Technetium Bone Scanning in the Diagnosis of Osteomyelitis:

A Meta-analysis of Test Performance

BENJAMIN LITTENBERG, MD, ALVIN I. MUSHLIN, MD, ScM, and the DIAGNOSTIC TECHNOLOGY ASSESSMENT CONSORTIUM

Purpose: ***To determine the diagnostic performance of technetium bone scanning in the setting of possible osteomyelitis in the foot of a patient who has diabetes or other vasculopathy.***

Design: ***Meta-analysis.***

Data identification and study selection: ***To be eligible for inclusion, a report must have used intravenous technetium-99m methylene diphosphonate or a similar agent in humans over the age of 16 years, must have addressed possible osteomyelitis of the lower extremity with ulcer or soft-tissue inflammation in the setting of diabetes, neuropathy, or vasculopathy, and must have allowed the generation of a two-by-two table. A structured search of the MEDLARS database found 296 possibly eligible reports; ten met all the inclusion criteria.***

Data extraction and synthesis: ***The reported sensitivity and specificity of each report were converted to their logistic transforms and a straight line was fitted by weighted least-squares regression. The line was then back-transformed to yield a summary receiver operating characteristic curve. The false-positive rate of the bone scan is at best in the range of 10 to 20%. This occurs at sensitivities between 70 and 80%. The studies with increased sensitivity also reported sizable increases in the false-positive rate ranging from 20 to over 90%. Even small increases in sensitivity have necessitated large sacrifices in specificity. Seven of the ten studies reported specificities under 70%.***

Conclusions: ***Published data defining the effectiveness of technetium bone scanning for the diagnosis of osteomyelitis in the impaired foot indicate relatively poor performance. In many clinical situations, the specificity of the bone scan will not be high enough to confirm the diagnosis of osteomyelitis.***

Key words: ***osteomyelitis; radionuclide imaging; diabetes mellitus; technetium bone scanning; meta-analysis.*** **J GEN INTERN MED 1992;7:158–163.**

DISTINGUISHING SOFT-TISSUE INFECTIONS from deeper bone infections in diabetes patients and other patients with vascular or neurologic deficits has significant management and prognostic implications. Unfortunately, clinical examination and plain radiography alone may not suffice to make the diagnosis. Many clinicians rely on radionuclide scanning with technetium isotopes, especially technetium-99m methylene diphosphonate (Tc99m MDP), to help determine whether the bone is infected in order to assist in the planning of antibiotic therapy. Negative scans indicate cellulitis and may result in a decision to treat with a short course of antibiotics. A positive scan showing accumulation of the radioisotope in the bone itself may lead to a commitment to a prolonged course of antibiotics for presumed osteomyelitis. How accurate this test is in making these clinical distinctions may, therefore, have important implications for the outcomes and costs of care.

We reviewed the literature on the performance of technetium bone scanning for the diagnosis of osteomyelitis of the lower extremity in patients with underlying vascular insufficiency or peripheral neuropathy. To estimate the test's performance and its value for differentiating osteomyelitis from soft-tissue infection in this situation, we used meta-analytic methods developed specifically for summarizing the performances of diagnostic test results from multiple studies.

METHODS

Literature Review

Our goal was to find all published reports of the performance of technetium bone scanning in diagnosing osteomyelitis in impaired feet and ankles. Impairment is most commonly due to diabetes, but studies including patients with other conditions causing vascular disease or peripheral neuropathy were eligible for inclusion. All eligible reports met these criteria:

1. The study was of humans more than 16 years old.
2. Intravenous Tc99m MDP or a similar agent was used.

Received from the Dartmouth–Hitchcock Medical Center, Hanover, New Hampshire (BL), and the University of Rochester Medical Center, Rochester, New York (AIM).

