ELSEVIER

Journal of Clinical Epidemiology 53 (2000) 852–862

Journal of Clinical Epidemiology

Meta-analysis of diagnostic tests for acute sinusitis

Eric A. Engels, M.D., M.P.H.[a,*], Norma Terrin, Ph.D.[a], Michael Barza, M.D.[b], Joseph Lau, M.D.[a]

[a]*Division of Clinical Care Research, Department of Medicine, New England Medical Center, Tufts University School of Medicine, 750 Washington Street, Boston, MA 02111, USA*

[b]*Division of Geographic Medicine and Infectious Diseases and the Tupper Research Institute, Department of Medicine, New England Medical Center, Tufts University School of Medicine, 750 Washington Street, Boston, MA 02111, USA*

Received 2 August 1999; received in revised form 21 January 2000; accepted 21 January 2000

Abstract

To facilitate management of acute sinusitis, we conducted a meta-analysis of published studies comparing diagnostic tests for this disorder. Thirteen studies were identified through literature search. Based on sinus puncture/aspiration (considered most accurate), 49–83% of symptomatic patients had acute sinusitis. Compared with puncture/aspiration, radiography offered moderate ability to diagnose sinusitis (summary receiver operator curve [SROC] area, 0.83). Using sinus opacity or fluid as the criterion for sinusitis, radiography had sensitivity of 0.73 and specificity of 0.80. Studies evaluating ultrasonography revealed substantial variation in test performance. The clinical evaluation, particularly risk scores formally incorporating history and physical examination findings, had moderate ability to identify patients with positive radiographs (SROC area, 0.74). Many studies were of poor quality, with inadequately described test methods and unblinded test interpretation. In conclusion, acute sinusitis is common among symptomatic patients. Radiography and clinical evaluation (especially risk scores) appear to provide useful information for diagnosis of sinusitis.

Keywords: Sinusitis; Diagnostic tests; Meta-analysis; Receiver operator curve; Radiography; Clinical examination

1. Introduction

Acute sinusitis is one of the most common conditions for which patients seek care from physicians [1]. Annually in the United States, almost three million patients are diagnosed with acute sinusitis [2]. Because acute sinusitis may often be caused by bacterial infection, patients with symptoms compatible with this syndrome frequently receive antibiotics. As a result, acute sinusitis is the fifth most common reason for outpatient antibiotic prescription [3].

The most accurate and cost-effective methods for diagnosing acute sinusitis remain uncertain [4]. Diagnostic tests utilized in this setting include radiography, computed tomography, magnetic resonance imaging, and, in some European countries, ultrasonography. Sinus puncture, performed by otolaryngologists, may be considered the "gold standard" test for this condition, and purulent secretions on aspiration provide direct evidence for sinus inflammation [4,5]. Nonetheless, sinus puncture can only rarely be clinically justified, given its cost, inconvenience, and associated patient discomfort. Radiography and ultrasonography are more easily incorporated into a general medical practice, but techniques and diagnostic criteria have varied [6–10]. Often the diagnosis of sinusitis is made on clinical grounds, but the accuracy of the clinical evaluation is unclear [11]. Inaccurate diagnosis of acute sinusitis leads to suboptimal clinical outcomes. Failure to treat patients with acute sinusitis with antibiotics may delay symptom resolution [12]. Inadequate antibiotic treatment can lead to chronic sinusitis and, in rare cases, serious bacterial complications [13]. On the other hand, unnecessary antibiotic use adds to medical costs and treatment side effects, and hastens the emergence of antibiotic-resistant microorganisms [14]. Reliable information on diagnostic test performance might allow development of more cost-effective treatment strategies for the management of patients with acute sinusitis symptoms.

We conducted a meta-analysis of published data to evaluate diagnostic tests for acute sinusitis. Summary estimates of how these tests perform provide a basis for their rational use and highlight areas where more research is needed.

2. Methods

To identify published studies comparing diagnostic tests in acute sinusitis, we conducted a MEDLINE literature

*Corresponding author. Viral Epidemiology Branch, Division of Cancer Epidemiology and Genetics National Cancer Institute, 6120 Executive Blvd., MSC 7248, Rockville, MD 20822. Tel.: 301-435-4722; fax 301-402-0817.

E-mail address: engelse@exchange.nih.gov (E.A. Engels)

0895-4356/00/$ – see front matter
PII: S0895-4356(00)00209-2

search for English language articles published between 1966 and October 1998, using the sensitive (but nonspecific) strategy "sinusitis (MeSH heading or textword) AND human." This strategy identified 4070 articles; after a review of identified titles and abstracts, several hundred of these articles deemed potentially relevant were retrieved. Detailed examination of retrieved articles identified 49 studies of diagnostic tests or clinical criteria. We also identified additional articles in reference lists of retrieved articles and review articles and through consultation with colleagues and experts.

To be included in the present meta-analysis, studies in this final screened group had to compare the ability of two or more tests to diagnose acute sinusitis. In general, subjects were recruited into these studies when they presented with symptoms consistent with acute sinusitis. Because there is no universally accepted definition of acute sinusitis, we included studies regardless of their subject eligibility criteria; some studies did not provide an explicit description of who was eligible. In this meta-analysis, we included only studies that evaluated each test on all study subjects, to avoid the problem of verification bias. Verification bias occurs when a reference test is applied only to high-risk subjects, leading to inaccurate estimates of sensitivity and specificity of the test being evaluated [15]. Diagnostic tests included: clinical examination (studies had to evaluate a composite measure such as overall clinical impression or a decision aid such as a risk score), radiography, ultrasonography, and sinus puncture/aspiration. Although we searched for studies evaluating computed tomography or magnetic resonance imaging, no eligible studies were identified.

Each included study allowed estimation of sensitivity and specificity for at least one test ("test of interest") in comparison to another ("reference test"). Data were obtained from each study identifying individuals as having or not having sinusitis, for each of the evaluated diagnostic tests. To calculate sensitivity and specificity in a comparison of two diagnostic tests, we decided which was the test of interest and which was the reference test. In this regard, included studies presented comparisons of diagnostic tests in a manner consistent with a "hierarchy" of accuracy: most accurate was sinus aspiration/puncture, followed by radiography, ultrasonography, and then clinical examination. This hierarchy reflects how these diagnostic modalities are used in practice, in that uncertainty following clinical examination can prompt a clinician to obtain a more "definitive" test, such as radiography, and uncertainty following radiography can, in some circumstances, prompt sinus aspiration. As a consequence of this hierarchy, for example, we derive estimates of sensitivity and specificity of ultrasonography with respect to radiography (and not vice versa). Many studies reported data only for "sinuses" and not for "patients;" these data were used in the analyses. For each study, test data were abstracted in duplicate, with discrepancies between reviewers resolved through discussion.

