Genetic Screening and Diagnosis

Jeff Struewing
Office of Population Genomics
NHGRI

Epidemiology for Genetics Researchers
July 18, 2008
NIH Natcher Auditorium

Genetic Screening and Diagnosis

(Lecture 7)

A. Bayes' Theorem
B. Test Performance Measures
C. Role of Prevalence (Prior Probability)
D. Clinical Breast Cancer Genetics

Bayes' Theorem

Bayes' theorem involves relationships between probabilities of events

Basic to the scientific method (...how likely is my theory, given new evidence...)

Important in epidemiology and statistics, especially in screening

Bayes' Theorem

MANY ways to state the theorem

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

$$P(A|B) = \frac{P(A \& B)}{P(B)}$$

Bayes' Theorem

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

P(A|B) is the posterior or conditional probability of A, given B

P(A) is the prior or marginal probability of A

P(B) is the prior probability of B (a normalizing constant)

Bayes' Theorem

P(A | B)

How probable is it that a woman has breast cancer if she has an abnormal screening mammogram?

A represents the true condition (breast cancer)

B represents the screening test (mammogram) result

Bayes' Theorem

Stepping back and framing the problem

We are interested in how new data affects the probability or likelihood of a state….

Does a woman have breast cancer?

Is a person likely to have an adverse drug reaction?

Is this SNP a significant susceptibility allele for diabetes?

Bayes' Theorem

Does a woman have breast cancer?

If I know nothing other than an adult female is before me, what it the chance she has breast cancer?

background prevalence of breast cancer - this is the prior or marginal probability

Bayes' Theorem

Does a woman have breast cancer?

If she has a mammogram and it is positive, I can recalculate her chance of having breast cancer?

--> posterior probability

Dependent on the performance characteristics of mammography (how good is the test) AND the background prevalence of the condition

Genetic Screening and Diagnosis

Evaluation of Genomic Applications in Practice and Prevention (EGAPP) is an initiative launched in 2004 to support a coordinated, systematic process for evaluating genetic tests and other genomic applications that are in transition from research to clinical and public health practice in the United States.

Criteria and Considerations for Prioritization and Selection of Evidence Review Topics

- **Health Burden (disease prevalence and severity, strength of association, effective intervention, etc.)**
- **Practice Issues (test availability, complexity, etc.)**

Completed topics include ovarian cancer, CYP450 and SSRIs, HNPCC, and breast cancer expression profiling.

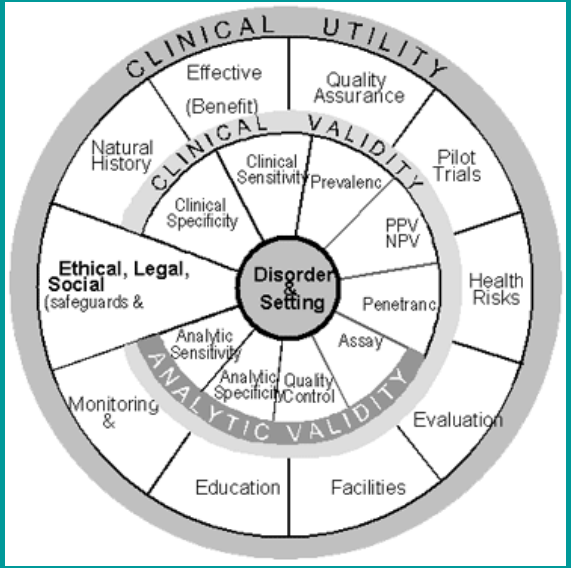
http://egappreviews.org

Genetic Screening and Diagnosis

- **EGAPP grew out of an earlier effort called ACCE - analytical validity, clinical validity, clinical utility, and ELSI**

ACCE Model System for Collecting, Analyzing and Disseminating Information on Genetic Tests

