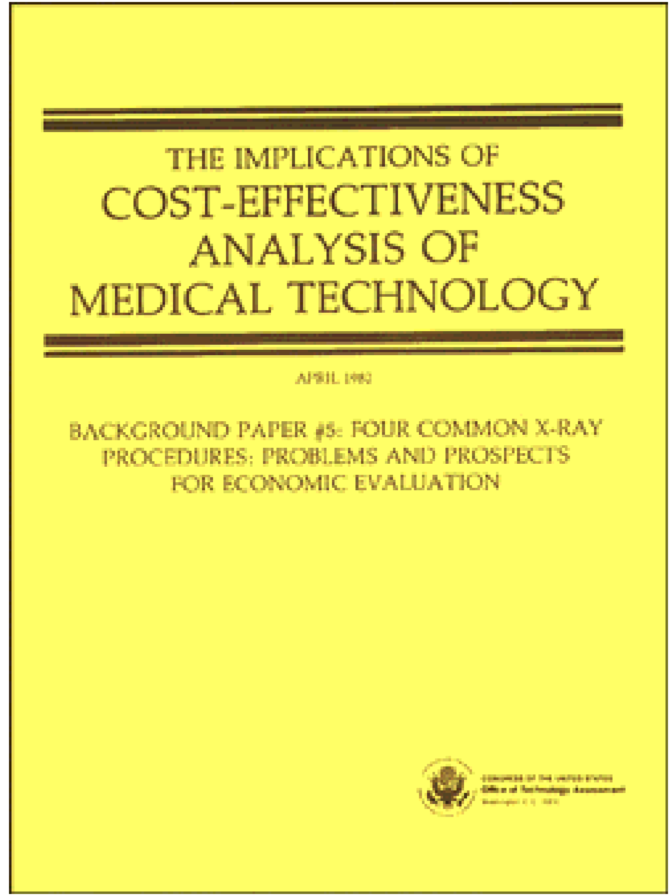
Four Common X-Ray Procedures: Problems and Prospects for Economic Evaluation

April 1982

NTIS order #PB83-165506



THE IMPLICATIONS OF
COST-EFFECTIVENESS ANALYSIS OF MEDICAL TECHNOLOGY

APRIL 1982

BACKGROUND PAPER #5: FOUR COMMON X-RAY PROCEDURES: PROBLEMS AND PROSPECTS FOR ECONOMIC EVALUATION

by
Judith L. Wagner, Ph. D., and Martha J. Krieger
The Urban Institute, Washington, D. C.

OTA Background papers are documents containing information that supplements formal OTA assessments or is an outcome of internal exploratory planning and evaluation. The material is usually not of immediate policy interest and does not present options for Congress to consider.

CONGRESS OF THE UNITED STATES
Office of Technology Assessment
Washington, D. C. 20510

Library of Congress Catalog Card Number 82-600532

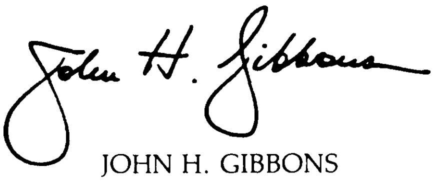
For sale by the Superintendent of Documents,
U.S. Government Printing Office, Washington, D.C. 20402

Foreword

This volume is a background paper for OTA'S assessment, The *Implications of Cost-Effectiveness Analysis of Medical Technology,* Prepared at the request of the Senate Committee on Labor and Human Resources, that assessment analyzes the feasibility, implications, and usefulness of applying cost-effectiveness and cost-benefit analysis (CEA/CBA) in health care decisionmaking. The major, policy-oriented report of the assessment was published in August 1980, In addition to the main report and this paper on the economic evaluation of diagnostic X-ray procedures, there are four other background papers: 1) a document which addresses methodological issues and reviews the CEA/CBA literature, 2) a psychotherapy case study, 3) 17 other case studies of individual medical technologies, and 4) a review of international experience in managing medical technology. Another related report was published in September of 1979: *A Review of Selected Federal Vaccine and Immunization Policies.*

Background Paper #.5: Four Common X-Ray Procedures: Problems and Prospects for Economic Evaluation was specifically requested by the Senate Committee on Finance. It was prepared under contract for OTA by Judith L. Wagner, Ph. D., and Martha J. Krieger, of the Urban Institute.

Since OTA does not make specific recommendations of policy action or endorse particular technologies, the views presented in this study are strictly those of the authors. OTA encouraged the authors to present balanced information and also subjected the study to an extensive review process. Initial drafts were reviewed by OTA staff and members of the advisory panel for the overall assessment, and subsequent drafts were then reviewed by more than 20 outside reviewers from Government agencies, professional societies, consumer and public interest groups, medical practice, and academia. Although the many reviewers for this background paper and the other parts of the assessment cannot be acknowledged individually, OTA is very grateful for their comments and advice.



JOHN H. GIBBONS
Director

Advisory Panel on The Implications of Cost-Effectiveness Analysis of Medical Technology

John R. Hogness, *Panel Chairman*
President, Association of Academic Health Centers

Stuart H. Altman
Dean
Florence Heller School
Bratzdeis University

James L. Bennington
Chairman
Department of Anatomic Pathology *and Clinical Laboratories*
Children's Hospital of *San Francisco*

John D. Chase
Associate Dean for Clinical Affairs
University of Washington School of Medicine

Joseph Fletcher
Visiting Scholar
Medical Ethics
School of Medicine
University of Virginia

Clark C. Havighurst
Professor of *Law*
School of Law
Duke University

Sheldon Leonard
Manager
Regulatory *Affairs*
General Electric Co.

Barbara J. McNeil
Department of *Radiology*
Peter *Bent* Brigham Hospital

Robert H. Moser
Executive Vice President
American College of Physicians

Frederick Mosteller
Chairman
Department of Biostatistics
Harvard University

Robert M. Sigmond
Advisor on Hospital Affairs
Blue Cross and Blue Shield Associations

Jane Sisk Willems
VA Scholar
Veterans Administration

OTA Staff for Background Paper #5

H. David Banta, *Assistant Director, OTA*
Health and Life Sciences Division

Clyde J. Behney, Health *Program Manager*

Clyde J. Behney, *Project Director, Implication of Cost-Effectiveness Analalysis of Medical Technology*

Kerry Britten Kemp, * *Editor*
Virginia Cwalina, *Research Assistant*
Lorraine Ferris, *Secretary*
Nancy L. Kenney, Secretary

OTA Publishing Staff

John C. Holmes, *Publishing Officer*

John Bergling | Kathie S. Boss | Debra M. Datcher | Joe Henson

*OTA contract personnel

AUTHORS' ACKNOWLEDGMENT

The authors of this background paper benefited from an earlier paper prepared by David Collier, M. D., for the Office of Technology Assessment. That paper was most helpful in directing the authors to the pertinent literature and providing summaries of the research field.

Contents

List of Tables

1 .

Overview and Summary

Overview and Summary

INTRODUCTION

This paper is about the economic evaluation of diagnostic procedures. To be precise, it is about whether and when the benefits from a diagnostic procedure are worth its risks and costs.

The issue of economic evaluation is explored in the context of four common diagnostic X-ray procedures that together in 1970 accounted for almost half of all diagnostic X-ray examinations in the United States:

- the chest X-ray,
- the skull X-ray,
- the barium enema study, and
- the excretory urogram.

Since these are all long-established and widely used radiological procedures, evidence of their benefits, risks, and costs should be **as** comprehensive as that of most other diagnostic procedures. The methods used to evaluate these procedures should (and do) represent the general state-of-the-art in the assessment of most diagnostic technologies.

The four X-ray procedures are also interesting in their own right. As high-volume procedures, they use up substantial health care resources. Table 1 summarizes the findings of a 1970 survey of diagnostic X-ray procedure use in the United States. Of the 130 million diagnostic X-ray procedures performed in that year, 60 million were for the four procedures of interest here. The average amount billed by a sample of radiologists in California in 1975 for each of the four procedures is shown in table 2. Though billed charges overstate the amount actually received and data from California do not represent the nation, the total burden of expenditures for X-rays is substantial.

The appropriateness of use of each procedure has been the subject of intense debate within the

Table 2.—Average Amount Billed by Radiologists for Selected X-Ray Procedures, California Medicaid Program, 1975

Procedure	Average billed amount
Chest X-ray:	
Single	$13.03
Double	22.73
Complete	31.61
Barium enema:	
Colon, barium enema	45.32
With air contrast	64.29
Air contrast (independent procedure)	NA
Excretory urogram:	
Routine intravenous pyelography	53.51
Extended hypertensive	59.60
Infusion, DIP	70,00
Skull X-ray:	
Limited series	21.71
Complete (minimum of four views)	41.53

NA = not available

SOURCE Urban Institute sample of approximately 5.0130 solopract It loners, including 177 radiologists (32 6 Percent of the solo radiologists in Call fornia)

Table I.— Estimated Number of Diagnostic X-Ray procedures in the United States, 1970 (in thousands)

Type of examination	All	Hospitals	Private office: Radiologists	Private office: Others	Private groups	Health agencies and others	Average number of films per examination (all sources)
All radiologic procedures	129,070	81,688	3,334	20,419	8,923	14,708	2,4
Skull	4,220	3,616	NA	NA	NA	NA	4.03
Chest radiograph	48,569	32,491	NA	7,565	3,207	4,293	1.7
Barium enema	3,428	2,774	NA	NA	NA	NA	3.!34
Excretory urogram	3,996	3,413	NA	NA	NA	NA	5.31

NA = not available

SOURCE Department of Health, Education and Welfare Public Health Service. *Population Exposure to X Rays,* U S 7970 November 1973 (30) medical profession. These debates have virtually always been brought about by the findings of evaluative research. Indeed, evaluations of the benefits of these procedures seem to have raised rather than answered questions about proper indications for their use, especially when the results have contradicted prevailing patterns of medical care. Thus, a critical review of the evaluative research on each of the four diagnostic technologies is a good way to characterize the current controversies surrounding each of these procedures.

COST-BENEFIT AND COST-EFFECTIVENESS ANALYSIS IN THE EVALUATION OF DIAGNOSTIC X-RAY PROCEDURES

Cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA) are methods to assist in allocating scarce resources among alternative uses. These methods were developed primarily to evaluate large public-sector investments such as highways, dams, and airports. When applied to diagnostic procedures, they are intended to provide information on two related questions: 1) Under what circumstances should the procedure be performed? and 2) How much investment in capacity to perform the procedure is justified? The answer to the second question rests on thorough study of the first, for only by knowing when a procedure should be performed can one assess how much investment in capacity is justified.

Applying the principles to medical procedures, CBA would enumerate and place a value on all benefits (both positive and negative) derived from performing a procedure on patients with a specified set of conditions and would compare those benefits to the cost of performing the procedure. The resulting net social benefit would indicate whether the procedure should be performed under the specified conditions. In traditional CEA, a measure of procedure effectiveness would be designated, and the ratio of that single measure to cost would be the critical item for resource allocation. Lives saved, life-years saved, quality-adjusted life-years saved, disability saved, and age-adjusted disability days saved are measures of effectiveness often chosen in studies of health care programs.

One can generalize the notion of economic evaluation to a social accounting framework, in which all dimensions of effect as well as cost are identified and their values estimated. It is difficult to argue with the soundness of knowing the direction and magnitude of all effects resulting from the performance of a procedure on a particular patient. However, the ideal of comprehensive and accurate evaluation is seldom met and may not be worth its own costs. Virtually all good evaluative studies are limited in the dimensions of cost and/or effectiveness under investigation.

The critical weakness of most studies reviewed in this paper is that they fail to consider one or more important implications of the procedure under study. By not dealing with these important dimensions of effect or cost, they leave open the possibility that their conclusions will be criticized or, worse, ignored. Yet, the paradox is that the conditions necessary to produce accurate information on the full array of effects and costs may not be achievable. The methodological and ethical problems of evaluative research, as well as its costs, frequently are barriers that cannot and perhaps even should not be overcome.

ORGANIZATION OF THIS BACKGROUND PAPER

This background paper is divided into two parts. The purpose of the first part is to summarize the different evaluative models underlying studies of the four diagnostic X-ray procedures and to lay out the strengths and weaknesses of each method. That part also identifies the conditions under which these models are likely to provide information that can affect patterns of medical care. The second part of the paper contains four separate chapters summarizing what is known about the utilization, costs, risks, and benefits of each procedure, with particular emphasis on the evaluative methods employed.

The review of the evaluative literature of the past decade is by no means comprehensive. To be included here, a study must have provided evidence pertaining to at least one ot the following two interrelated questions. Under what conditions, if any, should a particular procedure be performed? And, how should the procedure be performed? Though many clinical studies address these kinds of issues, further conditions were required of those included in the review: measures of benefit, risk, or cost had to be specified explicitly; two or more alternative diagnostic strategies had to be compared either explicitly or implicitly; and a large enough number of cases had to be analyzed to draw meaningful inferences. The application of these additional filters eliminated editorial opinions and case reports from the review.

SUMMARY OF FINDINGS

This study has provided insight into three important questions. What influence has the evaluation literature had on the use of each of the four diagnostic X-ray procedures? What factors limit the influence of evaluative findings on medical practice? And, what directions might evaluative research take to increase its influence over medical decisionrnaking?

We cannot be precise about the influence of the evaluative literature on medical practice, because data are unavailable on rates of use of X-rays over time. But it is possible to infer from the clinical literature whether a consensus has developed in response to evaluative findings. The influence of evaluative studies of X-rays in screening (symptomatic) contexts appears to be strong, but evaluations of X-rays in diagnostic contexts, where patients present with complaints or symptoms, seem to have little impact on medical standards or practice. In either context, the more dramatic the results of the evaluation, the more likely is the study to have an impact. For example, studies demonstrating very low diagnostic yield of skull X-rays in emergency rooms created general concern and have led to some change in practice in a few centers (99,126).

The reasons for the limited influence of evaluations lie partly in the evaluative studies themselves and partly in the health care system. Study methods often are so flawed that the resuits cannot be trusted. Patient selection bias due to uncontrolled study designs are a major problem. Radiologic methods are often unstandardized. All too often, a procedure is evaluated in a group of patients so heterogeneous in its presenting signs, symptoms, and risk factors that the results offer no guidance at all about who should be X-rayed and who should not. The evaluative criteria rarely include the ultimate benefits of the procedure. Time and time again, as studies are reviewed in subsequent chapters of this background paper, we conclude that the study findings are inadequate because the implications of the X-ray results for patient health and well-being and for medical costs are unknown.

A more fundamental barrier to the use of evaluation in decisions about X-rays lies in the conflict between the individual patient's best interest and society's best interest. The individual patient, who seldom has to pay the full cost of the procedure and often need not pay at all owing to the availability of insurance or other third-party payment, need not consider the costs against the benefits of the examination. The public as a whole, however, must make these tradeoffs. This may explain why the results of evaluations of X-ray screening programs are more influential than are evaluations of diagnostic uses of X-ray. Screening programs are often funded by public health agencies, not by insurance coverage.

How can evaluations have a greater impact? In the diagnostic context, investigators can begin to consider the effect of the diagnostic strategy on therapy and ultimately on patient outcomes. Attention should be given to the clinical significance of missing disease if an X-ray is not ordered. Independently, research by economists and social scientists on the reassurance value of negative X-rays might put these benefits into proper perspective. Greater care might be taken in separating evaluations for patient groups with different presenting conditions. The appropriate level of aggregation of patient characteristics should be considered explicitly in the study design. To this end, studies of the diagnostic yield of X-ray and symptoms will be suggestive of patient groups where more thorough, outcome-oriented analysis is warranted.

Still, these new research directions will have no effect on decisions about the use of X-rays until both physicians and patients accept as reasonable the possibility that the diagnostic process might be truncated when the costs of pursuing additional information outweigh its potential benefits.

2.
Evaluation Model

2
Evaluation Models

To understand the strengths and weaknesses of evaluation, one must keep in mind its fundamental purpose: to inform those who make decisions. The inferences drawn from an evaluative study of a diagnostic procedure are important, because they may influence decisions about its use. Therefore, one must carefully examine the assumptions underlying any approach to evaluation to determine the extent to which they provide direction or misdirection for decisions about the use of a procedure.

Ideally, one would want an evaluative study of a diagnostic procedure to provide precise measurement of the full array of medical and nonmedical benefits, risks, and costs resulting from alternative diagnostic strategies applied to a patient with specific signs, symptoms, and risk factors. Additionally, the ideal study would provide some method for comparing one kind of benefit or cost with another. Then, the decisionmaker could apply the findings in a relatively straightforward fashion to choose among alternative strategies.

Reality is not so accommodating. No studies of medical procedures deal with all possible benefits, risks, and costs. Nor do they often provide a method for collapsing multiple dimensions of benefit and cost into a single composite measure. Often, they aggregate findings over patients with diverse signs, symptoms, and risk factors, thereby limiting their usefulness for diagnostic decisions. Research costs, unfeasibility of measurement, and inability to make value tradeoffs dictate that all evaluative studies stop far short of the ideal. Yet, even partial approaches provide evidence that is useful. For example, most studies of the four X-ray procedures discussed here do not explicitly consider issues of cost, but some do reach important findings about the magnitude of the clinical benefits to be derived from the procedure. Indeed, the most important feature of diagnostic-procedure evaluation is not whether it has considered all possible dimensions of benefit, risk, and cost but what it uses as the "endpoint" of the evaluation (79). A study of the cost or accuracy of diagnosis may carry a message for medical decisionmakers that is quite different from a study that chooses as its endpoint the patient's ultimate health status or costs.

In our review of the literature on the four X-ray procedures, we were able to classify all studies into one of five broad categories based largely on their evaluative endpoints:

- studies of diagnostic "efficiency,"
- studies of diagnostic yield,
- studies of high-yield criteria,
- studies of diagnostic information, and
- studies of outcomes.

