Insurance Testing

THE IMPACT OF CANCER BIOLOGY, LEAD TIME BIAS, AND LENGTH BIAS IN THE DEBATE ABOUT CANCER SCREENING TESTS

BRIAN R. KAY, MD
Associate Medical Director
The Principal Financial Group
Des Moines, Iowa

DAVID L. WITTE, MD, PHD
Director of Clinical Laboratories
Laboratory Control, Ltd.
Ottumwa, Iowa

Tumor marker screening by insurers is one of the most significant and controversial developments in insurance testing. The debate is ongoing. Key dimensions of the issue have been recently described.[1,2] We do not wish to reiterate any of those dimensions, nor do we wish to comment favorably or unfavorably on tumor marker screening by insurers. Rather, we believe there are some important and interesting theoretical considerations for the evaluation of cancer screening tests in clinical or insurance medicine. This article illustrates the biologic variability of two malignancies which may not be widely appreciated. New work has shown dichotomous populations within groups of patients known to have monoclonal gammopathy of unknown significance (MGUS) and childhood neuroblastoma. In addition, this article elucidates the role of *lead time bias* and *length bias* in the interpretation of screening tests. This has significance for all screening tests, not just cancer screening tests or tumor marker screening tests.

Effective test utilization requires careful data-driven reasoning in several disciplines. Cancer biology is highly variable and our currently incomplete knowledge is continually expanding. Wide-population-based objective clinical epidemiologic data are now being collected which will more accurately reflect average outcome (e.g., MEDIS groups have been mandated in several states). The statistics of testing must be carefully evaluated for individual predictive value, population outcome predictions, and multiple sources of bias.

The biology of cancer is enormously variable even between individuals with the same malignancy. Advances in molecular biology provide new non-morphologic molecular indicators of prognosis in many forms of cancer. Clinical medicine continues to have few tools to detect cancer until some time after its biological onset. Biological onset may have a variable relationship to the clinical onset. The natural history of a cancer's biologic course is difficult to study in human populations. Detection methods and therapeutic modalities change, which alters the natural history during the course of our collective experience.[3] The best data are derived from longitudinal inclusive cohort studies which are difficult in the medically mobile American society. These problems have been discussed previously.[4,5] Two illustrative examples offer insight into biological variability and test utility (childhood neuroblastoma and adult plasma cell dyscrasia).

Adequate longitudinal cohort data exist showing childhood neuroblastoma to be a heterogeneous disease.[6,7] At one extreme is a non-aggressive biology that may regress spontaneously and is curable in 90% of cases, even if detected after symptoms develop. This non-aggressive tumor has identifiable genetic markers and is thought to have a longer interval from biologic onset to clinical detection. At the other extreme is a very aggressive biology which is genetically different, most frequently unresponsive to therapy, thought to have a shorter interval from biological onset to clinical detection, and is highly unlikely to regress spontaneously. These two processes do not appear to interconvert.

Both types of neuroblastoma can be detected by a technically practical urine test for elevated catecholamines and metabolites.[6] The current consensus in America[6,7] is that the urine test is not being implemented because positive tests are much more likely to detect the indolent, responsive, non-aggressive tumors which respond well even after detected clinically, and early detection appears to have little benefit for the aggressive tumors. This is an example of careful reasoning and recognition of the *length bias* in cancer screening strategies discussed below.

Cohort studies require longer duration in adults. Robert Kyle[8] has studied a cohort with "incidentally" found monoclonal gammopathy. All 241 patients' malignancies were discovered before 1971 and all were determined to be monoclonal gammopathy of unknown significance (MGUS) at time of discovery. Known cases of multiple myeloma, macroglobulinemia, amyloidosis, lymphoma, etc., were excluded from the cohort. These patients have been followed since 1971, a monumental achievement. No quantitative change in paraprotein has occurred in 57 (24%) of these people. An increased protein without overt symptoms was observed in 7 (3%). Multiple myeloma requiring intervention occurred in 53 (22%), with a median onset of three years after discovery of the gammopathy. More than half (124; 51%) died of causes unrelated to their gammopathy. This illustrates the varied course of plasma cell dyscrasias and the necessity of cohort studies to determine variability in cancer biology.