Supported by a grant from the John A. Hartford Foundation to the Diagnostic Technology Assessment Consortium. The Consortium includes: Dartmouth–Hitchcock Medical Center: W. Blair Brooks, MD, Terry Hurlbut III, MD, Benjamin Littenberg, MD, Andre Midgette, MD, MPH, David Smith, MD, Harold C. Sox, Jr., MD, and Carole Toselli, MD; Massachusetts General Hospital: Albert Mulley, Jr., MD; Stanford University: Lincoln E. Moses, PhD; University of Pennsylvania: Bruce Kinosian, MD, and J. Sanford Schwartz, MD; University of Rochester Medical Center: Daniel Kido, MD, Alvin I. Mushlin, MD, and Charles E. Phelps, PhD; and University of Washington: Richard Hoffman, MD, Dan Kent, MD, and Eric Larson, MD. Dr. Littenberg is an American College of Physicians George Morris Piersol Teaching and Research Scholar.

Address correspondence to Dr. Littenberg: Technology Assessment Program, 564 Borwell Building, Lebanon, NH 03756.

Reprints are not available.

3. The report addressed possible osteomyelitis of the distal lower extremity with ulcer or soft-tissue inflammation in the setting of diabetes, neuropathy, or vasculopathy.
4. A complete count of true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) patients (a two-by-two performance table) could be generated from the data in the paper.
5. The study included at least ten subjects, including at least one reference-positive and one reference-negative subject.

Literature Searches

We conducted four independent literature searches. Two were performed by physician-investigators who were familiar with the medical literature and with the clinical problem. Two more searches were performed by professional medical librarians with extensive expertise in medical bibliography and computerized medical literature databases but only a layman's appreciation of the clinical problem.

The four searchers constructed their own strategies based on key words, text words, and other searching devices of their own choosing and applied their strategies to the MEDLARS database. For example, one of the searchers used this strategy:

(scan$ or radionuclid$ or radioisotop$ or technetium or 99mtc)

and (foot or feet or leg or legs or ulcer or ulcers or bone)

and osteomyelitis

The dollar sign ($) is a "wild card" symbol that causes the computer to search for any character string. "scan$" retrieves "scan," "scans," "scanning," etc.

The four searches were then merged into a single list of references (many with abstracts as well as MeSH key words and bibliographic source data). Two physician-investigators reviewed the merged list and marked each entry according to the predefined eligibility criteria. If the abstract, key word, and title were not adequate to determine eligibility, we obtained the report from the library and reviewed it in full. We obtained English translations as necessary.

We also reviewed the reference lists of all included articles after the initial MEDLARS search and examined the citations for possible eligible reports. Each excluded report was coded with the first reason determined for exclusion. We did no further analysis of excluded papers.

Available Studies and Their Quality

In all, we examined 296 articles (the vast majority from the MEDLARS database) and found ten studies that were eligible for inclusion in this summary. Reports were most commonly excluded because they did not contain original data (38 reports), they were not actually about osteomyelitis (99), they did not investigate technetium bone scanning (60), or they did not address pedal inflammation (28). Other reports were excluded because they did not study adults (14), had fewer than ten subjects (34), did not study humans (10), or were not available (2). One additional study[1] was excluded after preliminary analysis because it used a substantially lower dose of isotope (6 mCi) than did the others (18 to 22 mCi). Not surprisingly, it reported substantially lower performance than did the full-dose reports. The remaining eligible reports are described in Tables 1 and 2.

Most reports were from departments of radiology or nuclear medicine. Two reports came from departments of podiatric surgery.[3,4] The sample sizes ranged from the minimum required for inclusion (ten)[3] to 94.[5] A total of 305 data points were included in the ten studies. Because five of the studies counted an impaired limb rather than a patient as the unit of analysis, the total number of patients studied is not known. Six studies used retrospective designs; four were prospective.

Nine of the included studies used technetium-99m methylenediphosphonate or hydroxymethylenediphosphonate in doses ranging from 18 to 22 mCi. Seven of these studies employed three-phase techniques, one used a four-phase method,[2] and one used single-phase scans.[4] One study[3] did not report the technical details of the bone scans.

