Special Report:

American College of Radiology Diagnostic Efficacy Studies

JOHN W. LOOP[1] AND LEE B. LUSTED[2]

Questions about costs and benefits of radiologic diagnostic methods have lead to the formation of a standing American College of Radiology Committee on Efficacy. In the past several years this committee has defined a hierarchy of efficacies associated with radiographic procedures and developed methods for assigning a numerical value to diagnostic efficacy. A national study of common x-ray examinations in the emergency setting has been conducted. These studies suggest that diagnostic thinking of clinicians was influenced by the results of the x-ray procedures they requested in more than 92% of cases. Medicolegal considerations were a prime concern of the clinician only 6% of the time overall when selecting the most frequent radiologic procedures conducted in hospitals. Considerable uncertainty about clinical diagnoses was typically present at the time radiologic examinations were requested; three-fourths of the time the most important diagnosis under consideration was judged to be less likely after x-ray examination than before. About one-eighth of the time the radiographic information focused attention on a new "most important" diagnosis. The committee plans further studies to include other imaging modalities and practice settings.

The enthusiastic endorsement of computed tomography by radiologists has heightened the problem of cost justification for medical diagnostic technology in particular and for health care in general. The issue is not new, however. A decade ago physicians were advised that public policy would soon require them to carefully consider the "economic as well as medical consequences of their decisions" [1]. They were enjoined to resist the "technological imperative," the practice of providing the best care technically possible, regardless of cost. Just how an economic imperative could be ethically substituted for a technical one has since been a subject of concern, with more questions raised than answered.

Cost justification from the radiologist's prospective requires that two separate questions be resolved: (1) Are radiological examinations conducted in a cost efficient manner? and (2) Are patients selected prudently for radiologic procedures? Most radiologists will feel more comfortable answering the first question than the second. For various reasons, it has not been their practice to second-guess the physician who refers patients to them. At the least, they have believed, such a practice could become very demanding of their time, and thus expensive, while conferring uncertain benefits to the patient.

Assuming the economics of health care compel a more exacting selection of patients for radiologic examinations, who ought to make the selection? What should be the selection criteria? What will the implications be for radiologic practice and radiologic training? These and related questions were the background for the decision of the American College of Radiology to undertake a long term study of radiologic efficacy.

In 1971 an Efficacy Studies Committee chaired by Lee B. Lusted was authorized by the chancellors of the American College of Radiology. This committee was instructed to formulate reports and recommendations to the college and its membership for guidance in the efficacious use of radiology in medical diagnosis.

Some important background work had already been carried out by Bell and Loop [2], who had investigated the question of cost effectiveness of plain skull radiography for trauma. Their main conclusion was that in the absence of certain key signs or symptoms, the clinician had little to gain diagnostically from subjecting his patient to skull x-ray examination, and the patient had even less prospect of therapeutic benefit. Lusted had already addressed similar issues at a theoretical level [3, 4].

Planning

The committee began to meet regularly to formulate a plan to accomplish the following: to refine the definition of efficacy; to develop a method of measuring efficacy; and to study the efficacy of procedures used in diagnostic radiology. To practically define efficacy proved not a simple task. One indicator of the effectiveness of a diagnostic test is its yield, that is, the proportion of abnormal to normal examinations. Yield does not fully express efficacy, however, since diagnostic information has a different value for different diseases (e.g., for those which can be cured by treatment compared to those which cannot). Further, negative examinations may not be counted as yield, yet a negative examination can prevent unnecessary diagnostic studies or treatment. It can also contribute to differentiating among diseases

Received March 14, 1978; accepted after revision April 4, 1978.

This work was supported in part by contract no. HSM 110-72-293 with the Health Services and Mental Health Administration, Department of Health, Education, and Welfare, and grant no. 1 R 18 HS01546 from the Center of Health Services Research HRA/DHEW.

[1] Department of Radiology, University of Washington, Harborview Medical Center, 325 Ninth Avenue, Seattle, Washington 98104. Address reprint requests to J. W. Loop.

