National Cancer Institute

How Much Do SNPs Improve Models to Predict Breast Cancer Risk?

Public Health Genomics Seminar
January 28, 2009

Mitchell H. Gail

Biostatistics Branch
Division of Cancer Epidemiology and Genetics

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health

Outline

- Models to predict breast cancer risk
  - BCRAT (Gail model 2)
  - BCRATplus7
- Improvements from BCRATplus7
  - Discriminatory accuracy (AUC)
  - Deciding to take tamoxifen
  - Deciding to have a mammogram
  - Allocating scarce public health resources for mammography
  - Reclassification

Breast Cancer Risk Assessment Tool (BCRAT)

- **The NCI's BCRAT or "Gail Model 2"**
  - **Risk factors in BCRAT**
    - **Age**
    - **Age at first live birth**
    - **Age at menarche**
    - **Number of mother/sisters with breast cancer**
    - **Number of previous benign breast biopsies and whether atypical hyperplasia present on any**
  - **Well calibrated**
  - **Discriminatory accuracy modest**

SNPs Associated with Breast Cancer

Location	Disease Allele Frequency	Odds Ratio per Allele	Reference
FGFR2	0.38	1.26	1
TNRC9 (or TOX3)	0.25	1.20	1
MAP3K1	0.28	1.13	1
LSP1	0.30	1.07	1
CASP8	0.87	1.136	2
8q	0.40	1.08	1
2q35	0.497	1.20	3
		Geometric mean 1.15	

1. Easton et al., Nature 2007;447:1087-1095
2. Cox et al., Nature Genetics 2007;39:352-358
3. Stacey et al., Nature Genetics 2007;39:865-869

Key Assumptions

- **Hardy-Weinberg equilibrium**
- **Linkage equilibrium across SNPs**

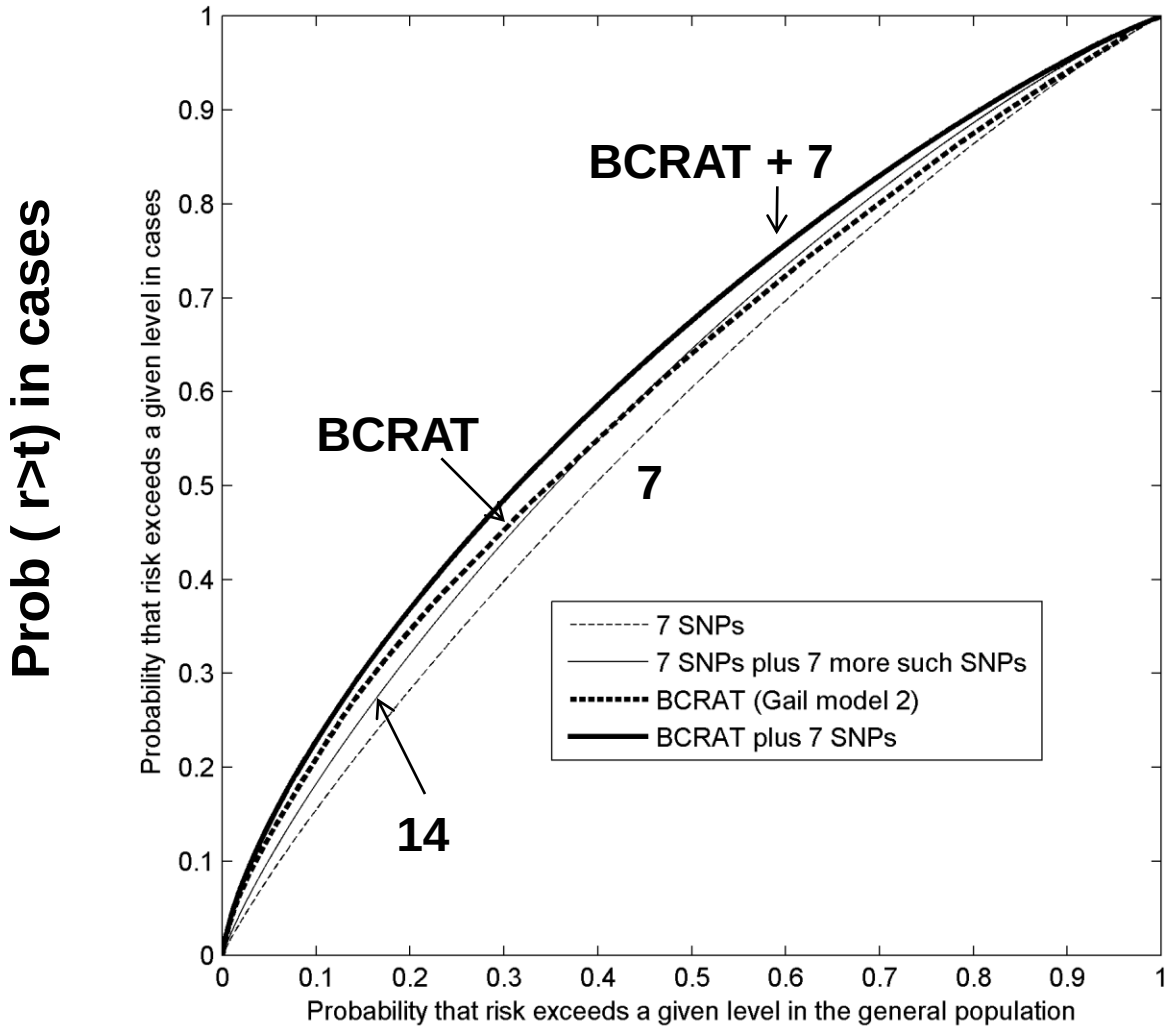
$$P(\mathbf{X}) = \prod_{i=1}^{7} p_i(X_i)$$