When studies presented test performance data for more than one threshold or "cutpoint" for a test of interest, we extracted data for each cutpoint separately. For instance, for the clinical examination compared with radiography, data for overall clinical impressions of "intermediate probability" and "high probability" were included separately.

For each combination of test of interest and reference test, a summary receiver operator characteristic (SROC) curve was derived [16]. The SROC curve, combining data from all of the relevant studies, displays the trade-off between a test's sensitivity and specificity. Multiple data points from studies that provided data at different cutpoints were used to derive these curves; because these observations were therefore not independent, we did not calculate confidence intervals around SROC curves. Each data point was weighted by the inverse of the estimate's variance. When studies provided estimates of specificity over a wide range, the area under the SROC curve was calculated by extending the curve over the total possible range of specificities from 0 to 1. For subsets of studies that provided estimates for a narrow range of values, we calculated random effects-weighted averages for sensitivity and specificity [17]. Statistical calculations were performed using Meta-Test (version 0.6, available upon request from Joseph Lau).

From each study we also abstracted the country where the study was performed and publication date; subject age and duration of symptoms; location of the study (hospital, office practice, or emergency department); and specialty of physicians evaluating patients (primary care physicians or otolaryngologists). We noted whether the study report stated that each diagnostic test was evaluated in a manner blinded to the results of the other evaluated tests and whether the reports clearly defined the criteria for a positive test result.

3. Results

3.1. Description of studies and study subjects

Thirteen studies were included, five of which provided comparisons of more than two tests (Table 1). Eleven studies were conducted in Europe, including all of those utilizing sinus puncture as a reference test. Two studies were conducted in the United States [18,19]. Six studies recruited patients from a hospital or emergency room setting, three recruited from physicians' offices, one recruited from both settings, and three did not describe the study setting. In most studies, the physicians evaluating the subjects were otolaryngologists (Table 1).

The study by Jannert and colleagues was the only study restricted to children [20], although three other studies included some children $\leq$10 years of age [8,21,22]. Six described symptoms that subjects needed to have for study inclusion (typically nasal symptoms or headache); the remainder included individuals when they or their physicians suspected acute sinusitis. Only four studies stated how

Table 1
Description of studies of diagnostic tests

Study	Country, Year	Setting/ Specialty[a]	Patient characteristics: Age, years	Patient characteristics: Symptoms on entry[b]	Symptom duration, days	Study characteristics: Blinded	Study characteristics: Clearly defined positive tests[c]	Number of patients/ number analyzed[d]	Number of cutpoints or datasets available
Radiography, compared with puncture									
McNeill [8]	N.Ireland, 1962	Hosp / ENT	>9	MD	–	No	Yes	150 / 242 S	3
Revonta [21]	Finland, 1980	Off / ENT	7–71	MD	–	Partial	Yes	–/ 230 S	2
Kuusela [7]	Finland,1982	Hosp / ENT	adult	–	–	No	Yes	105/ 156 S	1
van Buchem [4]	Holland. 1995	Off, Hosp / Prim, ENT	>17	HA, Nas	–	Yes	Yes	40 / 62 S	1
Laine [6]	Finland, 1997	Off / Prim	16–68	MD	<30	Partial	No	39 / 72 S	1
Savolainen [9]	Finland, 1997	Hosp / ENT	17–68	Nas, other	<30	Partial	Yes	161 / 234 S	2
Ultrasonography, compared with puncture									
Revonta [21]	Finland, 1980	Off / ENT	7–71	MD	–	Partial	Yes	– / 430 S	3
Kuusela [7]	Finland, 1982	Hosp / ENT	adult	–	–	No	Yes	105 / 156 S	2
van Buchem [4]	Holland, 1995	Off, Hosp / Prim, ENT	>17	HA, Nas	–	Yes	No	113 / 159 S	2
Laine [6]	Finland, 1997	Off / Prim	16–68	MD	<30	Yes	Yes	39 / 72 S	1
Savolainen [9];	Finland, 1997	Hosp / ENT	17–68	Nas, other	<30	Partial	Yes	161 / 234 S	2
Clinical exam, compared with puncture									
Berg [23]	Sweden, 1988	Hosp / ENT	average 38	Nas	<90	Partial	No	155 / 155 S	5
Ultrasonography, compared with radiography									
Berg [22]	Sweden, 1985	Hosp / ENT	10–75	Nas	–	Partial	No	105 / 90	1
Rohr [19]	USA, 1986	– / –	18–74	MD	–	Yes	Yes	99 / 198 S	2
Jensen [24]	Sweden, 1987	Hosp / ENT	adult	Nas, other	–	Yes	Yes	138 / 253 S	2
Clinical exam, compared with radiography									
Axelsson [25]	Sweden, 1976	– / ENT	average 35	PT, MD	–	No	No	164 / 310 S	1
Jannert [20]	Sweden, 1982	– /–	0–15	MD	–	No	Yes	175 / 175	3
Williams [18]	USA, 1992	Off / Prim	>17	HA, Nas, PT	<90	Yes	Yes	247 / 247	7

[a] Off, office; Hosp, hospital or emergency room; Prim, primary care physician; ENT, otolaryngologist
[b] HA, headache; Nas, nasal symptoms; PT, patient-suspected; MD, doctor-suspected
[c] Study clearly describes criteria for positive results for both reference test and test of interest.
[d] "Number of patients" refers to the total number of patients. "S" indicates that the unit of analysis is the sinus, and that "number analyzed" refers to number of sinuses.

long subjects could have symptoms before evaluation: two studies limited duration of symptoms to 30 days [6,9], and two limited duration of symptoms to 90 days [23,18].

Only four studies stated that interpretation of both the reference test and test of interest occurred under blinded conditions [4,18,19–24]. Five other studies [6,9,21–23] described blinded interpretation of the test of interest, but investigators interpreting the reference test were not blinded to the results of the test of interest (in the study by Laine *et al.*, interpretation of sinus aspiration results occurred with knowledge of radiography but not ultrasonography results). Most studies adequately described criteria for positive tests, in a manner that would allow duplication of study results, although three studies did not describe positive criteria for reference tests [22,23,25] and four did not describe positive criteria for at least one test of interest [4,6,23,25]. Three studies used the patient as the unit of analysis for comparing diagnostic tests [18,20,22]. In the remaining 10 studies, the sinus was the unit of analysis.

