Potential Usefulness of Single Nucleotide Polymorphisms to Identify Persons at High Cancer Risk: An Evaluation of Seven Common Cancers

Ju-Hyun Park, Mitchell H. Gail, Mark H. Greene, and Nilanjan Chatterjee

All authors: National Cancer Institute, Bethesda, MD.

Submitted October 27, 2011; accepted February 28, 2012; published online ahead of print at www.jco.org on May 14, 2012.

Supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Bethesda, MD.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Mitchell H. Gail, MD, PhD, National Cancer Institute, 6120 Executive Blvd, Rockville, MD 20852-7244; e-mail: gailm@mail.nih.gov.

© 2012 by American Society of Clinical Oncology

0732-183X/12/3017-2157/\$20.00

DOI: 10.1200/JCO.2011.40.1943

A B S T R A C T

Purpose

To estimate the likely number and predictive strength of cancer-associated single nucleotide polymorphisms (SNPs) that are yet to be discovered for seven common cancers.

Methods

From the statistical power of published genome-wide association studies, we estimated the number of undetected susceptibility loci and the distribution of effect sizes for all cancers. Assuming a log-normal model for risks and multiplicative relative risks for SNPs, family history (FH), and known risk factors, we estimated the area under the receiver operating characteristic curve (AUC) and the proportion of patients with risks above risk thresholds for screening. From additional prevalence data, we estimated the positive predictive value and the ratio of non-patient cases to patient cases (false-positive ratio) for various risk thresholds.

Results

Age-specific discriminatory accuracy (AUC) for models including FH and foreseeable SNPs ranged from 0.575 for ovarian cancer to 0.694 for prostate cancer. The proportions of patients in the highest decile of population risk ranged from 16.2% for ovarian cancer to 29.4% for prostate cancer. The corresponding false-positive ratios were 241 for colorectal cancer, 610 for ovarian cancer, and 138 or 280 for breast cancer in women age 50 to 54 or 40 to 44 years, respectively.

Conclusion

Foreseeable common SNP discoveries may not permit identification of small subsets of patients that contain most cancers. Usefulness of screening could be diminished by many false positives. Additional strong risk factors are needed to improve risk discrimination.

J Clin Oncol 30:2157-2162. © 2012 by American Society of Clinical Oncology

INTRODUCTION

Many models, such as the Breast Cancer Risk Assessment Tool (BCRAT),¹ are available to project cancer risk over a defined time interval. BCRAT has been criticized because it has limited discriminatory accuracy,² often measured as the probability (area under the curve [AUC]) that a person with disease will have a higher risk than a person without disease. High discriminatory accuracy is required for deciding who should receive an intervention.³ Otherwise, if one focuses exclusively on those at highest risk, one will only offer the intervention to a small proportion of those who have or will develop disease.⁴⁻⁶

Recent molecular and genetic epidemiologic studies have raised hopes for developing risk models with high discriminatory performance. Cancer-associated single nucleotide polymorphisms (SNPs) have been discovered through genome-wide association studies (GWASs). Although individual SNPs

confer only modest risk (typical odds ratio per risk allele < 1.3), it has been speculated that cumulatively, they may provide important risk stratification.⁷ However, recent studies have shown that established susceptibility SNPs have low discriminatory power (AUC approximately 0.60 for breast and from 0.64 to 0.67 for prostate cancers).⁸⁻¹⁴ It is hoped that large meta-analyses of existing and new GWASs will uncover additional susceptibility SNPs and improve discriminatory accuracy.

We evaluated the potential utility and limits of various cancer risk models based on known epidemiologic risk factors and on additional susceptibility SNPs that will be detected with large, but realistic, GWAS sample sizes. We extended previously published methods¹⁵ to estimate the numbers and effect sizes of current and yet-to-be-discovered susceptibility SNPs for cancers of the breast, prostate, colorectum, ovary, bladder, pancreas, and brain (glioma). We assessed the role of these SNPs by

evaluating new and previously used criteria for screening applications. Our work is new in that it attempts to determine the usefulness of currently known and foreseeable susceptibility SNPs for risk-based screening for common (breast, prostate, colorectum, bladder, pancreas) and uncommon (ovarian, glioma) cancers.

