risk	No.	Changed Use of Vitamins/Herbal Supplements, %	*P*	No.	Changed Use of Vitamins/Herbal Supplements, %	*P*	No.	Changed Use of Vitamins/Herbal Supplements, %	*P*
Breast cancer risk			.79			.99			.99
Not elevated	375	24.5		96	14.6		279	28.0	
Elevated	44	22.7		11	9.1		33	27.3	
Colorectal cancer risk			.53			.42			.39
Not elevated	524	19.5		177	10.7		347	23.9	
Elevated	166	21.7		49	6.1		117	28.2	
Prostate cancer risk			.008			.68			.001
Not elevated	207	11.6		89	7.9		118	14.4	
Elevated	64	25.0		28	3.6		36	41.7	

Abbreviation: PGT, personal genomic testing.

Table 7. Unadjusted Associations Between Genetic Risk Scores and Participants' Cancer-Specific Screening at 6 Months, Overall and Separately by Prior Screening

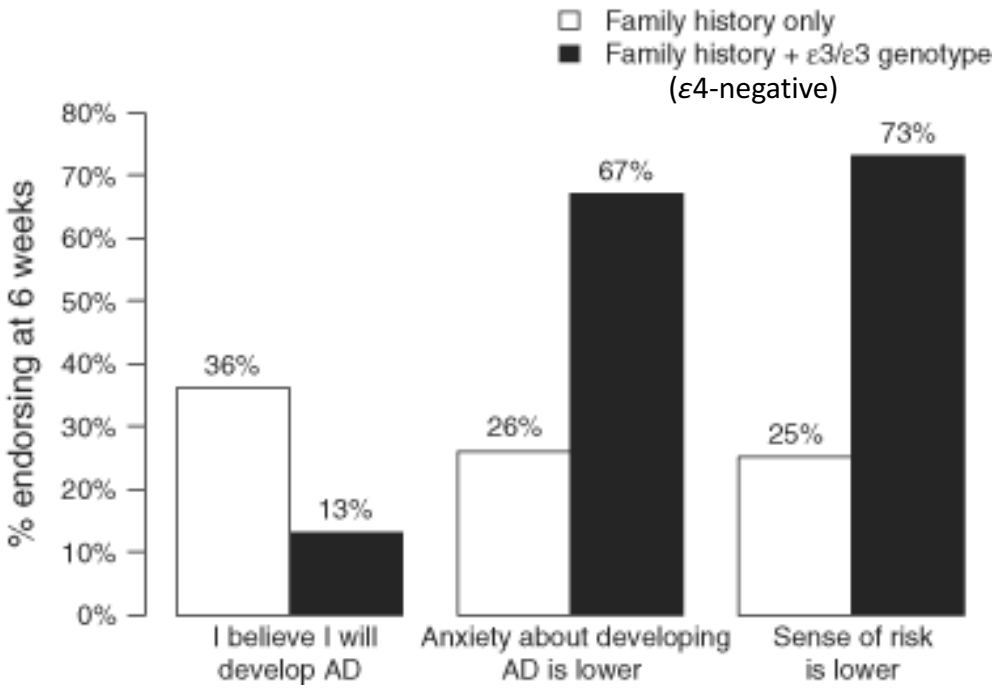
	Overall			No Prior Cancer-Specific Screening			Prior Cancer-Specific Screening		
PGT Cancer Risk	No.	Screened, %	*P*	No.	Screened, %	*P*	No.	Screened, %	*P*
Breast cancer risk			.22			.99			.25
Not elevated	386	27.2		155	0.6		231	45.0	
Elevated	52	19.2		22	0.0		30	33.3	
Colorectal cancer risk			.52			.99			.41
Not elevated	548	6.2		342	2.0		206	13.1	
Elevated	171	7.6		108	1.9		63	17.5	
Prostate cancer risk			.048			.007			.99
Not elevated	216	16.7		151	0.7		65	53.8	
Elevated	65	27.7		40	10.0		65	56.0	

Abbreviation: PGT, personal genomic testing.

