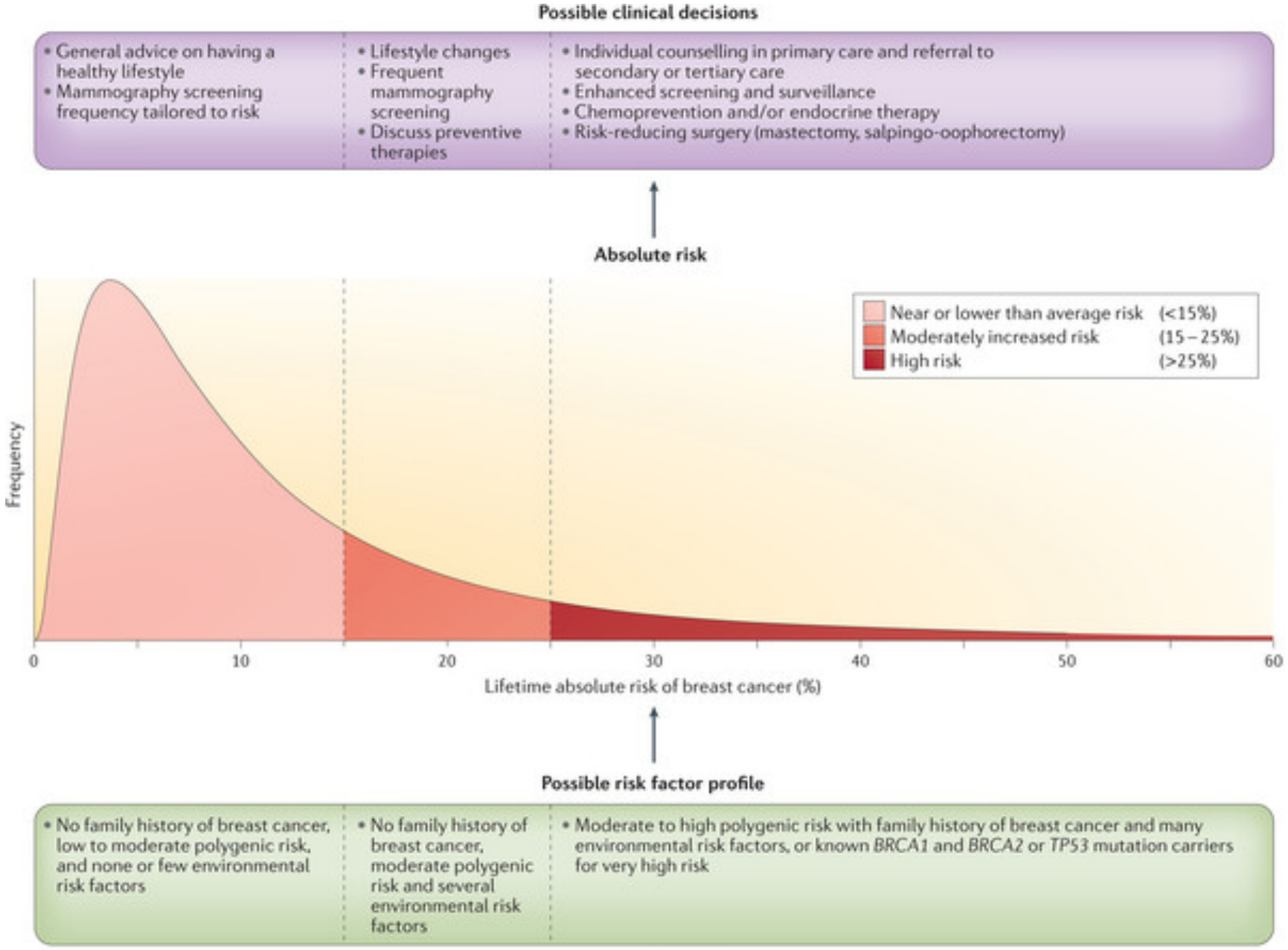
Risk prediction

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

Abstract Send to:

N Engl J Med. 2008 Feb 28;358(9):910-9. doi: 10.1056/NEJMoa075819. Epub 2008 Jan 16.

Cumulative association of five genetic variants with prostate cancer.

Zheng SL[1], Sun J, Wiklund F, Smith S, Stattin P, Li G, Adami HO, Hsu FC, Zhu Y, Bälter K, Kader AK, Turner AR, Liu W, Bleecker ER, Meyers DA, Duggan D, Carpten JD, Chang BL, Isaacs WB, Xu J, Grönberg H.

Author information

Abstract

BACKGROUND: Single-nucleotide polymorphisms (SNPs) in five chromosomal regions--three at 8q24 and one each at 17q12 and 17q24.3--have been associated with prostate cancer. Each SNP has only a moderate association, but when SNPs are combined, the association may be stronger.

METHODS: We evaluated 16 SNPs from five chromosomal regions in a Swedish population (2893 subjects with prostate cancer and 1781 control subjects) and assessed the individual and combined association of the SNPs with prostate cancer.

RESULTS: Multiple SNPs in each of the five regions were associated with prostate cancer in single SNP analysis. When the most significant SNP from each of the five regions was selected and included in a multivariate analysis, each SNP remained significant after adjustment for other SNPs and family history. Together, the five SNPs and family history were estimated to account for 46% of the cases of prostate cancer in the Swedish men we studied. The five SNPs plus family history had a cumulative association with prostate cancer (P for trend, 3.93x10(-28)). In men who had any five or more of these factors associated with prostate cancer, the odds ratio for prostate cancer was 9.46 (P=1.29x10(-8)), as compared with men without any of the factors. The cumulative effect of these variants and family history was independent of serum levels of prostate-specific antigen at diagnosis.

CONCLUSIONS: SNPs in five chromosomal regions plus a family history of prostate cancer have a cumulative and significant association with prostate cancer.

