RISK: Health, Safety & Environment (1990-2002)

Volume 3
Number 3 *RISK: Issues in Health & Safety*

Article 14

June 1992

Book Review

Darold T. Barnum

Book Review

Erratum
The citation for this review is *3 RISK 259 (1992)* in most commercial databases.

HELENA CHMURA KRAEMER, EVALUATING MEDICAL TESTS: OBJECTIVE AND QUANTITATIVE GUIDELINES. (Sage Publications 1992). [295 pp.] Foreword A. John Rush, M.D. Acknowledgments, index, list of statistical notations, preface, references, and about the author. ISBN 0-8039-4611-2 (cloth); 0-8039-4612-0 (paper). [Cloth $45.00, paper $22.50. 2455 Teller Road, Newbury Park CA 91320.]

The purpose of medical tests is to detect whether a patient has a disorder. However, the various tests for a disorder often vary immensely in accuracy and cost. Because of the consequences of incorrect diagnoses and the fact that testing accounts for about one-third of all medical expenses, the importance of evaluating tests for accuracy and cost is evident.

Kraemer has written an excellent book that develops a procedure for evaluating the performance of medical tests. The methodology considers the probabilities of incorrect results, the consequences of such errors, and costs. Kraemer argues correctly that future evaluation of medical tests should be based on systematic, completely objective and empirical methodologies. This book takes a significant step in that direction.

The book would be beneficial for anyone concerned with evaluating medical test performance. However, to profit from it, readers will need a moderately good grasp of basic statistics and of current approaches to assessing medical tests. Thus, this work is essential reading for those with a serious interest in methodologies for assessing medical test quality, and who have or are willing to develop a sound understanding of the field.

In the first five chapters, Kraemer thoughtfully reviews the basics of medical testing procedures. She discusses such topics as disorder vs. diagnosis; testing protocols, responses and referents; and population and sampling concepts. She also considers sensitivity, specificity, predictive values, efficiency, and statistical tests for significance. She devotes substantial discussion to the variability of diagnoses, and constructs an index measuring diagnosis reproducibility. She thereby demonstrates that the outcomes of diagnoses and gold standards do not perfectly identify the presence or absence of a disorder.

In Chapter 6, Kraemer critically analyzes the signal detection approach to test evaluation. She identifies problems with sensitivity and specificity, develops quality coefficients for each to correct the problems, and presents test Receive Operating Characteristic (ROC) curves for sensitivity/specificity and test Quality Receiver Operating Characteristic (QROC) curves for her quality coefficients. She clearly explains ROC and QROC curves, and why tests on them are preferable to those below the curves. She then takes the important step of developing ROC and QROC curves for the diagnosis, that is diagnosis ROC and QROC curves, and demonstrates how the best tests will be those lying closest to or above the diagnosis ROC or QROC curves. Her procedure provides an objective method, based on empirical evidence, for evaluating medical tests, and is a significant contribution.

Also in this chapter, Kraemer shows that a random test will have a sensitivity equal to the probability of a positive test result. (Test sensitivity measures the percentage of those with a disorder who test positive for it.) For example, assume a patient is tested by flipping a fair coin, and classified as positive if a head appears. Then, 50% of those with the disorder will test positive, resulting in an expected sensitivity of 50% by random chance. To correct for this problem, Kraemer develops a sensitivity quality coefficient, which measures the percentage of those with the disorder who test positive after deleting the random element. It equals (Sensitivity – Random Percentage)/(1 – Random Percentage). In the preceding example, where sensitivity was 50% by random chance, the sensitivity quality coefficient would be zero, since (0.50 –0.50)/(1 – 0.50) = 0. Now consider another test, which also has a 50% random percentage, but in which 75% of those diagnosed with the disorder test positive. For this test, the quality coefficient is 50%, since (0.75 – 0.50)/(1 – 0.50) = 0.50. In a third test with a random percentage of 50%, in which all of those diagnosed with the disorder test positive, sensitivity is 100%. And, (1 – 0.5)/(1 – 0.5) = 1, meaning that the quality coefficient is also 100%. Kraemer develops a similar quality index for specificity.

However, the book seems to suggest that the two quality coefficients always are preferable to sensitivity and specificity. This is not so. Assume for example two screening tests for a serious disease that is curable if detected. One test has a random level of 0.20 and a sensitivity of 0.40, while the other has a random level of 0.90 and a sensitivity of 0.91. The sensitivity quality index of the first test is 0.25, and of the second 0.10. For the first test, 40% of those with the disease will be detected, while with the second test, 91% of those with the disease will be detected. Yet, using only the quality indices, the first test incorrectly would be considered the best. My example is extreme and ignores specificity. But, it suggests that there may be cases where sensitivity is a better measure than its quality index, or at least that both should be considered. I don't think that Kraemer would disagree with my comments, and her final model avoids the problem. But, readers might draw the wrong conclusions about the universal superiority of the quality indices from the Chapter 6 presentation.

In Chapter 7, Kraemer covers the second major approach to test evaluation the Predictive Value, or Bayesian, approach. She argues that there is no fundamental difference between this approach and the signal detection method when sensitivity and specificity are replaced by her quality coefficients for these two terms. And, in Chapter 12, she discusses the third major technique, the Multivariate Discrimination approach. In both cases she provides excellent brief descriptions and critiques of the procedures.

Another contribution to test evaluation is covered thoroughly in Chapter 9. Here, Kraemer incorporates into her model the utilities of correct and incorrect test results, and the cost of the test itself. By the end of this chapter, she provides a performance index for medical test evaluation that includes test accuracy, relative clinical benefit, fixed test cost, and potential clinical benefit. Because this final model is logically satisfying and requires data that normally would be available, it is of practical value.

A further significant contribution is Kraemer's development of a procedure for evaluating batteries of medical tests by using optimal sequences, covered primarily in Chapter 11. She also discusses prognostic tests, and concludes with a summary of the past, present and future of medical testing. In her final chapter, she argues convincingly for a standardized procedure for test evaluation, with mandatory requirements along the lines of the standardization provided by the Food and Drug Administration for the evaluation of new drugs.

In summary, this book provides a unique and thought-provoking examination of medical test evaluation, illustrating what is wrong with many current methods. By offering a consistent and logical alternative, it makes an important contribution toward improved evaluations of medical test quality.

Darold T. Barnum†

† Dr. Barnum is a Professor of Management at the University of Illinois at Chicago, and Director of the University of Illinois Center for Human Resource Management.