3.2. Findings on sinus puncture

We included seven studies that used sinus aspiration, a direct method for visualizing sinus contents, as the reference test. Four studies aspirated sinus contents through a puncture of the inferior meatus [4,8,21,23], whereas three did not describe technique [6,7,9]. Criteria for sinusitis on puncture were aspiration of mucus or pus [6,8,9,21,23], "discharge" [7], or fluid [4]. Two studies reported results of microbiological cultures of aspirates, but these results were not used as part of the criterion for sinusitis [4,9], Van Buchem and coworkers [4] reported that pathogenic bacteria were isolated from sinus aspirates in 34% of patients (95% confidence interval 25–42%). Savolainen and colleagues reported that bacteria were isolated from 65% of punctured sinuses [9].

Four studies that used sinus puncture as a reference test provided sufficient data to estimate sinusitis prevalence among enrolled patients (in whom the clinical diagnosis of sinusitis had been considered plausible, based on entry criteria described in Table 1). Two studies [4,6] that included patients from general medical practices provided prevalence estimates of 49% and 51%. Two studies [7,9] that restricted enrollment to hospital-based otolaryngology clinics estimated prevalence at 53% and 83%.

3.3. Sinus radiography compared with sinus puncture

Six studies compared sinus radiography with sinus puncture [4,6–9,21]. Studies used a series of two or three radiographs, all including the occipito-mental (Water's) view. Because three studies provided more than one comparison of test performance, there were 10 values of sensitivity and specificity available for analysis.

Fig. 1 displays these 10 values of sensitivity and specificity and the SROC curve derived from them. The 10 data points appear to be well-described by the curve. Shown as black ellipses are the five values of sensitivity and specificity using the criterion "sinus fluid or opacity" to define positive radiographs. Shown as gray ellipses are the three values based on the criterion "sinus fluid or opacity or mucous membrane thickening" to define positive radiographs. The remaining two estimates are shown as white ellipses: one used the criterion "sinus opacity," while the explicit criterion was not available for the other. The area under the weighted SROC curve is 0.83.

Adding "mucous membrane thickening" as part of the criterion for a positive radiograph appeared to increase the sensitivity of radiography and decrease its specificity. Random effects estimates for sensitivity and specificity, using "fluid or opacity" as the definition of a positive radiograph, were 0.73 (95% confidence interval, 0.60–0.83) and 0.80 (0.71–0.87), respectively. With the definition of positive radiograph "sinus fluid or opacity or mucous membrane thickening," the estimates for sensitivity and specificity were 0.90 (0.68–0.97) and 0.61 (0.20–0.91), respectively. With positive radiographs defined as "sinus opacity," specificity increased only slightly to 0.85 (0.76–0.91), but sensitivity decreased dramatically to 0.41 (0.33–0.49).

3.4. Clinical examination compared with sinus puncture

A single study compared clinical examination with sinus puncture [23]. This study provided data for clinicians' overall impressions, and also for a risk score derived from the number of findings present from a four-item list: purulent rhinorrhea with unilateral predominance, local pain with unilateral predominance, bilateral purulent rhinorrhea, and presence of pus in the nasal cavity.

Fig. 2 (left panel) plots the five values of sensitivity and specificity obtained from this report. Shown as four separate gray ellipses are values of sensitivity and specificity for the risk score, derived using cutpoints of one, two, three, or four positive items out of the four-item list. The risk score appears to have better discrimination than the overall clinical impression (shown as a white ellipse); the SROC curve, fit only to the risk score values, has an area under the curve of 0.91.

Unfortunately, several difficulties with the study by Berg *et al.* [23] cast doubt on its internal validity. First, the reference test, sinus puncture and aspiration, is poorly described in the report. It is not clear whether radiography was used in conjunction with aspiration to identify subjects with sinusitis. Second, the report leaves unclear how "purulent rhinorrhea with unilateral predominance," "bilateral purulent rhinorrhea," and "pus in the nasal cavity" can be mutually independent risk score predictors of sinusitis.

3.5. Clinical examination compared with sinus radiography

Three studies compared clinical examination to sinus radiography. The Axelsson study compared a physician's overall clinical impression with sinus radiography [25]. The study by Jannert and colleagues evaluated a clinical risk score for children, with 0–3 of the following findings: puru-

