
Re: Discriminatory Accuracy From Single-Nucleotide Polymorphisms in Models to Predict Breast Cancer Risk

The advent of genome-wide association studies to identify low-penetrance common susceptibility alleles heralds the possibility of incorporating panels of gene variants into existing risk prediction models and of assessing improvement in model performance. However, to date, the updated models have shown only modest improvements in discrimination.

Gail (1) had previously shown that adding seven single-nucleotide polymorphisms (SNPs) identified from genome-wide association analyses to the original Breast Cancer Risk Assessment Tool had yielded only a modest improvement in area under the curve (AUC) from 0.607 to 0.632. Gail (2) now reports that inclusion of 11 SNPs exhibits an even smaller improvement in the AUC (0.637) than that of the BRACTplus 7 model.

The receiver operating characteristics curve may not be sensitive to differences in probabilities between models and, therefore, may be insufficient to assess the impact of adding new predictors. A very

large independent association of the new marker is required for a meaningful improvement in AUC, and a substantial gain in performance may not yield a substantial increase in AUC. One suggested statistic for comparing nested models is the net reclassification index that is useful when risk categories are defined, and there is a consensus as to clinically meaningful cut points (3). The net reclassification index quantifies overall improvement in model sensitivity and specificity. A net improvement in risk classification implies upward reclassification of case patients and downward reclassification of control subjects.

We evaluated these metrics in our own internally validated risk prediction model for lung cancer that incorporated easily attainable epidemiological and clinical variables (4). In a genome-wide association analysis of 315 450 tagging SNPs in 1154 patients with lung cancer who were current and former smokers and were of European ancestry and 1137 frequency-matched control subjects (5), two SNPs, rs1051730 and rs8034191, that mapped to a region within 15q25.1 (which encompasses the nicotinic acetylcholine receptor subunit genes *CHRNA3* and *CHRNA5*) were strongly associated with risk (odds ratio [OR] = 1.32, 95% confidence interval [CI] = 1.24 to 1.41, $P = 3.15 \times 10^{-18}$ for rs8034191; and OR = 1.32, 95% CI = 1.23 to 1.39, $P = 7.00 \times 10^{-18}$ for rs1051730). In a subsequent meta-analysis (6) involving the UK genome-wide association study, the International Agency for Research on Cancer genome-wide association study, and our Texas genome-wide association study, the strongest associations remained for SNPs mapping to 15q25.1 (ie, rs1051730, $P = 2.83 \times 10^{-19}$; and rs8034191, $P = 4.03 \times 10^{-19}$). There was also consistent evidence for a new disease locus at 5p15.33 (ie, rs401681, $P = 4.40 \times 10^{-6}$). This locus contains two known genes: *TERT* (human telomerase reverse transcriptase) gene and *CLPTMIL* (cleft lip and palate transmembrane 1-like) gene.

We therefore added one SNP from the 15q25.1 locus (ie, rs1051730, which was used because it was in strong linkage disequilibrium with rs8034191) and two SNPs from the 5p15.33 region (ie, rs2736100 and rs401681) to the baseline model and assessed discrimination improvement. Our AUC for the baseline epidemiological-clinical model including

1016 case patients and 1111 control subjects was 0.661 (95% CI = 0.64 to 0.68). With addition of the three SNPs, the AUC showed modest, yet statistically significant, improvement to 0.673 (95% CI = 0.65 to 0.70, $P = .01$). We defined risk categories on the basis of the lower and upper quartiles of predicted risk from our baseline model as proposed by Bach et al. (7): low (predicted risk <8%), intermediate (predicted risk = 8%–50%), and high (predicted risk >50%). The resulting net reclassification indices were 0.152 (95% CI = 0.112 to 0.193) overall, 0.089 (95% CI = 0.048 to 0.130) for case patients, and 0.064 (95% CI = 0.023 to 0.105) for control subjects (all statistically significant at the 0.2% level), indicating that the SNPs modestly improved both sensitivity (9%) and specificity (6%). Although it could be argued that models providing a continuous score are more appropriate in the clinical setting, it is likely that a variety of additional summary measures evaluating model performance will be needed to assess these multigenic models.

MARGARET R. SPITZ
C. I. AMOS
A. D'AMELIO, Jr
Q. DONG
C. ETZEL

References

1. Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. *J Natl Cancer Inst.* 2008;100(14):1037–1041.
2. Gail MH. Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. *J Natl Cancer Inst.* 2009;101(13):959–963.
3. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* 2008;27(2):157–172.
4. Spitz MR, Hong WK, Amos CI, et al. A risk model for prediction of lung cancer. *J Natl Cancer Inst.* 2007;99(9):715–726.
5. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet.* 2008;40(5):616–622.
6. Wang Y, Broderick P, Webb E, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet.* 2008;40(12):1407–1409.
7. Bach PB, Elkin EB, Pastorino U, et al. Benchmarking lung cancer mortality rates in current and former smokers. *Chest.* 2004; 126(6):1742–1749.

Funding

Flight Attendant Medical Research Institute and National Cancer Institute (CA55769, CA070907, CA93592, CA016672, CA121197, CA133996, and CA123208); cancer prevention fellowship K07CA093592 from the National Cancer Institute, National Foundation for Cancer Research; DAMD17-02-1-0706 (TARGET), a grant from the Department of Defense to Dr Waun Ki Hong.

Notes

The authors of this article had full responsibility for the design of the study (M. R. Spitz, C. I. Amos, and C. Etzel), the collection of the data (M. R. Spitz and C. I. Amos), the analysis and interpretation of the data (M. R. Spitz, C. I. Amos, A. D'Amelio, Q. Dong, and C. Etzel), the decision to submit the manuscript for publication (M. R. Spitz, C. I. Amos, and C. Etzel), and the writing of the manuscript (M. R. Spitz, C. I. Amos, A. D'Amelio, Q. Dong, and C. Etzel).

Affiliation of authors: Department of Epidemiology, University of Texas M. D. Anderson Cancer Center, Houston, TX (MRS, CIA, AD, QD, CE).

Correspondence to: Margaret R. Spitz, MD, Department of Epidemiology, University of Texas M. D. Anderson Cancer Center, 1155 Pressler Unit 1340, Houston, TX 77030 (e-mail: mspitz@mdanderson.org).

DOI: 10.1093/jnci/djp394

© The Author 2009. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org.

Advance Access publication on November 5, 2009.