Protected specimen brush or bronchoalveolar lavage to diagnose bacterial nosocomial pneumonia in ventilated adults: A meta-analysis

Annik de Jaeger, MD; Catherine Litalien, MD, FRCP; Jacques Lacroix, MD, FRCP; Marie-Claude Guertin, BSc; Claire Infante-Rivard, MD, PhD

From the Pediatric Intensive Care Unit (Drs. de Jaeger, Litalien, and Lacroix and Ms. Guertin), Department of Pediatrics, Sainte-Justine Hospital, Université de Montréal, and the Joint Departments of Epidemiology and Biostatistics (Dr. Infante-Rivard), and Occupational Health, Faculty of Medicine, McGill University, Montréal, Québec, Canada.

CRITICAL CARE MEDICINE 1999;27:2548-2560

[Click here for reference links. (95 references linked.)]

Objective: We conducted a meta-analysis by using summary receiver operating characteristic curves to compare the diagnostic value for bacterial nosocomial pneumonia of the following: a) quantitative culture (colony-forming units per milliliter or CFU/mL) of respiratory secretions collected with a bronchoscopic protected specimen brush (PSB); b) quantitative culture of a bronchoscopic bronchoalveolar lavage (BAL); and c) the percentage of infected cells (IC) in BAL.

Data Sources: All studies published in the English or the French language, through January 1, 1995, on the evaluation of PSB or BAL for the diagnosis of pneumonia were considered for analysis. The relevant literature was identified through computer and reference searching and by experts in the field.

Study Selection: A study was included if at least two of three independent readers regarded its purpose as the evaluation of CFU-PSB, CFU-BAL, or IC-BAL for the diagnosis in human beings of bacterial nosocomial pneumonia in ventilated adults and if the study was prospective and published in a peer-reviewed journal.

Data Extraction: Three readers reviewed all published articles and decided whether to include each study; consensus was defined as agreement by at least two readers. The authors of each original article included in the meta-analysis were asked to complete a questionnaire in which they were asked to check and to correct the data extracted by one of the independent readers.

Data Synthesis: Summary receiver operating characteristic curves were used to compare the efficacy of three diagnostic tests. Eighteen studies on CFU-PSB (795 patients) were included, as well as 11 studies on CFU-BAL (435 patients) and 11 on IC-BAL (766 patients). The accuracy of these tests was not different. However, it seems that administration of previous antibiotics markedly decreased accuracy of CFU-PSB ($p = .0002$) but not the accuracy of CFU-BAL and that of IC-BAL.

Conclusion: Both PSB and BAL are reliable to diagnose bacterial nosocomial pneumonia. Because CFU-BAL and IC-BAL seemed more resistant to the effects of antibiotics, we recommend BAL rather than PSB if the patient is already receiving antibiotics.

Key Words: bronchoalveolar lavage; critical care; cross infection; hospital-acquired infection; intensive care unit; laboratory tests; meta-analysis; nosocomial pneumonia; pneumonia; protected specimen brush.
Bacterial nosocomial pneumonia (BNP) is a significant health problem. Its direct and indirect costs were estimated to be 1.1 billion dollars per year in the United States in 1985 (1). The rate of mortality associated with it is reported to be between 20% and 55% among adults treated in intensive care units (ICUs) (2). Its attributable mortality was estimated to be between 7.1% (2) and 27.1% (3).

There is no consensus on the best diagnostic tool for nosocomial pneumonia (4, 5). Clinical and radiologic features, such as fever, rales, purulent sputum, leukocytosis, or a new radiologic opacity, are clinically helpful to screen for cases of nosocomial pneumonia (6), but their individual validity as diagnostic tools is not considered to be high enough. Even isolation of a bacteria from tracheal secretions is not a reliable diagnostic marker of pneumonia, because colonization of the trachea and contamination of the sample occur frequently. Clinical criteria and clinical scores have been suggested (7, 8); one study (9) suggests that the score of Pugin et al. (8) is valid, but more studies are needed before any conclusion can be drawn. Many authors consider that the best diagnostic criteria are based on quantitative cultures, with results expressed in colony-forming units per milliliters (CFU/mL), of bronchial secretions collected with a protected specimen brush (PSB) (10-56) or of secretions collected by a bronchoalveolar lavage (BAL) (6, 8, 13-15, 17, 22, 23, 25, 31-35, 37, 38, 41-43, 45, 46, 48-50, 55, 57-90). However, the validity of these procedures is still questioned. First, it is debated which technique is the most predictive. Second, PSB and BAL are frequently performed among patients who are already receiving antibiotics, and the effect of antibiotics on the validity of both techniques is unknown. Third, one can ask if the validity of PSB and BAL is really as good as it looks in the literature when the accuracy of these procedures is compared with that reported at the time of autopsy. We undertook a meta-analysis to address these questions.