[2] Department of Radiology, University of Chicago, Chicago, Illinois 60637.

which are typically associated with negative radiographic results in different frequencies [5].

The fullest and most long-range expression of efficacy ought to include some measure of the influence of the examination on the final outcome of the episode of ill health. The efficacy concept could even be broadened to express the costs and benefits of the procedure to other persons such as the patient's relatives, employer, health care providers, health cost underwriter, or attorney.

Abstractly, statistical decision theory offers a framework for the definition of efficacy. A decision maker is involved who must make choices in the face of uncertainty. Theory guides this decision maker, typically a physician in clinical situations, to assign some numerical probability for the presence of each diagnosis which is suggested by all the evidence available at the time. A value judgment can then be attached to every scenario that can be expressed as a consequence of alternative treatment plans. The costs and risks of any therapy (or no therapy) must be balanced against the costs and risks of applying that therapy for an incorrect diagnosis. A trade-off can be conceived of between the cost of gaining more diagnostic certainty and the consequences of inappropriate treatment. Maximum utility is achieved at a balance point between these costs and risks.

A formal system of mathematical analysis exists for examining decisions of this type [6, 7]. The practicality of applying such formulations to the clinician in a busy practice needed to be tested. Such a practical test was carried out at the University of Michigan in 1972 and 1973 [8, 9], in cooperation with the American College of Radiology Efficacy Committee.

This team was interested in the clinical decision background of intravenous urography. Their original aim was to measure, if possible, the most comprehensive expression of efficacy in its theoretic framework: Was the patient better off as a result of the procedure having been performed? This level of efficacy, outcome efficacy (or E-3), proved not to be measurable within the limitations of their study. It would have required long term evaluation of randomized groups of patients, since alternative therapies would have to be included.

A more limited approach to efficacy measurement could be contemplated based on the degree to which clinical management was influenced by the intravenous urogram. Such a measurement of therapy planning efficacy was termed therapeutic efficacy (E-2). The Ann Arbor investigators [8, 9] were not able to overcome the difficulties associated with practical expression of this utility. Clinicians balked at the prospect of formulating a treatment plan for a patient with, say, hematuria, who had not had a urographic contrast study. This experience resulted in the deferral by the American College of Radiology Efficacy Committee of attempts to measure efficacy at this level.

The approach finally selected for the pretest emphasized the influence of the radiographic information on the diagnostic thinking of the clinician. This level of efficacy, termed diagnostic efficacy (E-1), is less comprehensive than either E-2 or E-3 but has the advantage of pinpointing the contribution of the radiology procedure most sharply and soonest after its performance. Clinicians were required to formulate a differential diagnosis and assess in numerical terms the probability that each disease under consideration was present. This the clinicians were able to do with reasonable fidelity after preliminary instruction.

The analytic technique employed in the Michigan pretest (and in later studies) for measuring E-1 required that the probability of diagnosis be estimated both before and after radiographic results were known. These estimates could take the form of either percentage probability or odds. In the form of odds, the initial odds (IO) for a diagnosis multiplied by a measure of relevant radiologic information yields final odds (FO) for this diagnosis. Transposing, FO/IO = measure of radiographic information = likelihood ratio. This latter expression is a form of Bayes's theorum, a venerable rule in probability theory [9].

Radiologic information is thus shown to be equivalent to the ratio of final to initial odds, also termed the likelihood ratio. The numerical value of this ratio can be either larger or smaller than unity depending on the influence of the information on final odds. Only where the final and initial odds are identical is the likelihood ratio equal to one.

It is convenient to convert these ratios to their logarithms. In this form radiographic procedures which are irrelevant to the diagnosis of concern have an information value of zero. All which contribute have a nonzero value. The committee decided to ignore some of the E-1 which could be detected by limiting numerical efficacy estimates to only two of all the diagnoses that might conceivably be present in each case: the *most likely* diagnosis and the *most important* (e.g., threatening) diagnosis. In information theory jargon, this limitation would be termed "pruning the decision tree."

