Sensitivity and Specificity of Common Scintigraphic Procedures

Michael L. Goris

SENSITIVITY AND SPECIFICITY OF COMMON SCINTIGRAPHIC PROCEDURES

A Review of Clinical Efficacy

MICHAEL L. GORIS, M.D., Ph.D.

Professor of Radiology
Department of Radiology
Division of Nuclear Medicine
Stanford University Medical Center
Stanford University School of Medicine
Stanford, California

YEAR BOOK MEDICAL PUBLISHERS, INC.
CHICAGO

0 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging in Publication Data

Goris, Michael L.
Sensitivity and specificity of common scintigraphic procedures.

Bibliography: p.
Includes index.
1. Radioisotope scanning. 2. Diagnosis, Radioscopic. I. Title.
RC78.7.R4G676 1985 616.07'575 85-2518
ISBN 0-8151-3809-1

Sponsoring editor: Daniel J. Doody
Editing supervisor: Frances M. Perveiler
Production project manager: Sharon W. Pepping
Proofroom supervisor: Shirley E. Taylor

To Steven and Nicolas Goris

I was no prophet, neither was I a prophet's son; but I was an herdsman, and a gatherer of sycamore fruit.

(AMOS 7-14)

FOREWORD

NEARLY FORTY YEARS ago I learned the principles of the then relatively new field of electrocardiography by studying a wonderful little monograph. The author set forth in simple language the underlying electrophysiology and then proceeded to explain the graphic patterns of various rhythm and conduction disorders, chamber hypertrophy, infarction, and pericardial disease. On the last page of the book there was a single sentence:

"No cardiac examination is complete without an electrocardiogram."

As a fourth-year student, I took the statement to be a warning about the need for thoroughness in the diagnostic evaluations that would characterize my future work. Peers and superiors would expect me to "order an EKG" on every patient undergoing evaluation of the heart. I remember the uneasiness that sentence evoked in me. I knew that there was something wrong about the dictum, but few beginners in medicine in those days dared to question such advance.

During the years that followed, an extraordinary number of tests of the anatomy and of the function of the heart became available, and the list continues to grow with no end in sight. Some procedures, such as x-ray kymography and ballistocardiography, obsolesced quickly and have virtually disappeared from the scene. Others, like phonocardiography, gradually fell into disuse. Still others assume permanent positions in the diagnostic armamentarium: right heart catheterization, left heart catheterization, arterial blood gas determination, lipoprotein analysis, serum enzyme measurements, ultrasonography, angiography, radionuclide imaging, digital subtraction angiography, Doppler flow studies, dynamic computed tomography, magnetic resonance imaging, etc. Old admonitions about complete evaluations became increasingly disturbing, and choices about diagnostic approaches were clouded by mounting threats of malpractice litigation. Health care costs soared.

It is apparent that absolute completeness is neither a possible nor a desirable goal. Many tests are redundant and many provide either contradictory or inconclusive results. Furthermore, every test has at least two drawbacks: the hazard of a false-positive result, and the

onus of significant economic costs. Fortunately, during the past decade, the methods of decision analysis have been increasingly applied to the complex issues surrounding clinical choices. Approaches based on the accuracy and the predictive value of tests have been developed, and these often make it possible to formulate efficient plans, individualized to suit the unique circumstances that characterize each patient.

Dr. Goris and his colleagues have used these methods in a rigorous manner to assemble reasonable and efficient guidelines for the use of scintigraphic procedures. He is a world authority in the field of nuclear medicine and brings together in this work an encyclopedic knowledge of this discipline and a masterful use of the methods of decision theory. This is the kind of contribution that is needed to unravel the quandaries faced by my colleagues, the clinicians.

EDWARD RUBENSTEIN, M.D.
PROFESSOR OF MEDICINE
STANFORD UNIVERSITY SCHOOL OF MEDICINE
STANFORD, CALIFORNIA

CONTENTS

INTRODUCTION

THIS BOOK IS PRIMARILY intended for the clinician, who first has to decide if a given scintigraphic procedure is likely to help in the diagnosis, and second has to gauge the significance of the procedure's outcome. This book is therefore not an introduction to nuclear medicine or to scintigraphy in general, but to the clinical applications of scintigraphic procedures.

The chapters are organized accordingly. Each chapter considers the application of one scintigraphic procedure for the evaluation or diagnosis of one disease entity. The procedure is briefly described, chiefly in respect to the subsumed physiopathologic mechanism on which it is based. Methodologic variations are mentioned only to explain variations in results and important limitations. But the major emphasis in each chapter is on the performance of the procedure.

The intrinsic performance of a diagnostic procedure results from the relative frequency of positive outcomes in affected vs. unaffected subjects.