METHODS

Estimating the Number of Foreseeable SNPs and Their Distribution of Effect Sizes

We extended our previous methodology¹⁵ for estimating the number of underlying susceptibility SNPs and their distribution of effect sizes. The major steps were: (1) identifying the largest GWAS, termed the current study, for each of the cancers and listing observed susceptibility SNPs that have been discovered through these studies; (2) from the design of the discovery studies, computing the power to detect SNPs with given effect sizes; (3) estimating an underlying common density of effect sizes across different cancers by weighting each observed SNP by the inverse of its power of detection; and (4) estimating the total number of underlying susceptibility SNPs for each cancer from the common estimated effect-size density and the observed numbers of established SNPs for that cancer. We also refer to established susceptibility SNPs as known susceptibility SNPs. Details are provided in the Data Supplement, but we comment further here.

For each cancer, we identified the largest GWAS to date and constructed a list of observed susceptibility SNPs that could be considered to have been detected by this study. All independent SNPs that reached genome-wide significance according to specified criteria for these studies were included in the list of known susceptibility SNPs. Some studies used multistage designs and did not include follow-up of previously established susceptibility SNPs beyond the first stage. We included such previously established SNPs in our list if they reached the required threshold for follow-up in the first stage of the current study, on the assumption that these SNPs would have reached genome-wide significance had they undergone follow-up like all other SNPs meeting the same criterion. To minimize bias from the winner's curse, we estimated disease odds ratios per allele by excluding discovery stage data whenever replication phase data were available. Otherwise, we corrected for possible bias statistically.¹⁶

We extended the approach of Park et al¹⁵ for estimating the number of underlying susceptibility SNPs with various effect sizes from the set of established susceptibility SNPs. We defined effect size for a susceptibility SNP¹⁵ under Hardy-Weinberg equilibrium as its contribution to the genetic variance of the trait, namely $gv = 2\beta^2f(1 - f)$, where β is the log-odds ratio coefficient under a linear trend model in risk alleles and f is the risk allele

frequency. The power to detect a susceptibility SNP depends on β and f only through gv .¹⁵ We assumed the underlying density function for gv would be the same for the different cancers, but the total number of susceptibility SNPs could differ among cancers. We obtained nonparametric estimates of the underlying common density using a weighted kernel smoothing method¹⁵ that weighted each observed SNP with a given effect-size gv by the inverse of its power of detection. Thus, each observed SNP was multiplied to represent the population of underlying SNPs with similar effect size, leading to an estimate of the density of effect sizes. For each cancer, we used this density estimate and the power to detect a given effect size in the corresponding current study to estimate the total number of underlying susceptibility SNPs by equating the observed and expected numbers of SNP discoveries (Data Supplement).

We used the estimated numbers of underlying susceptibility loci and the effect-size distribution to project the number of additional SNP discoveries from future larger studies (Data Supplement). To define realistic limits for future discoveries, we considered future studies twice or three times as large as the current GWAS for a particular cancer. Hereafter, we use the term foreseeable SNPs to denote the sum of known SNPs plus the SNPs expected to be detected from a GWAS three times as large as the largest currently available GWAS. For current studies that used multistage designs, we defined an effective sample size that would correspond to a single-stage study but that was expected to lead to the same number of discoveries as the original multistage design.

Criteria for Evaluation Risk Models With Susceptibility SNPs

From the numbers of foreseeable SNPs with different effect sizes, we estimated various measures of utility for risk models based on a log-normal distribution of risk. We assumed that the logarithm of risk was a linear combination of the numbers of risk alleles at each SNP locus. Hence, the distribution of risks in patient cases and controls could be characterized by the total genetic variance explained by the susceptibility SNPs. Using the number of foreseeable SNPs with different effect sizes, we estimated the AUC that plots the proportion of patients above a risk threshold against the proportion of the general population above that threshold, as the threshold varied.³ This area was the probability that a randomly selected patient would have a larger risk than a randomly selected member of the general population. For a rare disease or for incidence of a common cancer on a short time interval, this AUC was nearly equal³ to the area under the receiver operator characteristic (ROC) curve.¹⁷ Likewise, we estimated the proportion of case patients with risks above the $(1 - p)$ th percentile of risk in the general population, termed the "proportion of cases followed" (ie, PCF_p) by Pfeiffer et al⁵ (Data Supplement). To assess the additional contribution of known epidemiologic risk factors, we again assumed a log-normal model for total risk and no gene-environment interaction on the log risk scale. Under such a model, the total variance of the risk score could be decomposed into a component