For a reference ("gold standard") test, most of the studies relied on combinations of pathologic findings following surgery or biopsy and clinical follow-up after treatment. One study relied, in part, on other imaging modes,[2] and one used discharge information[8] to determine the actual diagnosis. Nearly half of the studies appeared to have interpreted the bone scans without reference to the gold standard, but did not actually state that the reference test had been interpreted independent of the scan. Four studies did not address the issue of independence at all. Only two eligible reports[7,9] clearly maintained independence. Seven of the ten studies had assembled their patients through referrals to radiology: entry to the study was by virtue of a request for an imaging test. Two studies had used the reference test to obtain patients.[3,7] Only one study appeared to have used a clinical complaint related to possible osteomyelitis as the enrollment criterion.[4] Most of the reports provided some description of the populations studied. None, however, described comorbidity to any significant degree. Half of the reports failed to supply descriptive data concerning gender and age of

the subjects. Two of the reports did not indicate what fractions of their populations had diabetes.

Meta-analysis

Data reporting the performance of bone scanning were extracted from each eligible report and used to generate a summary receiver operating characteristic (SROC) curve. The method for deriving the SROC curve is described in Appendix A and elsewhere.[12] The resulting curve describes how the test's performance in those with osteomyelitis [sensitivity or true-positive rate (TPR)] varies with its performance in those without osteomyelitis [false-positive rate (FPR) or 1 − specificity].

Differences among the reported accuracies are due to several factors. First, some investigators may have used a stricter threshold or cutoff to declare a test "positive." These studies may report better specificity but sacrifice some sensitivity compared with those using a more lax threshold to define a positive test, and vice versa. Second, the reports may differ because of random variations in the performance (accuracy) of the test. Third, differences in the populations studied, the diseases sought, the settings in which the test is employed, and the methods of the index and reference tests may have profound effects on reported sensitivity and specificity. All these differences may lead to heterogeneity among the eligible reports that should argue against considering them as all estimating one underlying sensitivity and one underlying specificity. Rather, these factors (especially threshold) may cause the reported test performance to vary among studies in a way best described by an ROC curve that is consistent with the

TABLE 1

The Eligible Reports Used in the Meta-analysis

Report*	Types of Patients	Prevalence of Diabetes (%)	Number of Bone Scan Phases	Dose of Tc99m† (mCi)	Reference Test	Study Design
Alazraki et al. 1985[2]	Patients with lower-extremity ulcers and possible osteomyelitis, referred for bone scan	?	4	20	Clinical course, surgical bone specimen, x-rays, gallium scans, and computed tomography	Retrospective
Caprioli et al. 1986[3]	Biopsy subjects with suspected osteomyelitis and pedal ulceration	?	?	?	Percutaneous bone culture	Retrospective
Hetherington 1982[4]	Patients with peripheral neuropathy with foot ulceration admitted to podiatry service	64	1	18	Bone biopsy and culture	Prospective
Keenan et al. 1989[5]	Long-standing diabetic patients referred for evaluation of possible infection	100	3	22	Bone culture, histologic examination, or outpatient follow-up	Prospective
Maurer et al. 1986[6]	Patients with possible osteomyelitis and radiographic evidence of osteoarthropathy referred for indium and technetium scanning	100	3	20	Surgical bone specimen or 4-month follow-up	Retrospective
Park et al. 1982[7]	Patients with a variety of infections referred for bone biopsy or surgery	100	3	20	Bone biopsy	Retrospective
Schauwecker et al. 1988[8]	Patients with suspected osteomyelitis with radiographic changes consistent with neuropathic foot disease	74	3	?	Bone histology or discharge diagnosis and follow-up	Prospective
Segall et al. 1989[9]	Mixed population of patients with underlying pathologic conditions referred for bone scan	54	3	21	Surgical or biopsy specimen of bone	Retrospective
Seldin et al. 1985[10]	Patients with suspected osteomyelitis with bone scan and "proven diagnosis" available	90	3	20	Bone biopsy, culture, or both	Retrospective
Yuh et al. 1989[11]	Consecutive patients with suspected osteomyelitis and/or nonhealing foot ulcers	100	3	22	Pathologic or bacteriologic specimen of bone	Prospective

*For complete reference citations, see the reference list. †Technetium-99m.