The relative frequency of positive outcomes in affected subjects is the *sensitivity*. The frequency of positive outcomes in unaffected subjects is the *nonspecificity*.

True positive rates, true negative rates, and accuracy, on the other hand, are a function of the population mix, in addition to being a function of the intrinsic characteristics of the procedure.

The relation between these variables is best illustrated by a two-by-two contingency table. In this table we have two column entries, for the affected and unaffected patients, respectively. The two row entries are for the positive and negative outcomes of the test or procedure:

	Affected	Unaffected	Total
Positive	a	b	a + b
Negative	c	d	c + d
Total	a + c	b + d	a + b + c + d

where *a, b, c,* and *d* represent the number of observations in the four categories (affected and positive = *a;* affected and negative = *c;* unaffected and positive = *b;* unaffected and negative = *d*). The following definitions are used throughout the book:

SENS (sensitivity) = a/(a + c)
NSPEC (nonspecificity) = b/(b + d)
PREV (prevalence) = (a + c)/(a + b + c + d)
TPR (true positive rate) = a/(a + b)
TNR (true negative rate) = d/(c + d)
ACC (accuracy) = (a + d)/(a + b + c + d)

A better term for the true positive rate is the *positive predictive value* (PPV), or the likelihood that an individual with a positive outcome is in fact affected. In the same manner, the TNR is equivalent to the *negative predictive value,* or the likelihood that an individual with a negative outcome is truly unaffected.

The formal relationship between those parameters is defined by a simple equation popularly known as Bayes' theorem:

$$\text{PPV} = \frac{\text{SENS} \times \text{PREV}}{\text{SENS} \times \text{PREV} + (1 - \text{PREV}) \times \text{NSPEC}}$$

and conversely:

$$\text{NPV} = \frac{(1 - \text{PREV}) \times (1 - \text{NSPEC})}{(1 - \text{PREV}) \times (1 - \text{NSPEC}) + \text{PREV} \times (1 - \text{SENS})}$$

This abstract formulation can easily be expressed in common language: A positive test result is more likely to predict disease if the prevalence is high and/or if the nonspecificity is low. Conversely, a negative test result is more likely to indicate absence of disease if the prevalence is low and/or the sensitivity is high. The noteworthy aspects of these statements are that the positive predictive value and

hence the true positive rate, as well as the negative predictive value and hence the true negative rate, are a function *not only of the test but also of the population mix.*

This is easily illustrated with the following table with four hypothetical cases, all with a sensitivity of 90%, two with a nonspecificity of 10%, and two with a nonspecificity of 30%. For each nonspecificity the prevalence is 90% for one case and 10% for the other:

	CASE 1	CASE 2	CASE 3	CASE 4
Sensitivity	.90	.90	.90	.90
Nonspecificity	.10	.10	.30	.30
Prevalence	.90	.10	.90	.10
True positive rate	.99	.50	.96	.25
True negative rate	.50	.99	.44	.75

With constant sensitivity and specificity, but a prevalence varying from 90% to 10%, we note the decrease in positive predictive value (true positive rate) and the increase in negative predictive value (true negative rate). Incidentally, we note that an increase in nonspecificity from 10% to 30% between case 2 and case 4 reduces the positive predictive value from 50% to 25%.

The accuracy is also a function of the population, as it is of the intrinsic test characteristics. However, at the limit, accuracy is misleading. Indeed, in a population with a prevalence of 10% a test with a sensitivity of 0% and a specificity of 100%, i.e., a test that is *never* positive, would yield an accuracy of 90%.

It is not sufficient to know what the reported values of the sensitivity and the specificity are. It is also necessary to know the reliability of those values.

In this book the nonspecificity and the sensitivity are reported as fractions followed by the number of observations in parenthesis, i.e., 92 (50) means that the sensitivity was 92% and that this value is derived from 46 positive studies in 50 affected subjects.

A number of publications are not included in this review, because the data did not allow the distillation of the information needed. In many cases sensitivities and specificities are given as fractions, but the number of affected and unaffected subjects are not given. In other cases, those number are available, but the results are expressed as accuracies, while separate sensitivities and specificities are withheld.

The formulation of Bayes' theorem given previously cannot be applied to the clinical situation without dangerous oversimplifications. The limitations and dangers of this approach will be discussed further in chapter 11, and some will be pointed out in the chapters on clinical applications. But a few caveats are in order here: A particular procedure does not have a single sensitivity for a particular disease. More often than not, the sensitivity is higher for more advanced disease. Conversely, if there are degrees in the detected abnormalities, the more extreme outcomes often tend to be less sensitive (less frequent) but more specific. Binary classifications, which seem favored by many authors and even more so by journal editors, and which we had to accept to report the published results, hide the fact that the outcome can be more or less positive, and that the predictive values or the confidence of the predictions are variable from case to case even in a given population.