Monoclonal gammopathy is easily detected by serum protein electrophoresis and immunoelectrophoretic methods. The clinical significance is readily determined by bone marrow

biopsy and skeletal radiology. Screening populations for monoclonal gammopathy will find many more MGUS than rapidly impending myeloma.[8] Testing identifies a group with increased mortality, i.e., 22% developed myeloma. However, in 51%, the finding presumably had no significance related to mortality. The decision making regarding benefits of screening to a population and to individuals is actively discussed in clinical medicine[9] as well as insurance medicine.[4]

The biological variability of cancer induces predictable biases[11] in any study of screening tests. The more easily understood are the *selection bias* among those tested and *diag-nostic bias* by the test interpreters (frequently radiologists or pathologists). More subtle biases that falsely suggest a survival benefit in the screened population are *lead time bias* and *length bias*. In addition, the "Will Rogers Phenomenon"[12] complicates comparison of studies done at different times. That is, newer, more sensitive detection systems classify cases as "bad" when earlier methods would have classified them as "good."

Lead time bias results from moving the detection earlier in the course. However, if therapy fails to change the time from biological onset until death, then survival after detection is erroneously biased toward longer survival in the screened population. No real change in life years compared to an unscreened population has occurred.

We believe the *length bias* is of greatest interest in insurance medicine. *Length bias* occurs because slower growing, less aggressive tumors have a longer course than aggressive, rapidly fatal tumors. Therefore, anytime a population is screened, most of the cases found will be nonaggressive. This length bias may artifactually make survival appear longer in patients detected in an early screening program than those detected clinically. Vigorous discussion continues regarding the influence of length bias on clinical screening programs.[9,13]

Figure 1 is an adaptation from Hulka[11] and illustrates the variability in cancer biology and the potential impact of testing upon the identification of excess mortality for a hypothetical cancer. The figure shows biological onset (B) at the same time in the two courses. This is an over simplification. The age of incidence and relative frequency of the two courses must be determined from longitudinal cohort data and are required to make reasonable judgments. If a population is screened, the preclinical-test-detected (PC–TD) cases will be biased toward an increased relative frequency of the nonaggressive tumors, each with fewer years of life lost than the aggressive tumors.

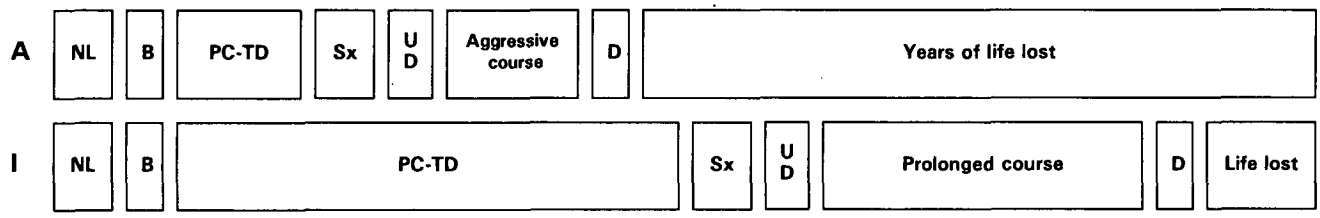
The utility of screening also requires data regarding predictive value of tests and prevalence of disease, which has been discussed elsewhere.[14,15] Frequently careful quantitative epidemiologic interpretations are counter-intuitive. Remember, with a test that is 95% specific and 95% sensitive for a disease with a prevalence of 0.001, only one positive test among 51 positives will represent a true positive. Both clinical and insurance medicine have responsibilities to all, be they false positives, false negatives, true positives, or true negatives. We must remember that the superb sensitivity and specificity achieved in anti-HIV testing is rare in laboratory medicine.

When cancer screening tests are undertaken in clinical medicine, it is always presumed that "prognosis is a function of detection."[11] That is, early detection postpones death due to this cause or increases years of high quality life. Lead time bias and length bias must always be considered and are difficult to exclude.[13] Very few effective cancer screening programs exist and many of those in place endure continuing debate. Accumulating adequate data to unequivocally support any cancer detection system is a large undertaking.