Comment in

Five genetic variants associated with prostate cancer. [N Engl J Med. 2008]
Five genetic variants associated with prostate cancer. [N Engl J Med. 2008]
Five genetic variants associated with prostate cancer. [N Engl J Med. 2008]
Five genetic variants associated with prostate cancer. [N Engl J Med. 2008]
Five genetic variants associated with prostate cancer. [N Engl J Med. 2008]
Five genetic variants associated with prostate cancer. [N Engl J Med. 2008]
Five genetic variants associated with prostate cancer. [N Engl J Med. 2008]
Re: cumulative association of five genetic variants with prostate cancer. [Eur Urol. 2008]
Complexities of prostate-cancer risk. [N Engl J Med. 2008]
Re: Cumulative association of five genetic variants with prostate cancer. [Eur Urol. 2008]
Words of wisdom. Re: Cumulative association of five genetic variants with prostate cancer. Zheng SL, Sun J, Wiklund F, et al. [Eur Urol. 2008]

PMID: 18199855 [PubMed - indexed for MEDLINE] Free full text

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

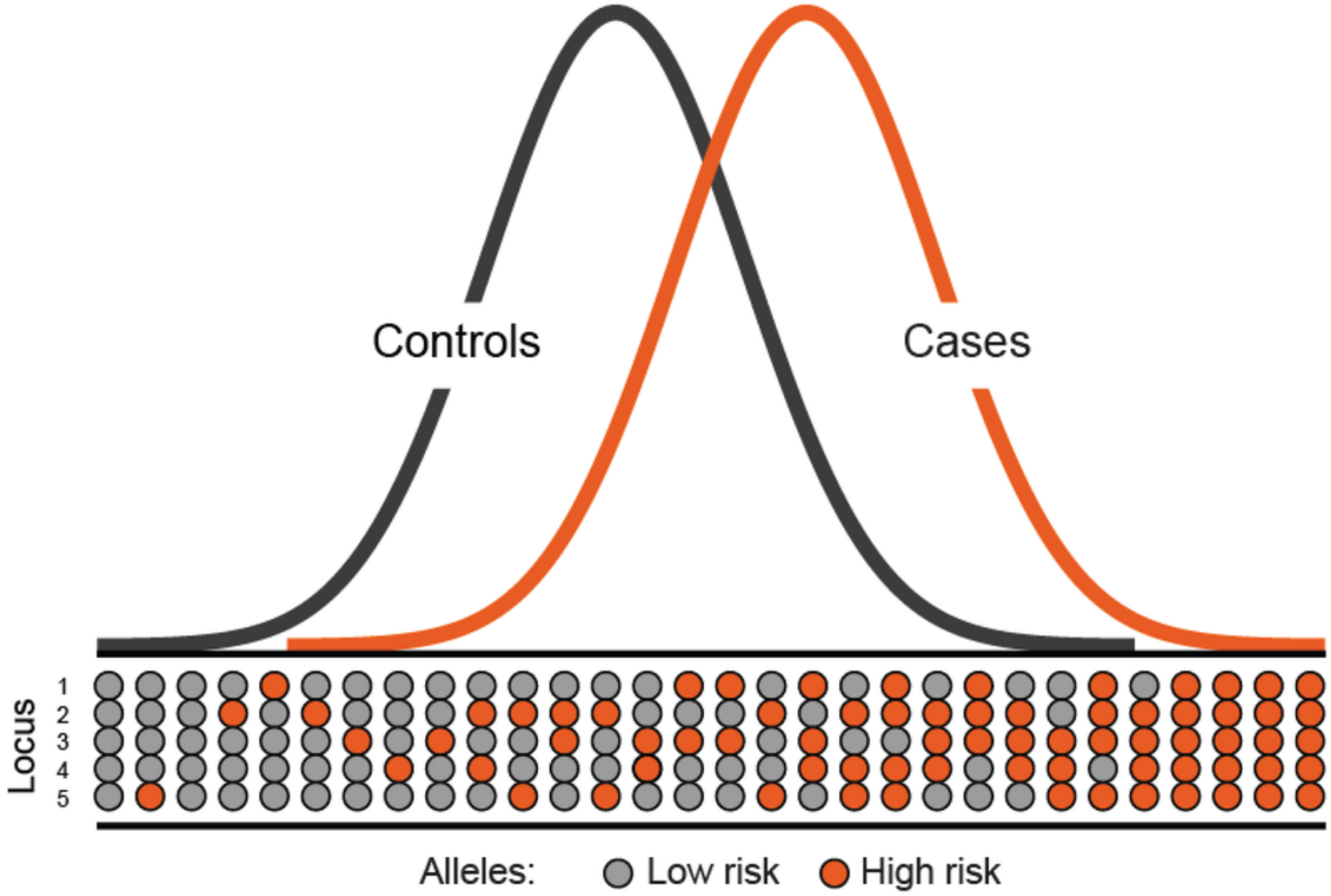
$$GRS_i = \sum_{j=1}^{m} \beta_{ij} G_{ij}$$

Generating a genetic risk score

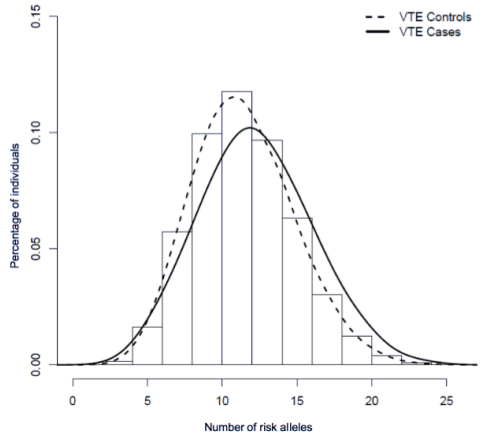
- If you are using a weighted score, do not use βs from your own data -> model overfitting