Gray, JCO 2017

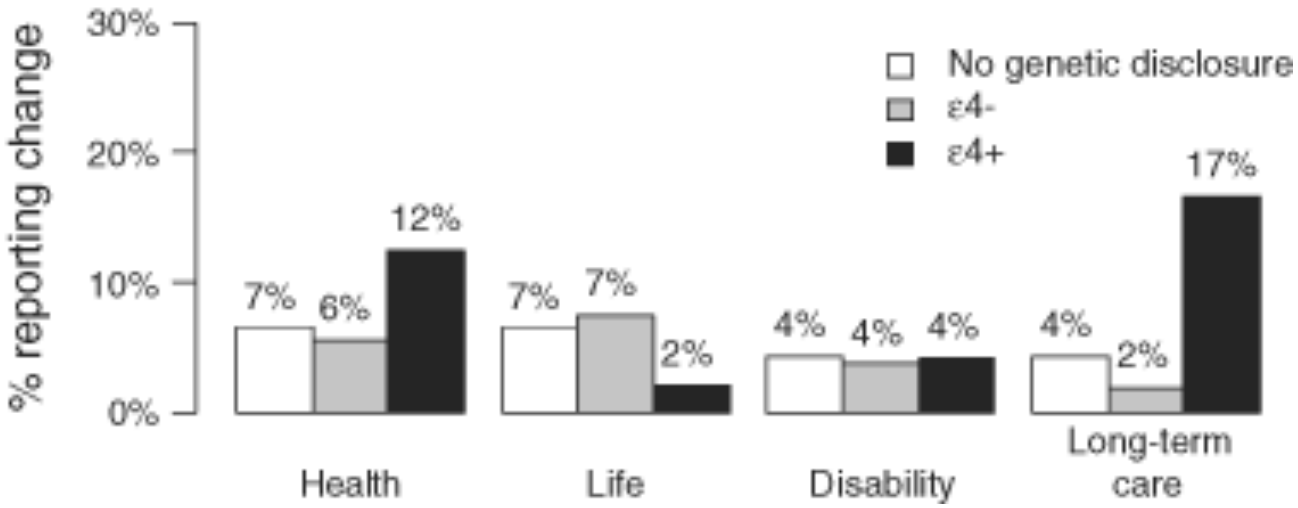
Alzheimer's Disease and APOE

Perceived risk 6 weeks after genetic testing



All women had a 29% life-time risk of developing Alzheimer's

Changes in insurance



Roberts, Clin Genet 2011

Closure of high-tech medical firm Arivale stuns patients: 'I feel as if one of my arms was cut off'

April 26, 2019 at 6:36 pm | *Updated April 26, 2019 at 7:42 pm*

"We start by getting to know you — the whole you — at the deepest level by looking at your genome, your blood, gut microbiome, and lifestyle," the company declared in its early marketing materials. "Then we connect you with a coach who will help explain your data and provide you with clear, actionable, lifestyle recommendations."

Much of what Arivale was selling was future health, whereas consumers "on the whole are looking for immediate gratification," Lewis said.