- **Additive effects of disease alleles**
- **Odds ratios multiply across SNPs**

$$rr(\mathbf{X}) = \prod_{i=1}^{7} (OR_i)^{X_i}$$

- **SNP ORs multiply BCRAT ORs**
- **SNPs independent of factors in BCRAT**

ROC-type Plots



Prob (r>t) in general population

Comparisons of Discriminatory Accuracy

Model	Age-specific AUC
7-SNPs	0.574
14-SNPs	0.604
BCRAT	0.607
BCRAT+ 7-SNPS	0.632

Decision to Take Tamoxifen in 100,000 Women Aged 50-59

Health Outcome	Relative Risk	# Cases If No Tamoxifen	# Cases If All Tamoxifen
Invasive Br. Ca.	0.51	246.6	125.8
Hip Fracture	0.55	101.6	55.9
Endometrial Ca.	4.01	81.4	326.4
Stroke	1.59	110	174.9
Pulmonary Emb.	3.01	50	150.5
Total		589.6	833.5

Threshold Risk *r** for Optimal Decision

Expected net benefit from tamoxifen for woman with BC risk *r*

$$r(1-0.51)+101.6(1-0.55)+81.4(1-4.01)+110.0(1-1.59)+50.0(1-3.01)$$
$$=0.49r-364.7.$$

Expected net benefit positive if *r* > 364.7/0.49=774.3 ≡ *r

Life-Threatening Events with Various Prevention Strategies

Strategy	Expected Life-Threatening Events
All get tamoxifen	833.5
None get tamoxifen	589.6
BCRAT > r*	588.2
BCRAT+7 SNPs > r*	587.8
Perfect Model	469.7

Percentage Improvement in Expected Events vs BCRAT

- **For women aged 50-59**
  - **BCRATplus7 0.07%**
  - **Perfect model 20.1%**
- **For women aged 40-49**
  - **BCRATplus7 0.81%**
  - **Perfect model 29.0%**

Losses in population screening to recommend mammography

Screening recommendation	Breast cancer present	No breast cancer
No mammography	C01=1271	C00=0
Mammography	C11=0.75 x 1271=953	C10=1

$$\text{threshold} \equiv r^* = 241.4 \text{x} 10^{-5} \left(\text{risk in women aged 50-54years}\right)$$

$$sens = \Pr(\text{estimated risk} > r^* \mid \text{detectable breast cancer})$$

$$spec = \Pr(\text{estimated risk} \leq r^* \mid \text{no detectable breast cancer})$$

$$\text{Expected Loss} =$$

$$C_{11}\mu(sens) + C_{01}\mu(1-sens) + C_{10}(1-\mu)(1-spec) + C_{11}(1-\mu)(spec)$$