The committee further proposed to gather enough information about each case studied to convince itself that the diagnostic estimates were plausable and to determine how well signs and symptoms alone might substitute for clinical judgments. It also wanted to design a study in which utilities for someone other than the patient could be expressed.

The committee was committed to investigate the most economically significant x-ray procedures. As expected, those proved to be the most commonly performed examinations: chest, extremities, lumbar spine, cervical spine, abdomen, skull, and excretory urogram. Together these procedures represent about 90% of the volume of hospital x-ray examinations in the United States. In order to accommodate collection of data, distinctive questionnaires were developed for each type of radiographic procedure which displayed a checklist of signs and symptoms frequently expected when the procedure was requested (fig. 1). These plans together with a detailed proposal for conducting definitive testing were developed through a contract between the American College of Radiology and the Health Services and Mental Health Administration, DHEW, which ran from June 1972 until a final report in May 1974 [10].

A Nationwide Test

The Michigan pretest had suggested the limits of cooperation which could be expected from busy clinicians in supplying written numerical probability estimates for tentative diagnoses. It also helped shape the sampling strategy eventually employed for the national study. An administrative structure was established with headquarters at the Chicago office of the American College of Radiology. The organization extended through regional and local radiologic leadership to cooperating clinicians.

Training methods and materials, developed under auspices of the central project management, flowed out through this structure, and subsequently data concerning sample cases moved in the opposite direction. Recruiting relied heavily on personal friendships and established referral patterns, rather than following a more traditional probability sampling scheme. The sampling was not haphazard, however. Conscious attempts were made to spread the data base geographically and by practice style. Thus 48 hospitals in 21 states were sampled, including a mix of large and small and teaching and nonteaching institutions.

Because resources were limited, the initial national study focused on the hospital emergency room and its clientele. The belief was that since radiologic results are usually available promptly in this setting, their effect would be minimally diluted by changes in diagnostic certainly caused by other information or the passage of time.

From one to eight clinicians participated in each hospital, and they contributed a median of 14 randomly selected cases of one or more x-ray procedure types. The final data base included 8,658 cases (table 1). While the methodology employed in the national study has been reported [11, 12], a few highlights are of interest here.

The study was designed to protect the confidentiality of information while preserving at the local level the ability to recover and review case records in follow-up studies. In each case the radiologist's diagnosis was accepted as final for purposes of the study. Selected chart review has very rarely shown contradiction of this diagnosis.

The information collected on the questionnaire (fig. 1) for each procedure included a checklist of clinically pertinent signs and symptoms. The clinician was free to and was encouraged to enter other historic (as reported) or objective (as found) data of possible interest. Utilities other than the patient's health could also be recorded on the collection form. Of interest here was the influence, if any, of patient's health insurance status. Medical, legal, or research interests of primary importance could be identified as well and related to efficacy. Rationales for the procedure other than establishing a diagnosis could similarly be recorded (e.g., to locate a foreign body known to be present).

TABLE 1

Distribution of Cases

Procedure	No. Cases
Skull	1,039
Cervical spine	958
Chest	2,627
Abdomen	957
Excretory urogram	285
Lumbar spine	807
Extremities	1,985
Total	8,658

Findings

A report which contains the analysis of results and conclusions of the initial study through May 1977 is available from the American College of Radiology (20 North Wacker Drive, Chicago, IL 60606; $10). Further reports are planned since these data represent the largest single collection of numerical probability estimates ever assembled in a medical application.

Some highlights of this initial study were presented as an exhibit at the annual meeting of the Radiological Society of North America, Chicago, November 1977. While further refinement of numerical estimates can be predicted, several general conclusions are now well established: (1) the influence of x-ray examinations on clinicians diagnostic thinking can be measured; (2) medicolegal reasons were cited only about 6% of the time as the prime reason for requesting x-ray examinations; (3) x-ray examination had an impact on diagnostic thinking in as many as 92% of situations; (4) diagnostic uncertainty at the time of x-ray request was substantial; (5) the radiologic examination reassured the physician; and (6) results for different procedures and training levels were similar.