National Office of Public Health Genomics, CDC



http://www.cdc.gov/genomics/gtesting/ACCE.htm

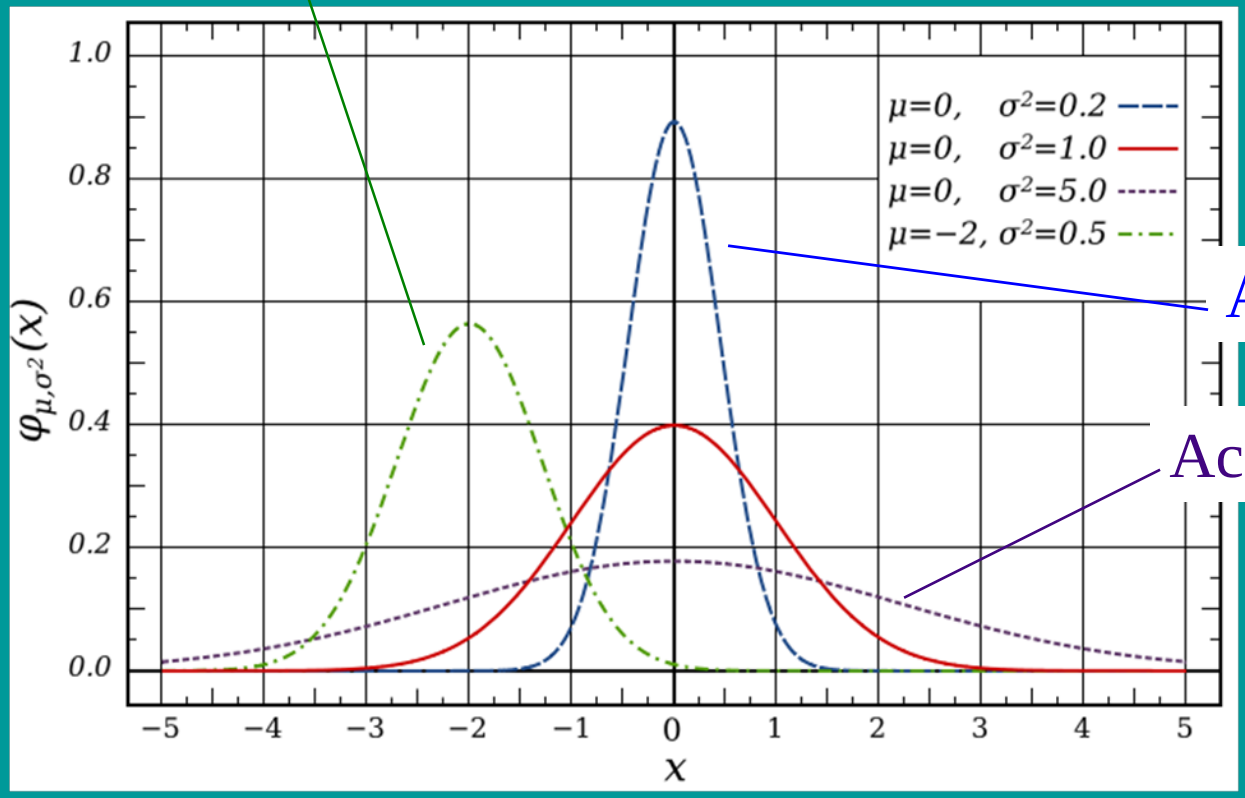
Measures of Test Performance

- **Analytical and clinical validity**
  - Reliability
  - Accuracy
  - Sensitivity
  - Specificity
  - Positive predictive value
  - Negative predictive value
- **Clinical utility**
  - Effective interventions

Measures of Test Performance

Accuracy & Reliability

Inaccurate but reliable test



Accurate and reliable test

Accurate but unreliable test

X = true value

Measures of Test Performance

Sensitivity

How good is a screening test at identifying disease?

Specificity

How good is a screening test at identifying the non-diseased state?

"inherent" properties of a screening test

Sensitivity and Specificity of Mammography

Sensitivity

- ~ 80%
- 90% in women over age 60
- 60% in women under age 40
- ~100% in fatty breasts
- 45% in dense breasts