Table 1. Susceptibility SNPs and Associated GV From Doubling and Tripling Size of Largest Available GWAS and Risk Information for FH

Cancer Site	Effective Sample Size of Largest GWAS (No.)*	FH†		Known SNPs		Foreseeable SNPs Discovered With GWAS Sample Size Doubled		Foreseeable SNPs Discovered With GWAS Sample Size Tripled	
		Prevalence (%)	RR	No.	GV‡	Estimated No.§	GV‡	Estimated No.§	GV‡
Breast	18,163	10.9	1.8	18	0.126	39	0.173	70	0.239
Prostate	29,262	7.0	2.7	29	0.286	66	0.363	93	0.418
Colorectum	16,195	5.1	2.25	14	0.085	29	0.120	54	0.175
Ovary	17,591	0.9	3.1	4	0.041	6	0.046	13	0.059
Bladder	13,364	2.9	2	10	0.118	16	0.137	29	0.170
Glioma	6,431	0.7	1.77	5	0.121	10	0.155	18	0.191
Pancreas	7,784	1.9	2.93	4	0.074	9	0.099	17	0.129

Abbreviations: FH, family history; GV, genetic variance; GWAS, genome-wide association study; RR, relative risk association with positive FH; SNP, single nucleotide polymorphism.

*References to these largest GWASs are included in the Data Supplement.

†References for the prevalence of a positive FH (none v at least one affected first-degree relative) and the associated RR are included in the Data Supplement.

‡GV is the sum of individual SNP gvs, where for each $gv = 2\beta^2f(1 - f)$, β is the log odds per risk allele and f is the risk allele frequency.

§Estimated No. is the sum of the No. of known SNPs plus the No. of SNPs expected to be newly discovered from the future GWAS.

Table 2. Estimated AUCs Based on Known SNPs, Foreseeable SNPs, and Models With SNPs and Epidemiologic Risk Factors

Cancer Site	AUC				
	FH Only	Known SNPs	Foreseeable SNPs	FH and Known SNPs	FH and Foreseeable SNPs
Breast	0.536	0.599	0.635	0.613	0.646
Prostate	0.549	0.647	0.676	0.668	0.694
Colorectum	0.528	0.582	0.616	0.598	0.629
Ovary	0.509	0.557	0.568	0.564	0.575
Bladder	0.514	0.596	0.615	0.602	0.620
Glioma	0.503	0.597	0.621	0.598	0.622
Pancreas	0.517	0.576	0.600	0.588	0.610

Abbreviations: AUC, area under the curve; FH, family history; SNP, single nucleotide polymorphism.

associated with the SNPs and a component associated with the epidemiologic factors. We estimated the variance associated with epidemiologic risk factors by inverting AUC estimates obtained from known risk models such as BCRAT. Because the log-normal calculations may not have been accurate for evaluating influence of a single strong binary risk factor such as family history (FH), we used exact calculations (Data Supplement).

For colorectal,¹⁸ ovarian,¹⁹ and breast cancers,²⁰ we used data on the prevalence of previously undetected cancer that is detectable by a screening protocol. Using these prevalences, we computed the positive predictive value (PPV_p) and related criteria. We assumed that the risk of prevalent disease was proportional to the risk of disease incidence, because most available risk models are for disease incidence, and because prevalence and incidence are proportional in the steady state. Data support this assumption for breast cancer.²⁰ If we were to screen the members of the general population with risks above the $(1 - p)$ th percentile of risk, ξ_{1-p} , then PPV_p would be the proportion of individuals with risks above ξ_{1-p} who had screen-detectable cancer (prevalent cases). We defined the false-positive ratio, $(1 - \text{PPV}_p)/\text{PPV}_p$, as the ratio of non-patient cases to patient cases among those with risks above the threshold ξ_{1-p} . A high false-positive ratio may indicate that the cost of screening healthy individuals outweighs potential benefit.

Although we have defined PCF_p, PPV_p, and the false-positive ratio for detecting prevalent cancer, these quantities also apply to prospective studies of disease incidence. In that context, PPV_p is the proportion of the population with risks above the $(1 - p)$ th percentile of risk in the general population, ξ_{1-p} , who will develop incident cancer over a defined time interval. Other quantities were defined analogously. In our analyses, the AUC, PCF_p, PPV_p, and false-positive ratio pertained to people with the same age but varying SNP genotypes and epidemiologic factors. Thus, these quantities describe the ability of SNPs and other factors except age to differentiate risk.

