(D-CA), who is likely to become chairman of the House Committee on Veterans' Affairs. "They talk about the seamless transition, but there is no such thing. The proactive approach is just not part of their culture."

"I give the military and the VA credit" for creating programs to treat brain-injured veterans, "but there are not enough of them, and I think that's the bottom line," said Gene Bolles, an assistant professor of neurosurgery at the University of Colorado at Denver, who treated soldiers wounded in

Afghanistan and Iraq at the military's Landstuhl Regional Medical Center in Germany from 2001 to 2003. "The best thing the military could do is to recognize that this is a serious problem, help them get jobs, and give them the disability [payments] that they deserve."

An Interview with Jason Pepper and Harriet Zeiner can be heard at www.nejm.org.

Dr. Okie is a contributing editor of the *Journal*.

1. Defense Manpower Data Center Statistical Information Analysis Division. Global war on terrorism — Operation Iraqi Freedom by casualty category within service, March 19, 2003 through November 4, 2006. (Accessed November 30, 2006, at http://siadapp.dior.whs.mil/personnel/CASUALTY/castop.htm.)
2. *Idem.* Global war on terrorism — Operation Enduring Freedom by casualty category within service, October 7, 2001 through November 4, 2006. (Accessed November 30, 2006, at http://siadapp.dior.whs.mil/personnel/CASUALTY/castop.htm.)
3. Okie S. Traumatic brain injury in the war zone. N Engl J Med 2005;352:2043-7.
4. Healthcare inspection: health status of and services for Operation Enduring Freedom/Operation Iraqi Freedom veterans after traumatic brain injury rehabilitation. Washington, DC: Department of Veterans Affairs Office of Inspector General, July 12, 2006. (Report no. 05-01818-165.)

Copyright © 2006 Massachusetts Medical Society.

STATISTICS AND MEDICINE

The Limitations of Risk Factors as Prognostic Tools

James H. Ware, Ph.D.

Identification of new risk factors for specific diseases is an enduring theme in medical research. Advances in molecular biology, genetics, and computational biology are accelerating the pace of this work. The research seeks to increase our understanding of the causes of diseases, but there is also hope that the recognition of new risk factors will lead to improved methods for identifying persons who are in the early stages of, or at high risk for, the diseases of concern. Research has shown, however, that a risk factor must have a much stronger association with the disease outcome than we ordinarily see in etiologic research if it is to provide a basis for early diagnosis or prediction in individual patients.[1,2] A simple example will illustrate the problem.

Suppose that the goal is to identify persons at high risk for acquiring a disease (the outcome) within 5 years. We will assume, as shown in the figure, that the risk factor is normally distributed in the group of persons who will not experience the outcome. We will further assume that the risk factor is also normally distributed with the same variance, but with a mean 0.5 SD that is larger among persons who will have the outcome than among the group of persons who will not. (The results that follow do not depend on the values of the mean end variance in the group of persons without the outcome.) Finally, let us suppose that the cumulative incidence of the outcome within 5 years is 5%.

Given these assumptions, we

can use standard methods to calculate the conditional probability that a person with a given value of the risk factor will have the outcome.[3] For example, an event will occur with probability 0.081 in persons whose value for the risk factor falls at the 90th percentile of the distribution in the event-free group, and with probability 0.024 in persons with a value at the 10th percentile. Thus, the odds ratio for the outcome in those at the 90th percentile, relative to those at the 10th percentile, of the distribution in the event-free group is $(0.081 \div 0.919) \div (0.024 \div 0.976) = 3.58$. In most epidemiologic studies, a risk factor with an odds ratio of this magnitude would be of considerable interest.

Despite the strong association between the risk factor and the

disease outcome, it does not follow that the risk factor provides a basis for an effective prediction rule for individual patients. Consider a prognostic test based on this risk factor with a cutoff value for a "positive" finding of 1.645 SD above the mean in the event-free group. Such a test would have a false positive probability of 0.05 — that is, a specificity of 0.95. The other important property of a prognostic test is its sensitivity, or the probability of a positive test result among those who will have the event. Standard methods for calculating tail probabilities of the normal distribution show that the sensitivity of a test based on this cutoff value would be only 0.13. That is, only 13% of persons who will have the outcome within 5 years will be identified by a positive result on the proposed test.