Specificity

- ~ 90%

Sensitivity and Specificity of Mammography

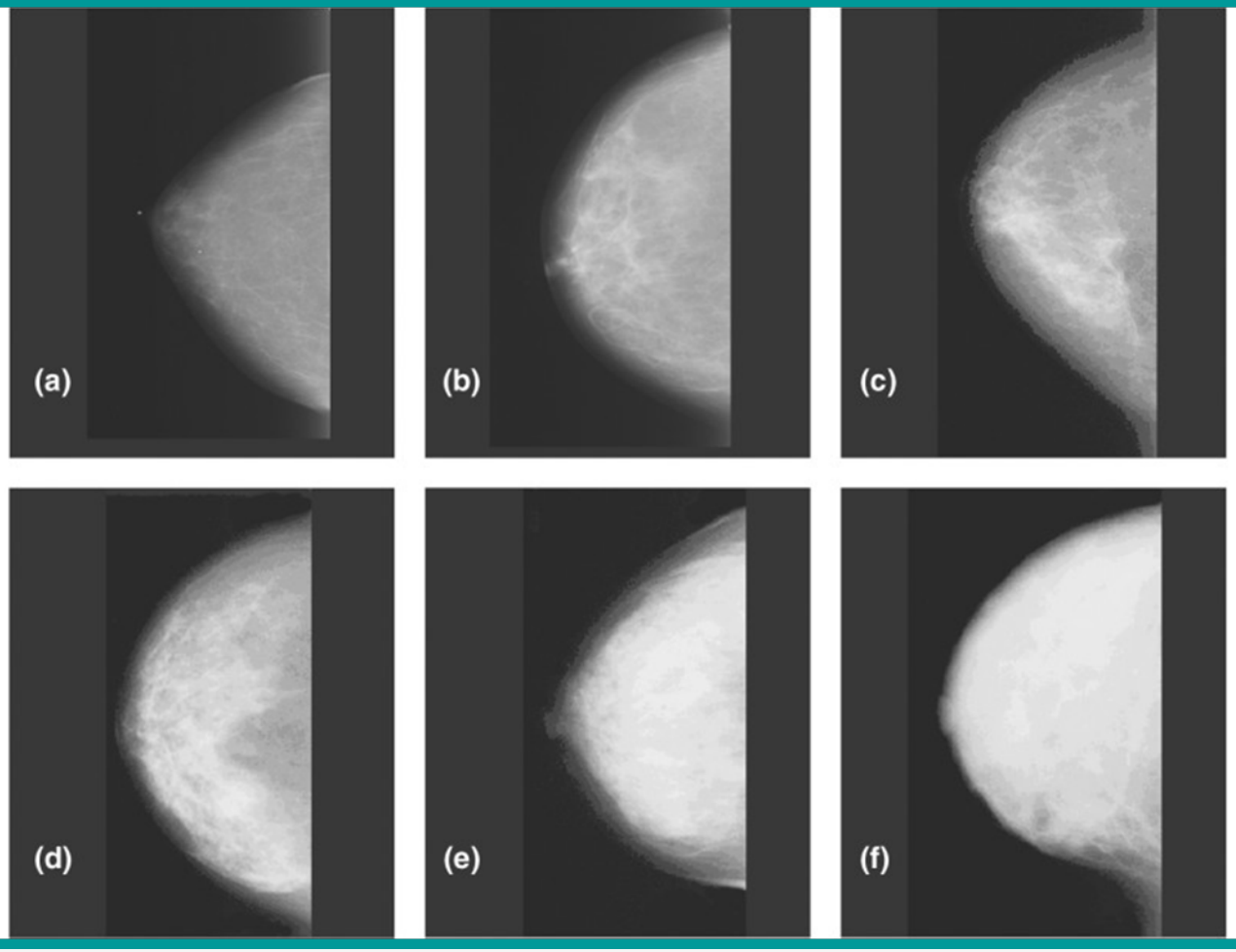


Figure 1. A six-category system for classifying mammographic density. The categories describe the fraction of fibroglandular tissue in the breast as judged by an observer and are: **(a)** 0, **(b)** <10%, **(c)** 10–25%, **(d)** 26–50%, **(e)** 51–75%, **(f)** >75%. Reproduced from [1] with permission from American Association for Cancer Research.

Breast Cancer Research 2008, 10:209doi:10.1186/bcr2102

Measures of Test Performance

Positive Predictive value

If a screening test is positive, what is the chance disease is really present?

Negative Predictive Value

If a screening test is negative, what is the chance disease is really absent?