Each of these approaches is discussed in detail below. Each has its advantages and disadvantages. One must consider an evaluative model in terms of the cost and feasibility of analysis and the reasonableness of the assumptions about factors outside its scope. In short, no single category of evaluation is best in all cases; each has a place in the evaluator's bag of techniques.

STUDIES OF DIAGNOSTIC EFFICIENCY

Diagnostic efficiency is used here to refer to the capacity of a test to meet its immediate objective: correct diagnosis, A perfectly efficient test is one in which either of two kinds of error would never occur: 1) the detection of disease when in truth none is there (false positives), and 2) the failure to detect disease when it exists (false negatives). A perfectly inefficient test would always be in error. Between these two extremes lie an infinite number of combinations of frequency of the two kinds of error.

The classic measures of diagnostic efficiency corresponding to each of the two errors are sensitivity and specificity. Sensitivity is the proportion of individuals with disease whose test results are positive (true-positive rate); specificity is the proportion of normal individuals whose test results are negative (true-negative rate).

Knowledge of these error rates is essential to the appropriate interpretation of test results by clinicians. Only by knowing the test's sensitivity and specificity and the prevalence of the suspected disease in the population can the clinician accurately interpret a positive or negative finding on a test (55). Though probably no clinician is prepared to perform sophisticated analyses of pretest and posttest disease probabilities for every test he or she orders, implicit processing of such information does take place, and accurate assessments of these two components of diagnostic efficiency in different situations are useful in this regard.

The use of diagnostic efficiency to assist in the choice among alternative diagnostic strategies is another matter altogether. Here, we say that if one test is both more sensitive and more specific than another, it is more efficient and may be the procedure of choice, although differences in the cost of each test should also be considered before such a decision can be made.

The comparison among tests is frequently not so straightforward. The more sensitive test is often less specific than its competitor. Then, it is impossible to avoid considering the implications of the two kinds of errors. If the consequences of a false positive are major—perhaps the performance of expensive or risky followup tests or even application of inappropriate and dangerous therapy—high specificity is desirable. Conversely, if the implication of a false negative is dire—the deterioration of the patient's condition to an unredeemable state—then a highly sensitive test may be preferred. Indeed, whether either test is worth performing at all cannot be determined by examination of its sensitivity and specificity alone. One must know the implications of each kind of test result (true positive, false positive, true negative, false negative) for the health and well-being of the patient and for the cost of medical care (54). Nevertheless, these components of diagnostic efficiency must be known if such analysis of the test's implications for outcomes is to take place. The problem with studies using these efficiency measures is that they often draw inferences about the usefulness of one diagnostic strategy versus another without further analysis.

Summary indexes of diagnostic efficiency have been developed by collapsing sensitivity and specificity into composite measures. The most common is diagnostic accuracy, * defined as the proportion of all test results that are correct, and measured as the sum of true-positive and true-negative results divided by the total number of tests. This index of efficiency is attractive because it is comprehensible, but it is more dangerous than the separate use of sensitivity and specificity because it assumes that the two kinds of testing error (false positive and false negatives) are equally important. This assumption is arbitrary and, for the vast majority of diagnostic tests, invalid. By using it, the investigator has bought into the equal valuation of the two kinds of error. Inferences, usually incorrect, are almost inevitable.

Diagnostic accuracy has also been criticized because, unlike sensitivity and specificity, it systematically varies with the prevalence of disease in the sample of patients under study. If the sensitivity of a test exceeds its specificity, then the higher the prevalence of disease in the study sample, the higher will be the measured diagnostic accuracy. This property is viewed as dangerous by some, for it implies that by manipulating the selection of patients under study, the investigator has the power to predetermine measured accuracy (41). However, the source of this problem is not the measure itself but the inappropriate generalization of study findings to populations not represented by the study sample. If one test proves more accurate than another in a random sample of men over 65 who

*Semantic problems abound in the literature. Although accuracy is generally defined as above, the same measure is referred to by at least one author as "validity" (41).

are hospitalized for suspected lung disease, and if there are no suspected sampling biases, * then the test will also be more accurate in all such patients. In any event, this problem is minor compared to the assumptions of equal importance of all kinds of error.

Another index of diagnostic efficiency is the likelihood ratio, L, defined as the ratio of the true-positive rate (sensitivity) to the false-positive rate (1 – specificity), or,

$$L = \frac{\text{Sensitivity}}{1 - \text{Specificity}}$$

Since sensitivity and specificity take on values between O and 1, L must lie between O and 8. L can also be interpreted as the rate at which the odds in favor of disease prior to having test results are translated into odds in favor of disease after the test results are known. * * Thus, it is a measure of the information content of the test, Very high or very low values of L imply high information content, while a value of L near 1 implies that a test adds little information to assist in diagnosis.

Unlike diagnostic accuracy, the likelihood ratio does not vary with disease prevalence, but it, too, makes implicit assumptions about the relative importance of the different kinds of testing errors. L is essentially a measure of the ability of a test to remove uncertainty, but uncertainty of different kinds may be more or less important to the patient. By collapsing sensitivity and specificity into a single composite index, one loses the ingredients necessary for such an analysis.

Other measures of diagnostic efficiency such as the predictive value of a positive test (55) are also available. They are essentially variations of the indexes described here, and they also suffer from the general limitations described above.

To summarize, measurement of the basic constituents of diagnostic efficiency—sensitivity and specificity—is a critical first step in obtaining information necessary for decisions about the use of a diagnostic test, but it is generally an insufficient guide for decisionmaking unless much is known or can be assumed about the importance of each kind of testing error. In any case, the use of summary indexes of diagnostic accuracy in the absence of reporting on sensitivity and specificity is inadequate.

*Most studies are based on patients at a single hospital or institution, The sample may be representative of those presenting at the hospital, but there is always a problem in generalizing beyond the particular institutions, for patient populations vary widely among hospitals.

**The mathematical derivation of this interpretation is given by McNeil and Mellins (81],

STUDIES OF DIAGNOSTIC YIELD

Diagnostic yield is defined as the proportion of all test results found to be positive or abnormal. This measure of effectiveness is employed in evaluating the usefulness of a diagnostic X-ray procedure in a group of individuals with a specified set of signs, symptoms, or risk factors. Investigators typically compare yields of two or more alternative diagnostic strategies, one involving the X-ray procedure, the other(s) not. If the diagnostic yield of the strategy using the X-ray procedure is low relative to competing strategies, the inference is that the X-ray procedure is unjustified.

The concept of diagnostic yield is employed most frequently in evaluating the usefulness of an X-ray procedure as a screening tool. Here, the study population is defined by demographic and behavioral risk factors such as age, sex, occupation, history of smoking, or admission to a hospital. Symptoms or physical signs causing suspicion of a disease detectable by the X-ray are absent. The X-ray procedure has potential for detecting radiographic signs of occult (symptomatic) disease, presumably in early and more manageable stages than would appear with symptoms. The diagnostic strategies in contention are generally straightforward: screening v. no screening. The diagnoslic yield of the screening procedure thus represents the net difference in the number of cases detected between the screening and no screening options.

The cost of the screening program is often introduced as an element of the analysis. The case-finding cost (i.e., cost per abnormal) is estimated as the unit cost of the procedure divided by diagnostic yield. High case-finding costs are generally interpreted as evidence against the use of an X-ray procedure on the class of individuals in question.

Just as measures of diagnostic efficiency give inadequate consideration to the implications of different kinds of errors, so too does diagnostic yield. The negative finding, whether correct or incorrect, appears to have no value. Yet, negative results may point the way to other possible diagnosis and often reassure the patient, a function of considerable value in some situations (51).

Diagnostic yield is also insensitive to the potential significance of positive findings. When abnormalities detected by the X-ray are already known or could have been detected by simpler diagnostic approaches not considered, or when knowledge of an abnormality does not affect subsequent management of the case because it is either clinically insignificant or not amenable to treatment, then the use of diagnostic yield as an evaluative criterion overestimates the clinical value of a test. Consequently, many researchers use a modified definition of diagnostic yield, counting as abnormal only those cases whose test results are "important," "clinically significant, " or unknown prior to the test. This approach attempts to incorporate into the definition of "abnormal" some consideration of the health and economic implications of the finding.

When is the diagnostic yield a useful evaluative endpoint? Three conditions must hold: 1) it must be reasonable to ignore the reassurance value of a negative finding, 2) a false-positive result should not imply costly or risky followup procedures, and 3) a true-positive result should be likely to significantly influence therapy and outcome. Under these conditions, comparison of the diagnostic yield of alternative tests or testing strategies will generally result in appropriate medical decisions. Few diagnostic testing alternatives meet all of these conditions. To the extent that they do not, the results of such analyses should be interpreted cautiously.

STUDIES OF HIGH-YIELD CRITERIA

The "high-yield criteria" approach is an interesting variation on diagnostic yield. Whereas studies of diagnostic yield attempt to identify the best diagnostic strategy for individuals or patients with prespecified presenting conditions, studies of high-yield criteria have as their objective the identification of the presenting conditions (signs, symptoms, risk factors, etc.) that would justify a diagnostic strategy (namely, the use of the X-ray test). A diverse set of individuals is partitioned into two groups: those for whom the diagnostic X-ray strategy is preferred to other strategies, and those for whom the X-ray strategy is inferior to other strategies. The high-yield criteria are the factors used to partition the universe of patients.

This approach is most frequent in studies of the use of X-ray procedures in symptomatic patients, where there are many combinations of presenting signs and symptoms that are potential indications for the procedure. The analysis proceeds on the assumption that if presenting conditions can be identified which account for the vast majority of positive findings, then the diagnostic yield can be improved by limiting the procedure only to patients with these conditions. The high-yield conditions so identified become X-ray referral criteria.

The performance of a set of high-yield criteria must be assessed, but this cannot be accomplished by applying the new criteria to the sample of patients used to create them. Testing on independent patient samples taken from the same population is necessary, and performance, as measured by the number of positives who were missed and the number of X-rays saved, will always deteriorate from the original data set (86). Moreover, if the criteria are applied in patient populations that are inherently different from the sample used to create them, performance may increase or decrease markedly.

Virtually all high-yield criteria studies analyze data on a sample of patients referred for the X-ray procedure under study. The sample is biased in that patients with the same presenting conditions who are not referred for X-ray are not included in the study. The result is probably to overestimate diagnostic yield in patients with moderate signs and symptoms.

The greatest weakness of studies of high-yield criteria stems from their use of diagnostic yield as the relevant evaluative endpoint, The problems that plague diagnostic yield as a valid measure are equally germane to this discussion. How can one logically choose among signs and symptoms to maximize the diagnostic yield of a test when the implications of an abnormal finding (or of a normal finding) are unknown or their benefits doubtful? These limitations should be kept in mind in assessing whether the decision situation is appropriate to the application of this technique.

STUDIES OF DIAGNOSTIC INFORMATION

An alternative to the use of objectively measurable endpoints, such as sensitivity, specificity, or diagnostic yield, is to measure the impact of the test on the probabilities that physicians subjectively assign to possible diagnoses. If a test finding, positive or negative, has very little impact on such subjective probabilities, then it may be assumed that it has little impact on therapy and outcomes. The information value of the test, then, is measured as the degree of change in physicians' subjective probabilities brought about by the performance of the test.

The information value of the test is attractive as an evaluative measure because it makes a connection, admittedly loose, with therapy and outcomes, and because it considers both positive and negative test findings as important. But the use of physicians' subjective probabilities as the basis for measuring information content is troublesome. Since the likelihood ratio, discussed above, is a completely analogous measure of information content based on objective measures (sensitivity and specificity), it is questionable why one would wish to introduce into an evaluation of a diagnostic strategy the inaccuracy inherent in subjective probability assessments by physicians.

The design of a study to assess the information content of a test is difficult at best. Contrived clinical situations would be needed to protect against biased posttest probability assessments by physicians who had committed themselves to ordering the test. Though the design of studies of test sensitivity and specificity is also difficult for interpretive tests such as X-rays, the introduction of respondent bias is an unnecessary complication.

This approach assumes that the alternative strategy to the test is a state of zero information. The information content Of the test is measured at two points in time: once before and once after its performance. The change in probabilities between these times is attributed wholly to the test results. The assumption is that probabilities would not have changed over time in the absence of the X-ray, or that if more time were allowed to elapse, more information would not be forthcoming as the clinical course progressed. The more appropriate comparison would be between the information available immediately from the X-ray and the information that would be obtained in time as further clinical information emerges. This tradeoff between early and late information can only be compared by looking at the implications of a delay in therapy for patient outcomes and costs.

It is difficult to imagine how the information content of a test could be used to influence medical decisions. Supppose one found that over half of the procedures performed added little information. Would that imply that the procedure is being wasted? Not necessarily, for if the cost of missing a rare disease is high (that is, if lives are lost), then one might be willing to use a test to rule out the disease, depending on the dollar costs and medical risks of the test. But using a test in this way would insure that a substantial proportion of cases would have similar pretest and posttest probabilities and, hence, index values close to zero.

STUDIES OF OUTCOMES

The inadequacies of diagnostic endpoints can be overcome only by research into the implications of test findings for patient outcomes, including mortality, morbidity, and quality of life. A few studies have attempted to relate the findings from diagnostic X-rays to ultimate outcomes of the medical process. Outcome measures such as 1- or 5-year survival rates and years of life saved have been applied in a few studies. Alternative diagnostic strategies are compared for their ultimate impact on health outcomes and sometimes on the cost of achieving those outcomes.

Studies of this kind require that either a great deal of evidence be available in the literature on the outcome of alternative therapies or expensive prospective and well-controlled studies will be performed to obtain outcome data. Given these requirements, it is understandable that few studies fall into this category.

3 .
Chest X-Ray

3. Chest X-Ray

UTILIZATION, COSTS, AND CONTROVERSIES

In 1970, approximately 48 million radiographic X-ray examinations of the chest were performed, accounting for almost one-half of the total volume of X-ray examinations in the United States. About 64 percent of these chest X-ray examinations were performed in hospitals; the rest were performed in private offices or groups, or in health agencies (31). Though there are no available data listing indications for chest X-rays in various settings, the common practice of routine chest X-ray on admission to hospitals or prior to surgery is likely to contribute substantially to the use of chest X-ray, (In 1977, there were approximately 34 million admissions to short-term hospitals in the United States.) In many hospital emergency rooms and outpatient departments, chest X-rays are performed routinely if a complaint refers to the chest in any way (10,76),

Although the volume of procedures is high, chest X-rays accounted for less than one-third of the total films used in 1970, owing to a below-average number of films per examination. The mean number of films used per chest examination was 1.7, whereas the average for all body areas is 2.4 films per examination. As with all procedures using ionizing radiation, the chest X-ray subjects the patient to radiation exposure.

The biological effect of radiation is expressed in "rads," i.e., the amount of energy absorbed at particular points in the body, such as the bone marrow, thyroid, and gonads. All of these points are associated with different effects, such as leukemia, malignant tumors, possible impairment of fertility, or genetic effects (28). The median gonad dose to males and females from diagnostic chest radiographs is very small, less than 0.5 millirad per exam. Photofluorographic chest exams to females have a median gonadal dose of 1 millirad per exam, also a low dosage, The bone marrow dose produced by chest radiographs is relatively low too, at about 10 millirads per exam (28). The radiation exposure on each chest film is also relatively low. On average, the radiograph exposed the patient to 47 millirads compared to an average exposure of 310 millirads for skull films and 70 millirads for forearm films. Thus, although the volume of chest examinations is high relative to other X-ray procedures, the total dose to which the population was exposed represented a much lower proportion of the total dose exposure from all medical X-rays.

Chest X-rays are also inexpensive relative to other diagnostic X-ray procedures. In a 1975 sample of physician charges for X-ray procedures in California, the average bill for a two-view chest X-ray was $25.15, compared, for example, to $47.40 for a complete skull examination (121).

Information on the setting in which chest X-rays are performed is scanty, but some limited evidence based on California medicaid claims data is available. All medicaid claims submitted by physicians in the first quarter of 1978 were compiled for specific chest X-ray procedures. The claims indicate the location of service and specialty of the physician performing the service. Tables 3 and 4 show the distribution of chest X-ray claims submitted by physicians and radiologists in California, Although these data do not include all such procedures performed (i. e, X-ray procedures performed in hospitals which bill directly for both technical and professional charges are not included), they do show the relative importance of simple v. more complicated chest X-ray procedures. Notice that two-view chest X-rays are much more common relative to single-view examinations in the ambulatory care setting than in the inpatient setting (table 3). This may result from the frequent use of portable X-ray equipment on critically ill patients in hospitals where it is technically infeasible to obtain more than one view.

Table 3.—Medicaid Chest X-Ray Claims in California Submitted by Physicians in First Quarter of 1978, by Location of Service

	Inpatient[a]	Outpatient[a]	Office	Other	Total
71010 Chest: single view.	12,605 (38.4°/0)	3,183 (1 2.4%)	13,230 (12,1 0/~)	3,456 (22.900/.)	*32,474 (17.8%)*
71020 Chest: two views.	19,753 (60.20/.)	22,295 (86.6%)	95,037 (86.90/.)	11,639 (77.10°/0)	*148,724 (81 .7°/0)*
71021 Chest: three views.	207 (00.6%)	129 (00.5°/0)	262 (00.2%)	2 (00.01%)	*600 (00.3%)*
71030 Chest: complete (minimum of four views).	255 (00.8°/0)	123 (00.5°/0)	789 (00.7°/0)	—	151 (00.1 °/0)
Total	32,820	25,730	109,318	15,097	181,949

[a]Recorded only when hospital and physician bill separately for the service (split billing arrangement)

SOURCE Urban Institute sample of 5,000 solo practitioners, including 177 radiologists

Table 4.—Medicaid Chest X-Ray Claims in California Submitted by Radiologists in First Quarter of 1978, by Location of Service

	Inpatient[a]	Out patienta	Office	Other	Total
71010 Chest: single view.	11,836 (39.5°/0)	3,127 (12.4°/0)	1,371 (04.20/o)	34 (03.70%)	16,503 (18.30/o)
71020 Chest: two views.	17,695 (59.O%)	21,944 (86.7%)	31,005 (94.4°/0)	878 (95.00°/0)	72,594 (80.4%)
71021 Chest: three views. . . . ., .	204 (00.7%)	126 (00.5%)	168 (00.5%)	1 (00.01 °/0)	499 (00.6%)
71030 Chest: complete (minimum of four views).	249 (00.8°/0)	121 (00.50/0)	288 (00.970)	11 (01.2070)	675 (00.7°/0)
Total	29,984	25,318	32,832	924	90,271

[a]Recorded only when hospital andphysicianbill separately for the service (split billing arrangement)

SOURCE Urban Institute sample of 177 solo radiologists (326 percent of solo radiologists in Call fornia)

Virtually all medicaid chest X-ray examinations in California hospitals in 1978 were performed by radiologists. Other kinds of physicians played a more important role in performing chest X-rays in physicians' offices. About 70 percent of all medicaid chest X-rays in physicians' offices were performed by nonradiologists.

