EDITORIAL

The Power of Prediction

Harry B. Burke, MD, PhD

Department of Medicine, George Washington University, Washington, DC.

If knowledge is power, then the ability to predict and control the future is the ultimate power. The science of medicine provides the power to chart patients' futures and to change their fate.

When a patient receives a treatment, there is an implicit prediction that over an interval of time the patient will benefit from the treatment, that it will change the patient's risk of incident disease or prognosis. It is extremely difficult to make these implicit predictions explicit, but it is important to do so to determine the best treatment for an individual patient. The challenge is to discover, model, and validate biological factors in order to capture the disease process represented by the factors, and to use the power of the disease to predict its outcomes. Ideally, a therapy should only be given to those patients who will experience a greater benefit than harm.

Cancer prediction is divided into 3 domains: risk, diagnosis, and prognosis. Prognosis, the ability to predict the survival of cancer patients, has long been an active area of inquiry. Fifty-five years ago, the TNM staging system was created to predict when cancer patients with solid tumors receiving surgery would, on average, be expected to die of their disease.[1] The patients were stratified into 4 stages based on increasing anatomic extent of disease using 3 prognostic factors: tumor size, lymph node status, and metastases. The 4 stages were created using historical cohorts of patients with known outcomes. A new patient was placed in 1 of the 4 stages based on the values of his or her 3 prognostic factors; the new patient would have the same prognosis as the average of the historical cohort of patients in that stage.

The TNM staging system was an improvement over previous cancer prediction systems because it was systematic (based on the anatomic extent of disease) and its 3 prognostic factors were relatively reproducible. Furthermore, it could be applied to more than 1 solid tumor type and its predictions were better than chance at predicting death from cancer. However, its predictive accuracy rested on patients being detected relatively late in the disease process and on their receiving only surgery. Finally, it assumed that there were only 3 prognostic factors.

Three events significantly eroded the accuracy of TNM and its clinical utility. The first was the evolution of screening and early detection, which detected premalignant lesions and early disease, thereby reducing the informativeness of the TNM variables and including many patients whose disease would not be lethal. The second was new therapies, including chemotherapy and molecular

See referenced original article on pages 907-15, this issue.

Address for reprints: Harry B. Burke, MD, PhD, George Washington University School of Medicine, 2300 I Street, NW, Suite 558, Washington, DC 20037; Fax: (202) 994-8972; E-mail: domhbb@gwumc.edu

Received March 25, 2008; accepted April 9, 2008.

DOI 10.1002/cncr.23675
Published online 9 July 2008 in Wiley InterScience (www.interscience.wiley.com).

therapies, that modified the patient's prognosis but were not taken into account in the TNM system's "surgical therapy only" predictions. The third was the discovery of powerful molecular predictive factors–factors that are not, and cannot be, be taken into account in the TNM system.

Currently, the majority of women with breast cancer present with stage I and II disease. The TNM system is no better than flipping a coin (receiver operating characteristic [ROC] of 0.57) at predicting 5-year disease-specific survival for women with early-stage breast cancer.[2] Furthermore, the TNM system can produce "outcome crossover," in which patients in better outcome stages have a worse prognosis than patients in poorer outcome stages.[3] Therefore, the TNM system has ceased to be a reliable predictor of disease-specific outcome for many solid tumors.

There is a critical need for better outcome prediction in cancer, for a system that relies on the biology of the disease rather than the time of the disease's detection; a system where the variables are strong predictors because they represent the unfolding disease process (biological determinism) itself rather than those that mark the state of the disease at detection (temporal determinism).[3]

We must move from a model of temporal determinism to one of biologic determinism. Carcinogenesis is not defined by what stage of disease the patient is in at the time of detection but rather by the molecular (genomic and proteomic) characteristics of the tumor and the host. Biologic determinism takes the view that the anatomic location of the disease at detection is more related to our methods of detection than to the tumor itself. All patients are at risk of metastatic disease; some are further along a biological metastatic pathway at detection than others. In this view, treatment should be driven by the molecular biology of the tumor and host and not the tumor's location at detection.[3]

The scientific assessment of cancer risk has a more recent etiology than prognosis. The literature before the Gail model was dominated by 2 approaches: relative risk models, which were difficult to use clinically because a woman's risk varied with the reference population; and lifetime risk models, in which a woman's risk varied with her birth cohort and in which there was no recognition of her risk at a specific age.

The first widely recognized attempt to predict a person's risk of cancer occurred nearly 20 years ago with the publication of a new statistical method by Dupont[4] and a clinical model by Gail et al.[5] Gail et al used a subset of the Breast Cancer Detection Demonstration Project (BCDDP) data, combining several risk factors (age at menarche, age at first live birth, number of previous biopsies, and number of first-degree relatives with breast cancer) into a model that provided a woman with her absolute risk of incident breast cancer over a specified interval of time.

The original Gail model has been evaluated and found to require improvement.[6] To increase its predictive accuracy, the original Gail model was modified by examining the model results and, among other things, substituting age-specific invasive breast cancer incidence rates from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program for the BCDDP rates and by the use of attributable risk estimates from the SEER program to obtain the baseline hazards ratios ("Model 2"). Several validation studies of Model 2 have demonstrated an ROC for the Gail model of 0.58[7] and 0.59.[8] This is unfortunate because an ROC of 0.58 to 0.59 is essentially chance prediction and not clinically useful.