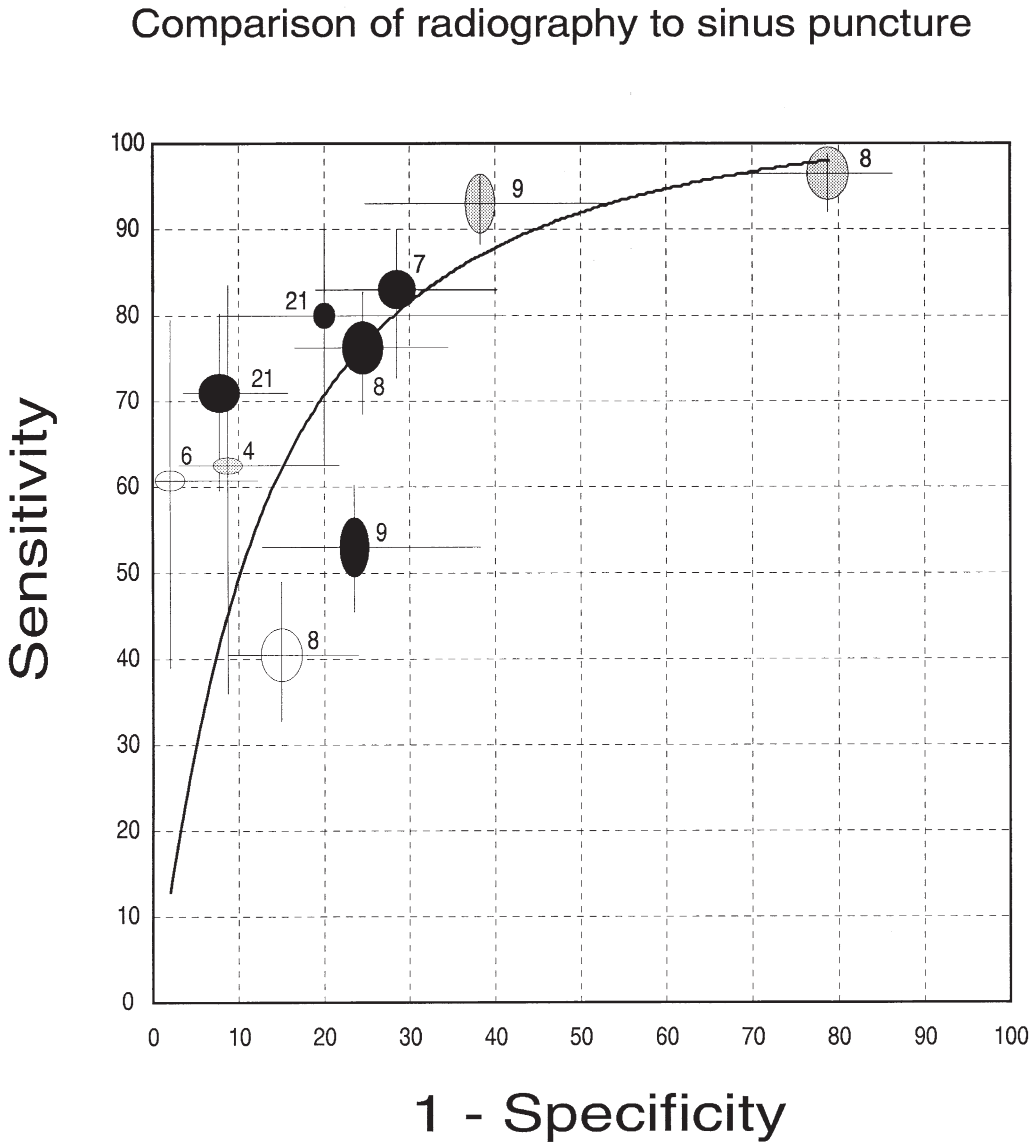

Fig. 1. Summary receiver operator characteristic curve for sinus radiography, compared with sinus puncture/aspiration for the diagnosis of acute sinusitis. Each ellipse corresponds to a study estimate of sensitivity and specificity; the area of each ellipse is proportional to the study's size, and horizontal and vertical lines associated with each ellipse correspond to 95% confidence intervals for the study's estimates. Black ellipses provide estimates using "sinus fluid or opacity" as the radiographic criterion for sinusitis. Gray ellipses provide estimates for "sinus fluid or opacity or mucous membrane thickening" as the radiographic criterion for sinusitis. The two white ellipses provide estimates for "sinus opacity" and for an unspecified diagnostic criterion. Numbers next to each ellipse refer to the study that provided data for that estimate, as identified in Table 1 and in the references.

lent nasal secretions on examination, history of upper respiratory infection during the previous two weeks, and sinus pain or tenderness [20]. The study by Williams *et al.* evaluated a clinical risk score for adults, with 0–5 of the following findings: maxillary toothache, abnormal transillumination, poor response to decongestants, purulent secretions on examination, and history of colored nasal discharge [18]. The Williams study also included data for overall clinical impressions of "intermediate" or "high" probability of sinusitis. There were therefore 11 values of sensitivity and specificity available for analysis.

Fig. 2 (right panel) displays these 11 values of sensitivity and specificity and the SROC curve. The points are well-described by the SROC curve. Values based on various cutoffs for the two risk scores, shown as gray ellipses, appear to have similar discrimination to the overall impressions of clinicians, shown as white ellipses. The area under the weighted SROC curve is 0.74.

3.6. Ultrasonography compared with sinus puncture or radiography

Eight reports [4,6,7,9,19,21,22,24] evaluated the ability of ultrasonography to diagnose acute sinusitis (Table 1). Seven studies utilized A-mode ultrasonography [6,7,9,19,21,22,24], with most using the presence of a "back wall echo" as a posi-