RESULTS

Foreseeable Numbers of SNPs and Genetic Variance

The number of previously established SNPs ranged from four for ovarian and pancreatic cancers to 29 for prostate cancer (Table 1). Tripling the size of the GWAS would result in more than three times as many foreseeable susceptibility SNPs for all but one of these cancers, but the genetic variance explained would increase by approximately two-fold or less, because the effect sizes of new susceptibility SNPs detected by the larger GWAS would be smaller than those in the current GWAS.¹⁵ This is because virtually all the SNPs with large effect sizes have already been detected by the current studies for these cancers, which have near-perfect power to detect such SNPs.

AUC

The AUC for FH alone ranged from 0.503 for glioma to 0.549 for prostate cancer (Table 2). The low AUC values for FH could have resulted from low relative risks, from low prevalence of positive FH, as in glioma and ovarian cancer (Table 1), or from the fact that FH is dichotomous. The AUCs for known susceptibility SNPs exceeded those from FH for each cancer but were nonetheless less than 0.6, except for prostate cancer, with an AUC of 0.647. Tripling the GWAS sample sizes led to AUCs of 0.6 or more, except for ovarian cancer. Adding FH to the foreseeable susceptibility SNPs led to AUCs ranging

Table 3. PCF_p in the 10% and 30% of the Population at Highest Risk

Cancer Site	Top 10%*					Top 30%†				
	Known SNPs	Foreseeable SNPs	FH and Known SNPs	FH and Foreseeable SNPs	Epidemiologic Risk Factors and Foreseeable SNPs	Known SNPs	Foreseeable SNPs	FH and Known SNPs	FH and Foreseeable SNPs	Epidemiologic Risk Factors and Foreseeable SNPs
Breast	0.177	0.214	0.197	0.230	0.255	0.433	0.486	0.456	0.503	0.539
Prostate	0.228	0.263	0.264	0.294		0.504	0.549	0.537	0.576	
Colorectum	0.161	0.194	0.190	0.215	0.240	0.408	0.458	0.434	0.478	0.520
Ovary	0.140	0.149	0.153	0.162		0.373	0.389	0.384	0.399	
Bladder	0.174	0.192	0.185	0.201	0.332	0.428	0.455	0.438	0.464	0.627
Glioma	0.175	0.199	0.177	0.201		0.430	0.465	0.431	0.467	
Pancreas	0.156	0.178	0.178	0.197		0.400	0.434	0.418	0.450	

Abbreviations: FH, family history; PCF_p, proportion of cases followed; SNP, single nucleotide polymorphism.

*The proportion of the population at highest risk is $p = .1$.

†The proportion of the population at highest risk is $p = .3$.

Table 4. PPV and False-Positive Ratio for 10% of the Population at Highest Risk

Cancer Site	Prevalence × 10 ⁴ *	Known SNPs		Foreseeable SNPs		FH and Known SNPs		FH and Foreseeable SNPs	
		PPV × 10 ⁴	False-Positive Ratio	PPV × 10 ⁴	False-Positive Ratio	PPV × 10 ⁴	False-Positive Ratio	PPV × 10 ⁴	False-Positive Ratio
Colorectum	29	47	213	56	177	55	181	62	159
Ovary	10	14	706	15	662	15	644	16	610
Breast									
Age 50 to 54 years	31	55	180	67	148	62	161	72	138
Age 40 to 44 years	16	27	364	33	300	31	326	36	280

Abbreviations: FH, family history; PPV, positive predictive value; SNP, single nucleotide polymorphism.

*Prevalence was estimated from studies by Weissfeld et al¹⁸ for colorectal cancer, Buys et al¹⁹ for ovarian cancer, and Gail²⁰ for breast cancer.

from 0.575 for ovarian cancer to 0.694 for prostate cancer. For breast,¹ colorectal,²¹ and bladder cancers,²² we found validated epidemiologic models with more elaborate FH data and with other risk factors. Combining these epidemiologic risk models with foreseeable SNPs yielded AUCs of 0.670 for breast, 0.658 for colorectal, and 0.726 for bladder cancers.