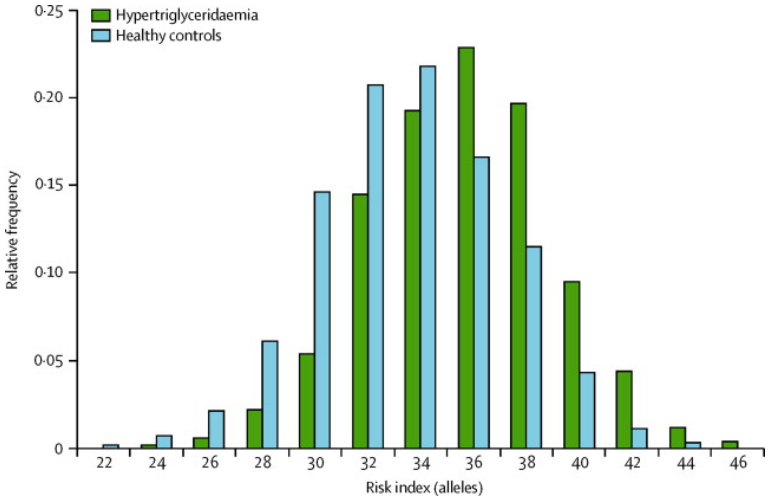
- Need to handle missing data
  - Complete case analysis (remove all samples with ≥1 SNP missing)
  - Impute
    - LD (do not always have this information, e.g. only GRS SNPs were genotyped)
    - Expected value based on allele frequency (PLINK)
    - Sampling from your data conditioned on some variables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

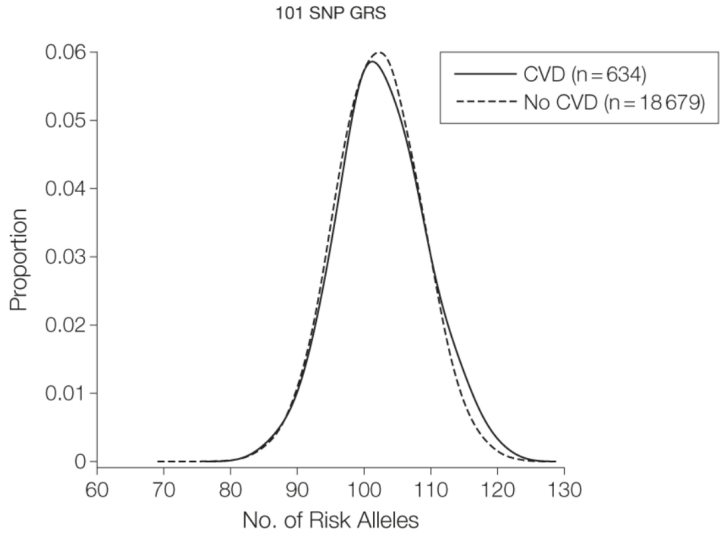
Distribution of genetic risk scores (GRS)

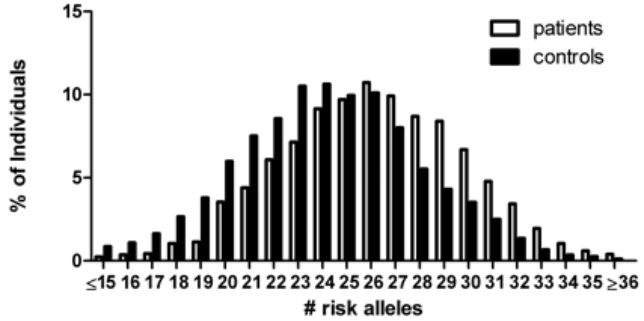


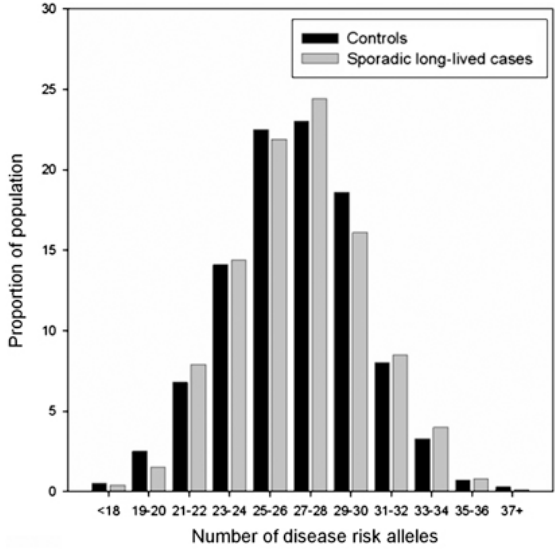
Distribution of GRS for complex diseases