Measures of Test Performance

Positive Predictive Value &

Negative Predictive Value

Depend on inherent performance characteristics of the screening test, AND how common the disease is in the screened population

Measures of Test Performance

		Disease State (Truth/Reality)	
		+	-
Screening Test	+	a	b
	-	c	d

Measures of Test Performance

		Disease State (Truth/Reality)	
		+	-
Screening Test	+	a true +	b false +
	-	c false -	d true -

Measures of Test Performance

		Disease State +	Disease State -
Screening Test	+	a	b
	-	c	d

Sensitivity = a / (a + c)

Specificity = d / (b + d)

Positive Predictive Value [PPV] = a / (a + b)

Negative Predictive Value [NPV] = d / (c + d)

Measures of Test Performance

		Disease State	
		+	-
Screening Test	+	a	b
	-	c	d

D_+ = true presence of disease

(a + c)

D_- = true absence of disease

(b + d)

S_+ = screening test positive

(a + b)

S_- = screening test negative

(c + d)

Measures of Test Performance

Screening Test	Disease State +	Disease State -
+	a	b
-	c	d

D_+ = true presence of disease
(a + c)

D_- = true absence of disease
(b + d)

S_+ = screening test positive
(a + b)

S_- = screening test negative
(c + d)

Bayes' theorem (probabilities)

Sensitivity = $P(S_+ \mid D_+)$

Specificity = $P(S_- \mid D_-)$

PPV = $P(D_+ \mid S_+)$

NPV = $P(D_- \mid S_-)$

Prevalence, i.e. prior probability = $P(D_+)$

Positive Predictive Value

Screening Test	Disease State +	Disease State -
+	a	b
-	c	d

$PPV = P(D_+ \mid S_+) = a / (a + b)$

- **Our screening test is positive - what is the chance it is "real"**

Depends on test sensitivity, specificity, and prevalence

Prevalence and Predictive Value

- **Assume screening test**
  80% sensitivity and 90% specificity

Disease Prevalence (per 1000)	Positive PV (%)	Negative PV (%)
1	0.8	99.98
1.5	1.2	99.97
3	2.4	99.93
10	7.5	99.78
100	47.1	97.59

Sensitivity and Predictive Value

- **Assume disease prevalence is constant at 3 per 1000**
  (also constant specificity of 90%)

Test Sensitivity (%)	Positive PV (%)	Negative PV (%)
45	1.3	99.82
60	1.8	99.87
80	2.4	99.93
95	2.8	99.98
99.5	2.9	99.99

Clinical Utility

- **Are there effective interventions?**
  - **Clinical disease screening (like cancer)**
  - **Genetic screening to identify individuals prone to disease or adverse events (BRCA1/2 testing to target screening and prevention, CYP450 screening for adverse drug events)**

Clinical Considerations

- **What is the cost (worry, invasive tests, complications of diagnostic tests, etc.) for a false positive test?**
  - **If high, want to have test with extremely high specificity, or tailor screening to population with high prevalence (prior probability)**
- **What is the cost of a false negative screening test?**
  - **If high, want to have test with extremely high sensitivity**

Genetic Screening and Diagnosis of Breast Cancer

- Understanding of genetics of breast cancer fairly advanced
- There are numerous clinical cancer screening tests available
- There are drastic but preventive measures available

Genetics of Breast Cancer

Major classes/representative genes	Prevalence	Rel Risk
Rare, high risk		
BRCA1 & *BRCA2*	3 per 1000	5 - 10
Rare, low/mod risk		
CHEK2, *PALB2*, *BRIP1*, *ATM*	~ 1%	2 - 2.5
Common, low risk		
mostly GWAS hits - *FGFR2*, *TNRC9*, *CASP8*, etc.	20% - 50%	1.1-1.3

Clinical Testing for Breast Cancer Genes

- *BRCA1*/*BRCA2*
  - Commercially available (~$4000)
  - Clinical validity and utility
  - Complex analytically, but extremely low false +
  - Non-zero but unknown false -

- Is there a role for testing for lower-penetrance gene mutations???
  - Analytical validity can/should be extremely high
  - Generally (!) not clinically available
    - *CHEK2*, direct-to-consumer testing using SNP chips
  - Clinical validity and utility less clear

Clinical Breast Cancer Screening and Prevention

- Self breast exam
- Clinical breast exam
- Ultrasound
- Mammography
- MRI

- Chemoprevention
- Risk-reducing surgery
- Oophorectomy

Predictive Value of Screening Mammography

- **Assume 80% sensitivity and 90% specificity, women age 40 and older**

Disease Prevalence (per 1000)	Positive PV (%)	Negative PV (%)
1.5 (FH -)	1.2	99.97
3 (FH +)	2.4	99.93

Predictive Value of Screening

Is Mammography Adequate for Screening Women with Inherited BRCA Mutations and Low Breast Density?