Seattle Times, April 26 2019

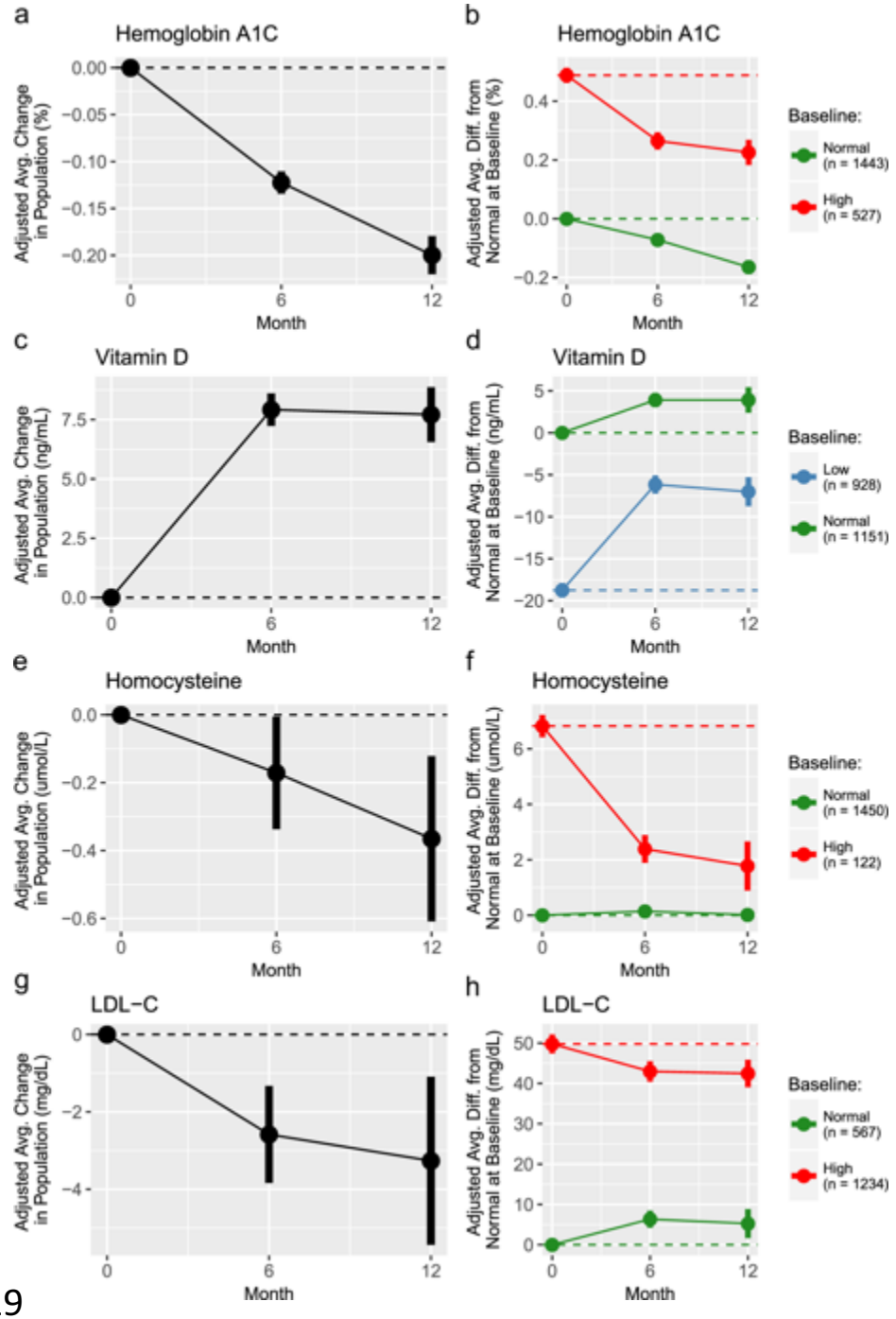


Fig 1. Longitudinal changes for select clinical markers. Panels a,c,e, and g: Adjusted changes for the average participant in the entire study population. Panels b,d,f,h: Adjusted average differences from the 'normal at baseline' strata at baseline for each baseline strata over time in the program.