TABLE 2

The Reported Results of the Eligible Studies*

Report†	Prevalence (%)	TP	FP	FN	TN	TPR	FPR
Alazraki et al. 1985[2]	25	4	2	1	13	0.80	0.13
Caprioli et al. 1986[3]	60	6	3	0	1	1.00	0.75
Hetherington 1982[4]	86	12	2	0	0	1.00	1.00
Keenan et al. 1989[5]	40	38	35	0	21	1.00	0.62
Maurer et al. 1986[6]	31	3	4	1	5	0.75	0.44
Park et al. 1982[7]	58	20	3	1	12	0.95	0.20
Schauwecker et al. 1988[8]	49	17	18	0	0	1.00	1.00
Segall et al. 1989[9]	42	7	8	3	6	0.70	0.57
Seldin et al. 1985[10]	53	15	2	1	11	0.94	0.21
Yuh et al. 1989[11]	62	17	9	1	2	0.94	0.82

*TP = number of true-positive subjects; FP = number of false-positive subjects; FN = number of false-negative subjects; TN = number of true-negative subjects; TPR = true-positive rate = TP/(TP+FN); FPR = false-positive rate = FP/(FP+TN).

†For complete reference citations, see the reference list.

reported data but allows for differences among the studies. Further, the impact of these factors on the ROC curve may be assessed by sensitivity analysis (Appendix A).

RESULTS

The estimated SROC curve appears in Figure 1. This curve results from a straight line with slope (b) of −0.103 and intercept (i) of +2.243 based on the logistic transforms of the operating characteristics of the bone scan from the ten studies (Appendix A). We did not extrapolate the estimated SROC curve beyond the region where eligible reports provided information. Because none of the eligible reports presented data in the lower left region of the SROC space, the curve does not extend to the origin.

As can be seen in Figure 1, the FPR of the bone scan is at best 20%. This occurs at sensitivities (TPRs) from 70 to 80%. The studies with increased sensitivities also report sizable increases in the FPRs ranging from 20 to over 90%. Even small increases in sensitivity necessitate large sacrifices in specificity. Seven of the ten studies reported FPRs over 30%.

We compared the summarized test performances in subgroups of reports based on the year of publication, the prevalence of disease in the population studied, whether a prospective or a retrospective study design was used, the sample size, the reference test method (pathology report vs. clinical follow-up), whether the criteria for a positive scan were noted, whether analysis was by patient or by lesion, and the prevalence of diabetes in the study groups. In spite of the large apparent differences between the subgroups compared, none of these factors had a statistically significant effect on the SROC curve.

Six reports indicated the apparent causes of false-positive scan results. Among the 50 false-positive results in those six series, 14 were called "neuropathic" by the authors, four were labeled "neurotrophic," two were at the site of a fracture, two were due to "gout," and one was noted to be a "Charcot's joint." Twenty-seven were not explained.

DISCUSSION

The available literature about the performance of bone scans in this clinical situation is seriously limited. Although most papers supply details about the technical aspects of the bone scans, few studies ensured that the subjects formed an unbiased sample of a relevant population. None of the studies scanned an entire population of subjects selected because of a clinical suspicion of osteomyelitis and then determined the final diagnosis in a reliable and complete manner. Most commonly, these studies were based on retrospective collections of patients referred for bone scans who also had bone biopsies.

All the studies suffered from problems in obtaining an adequate gold standard. First, there is controversy about what constitutes a perfect reference test, even under ideal circumstances. The presence of bacteria on bone culture or histopathologic examination is persuasive (although not absolutely convincing) for the diagnosis of osteomyelitis. A negative result is less satisfactory, as it is easy to imagine the trocar's missing the site of infection. Untreated osteomyelitis rarely resolves[13] and often progresses to obvious manifestations of underlying bone infection. However, what of the patient whose condition worsens on antibiotic therapy? Or improves? How shall these cases be judged? It may be neither practical nor ethical to deny the use of antibiotics to ensure that the clinical course provide a valid indication of the true extent of the disease. For this analysis, we accepted as an "adequate" reference test biopsy, long-term clinical follow-up (with or without therapy), or both.