The nonspecificity, or the likelihood of a positive outcome in an unaffected population, is heavily influenced by the composition of this group. A test result could almost always be negative in a healthy population, but more often positive in a clinic population composed of people with complaints or diseases different from those for which one tests.

Finally, the simplified formulation given above does not accommodate multiple outcomes from one procedure, an aspect that will be discussed in chapter 11.

A final word on terminology: The early imaging systems were moving detector systems, which "scanned" an area of the body. Those systems are now largely replaced by scintillation cameras, which have large, nonmoving detectors. The term *scanning,* loosely applied to imaging of radioactivity, is therefore inaccurate. There are exceptions: in some procedures whole body imaging is used, and the large detector is used to scan the whole body. However, we have used the general term *scintigraphy,* since most detectors used today are scintillation detectors. But the reader should not be distressed by the wide variety of terms used in the bibliography.

In ten chapters of this book we present a number of applications of scintigraphic procedures that in the aggregate represent a large fraction of clinical nuclear medicine, but the list is not exhaustive. There are a number of obvious omissions, e.g., gallium scintigraphy for the detection of focal inflammation is not included. We preferred to describe the use of labeled leukocytes for this application (chapter 8).

Brain scintigraphy is not described, having been largely replaced by computed axial tomography. Renography (chapter 4) is discussed only in relation to the diagnosis of renovascular hypertension, even though it has applications in other renal diseases. The applications of scintigraphic ventriculography include, in addition to the diagnosis of coronary artery disease (chapter 1), the evaluation of cardiomyopathy, including adriamycin-induced cardiomyopathy. In those two cases the selected applications are the diagnostic ones rather than the evaluative ones.

The hope is that, with all its shortcomings, this book contributes to a rational and efficient use of scintigraphic procedures.

1

THE DETECTION OF CORONARY ARTERY DISEASE USING STRESS (EXERCISE) SCINTIGRAPHIC VENTRICULOGRAPHY

WITH JOSIANE BRETILLE, DOCT. MED.

FOR SCINTIGRAPHIC VENTRICULOGRAPHY a radioactive intravascular tracer is injected intravenously. The analysis is based on the detection of changes in count rates during the cardiac cycle in the region of the left ventricle. One usually distinguishes two methods. In the first one the data are acquired during the first passage of a radioactive bolus through the central circulation (first-pass nuclear angiocardiography, or FPNA). In the second, data are acquired with the tracer homogeneously distributed within the vascular system, or at equilibrium. The cardiac cycle and the acquisition are synchronized using a gating mechanism based on the detection of the R wave

on electrocardiogram (ECG) (equilibrium [ECG] gated nuclear angiocardiography, or EGNA).

The results are not immediately available in either method, but require a variable amount of data processing or information extraction. The differences in results one might expect between both methods tend to be moderated by appropriate processing techniques. Methodologically, the major difference is the fact that FPNA favors punctate measurements, while EGNA favors continuous measurements. In both cases the data consist of an "image" of a representative cardiac cycle. The images obtained through EGNA measurements are an average of measurements obtained during a large number of cardiac cycles. In contrast, FPNA images are based on measurements of six or even fewer cardiac cycles.

The most commonly derived parameter of left ventricular function is the ejection fraction (EF), either at rest (REF) or during stress. One can estimate end-diastolic and end-systolic volumes (with some difficulties), or the relative changes in those values during the exercise measurements vs. the resting state (with relative simplicity). One group[2] has proposed an ejection rate, which is the EF divided by the systolic time interval, but this parameter has not proved to be particularly useful.

In addition to those global measures, regional ventricular kinetics can easily be evaluated by a variety of means. The most popular method is purely visual and analogous to the radiologic cine method. The data from the single representative cycle (which is an average in itself) is visualized in an apparent endless loop. In this way noise is easily distinguished from phenomena that have the same periodicity as the cardiac cycle. Dyskinesis is particularly easy to discern. Alternatively—and again in analogy to the radiologic methods—one can trace the outer edges of the ventricular cavity in the end-diastolic and the end-systolic frames and show them in superposition. Approaches that are more specific for the scintigraphic methods include the computation of functional images and the analysis of the data based on the first component of a Fourier series that would fit the kinetics of individual pixels.

The global ventricular behavior during stress is characterized by the absolute value of the EF at the end point of the stress (EEF), or by the change between this value and the resting value (dEF = EEF − REF).

Most authors use a binary classification of symptoms and signs,