
Joann G. Elmore, Suzanne W. Fletcher

It’s tough making predictions, especially about the future.
—Attributed to many individuals, including Yogi Berra. But, as Yogi himself said, “I really didn’t say everything I said.” (1)

Prediction is the foundation of cancer risk assessment. Risk prediction is accomplished by identifying characteristics that are associated with a high or low risk of developing a disease and then combining those characteristics in a statistical model to produce a probability estimate of developing the disease over a given period. Historically, demographic and clinical risk factors have been used in risk prediction models; more recently, genetic makeup has been added to certain models (2). Cancer risk prediction models have been used to estimate the costs of the population burden of cancer, plan intervention trials, create benefit–risk indices, and design prevention strategies for at-risk populations (3).

There is increasing interest in using risk prediction models to help individual patients to estimate their personal chance of being diagnosed with breast cancer. “What is my risk of getting cancer?” is a question that clinicians frequently encounter in their everyday practice. The 2004 Institute of Medicine report on breast cancer screening identified individual risk assessment as essential to improving early detection of breast cancer (4).

The best-known model for predicting an individual woman’s chance of being diagnosed with breast cancer is the Gail model (5,6). This model includes the following risk factors: current age, race, age at menarche, age at first live birth, the number of first-degree relatives with breast cancer, the number of previous breast biopsy examinations, and presence of atypical hyperplasia. The model predicts a woman’s likelihood of having a breast cancer diagnosis within the next 5 years and within her lifetime (up to age 90 years). This and similar risk prediction models are readily available to clinicians and patients around the world through the Internet (7,8). A version of the Gail model available on the National Cancer Institute’s Web site (http://www.cancer.gov/bcrisktool/) is viewed 20,000 to 30,000 times each month (Rehmert JH: personal communication).

The Gail model was developed and validated in the United States (5,6,9–11). However, breast cancer incidence rates vary fourfold by geographic location, with some of the highest rates in the United States and northern Europe (12). Given this wide variation in breast cancer incidence, it was not known how well the Gail model would perform internationally. In this issue of the Journal, Decarli et al. (13) used data from a multicenter case-control study in Italy and from Italian cancer registries to develop a new breast cancer risk prediction model that used the same risk factors as the Gail model. Decarli et al. then tested the relative predictive accuracy of the Italian and Gail models by using independent data from a cohort study in Florence, Italy (14,15). They found that the two models produced similar results.

Cancer risk prediction models are commonly assessed in two ways: by measuring their performance at the population level and at the level of the individual woman. Decarli et al. assessed each model’s performance at the population level by comparing the number of women in their study who the model estimated (E) would develop breast cancer with the number of women who actually were diagnosed with breast cancer (observed [O]). The Italian and Gail models estimated that 186 and 180 women, respectively, would develop breast cancer. The actual number was 194. Therefore, the overall E/O ratios for the Italian and Gail models were similar (0.96, 95% confidence interval [CI] = 0.84 to 1.11, and 0.93, 95% CI = 0.81 to 1.08, respectively).

Decarli et al. also assessed each model’s performance at the level of the individual woman. A model that discriminates well at this level should consistently predict a higher risk of breast cancer for women who will be diagnosed with the disease than for women who will not. Decarli et al. randomly selected pairs of women, one of whom was diagnosed with breast cancer and one of whom was not, to determine the frequency with which each model calculated a higher risk for the woman who developed breast cancer. The resulting calculation produced a concordance statistic, whose value could range from 0.50 (equivalent to a coin toss) to 1.0 (perfect discrimination). The concordance statistics for the Italian and Gail models were essentially the same, approximately 0.59 (with 95% confidence intervals that ranged from 0.54 to 0.63). In other words, for 59% of the randomly selected pairs of women, the risk estimated for the woman who was diagnosed with breast cancer was higher than the risk estimated for the woman who was not. Unfortunately, for 41% of the pairs of women, the woman with breast cancer received a lower risk estimate than her cancer-free counterpart. Thus, for any given woman, the two models were better at prediction than a coin toss—but not by much.

Figure 1 illustrates the problem of a similarly low concordance statistic (0.58) noted by Rockhill et al. when they applied the Gail model in the Nurses’ Health Study (10; Rockhill Levine B: personal communication). There is no place along the x-axis where one can clearly separate the group of women with breast cancer...
cancer from the group without. If, instead of proportions, the figure used actual numbers of women, the curve for the 80,755 women who did not develop breast cancer would swallow up the far smaller group of 1,354 women who did. Rockhill et al. also reported a sensitivity of 0.44 and specificity of 0.66 using a 5-year risk of developing breast cancer of 1.67% as their cut point. This result illustrates a conundrum that all clinicians know: many “low-risk” women develop breast cancer, while many “high-risk” women do not. In contrast, Fig. 2 illustrates what happens when a fictitious risk prediction model clearly differentiates between the two groups and has a concordance statistic of 1.0.

Recent attempts to improve the Gail model by adding information on other risk factors, such as breast density, have improved the concordance statistic somewhat by bringing it up to 0.66 (16, 17). However, in most situations, even a concordance statistic of 0.66 is still too low to make management decisions for individual patients.

Why is it so difficult to develop worthwhile breast cancer prediction models for individuals? First, the risk factors used in current models are widely prevalent throughout the population and are neither highly sensitive nor highly specific. In addition, a risk factor must be very strongly associated with a disease (with a relative risk of about 200) to be worthwhile for screening (18), and the same appears to be the case for accurate prediction using combinations of risk factors. Most risk factors for breast cancer are relatively weak. Even “strong” risk factors, such as older age, mammographically dense breasts, and radiation exposure, are associated with relative risks of less than 10. [Deleterious BRCA1 mutations in young women may be an exception (19).]

Current breast cancer risk prediction models perform well for populations but poorly for individuals. Cancer risk information, now readily available through the Internet, can show an individual that she is a member of a group that is at higher risk for a cancer diagnosis compared to the average population. This is valuable information, but it must be interpreted carefully (10, 20, 21). Both clinicians and patients must understand the numeric information resulting from breast cancer risk prediction models in order to use them effectively (20, 22). Because we still cannot predict accurately enough which individual woman will or will not develop breast cancer, there is much work yet to do in the field of cancer risk prediction.

REFERENCES


**Fig. 1.** Ability of the Gail et al. breast cancer risk prediction model to discriminate between women who were diagnosed with breast cancer and women who were not diagnosed in the Nurses’ Health Study. Concordance statistic = 0.58 (95% confidence interval = 0.56 to 0.60). **Solid line** = women with no diagnosis of breast cancer; **dashed line** = women with a breast cancer diagnosis (Reprinted with permission from Rockhill Levine B).

**Fig. 2.** Ability of a fictitious cancer risk prediction model to discriminate between women who will be diagnosed with breast cancer and women who will not. Concordance statistic = 1.0. **Solid line** = women with no diagnosis of breast cancer; **dashed line** = women with a breast cancer diagnosis.


NOTES

J. G. Elmore has served as an expert witness on the topic of risk communication. The authors thank R. Harns for editorial guidance.