In a study published in this issue of *Cancer*, Chen et al,[9] using data from the Women's Health Initiative (WHI), found the Gail model to have an ROC of 0.55. Chen et al then added another clinical risk factor,bone mineral density, to the Gail model and achieved an ROC of 0.62. If the threshold for clinical utility is set at 0.60, the Gail model does not achieve it, although Chen et al do–but not by much. A helpful heuristic is that a validated ROC of 0.60 is usually required to surmount a predictive model's variance, 0.65 is a good clinical start, 0.70 is a clinically useful model, and 0.75 and above indicates excellent clinical accuracy.

The research of Chen et al highlights how difficult it is to create accurate risk models. Although prognostic models require relatively large datasets, typically they must follow patients over long periods of time, and they require a sufficient event rate (outcomes such as disease recurrence or death); they are strengthened by the finding that all the patients have the disease. The fact that all the patients have the disease produces a relatively homogeneous patient population and creates factors that are related to the disease, albeit by time (eg, tumor size). This relative homogeneity reduces interpatient variance and the disease-related variables can be relatively powerful prognostic factors.

These advantages do not exist in risk modeling because the majority of the people in the study cohort, even in high-risk cohorts, will never exhibit the disease. Because of the heterogeneity of the population and the relatively low disease frequency,

risk studies require very large populations, there tend to be many putative risk factors, and the factors tend to have high variances. Furthermore, most putative risk factors will not have any predictive power and even powerful predictors can appear weak because of the low event rate, colinearity with other factors, and high factor variance in the study cohort. Therefore, investigators who wish to create risk prediction models, even with large populations, usually use either a case–control design or a retrospective cohort, or wait a long time for there to be a sufficient number of incident cases. However, even case–control risk studies can require more than 1000 cases and 3000 controls.[10]

Since the early attempts at predicting cancer prognosis, several molecular risk and prognostic factors have been discovered. Breast cancer's prognostic factors include estrogen and progesterone receptor status as well as HER-2 status. In terms of risk factors for prostate cancer, there is prostate-specific antigen (PSA). These molecular factors should be included in the relevant predictive models.

Gene expression has been enlisted to predict outcomes in all major solid tumor types and whole genome analysis is being explored to assess both risk of disease and prognosis. In terms of clinical outcome prediction, gene expression uses relatively small datasets to assess a large number of potential predictive factors. Gene-based prediction models are being used clinically[11,12] but they are very difficult to validate.[13] Although there have been investigations into the use of serum proteins as predictive factors, the large-scale analyses performed to date have not yielded fruitful results.

Where does this leave the field of cancer prediction? It suggests that today clinical and tumor-related factors, by themselves or in combination, yield relatively poor predictions. What is required are molecular factors and, when validated, genomic and proteomic factors that can be added to the clinical and tumor-related predictive factors.

The often-heard argument is that the current methods may be inadequate, and even wrong, but we have to do something. This is a specious claim. It is better to use weak, but correct, clinical, tumor-related, and molecular factors than to claim genomic and proteomic power that we do not possess. For in the end, power must be used wisely and, more importantly, we cannot pretend to have power when we do not. We are not the Wizards of Medicine.

Finally, the ability to predict the future must be tempered with an existential humility. Within the hubris of our denial lurks the certainty of death, where knowledge means nothing, the will to fight means nothing, and the skill of the physician means nothing.[14] The end must come to us all, but it is our duty to forestall the patient's ultimate fate with whatever powers and arts we possess–including the power of prediction.

REFERENCES

1. Gospodarowiez M, Benedet L, Hutter RV, Fleming I, Henson DE, Sobin LH. History and international developments in cancer staging. *Cancer Prev Control*. 1998;2:262-268.
2. Burke HB, Henson DE. Histologic grade as a prognostic factor in breast carcinoma. *Cancer*. 1997;80:1703-1705.
3. Burke HB. Outcome prediction and the future of the TNM staging system. *J Natl Cancer Inst*. 2004;96:1408-1409.
4. Dupont WD. Converting relative risks to absolute risks: a graphical approach. *Stat Med*. 1989;8:641-651.
5. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. 1989;81:1879-1886.
6. Spiegelman D, Colditz GA, Hunter D, Hertzmark E. Validation of the Gail et al. model for predicting individual breast cancer risk. *J Natl Cancer Inst*. 1994;86:600-607.
7. Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst*. 2001;93:358-366.
8. Decarli A, Calza S, Masala G, Specchia C, Palli D, Gail MH. Gail model for prediction of absolute risk of invasive breast cancer: independent evaluation in the Florence-European Prospective Investigation Into Cancer and Nutrition cohort. *J Natl Cancer Inst*. 2006;98:1686-1693.
9. Chen Z, Arendell L, Aickin M, Cauley J, Lewis CE, Chlebowski R. Hip bone density predicts breast cancer risk independently of Gail score: results from the Women's Health Initiative. *Cancer*. 2008;113:907-915.
10. Santen RJ, Boyd FN, Chlebowski RT, et al. Critical assessment of new risk factors for breast cancer: considerations for development of an improved risk prediction model. *Endocr Relat Cancer*. 2007;14:169-187.
11. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst*. 2006;98:1183-1192.
12. Wolf I, Ben-Baruch N, Shapira-Frommer R, et al. Association between standard clinical and pathologic characteristics and the 21-gene recurrence score in breast cancer patients. *Cancer*. 2008;112:731-736.
13. Simon E. Development and evaluation of therapeutically relevant predictive classifiers using gene expression profiling. *J Natl Cancer Inst*. 2006;98:1169-1171.
14. Rieff D. Swimming in a Sea of Death, as quoted by Katie Roiphe, Without Metaphor. *New York Times*. February 2, 2008:11.