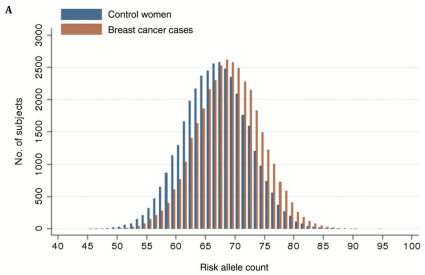


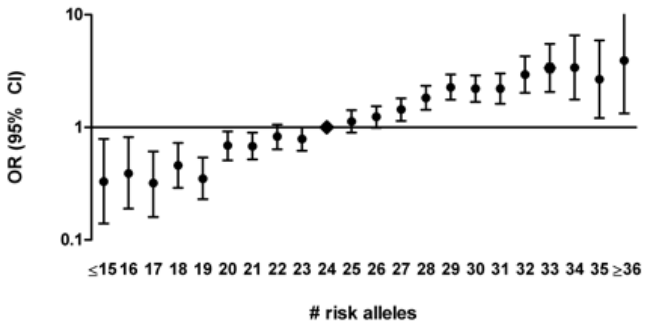


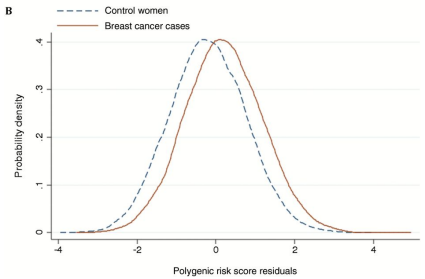




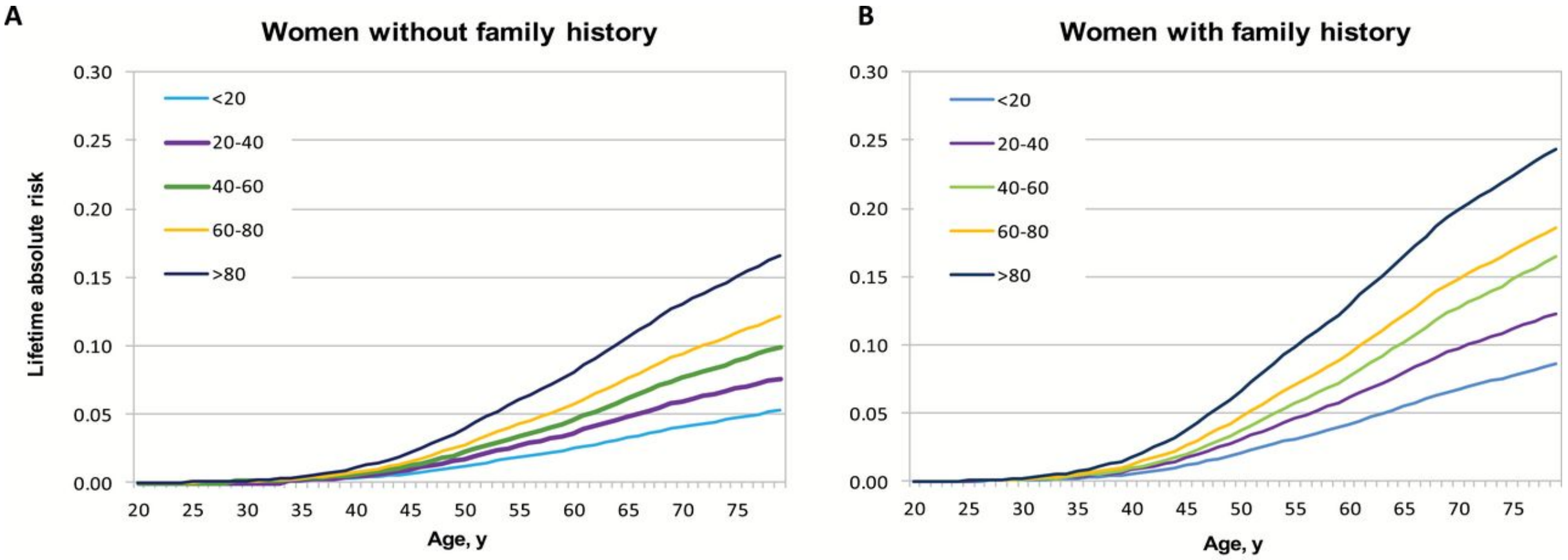








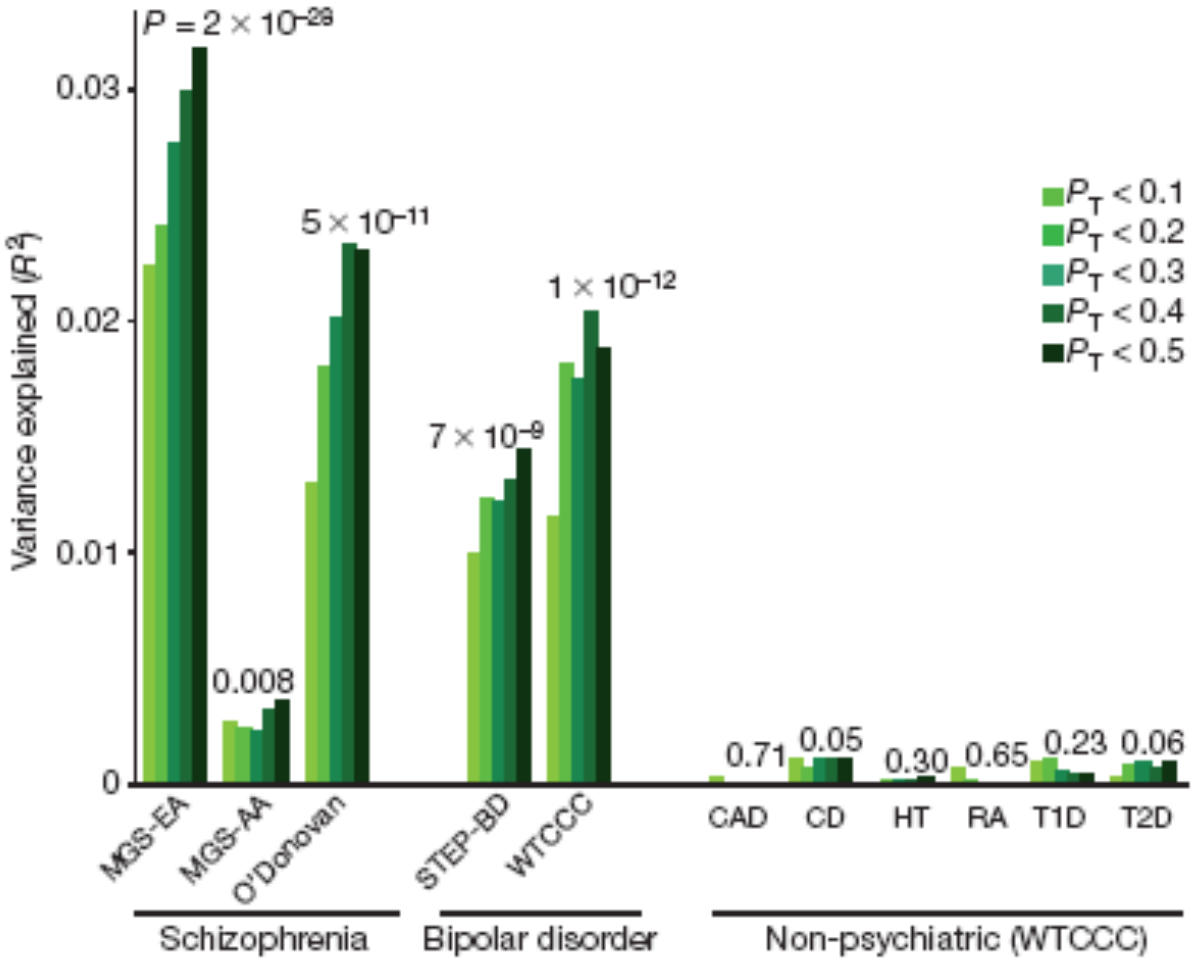
Lifetime risk of breast cancer based on a genetic risk score (77 SNPs) in women of European origin