PCF_p With Screening or Preventive Intervention

Using the previous risk models, one can identify the 10% of the population with risks above the 90th percentile of risk in the general population, $\xi_{1-p} = \xi_{0.9}$. If risk were highly concentrated, the top 10% would have a large PCF_p (PCF_{0.1}). On the basis of known SNPs, the PCF_{0.1} values ranged from 0.140 for ovarian cancer to 0.228 for prostate cancer (Table 3). For foreseeable SNPs, the values ranged from 0.149 for ovarian cancer to 0.263 for prostate cancer. Adding FH to foreseeable SNPs resulted in PCF_{0.1} values ranging from 0.162 for ovarian cancer to 0.294 for prostate cancer. Thus, even for prostate cancer, fewer than 30% of the men developing or having cancer in a given age stratum would be identified for preventive intervention or screening. The PCF_{0.1} values were somewhat higher for breast, colorectal, and bladder cancers when epidemiologic risk models superior to FH alone were combined with foreseeable SNPs, but even for bladder cancer, fewer than 35% of patients would be identified for preventive intervention or screening (Table 3). If one were to lower the intervention or screening threshold to include the 30% of the population at highest risk, then 40% to 63% of people developing cancer would exceed this threshold for models based on foreseeable SNPs combined with either FH or with more discriminating epidemiologic models (Table 3).

Positive Predictive Values and False-Positive Ratios

For colorectal, ovarian, and breast cancers, we calculated the PPV_p and false-positive ratio for a threshold that targeted the 10% of the population at highest risk (Table 4) or the 30% of the population at highest risk (Table 5). PPV_{0.1} increased with disease prevalence, but even in the model with FH and foreseeable SNPs, PPV_{0.1}s ranged only from 0.0016 to 0.0072 (Table 4). Thus, the vast majority of patients targeted for screening would not have disease, as reflected in high false-positive ratios. For breast cancer in women age 50 to 54 years, 138 would require screening to detect one breast cancer. This ratio was 610 for ovarian cancer, which has lower prevalence.

The PPV_{0.3}s were smaller, the false-positive ratios were larger, and the screening ratios were much larger if the top 30% of the population with highest risk were to be screened (Table 5), instead of the top 10% (Table 4). For example, for breast cancer in women age 50 to 54 years, and based on the model with FH and foreseeable SNPs, the PPV_{0.3} was 0.0052 (Table 5) instead 0.0072 (Table 4); the false-positive ratio was 190 instead of 138.

DISCUSSION

Despite the discovery of many cancer susceptibility SNPs, their utility for risk prediction has been limited.^{8,11-14,20,23} Recent meta-analyses²⁴⁻²⁶ have discovered many additional loci, and ongoing consortia will likely discover additional cancer susceptibility SNPs, raising the possibility of enhanced risk prediction. We extended previous methods¹⁵ and used existing and new criteria to evaluate the usefulness of foreseeable SNPs for selecting high-risk patients

Table 5. PPV and False-Positive Ratio for 30% of the Population at Highest Risk

Cancer Site	Prevalence × 10 ⁴ *	Known SNPs		Foreseeable SNPs		FH and Known SNPs		FH and Foreseeable SNPs	
		PPV × 10 ⁴	False-Positive Ratio	PPV × 10 ⁴	False-Positive Ratio	PPV × 10 ⁴	False-Positive Ratio	PPV × 10 ⁴	False-Positive Ratio
Colorectum	29	39	253	44	225	42	238	46	216
Ovary	10	13	795	13	763	13	773	13	744
Breast									
Age 50 to 54 years	31	45	221	51	196	48	209	52	190
Age 40 to 44 years	16	22	446	25	397	24	424	26	384

Abbreviations: FH, family history; PPV, positive predictive value; SNP, single nucleotide polymorphism.

*Prevalence was estimated from studies by Weissfeld et al¹⁸ for colorectal cancer, Buys et al¹⁹ for ovarian cancer, and Gail²⁰ for breast cancer.

for screening or other interventions. The SNPs that will be discovered will tend to have small effect sizes. Even the most optimistic foreseeable models had modest discriminatory accuracy, high false-positive rates, and low cancer detection probabilities (PCF_p) for most risk-based interventions.

