Whether information on genetic variants can be used for better prediction of disease risk is of current clinical interest. Previous studies have shown that mutations in the BRCA1 and BRCA2 genes are associated with high risk of breast cancer (1). Theoretical considerations suggest that further work will lead to the discovery of more genes that are involved in breast carcinogenesis (2). However, it requires more than proof of association to assess improvement in risk prediction by including genetic information. In this issue of the Journal, Meaillife et al. (3) use data from the Women’s Health Initiative Clinical Trial to provide additional support for including genetic information in risk prediction models for breast cancer. This study finds that adding a small number of single-nucleotide polymorphisms to the well-known and established Gail model can improve the prediction of breast cancer. The authors (3) indicate that previous studies that added single-nucleotide polymorphisms to the Gail model for breast cancer risk prediction suffer from methodological limitations or are merely theoretical in nature.

What statistical methods are appropriate to use in comparing risk prediction models has been subject to much recent debate. Standard analyses comparing receiver operating characteristic curves have been criticized as being insensitive to meaningful changes in predicted risk and not well suited to prospective data (4,5). Meaillife et al. (3) use newer methods based on reclassification, or changes in risk strata, which have become relatively widespread in the clinical literature. A reclassification table sorts the predicted probabilities from each risk prediction model into clinically relevant risk categories and tabulates one against the other. The extent and accuracy of any reclassification, or change in risk category, indicate whether a model can generate improvements in risk stratification, and ultimately, decisions regarding therapies. The authors compute the net reclassification improvement (NRI), which compares the proportion of case patients appropriately moving up vs down a risk category with the corresponding proportions for control patients (6).

There are several questions regarding how these methods perform and how they are affected by study design, some of which are addressed in this article (3). First, how should relevant risk categories be defined? It is known that reclassification statistics, such as the percent reclassified or the NRI, can vary according to how the risk strata are defined (7). For cardiovascular disease, an expert panel has established levels of predicted risk to guide the use of lipid-lowering therapy (8); these levels are often used to define the risk strata used in reclassification for cardiovascular disease. Although established risk strata to guide therapy are not available for all diseases, the boundaries of the risk strata should in general suggest potentially different treatment decisions or otherwise indicate a qualitative difference in risk. In breast cancer, patients with a Gail model 5-year risk greater than 1.66% are offered chemoprevention (9). In this article (3), the authors chose cut points of 1.5% and 2.0% for Gail risk estimates, that spanned the risk level of 1.66%, and examined the question of defining risk strata by computing the NRI for different boundaries. Their results show that the value of the NRI tends to be larger when the cut points are widely spaced, and smaller when the cut points are closely spaced. The authors also found little variation in the NRI for a broad range of cut points. Another study found that the NRI followed a bimodel distribution when a single cut point was used, with a small trough when the cut point was near the overall prevalence of disease (10). Thus, the NRI is at least partially dependent on the chosen risk boundaries.

Second, there are alternative methods to assess the accuracy of reclassification. The reclassification calibration test directly compares the observed rates with the expected rates within the reclassification table (7). Although the chi² statistic used in the test for reclassification calibration will vary randomly when using different risk categories, it is less dependent on the category definition and should have less systematic bias than the NRI. It seems a natural metric for cohort studies, and could be extended to nested case-control studies such as this (3).

Third, case-control studies introduce complexities into the estimation of absolute risk. In simple random sampling of case patients and control subjects, the predicted probabilities can be adjusted using Bayes theorem to reflect the overall population probability of disease, and the authors have made a comparable adjustment here (3). This calibration of the mean allows the computation of the NRI within the absolute risk categories. It is also possible to compute an NRI based on risk quantiles, such as quartiles or terciles, rather than categories of absolute risk. These may not be as interpretable clinically but could still provide a test of whether the cases are shifted to qualitatively higher risk categories than control subjects.

Fourth, this study (3) used matched case-control data. Matching distorts the distribution of predicted risk, typically among the control subjects. Janes and Pepe (11) have shown that matching case patients and control subjects also distorts the estimated receiver operating characteristic curve along with its associated area under the curve. Matching does the same for the reclassification table, particularly for the control subjects. When the data are from a nested case-control study, an adjustment using the overall cohort could help correct the distortion.

Finally, the authors compute the NRI for a subgroup of women who are at an intermediate risk of breast cancer with an estimated
Gail risk between 1.5% and 2.0%. However, when restricted to such an intermediate risk subgroup (12), the NRI can be biased because the estimated effect can be positive even in the null situation with no improvement in risk prediction. A suggested adjustment is to subtract the statistical expected value for case patients and control subjects when there is no improvement, that is, when the off-diagonal elements of the reclassification table are equal (N. R. Cook and N. P. Paynter, unpublished observations). The resulting adjusted NRI is then unbiased with an appropriate type I error.

Whether these new methods will find genetics useful for risk prediction of breast cancer remain under investigation. As noted by the authors (3), a previous article by Wacholder et al. (13) also showed a modest improvement in prediction of breast cancer risk with the addition of 10 genetic variants to the Gail model. In other complex diseases, such as diabetes (14) and cardiovascular disease (15), an improvement in risk prediction with the addition of genetic variants was not observed, even after examining risk reclassification. However, in both diabetes and cardiovascular disease, the prediction achieved by clinical risk scores was higher, with area under the curves in the range of 0.72 to 0.80 (14,15). These clinical risk models may include more causal intermediates, setting a higher standard for the addition of genetic markers.

The promise of genetics continues to evolve because of the dynamic nature of the field. Genome-wide analyses in increasingly larger consortia continue to be published, while advancements of technology allowing increasingly finer mapping of genes at a cheaper cost are simultaneously being implemented. Although the work of Mealiffe et al. (3) suggests that we may already have enough information to suggest some clinical utility for genetic variants, we look forward to seeing how the research in the next few years leads to better identification of women at high risk for breast cancer.

References

Affiliation of authors: Division of Preventive Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA (NRC, NPP).