Nasim Mavaddat et al. JNCI J Natl Cancer Inst 2015;107:djv036

JNCI

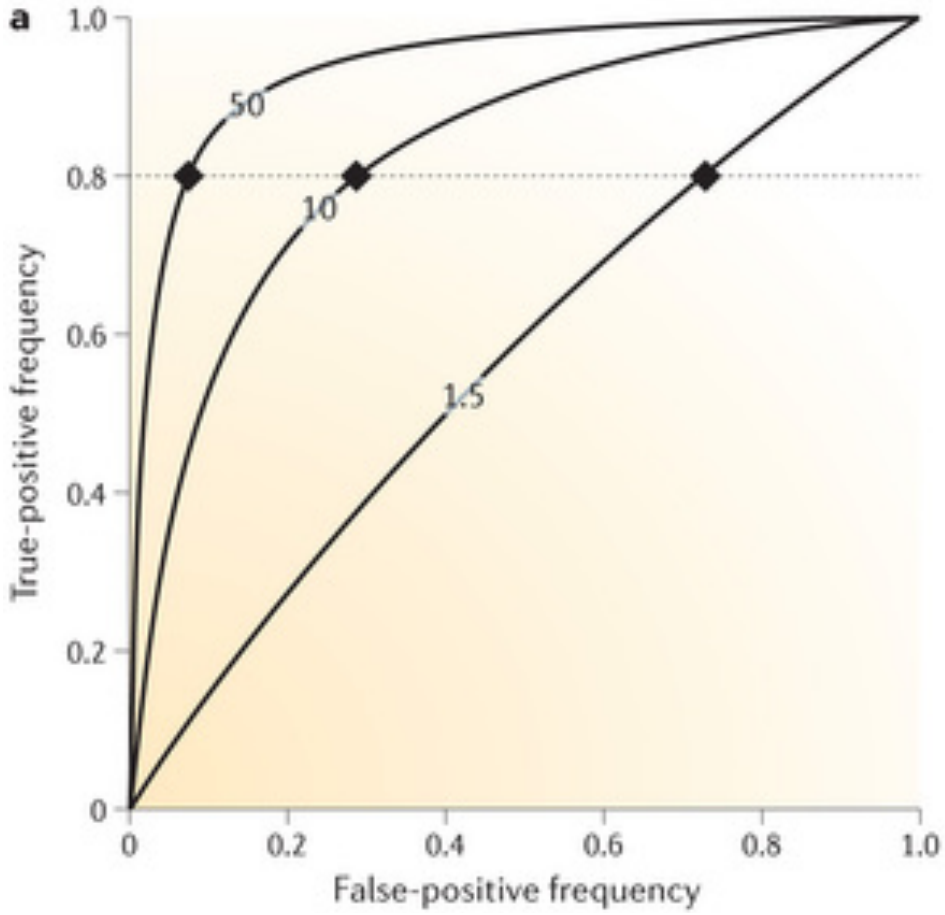
Going beyond genome-wide significant SNPs



Purcell, Nature 2009

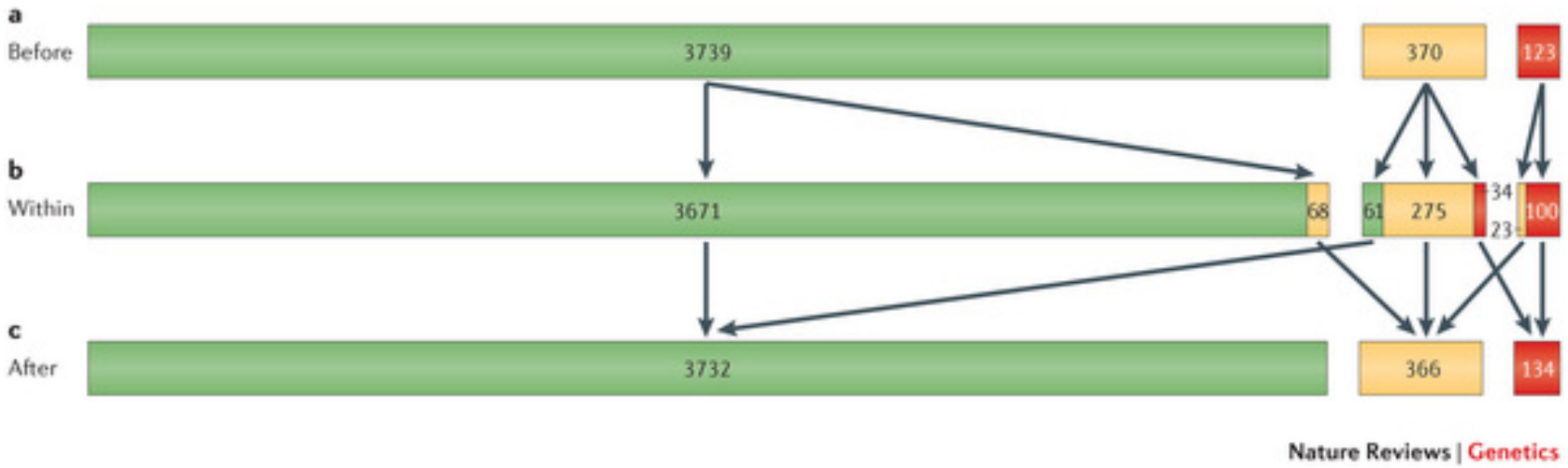
Measures of risk prediction performance (i)