Pashayan et al¹⁰ compared screening based on age thresholds alone with using a risk model based on age and SNPs. Those with 10-year prostate cancer risk of 2% or greater and with 10-year breast cancer risk of 2.5% or greater were targeted for screening. We agree with their broad conclusion that adding risk factors to a model based on age alone improves discriminatory accuracy and performance of the screening program. Our results (PCF_p in Table 3) were not comparable to results listed in their Tables 2 and 3, because we considered the usefulness of SNPs for a population of a given age, whereas Pashayan et al compared age plus SNPs with age alone. Pashayan et al estimated that if one screened 49,798 of 100,000 women based on a risk model with SNPs and age, one would detect 172 of the 236 breast cancers, for a yield of $172/236 = 0.729$. We calculated $PCF_{0.49798} = 0.637$. Because the AUC values for SNPs are nearly identical in these analyses, the difference $0.729 - 0.637 = 0.092$ reflects the wide age range (35 to 79 years) in the population that Pashayan et al studied.

Our calculations indicated large numbers of false-positive and false-negative screening results, even for models that combined epidemiologic information and foreseeable SNPs (Tables 3 to 5). Even if we were to screen 30% of the population, approximately half or more of the patient cases in a given age group would be missed. High false-positive ratios indicate that large numbers of healthy people would require screening to detect one patient case.

The utility of risk-based screening depends heavily on potential costs from screening healthy people as well as on potential benefits from screening those found to have disease. One example of the former costs is the need to evaluate false-positive mammograms with further clinical testing. In the case of prostate cancer, the benefits of early detection are unproven.^{27,28} The utility of risk models is smaller for less common cancers, such as cancers of the ovary, pancreas, and brain, because the positive predictive value is lower and the false-positive ratio higher (Tables 4 and 5).

The clinical utility of a model for implementing a risk-based intervention strategy depends on the losses associated with various decisions and disease states.^{3,29-31} It can be shown (Data Supplement) that the approximate expected loss decreases linearly in PCF_p and PPV_p and increases with $(1 + \text{false-positive ratio})$. Thus, if loss values were specified, the metrics we studied could be used to assess expected loss.

We assumed that the effects of SNPs and epidemiologic risk factors acted multiplicatively. Although it is possible that models incorporating gene-gene and gene-environment interactions may enhance discriminatory power, there is little evidence for such interactions.^{13,23,32} We may have overestimated the predictive performance of models that included both SNPs and FH, because we assumed that SNPs were uncorrelated with FH. Such bias is probably small, because foreseeable SNPs typically explain less than 20% of cancer heritability.¹⁵ Log-normal models approximate the distribution of risk well when log risk is determined by many risk factors,⁶ such as 10 or more SNPs, but the approximation is less adequate if the number of factors is small and/or a single factor has a large effect. That is why we used an exact approach to combine the effects of family history with SNPs.

GWAS discoveries help elucidate pathways for cancers and other complex diseases, but care is needed to assess their utility in public health applications. Our results for selected cancers indicate that foreseeable SNPs are unlikely to yield high discriminatory accuracy and identify small subgroups of individuals that contain most patient cases. Statistical approaches that incorporate gene-gene or gene-environment interactions or include susceptibility SNPs that do not meet stringent genome-wide significance thresholds^{33,34} may improve discriminatory performance. New detection platforms hold promise and may reveal SNPs not covered by current technologies and other genetic variants, such as rare SNPs and copy-number variants.^{7,23,35} From whatever source, strong risk factors are needed to improve discriminatory accuracy for risk-based screening and other risk-based interventions.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: Ju-Hyun Park, Mitchell H. Gail, Nilanjan Chatterjee

Administrative support: Nilanjan Chatterjee

Collection and assembly of data: Ju-Hyun Park, Mitchell H. Gail, Nilanjan Chatterjee