As expected, radiologists tend to perform more intensive chest X-ray examinations than do other kinds of physicians. Only 4 percent of chest X-rays performed by radiologists in their offices were for single-film examinations, while 15 percent of chest X-rays performed by other physicians were single-film procedures.

The high volume of chest X-ray procedures is due in large part to its importance as a screening test. Chest X-rays have been used to screen for cardiopulmonary disease in the general population and in selected high-risk groups. They have also been advocated for use in healthly populations as a baseline measure for evaluation of future radiographs (14,42). Chest X-ray screening has most frequently been advocated for detection of tuberculosis (TB), lung cancer, and cardiovascular disease. Its use has been encouraged by hospital and nursing home policies which have dictated routine admission or preoperative chest radiographs, by laws or regulations which have required chest X-rays of workers in high-risk occupations, by colleges and summer camps which have required prematriculation chest X-rays, and by public health agencies which have made chest X-ray examinations available to selected high-risk groups and sometimes to the public at large. These screening uses have had pervasive effects on the American public. Over 80 percent of all noninstitutionalized Americans over 17 years of age in 1973 had had at least one chest X-ray procedure. Thirty-one percent had received the examination in the previous year (29). There is no source of data, recent or old, on the proportion of chest X-ray examinations performed for screening purposes. However, the Bureau of Radiological Health reported that in 1.970, about 20 percent of all radiologic examinations of the thorax were conducted at health agencies. * It may be assumed

*About 9 percent of these chest X-ray examinations were photofluorograms, a procedure whose appropriateness relative to the radiograph has been questioned (31).

that the vast majority of these examinations were performed for screening purposes. Since a large but unknown proportion of the chest X-rays performed at other locations such as hospitals and physicians' offices are also for screening purposes, the contribution of screening to total chest X-ray volume in 1970 is likely to have been well above 20 percent.

The popularity of chest X-ray as a screening procedure has declined in the past decade, partly in response to the secular decline in the prevalence of TB, one major target of chest X-ray screening programs (102, 115). The change in attitudes about the appropriate place of chest X-rays is also a result of the myriad research studies documenting low diagnostic yield and high case-finding costs of many chest X-ray screening programs, even those targeted at high-risk groups. In the 1970's, public and quasi-public bodies issued recommendations reversing policies toward chest X-ray screening programs; these recommendations have rested on the studies of diagnostic yield in various populations (27,94), Today, the literature on chest X-ray screening for TB appears to be focused on workers in foodhandling occupations (18,108) and on high-prevalence groups, such as refugees from Southeast Asia, where a very high proportion would be expected to react positively to the tuberculin skin test due to high levels of exposure to the tuberclin bacillus **(31,67).** Nevertheless, selective screening programs using chest X-rays still exist. For example, all immigrants are required to have chest X-rays, and in many States, teachers', hairdressers', nurses', and other health professionals' X-ray screening programs continue. Currently, there is much debate over whether the X-ray should be used to screen high-risk groups for lung cancer (4,44).

Although most of the controversy over the chest X-ray examination has centered on its use on asymptomatic people, there has been some concern about the use of chest X-ray in diagnostic contexts. In particular, the frequency of radiographic followup of patients with cardiopulmonary disease has been raised as an issue in the evaluative literature. How often, for example, should a patient hospitalized for pneumonia be subjected to a chest X-ray, both during the hospital stay and after discharge? Hospital policy or medical practice may dictate too frequent followup examinations, * thus subjecting the patient to unnecessary radiation and medical costs.

There is also substantial question about appropriate radiological methods, Some of these issues are the number of views (hence, films) needed, the appropriateness of photofluorography radiography, the necessary level of expertise or credentials of the reader, and the optimal number of readers. Clinical studies have compared the additional diagnostic yield obtained from the two-view chest examination with that obtained from the single-view procedure (105). Other studies have assessed the diagnostic efficiency of alternative reader and equipment configurations (60).

Aside from these questions of appropriate radiological method, evaluation of the usefulness of the chest X-ray in symptomatic patients has been limited. To our knowledge, there have been no attempts to evaluate diagnostic efficiency or outcomes of chest X-rays in patients with particular kinds of signs or symptoms; only one study of high-yield criteria for ordering chest films has been attempted to date (10).

*It is possible that medical practices dictate too few followup examinations and thereby subject the patient to poorer expected health outcomes. Problems in the direction of too few radiographs do not appear to hold much attention in the literature.

CHEST X-RAY EVALUATIONS

This section summarizes the most prominent evaluation studies of the chest X-ray in the last 10 years, The purpose of this review is not only to document the evidence available to support decisions about the use of chest X-rays, but also to identify and assess the importance of methodological problems inherent in the literature.

The review is organized into these general categories: 1) studies of chest X-ray as a screening device, 2) studies of diagnostic uses of chest X-ray, and 3) studies of radiological method. Because the screening literature is large, it has been further divided into studies of community-based chest X-ray screening for TB, community-based screening for lung cancer, and X-ray screening for cardiopulmonary problems in selected patient populations.

Studies of Community-Based TB Screening Programs

The use of the chest X-ray as a mass screening tool for the detection of tuberculosis was common in public health departments in the 1940's and 1950's. As information accumulated on the low diagnostic yield of chest X-rays, however, the cost effectiveness of these programs was increasingly questioned (62,73,113). Table 5 summarizes the findings of studies published in the early 1970's on the usefulness of chest X-ray screening for TB. Virtually all of these studies used diagnostic yield as the principal evaluative criterion. The diagnostic yield of TB cases is calculated from a series of chest X-rays on large numbers of individuals. In those with neither symptoms nor high-risk factors such as alcohol abuse, the TB yield was found to be universally below 0.5 percent. Targeting screening programs to high-risk locales did not raise the yield.

The observed diagnostic yield of a test varies directly with the prevalence of the disease in the population studied. As disease prevalence decreases, diagnostic yield will decrease and case-finding costs will rise. After definitive antibiotic therapy for TB became available in the United States, the mass chest X-ray programs experienced rapidly decreasing yields.

The use of diagnostic yield or case-finding cost as a criterion for evaluating a screening program requires an implicit assumption about the implications of finding a case. If, for example, it is concluded that a screening program with a diagnostic yield of 1 percent is unjustified, one must assume that the benefits to the 1 percent of cases found do not outweigh the cost of the screening program. In the case of TB screening, the rarely expressed assumption is that TB discovered in its early symptomatic stages is curable and that the risks of disease communication by undetected active cases do not warrant the cost and radiological risk of the screening pro-

Table 5.—Screening for Tuberculosis

Study[a]	Year(s) of data collection	Population description	Number of cases	Sample	TB yield	Case-finding cost	Comments
Lewis (73)	1969	Ghetto population	109,000	Cleveland mobile X-ray mass screening program for TB	0.01%		This O .01% was 5% of the new cases reported that year
Swallow and Sbarbaro (113)	1965-70	General urban population (Denver)	—	All TB screening and case- finding pro. grams in Denver for 5 years are retro. spectively analyzed	a) Detected O 010, of new active disease b) Inactive TB detected in O.3% of people X-rayed c) Other abnormal conditions found at rate of 3 3° ,	$8,115 $ 372 $ 35	Only 13.5% of new active cases is found by screening
Retchman (102)	1970	General urban population (New York City)	283,000	–	0 02",		
Felngold (39)	1972-75	Primarily elderly and chronically Ill outpatients	48,000	Over 3 years, cases taken from the general medical out patient Clinic of an urban hospital	0.05%	$4091	50% of the new cases were in alcohol abusers
Horwttz and Darrow (62)	1972	Danish adults	677,800	Those appearing in Danish chest clinic for mass screening	o 02" o		Symptomatic patients examined by chest X ray in same clinic had TB yield of O 3° ,

aNumbers in parentheses refer to references in the list that appears at the end of this background paper gram. If TB were incurable once it reached the clinical stage but completely curable in the preclinical stage, a different consensus might be reached about the minimum level of diagnostic yield required to justify a screening program,

The reliance on diagnostic yield as a criterion for evaluating chest X-ray screening programs is curious, especially in light of the exampIes of good outcome-oriented evaluations of other screening and preventive technologies that are available in the literature (23,96,109). The development of a cost-benefit evacuative framework for analyzing the value of chest X-ray screening programs in populations with varying TB prevalence rates would be useful in resolving issues that are likely to continue to arise as new high-risk groups are identified.

Public health agencies and professional societies and associations have been influenced by the studies of diagnostic yield. In 1972, the Department of Health, Education, and Welfare issued a policy statement recommending against chest X-ray screening programs in the general population and in favor of limiting its use as a screening tool to adults in selected high prevalence populations. In other groups, the chest X-ray would be a followup procedure to positive reactions on tuberculin skin tests (27). This statement superseded a department policy dating back to 1958, which endorsed community-based chest X-ray screening programs, particularly when targeted to high-risk groups. Prior to the 1972 statement, the National Tuberculosis and Respiratory Disease Association, the American College of Chest Physicians, and the American College of Radiology had issued similar statements against mass chest X-ray screening programs.

Lung Cancer Screening

The question of whether periodic chest X-ray screening for lung cancer is useful in high-risk groups (generally comprising smokers over 40 years of age) has been analyzed by evaluative criteria that differ markedly from those used to analyze TB screening. Research on lung cancer screening has gone beyond calculation of diagnostic yield to consider the net effect of periodic screening on 2-, 3-, and 5-year survival rates and in some cases mortality rates from lung cancer,

This emphasis on final outcomes of the disease process results from the lack of a viable therapy for almost all but the earliest localized lung cancers. For these, resection (removal) of the diseased lung offers the only hope of cure. To be effective, then, a screening strategy must be able to detect cancers while they are still locaIized. To the extent that periodic screening does, indeed, uncover localized lung cancers, it is a lifesaving measure, At present, however, the evacuation and staging of lung cancers is imprecise, and the ultimate proof of cure comes only with time. Thus, diagnostic yield, or even the yield of apparently respectable cancers, is inadequate. Most investigators have compared 5-year lung cancer survival rates in screened and unscreened populations,

Lung cancer screening programs vary in their particulars, The testing protocols of some programs have involved a chest X-ray only, while others have included sputum. The frequency of screening has also varied. Annual, semiannual, and even more frequent examinations have been offered to participants,

Table 6 summarizes the results of the principal studies of lung cancer screening in the past 15 years. Randomized lung cancer screening studies are ongoing in three institutions— the Mayo Clinic, Memorial Sloan Kettering, and The Johns Hopkins University–as part of an evaluation of lung cancer detection methods sponsored by the National Cancer Institute (see table 6). The final results of these studies are not yet available.

The studies of lung cancer screening presented in table 6 are difficult to compare because of the different time periods covered, screening protocols undertaken, and biases inherent in their design. Most studies have not been randomized, leaving open the possibility that participant self-selection may have distorted the results. It is also difficult to interpret the 2- or 5-year survival rate when survival is counted from the time of disease detection. Screened groups may be expected to have higher survival rates simply because their disease is detected earlier, independent of the effect of surgery. Although the measured

Table 6.—Screening for Lung Cancer

Study[a]	Year(s) of data collection	Patient description	Number of cases	Patient sample selection	Yield	Screening cost per cure	Outcome/overall mortality
Boucot (15)	1951-61	Men, 45 years and older	6,136	Prospective study of asymptomatic male volunteers in biannual screening over 1O-year period.	**2%** (excluding results of Initial tests)	$83,000	80/0 of detected cases survived 5 years
Gilbertsen and Lillehei (50}	1950-69	Adults, 45 years and older	13,000	Prospective study of asymptomatic adults in routine annual screening program at University Medical Center	0 080/0	—	10% of detected cases had 5-year survival
Brett (16)	1965-68	Men, 40 years and older, in British factories	29,723	Over 3 years. Test group was given semiannual X-rays	0.3% (excluding results of Initial tests)		15% had a 5-year survival rate
				Control group radio. graphed at beginning and end of study	O.3% (excluding results of Initial tests)		**6%** had a 5-year survival rate
Gryzbowski and Coy (53)	1969	Men-over 40, smokers, and one additional risk factor	2,112	Screening in Vancouver by X-ray and sputum cytology	0. 47% (found by X-ray)	—	
Lilienfeld (74)	1958-61	Residents of Veterans Adminis. tration Domicil. iaries, who were at high risk of developing lung cancer	14,607	Residents were screened approximately every 6 months for 3 years, by chest X-ray and/or sputum cytology	0.01%	$1.000 for-a workup on a SUSPICIOUS X-ray	**12%** 3-year survival rate in respectable cases. 13°/0 of the diagnosed cases were respectable
Borrie (14)	1976			Tokyo Health Control Center One year	11 per 100,000 (0 01 %) for patients over 60 105 per 100,000 (1%)		
Stitik and Tockman (1 12)	1973-78	Men, ever 65 years old, smoke a pack of cigarettes per day or more	10,362	'Randomized clinical trial, volunteers selected from Maryland records Control group IS given annual chest X-rays (4 views). Study group IS given annual chest X-rays and sputum cytology every 4 months	1.1'/. "suspected cancer. 0.3% confirmed cancer (results of Initial screening)		
Fontana (43) (and as reported by American Cancer Society)	1971-76	Men, 45 years and older, who smoked one or more packs of cigarettes a day	11,000	Randomized clinical trial—control group is offered an annual chest X-ray. Study group IS screened with sputum cytology and chest X-rays every 4 months	—	—	
Melamed, et al (88)	1971-76	Men, over 40 cigarette smokers, at a high risk	6,612	Randomized clinical trial—control group 3,387 men given annual chest X-ray	O.03% early cancer	—	
				Study group 3,325 men given annual chest X-ray and sputum cytology every 4 months	0.03% early cancer found by X ray		
Dales, et al (22)	1964-75	Men and women aged 35 to 54 at beginning of study, residents of San Francisco Bay Area and members for > 2 years of Kaiser Foundation Health Plan	10,713	Study group of 5,156 members urged to take annual multiphasic health checkups (including chest X-ray) Control group of 5,557 members not so urged	—	Savings of over $2,100 per study group, men 45 to 54 years of age No net savings associated with younger men or women	Death rate per 1,000 for 11 year study period due to cancer of bronchus and lung, Control 47 Study 49

[a]Numbers in parentheses refer to references in the list that appear at the end of this background PaPer

survival rates can be corrected for this “leadtime bias” (125), the lung cancer mortality rate is a more appropriate outcome measure. The ongoing randomized studies are expected to measure the impact of their screening protocols on lung cancer mortality.

Recognizing these limitations, the findings to date do not support the notion that X-ray screening programs have a favorable impact on lung cancer survival. In a recent study of the issue, the American Cancer Society (ACS) reached the same conclusion (4). In addition to the results reported in table 6, the authors of that study had access to some preliminary results of the Mayo Clinic lung cancer study which showed no difference in mortality rates between screened and unscreened populations. The ACS study emphasized the long and costly search for cancer sometimes initiated by cost, morbidity, and time needed to localize a tumor when the sputum cytology is positive or suspicious and the chest X-ray is negative. In the absence of evidence suggesting beneficial effects on mortality, these problems of followup and the cost of screening led ACS to recommend against periodic screening.

The ACS action has been controversial. The principal investigators of the Mayo Clinic study have argued that the recommendation was premature because it used preliminary results (44), More recent data may indeed show improvement in mortality rates (90), Moreover, according to the investigators in the lung cancer detection study, the need for long and costly followup is a relatively rare event. *

While all parties to the debate agree that mass lung cancer screening programs are not justified by the evidence available at this time, there is sharp disagreement as to how the physician should advise the individual patient seeking a physical examination. The ACS board believes that physicians should discourage periodic lung cancer detection procedures, because their benefits have not been demonstrated, while their costs are high. Other-s believe that since screening is the only possibility for detecting lung cancer early, cost should not be considered in the individual decision (44).

To some extent, the issue can be resolved by considering who pays for the screening examination. If the patient is fully and fairly informed about the evidence on benefits and risks and is prepared to pay for the procedure out-of-pocket, then by definition the procedure is worth its costs to that patient. If, however, the screening test is covered by third-party payers such as public or private insurers, the issue of physician responsibility is, or at least should be, more complex. Since the patient is subsidized, the patient may demand a test for which he or she would not otherwise be willing to pay. To the extent that the ACS interpretation of the benefits, risks, and costs of lung cancer screening is accurate, promulgation of this information and its recommendations is a valuable service to both patients and third-party payers.

Studies of Chest X-Ray Screening of Selected Patient Populations

The use of chest X-ray as a routine screening test for cardiopulmonary abnormalities in patients with unrelated complaints has been studied many times. Table 7 presents summaries of the most important studes. Diagnostic yield and case-finding cost appear to be the most widely used criteria for evaluating these uses of the chest X-ray, but *many* studies differentiate between the total yield and the yield of clinically significant findings.

In studies of routine hospital admission and preoperative X-rays, the diagnostic yield differed widely with age of patient (105), but, even within particular age groups, the variation in diagnostic yield from study to study is high. For example, one study of admission chest X-ray yield from study to study is high. For example, one study of admission chest X-ray yield in a geriatric hospital reported 2 percent abnormalities (56), while another detected abnormalities in 16 percent of patients over 40 years of age (105). Such variation may be due to differences in patient case mix, reader definition of abnormality, radiological method, or reader competence.