Zubair, Sci Rep 2019

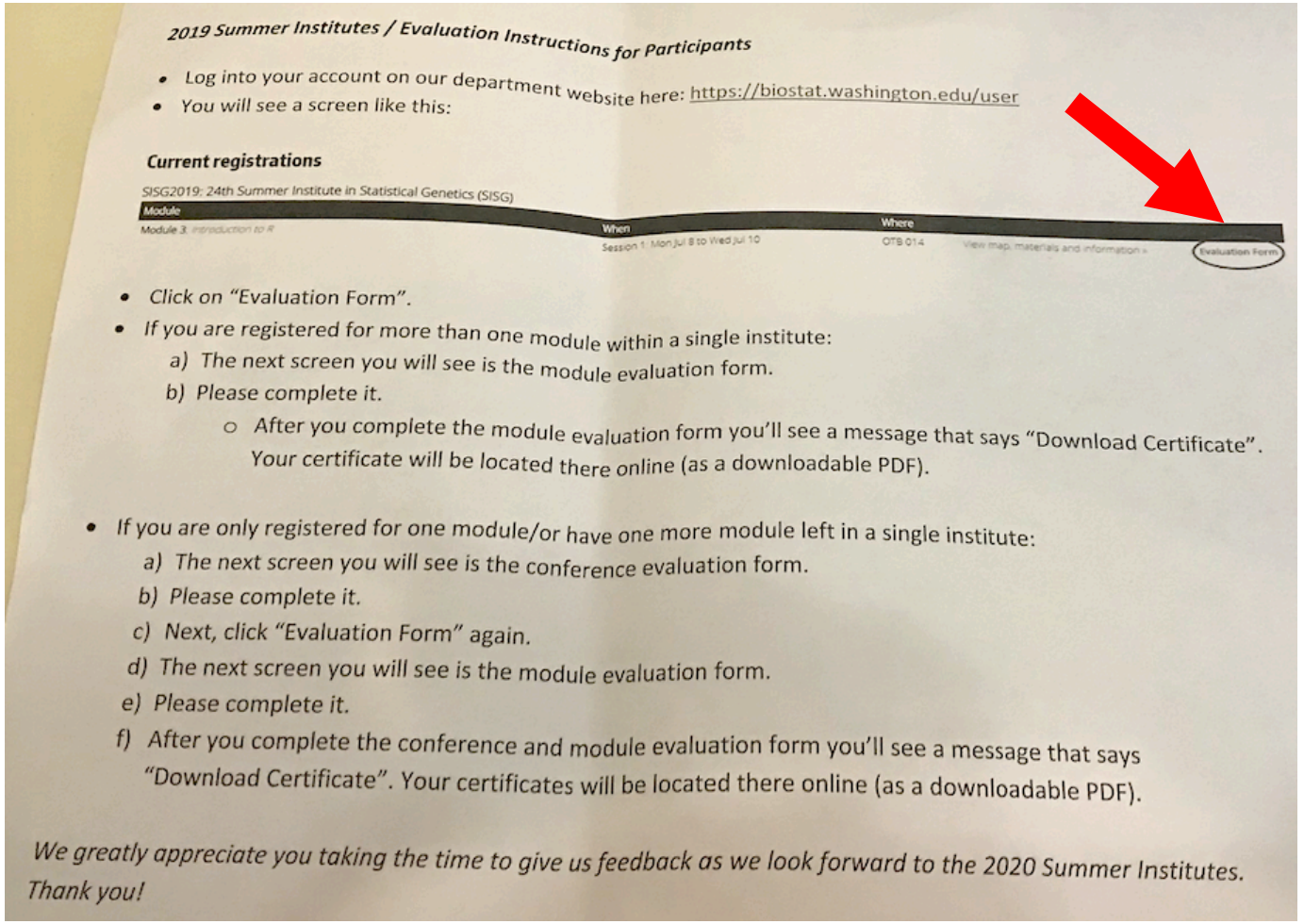
Thank you!

afohner@uw.edu

saralind@uw.edu

- Course Evaluations!

https://biostat.washington.edu/user



2019 Summer Institutes / Evaluation Instructions for Participants

- *Log into your account on our department website here: https://biostat.washington.edu/user*
- *You will see a screen like this:*

Current registrations

SISG2019: 24th Summer Institute in Statistical Genetics (SISG)

Module	When	Where	
Module 3: Introduction to R	Session 1: Mon Jul 8 to Wed Jul 10	OTB 014	View map, materials and information » Evaluation Form

- *Click on "Evaluation Form".*
- *If you are registered for more than one module within a single institute:*
  a) *The next screen you will see is the module evaluation form.*
  b) *Please complete it.*
     - *After you complete the module evaluation form you'll see a message that says "Download Certificate". Your certificate will be located there online (as a downloadable PDF).*

- *If you are only registered for one module/or have one more module left in a single institute:*
  a) *The next screen you will see is the conference evaluation form.*
  b) *Please complete it.*
  c) *Next, click "Evaluation Form" again.*
  d) *The next screen you will see is the module evaluation form.*
  e) *Please complete it.*
  f) *After you complete the conference and module evaluation form you'll see a message that says "Download Certificate". Your certificates will be located there online (as a downloadable PDF).*

We greatly appreciate you taking the time to give us feedback as we look forward to the 2020 Summer Institutes. Thank you!