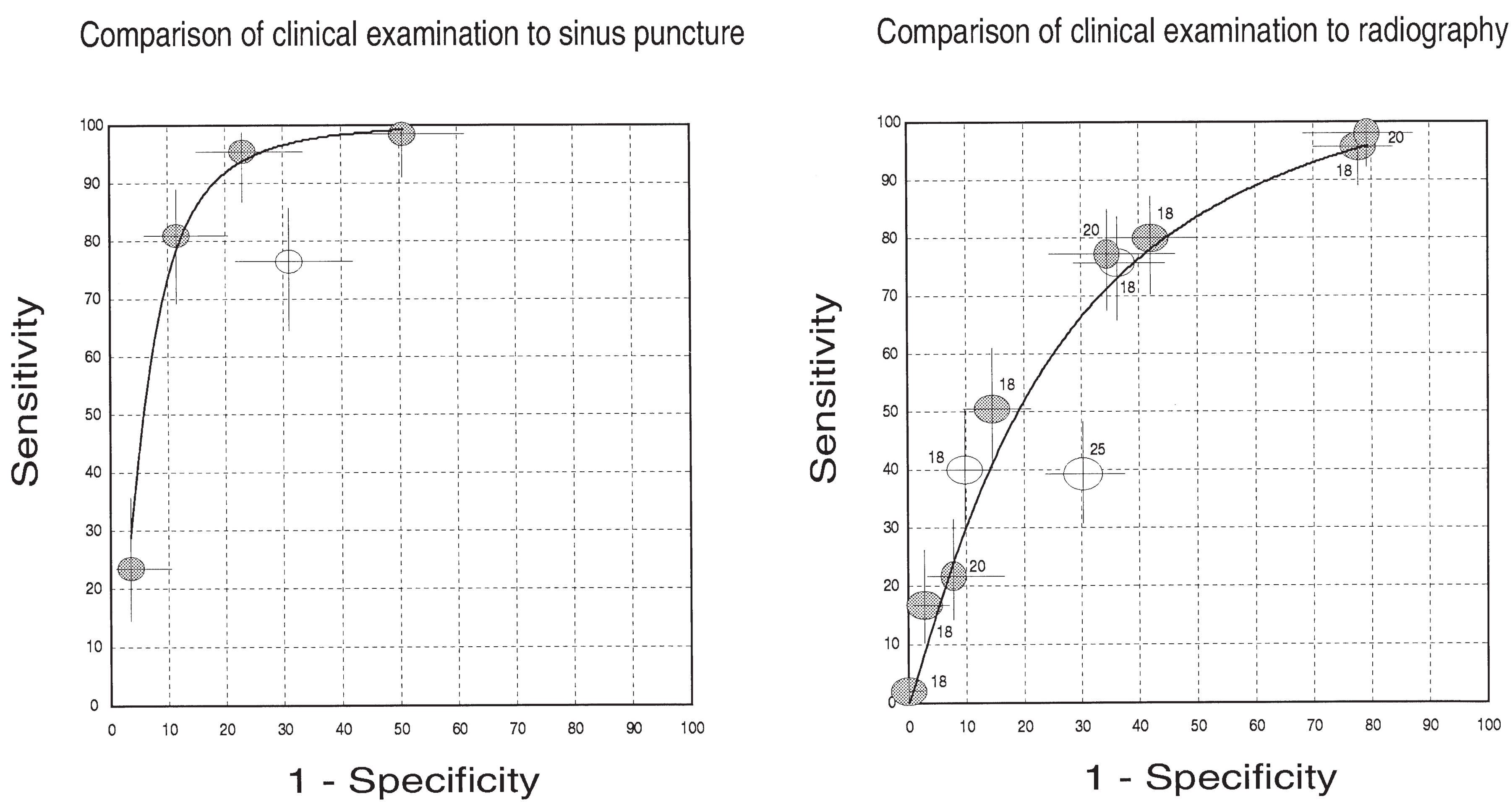

Fig. 2. Summary receiver operator characteristic curves for clinical examination compared with sinus puncture (left panel) or radiography (right panel). Study estimates of sensitivity and specificity are represented by ellipses, as in Fig. 1. In each panel, gray ellipses correspond to estimates for risk scores, which are decision aids in which the likelihood of sinusitis is suggested by the presence or absence of a series of clinical findings. Using these risk scores, various values of sensitivity and specificity are displayed, for each of the several possible cutpoints defined by the number of positive risk score items. White ellipses correspond to estimates based on clinicians' overall clinical impressions. In the left panel, the summary receiver operator curve is fit only to the risk score estimates. As in Fig. 1, numbers next to ellipses identify the study providing data (see Table 1); for the left panel, all data are derived from Berg *et al.* [23].

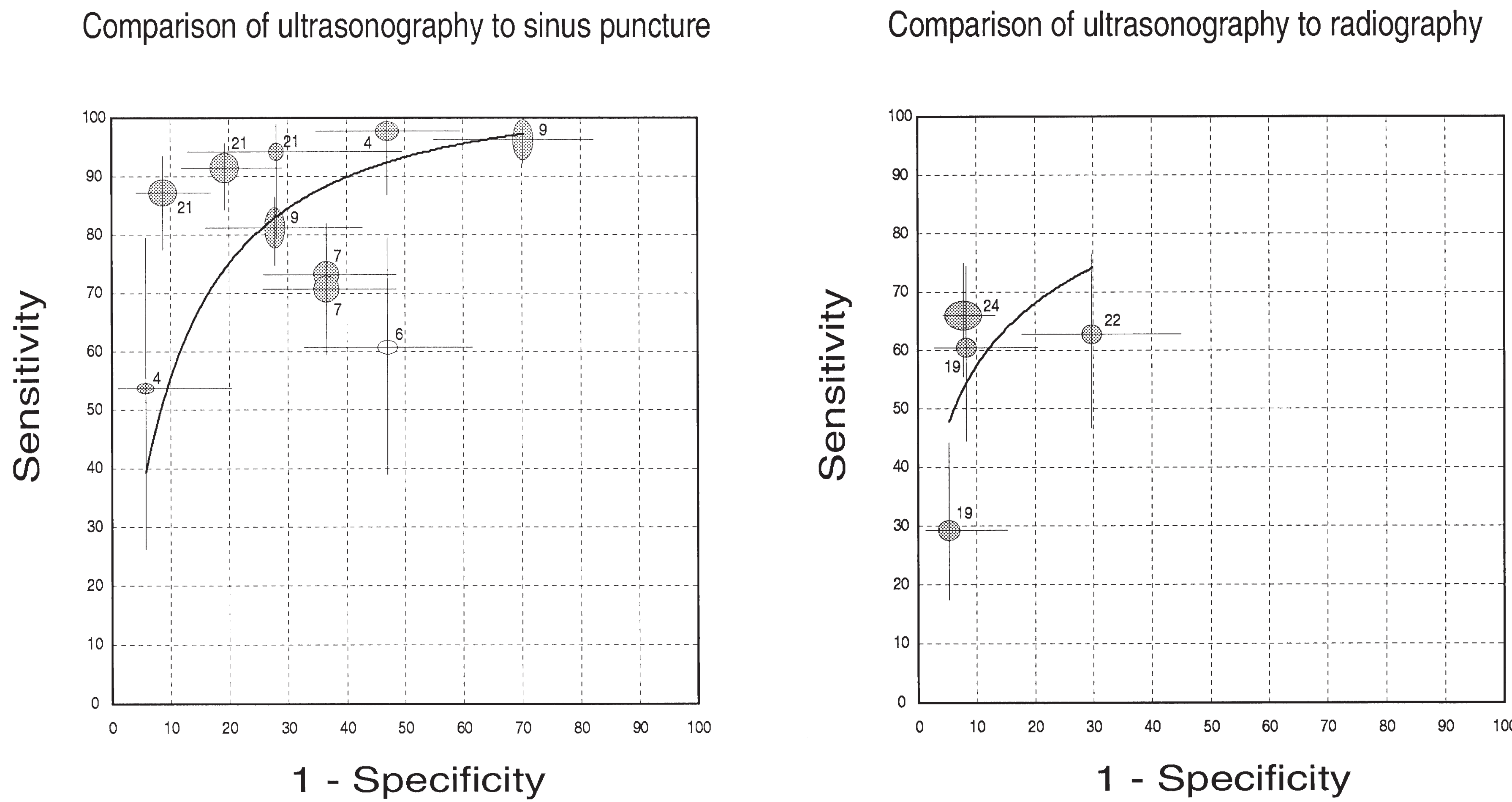

Fig. 3. Summary receiver operator characteristic curves for ultrasonography compared to sinus puncture (left panel) or radiography (right panel). As in Fig. 1, study estimates of sensitivity and specificity are represented by ellipses. All study estimates are shown as gray ellipses, except for that from the study by Laine *et al.* [6] (white ellipse, left panel), in which ultrasonography was performed and interpreted by untrained personnel. As in Fig. 1, numbers next to ellipses identify the study providing data (see Table 1).