Rachel Z. Bigenwald,[1] Ellen Warner,[1] Anoma Gunasekara,[2] Kimberley A. Hill,[1] Petrina A. Causer,[5] Sandra J. Messner,[4] Andrea Eisen,[1,4] Donald B. Plewes,[2,6] Steven A. Narod,[7] Liying Zhang,[3] and Martin J. Yaffe[2,6]

Cancer Epidemiol Biomarkers Prev 2008;17(3). March 2008

- **Very high risk women (i.e., *BRCA1/BRCA2* mutation carriers)**

  -annual screening recommended starting age 25-35

  Prevalence ~~ 1 per 100 screen

Method (sensitivity)	Positive PV (%)	Negative PV (%)
Mammogram (~26%)	2.6	99.2
MRI (90%)	8.3	99.89

Clinical Breast Cancer Genetics

SPECIAL ARTICLE

Polygenes, Risk Prediction, and Targeted Prevention of Breast Cancer

Paul D.P. Pharoah, Ph.D., Antonis C. Antoniou, Ph.D., Douglas F. Easton, Ph.D., and Bruce A.J. Ponder, F.R.S.

N ENGL J MED 358;26 WWW.NEJM.ORG JUNE 26, 2008

Clinical Breast Cancer Genetics

Table 1. Established Common Breast-Cancer Susceptibility Alleles.*

dbSNP No.	Gene†	Chromosome	Risk-Allele Frequency‡	Relative Risk per Allele‡	Fraction of Total Variance in Risk Explained§	Population Attributable Risk§	Study
					%		
rs2981582	*FGFR2*	10q	0.38	1.26	1.7	19	Easton et al.,[26] Hunter et al.[27]
rs3803662	*TNRC9, LOC643714*	16q	0.25	1.20	0.9	10	Easton et al.[26]
rs889312	*MAP3K1*	5q	0.28	1.13	0.4	7	Easton et al.[26]
rs3817198	*LSP1*	11p	0.30	1.07	0.1	4	Easton et al.[26]
rs13281615	None known	8q	0.40	1.08	0.2	6	Easton et al.[26]
rs13387042	None known	2q	0.50	1.20	1.2	19	Stacey et al.[28]
rs1053485	*CASP8*	2q	0.86	1.13	0.3	20	Cox et al.[25]

* *CASP8* denotes caspase 8, dbSNP database of single-nucleotide polymorphisms, *FGFR2* the fibroblast growth factor receptor 2 gene, *LOC643714* a hypothetical protein LOC643714, *LSP1* lymphocyte-specific protein 1, *MAP3K1* mitogen-activated protein kinase kinase kinase 1, and *TNRC9* trinucleotide repeat containing 9.

† These genes are within the linkage-disequilibrium block or blocks defined by the associated variant and are plausible candidates for the causal gene.

‡ Values are from published data cited in the Study column.

§ See the Supplementary Appendix for details.

Clinical Impact of Identifying Genetic Risk Factors

- **Assuming we could reduce disease risk by 40% by identifying genetically susceptible individuals and implementing preventive measures:**

	Pop'n prev.	Case prev.	Pop'n reduction in disease burden
BRCA1 & *BRCA2*	0.003%	3%	0.7%
CHEK2:1100delC	0.7%	1.5%	0.7%
Common, OR=2	5%	9.5%	4%
FGFR2 SNP	38% maf	42% maf	2.6%

Clinical Impact of Identifying Genetic Risk Factors

- Take 7 known common, low-risk breast cancer alleles
- Assume multiplicative risk model and apply to UK population
- Plot population distribution of risk strata based on # of high-risk genotypes