It is feasible to measure the influence of x-ray information on the diagnostic thinking of physicians. This conclusion is supported by the internal consistency of the case data and by other findings of the study. Clinicians in the emergency room gave a good account of their diagnostic thinking. Numerical probability estimates appear to be superior to such terms as highly likely, very likely, moderately likely . . . highly unlikely, in principle and in this type of application. However, 9% of the questionnaires had to be rejected for various reasons, and some physicians would not cooperate on this type of sampling.

The vast majority of the x-ray procedures studied produced information which influenced diagnostic thinking. Overall not more than 8% of examinations seemingly had no influence in changing probability assessments for the most important and most likely diagnoses or in requiring formulation of new diagnoses. Refined analysis suggests that the actual figure may be

Last Name ______________ First Name ______________ Patient I. D. (if known) ______________

Date of Birth (estimate if necessary) ______________ Sex ______________ Case Number ______________

AMERICAN COLLEGE OF RADIOLOGY - EFFICACY STUDY: ABDOMEN-EMERGENCY

PART I (TO BE COMPLETED BY CLINICIAN BEFORE RADIOLOGIC PROCEDURE)
(See CLINICIAN'S HANDBOOK for guidance in completing this form.)

A. Clinical Data: For each entry check one box only. (Y-Yes, N-No, ?-Equivocal, ND-Not Determined)

Y	N	?	ND	WAS REPORTED
				Recent Trauma
				Indigestion
				Abdominal Pain
				Nausea or Vomiting
				Change in Bowel Habits
				Hematemesis or Melena
				Fever or Chills
				Abnormal Urination
				Abnormal Menstruation
				Other ______________ (Specify)

Y	N	?	ND	WAS FOUND
				Physical Evidence of Injury
				Abdominal Tenderness
				Spasm or Guarding
				Mass
				Enlarged Organ
				Jaundice
				Distention or Abnormal Bowel Sounds
				Evidence of Alcoholism
				Abnormal Urinalysis
				Previous Abdominal Surgery
				Other ______________ (Specify)

B. What is your patient's one PROBLEM that causes you to request this examination? ______________ Code: ___ ___ ___ . ___

C. 1) For the problem in B, state the most important prospective radiographically-relevant DIAGNOSIS which prompts this procedure. ______________ Code: ___ ___ . ___ ___ ___

2) What are your odds or probability estimate that the diagnosis in "C-1" will prove correct? ______

D. 1) For the problem in B, state the most likely prospective radiographically-relevant DIAGNOSIS ("normal" may be used) which prompts this procedure (only if different than the diagnosis in C). ______________ Code: ___ ___ . ___ ___ ___

2) What are your odds or probability that the diagnosis in "D-1" will prove correct? ______

E. What is the one major reason for this procedure? (Check one box only)

- ☐ Prove part normal
- ☐ Confirm no change
- ☐ Institutional policy
- ☐ Confirm diagnosis
- ☐ Show change in disease or healing
- ☐ Teaching or research
- ☐ Investigate diffuse suspicions
- ☐ Assess length, position, etc.
- ☐ Medical-legal
- ☐ Other ______________

F. Are you presently aware of patient's medical insurance status?

☐ Not Aware Believe patient is: Insured ☐ Not Insured ☐

Your Name ______________ (Please Print) and/or ACR I. D. Number ______________ Date Filled Out ______

RETURN TO RADIOLOGY AFTER COMPLETING PART II

A

NOT A PART OF MEDICAL RECORD

2/75

Fig. 1A.—Front page of typical questionnaire. These data were supplied by clinician before x-ray procedure.