Sample sizes in this literature are small. The largest report had only 94 subjects, and the mean was fewer

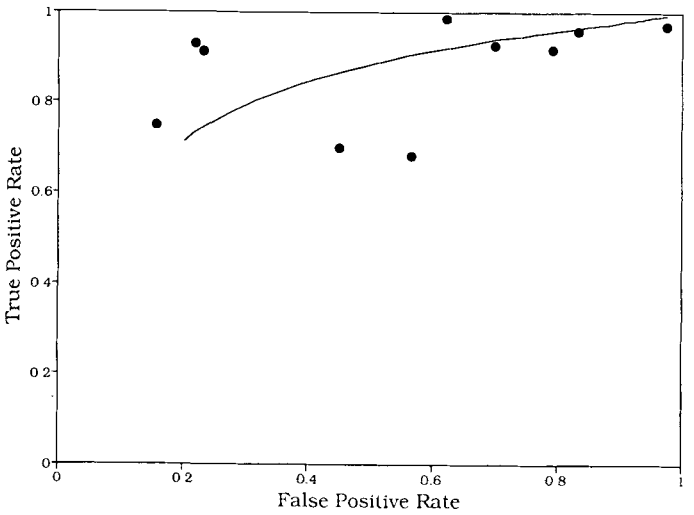

FIGURE 1. The estimated summary receiver operating characteristic (SROC) curve based on the ten reports studied. The vertical axis represents the true-positive rate; the horizontal axis represents the false-positive rate. Each of the ten reports is represented as a small filled circle. The position of each report indicates its reported performance. The parameters of the curve (see the text) are b = −0.103 and i = +2.243.

than 31. The biologic and clinical heterogeneity of osteomyelitis demands that more patients be studied to uncover any inherent differences in test performance between relevant clinical subsets.

Missing any relevant literature could have altered our estimated test performance. To prevent this, we performed an exhaustive literature search. Nonetheless, it is possible that an eligible report eluded our attention.

The substantial heterogeneity among the known eligible reports, coupled with their less-than-optimal design characteristics, points to the need for further prospective evaluation of this technology before a stable estimate of performance can be produced. It is even more difficult to evaluate the influences of study and population characteristics. With only ten eligible studies, only very large differences between subgroups can reach statistical significance. In other words, this meta-analysis lacked statistical power to reject hypotheses based on subgroup analysis. However, the fact that multiple sensitivity analyses failed to find significant differences argues that the overall estimated SROC curve (based on all ten eligible reports) is an appropriate summary of the available literature.

Despite these limitations, our analysis provides some useful findings. It confirms the clinical impression that technetium bone scanning is not highly accurate. In particular, the scan is often unable to permit distinction of osteomyelitis from competing diagnostic entities such as soft-tissue infection, neurotrophic lesions, gout, and stress fractures. In many cases, the scan is positive absent any bone or joint abnormality.

For distinguishing the presence or absence of osteomyelitis, the SROC curve provides the best current summary of the performance of the technology. It is more informative than a single fourfold table based on pooling all the available data. Rather than ignoring the diversity of results in the literature, the SROC curve takes into account the range of possible combinations of sensitivities and specificities that best describes the included reports. We chose to represent performance as a SROC curve, rather than as a point estimate, because we believe that different thresholds (and other differences) are generally present in different reports.

The SROC curve resulting from this meta-analysis (Fig. 1) suggests that technetium bone scanning is less accurate than previously expected. To achieve a TPR (sensitivity) of 90% would necessitate that over 50% of patients with soft-tissue infection or other inflammatory conditions be misdiagnosed as having osteomyelitis (FPR = 54%). Likewise, altering the diagnostic threshold such that the FPR is only 25% means that nearly a fourth of the patients with bone involvement would be missed (sensitivity = 77%).