UK Population Distribution of Risk Strata based on 7 Breast Cancer SNPs

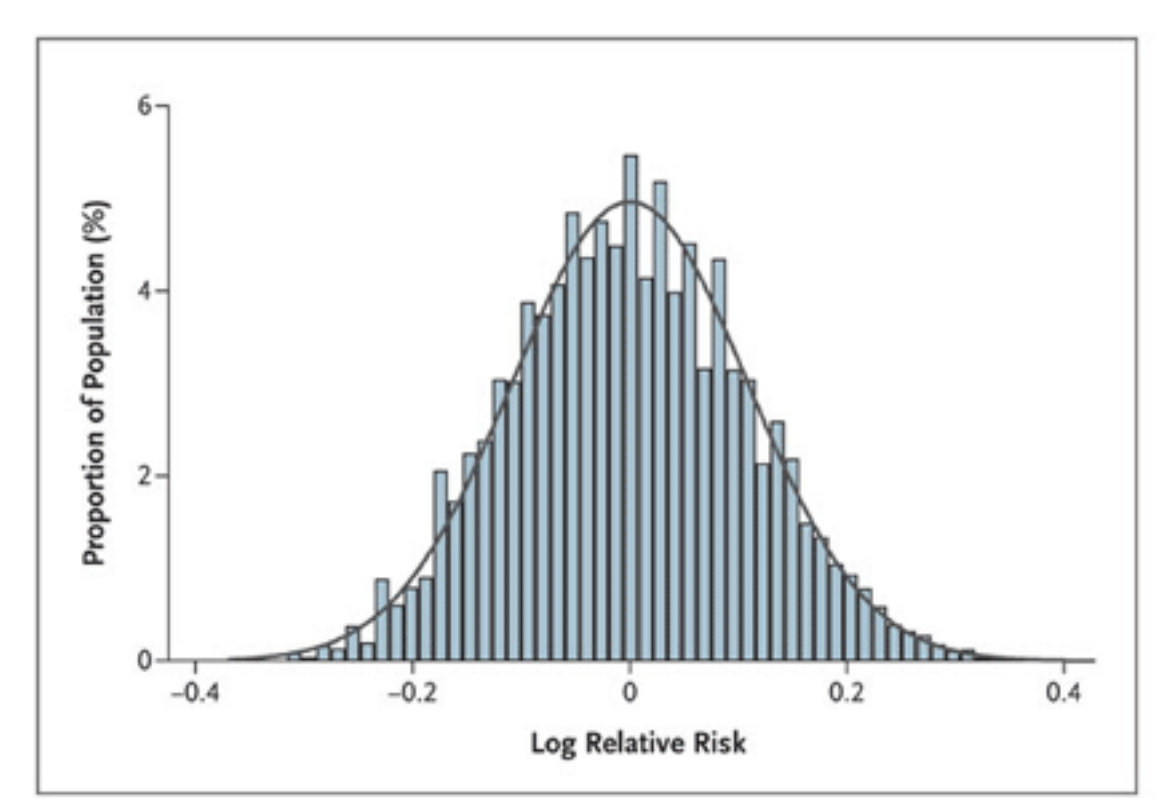
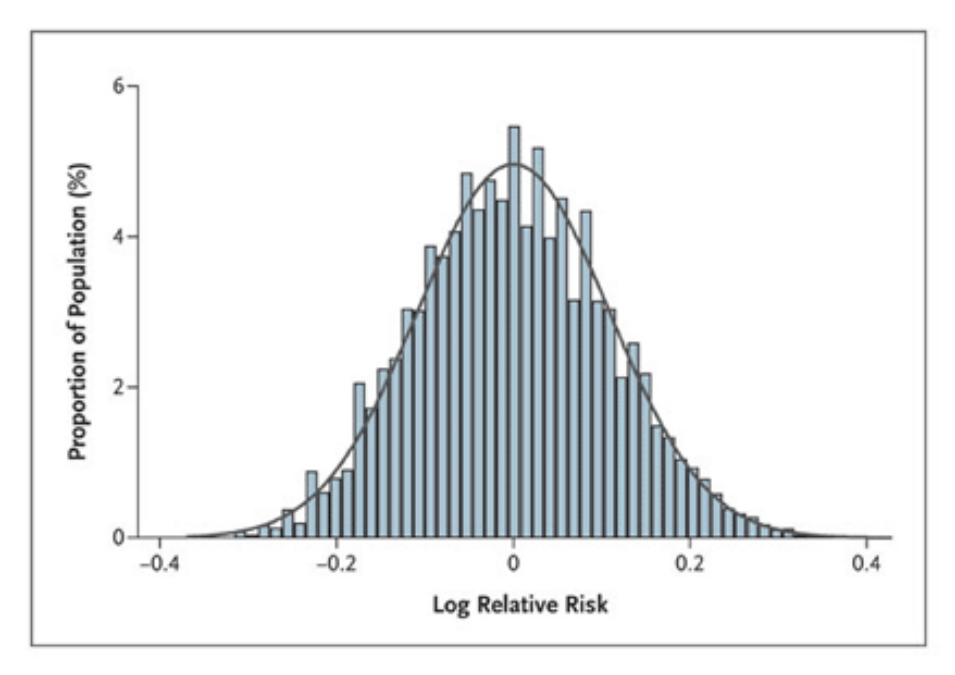


Figure 1. Distribution of Genetic Risk in the Population.
The log relative risk scale of −0.4 to 0.4 is equivalent to 0.4 to 2.5 on the relative risk scale.