The reasons for the difference in performance as a risk factor and as a prognostic test are apparent in the figure. If the distributions of the risk factor differ between the group that will have the outcome and the group that will not, the risk factor is associated with the outcome. If the sample size is sufficiently large and the model is properly specified, one can expect to show that the risk factor is a statistically significant contributor to a prediction model for the outcome. For the risk factor to perform well as a prognostic test for the individual patient, however, the distributions in the two groups must be sufficiently well separated to permit the selection of a cutoff value that will discriminate between the two groups with high sensitivity and specificity. In our example, the mean value in the group that will
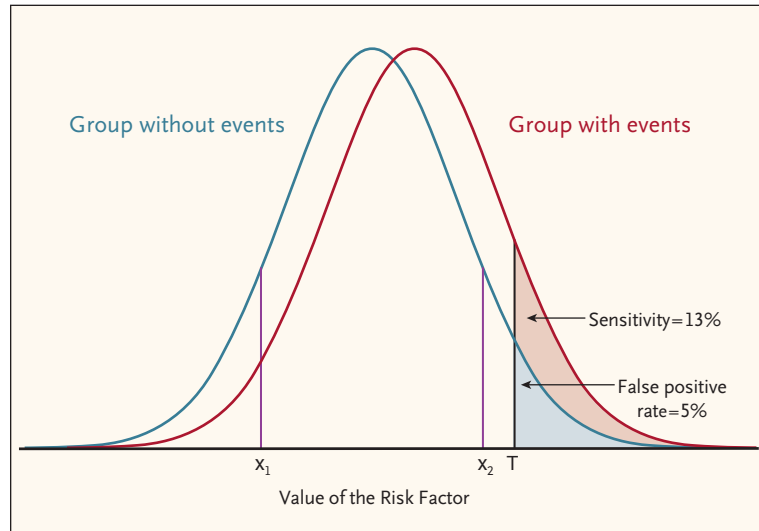
Normal Probability Density Functions of the Risk Factor among Persons Who Will Not Have the Event (Blue) and among Those Who Will (Red).

The 10th and 90th percentiles of the distribution in the event-free group are labeled as x_1 and x_2, respectively. The cutoff value for a diagnostic test with 95% specificity (false positive probability=0.05) is labeled as T. To calculate the conditional probability of membership in the group with an event, given a value of the risk factor, one assigns weights to the probability densities in the two groups according to their population prevalence of either 5% or 95%.

have the event must be 2.12 SD higher than the mean in the event-free group in order for a prediction rule to achieve a sensitivity of 0.80 and a specificity of 0.90. These values correspond to an odds ratio of 228 for persons at the 90th percentile of the distribution in the event-free group relative to those at the 10th percentile.

In this issue of the *Journal*, Wang and colleagues (pages 2631–2639) attempt an even more ambitious task. They seek to identify biomarkers that contribute to prediction models for death from any cause and major cardiovascular events after controlling for a set of established risk factors. They consider 10 biomarkers that have been proposed as risk factors for these outcomes.

In their analysis of death from any cause, Wang and colleagues

first perform a proportional-hazards regression analysis to identify the biomarkers that contribute to a multivariate regression model. They find that 5 of the 10 markers make a statistically significant contribution to the multivariate model. They then compute a "biomarker score," defined as a weighted sum of the biomarkers, in which the weights are the estimated regression coefficients in the proportional-hazards regression model. This calculation is appropriate because this weighted sum measures the cumulative contribution of these biomarkers to the estimated hazard function.

When study participants are stratified into quintiles on the basis of this biomarker score, the unadjusted 5-year mortality rate in the highest quintile is about six

times as great as the 5-year mortality rate in the two lowest quintiles (see Fig. 2A of the article by Wang et al.). After adjustment for conventional risk factors, the hazard rate in the highest quintile is 4.08 times the hazard rate in the two lowest quintiles (see Table 2 of the article by Wang et al.).

Despite this significant contribution to the proportional-hazards regression model, the proposed biomarker score adds little to the sensitivity and specificity of a prognostic test for death within 5 years. The usual measure of the performance of a prognostic test is the receiver-operating-characteristic (ROC) curve, which plots the sensitivity of the test against 1 minus the specificity for all possible cutoff values. The area under that curve, known as the C statistic, is the proportion of pairs for which the model assigns higher probability to the person who will have the event than to the person who will not.[4] Wang et al. find that the C statistic is increased only from 0.795 to 0.816 by the addition of the five biomarkers to the model that is based on the conventional risk factors. For a cutoff value chosen to achieve a false positive probability equal to 0.05, we see from their Figure 3A that the sensitivity of the hypothetical test is increased from about 0.30 to 0.42 by the addition of the biomarkers. Thus, this group of biomarkers makes a substantial contribution to the proportional-hazards model for predicting death from any cause, but it is of limited value for the risk stratification of individual patients.

This scenario has unfolded repeatedly as we have discovered new biologic variables that lie on the complex pathway leading to chronic disease and death. The work of Wang and colleagues, however, shows us how difficult it is to achieve effective risk stratification with respect to multifactorial disease processes. Much work remains to be done before biomarkers of the type the authors consider here can provide a basis for prognostic evaluation of the individual patient.

Dr. Ware is the dean for academic affairs and a professor of biostatistics at the Harvard School of Public Health, Boston, and a statistical consultant to the *Journal*.

1. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 2004;159:882-90.
2. Wald NJ, Hackshaw AK, Frost CD. When can a risk factor be used as a worthwhile screening test? BMJ 1999;319:1562-5.
3. Anderson TW. An introduction to multivariate statistical analysis. 3rd ed. Hoboken, NJ: Wiley-Interscience, 2003.
4. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat Med 2004;23:2109-23.