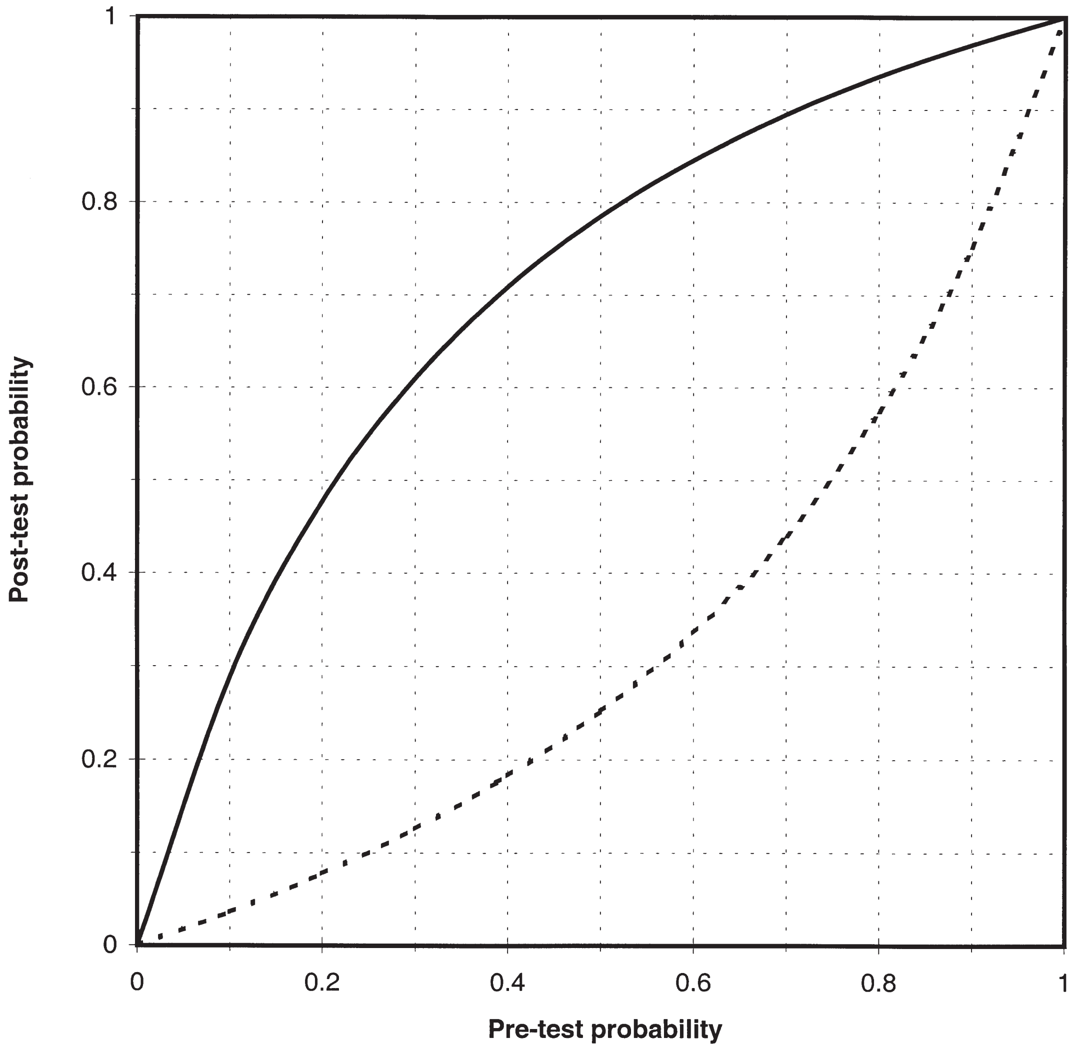

Fig. 4. The effect of sinus radiography results on the probability of acute sinusitis. Likelihood ratios for acute sinusitis, using the diagnostic criterion "sinus fluid or opacity," are 3.65 for positive radiographs and 0.34 for negative radiographs. Application of these likelihood ratios to a range of possible pre-test probabilities for acute sinusitis (corresponding to various levels of sinusitis prevalence; horizontal axis) yields post-test probabilities (vertical axis). Post-test probabilities given a positive radiograph are illustrated by the solid curve, and post-test probabilities given a negative radiograph are shown by the dashed curve.

tive criterion for sinusitis [7,9,19,21,22,24]. Two studies provided additional data on "mucous membrane thickening" as a criterion [9,24]. One study did not describe technique or criteria for sinusitis [4].

Five reports compared sinus ultrasonography with puncture/aspiration [4,6,7,9,21]. Reports provided data on more than one set of patients or for more than one diagnostic cutpoint, so there were 10 values of sensitivity and specificity available for analysis (Fig. 3, left panel). Of note, these points do not appear well-described by the SROC curve, implying that variability in test performance is present. Performance appeared poorest in the study by Laine and colleagues (shown as a white ellipse in Fig. 3, left panel); this was the only study in which untrained primary care physicians performed and interpreted the ultrasounds [6]. Performance was best in the study by Revonta and colleagues [21], with ultrasound simultaneously having good sensitivity and specificity (Fig. 3, left panel).

Three reports [19,22,24] provided data for five comparisons of ultrasonography to sinus radiography (Fig. 3, right panel). It is difficult to interpret these comparisons, because the five data points fall close together in the SROC plot, and it is unclear how well the SROC curve describes them or how the SROC curve can be extrapolated.

4. Discussion

Acute sinusitis is common among patients presenting for evaluation of symptoms consistent with this disorder. The studies that used sinus aspiration as a reference test demonstrated that approximately half of symptomatic patients in a general medical clinic will have acute sinusitis [4,6]. In otolaryngology clinics, where referred patients would more often be expected to have sinusitis, the prevalence of acute sinusitis was as high as 83% [9].

Sinus puncture and aspiration may be the most valid "gold standard" for these patients, because the procedure allows direct examination of sinus contents. Indeed, a previous meta-analysis of diagnostic tests for acute sinusitis by de Bock and colleagues found that sinus puncture was the most accurate means of diagnosing sinusitis [26]. Nonetheless, the pain, need for referral, and cost attendant to sinus puncture preclude its use in routine diagnosis and prompt an examination of how well other tests perform in comparison. In the present study, we therefore summarized data on test performance for sinus radiography, ultrasonography, and the clinical examination. Using different methods than those used in the de Bock meta-analysis [26], we constructed SROC curves that allow examination of the inherent trade-off in sensitivity and specificity for each diagnostic modality.

4.1. Radiography and ultrasonography

Sinus radiography discriminates fairly well between patients who have acute sinusitis, as documented by puncture, and those who do not (Fig. 1, area under SROC curve 0.83). Furthermore, inter-observer agreement is good for four-view sinus radiographs, particularly in the diagnosis of maxillary sinusitis [10]. Based on our results, the radiographic criterion "fluid or opacity" has a likelihood ratio positive of 3.65 and a likelihood ratio negative of 0.34 for acute sinusitis [27]. As shown in Fig. 4, application of these likelihood ratios to estimates of sinusitis prevalence yield post-test probabilities that can be helpful to physicians making treatment decisions. For example, given a prevalence of sinusitis of 50% among symptomatic primary care patients, a radiograph showing sinus fluid or opacity would increase the post-test probability of sinusitis to 78%, while a negative radiograph would decrease it to 25%.

Studies included in this report also provided data for evaluation of the radiographic criterion "sinus fluid or opacity or mucous membrane thickening." Likelihood ratios for this criterion are 2.31 for positive results and 0.16 for negative results. However, our estimate of the specificity of radiography based on this criterion was imprecise (95% confidence interval 0.20–0.91), so these likelihood ratios (and post-test probabilities based on them) are also imprecise. The diagnostic information provided by radiographically documented mucous membrane thickening therefore remains uncertain.