Pharoah et al NEJM Summary



- **Based on average lifetime risk of 9.4% in UK, the extremes of the 7 SNP risk strata have rates of 4.2% to 23% (but very few people in the tails of this distribution)**

UK Population Breast Cancer Risk and Proportion of Cases Explained

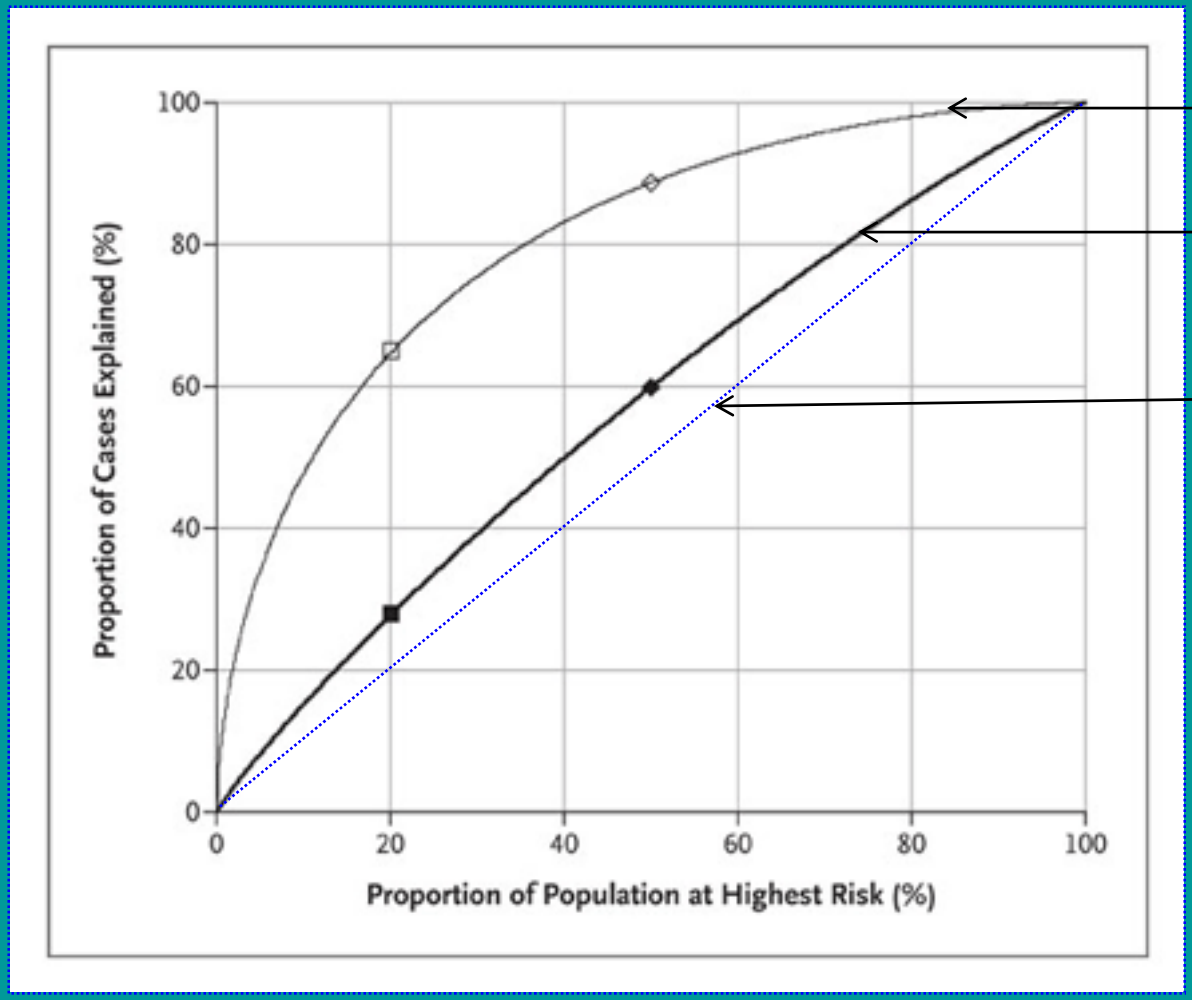
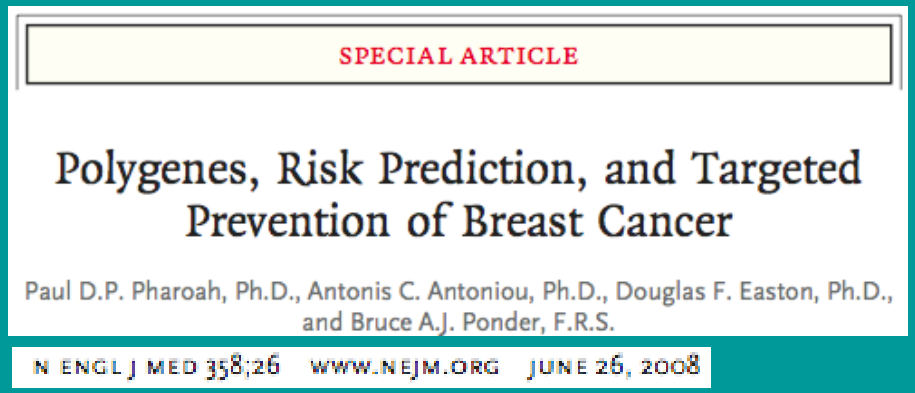