These results have implications for clinicians using this test in this difficult diagnostic situation. Given the range of possible operating characteristics for this test, it is imperative that the sensitivity and specificity values in each local community or hospital be known before a bone scan is interpreted. We expect that in many clinical settings, the specificity of the bone scan will not be high enough to confirm the diagnosis of osteomyelitis. In other words, the clinician should not recommend a costly, risky, or noxious therapy on the basis of a positive bone scan alone. Furthermore, unless the prior probability of osteomyelitis is very low, the reassurance provided by a negative bone scan will be limited. At present, the test is probably best used in patients strongly suspected of having osteomyelitis for whom a positive scan will be followed by a biopsy to confirm the presence of bony infection. For those with a negative scan, if conservative treatment with oral antibiotics is chosen, the clinical plan should include close patient follow-up to ensure that osteomyelitis has not been falsely missed. Further conclusions about its role in diagnosis will have to await a full analysis of the clinical implications and costs of the errors produced by the test versus the help it provides in diagnosis. Until then, the information now available from this meta-analysis should serve to inform clinicians about the inherent diagnostic ability and limitations of technetium bone scanning.

ADDENDUM

Since we prepared the manuscript, an additional eligible report was published.[14] This report focused primarily on leukocyte scanning with Indium In111, but provided data about the use of Tc99m as well. Eighteen of 26 patients with osteomyelitis were detected by

bone scanning (TPR = 0.69). Five of 13 without osteomyelitis tested positive (FPR = 0.38). The recalculated summary ROC curve with the additional study included has a slope (b) of −0.012 and an intercept (i) of 1.867. This indicates slightly lower test performance. The difference may not be clinically significant.

The authors are indebted to Gloria Linder, MLS, of Stanford University and Kathryn Nesbitt, MLS, of the University of Rochester for superlative library services.

REFERENCES

1. Ignatov A, Dyankov L, Velichkov L, Pampulov L, Gavrailov M, Andreyev D. Radioisotopic diagnosis of ossa pedis changes in diabetes. Med Radiol (Mosk). 1979;24(6):34-8.
2. Alazraki N, Dries D, Datz F, Lawrence P, Greenberg E, Taylor A Jr. Value of a 24-hour image (four-phase bone scan) in assessing osteomyelitis in patients with peripheral vascular disease. J Nucl Med. 1985;26:711-7.
3. Caprioli R, Testa J, Cournoyer RW, Esposito FJ. Prompt diagnosis of suspected osteomyelitis by utilizing percutaneous bone culture. J Foot Surg. 1986;25(4):263-9.
4. Hetherington VJ. Technetium and combined gallium and technetium scans in the neurotrophic foot. J Am Podiatr Med Assoc. 1982;72(9):458-63.
5. Keenan AM, Tindel NL, Alavi A. Diagnosis of pedal osteomyelitis in diabetic patients using current scintigraphic techniques. Arch Intern Med. 1989;149:2262-6.
6. Maurer AH, Millmond SH, Knight LC, et al. Infection in diabetic osteoarthropathy: use of indium-labeled leukocytes for diagnosis. Radiology. 1986;161:221-5.
7. Park H-M, Wheat LJ, Siddiqui AR, et al. Scintigraphic evaluation of diabetic osteomyelitis: concise communication. J Nucl Med. 1982;23:569-73.
8. Schauwecker DS, Park HM, Burt RW, Mock BH, Wellman HN. Combined bone scintigraphy and indium-111 leukocyte scans in neuropathic foot disease. J Nucl Med. 1988;29:1651-5.
9. Segall GM, Nino-Murcia M, Jacobs T, Chang K. The role of bone scan and radiography in the diagnostic evaluation of suspected pedal osteomyelitis. Clin Nucl Med. 1989;14:255-60.
10. Seldin DW, Heiken JP, Feldman F, Alderson PO. Effect of soft-tissue pathology on detection of pedal osteomyelitis. J Nucl Med. 1985;26:988-93.
11. Yuh TC, Corson JD, Baraniewski HM, et al. Osteomyelitis of the foot in diabetic patients: evaluation with plain film, 99mTc-MDP bone scintigraphy, and MR imaging. AJR Am J Roentgenol. 1989;152:795-800.
12. Moses LE, Littenberg B, Shapiro D, Rabinowitz D. Some data-analytic methods for combining estimates of sensitivity and specificity into a single ROC curve for a diagnostic test. Submitted to Statistics in Medicine, 1990.
13. Bamberger DM, Daus GP, Gerding DN. Osteomyelitis in the feet of diabetic patients: long-term results, prognostic factors, and the role of antimicrobial and surgical therapy. Am J Med. 1987;83:653-60.
14. Newman JG, Waller J, Palestro CJ, et al. Unsuspected osteomyelitis in diabetic foot ulcers. JAMA. 1991;266:1246-51.