- 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

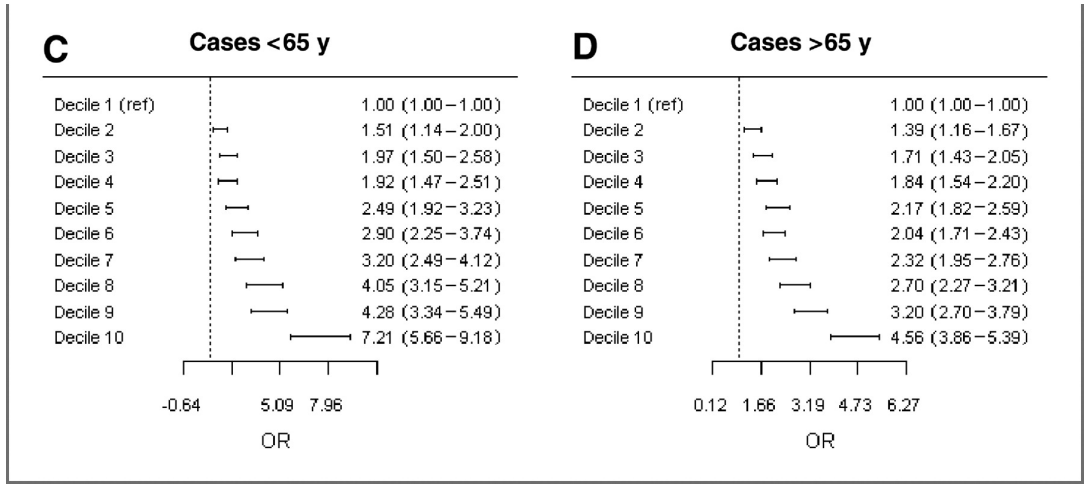
Reclassification statistics and outcome data show improvement in classification

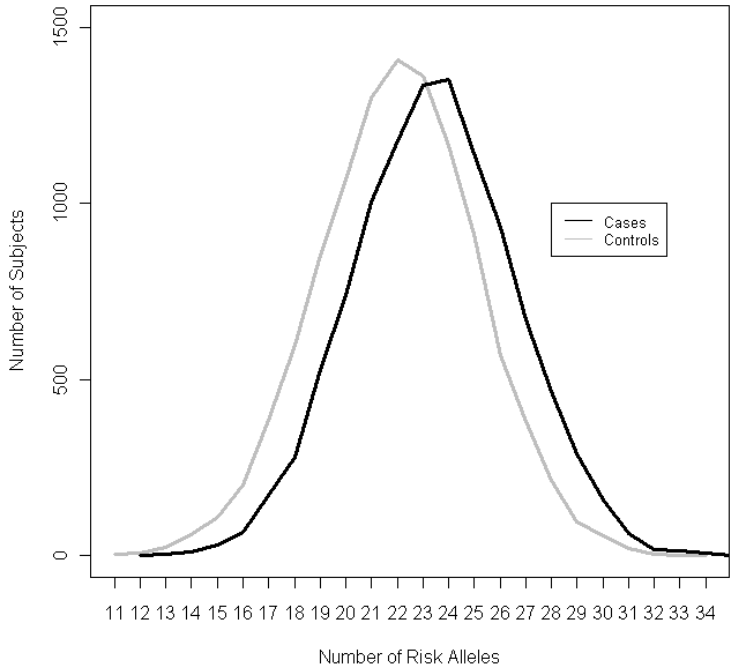
Manolio, Nat Rev Genetics 2013

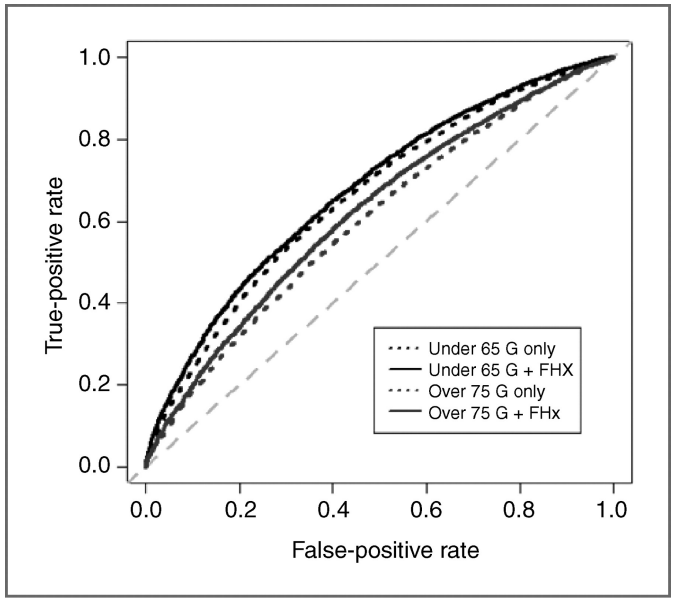
Two empirical examples

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)

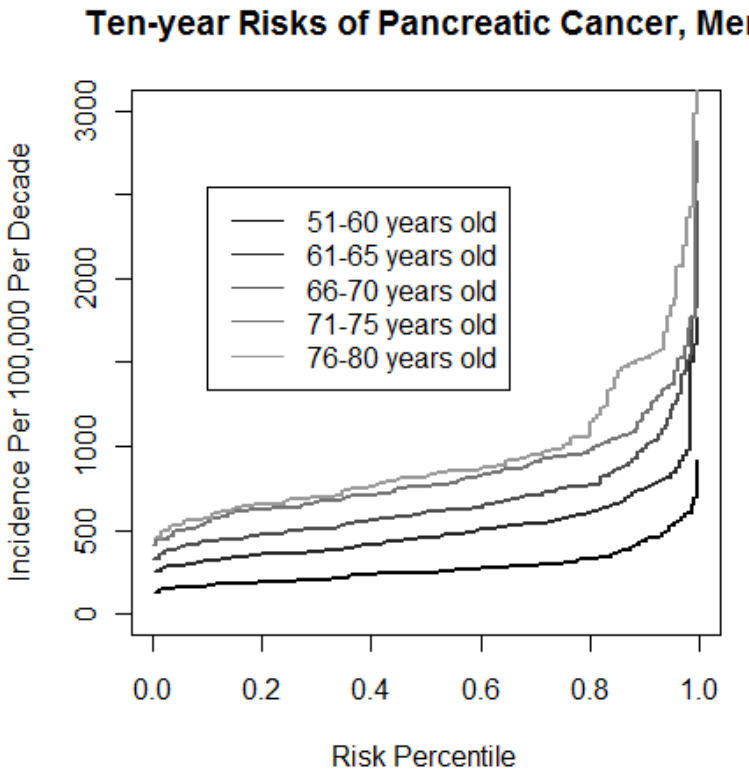
Age	Family history	No information on genetics	10th percentile	30th percentile	50th percentile	70th percentile	90th percentile
50	Negative FH	0.020	0.008	0.012	0.017	0.023	0.034
	Positive FH	0.042	0.016	0.027	0.038	0.049	0.067
60	Negative FH	0.064	0.029	0.043	0.056	0.075	0.109
	Positive FH	0.134	0.057	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