*It is estimated by Robert Fontana M.D., of the Mayo Clinic, that only 15 percent of positive sputum findings are accompanied by negative chest X-rays (personal communication).

Table 7.—Chest X-Rays for Selected Patient Populations

Study[a]	Year(s) of data collection	Population description	Number of cases	Patient sample selection	Yield (as a percent of N)	Other measures of efficacy effectiveness: Case finding cost	Other measures of efficacy effectiveness: Effect on patient management	Comments
Feingold (39)	1972	Routine hospital admission chest X-ray	39,017	Hospital admissions for the year	2 cases (insignificant percentage)			
Sagel. et al (105)	October 1973 to March 1974	Hospital admissions and preoperative patients	10,597 a) 6,063 521 3,689 b) 1,996 C) 2,538	Prospective study of chest X-ray taken over a 6-month period, in three categories, a) Routine under 20 years over 40 years b) Possible chest abnormality c) Suspected chest abnor mantles	(For serious" abnormalities) a) 16% 0° o 26° O b) 34° o C) 69%		Effect not measured Four surgical cases were postponed but the effect 'on the majority of cases IS unknown	—
Hammar (56)	1954-69	Hospital admissions in a predominantly geriatric hospital population	36,475	A retrospective review of hospital admission records at a Minnesota hospital	2% abnormalities detected O 08°, new cases TB		Delayed surgery (in 0.04%	
Bartha and Nugent (8)	1974	Adults with hypertension	102	All patients entering clinic over a year's time with hypertension and having a roentgenogram	24%	$106 (estimated at $25 per examination)	No Influence on hypertension management	
Steel, et al (111)	1974	Gertatric patients	195	Seen in clinic for placement and evaluation	Abnormalities of aorta 11 20/0 Of heart and lungs 61%		–	
Pollen, et al (20)	1969	Members of a San Francisco health maintenance organization	44,663	These multiphasic examinations were performed during a 12 month period	7.4% under 40 21 "o, over 60 19 2"o	$ 6.20b $21.90 $2.40	–	b
Sane, et al (106)	November 1974 to June 1975.	Children (to 19 years), preoperative	1,500	A prospective study Of 1,500 consecutive pre operative children	7 4°, with abnormalities: 4.7% unsuspected significant abnormalities	**$316** per Significant finding (estimated at $15 per examina tion}	3 8°. postponed or changed surgical strategy	–
Brill, et al (17)	—	Children up to 18 years old in low income area	1,000	Consecutive inpatients and outpatients in pediatric clinic as part of screening program—excludes patients with positive TB tests or known chest abnormal {ties	6%: 4% were minor skeletal abnormalities	–	No medical or surgical treatment rendered	
Farnsworth, et al (38)	–	Children 15 months to 14 years, entering elective surgery	350	Routine preoperative chest roentgenograms reviewed retrospectively Chosen chronologically, they ex eluded cases of thoracic surgery, chest disorders and major trauma	880,0		Surgery never cancel led, preoperative diagnosis never changed	–
Royal College of Radiologists (103)	1977	Patients in nonacute, noncardiopulmonary surgery	10,619	8 British hospitals participated in the prospective study of chest X-ray	NA		No effect on patient management	—
Loder (75)	1977	British inpatients.	1,000	Over a 2 month period, preoperative patients	Over 30 11. 5% Under 30 1 .14%	–	Frequent changes in management of older patients	—

Table 7.—Chest X-Rays for Selected Patient Populations—Continued

						Other measures of efficacy effectiveness		
Study[a]	Year(s) of data collection	Population description	Number of cases	Patient sample selection	Yield (as a percent of N)	Case finding cost	Effect on patient management	Comments
Petterson and Janouver (98)	1976	Middle class hospital population	1 530	All preoperative patients	—	$15,000 per postponed surgery	Postponed surgery in 2 cases	—
Rees, et al. (101)	1976	Preoperative patients electing non cardiopulmonary surgery in males	667	Radiography performed on consecutive preoperative noncardiopulmonary surgery	<30 years of age number appreciable abnormality	—	—	38% had chest X ray in previous year
Bone brake et al. (1 1)	1966-75	Pregnant women given prenatal chest roentgenograms	11,725	Retrospective review of all available delivery records at Mayo Clinic	O .6%	$.2,773 (estimated at $1750 per examina - tion)	No beneficial change In patient manage ment	—
Mattox (87)	1972	Pregnant women given prenatal chest X ray	1,239	Retrospective review of patient records at Unversity Medical Center	1 3°, (significant abnormal findings)	$1,176 (approximately)	–	All findings would have been discovered without X rays except 1 benign noninfectious case of T B — case cost approximately **$20000** to find
Schneider and Dykan (108)	1977	Sample of job candidates in New York City hospital	3500	Chest X-rays were given to 14% of this sample—those over 40 with a positive TB skin test	O	—	—	
Brubaker (18)	1971	Job applicants in food manufacturing plants were given routine physicals	29	During the year 842 applicants were given physi - cat examinations 29 chest X rays resulted	14°, (4 findings out of 29 examinations)	$72		—

aNumbers in parentheses refer to references in the list that appears at the end of this background paper
bCost of X-ray examination did not include overhead

Frequently, the studies of chest X-rays in hospital medical or surgical inpatients have reported on the proportion of cases in which the chest X-ray changed patient management. With the exception of two studies, one on chest X-rays in children prior to surgery **(106)** and one on inpatients in a British hospital (7.5), the chest X-ray appears to have negligible effects on plans for surgery *or* other aspects of patient management.

A study by Sane, et al. (106) of preoperative chest X-rays in children highlights the difficulty of drawing conclusions for medical practice from studies of diagnostic yield. In the study, 4.7 percent of all chest X-ray examinations showed previously unsuspected, clinically significant abnormalities. The cost of detecting each abnormality was estimated at $316. * The authors concluded that this "low" case-finding cost justifies the use of the preoperative chest X-ray. In a comment on the study, Neuhauser laid out the kinds of information on subsequent benefits and costs that would be needed to make such a judgment with confidence (95). One would need to know how the detection of abnormalities translates not only into changes in patient management but also into ultimate patient outcomes. Even when the yield is so low as to be negligible, the diagnostic yield alone is insufficient. For example, in a recent smaller study of pediatric preoperative chest *X-rays, only* 1 *of*

● The cost of each chest X-ray examination was assumed to be $15, 350 X-ray examinations revealed a clinically consequential finding which could not have been anticipated on the basis of the clinical history and examination (38). To make sense of these results, one would need to know whether the clinically significant finding had an impact on mortality or morbidity,

In summary, the studies of chest X-ray screening in selected patient populations reveal wide interinstitution variation in diagnostic yield for similar types of patients and general use of an informative but incomplete evaluative measure—the diagnostic yield—as the basis for inferences about the appropriate place of chest X-ray in medical practice. It is no wonder, then, that this literature has had relatively little impact on routine X-ray policies of hospitals.

Studies of Patient Management and Followup

How often should a patient with a cardiopulmonary disease be X-rayed for purposes of monitoring patient progress? To our knowledge, this question has been addressed only in the case of tuberculosis and pneumonia.

The issue for tuberculosis has been framed in terms of the need for periodic radiographic and sputum examination in TB patients who have completed a program of modern chemotherapy. The diagnostic yield of TB on recalled patients was found to be negligible in three studies (2,6,33), leading the Center for Disease Control and the American Thoracic Society in 1974 to recommend against such a policy (5,25).

The annual number of hospital discharges for pneumococcal pnemonia in the United States has been estimated at approximately 382,000 (31). If all such pneumonia patients were X-rayed one more time than is necessary over the course of the disease, **382,000** unneeded X-ray examinations would be conducted. Thus, the question of optimal frequency of X-ray followup has major cost implications.

Jay and his colleagues prospectively followed 80 consecutive patients admitted to a teaching hospital for acute bacterial pneumonia (65). Typically, these patients were subjected to chest X-ray examinations every 1 to 3 days in the hospital and every 2 weeks after discharge until radiographic abnormalities disappeared. In the 72 surviving patients, the standard radiographic signs of pneumonia had disappeared in all patients by the eighth week from admission, but after 4 weeks, 36 percent of radiographs were still positive for pneumonia, At 4 weeks, almost 50 percent of the radiographs showed some form of abnormality related to pneumonia, thus necessitating continued followup. The authors concluded that the appropriate interval for radiographic followup after diagnosis is once at the time of discharge from the hospital and subsequently at 6-week intervals.

A more recent study of radiographic followup of pneumonia in children up to 15 years of age showed that after 4 weeks 80 percent of the radiographs were normal, whereas complete resolution of radiographic findings occurred in all patients after 6 weeks. These authors also concluded that in the absence of persistent signs or symptoms, the interval between radiographs should be 6 weeks (52).

Studies of changes in diagnostic yield over time in followup examinations are productive in identifying areas where significant savings can occur. But as in other contexts, diagnostic yield must be interpreted cautiously. In the instance of pneumonia, evidence from the chest X-ray is not generally used to alter therapy except when symptoms persist over a long period of time. In other conditions, the chest X-ray may be critical for therapeutic decisions, and changes in the radiograph, while occurring in few patients, may be important signals for therapeutic strategy. Thus, in any study of chest X-ray as a followup test, the implications of normal or abnormal findings for therapy and outcomes must be understood and accounted for.

Studies of the Diagnostic Uses of Chest X= Ray

Under the auspices of the American College of Radiology, Lusted conducted a study of the information content of chest roentgenograms (and other X-ray procedures) in hospital emergency rooms (76). Information content was measured by an index of the change in physicians' assessments of the probability of disease due to the chest X-ray. A large change in these subjective probability estimates would mean that the X-ray provided valuable diagnostic information. While the value of such information can only be assessed in terms of its impact on therapy and outcomes, its existence is a necessary condition for these ultimate results.

Before each chest X-ray procedure was performed, the ordering physician was asked to estimate the probability of the most significant disease and the most likely condition. After the physician had reviewed the chest X-ray, these probabilities were estimated again. Using a logarithmic index, the authors concluded that in about 90 percent of all chest examinations, these probabilities changed sufficiently to infer high information content. *

The study contained some serious design flaws. The selection of participating institutions and physicians was not random and may well have been seriously biased. When the ordering physician was asked to revise his initial probability estimates after the X-ray, he was reminded of his early assessment, thus probably strengthening an already existing bias toward large changes in probability assessments. The index of probability change used in the study is also difficult to interpret in terms of real gains in information.

Even with these problems corrected, the findings of the study would offer little in the way of information on the appropriate uses of chest X-rays in emergency rooms, because the results are aggregated across all presenting signs and symptoms. Whether chest X-rays overall provide much or little information is of no consequence, for the information content most likely varies widely among different kinds of patients seen in emergency rooms. The objective of evaluation is to discover which indications are justified and which are not.

This objective has been pursued in a recent study of symptomatic patients receiving chest X-rays in an emergency room (10). Over 1,100 consecutive patients presenting at an emergency ward with complaints related to the chest were followed prospectively. Their clinical signs, symptoms, and some risk factors were recorded, and the relationship between these findings and the X-ray results was studied. In patients over 40 years of age, 37 percent of X-rays taken in those with normal physical examinations showed acute radiographic findings such as enlarged heart, pneumonia, plural effusion, and congestive heart failure. In younger patients without positive clinical signs, only 4.8 percent of all X-rays had acute findings (most frequently pneumonia). If X-rays had been withheld from younger patients (under 40) with normal physical examinations, 16 percent of the X-rays would have been avoided, while approximately 1.5 percent of acute radiographic X-ray findings would have been missed. These cases (typical pneumonia) were clinically important, however, and the authors recommended that X-rays be provided if symptoms persist in the younger group.

*The computed index of information content was close to zero (absolute value less than 0.25) in 9.5 percent of all cases.

The high yield of abnormal chest X-ray findings in symptomatic patients may explain why the use of chest X-rays in this patient group is not controversial. The seriousness of some findings and the importance of therapy to the outcome of diseases like pneumonia justify the presumption that the benefits in lives saved or reduced morbidity outweigh the cost of X-rays, particularly in older patients with chest complaints who present at emergency rooms.

Studies of Radiological Method

Potential variations in radiological method can affect diagnostic efficiency, diagnostic yield, costs, and outcomes to differing degrees for each diagnostic or screening application of the chest X-ray. The type of equipment and film used, the number of views taken, technician competence, number and competence of readers, and access of readers to patient information are all potentially important aspects of radiological method.

Because the way in which these elements are combined is important in assessing the validity of an X-ray evaluation study, such information should always be given in the study results. Unfortunately, this is not always easy. Retrospective studies often are unable to determine these elements from the medical record. Nevertheless, inherent in any evaluation study is the question of whether the results would have differed significantly had the method varied.

The question of radiological method has been addressed directly in a few studies. Sagel and his colleagues (105) investigated the usefulness of including the lateral view in addition to the frontal view in routine hospital admission chest X-ray examinations (105). In a retrospective study of over 10,000 chest X-ray examinations, the additional information provided by the lateral view was assessed. For patients between 20 and 39 years of age without reasonable possibility of chest disease, a potentially serious abnormality was seen only on the lateral view and not on the posteroanterior film in only one case (0.05 percent). When chest disease was considered a possibility, the additional diagnostic yield uniquely contributed by the lateral film was 2 percent. In older patients (40 and above) with no suspicion of chest disease, the additional yield of the lateral field was 0.9 percent. The authors concluded that chest X-ray examinations with lateral views should be limited to older patients and those with a reasonable probability of chest disease.

One cannot assess the validity of this recommendation without knowing more about the implications for outcomes of the abnormalities uncovered, It must also be assessed in light of the question of the need for a chest X-ray at all in routine hospital admissions.

The optimal number of readers of chest X-ray films has been addressed by a number of authors, particularly those interested in the screening uses of chest X-ray (60). Radiological findings are interpretive; as such, they are subject to observer error both in false positives and false negatives. A false positive will entail additional followup and possibly even inappropriate therapy, whereas a false negative can deny or delay the initiation of needed therapy, Retrospective studies of TB X-ray screening programs showed that between 20 and 30 percent of positive cases were missed by single reading (60). Adding an independent second reader, however, increases the probability of false positives.

Inherent in any dual-reader configuration is the need to resolve disagreements, Arbitration, independent third reader with majority vote, “believe positive” or “believe negative” are examples of such rules. Hessel, Herman, and Swenson (60) have analyzed the effect of alternative methods for resolving reader disagreements on the ultimate accuracy of chest radiographs with two readers. Using the results of independent readings by eight radiologists on 100 randomly selected chest radiographs, accuracy increased from 43.3 percent with an individual reader to 50 to 54,7 percent under dual-reader systems with various schemes for resolving disagreements. Specific methods for resolving disagreements had differential effects on the false-positive and false-negative rates. Consequently, the best method of resolution would depend on the implication of each of these two kinds of errors on patient outcomes and costs.

4
Skull X-Ray

4. Skull X-Ray

UTILIZATION, COSTS, AND CONTROVERSIES

In 1970, approximately 4.2 million radiographic skull examinations were performed, representing 3 percent of the total number of X-ray examinations in the United States in that year. Over 86 percent of these procedures were performed in hospitals (27). The average number of films used per examination was four, and in 1970, the median gonadal dose per exam for both sexes was less than 0.5 millirad. The bone marrow dose was in the medium range, at 78 millirads per exam.

Skull X-rays are ordered to detect and evaluate abnormalities of the head. They provide strong evidence of skull fracture and in some cases provide clues about abnormal intracranial conditions. They have held a traditional place in the evaluation of patients with head injury or sudden onset of unconsciousness or coma.

Physicians' claims for skull X-rays performed on California medicaid patients in 1978 are shown in table 8. The limited skull examination plays a minor role in radiologic practice in California. Over 85 percent of skull examinations were comprehensive. Comparing this table with claims by physicians shown in table 9 reveals that radiologists perform most hospital skull X-ray procedures but only 38 percent of such examinations performed in physicians' c) ffices.

In the past decade, the skull X-ray has been criticized as an overused and not very valuable radiological procedure (1,66,85,107). This criticism has stemmed from studies demonstrating the low yield of positive skull X-rays (9) and low sensitivity of the skull X-ray in detecting clinically important intracranial abnormalities such as subdural hema torna (114) and cortical a trophy (119); from the general unimportance of detecting a fracture in the absence of any clinical evidence of intracranial damage (58,64); and from the arrival of a noninvasive diagnostic imaging technology that offers vastly superior in-

Table 8.— Medicaid Skull X-Ray Claims in California Submitted by Physicians in First Quarter of 1978, by Location of Service

	Inpatient[a]	Out patienta	Office	Other	Total
70250 Skull: limited.	353–(2 1.7 %)	403 (10.8°f0)	893 (14.6%)	55 (07.7%)	1,704 (14.0%)
70260 Skull: complete	1.274 (78.3°/0)	3,319 (19.2%)	5,219 (85.4%)	657 (92.3%)	10,469 (86.0°0)
Total.	1,627	3,722	6,112	71 2	12,173

[a]Recorded only when hospital and physician bill separately for the service (split billing arrangement)

SOURCE Urban Institute sample of 5000 solo practitioners, including 177 radiologists

Table 9.—Medicaid Skull X-Ray Claims in California Submitted by Radiologists in First Quarter of 1978, by Location of Service

	Inpatient[a]	Out patienta	Office	Other	Total
70250 Skull. limited	323 (21.3%)	400 (1 2,40/')	195 (08.4%)	409 (84.3%)	927 (12 30.)
70260 Skull: complete	1,192 (78.7%)	2,836 (87.6%)	2,125 (91.6°6)	76 (15. 7%)	6,629 (87.7%)
Total	1,515	3,236	2,320	485	7,556

[a]Recorded only when hospital and physician bill separately for the service (split billing arrangement)

SOURCE Urban Institute sample of 177 solo radiologists (32 6 percent of all solo radiologists in California) formation on intracranial disease and injury, namely, the computed tomographic (CT) scanner (68,126).