Figure 2. Proportion of Cases of Breast Cancer Explained by the Proportion of the Population at Highest Risk for the Disease.

Estimates based on currently known susceptibility alleles are indicated by the thick line. Estimates based on the best-case scenario, in which all possible breast-cancer susceptibility alleles are known, are indicated by the thin line.[6] The graph shows that the half of the population at highest risk for breast cancer on the basis of the genotype at seven known susceptibility loci accounts for 60% of all cases of breast cancer (solid diamond) and the 20% at the highest risk account for 28% of all cases (solid square). If all possible susceptibility alleles were known, the respective proportions, based on the genotype, would be 88% (open diamond) and 64% (open square).

Implications



SPECIAL ARTICLE

Polygenes, Risk Prediction, and Targeted Prevention of Breast Cancer

Paul D.P. Pharoah, Ph.D., Antonis C. Antoniou, Ph.D., Douglas F. Easton, Ph.D., and Bruce A.J. Ponder, F.R.S.

N ENGL J MED 358;26 WWW.NEJM.ORG JUNE 26, 2008

- **Routine screening begins at age 50, when the 10-year risk is 2.3%**
- **Screening could be tailored to identify genetic risk groups that meet this threshold, and screening initiated at different ages**

Pharoah et al NEJM Summary

Table 2. Absolute Risks of Breast Cancer According to Percentile of Population.*

Percentile of Population	Relative Risk	Lifetime Risk†	10-Yr Risk at 50 Yr of Age†	Age at Which 10-Yr Risk ≥2.3%
		%		*yr*
5	0.63	6.1	1.5	NA‡
10	0.69	6.7	1.6	NA‡
20	0.77	7.4	1.8	NA‡
40	0.90	8.6	2.1	53
60	1.03	9.7	2.4	49
80	1.20	11.0	2.7	45
90	1.35	12.0	3.0	43
95	1.49	14.0	3.4	41

* The relative risks are based on the risk distribution of seven known breast-cancer susceptibility loci.
† The absolute risks (lifetime risk and 10-year risk at 50 years of age) are estimated from the relative risks and age-specific breast-cancer incidence and all-cause mortality in England and Wales in 2004.
‡ NA denotes not applicable. The 10-year risk of breast cancer increases with age and peaks at approximately 60 years of age.[29] It then decreases because the mortality from other causes increases faster than the incidence of breast cancer. Thus, the maximum 10-year risk among some women is less than the 2.3% threshold.

Genetic Screening and Diagnosis

(Lecture 7)

1. Evaluating genetic and other screening tests involves conditional probabilities

2. In addition to considerations like test sensitivity and predictive value, many factors impact the clinical utility and acceptability of any screening test

Genetic Screening and Diagnosis

(Lecture 7)

3. Much of the inherited component of chronic disease is yet to be discovered

4. Clinical screening and interventions need to be tested specifically in genetically susceptible

5. Even complete knowledge of genetic susceptibility factors is not deterministic - i.e., many cases are due to non-genetic factors