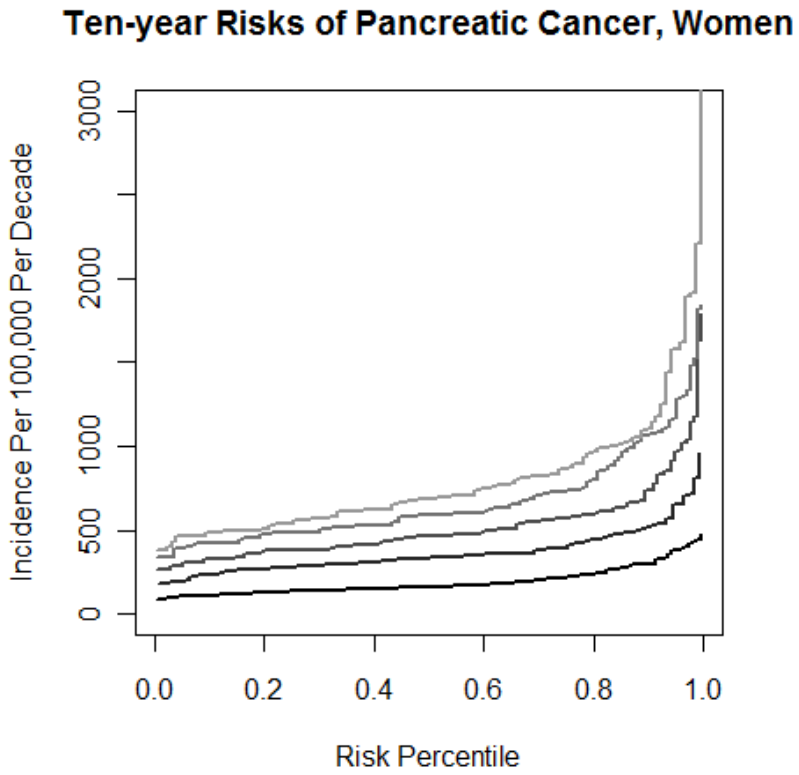
NOTE: Quintiles of genetic risk were based on the distribution in controls. All calculations are based on regression parameters estimated in the imputed data set. Incidence rates are based on SEER data.
Abbreviation: FH, family history.

Lindström et al, CEBP 2012

Pancreatic cancer - Risk model performance

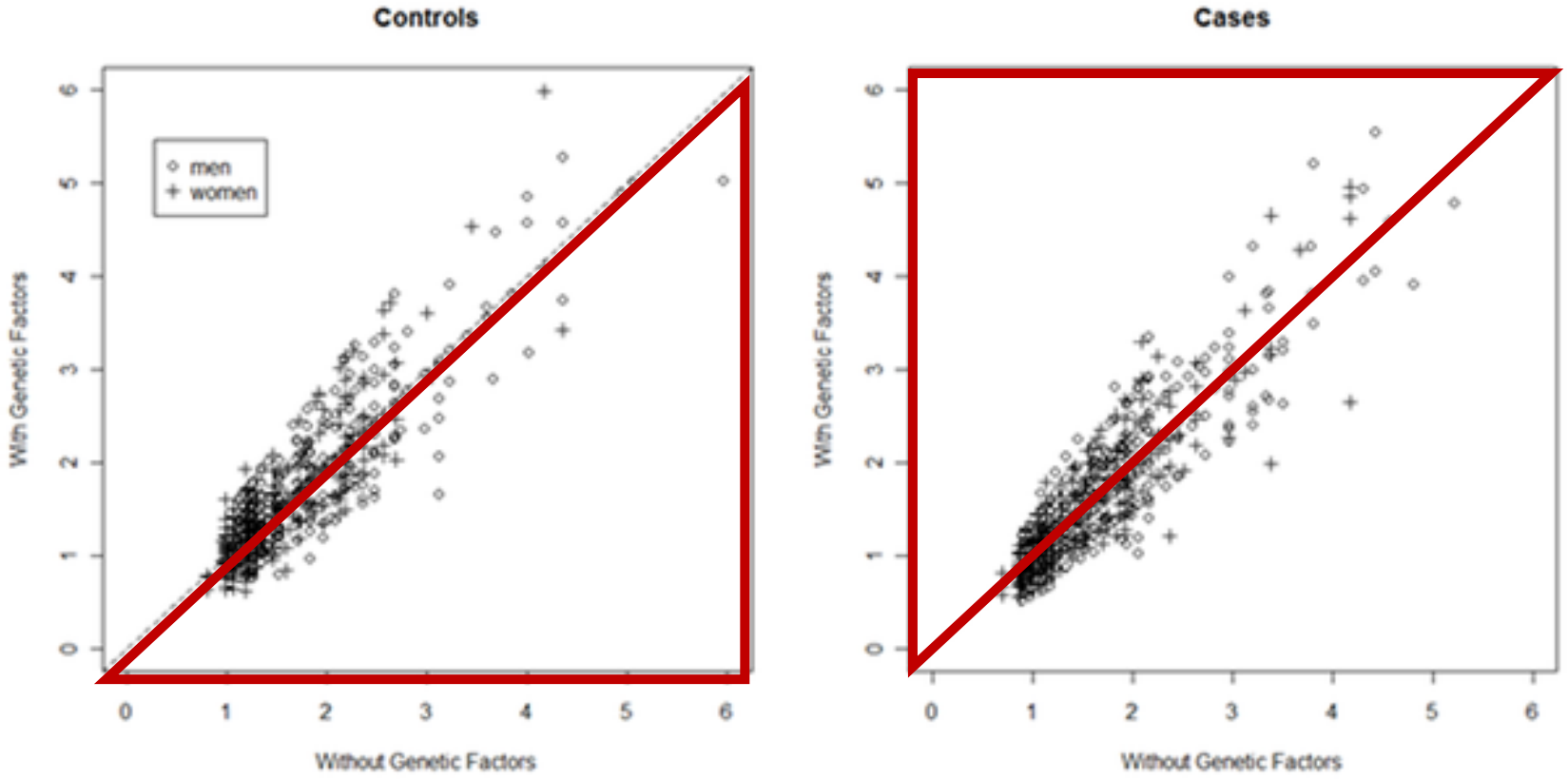
Model 1: Non-genetic risk factors	Model 2: Genetic risk factors	Model 3: Non-genetic and genetic risk factors
AUC=0.57 (0.55-0.59)	AUC=0.58 (0.56-0.60)	AUC=0.61 (0.58-0.63)