Although there are no national data showing the extent to which the use of the skull X-ray has been affected by these factors, anecdotal evidence from a few institutions with CT scanners indicates that many patients, particularly head trauma victims who would have been given skull radiographic examinations, are examined by CT instead. In one hospital, it was reported that 24 percent of patients suffering from acute head injury were spared a skull roentgenogram after the introduction of CT (126). Even without CT scanning, policies regarding skull X-rays in two emergency rooms reduced the rate of use of the procedure by 29 percent (99,100).

SKULL X-RAY EVALUATION

The evaluation literature has had an important place in framing the debate about the appropriateness of the skull X-ray. Two kinds of studies have provided evidence about the usefulness of this procedure: 1) studies of diagnostic efficiency, and 2) studies of high-yield criteria for ordering skull examinations in emergency rooms.

Diagnostic Efficiency of Skull X-Ray Examinations

The diagnostic efficiency of a radiologic test can only be assessed in terms of the diagnoses of interest. If one is concerned with the diagnostic efficiency of the skull X-ray in detecting skull fractures, then the skull X-ray is generally the definitive test. Except for examination during an operation or postmortem, there is generally no other method for determining whether a skull fracture exists. Thus, for skull fractures, the observed true-positive rate of the skull X-ray approaches 100 percent. Specificity (true-negative rate) is also presumably high.

The diagnostic efficiency of the skull examination becomes more questionable when one considers other, more clinically significant, diagnoses. Except for depressed fractures, which sometimes require surgery, the mere presence of a fracture is unlikely to influence therapy in head injury patients. The more important question is how well a test can detect intracranial abnormalities. Here, the skull X-ray does not perform well.

In a 1971 study of 100 patients with acute subdural hematoma, Talalla and Morin reported the findings of skull X-rays taken in 50 of these cases (114). * While 60 percent of the skull X-rays were abnormal, in only 5 percent of the examinations was a subdural hematoma identified. The rest showed fractures or other abnormalities unspecific to the diagnosis. A similar study of subdural hematomas of traumatic origin in England showed that for those cases where a skull X-ray was performed, the positive rate was 48 percent for simple hematomas and 80 percent for complicated hematomas (64).

These pre-CT studies cast doubt on the ability of the skull X-ray to differentiate between patients with and without serious injury. A later study comparing the skull X-ray examination with CT scanning in patients with head trauma draws an even sharper picture of the limitations of the skull examination.

The results of skull examinations taken in 76 percent of 285 consecutive acute head trauma patients who received CT examinations were recorded by Zimmerman and his colleagues (126). Of those patients with CT evidence of significant intracranial abnormality, skull films were normal in 31 percent of children and 33 percent of the adults. Of those patients with negative CT

* Skull X-rays were not taken in urgent situations, or when the patient did not survive long enough, Thus, the efficiency of skull X-ray is probably underestimated for the population as a whole. However, since the procedure is only feasible for the subset of patients who did receive it, this group should be considered representative of the relevant population,

scans, 23 percent of the adults and 32 percent of the children showed fractures on the skull film. * If intracranial damage is considered, it appears that approximately one-third of abnormal skull examinations are unrelated to the diagnosis of interest. In detecting significant intracranial damage, CT scanning is the definitive procedure in the same sense that the skull X-ray is for fracture.

Even lower estimates of skull X-ray sensitivity were obtained in studies of patients with nonacute problems suggestive of intracranial disease. A study of patients admitted to a psychiatric hospital with symptoms suggestive of organic brain disease showed that the skull X-ray was abnormal in only **6** percent of cases found to be abnormal by CT scanning (119). A comparative study of skull radiographs, CT scans, and radionuclide bone scans in detecting cancerous metastasis to the skull (calvaria) showed that the skull X-ray detected only 55 percent of the calvarial metastasis identified by one or more of the three procedures (68). The skull X-ray was uniqueIy responsible for the detection of 1 of the 32 total calvarial lesions.

Were it not for the difference in the cost of the skull X-ray and CT head scans, the evidence comparing the diagnostic efficiency of skull X-ray and CT scanning in identifying intracranial abnormalities would argue for virtual replacement of skull films by CT scans in the conditions studied. However, the technical cost of a skull examination was estimated at about 1] percent of that of a CT examination in **1977 (48). Consequently, the appropriate** place of these two examinations in the applications discussed is not so straightforward. At the very least, it may be argued that head injury patients who received CT examinations should not **also** receive skull X-rays as a routine procedure.

High-Yield Criteria for Skull X-Rays

Historically, skull X-ray examinations have been used as a standard radiological procedure in the evaluation of patients with head injury. Over time, the procedure has come to be viewed as necessary to the provision of quality care, and it has been claimed that many such examinations are ordered by emergency room physicians for medico-legal reasons or because the patient or patient's family request it, and not because the physician sees the examination as contributing greatly to his or her information **(1,66).** The yield of abnormal findings in skull X-rays ordered in this way is low. Studies of diagnostic yield in hospital emergency rooms in this country and in Britain have reported rates ranging from under 2 percent to about 8.5 percent (9, 32,36). It is not surprising, then, that the search for criteria for ordering skull X-ray procedures to improve diagnostic yield would have begun with the emergency skull X-ray.

The first attempt to determine which presenting signs, symptoms, or risk factors would be good predictors of an abnormal skull X-ray finding was reported by Bell and Loop, who identified 21 objective and subjective findings present in at least 10 percent of positive skull X-ray examinations obtained in 1,500 consecutive skull examinations in two hospital emergency rooms (9). These 21 attributes, termed "high-yield findings, " included such elements as presence of unconsciousness for more than 5 minutes (present in 41 percent of all skull fractures found by radiograph); discharge from the ear (present in 30 percent); accident at work or gunshot wound (present in 15 percent); and serious suspicion of **a** fracture (present in 76 percent).

Only one patient with a positive skull fracture did not have at least one of the high-yield findings, and the radiograph made no difference to the management of the patient's care. Had skull X-ray examinations been limited only to those patients with the high-yield findings, however, approximately 29 percent of the 1,500 **patients would not have had the examination.**

The performance of the 21 high-yield findings was recently studied in a sample of 594 cases of head trauma in a military hospital's *emergency* room (32). Of the 17 skull fractures detected, 7 had none of the high-yield findings. All 7 were children under 17 years of age. The patient sam-

*Because the patient populations included only those for whom a CT scan was ordered, patients who are bound to be sicker than most, results probably overestimated the sensitivity of the skull X-ray and underestimated its specificity, although the magnitude of the effect is not known.

ple in this study was 55 percent children, compared to 9.7 percent in the Bell and Loop study (9), indicating that in children there may be a different relationship between symptoms and X-ray results. With one exception, none of the high-yield findings was significantly correlated with positive skull radiographs (p **0.10),** but the authors did not report on the number of radiographs that would have been saved had the criteria been applied to the patients in the study.

A recent British study of 504 head injury patients found that six of the nine skull fractures contained one or more of just seven findings: headache/concussion, vomiting, loss of conscious, focal or general signs of central nervous system involvement, scalp hematoma, scalp laceration, and ear bleeding (36). These findings were selected arbitrarily by the authors, and their performance in terms of saved radiographs was not assessed.

The fundamental weakness of studies of high-yield criteria for skull X-rays stems from their emphasis on diagnostic yield in the face of overwhelming evidence that X-ray findings in patients with head injury mean very little to patient management. Detection of a fracture is useful only to the extent that it indicates potential cerebral damage. Those with serious intracranial injury often have no fracture, while those with fracture and no other clinical findings frequently require no management save observation (36, 58). In some areas, hospital admission for observation is automatically prompted by a positive skull X-ray **(36), thus** raising the costs issuing from the use of this procedure on all head injury patients.

One must question the use of a method such as high-yield criteria to reduce the number of examinations of a procedure that offers little information relevant to therapy when the more appropriate strategy may be to eliminate skull X-rays altogether in favor of CT scanning, which directly detects the important conditions. Because CT scans are expensive, however, identification of high-yield criteria for ordering CT scans on head injury patients is of utmost urgency. One such study has been performed to date **(57).**

5.
Barium Enema

5. Barium Enema

UTILIZATION, COSTS, AND CONTROVERSIES

Barium enema (BE) is a generic term referring to radiological studies of the colon and rectum using contrast materials. The conventional procedure, sometimes called a single-contrast enema (SCE), consists of fluoroscope-guided radiography of the colon while it is filled with a barium contrast solution. The double-contrast enema (DCE) is performed while a thin layer of contrast solution coats the colon lining and air is insufflated into the colon. The purpose of the BE examination is to detect diseases of the colon, principally colitis, regional enteritis, diverticulitis, polyps, and carcinoma.

In 1970, an estimated 3.5 million BE examinations were performed in the United States. About 80 percent were performed in hospital settings (31), The examination involves a high radiation dose because of the multiple views and the need for fluoroscope to localize the spot films, It is estimated that in 1970, there were an average 3.9 films per examination. As might be expected, because of the direct exposure of the gonads to primary beam irradiation, the BE has a higher gonad dose for both sexes than do the chest and skull exams. For males, this dose is 22 millirads per exam (a medium dose); for females it is 574 millirads. For females, only the lumbar spine exam results in a higher gonadal exposure (28).

Of all physicians, radiologists predominate in the administration of BE regardless of the setting in which they are performed. Tables 10 and 11 present data on the California medicaid claims for BE examinations. In California in 1978, 97.8 percent of all physicians' claims for BE to the medicaid program were filed by radiologists. Virtually 100 percent of all physicians' claims for BE on inpatients and outpatients were submitted

Table IO.— Medicaid Barium Enema Claims in California Submitted by Physicians in First Quarter of 1978, by Location of Service

		Inpatient[a]	Outpatient[a]	Office	Other	Total
74270	Colon: barium enema	688 (89.9%)	1,227 (95.6%)	4,050 (84.0%)	61 (100%)	6,026 (82.0%)
74275	Colon: combined with air contrast	62 (08.1%)	25 (01.9%)	694 (14.4%)	—	781 (11.3%)
74280	Colon: air contrast	15 (02.0%)	31 (02.4%)	75 (01.6%)	—	121 (01.7%)
	Total	765	1,283	4,819	61	6,928

[a]Recorded only when hospital and physician bill separately for the service (split billing arrangement).

SOURCE Urban Institute 1980 Sample of 5000 solo practitioners including 177 radiologists

Table 11 .—Medicaid Barium Enema Claims in California Submitted by Radiologists in First Quarter of 1978, by Location of Service

		Inpatient[a]	Outpatient[a]	Office	Other	Total
74270	Colon: barium enema	636 (89.2%)	1,206 (85.2%)	3,141 (83.9%)	57 (100%)	5,040 (84.50/.)
74275	Colon: combined with air contrast	62 (08.7%)	213 (15.1%)	545 (14.6%)	—	820 (13.8%)
74280	Colon: air contrast	15 (02.1%)	31 (02.2%)	56 (01.5%)	—	102 (01.7%)
	Total	713	1,415	3,742	57	5,962

[a]Recorded only when hospital and physician bill separately for the service (split billing arrangement).

SOURCE: Urban Institute, 1980 Sample of 177 radiologists (32.6 percent of solo radiologists in California).

by radiologists; almost 97 percent of claims for BE examinations in physicians' offices were made by radiologists. *

The conventional SCE is much more frequent than the more involved double contrast or combined studies. MacEwan reported that in Canada's Manitoba Province in 1974, 95 percent of all BEs were SCES (77). California medicaid claims data from a sample of approximately 5,000 solo practitioners substantiate the dominance of the conventional examination over DCE but reveal a slight decline in its proportionate use from 1973 to 1978.** In this sample of physicians, SCE exams accounted for **86.2** percent of BE procedures performed in physicians' offices in 1973; by 1978, this proportion had diminished slightly to 84.1 percent. The relative proportion of conventional BE to air-contrast procedures decreased more appreciably in other settings. From 1973 through 1978, the use of SCE examinations declined from 96 to 89 percent in inpatient settings, and from 94.5 to 83 percent in hospital outpatient settings. An earlier survey of leading institutions and practitioners found a similar trend. The number of radiologists using only the DCE increased between 1966 and 1968 (84).

BE is a complicated procedure for both the patient and physician. Adequate preparation of the patient to insure complete evacuation of fecal matter must precede the application of the contrast enema. This preparation usually involves dietary restriction for 1 to 2 days, administration of cathartics, and cleansing enemas (82). Inadequate or hasty patient preparation seriously compromises the ability to diagnose accurately (82,91).

The radiologist is faced with numerous other choices in technique, ranging from patient positioning to methods for compressing the colon. One textbook states that "probably no two radiologists perform the barium enema in exactly the same way, [but] opinion and practice on most of the important steps of the examination are amazingly uniform" (82).

*Urban Institute, unpublished data, 1980.
* ● Urban Institute, unpublished data, 1980

The importance of radiological method, particularly patient preparation, to the success of the BE implies high variation in diagnostic efficiency among physicians and facilities. This variation presents an important evaluative dilemma. If the procedure does not perform well in a study, critics are quick to argue that radiological technique was inadequate (3). If BE is not evaluated at its best, physicians will be loathe to accept study findings, in the belief that the performance in their own institutions will surpass that of the study. However, evaluating performance of the BE only in the best centers, as advocated by some (3,72), may overestimate its general value as a diagnostic tool.***

BE is well established as the mainstay of diagnostic methods in the colon (59). The development of flexible fiberoptic endoscopy of the colon (colonoscopy) in the past decade has increased the ability to detect certain diseases of the colon, but the high cost of colonoscopy relative to BE and the general reliance of colonoscopists on prir BE films to guide their own examinations imply that the BE will not be replaced by endoscopy to any significant degree in the foreseeable future (92). Indeed, the potential for conflicting findings on the two examinations may lead to increasing numbers of followup X-ray procedures.

In recent years, investigators have used the findings of colonoscopy to assess the sensitivity of BE in detecting polyps and carcinomas of the colon. Prior to the availability of colonoscopy, the sensitivity of BE could only be surmised, since there was no independent method for demonstrating false negatives from BE. Now, at least some tumors and polyps not found on BE have been found by colonoscopy. The results of these studies of the diagnostic efficiency of BE and colonoscopy are discussed later in this section.

At present, radiologists disagree about whether and where the DCE should be used instead of, or in addition to, the single-contrast procedure (69,83). The choice between the SCE and the DCE is important for medical cost and radiological risk. The DCE examination is gen-

* • *This general issue of "efficacy" v, "effectiveness" has been discussed elsewhere (97).

erally more time consuming than the SCE and requires careful patient preparation. If both procedures are performed, costs and radiation exposure to the patient are almost doubled, The DCE procedure appears to be more sensitive than the SCE in detecting polyps and carcinomas when patient preparation has been meticulous (69), but reliability is claimed to be low in less ideal situations (82). Some radiologists believe that the DCE procedure should be used routinely instead of the SCE except in certain contraindicated situations; others believe that it should be a standard second procedure; still others believe that it is indicated when the conventional examination gives unsatisfactory or equivocal results (82). Often, a combined procedure of SCE followed by partial evacuation of the contrast solution and performance of an air study is done, but the barium contrast solution used for the single-contrast procedure is not ideal for a double-contrast procedure (82). Some recent studies of the diagnostic efficiency of the SCE and DCE have shed light on the issue. They are reported later in this section.

Because BE remains the only practical first test for comprehensive investigation of the colon in patients with signs or symptoms suggestive of colonic disease, it is particularly suitable for the high-yield criteria evaluation technique. One such study is described later in this section.

The Diagnostic Efficiency of BE

The ability of various diagnostic procedures to detect the existence of color-tic polyps is of great interest because of the presumption that a large proportion of colon cancers originate in polyps (33). About 100,000 new cases of colon cancer are diagnosed each year; if they are detected while localized, 5-year survival rates are high. Otherwise, patients have about a 17-percent 5-year survival rate (47). The detection and subsequent monitoring or removal of polyps is viewed by some as the best available method of cancer control.

Table 12 presents a summary of studies of the sensitivity of BE in detecting colonic polyps. The recent studies of the number of colonic polyps missed on BE but found on subsequent colonoscopy have two important findings. First, the double-contrast examination appears to be much more sensitive than the standard examination in detecting polyps, especially small lesions. Second, as polyp size increases, the sensitivity of the double-contrast examination approaches that of colonoscopy. Though the studies from which these conclusions were drawn suffer from serious design flaws, these findings are too pervasive and strong to be discounted.

What do these finding imply for the use of BE in the detection of polyps? Unfortunately, the information is incomplete. We do not know how the DCE and SCE compare on specificity. If the DCE examination were to yield a very high proportion of false-positive results, thereby necessitating more expensive followup, its apparent superiority to the SCE would need to be explored further. Nevertheless, at this time, in patient groups for which the detection of polyps is the goal, the double-contrast examination appears to offer greater diagnostic efficiency.

These studies can also be interpreted as measuring the additional contribution of colonoscopy to the detection of polyps. Does a colonoscopy following a BE make a significant contribution to polyp detection? BE sensitivity results presented in table 12 indicate that colonoscopy offers large proportional improvements in sensitivity only for small polyps. Current evidence suggests that polyps less than 1 cm in size have **a 1-** to 2-percent incidence of cancer (3). Thus, only 1 or 2 in 1,000 of the small polyps missed by BE could be expected to be cancerous. This additional yield of extra cancers would have to be considered against the high cost of colon-Oscopy.

Several important limitations Of the design are common to all of the studies summarized in the table. * First, each was subject to serious patient selection biases. Patients with negative BEs whose physicians saw no reason to refer to colonoscopy were excluded from every study. In one study, the only patients selected were those

'The studies also have been criticized on the ground that the colonoscopists had access to BE findings before performing their own examinations (3, 72). Because the studies were not Intended to compare colonoscopy and BE as direct substitutes for another this criticism is of no concern here.