The appendix follows on the next page

Appendix A

Curve Estimation and Analysis

To construct the best estimate of the performance of bone scans, the true-positive rate (TPR) and the false-positive rate (FPR) from each study were transformed into their corresponding logistic equivalents and weighted least-squares regression was performed.[12] First, a two-by-two table of performance was constructed for each eligible report:

	Osteomyelitis Present	Osteomyelitis Absent
Bone Scan Positive	Number of true-positive subjects (TP)	Number of false-positive subjects (FP)
Bone Scan Negative	Number of false-negative subjects (FN)	Number of true-negative subjects (TN)

To avoid taking the logarithm of zero, a correction factor of 0.5 was added to each cell. Then TPRs and FPRs were calculated for each table:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

The TPRs and the FPRs were converted to their logistic equivalents:

$$\text{logit}(TPR) = \ln\left(\frac{TPR}{1 - TPR}\right)$$

$$\text{logit}(FPR) = \ln\left(\frac{FPR}{1 - FPR}\right)$$

Two additional measures were derived:

$$D = \text{logit}(TPR) - \text{logit}(FPR)$$

$$S = \text{logit}(TPR) + \text{logit}(FPR)$$

D is a measure of the performance of the test in that it is related to how well the test (at the given value of S or cutoff) distinguishes between the "sick" and the "well." S is related to the diagnostic cutoff chosen by the original investigators in their assessment. Investigators with "strict" criteria for calling a scan "positive" will tend to report a low TPR (low sensitivity) as well as a low FPR (high specificity). If the threshold is low or "lax," sensitivity improves, specificity declines, and both TPR and FPR go up. The variances of S and D are equal and are calculated as:

$$\text{Var} = \frac{1}{TP} + \frac{1}{FP} + \frac{1}{FN} = \frac{1}{TN}$$

Weighted least-squares regression was performed with S as the independent variable against D as the dependent variable and the inverses of the variances serving as the weights. The regression generally yields a line with slope (b) near zero whose height (the intercept, i, of the regression analysis) is a measure of the overall performance of the test. The line was back-converted to the familiar representation of FPR vs. TPR according to this formula:

$$TPR = \frac{1}{1 + \dfrac{1}{e^{\frac{i}{1-b}} \cdot \left(\dfrac{FPR}{1 - FPR}\right)^{\frac{1+b}{1-b}}}}$$

e is the base of the natural logarithm. This equation traces out a curve that is consistent with each of the included reports of TPR and FPR and serves as a summary receiver operating characteristic (SROC) curve. Changes in b lead to slightly different shapes of the SROC curve. As i increases, the SROC curve more closely approaches the shape of a "perfect test" that operates in the upper left corner of the graph with high TPR and low FPR. Sensitivity analyses were performed by recalculating and redrawing the SROC curve with selected studies omitted. Subgroups of reports were analyzed by comparing the distances of the transformed (linear) data from the overall regression line for individual subgroups. The averages of the two subgroups were compared using Student's t-test.