The eight studies that compared sinus ultrasonography with either puncture [4,6,7,9,21] or radiography [19,22,24] offered inconclusive information about how well ultrasonography identifies patients with sinusitis. As illustrated by the five studies comparing ultrasonography with puncture, performance of ultrasonography may vary substantially beyond what is expected simply from different test thresholds (Fig. 3, left panel; see [28]). This added variability may arise through differences in patient populations, ultrasonographic techniques, or medical personnel involved in diagnostic testing, and it calls into question the reliability of ultrasonography. As documented by the poor performance of ultrasonography when performed and interpreted by untrained personnel [6], extensive experience may be necessary before clinicians can appropriately utilize this test. Before ultrasonography could be accepted as a useful and reliable diagnostic tool, further studies comparing it with a reference test will be necessary.

4.2. Clinical examination

Four studies looked at the clinical examination as a diagnostic tool in identifying patients with acute sinusitis, comparing it with sinus puncture [23] or radiography [18,20,25]. The clinical examination, composed of items from the patient history and physical examination, provides a rapid, readily available, and reasonably inexpensive approach to the diagnosis of sinusitis.

Because of methodological problems in the study by Berg and colleagues [23], there are no reliable data for how well clinical examination compares with sinus puncture as a reference test. Furthermore, data comparing clinical examination with radiography cannot be easily interpreted as "true" estimates of sensitivity and specificity, because radiography itself is an imperfect test and not an ideal reference standard. However, based on data from three studies, it is possible to conclude that the clinical examination does offer moderate ability to identify patients who will have a positive radiograph (Fig. 2, right panel, area under SROC curve 0.74). In turn, this implies that some components of the clinical examination may be useful for determining which patients would benefit most from radiography as part of the diagnostic evaluation. For example, it may be reasonable to defer radiography if there is evidence on clinical evaluation that the patient would very likely have a negative radiograph. An approach in which clinical findings dictate which patients undergo sinus radiography may have important cost implications and deserves study.

An important type of clinical examination tool is the risk score. With a risk score, a clinician determines the presence or absence of each of a series of symptoms and signs; the probability of sinusitis increases with each additional positive finding, and the clinician can use this information to help decide if sinusitis is present. Three studies provided data on performance of risk scores, each consisting of three to five separate clinical symptoms and signs. One study suggested that a risk score performs better than a physician's overall clinical impression [23]. On the other hand, two studies (one evaluating diagnosis of sinusitis in children) suggested equivalent discrimination between a clinician's overall impression and a risk score [18,20]. Because a risk score may be simple to apply, and it may depend less than a clinician's overall impression on experience and acumen, further work to develop risk scores that can assist clinicians by adding to or performing better than their overall impressions is warranted. Ideally new risk scores would be carefully described, easily reproducible, and prospectively validated against a reference test.

4.3. Limitations and future work

Some studies of diagnostic tests included in the present meta-analysis were poorly designed or inadequately reported. The most common problems were incompletely described study populations, poorly described test methods and criteria for a positive test (for both reference tests and tests under evaluation), and lack of blinding of investigators to results of one test when they were performing and interpreting another (Table 1). It is unclear how much these limitations of the primary studies affect the conclusions of our meta-analysis. However, study characteristics can affect assessment of diagnostic test performance [15,29]. Future evaluations of diagnostic tests in acute sinusitis should give careful attention to study design and reporting.

The cross-sectional modalities magnetic resonance imaging and, especially, computed tomography offer clear images of the paranasal sinuses, easily discriminating between bone, fluid, and air. Approximately three-quarters of patients clinically diagnosed with acute sinusitis in an emergency room [30] or office practice [31] who undergo computed tomography of the sinuses have abnormalities on the scans. Particularly for frontal, ethmoid, and sphenoid sinus disease, computed tomography may identify abnormalities missed by conventional radiographs [30]. However, computed tomography may not have adequate specificity for use in routine practice, because 87% of individuals with uncomplicated viral upper respiratory tract infections also have abnormal computed tomography findings, such as sinus opacification or air-fluid levels [32]. Because we identified no single study that provided both sensitivity and specificity estimates for these cross-sectional imaging techniques, it was impossible to examine trade-offs between sensitivity and specificity, and to compare these tests with other diagnostic modalities. Ultimately, documentation of the utility of computed tomography and magnetic resonance imaging will require rigorous comparisons of these tests with an even more fundamental diagnostic test, likely sinus puncture.

Because no study used “positive microbiological culture of sinus contents” as the criterion for acute sinusitis, we were unable to determine how well any diagnostic test identifies patients who have bacterial sinusitis and would thus directly benefit from antibiotic treatment. One included study found that one-third of patients with symptoms of acute sinusitis have microbiological evidence of bacterial infection [4], lower than earlier estimates of 40–60% [33,34]. Some patients with clinically or radiologically documented sinusitis will not have a bacterial infection, so providing antibiotics to all patients with a positive test may result in substantial over-treatment. Decision analyses that utilize our results regarding test performance might accurately model the trade-offs between patient symptoms, cost, and the antibiotic resistance attendant with overuse.

5. Conclusions

Better information on the performance of diagnostic tests may allow a more complete understanding of their role in patient management and research. It may be possible to construct a hierarchy of diagnostic tests, allowing the selection of the most appropriate test for each specific purpose. At the top of such a hierarchy would likely remain sinus puncture, the “gold standard” test. However, because of its invasive nature, puncture will likely be suitable only for research purposes. This meta-analysis points out that additional work is needed to define and characterize other tests in the middle of this test hierarchy. Computed tomography may provide images clear enough to serve in this role, perhaps as a reference standard for clinical work. Finally, it is likely that physicians will find little justification for using any imaging modalities, regardless of their accuracy, if the tests are expensive or difficult to obtain. Further attention to perfecting simple diagnostic tools, such as the clinical examination, at the bottom of a test hierarchy, will therefore improve the care of patients presenting with acute sinus symptoms.

Acknowledgments

This work was conducted by the New England Medical Center Evidence-based Practice Center under contract from the Agency for Healthcare Research and Quality (AHRQ) (contract no. 290-97-0019). Additional support was provided by AHRQ grants T32 HS00060 and R25 HS09796. Parts of this manuscript have been included in the evidence report provided to AHRQ. We thank the technical experts (Drs. Michael Benninger, Larry Culpepper, Aidan Long, Anne Meneghetti, Eli Rebeiz, Carol Sakala, Ellen Wald, John Williams Jr., and Ms. Rosalie Phillips) who contributed valuable advice to the evidence report. We also thank our colleagues, Helena Varonen, Marjukka Mäkelä, Seppo Savolainen, and Esa Läärä, from Finland, and Jörgen Hilden from Denmark, for generously sharing with us their meta-analysis of the diagnostic tests for acute sinusitis. We give special thanks to Dr. Varonen for her assistance in extracting data from a Finnish language article.