Table 12.—The Sensitivity of Barium Enema in the Detection of Polyps

Study[a]	Year(s) of data collection	Patient sample	Study design	Definition of sensitivity	Reported sensitvity
Thoeni and Menuck (1 16)	1974-76	210 patients undergoing colonoscopy subsequent to barium enema (BE)	Colonoscopists had results of BE at hand BE technique not standardized Both DCE and SCE performed	Proportion of confirmed polyps detected by test Polyps were confirmed by one or both examinations If seen on BE and not found on colonoscopy, repeat BE confirmed	DCE 8830,0 SCE 54 8°1,
Laufer, et al (71)		46 patients undergoing colonoscopy as followup to positive DCE	Colonoscopists had results of DCE at hand DCES were performed at single Institution with standard technique.	Proportion of confirmed polyps detected by X-ray Confirmed polyps defined as Identified on both DCE and colonoscopy, or on subsequent endoscopy or subsequent DCE examination	DCE 94%
Liencke, et al (72)	1971-74	64 patients undergoing colonoscopy subsequent to BE	Colonoscopists had results of BE at hand BE technique not standardized, not performed at single Institution Results were not reported accord. ing to type of BE examination performed Patients who had technically unsuccessful colonoscopies were excluded from study	Proportion of confirmed polyps detected by BE Colonoscope commonly used to arbitrate presence or absence of polyps, except when multiple X-rays, colonoscopic examinations or surgical examinations available	< 5mm In size BE 4400 5mm.1cm In size BE 440/0 1 2cm In size BE 77°4 2 or more cm In size BE 93%
Williams, et al (122)	1974	182 patients referred for colonoscopy because of positive BE or unexplain. ed bleeding with a normal BE	Colonoscoplsts had results of BE at hand BE technique not standarized • about 50% performed at same Institution with DCE . the rest were referred from other hospitals, most were SCE	Proportion of confirmed polyps detected by BE A polyp was confirmed if found on colonoscopy If reported on BE and not on colonoscopy, was assumed to be false positive due to poor preparation	< 5mm In size DCE. 730/0 SCE 13% 5mm- 1cm In size DCE 87% SCE 40% 1cm 2cm in size DCE 99% SCE. 75% > 2cm in size DCE 950/. SCE 79%
Wolff, et al (124)	·	500 patients referred for endoscopy after one or more BEs	Colonoscopists had results of BE at hand. Referrals for con firmation of abnormality classification	Proportion of polyps confirmed by colonoscopy found on BE	BE 80%

aNumbers in Parentheses refer to references in the list that appears at the end of this background Pa Per

with positive DCE examinations, The net effect is that the sensitivity of the X-ray examination is overestimated, because some lesions were undoubtedly missed in the group that was not referred. The extent of the overestimation is unknown, but the problem is more troublesome for small polyps.

In all but one study, the BE technique was not standardized. The X-rays were performed in different hospitals or facilities from the place of the colonoscopy. Patient preparation was not controlled or even monitored. While this lack of control makes it difficult to judge the quality of radiologic method, it may also mean that the findings are more representative of the diagnostic efficiency typical of a community center rather than of performance at an outstanding center.

These limitations do not negate the central implications of this literature: If a patient is suspected or at risk of colonic polyps or cancer, and if X-ray studies are planned, the best approach is a DCE followed by colonoscopy only if the X-ray examination is positive or suggests the need for further workup. * Though formal analyses of the costs of the single-and double-contrast procedures and of colonoscopy were not conducted, this result is based on implicit assess-

*This conclusion must be tempered by recognition of the differences in competence in the conduct of DCE.

ment of the cost of the three procedures in relation to their diagnostic efficiency. Notice, however, that the findings do not reveal the conditions under which the benefit from either kind of X-ray study of the colon is worth its costs. What patient signs, symptoms, or risk factors are likely to correlate highly with the existence of polyps, particularly large polyps or carcinomas of the colon? What are the ultimate benefits and costs of offering the double-contrast examination to these or other kinds of patients? These questions have not been addressed in the literature.

A recent study of the diagnostic efficiency of DCE in rectal cancer illustrates the limitations of diagnostic efficiency as the evaluative endpoint. The rectum has long been considered the province of the rigid sigmoidoscope, a relatively inexpensive and wide] y available diagnostic instrument. (In 1980, California's medicaid program paid $13.17 for a proctoscopy compared to $39 for a DCE.) SCE has never been particularly suitable for visualization of the rectum. With increasing acceptance of the double-contrast technique, however, the ability of X-rays to detect lesions in the rectum has increased markedly. Evers and her colleagues reviewed the records of go cases of cancer of the rectum in two institutions which routinely use the double-contrast method (3.5). In 66 of those cases, both proctoscopy and DCE were performed, although which came first and whether the results of the earlier precedure were available to those who performed the later procedure were not known. The observed true-positive rate of the DCE was 91 percent, while that of proctoscopy was 86 percent. Should patients presenting with symptoms suggestive of rectal carcinoma (constipation and bleeding) be examined by DCE instead of, or in addition to, the proctoscope? The results of the study are insufficient to answer this question, although the authors concluded that "the rectum is the province of the radiologist" (35). The specificity of the two examinations, the number of true-positive cases that would be found and the stage at which they are detected, the ultimate effect *on* survival, and the cost of the examinations and subsequent medical care would have to be known to fully understand the implications of an affirmative or negative response to the question.

High-Yield Criteria for BE

Gerson and colleagues have recently reported on an attempt to apply a high-yield criteria method to SCE (49). The study was intended to identify symptoms, signs, and laboratory findings which could be used as criteria for referral to SCE. Physicians referring patients to a university-affiliated radiology department for SCES over a 22-month period in 1974 and 1975 were required to fill out a requisition form containing information on *over* 40 specific patient attributes—signs, symptoms, and laboratory findings. The presence or absence of these findings was correlated with the outcome of the BE examination, An attribute was included in the high-yield criteria set if its likelihood ration, L,, was significantly greater than 1. L_i was defined as follows:

$$L_i = \frac{\text{proportion of abnormal BEs with attribute i}}{\text{proportion of normal BEs with attribute i}}$$

When all attributes with values of L significantly greater than 1 were identified, the "high-yield" set was considered complete. Only five findings were so identified: rectal mass, fever, abdominal mass, low hematocrit, and positive stool benzidine (a test for blood in the stool). Had these criteria been used in the sample of patients studied, 36 percent of the BEs would have been avoided, but 22 percent of diseased patients and 10 percent of those with cancer would have been missed. Thus, the high-yield criteria so selected did not perform well, especially since an even greater deterioration in performance could be expected in other patient samples.

Although these results are disappointirig, they do not provide definitive evidence against the use of high-yield criteria; performance may have been improved by using other methods or statistical approaches. If performance were to remain low after reanalysis, however, this would suggest that the symptoms suggestive of colonic disease are also suggestive of other problems, and that colonic disease occurs in too broad an array of symptom complexes to provide criteria that will perform well.

This study also highlights a potential danger inherent in implementing a high-yield criteria referral process for a test with such a wide referral base (unlike skull X-rays, where the potential population is limited to those with head injury). Suppose, for example, that the five findings had performed well and that they were accepted as referral criteria for barium enemas. The presence of fever would be one such criterion, Does this imply that all persons or even all persons over 40 presenting with fever should be referred? Surely not, but the mere existence of the referral list would be likely to provoke referral of more patients with marginal clinical indications, or isolated findings, with a consequent further deterioration of diagnostic yield. MacEwan has demonstrated that the physician's strong suspicions (probability > 0.50) of carcinoma of the colon is related to a high proportion of positive cases (7'7). The Gerson study did not include such subjective criteria in its initial attribute list, but they are probably surrogates for the existence of patterns of multiple-objective findings that are likely to occur together in the presence of disease. It would be worth exploring how combinations of symptoms contribute to the separation of normal from abnormal X-rays.

6 .
Excretory Urogram

6. Excretory Urogram

UTILIZATION, COSTS, AND CONTROVERSIES

Excretory urogram (ExU) refers to any procedure that provides for X-ray delineation of the urinary tract using an opaque contrast medium injected (or occasionally infused) into the blood (123). * As the contrast medium is excreted by the kidneys through the urinary tract, usually within 15 to 20 minutes of injection, the anatomy of the urinary tract organs can be delineated by X-ray films. Special types of ExU examinations, such as cystography (bladder studies) or urethorography, focus on particular sections of the urinary tract. The examination "must be tailored to meet the needs of the individual patient or clinical problem" (123).

The primary purpose of ExU is to detect obstructions (calcifications, tumors, etc.) and other diseases in the urinary tract. ExU is also a diagnostic tool for detection of renovascular disease, an underlying cause of hypertension in about 10 percent of hypertensive cases (13).

In 1970, approximately 3.9 million ExU examinations were performed, representing about 3 percent of the radiographic examinations performed in the United States that year. Over 85 percent of these procedures took place in hospitals, and the average number of films used per examination was 5.3. The ExU procedure subjects the patient to a relatively high gonad dose, and this dose is higher for women (448 millirads per exam) than for men (20 millirads per exam) (28).

Not only does ExU involve high radiation exposure to vulnerable organs, but the injection of the contrast medium can cause acute reactions. In a study of almost 33,000 consecutive patients receiving ExUS, 1.72 percent had acute reactions; 5 percent of these developed severe or life-threatening reactions, and one patient died (123). Patients with heart disease and diabetes are at higher risk for the development of undesirable side effects such as rapid heart beat or cardiac arrest (10,123).

As with the BE examination, radiologists perform most ExUS. According to California physicians' medicaid claims records for 1978,** radiologists performed 93 percent of all ExU exams on these patients. In hospitals (both inpatient and outpatient settings}, radiologists performed 100 percent of these exams, compared to 88 percent in physicians' offices.

In comparing the relative frequency of ExUS, hypertensive urograms, and drip infusion urograms, the standard procedure accounts for a large proportion of all urograms. Tables 13 and 14 present summaries of physicians' medicaid claims for ExU in 1978. California medicaid data indicated that from 1973 to 1976, the standard ExU comprised an average Of 84.5 percent of the ExUs performed in physicians' offices. In outpatient settings, the proportion was similar, at an average of 84.1 percent, and slightly lower for inpatient examinations (81. 7 percent). The data do not reveal any *trends* toward increases or decreases in the proportion of standard urograms performed during these 4 years.

The ExU examination is uncomfortable for the patient, and it requires some preparation, including prior dehydration and catharsis. Considered together with the morbidity, risk, and radiation exposure inherent in the procedure, and its cost, ExU is in little danger of becoming a routine or screening examination. Nevertheless, questions have arisen about the appropriateness of its use in certain patient groups such as women with urinary tract infections and about the potential for simplifying radioIogic methods

*The term excretory urogram is generally used interchangeably with intravenous pyelography (IVP) but the latter term is a misnomer and is avoided in modern texts (123).

**These are claims for physicians' services only. When the hospital bills directly for both technical and professional services, the claims are not included.

Table 13.—Medicaid Excretory Urogram Claims in California Submitted by Physicians in First Quarter of 1978, by Location of Service

	In patient[a]	Outpatient[a]	Office	Other	Total
74400 Excretory urogram.	789 (77.1 O/.)	1,011 (71.0°/0)	4,105 (82.20/.)	16 (84.2°/0)	5,921 (79.20/o)
74405 Excretory urogram: hypertensive.	19 (01 .90/o)	143 (lO.O°/O)	290 (05.80/.)	—	452 (06.0°/0)
74406 Excretory urogram: extended.	139 (13.6°/0)	128 (09.0%)	286 (05.7%)	3 (05.8°4)	556 (07.5°/0)
74410 Excretory urogram: infusion	76 (07.4%)	142 (lO.O°/~)	313 (06.0°/0)	—	531 (07.1 °/0)
Total	1,023	1,424	4,994	19	7,460

[a]Recorded onlywhen hospital and physician bill separately for the service (split bi l ling arrangement)

SOURCE. Urban Institute, 1980 Sample of 5,000 sole practitioners, including 177 radiologists

Table 14.—Medicaid Excretory Urogram Claims in California Submitted by Radiologists in First Quarter of 1978, by Location of Service

	Inpatient[a]	Outpatient[a]	Office	Other	Total
74400 Excretory urogram.	706 (88.30/.)	1,003 (77.9°/0)	2,031 (87.5%)	16 (lOO°/o)	3,756 (84.9°/0)
74405 Hypertensive ('69 RVS). ... , .	18 (O2.3%)	143 (11 .1%)	186 (08.0%)	—	347 (07.80/.)
74410 Infusion ('60 RVS).	76 (09.5°/0)	142 (1 1.0%)	103 (04.40/o)	—	321 (07.3°/0)
Total	800	1,288	2,320	16	4,424

[a]Recorded only when hospital and physician bill separately for the service (split billing arrangement)

SOURCE Urban Institute, 1980 Sample of 177 sole radiologists (326 percent of sole radiologists in Call fornia)

in order to save cost and reduce radiation exposure. A few clinical evaluative studies have addressed these issues.

The evaluative literature follows the following broad categories: 1) studies of diagnostic yield and high-yield criteria for ordering ExU, 2) studies of appropriate radiologic method, 3) studies of the role of ExU in screening for renovascular disease, and 4) studies of the information content of the ExU. Each is discussed below.

Diagnostic Yield and High= Yield Criteria

Van Woert and associates reported on a review of 200 consecutive hospital and clinic patients who had received ExUs during a 4-month period in 1955-56 (120). Fifty-five examinations (27.5 percent) were abnormal. The investigation correlated ExU results with prior signs, symptoms, or findings. In 50 percent of patients with a history of colicky pain, ExU was positive. Other signs or symptoms found to be important predictors of abnormality were palpable renal mass (5o percent); hematuria (40 percent); white blood cells in urine (25 percent); and costovertebral angle tenderness (20 percent). Moreover, combinations of these symptoms increased the probability of an abnormal ExU. Using this list as a starting point, the authors constructed a set of 10 indications for ordering an ExU. Use of the criteria constructed in this arbitrary fashion would have eliminated 25 percent of the 200 ExU examinations while at the same time missing 4 of the 55 abnormal ExU findings. In the opinion of the authors, none of the 4 was clinically significant, It must be noted, however, that the performance of the criteria was not tested on independent patient samples.

More recently, Mellins and associates studied over **1,622 patients** referred to a major teaching hospital for ExU (89). Referring physicians were asked to complete a questionnaire giving information on signs, symptoms, and laboratory findings in each patient. These attributes were then associated with ExU findings by constructing attribute likelihood ratios. Those with likelihood ratios suggesting a significant relationship between the attribute and ExU results were considered high-yield criteria. In patients without known prior genitourinary disease, these high-yield findings included gross hematuria (visible blood in urine); nocturia (excessive urination at night); elevated creatinine level; proteinuria (excess of serum proteins in urine); and elevated BUN (urea nitrogen) levels. Had these findings been used as referral criteria in the patient sample studied, 65 percent of ExU examinations would have been avoided, but 55 percent of abnormal ExU examinations would have been missed. The authors concluded that the use of the high-yield criteria so constructed would provide an unacceptably high false-negative rate (low sensitivity) and therefore offered no improvement over current referral processes.

An example of the potential benefits to be derived from further studies of the contribution of particular indications to diagnostic yield is given in two recent studies of diagnostic yield of ExU in women with recurrent urinary tract infection (37, 46). In the first, a review of 164 cases in which an ExU was performed at a university-affiliated hospital on women whose major indication for the examination was a history of recurrent urinary tract infection showed a diagnostic yield of 5.5 percent. Only one positive examination (0.6 percent) revealed a potentially significant abnormality, but it was considered unrelated to the symptoms of urinary tract infection. These results were supported by those of a similar study of **104 patients** referred for excretory urography **(46).** Although it is unknown how frequently ExU examinations are performed for this indication, promulgation of better criteria for ordering the procedure is likely to make some difference to total ExU volume.

Studies of Appropriate Radiologic Method

New variations in radiologic method are frequently tested and reported. Here, one element of radiological method with significant implications for the cost of medical care is reviewed: the number of films taken as part of the examination. In ExU, X-ray films are taken both before and at specific intervals after the injection of the contrast materials. Although these intervals vary with the clinical problem (e. g., the sequence of films is more rapid for hypertensive patients than for other patients), there is wide variation in the number and timing of films used in standard examinations,

Van Woert and colleagues addressed the question of film sequence in 1958 in a study whose purpose was to search for ways of decreasing gonadal radiation (120). In 87 positive ExU examinations for ureteral abnormalities, films were exposed at 5-, 10-, and 15-minute intervals after injection of the contrast. In only 3 of the 87 cases could the pathology be demonstrated on the early films but not on the 15-minute or subsequent films; the 3 cases involved clinically insignificant conditions. Thus, in none of the 87 cases was the 5- or 10-minute film essential to diagnosis of lower urinary tract disease. Gonadal shielding on the early films was recommended. It should be noted that the study was a review of past urograms, probably with reviewer knowledge of the original findings. It is not clear how an independent blind review of these cases in a larger sample containing both normal and abnormal urograms would have fared in detecting ureteral abnormalities on only the 15-minute film.

More recently Hillman and his associates investigated the sensitivity, specificity, and accuracy of ExU exams using different film sequences (61). Four experienced staff radiologists viewed 45 urograms, 21 normal and 24 abnormal, with abnormal cases representing a broad spectrum of diseases, The radiologists first reviewed a one-film examination (15 minutes); a three-film examination (1, 10, and 15 minutes); and a six-film examination (1, 5, 10, 10, 10, and 15 minutes). ' Mean sensitivity (true-positive rate) was 93, 88, and 91 percent on the one-, three-, and six-film urograms, respectively. Mean specificity (true-negative rate) was 67, 77, and 80 percent, respectively. Accuracy was 82, 83, and 85 percent. Thus, while specificity increased with additional films, sensitivity was not affected. However, the definitiveness of disease diagnosis (i. e., ability to correctly differentiate among possible

*The six-film examination involved three different views at 10 minutes, diseases) did increase as the number of films increased (75, 85, and 84 percent, respectively). On the basis of these findings, the authors recommended a policy of a one-film postinjection examination (at 10 minutes) followed by reinfection of contrast if the film indicates an abnormal. The authors made a rough estimate of national annual savings of $368 million if this policy were followed. Better estimation procedures would probably reduce this value substantially. Nevertheless, the authors have demonstrated that economies are possible through changes in radiographic method without significant changes in diagnostic efficiency.