Klein et al, PLoS One 2013

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

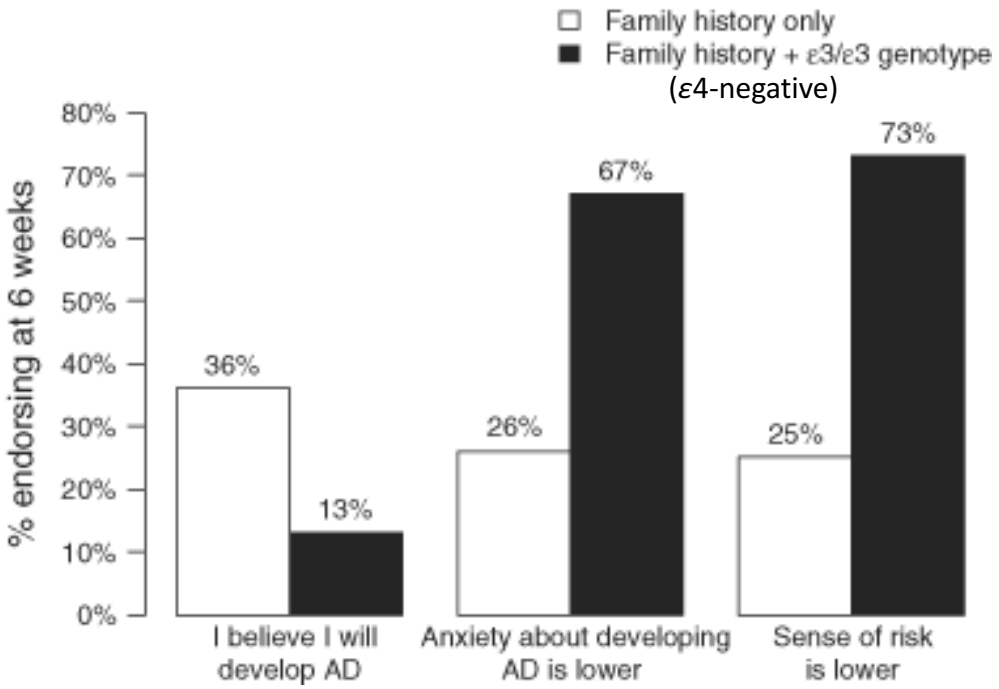


Fewer than 0.3% individuals had more than a 5% average lifetime risk.
No individual had an estimated lifetime risk above 7.5%.

Klein et al, PLoS One 2013

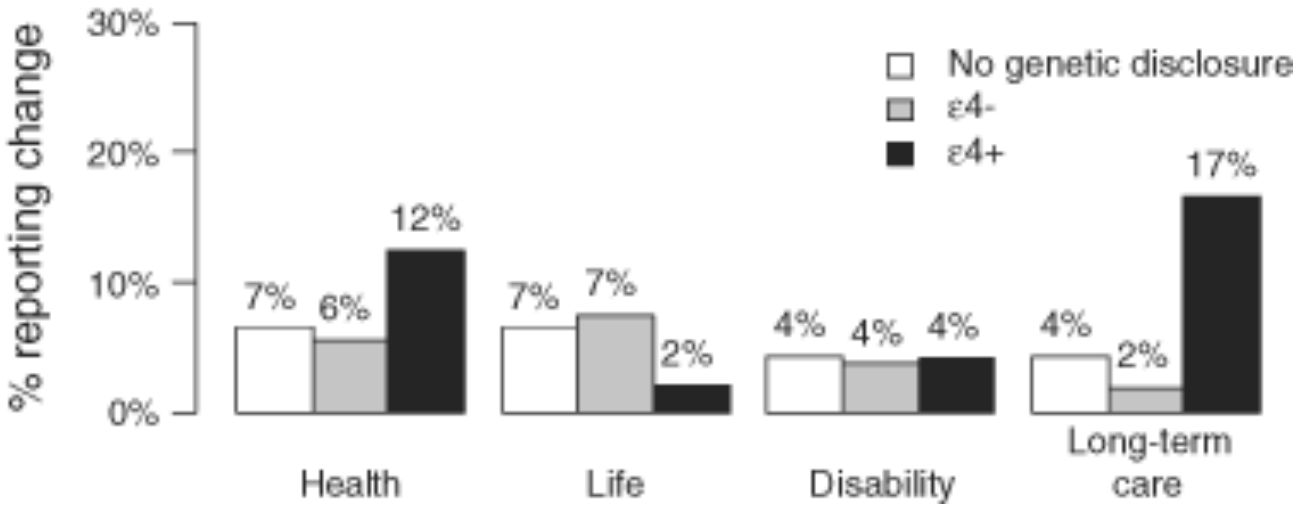
Alzheimer's Disease and APOE

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

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.

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