References

[1] Kennedy DW. First-line management of sinusitis: a national problem? Overview. Otolaryngol Head Neck Surg 1990;103:847–54.
[2] National Center for Health Statistics. National Ambulatory Medical Care Survey, 1990–1995 [CD-ROM]. Series 13.
[3] McCaig LF, Hughes JM. Trends in antimicrobial drug prescribing among office-based physicians in the United States. JAMA 1995; 273:214–9.
[4] van Buchem FL, Knottnerus JA, Schrijnemaekers VJ, Peeters MF. Primary-care-based randomised placebo-controlled trial of antibiotic treatment in acute maxillary sinusitis. Lancet 1997;349:683–7.
[5] Winther B, Gwaltney JM Jr. Therapeutic approach to sinusitis: anti-infectious therapy as the baseline of management. Otolaryngol Head Neck Surg 1990;103:876–8.

[6] Laine K, Maättä T, Varonen H, Mäkelä M. Diagnosing acute maxillary sinusitis in primary care: a comparison of ultrasound, clinical examination and radiography. Rhinology 1998;36:2–6.

[7] Kuusela T, Kurri J, Sirola R. Ultrasound in the diagnosis of paranasal sinusitis of conscripts: comparison of results of ultrasonography, roentgen examination, and puncture and irrigation. Ann Med Milit Fenn 1982;57:138–41.

[8] McNeill RA. Comparison of findings on transillumination, x-ray and lavage of the maxillary sinus. J Laryngol Otol 1962:1009–13.

[9] Savolainen S, Pietola M, Kiukaanniemi H, Lappalainen E, Salminen M, Mikkonen P. An ultrasound device in the diagnosis of acute maxillary sinusitis. Acta Otolaryngol Suppl (Stockh) 1997;529:148–52.

[10] Williams JW Jr, Roberts L, Distell B, Simel DL. Diagnosing sinusitis by X-ray: is a single view adequate. J Gen Intern Med 1992;7:481–5.

[11] Williams JW Jr, Simel DL. Does this patient have sinusitis? Diagnosing acute sinusitis by history and physical examination. JAMA 1993;270:1242–6.

[12] Lindbaek M, Hjortdahl P, Johnsen UL. Randomised, double blind, placebo controlled trial of penicillin V and amoxycillin in treatment of acute sinus infections in adults. BMJ 1996;313:325–9.

[13] Gwaltney JM Jr. Sinusitis. In: Mandell GL, Bennett JE, Dolin R, editors. Principles and Practice of Infectious Diseases. New York: Churchill Livingstone, 1995, pp. 585–90.

[14] U.S. Congress Office of Technology Assessment. Impacts of Antibiotic-Resistant Bacteria (OTA-H-629). Washington, DC: U.S. Government Printing Office, 1995.

[15] Irwig L, Tosteson ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F. Guidelines for meta-analysis evaluating diagnostic tests. Ann Intern Med 1994;120:667–76.

[16] Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Statist Med 1993;12:1293–316.

[17] Laird NM, Mosteller F. Some statistical methods for combining experimental results. Int J Technol Assess Health Care 1990;6:5–30.

[18] Williams JW Jr, Simel DL, Roberts L, Samsa GP. Clinical evaluation for sinusitis. Making the diagnosis by history and physical examination. Ann Intern Med 1992;117:705–12.

[19] Rohr AS, Spector SL, Siegel SC, Katz RM, Rachelefsky GS. Correlation between A-mode ultrasound and radiography in the diagnosis of maxillary sinusitis. J Allergy Clin Immunol 1986;78:58–61.

[20] Jannert M, Andreasson L, Helin I, Pettersson H. Acute sinusitis in children—symptoms, clinical findings and bacteriology related to initial radiologic appearance. Int J Pediatr Otorhinolaryngol 1982;4:139–48.

[21] Revonta M. Ultrasound in the diagnosis of maxillary and frontal sinusitis. Acta Otolaryngol Suppl (Stockh) 1980;370:1–55.

[22] Berg O, Carenfelt C. Etiological diagnosis in sinusitis: ultrasonography as clinical complement. Laryngoscope 1985;95:851–3.

[23] Berg O, Carenfelt C. Analysis of symptoms and clinical signs in the maxillary sinus empyema. Acta Otolaryngol (Stockh) 1988;105:343–9.

[24] Jensen C, von Sydow C. Radiography and ultrasonography in paranasal sinusitis. Acta Radiol 1987;28:31–4.

[25] Axelsson A, Runze U. Symptoms and signs of acute maxillary sinusitis. ORL J Otorhinolaryngol Relat Spec 1976;38:298–308.

[26] de Bock GH, Houwing-Duistermaat JJ, Springer MP, Kievit J, van Houwelingen JC. Sensitivity and specificity of diagnostic tests in acute maxillary sinusitis determined by maximum likelihood in the absence of an external standard. J Clin Epidemiol 1994;47:1343–52.

[27] Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology. A Basic Science for Clinical Medicine 2nd ed. Boston, MA: Little, Brown, 1991.

[28] Littenberg B, Moses LE. Estimating diagnostic accurary from multiple conflicting reports: a new meta-analytic method. Med Decis Making 1993;13:313–21.

[29] Begg CB. Biases in the assessment of diagnostic tests. Statist Med 1987;6:411–23.

[30] Burke TF, Guertler AT, Timmons JH. Comparison of sinus x-rays with computed tomography scans in acute sinusitis. Acad Emerg Med 1994;1:235–9.

[31] Lindbaek M, Johnsen ULH, Kaastad E, Dolvik S, Moll P, Laerum E, Hjortdahl P. CT findings in general practice patients with suspected acute sinusitis. Acta Radiologica 1996;37:708–13.

[32] Gwaltney JM Jr, Phillips CD, Miller RD, Riker DK. Computed tomographic study of the common cold. N Engl J Med 1994;330:25–30.

[33] Hamory BH, Sande MA, Sydnor A Jr, Seale DL, Gwaltney JM Jr. Etiology and antimicrobial therapy of acute maxillary sinusitis. J Infect Dis 1979;139:197–202.

[34] Evans FO Jr, Sydnor JB, Moore WE, Moore GR, Manwaring JL, Brill AH, Jackson RT, Hanna S, Skaar JS, Holdeman LV, Fitz-Hugh S, Sande MA, Gwaltney JM Jr. Sinusitis of the maxillary antrum. N Engl J Med 1975;293:735–9.