Studies of the Role of ExU in Screening for Renovascular Hypertension

It has been estimated that about 10 percent of persons with hypertension have the condition as a consequence of underlying renovascular disease (13). In particular, stenosis (narrowing) of the renal artery is often present, and in some cases, surgery can correct the resulting hypertension. Thus, differentiation of patients with essential hypertension (i. e., not secondary to another condition) from those with renovascular hypertension has been of interest to clinicians and researchers alike. In 1961, a cooperative study of renovascular hypertension was initiated, with the purpose of improving understanding of the disease, its detection, and therapy (13), Consequently, a great deal of information exists on the efficiency of alternative diagnostic methods and outcome of therapies.

ExUs show features suggestive of renovascular disease. The major features are a significant difference in length of the kidneys; a significant difference in the time until contrast appears at the calyx on each side; and a difference in concentration of contrast medium on each side (13). The difference-in-appearance time mandates a large number of early postcontrast films. A fairly uniform technique is to take X-ray films at intervals of 1, 2, 3, 4, 5, 10, 15 and 30 minutes after the rapid injection of contrast (13), The urogram does not provide definitive diagnosis, however. Angiography, an invasive, difficult, and expensive procedure can be used to prove renovascular disease. * Urography is one of several signs, symptoms, and laboratory findings potentially useful in screening for the disease. The central question involving the use of urography is whether it should be used as a general or selective screening tool for detection of renovascular disease in hypertensive patients.

Bookstein and colleagues reported on a study of diagnostic efficiency of ExU in 198 patients with proved unilateral renovascular hypertension and 771 cases of essential hypertension in the Cooperative Study (12,13). The ExU false-positive rate in essential hypertension was 11.4 percent. In patients with serious unilateral renovascular disease, the true-positive rate was 78.2 percent.

The information generated by the Cooperative Study enabled McNeil and her colleagues (80) to perform a cost-effectiveness analysis of two issues: 1) the relative merits of screening hypertensives with ExU v. the renogram, a radionuclide procedure; and 2) the relative merits of screening v. no screening for renovascular diseases in hypertensive patients, In the former study, the evaluative measure chosen was the cost per surgical cure of each screening alternative. In the latter, the authors calculated the cost of keeping a patient from incurring a morbid event associated with hypertension for 16 years.

In the study of the ExU v. the renogram, the cost per surgical cure was found to be **$15,000** for ExU, $14,000 for the renogram, and $20,000 if both are performed. These estimates, while based on some rather heroic assumptions about the cost of diagnosis and treatment,** would suggest that the renogram be considered the primary screening technique.

The more germane issue is whether a general screening program in hypertensives is worthwhile at all. Here, the alternative strategies are: 1) screening for renovascular disease by ExU, with surgical intervention in those renovascular

‘Recent advances in angiographic techniques that allow for direct visualization of the renal arteries with less invasive methods may make the ExU obsolete in the future.

** See Wagner (121) for a discussion of problem of cost measurement.

hypertensive found to be amenable to surgery; and 2) no screening and medical management of all patients. The cost of providing an additional well patient by the route of screening and surgery was $56,000. The cost of protecting women is higher. Cost also increases with decreasing diastolic blood pressures and with the rate of compliance with medical regimens.

The success of the authors in examining the impact of alternative diagnostic strategies on outcomes is due largely to the availability of a rich data base emanating from large-scale epidemiological and clinical studies. It might have been useful to know how the morbid and mortal events under the alternative strategies were dispersed through time, since bad outcomes that occur later are preferred to those that occur sooner. Nevertheless, information sufficient to make a decision about the usefulness of screening, and the appropriate diagnostic tools for screening, is present. Whether $56,000 per additional well person is too high or too low a price to pay for hypertensive screening is a political decision fraught with questions of equity and ethics. Should society pay to screen hypertensive in groups with historically low rates of compliance with medical regimens? These questions can be posed only because information is available to frame existing tradeoffs.

Studies of the Diagnostic Information From ExU

Thornbury, Fry back, and Edwards attempted to determine the extent to which ExU is useful in changin~ physicians' levels of uncertainty about the correct diagnosis (117). Their study was premised on the assumption that if a test substantially alters physicians' subjective probability assessments, then it has diagnostic value. In a prospective study of 67 patients referred to two outpatient clinics for ExU, the referring physicians were asked to estimate the probability of the most likely diagnosis prior to the test. After the results of the ExU were received, the physicians' ratio of posttest to pretest odds was constructed, and a logarithmic index of this ratio was used to measure the degree of change (either positive or negative) in these odds. An index value of zero would denote no effect of ExU on physicians' probabilities. Successively higher values would suggest larger impacts of the ExU on diagnostic certainty. Thirty-five of the sixty-seven patients had an index value of less than 0.5, indicating little or no change in probability estimates resulting from the information provided by the ExU. These results contrast with those of the American College of Radiology study, where only 34 percent of ExU examinations performed in participating emergency rooms received an index value of less than 0.75 (76). The differences in the two studies could result from inconsistencies in study designs or differences in the patient mix.

In any case, it is difficult to draw inferences for medical practice for these studies. The study sample had various presenting signs and symptoms; whether the ExU is useful for some of these and not for others was not explored. Thornbury and his associates (117) noted that about one-half of the cases in which no change in probabilities was recorded were ultimately diagnosed as essential hypertension, indicating that these tests were performed to screen for renovascular hypertension. Without information on the findings from all hypertension patients in the study and separate analysis of those with other presenting conditions, these studies are virtually meaningless.

Appendixes

Appendix A. —Glossary of Terms

Angiography. —Radiography of vessels after the injections of a radiopaque material.
Asymptomatic. —Without symptoms.
Calvarial.—Relating to the skull cap.
Calyx. —One of the branches of recesses of the pelvis of the kidney.
Costovertebral angle tenderness. —Tenderness in the ribs and bodies of the thoracic vertebrae.
Colicky pain.—Spasmodic pain in the abdomen.
Cortical atrophy .—Wasting of tissues in the outer portion of the cerebrum.
Diagnostic efficiency. —A general term to denote the ability of a test to diagnose correctly.
Diagnostic accuracy. —An index of diagnostic efficiency, defined as the proportion of test results that are correct.
Diagnostic yield. —The proportion of test results that are positive.
Evaluative endpoint. —The measure or measures chosen to denote the effectiveness of a diagnostic procedure.
Fluoroscope. —Use of a fluorescent screen in rendering X-ray shadows visible.
Gonad dose. –The amount of radiation absorbed by the gonads resulting from any part of the body being exposed to X-rays.
Hematoma. —A bruise confined to a particular organ or tissue.
Hematuria. —Blood in the urine.
High-yield criteria. —Signs, symptoms, or risk factors occurring in a patient which are related to high probability that a test result will be positive.
Incidence. —The rate of occurrence of a disease or condition in a defined time period.
Information value of a test. —The ability of a test to alter the probabilty of a disease, measured either subjectively or objectively.
Intracranial. —Within the skull.
Likelihood ratio. —The ratio of true-positive to false-positive test results.
Metastasis .—The shifting of a disease, or its local manifestations, from one part of the body to another.
Organic. —Relating to an organ.
Photofluorography. —Fluoroscope; the recording on film of fluoroscopic views.
Polyp.—A general term for any mass of tissue that projects outward from normal surface tissue.
Prevalence. —The relative frequency of a disease or condition in a population at any point in time,
Prospective studies. —Studies that follow patients and collect research data during the course ok treatment.
Rad.—Radiation absorbed dose, or the energy absorbed at a particular point in a substance.
Radiogram. —A record made by means of X-rays ~>r a radioactive substance.
Radiography. —The making of a radiogram.
Radiologic method. —Techniques in radiographer which are at the discretion of the physician, such as number of views, positioning, choice of contrast medium, length of exposure, patient preparation, etc.
Renovascular disease. —Disease of the blood vessels of the kidneys.
Risk factors .—Characteristics of an individual, such as race, sex, age, other demographic variables, genetic variables, or aspects of lifestyle which predispose the person to a particular disease.
Roentgenograph.—Examination of any part of the body for diagnostic purposes by using roentxen rays (i. e., X-rays).
Screening.—Performance of a test in an individual with no relevant symptoms.
Sensitivity .—The ability of a test to detect disease when it is present. Measured as the proportion of diseased individuals whose test results are positive.
Specificity .—The ability of a test to rule out disease when it is not present. Measured as the proportion of normal individuals whose test results are negative.
Stenosis.—A narrowing or stricture of a vessel or valve.
Subdural hematoma.—A bruise occurring beneath the dura, or outer envelope of the brain.
Symptomatic .—Any departure from the normal in function, appearance, or sensation experienced by the patient and indicative of disease.

Appendix B.—Description of Other Volumes of the Assessment

The overall OTA assessment, *The Implications of Cost-Effectiveness Analysis of Medical Technology,* consists of a main, policy oriented report plus five background papers. The present volume, *Four Common X-Ray Procedures: Problems and Prospects for Economic Evaluation,* is one of the background papers. The main report and the other background efforts are briefly described below.

The main report examines three major issues: 1) the general usefulness of CEA/CBA in decisionmaking regarding medical technology, 2) the methodological strengths and shortcomings of the technique, and 3) the potential for initiating or expanding the use of CEA/CBA in six health care programs (reimbursement coverage, health planning, market approval for drugs and medical devices, Professional Standards Review Organizations, R&D activities, and health maintenance organizations), and most importantly, the implications of any expanded use.

The prime focus of the report is on the applications of CEA/CBA to medical technology (i.e., the drugs, devices, and medical and surgical procedures used in medical care, and the organizational and support systems within which such care is provided). With the exception of the present background paper on psychotherapy, the report does not address psychosocial medicine. Other aspects of health, such as the environment, are not directly covered either. The findings of the assessment, though, might very well apply to health care resource decisionmaking in general, and with modification, to other policy areas such as education, the environment, and occupational safety and health.

The main report contains chapters on methodology, general decisionmaking, each of the six health programs mentioned above, and the general usefulness of CEA/CBA. It contains appendixes covering a survey of current and past uses of CEA/CBA by agencies (primarily Federal), a survey of the resource costs involved in conducting CEA/CBAs, a discussion of ethical issues and CEA/CBA, and a brief discussion of legal issues.

Background Paper #1:Methodological Issues and Literature Review, includes an in-depth examination of the decisionmaking context and methodology discussions presented in this report. A critique of the literature, a bibliography of over 600 items, and abstracts of over 70 studies and other articles are also included.

In order to help examine the applicability of techniques to assess the costs and benefits of medical technology, 19 case studies were prepared. All 19 are available individually. In addition, 17 of the cases are available collectively in a volume entitled *Background Paper #l: Case Studies of Medical Technologies.* Some of the cases represent formal CEAS (e.g., the case on bone marrow transplants), and some represent net cost or "least cost" analysis (e. g., the case on certain respiratory therapies). Other cases illustrate various issues such as the difficulty of conducting CEA in the absence of adequate efficacy and safety information (e. g., the case on breast cancer surgery), or the role and impact of formal analysis on policymaking (e. g., the case on end-stage renal disease interventions). The 17 case studies in *Back-ground Paper #2* and their authors are:

Artificial Heart
  Deborah P. Lubeck
  John P. Bunker
Automated Multichannel Chemistry Analyzers
  Milton C. Weinstein
  Laurie A. Pearlman
Bone Marrow Transplants
  Stuart O. Schweitzer
  C. C. Scalzi
Breast Cancer Surgery
  Karen Schachter
  Duncan Neuhauser
Cardiac Radionuclide Imaging
  William B. Stason
  Eric Fortess
Cervical Cancer Screening
  Bryan R. Luce
Cimetidine and Peptic Ulcer Disease
  Harvey V. Fineberg
  Laurie A. Pearlman
Colon Cancer Screening
  David M. Eddy
CT Scanning
  Judith L. Wagner
Elective Hysterectomy
  Carol Korenbrot
  Ann B. Flood
  Michael Higgins
  Noralou Roos
  John P. Bunker
End-Stage Renal Disease Interventions
  Richard A. Rettig
Gastrointestinal Endoscopy
  Jonathan A. Showstack
  Steven A. Schroeder

Neonatal Intensive Care
Peter Budetti
Peggy McManus
Nancy Barrand
Lu Ann LeRoy
Orthopedic Joint Prosthetic Implants
Judith D. Bentkover
Philip G. Drew
Periodontal Disease Interventions
Richard M. Scheffler
Sheldon Rovin
Respiratory Therapy
Richard M. Scheffler
Morgan Delaney

The 18th case study is Background Paper **#3: The** *Efficacy and Cost Effectiveness of Psychlotherapy.* The 19th case study is the present volume, *Back-groung Paper* **#5:** Four Common X-Ray Procedures: Problems and Prospects for Economic Evaluation. It was prepared by Judith L. Wagner, and Martha J. Krieger.

Backround Paper #4: The Management of Health Care Technology in Ten Countries is an analysis of the policies, programs, and methods, including cost-effectiveness and cost-benefit techniques, that nine industrialized nations other than the United States use to manage the effects of medical technology. The experience of these nine countries in managing medical technology is compared to that of the United States. The paper on the United States and the comparative analysis were prepared by OTA staff, assisted by Louise Russell. The authors of the papers on the nine foreign countries are:

United Kingdom
Barbara Stocking
Canada
Jack Needleman
Australia
Sydney Sax
Japan
Joel Broida
France
Rebecca Fuhrer
Germany
Karin A. Dumbaugh
Netherlands
L. M. J. Groot
Iceland
David Gunnarson
Duncan vB. Neuhauser
Sweden
Erik H. G. Gaensler
Egon Jonsson
Duncan vB. Neuhauser

A related report prepared by OTA and reviewed by the Advisory Panel to the overall assessment is A Review of Selected Federal Vaccine and Immunization Policies. That study, published in September of 1979, examined vaccine research, development, and production; vaccine efficacy, safety, and cost effectiveness; liability issues; and factors affecting the use of vaccines. Pneumococcal vaccine was used as a case study, and a CEA/CBA was performed.

References

References

1. Abrams, H. L., "The 'Overutilization' of X-Rays," *N. Eng. J.* Med. 300:1213, **1979.**
2. Albert, R. K., et al., "Monitoring Patients With Tuberculosis for Failure During and After Treatment," Am, *Rev. Resp. Dis.* 114:1051, 1976.
3. Amberg, J. R., et al., "Opinion: Colonic Polyp Detection: Role of Roentgenography and Colonoscopy," Radiology 125:225, 1977.
4. American Cancer Society, Inc., "ACS Report on the Cancer-Related Health Check-up," *Ca-A Cancer Journal for Clinicians,* 30, July/August 1980.
5. American Thoracic Society, "Discharge of Tuberculosis Patients From Medical Surveillance (Policy Statement)," Am. *Rev. Resp. Dis.* 113: 709, 1976.
6. Bailey, W. C., et al., "Evaluating the Need for Periodic Recall and Reexamination of Patients With Inactive Pulmonary Tuberculosis," *Am. Rev. Resp. Dis.* 107:854, 1973.
7. Balint, J., et al., *Gastrointestinal Bleeding: Diagnosis and Management* (New York: John Wiley & Sons, 1977).
8. Bartha, G. W., and Nugent, C. A., "Routine Chest Roentgenograms and Electrocardiograms—Usefulness in the Hypertensive Work-up," *Arch. Int. Med.* 138:1211, 1978.
9. Bell, R. S., and Loop, J. W., "The Utility and Futility of Radiographic Skull Examination for Trauma," N. *Eng. J. Med.* 284:236, 1971.
10. Benacerraf, B. R., et al., "An Assessment of the Contribution of Chest Roentgenography in Outpatients With Acute Chest Complaints: A Prospective Study," *Radiology,* forthcoming.
11. Bonebrake, C., R., et al., "Routine Chest Roentgenography in Pregnancy," *J.A. M.A.* 240:2747, 1978.
12. Bookstein, J. J., et al., "Radiologic Aspects of Renovascular Hypertension, Part 1: Aims and Methods of Radiology Study Group," *J. A.M.A.* 220:1218, 1972.
13. ______ "Radiologic Aspects of Renovascular Hypertension, Part 2: The Role of Urography in Unilateral Renovascular Disease," *J. A.M. A.* 220:1255, 1972.
14. Borrie, J., "Routine Chest X-Rays," letter, *N.Z. Med. J.* 86:240, **1977.**
15. Boucot, K. R., and Weiss, W., "Is Curable Lung Cancer Detected by Semiannual Screening?" *J. A, MA.* 224:1361, 1973.
16. Brett, G. Z., "Earlier Diagnosis and Survival in Lung Cancer," *Br.* Med. *J.* 4:260, 1969.
17. Brill, P. W., et al., "The Value of Routine Chest Radiography in Children and Adolescents," *Pediatrics* 52:125, *1973.*
18. Brubaker, W. W., "Pre-Employment Physical Examinations," *Penn. Med.* 75:53, 1972.
19. Clarke, A. M., and Jodes, I. S. L., "Diagnostic Accuracy and Diagnostic Delay in Carcinoma of the Large Bowel," *N.Z. J. Med.* 71:341, 1970.
20. Cohen, M. F., et al., "Dollar Cost per Positive Test for Multiphasic Screening," *N. Eng. J. Med.* 28:459, 1970.
21. Cooley, R. N., et al., "Diagnostic Accuracy of the Barium Enema Study in Carcinoma of the Colon and Rectum," *Am. J. Roentgenol.* 84:316, 1960.
22. Dales, L. J., et al., "Evaluating Periodic Multiphasic Health Checkups: A Controlled Trial," *J. Chron. Dis.* 32:385, *1979.*
23. ______ "Evacuation of a Periodic Multiphasic Health Checkup," *Meth. Infor. Med. 13:140, 1974.*
24. Department of Health, Education, and Welfare Public Health Service, Center for Disease Control, "Health Status of Indochinese Refugees," *Morbidity & Mortality* 28:385, 1979.
25. ______, Center for Disease Control, "Recommendation for Health Department Supervision of Tuberculosis Patients," Morbidity & *Mortality* 23:75, 1974.
26. ______, Food and Drug Administration, *Annotated Bibliography on the Selection of Patients for X-Ray Examination,* HEW publication No. (FDA) 73-8067, July 1978.
27. ______ Food and Drug Administration, "The Chest X-Ray as a Screening Procedure for Cardiopulmonary Disease: A Policy Statement," April 1973.
28. ______ Food and Drug Administration, Bureau of Radiological Health, *The Selection of Patients for X-Ray Examinations,* HEW publication No. (FDA) 80-8014, January 1980.
29. ______, Health Resources Administration, National Center for Health Statistics, *Current Estimates From the Health Interview Survey, United States* 1973, Vital and Health Statistics Series 1O-NO. (HRA) 75-1522, October 1974.
30. ______, Office of Health Research, Statistics, and Technology, National Center for Health Statistics, *Detailed Diagnoses and Surgical Pro-*

cedures for Patients Discharged From Short-Stay Hospitals: United States, 1977, DHEW publication No. (PHS) 79-1274, September 1979.