Losses in population screening to recommend mammography

Screening recommendation	Breast cancer present	No breast cancer
No mammography	C01	C00=0
Mammography	C11=0.75 x C01	C10=1

Cost of mammogram in woman without breast cancer
= C = C10 - C00 = 1

Benefit of mammography in woman with breast cancer
= B = C01 – C11

Optimal Threshold = Prevalence* =p*= C/(C+B)

(Pauker & Kassirer, NEJM 1975)

Backcalculation of losses assuming prevalence in women aged 50-54 y, $p^*=313x10^{-5}$, is the ideal threshold.

C00=0 ; C10 = 1, the unit of costs

$$p^* = 313x10^{-5} = C/(C+B) = 1/(0.25C01+1)$$

$$C01 = 1271, \text{ and } C11 = 0.75C01 = 953$$

$$r^* = p^*/1.3 = 241.4 \text{ x } 10^{-5}$$

Expected Losses[1] for 3 Models

	BRCAT	BCRAT + 7 SNPs	Perfect Model
Sensitivity	0.476	0.549	1.0
Specificity	0.678	0.638	1.0
Expected loss	3.834	3.801	2.991
% improved	Baseline	0.86%	22.0%

1. **Expected losses computed for 50-54 year old women with average BC prevalence of**

$$\mu = 1.3\text{x}241.4\,\text{x}10^{-5} = 313\text{x}10^{-5}.$$

Allocating Mammograms When Only Enough Money for Half the Population

Screen with	Proportion of lives saved compared to giving mammograms to all women	% Improvement
No Screen	**0.500**	
BCRAT	**0.632**	**Baseline**
BCRATplus7	**0.667**	**5.5%**

Five-year risk from BCRAT	Five-year risk from BCRATplus7					
	<1.0%	1.0 to <1.5%	1.5 to <2.0%	2.0 to <2.5%	≥2.5%	Total
<1.0%	29.4	8.0	0.6	0.0	0.0	38.0
1.0 to <1.5%	15.4	21.6	6.0	0.9	0.1	44.0
1.5 to <2.0%	0.2	3.0	3.7	1.9	0.9	9.7
2.0 to <2.5%	0.0	0.6	1.8	1.6	1.3	5.3
≥2.5%	0.0	0.0	0.2	0.4	2.3	2.9
Total	45.0	33.2	12.3	4.8	4.6	99.9

Cross-classification in Percent at the Threshold of 2%

		BCRATplus7		Total
		<2%	≥2%	
BCRAT	<2%	87.9	3.8	91.7
	≥2%	2.6	5.6	8.2
	Total	90.5	9.4	99.9

Conclusions

- **Very modest public health improvements** from BCRATplus7 for
  - Discriminatory accuracy (AUC) (4.1%)
  - Deciding whether to take tamoxifen (0.1% or 0.8%)
  - Deciding to have mammogram (0.8% or 0.1%)
  - Allocating scarce mammogram resources (5.5%)
- **Reclassification versus BCRAT** useful for individuals **if** BCRATplus7 is well calibrated
- BCRATplus7 **needs to be validated** in independent **cohort data** on individuals

Conclusions (continued)

- Usefulness of SNPs depends on the application, validity of model, and costs
- To achieve high discriminatory accuracy (AUC=0.8) would require hundreds of SNPs

References

- Fisher et al. JNCI 1998; 90:1371-1388
- Gail, JNCI 2008;100:1037-1041
- Gail & Pfeiffer, Biostatistics 2005;6:227-239
- Gail, Costantino et al., JNCI 1999;91:1829-1846
- Pharoah et al., Nat Genet 2002;31:33-36
- Pharoah et al., NEJM 2008;358:2796-2803
- Easton et al., Nature 2007;447:1087-1095
- Stacey et al., Nat Genet 2007;39:865- 869
- Cox et al., Nat Genet 2007;39:352-358