If cancer detection systems are independently studied by insurance medicine, the same biases need to be understood. Length biases may act as a positive attribute by lengthening

Figure 1

A Hypothetical Cancer with Variable Course



A ---------- Time line for aggressive tumor
I ---------- Time line for indolent tumor

NL ..... Biologically normal?
B ..... Biological onset
PC-TD ..... Preclinical test detectable and asymptomatic
Sx ..... Symptoms
UD ..... Usual detection time by clinical means
D ..... Death

the time for disease detection. Tests may identify populations with increased mortality but individual predictions may have great uncertainty.[16,17] Cumulative population outcomes may be estimated with some precision, but insurance medicine must deal effectively with its public image. The lay press[18] readily discusses company attempts to limit insurance risks while not discussing adverse effects of antiselection by the applicants. Any program requires clear protocols for dealing with both positive and negative tests and, remember, there may be many false positives and a few false negatives and no clues which is which.

As new technology becomes available, we are challenged to make sound quantitative decisions under uncertain conditions.[19] We need to seek unimpeachable data, search for biases, and discard inconclusive data.

Finally, we must remember the "framing of questions" or "views of the situation" can have a profound impact on any decision.[20] Logical and quantitative thought is needed to find tests that yield the most positive financial, medical, emotional, and market outcomes. We hope this editorial and bibliography will solicit discussion.

REFERENCES

1. Chambers D. Tumor Marker Screening Tests Post Dilemma for Insurance Industry. *Medi Resource* 1990(Sept./Oct.);2(5):1-3.
2. Chambers D. Tumor Marker Screening by Insurers. *Med Resource Special Supplement* 1991(May/June).
3. Singer RB. Feinstein's Report of the "Will Rogers" Phenomenon: Advances in Diagnosis, Resulting Stage Migration, and Their Impact on 6-Month Lung Cancer Mortality by Stage. *J Insur Med* 1991(Summer);22(2): 141-4.
4. Mills GM. A General Model for Conducting Protective Value Studies. *J Insur Med* 1991(Spring);23(1):12-15.
5. Singer RB, Kita MW. Guidelines for Evaluation of Follow-Up Articles and Preparation of Mortality Abstracts. *J Insur Med* 1991(Spring); 23(1):21-9.
6. Murphy SB, Cohn SL, Craft AW et al. Do Children Benefit from Mass Screening for Neuroblastoma?: Consensus Statement from the American Cancer Society Workshop on Neuroblastoma Screening. *Lancet* 1991(Feb.);337:344-6.
7. Tuchman M, Lemieux B, Woods WG. Screening for Neuroblastoma in Infants: Investigate or Implement? *Pediatrics* 1990(Nov.);86(5):791-3 *(editorial).*
8. Kyle RA, Lust JA. Monoclonal Gammopathies of Undetermined Significance. *Seminars in Hematology* 1989;26:176-200.
9. Habbema JDF, Van Oortmarssen GJ, Van Putten DJ. An Analysis of Survival Differences Between Clinically and Screen-Detected Cancer Patients. *Statistics in Med* 1983;2:279-85.
10. Chuong, JJH. A Screening Primer: Basic Principles, Criteria, and Pitfalls of Screening with Comments on Colorectal Carcinoma *J Clin Gastroenterology* 1983;5:229-33.
11. Hulka, BS. Cancer Screening; Degrees of Proof and Practical Application. *Cancer* 1988;62:1776-80.
12. Feinstein AR, Sosin DM, Wells CK. The Will Rogers Phenomenon; Stage Migration and New Diagnostic Techniques as a Source of Misleading Statistics for Survival in Cancer. *New Eng J Med* 1985;312:1604-8.
13. Flanders WD, Longini IM. Estimating Benefits of Screening from Observational Cohort Studies. *Statistics in Med* 1990;9:969-80.
14. Wesley D, Kita MW. Introduction to Probability Methods and Concepts. *J Insur Med* 1991(Spring);23;1:16-20.
15. Sox HC. Probability Theory in the Use of Diagnostic Tests: An Introduction to Critical Study of the Literature. *Ann Intern Med* 1986;104:60-6.
16. Laupacis A, Sackett DL, Roberts RS. An Assessment of Clinically Useful Measures of the Consequences of Treatment. *New Eng J Med* 1988(June);318:1728-33.
17. Silver, AL. Life Expectancies: Population or Person? *Ann Intern Med* 1989(Aug);111(3):257 *(editorial).*
18. Murray MJ. Is Cancer Your Family's Weak Link? *Coping* 1991(Spring); pp. 16-19.
19. Tversky A, Kahneman D. Judgment Under Uncertainty: Heuristics and Biases. *Science* 1974;185:1124-31.
20. Tversky A, Kahneman D. The Framing of Decisions and the Psychology of Choice. *Science* 1981(Jan);211:453-8.