31 ______, *Population Exposure to X-Rays, United States 1970*, DHEW publication No. (FDA) 73-804F, November 1973,

32, DeSmet, A. A., et al., "A Second Look at the Utility of Radiographic Skull Examination for Trauma," *Am. J. Roentgenol.* 132:95, *1979*.

33. Eddy, D. M., *Screening for* **Cancer:** *Theory, Analysis and Design* (Englewood Cliffs, N. J.: Prentice Hall, 1980).

34. Edsall, J., and Collins, G., "Routine Followup of Inactive Tuberculosis: A Practice To Be Abandoned," Am. *Rev. Resp. Dis.* 107:851, 1973,

35. Evers, K., et al., "The Double Contrast Enema in Carcinoma of the Rectum," presented at the Annual Meeting of the Radiological Society of North America, Atlanta, Ga., Nov. 29, 1978.

36. Eyes, B., and Evans, A. F., "Post-Traumatic Skull Radiographs, Time for a Reappraisal," *Lancet* 2:85, July 8, 1978.

37. Fair, W. R., et al., "Are Excretory Urograms Necessary in Evaluating Women With Urinary Tract Infection?" *Urology* 121:313, 1979.

38, Farnsworth, P. B., et al., "The Value of Routine Preoperative Chest Roentgenograms in Infants and Children," *J. A.M.A.* 244:582, *1980*.

39. Feingold, A, O., "Cost Effectiveness of Screening for Tuberculosis in a General Medical Clinic," *Pub. Hlth. Rpts.* 90:544, *1975*.

40. ______, "Routine Chest Roentgenograms on Hospital Admissions Do Not Discover Tuberculosis," *South. Med. J.* 70:579, 1977.

41. Feinstein, A. R., *Clinical Biostatistics* (St. Louis: C. V. Mosby Co., 1977).

42. Finkel, A. J., "Chest X-Ray Screening for Tuberculosis," *J, A.M.A.* 226:799, *1973*.

43, Fontana, R. S., "Early Diagnosis of Lung Cancer," Am. *Rev. Resp. Dis.* 116:399, 1977.

44. Fontana, R. S., and Sanderson, D. R., "American Cancer Society's Revised Guidelines for Annual Chest Film—A Dissenting Opinion," editorial, *J, A.M.A.* 244:592, 1980.

45. Forrest, J. V., and Sagel, S. S., "The Lateral Radiograph for Early Diagnosis of Lung Cancer," *Radiology 131:309*, *1979*.

46. Fowler, J. E., and Pulaski, E. T., "Excretory Urography, Cystography, and Cystoscopy in the Evaluation of Women With Urinary Tract Infection: A Prospective Study," *N. Eng. J. Med.* 304:462, *1981*.

47, Franklin, R., and McSwain, B., "Carcinoma of the Colon, Rectum, and Anus," *Ann. Surg.* 17: *811*, *1970*.

48. Gempel, P. A., et al., "Comparative Cost Analysis: Computed Tomography vs. Alternative Diagnostic Procedures, 1977 and 1980" (Cambridge, Mass.: Arthur D. Little, Inc., December 1977).

49, Gerson, D. E., et al., "The Barium Enema: Evidence for Proper Utilization," *Radiology 130: 297*, 1979.

50. Gilbertsen, V. A., and Lillehei, J., "The Chest X-Ray in the Diagnosis of Lung Cancer," *Br. J. Clin. Prac.* 23:149, 1969.

51. Gerry, G. A., et al., "The Diagnostic Importance of the Normal Finding," N. *Eng. J. Med.* 298:486, 1978.

52. Grossman, L. K., et al., "Roentgenographic Follow-Up of Acute Pneumonia in Children," *Pediatrics* 63:30, *1979*.

53. Gryzbowski, S., and Coy, P., "Early Diagnosis of Carcinoma of the Lung: Simultaneous Screening With Chest X-Ray and Sputum Cytology," *Cancer* 25:113, *1970*.

54. Habicht, J.-P., "Assessing Diagnostic Technologies," letter, *Science* 207:1414, 1980.

55. ______, "Some Characteristics of Indicators of Nutritional Status for Use in Screening and Surveillance," *Am. J. Clin. Nutr.* 33:531, *1980*.

56. Hammar, L. M., "Admission Chest Film Program," *Minn. Med.* 52:1899, *1969*.

57. Hanley, J. A., and McNeil, B. J., "Maximum Attainable Discrimination and the Utilization of Radiologic Examinations" (Boston: Sidney Farber Cancer Institute, August 1980).

58, Harwood-Nash, D., et al., "The Significance of Skull Fractures in Children," *Radiology 101: 151, 1971*.

59, Hendrix, T. R., and Saba, G. P., *The Radiographic Examination of the Colon*, Department of Health and Human Services, Public Health Service, Office of Health Research, Statistics, and Technology, National Center for Health Care Technology, 1980.

60. Hessel, S, J., et al., "Improving Performance by Multiple Interpretations of Chest Radiographs: Effectiveness and Cost," *Radiology* 127:589, *1978*,

61. Hillman, B., et al., "Simplifying Radiological Examinations: The Urogram as a Model," *Lancet* 1:1069, *1979*.

62. Horowitz, O., and Darrow, M. M., "Principles and Effects of Mass Screening: Danish Experi-

ence in Tuberculosis Screen ing, " *Pub. Hlth. Rpts.* 91:146, 1976.

63. Jacobs, J. C., "Chest X-Ray Screening for Tuberculosis, " letter, *). A.M.A.* 228:24, *1974.*

64. Jamieson, K. G., and Yuelland, J. D. N., "Surgically Treated Traumatic Subdural Hematomas," *J. Neurosurg.* 37:137, 1972.

65. Jay, S. J., et al., "The Radiographic Resolution of Streptococcus Pneumonia Pneumonia, " *N. Eng. J. Med.* 293:798, 1975.

66. Jergens, M. E., et al., "Selective Use of Radiography of the Skull and Cervical Spine, " *West. J.* Med. 127:1, 1977.

67 Jones, M. J., et al., "Infectious Diseases of Indochinese Refugees, " *Mayo Clinic Proc.* 55:482, 1980.

68. Kido, D. K., et al., "Comparative Sensitivity of CT Scans, Radiographs, and Radionuclide Bone Scans in Detecting Metastic Calvarial Lesions, " *Radiology* 128:371, *1978.*

69. Laufer, I., "The Double-Contrast Enema: Myths and Misconceptions, " *Castro. Radiol.* 1:19, *1976.*

70, ______, *Double Contrast Gastrointestinal Radiology With Endoscopic Correlation* (Philadelphia: W. B. Saunders Co., 1979).

71. Laufer, I., et al., "The Radiological Demonstration of Colorectal Polyps Undetected by Endoscopy, " *Gastroenterology* 70:167, 1976.

72. Leinicke, J. L., et al., "A Comparison of Colonoscopy of Roentgenography for Detecting Polypoid Lesions of the Colon, " *Gastro. Radiol.* 2:125, 1979.

73. Lewis, W. W., "Mobile X-Ray Units: A Critical Look," *NTRDA Bulletin* 57:45, *1971.*

74. Lilienfeld, A., et al., "An Evaluation of Radiologic and Cytologic Screening for the Early Detection of Lung Cancer: A Cooperative Pilot Study of the American Cancer Society and the Veterans Administration, " *Cancer Res.* 26:2082, *1966.*

75. Loder, R. E., "Routine Pre-Operative Chest Radiography: 1977 Compared With 1955 at Peterborough District General Hospital, " *Anesthesia* 33:972, 1978.

76. Lusted, L. B., "A Study of the Efficacy of Diagnostic Radiology Procedures" (Chicago: American College of Radiology, 1977).

77. MacEwan, D. W., et al., "Manitoba Barium Enema Efficacy Study, " *Radiology* 126:39, *1978.*

78. McKenzie, C. J. G., "A Two Year Followup of Persons With Non-Tuberculosis Chest Disease Found at 'Operation Doorstep' Vancouver 1964, " *Can. Med. Assoc. J.* 103:1019, 1970,

79. McNeil, B. J., "Pitfalls in the Requirements for Evaluations of Diagnostic Technologies, " in *Medical Tech) tology*, J. Wagner (cd.), Urban Institute Conference, NCHSR Research Proceedings Series, DHEW publication No. (PHS) 79-3254, Department of Health, Education, and Welfare, Public Health Service, Office of Health Research, Statistics, and Technology, National Center for Health Services Research, 1979.

80. McNeil, B. J., and Adelstein, S. J., "Measures of Clinical Efficacy: The Value of Case Finding in Hypertensive Renovascular Disease, " N, *Eng. J. Med.* 193:220, *1975.*

81. McNeil, B. J., and Mellins, H. Z., "Evaluation of Diagnostic Test in Uroradiology, " *Rad. Clin. N. Amer.* 17:175, 1979.

82. Margulis, A. R., "Examination of the Colon, " in *A limentary Tract Roentgenology*, A. R. Margulis and H. J. Burheene (eds.) (St. Louis: C. V. Mosby Co., 1973).

83. ______, "Is Double-Contrast Examination of the Colon the Only Acceptable Radiographic Examination?" *Radiology* 119:741, 1976.

84. Margulis, A. R., and Goldberg, H. I., "The Current State of Radiologic Technique in the Examination of the Colon: A Survey, " *Rad, Clin. N. Amer.* 7:27:42, *1969.*

85. Martin, G., "Skull X-Ray Policy, " letter, *Lancet* 2:572, 1978.

86. Marton, K., "Upper Gastrointestinal Examination, " presented at the National Conference on Referral Criteria for X-Ray Examinations, Bureau of Radiological Health, Food and Drug Administration, Public Health Service, Department of Health, Education, and Welfare, Washington, D. C., Oct. 25-27, 1978.

87. Mattox, J. H., "The Value of a Routine Prenatal Chest X-Ray, " *Obstet. & Gyn.* 41:243, 1973.

88. Melamed, M., et al., "Preliminary Report of the Lung Cancer Detection Program, " *Cancer 39*: 369, 1977.

89. Mellins, H. Z., et al., "The Selection of Patients for Excretory Orography, " *Radiology 130:293, 1979.*

90. Memorial Sloan Kettering Cancer Center, Office of Cancer Communications, Office of Cancer Control, "MSKCC Study Shows X-Rays and Sputum Cytology Can Help Reduce Lung Cancer Mortality Rates, " *Cancer Control Commun.* 3:1, 1980.

91. Miller, R., "Detection of Colon Cancer and the Barium Enema, " *], A.M.A.* 230:1195, *1974.*

92. Miller, R. E., and Lehman, G., "The Barium Enema: Is It Obsolete?" *J. A.M.A.* 235:2842, *1976.*

93. ______, "Polypoid Colonic Lesions Undetected by Endoscopy," *Radiology* 129:295, 1978.

94. National Tuberculosis and Respiratory Disease Association, "Chest X-Ray Screening Recommendations for TB-RD Associations, " *NTRDA Bulletin*, October 1971.

95. Neuhauser, D., "Cost-Effective Clinical Decision Making, " *Pediatrics* 60:756, 1977,

96. Neuhauser, D., and Lewicki, A. M., "What Do We Gain From the Sixth Stool Guaiac?" *N. Eng. J. Med.* 293:226, 1975,

97. Office of Technology Assessment, U. S. Congress, *Assessing the Efficacy and Safety of Medical Technologies*, GPO Stock No. 052-003-00593-0 (Washington, D. C.: U.S. Government Printing Office, September 1978).

98. Petterson, S. R., and Janower, M, L., "Is the Routine Preoperative Chest Film of Value?" *Appl. Radiol.* 6:70, 1977.

99 Phillips, L, A., "Final Report of the Demonstration Project on Emergency Facility Skull Radiography" (Seattle, Wash.: Washington State Professional Standards Review Organization, Aug. 7, 1978).

100. ______, *A Study of the Effect of High Yield Criteria for Emergency Room Skull Radiography,* U.S. Department of Health Education, and Welfare, Public Health Service, Food and Drug Administration, Bureau of Radiological Health, HEW publication No. (FDA) 78-8069, July 1978.

101. Rees, A. M., et al., "Routine Preoperative Chest Radiography in Non-Cardiopulmonary Surgery, " *Br. Med. J.* 1:1333, 1976.

102. Reichman, L. B., "Tuberculosis Screening and Chest X-Ray Films, " *Chest* 68 (3 Suppl.):448, 1975.

103. Royal College of Radiologists, "Preoperative Chest Radiology: National Study by the Royal College of Radiologists, " *Lancet* 2:83, 1979.

104. Sackett, D. L., and Holland, W. W., "Controversy in the Detection of Disease, " *Lancet* 2:357, 1975.

105, Sagel, S. S., et al., "Efficacy of Routine Screening and Lateral Chest Radiographs in a Hospital Based Population, " *N. Eng. J. Med.* 291:1001, 1974.

106 Sane, S. M., et al., "Value of Preoperative Chest X-Ray Examinations in Children, " *Pediatrics* 60:669, 1977.

107, Sarkies, N. J. C., "Skull X-Ray Policy," letter, *Lancet* 2:312, 1978.

108. Schneider, W. J., and Dykan, M., "The Preplacement Medical Evaluation of Hospital Personnel, " *J. Ocup. Med.* 20:741, 1978.

109, Schoenbaum, S. C., et al., "Benefit-Cost Analysis of Rubella Vaccination Policy, " *N. Eng. J. Med.* 294:306, 1975.

110. Stadalnik, R. C., et al., "Electrocardiographic Response to Intravenous Urography: Prospective Evaluation of 275 Patients, " *Am. J. Roentgenol.* 129:825, 1977.

111. Steel, K., et al., "Laboratory Screening in the Evaluation and Placement of Geriatric Patients," *J. Am. Ger. So.* 22:538, 1974.

112. Stitik, F. P., and Tockman, M. S., "Radiographic Screening in the Early Detection of Lung Cancer, " *Rad. Clin. N. Amer.* 16:347, 1978.

113. Swallow, J., and Sbarbaro, J. A., "Analysis of Tuberculosis Casefinding in Denver, Colorado, 1965 -1970," *Health Serv. Rpts.* 87:375, 1972.

114. Talalla, A., and Morin, M. A., "Acute Traumatic Subdural Hematoma: A Review of One Hundred Consecutive Cases, " *J. Trauma* 11:771, 1971.

115. Tenret, J., "Is a Systematic Annual Chest Film Useful?" *Conn. Med.* 39:604, 1975.

116. Thoeni, R. F., and Menuck, L., "Comparison of Barium Enema and Colonoscopy in the Detection of Small Colon Polyps, " *Diag. Radio).* 134:631, 1977.

117. Thornbury, J. R., et al., "Likelihood Ratios as a Measure of the Diagnostic Usefulness of Excretory Urogram Information, " *Radiology* 114: 561, 1975.

118. Torchia, M., and DuChez, J., *Chest X-Ray Screening Practices, A) I Annotated Bibliography,* U.S. Department of Health, Education, and Welfare, Public Health Service, HEW publication No. (FDA) 80-8116, March 1980.

119 Tsai, L., and Tsuang, M. T., "Computerized Tomography and Skull X-Rays: Relative Efficacy in Detecting Intracranial Disease, " *Am. J. Psychiatry* 135:1556, 1978.

120. Van Woert, I., et al., "Radiation Hazards of Intravenous Pyelography, " *). A.M.A,* 166:1826, 1958.

121 Wagner, J. L., "The Feasibility of Economic Evaluation of Diagnostic Procedures: The Case of CT Scanning, " in *The Implications of Cost-Effectiveness Analysis of Medical Technology/ Background Paper #2: Case Studies of Medical Technologies,* prepared by OTA, U.S. Congress (Washington, D. C.: U.S. Government Printing Office, April 1981).

122. Williams, C. B., et al., "Colonoscopy in the Management of Colon Polyps, " *Br. J. Surg.* 61: 573, 1974,

123. Witten, D. M., et al., *Emmett's Clinical LJrography*, vol. 1 (Philadelphia: W. B. Saunders co., 1977).
124. Wolff, W., et al., "Comparison of Colonoscopy and the Contrast Enema in Five Hundred Patients With Colorectal Disease, " *Am. J. Surg.* 129:181, 1975.
125, Zelen, M., and Feinleib, M., "On the Theory of Screening for Chronic Diseases, " *Biometrika 56: 601, 1969.*
126, Zimmerman, R. A., et al., "Cranial Computed Tomography in Diagnosis and Management of Acute Head Trauma, " *Am. J. Roentgenol.* 131: *27, 1978.*