Data analysis and interpretation: All authors

Manuscript writing: All authors

Final approval of manuscript: All authors

REFERENCES

- Costantino JP, Gail MH, Pee D, et al: Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst* 91:1541-1548, 1999
- Rockhill B, Spiegelman D, Byrne C, et al: Validation of the Gail et al model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst* 93:358-366, 2001
- Gail MH, Pfeiffer RM: On criteria for evaluating models of absolute risk. *Biostatistics* 6:227-239, 2005
- Gail MH: Personalized estimates of breast cancer risk in clinical practice and public health. *Stat Med* 30:1090-1104, 2011
- Pfeiffer RM, Gail MH: Two criteria for evaluating risk prediction models. *Biometrics* 2010:1541-0420, 2010
- Pharoah PD, Antoniou A, Bobrow M, et al: Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet* 31:33-36, 2002
- van Zitteren M van der Net JB, Kundu S, et al: Genome-based prediction of breast cancer risk in the general population: A modeling study based on meta-analyses of genetic associations. *Cancer Epidemiol Biomarkers Prev* 20:9-22, 2011
- Gail MH: Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. *J Natl Cancer Inst* 100:1037-1041, 2008
- Johansson M, Holmström B, Hinchliffe SR, et al: Combining 33 genetic variants with prostate-specific antigen for prediction of prostate cancer: Longitudinal study. *Int J Cancer* 130:129-137, 2012
- Pashayan N, Duffy SW, Chowdhury S, et al: Polygenic susceptibility to prostate and breast cancer: Implications for personalised screening. *Br J Cancer* 104:1656-1663, 2011
- Pharoah PD, Antoniou AC, Easton DF, et al: Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med* 358:2796-2803, 2008
- Salinas CA, Koopmeiners JS, Kwon EM, et al: Clinical utility of five genetic variants for predicting prostate cancer risk and mortality. *Prostate* 69:363-372, 2009
- Wacholder S, Hartge P, Prentice R, et al: Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* 362:986-993, 2010

14. Zheng SL, Sun J, Wiklund F, et al: Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 358:910-919, 2008
15. Park JH, Wacholder S, Gail MH, et al: Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* 42:570-575, 2010
16. Ghosh A, Zou F, Wright FA: Estimating odds ratios in genome scans: An approximate conditional likelihood approach. *Am J Hum Genet* 82:1064-1074, 2008
17. Pepe MS: *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York, NY, Oxford University Press, 2003
18. Weissfeld JL, Schoen RE, Pinsky PF, et al: Flexible sigmoidoscopy in the PLCO cancer screening trial: Results from the baseline screening examination of a randomized trial. *J Natl Cancer Inst* 97:989-997, 2005
19. Buys SS, Partridge E, Greene MH, et al: Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial: Findings from the initial screen of a randomized trial. *Am J Obstet Gynecol* 193:1630-1639, 2005
20. Gail MH: Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. *J Natl Cancer Inst* 101:959-963, 2009
21. Freedman AN, Slattery ML, Ballard-Barbash R, et al: Colorectal cancer risk prediction tool for white men and women without known susceptibility. *J Clin Oncol* 27:686-693, 2009
22. Wu X, Lin J, Grossman HB, et al: Projecting individualized probabilities of developing bladder cancer in white individuals. *J Clin Oncol* 25:4974-4981, 2007
23. Stadler ZK, Thom P, Robson ME, et al: Genome-wide association studies of cancer. *J Clin Oncol* 28:4255-4267
24. Frank TS, Deffenbaugh AM, Reid JE, et al: Clinical characteristics of individuals with germline mutations in BRCA1 and BRCA2: Analysis of 10,000 individuals. *J Clin Oncol* 20:1480-1490, 2002
25. Lango Allen H, Estrada K, Lettre G, et al: Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467:832-838, 2010
26. Teslovich TM, Musunuru K, Smith AV, et al: Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466:707-713, 2010
27. Andriole GL, Crawford ED, Grubb RL 3rd, et al: Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* 360:1310-1319, 2009
28. Hugosson J, Carlsson S, Aus G, et al: Mortality results from the Göteborg randomised population-based prostate-cancer screening trial. *Lancet Oncol* 11:725-732, 2010
29. Baker SG, Cook NR, Vickers A, et al: Using relative utility curves to evaluate risk prediction. *JR Stat Soc Ser A Stat Soc* 172:729-748, 2009
30. Pauker SG, Kassirer JP: Therapeutic decision-making: Cost-benefit analysis. *N Engl J Med* 293:229-234, 1975
31. Vickers AJ, Elkin EB: Decision curve analysis: A novel method for evaluating prediction models. *Med Decis Making* 26:565-574, 2006
32. Ciampa J, Yeager M, Amundadottir L, et al: Large-scale exploration of gene-gene interactions in prostate cancer using a multistage genome-wide association study. *Cancer Res* 71:3287-3295, 2011
33. Allen HL, Estrada K, Lettre G, et al: Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467:832-838
34. Purcell SM, Wray NR, Stone JL, et al: Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460:748-752, 2009
35. Chatterjee N, Park JH, Caporaso N, et al: Predicting the future of genetic risk prediction. *Cancer Epidemiol Biomarkers Prev* 20:3-8, 2011