PART II TO BE COMPLETED BY CLINICIAN AS SOON AS RADIOLOGIC RESULTS ARE KNOWN

G. Knowing the X-ray findings, now estimate the odds or probability that the:

1) "most important" diagnosis stated in "C-1" of Part I is correct ________

2) "most likely" diagnosis stated in "D-1", if any, of Part I is correct ________

H. Enter below any NEW diagnoses based on radiological findings?

1) most important new diagnosis ________ Code: __ __ . __ __ __)

2) most likely new diagnosis (include normal) ________ Code: __ __ . __ __ __)

Your Name ________ (Please Print) and/or ACR I. D. Number ________ Date Filled Out ________

SIGNIFICANT RADIOLOGIC FINDINGS (To be filled out by radiologist or referring physician):

TO BE COMPLETED BY RADIOLOGY

RADIOLOGIC PROCEDURE CODE: __ __ __ __ __

RADIOLOGIC DIAGNOSES CODES Dx1 __ __ . __ __ __ __ __ Dx2 __ __ . __ __ __ __

Dx3 __ __ . __ __ __ __ __

SETTING (check one) ☐ Screening ☐ Inpatient

☐ Emergency ☐ Outpatient

RETURN TO Dr. ________ IN RADIOLOGY AFTER COMPLETING PART II

B NOT A PART OF MEDICAL RECORD

Fig. 1*B*.—Reverse side of questionnaire. Questions were answered by clinician and local radiologist after x-ray results were available.

less than 5%. This contradicts the widespread belief that radiographic procedures are often carried out without yield, at least in the emergency setting. Cases with no difference in initial and final odds ranged from 13.1% of 1,039 skull examinations to 3.8% in 1,985 extremity examinations. If cases in which the initial probabilities are very long (e.g., 50:1 for or against) are disregarded, the percentage of valueless procedures is reduced.

At the time radiologic examination is requested, the tentative most important diagnosis is substantially uncertain. Only in one case in five is this diagnosis thought more likely than not to be present. The initial odds for the most important but relatively remote diagnosis were especially low in skull and cervical and lumbar spine, less so in chest, abdomen, and extremities, and least in excretory urography.

In about three-fourths of examinations the clinicians final probability for the tentative most important diagnosis was lower than his initial probability. Some differences in this trend were noted among procedures, but these were not striking. As might be predicted, radiographs are often used in the emergency setting to "rule out" dangerous conditions.

Clinicians express less uncertainty concerning their tentative most important diagnosis after x-ray examination than before. Judging from final odds, substantial uncertainty about final diagnosis persists even though the effect of x-ray information substantially reduces the degree of doubt.

Medicolegal reasons were cited in only about 6% of cases as the prime reasons for requesting x-ray examinations. Even in these cases diagnostic efficacy could be detected, but the apparent influence of radiographic information on diagnostic thinking was less than in other cases. Medicolegal reasons were cited as the primary reason for cervical spine, skull, and lumbar spine examinations in 19.4%, 14.3%, and 8.9% of cases, respectively. Least frequently cited was abdomen (1.5%).

In about 16% of cases, radiographic information led to a change of the initial most important or most likely diagnosis in favor of a new diagnosis. Relatively few of these seem to represent genuinely serendipitous findings.

The influence of x-ray information on diagnostic thinking was broadly similar for interns, resident physicians in training, and practicing physicians. Most other characteristics, such as the distribution of initial probabilities for diagnoses and the distribution across procedures, were similar for the three groups. This same conclusion holds for final probabilities.

There is a strong relationship between the clinician's initial probability and the subsequent fraction of positive diagnoses, whether the latter is measured by the clinician's final probabilities or by the radiologist's film interpretations. The calibration is imperfect, however. On the average, fewer positive diagnoses are reached than would have been expected from the initial probabilities. This tendency to overassessment is most constant and severe for suspected skull and cervical spine fracture, less so for other frequently cited diagnoses such as pneumonia or bowel obstruction.

The tentative findings for emergency room practice do not indicate that paramedical personnel guided by a handbook or a computer program of signs and symptoms would be an acceptable substitute, in general, for the judgment of the clinician who requests the x-ray examination. However, guidelines presented as tables and graphs relating patient signs and symptoms to initial odds or probabilities for disease could help clinicians and paramedics working under clinical supervision. The information to be expected from a particular x-ray examination then could be related to the diagnosis. Such guidelines are not available to emergency room physicians at present in any comprehensive form to our knowledge.

Discussion

These efforts are viewed as largely preliminary and do not pretend to be a defense of the status quo. But some level of practice which is both economically and technically acceptable to all parties does exist and needs to be identified. Who is responsible to identify this level? We have been reminded again recently that the alternative to physicians facing up to tough allocations will be explicit rationing of services based on administrative priorities which limit physician discretion [13].

These studies indicate that radiologists at least are prepared to face up to these problems. The whole question of radiologic cost-benefit has now joined the traditional scientific topics in our national forums and journals. We see this as a sign of vigor and health in the profession, fulfilling the best hopes expressed by Bell and Loop [2] in their pioneering paper.

The American College of Radiology Efficacy Committee claims no special insights as it has developed and applied the methodology briefly described above. The committee is prepared to embrace new methods should these prove to be superior. The hope is that future efficacy studies can proceed in concert, so that pitfalls avoided by one can be circumvented by all, and that results can be pooled effectively for common benefit.

ACKNOWLEDGMENTS

We wish to express our obligation to our co-investigators, the more than 400 fellow physicians who participated in these studies. Special thanks are due to R. S. Bell, R. D. Moseley, D. L. Searle, J. P. Steele, and J. R. Thornbury for radiologic leadership in these projects. W. Edwards, D. G. Fryback, M. Lahiff, H. V. Roberts and D. L. Wallace provided expert consultation in decision theory and statistics. Cheryl Gabel typed the manuscript.

REFERENCES

1. Fuchs VR: The growing demand for medical care. *N Engl J Med* 279:190–195, 1968
2. Bell RS, Loop JW: The utility and futility of radiographic skull examinations for trauma. *N Engl J Med* 284:236–239, 1971

3. Lusted LB: *Introduction to Medical Decision Making* Springfield, Ill., Thomas, 1968
4. Lusted LB: Decision-making studies in patient management. *N Engl J Med* 284:416–424, 1971
5. Gorry GA, Pauker SG, Schwartz WB: The diagnostic importance of the normal finding. *N Engl J Med* 298:486–489, 1978
6. Raiffia H: *Decision Analysis*. Reading, Mass., Addison-Wesley, 1968
7. Schlaifer R: *Analysis of Decisions under Uncertainty*. New York, McGraw-Hill, 1968
8. Thornbury JR, Fryback DG, Edwards W: Likelihood ratios as a measure of the diagnostic usefulness of excretory urogram information. *Radiology* 114:561–565, 1975
9. Thornbury JR, Fryback DG, Edwards W: Studies on the use of the intravenous pyelogram, in *Current Concepts in Radiology*, vol 2, edited by Potchen EJ, St. Louis, Mosby, 1975, pp 59–73
10. Lusted LB: Efficacy study; final report. Reference HSM 110-72-293. Submitted to Abraham Ringel, associate director, Division of Professional and Technical Development. Regional Medical Programs Service Administration, May 29, 1974
11. Lusted LB: A study of the efficacy of diagnostic radiologic procedures. Final report on diagnostic efficacy. Chicago, American College of Radiology, May 31, 1977
12. Lusted LB, Bell RS, Edwards W, Roberts HV, Wallace DL: Evaluating the efficacy of radiologic procedures by Bayesian methods; a project report, in *Models and Matrices for Decision Makers*, edited by Snapper K, Washington, D.C., Information Resources Press. In press, 1978
13. Mechanic D: Approaches to controlling the cost of medical care: short range and long range alternatives. *N Engl J Med* 298:249–254, 1978