MATERIALS AND METHODS

Definitions. This meta-analysis estimates the validity of microbiological tests on respiratory secretions collected by bronchoscopic PSB or BAL to diagnose BNP. By definition, PSB is performed with a protected catheter. BAL could be performed with many kinds of tubular devices, but the catheter should be pushed as far as possible into the airways, beyond the main bronchi. The media used to culture the collected respiratory secretions should allow the isolation of aerobic and anaerobic Gram-negative and Gram-positive bacteria. The cultures should be quantitative and the results reported in CFU/mL. We also looked at percentage of infected cells (IC; cells with bacteria inside) in bronchoscopic BAL samples.

We considered adequate for the diagnosis of BNP any reference standard, such as clinical data, radiologic data, blood culture, PSB culture if the test was CFU-BAL or IC-BAL, lung biopsy, and autopsy, inasmuch as it was well defined in the original study. Nevertheless, it was also decided before this meta-analysis was undertaken that one section would include only cases for which an autopsy was the reference standard.

Retrieving the Literature. Studies published between January 1979 and January 1995, on the diagnosis of nosocomial pneumonia with PSB or BAL, were retrieved from the literature according to the following steps. First, a primary search of the English and French literature was performed by computer, with MEDLINE, using the following key words: bronchoalveolar lavage, critical care, cross-infection, diagnosis, hospital-acquired infection, intensive care unit, meta-analysis, nosocomial pneumonia,
pneumonia, protected specimen brush, and prospective studies. Initially, all articles were considered, including reviews, overviews, editorials, monographs, symposia, book chapters, and clinical studies. Second, a manual search was done by scrutinizing the reference lists of the articles obtained during the first step. Third, colleagues and specialists were contacted to identify published studies. Fourth, all authors of the articles retrieved by computer and manual search, as well as other colleagues and specialists, were sent the list of retrieved articles and asked if they knew of any other study or meta-analysis on the same topic. The last process was repeated until no new article was found.

**Choice of Studies.** To be included in this meta-analysis, a study had to fulfill the following criteria: At least two of three independent readers regarded its main purpose as the evaluation of a diagnostic test for the diagnosis of BNP, namely, PSB or BAL, against a concurrent reference standard; also, it should be published as an article or an abstract in a peer-reviewed journal. A study was excluded if any of the criteria listed in Table 1 was met.

| Table 1. Causes of exclusion of 69 studies |

One objective of this meta-analysis was to evaluate change in accuracy of PSB and BAL, regardless of their duration or spectrum, when antibiotics were given in the 24 hrs before the sample was collected. Two strata were considered, one for those who were receiving antibiotics and another for those who were not. A letter was sent to authors of the original articles to ascertain how many patients should be placed in these respective strata. We included in this part of the meta-analysis only those studies for which an answer was received, unless the data available in the original article were clear enough to categorize the patients according to these strata.

**Research Selection.** The decision to include or to exclude an article was made by three readers (AD, CL, and JL) who determined independently if a study met the inclusion criteria. Full copies of all articles available for selection were given to the readers. Reviewers were not blinded as to details of authorship and study results, because "no evidence exists that blinding results in a decrease in bias" (91). Disagreement with regard to inclusion of a study was resolved by conference.

**Assessment of Methodologic Quality.** A set of criteria suggested by Chalmers et al. (92), to evaluate the quality of clinical trial, was adapted to assess the quality of studies evaluating the sensitivity and the specificity of a diagnostic test. We also added a criterion on reproducibility, as suggested by Cook et al (93). Table 2 outlines the criteria of quality evaluated independently by the three readers (AD, CL, and JL).
Table 2. Quality criteria of the 26 included studies

<table>
<thead>
<tr>
<th>Study Protocol</th>
<th>Data Collection</th>
<th>Data Interpretation</th>
<th>Statistical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>Question 2</td>
<td>Question 3</td>
<td>Question 4</td>
</tr>
</tbody>
</table>

The rate of agreement (concordance) between the experts, regarding the questions directed to them, was expressed as a percentage and by using a weighted kappa score (94).

**Data Collection.** Data regarding the sensitivity and the specificity of PSB or BAL were extracted from the studies by one of the authors (JL); they were recalculated, and they were displayed in two-by-two contingency tables (positive or negative test result vs. presence or absence of the disease). Four contingency tables were constructed; i.e., the first included all patients, the second included patients who were not receiving antibiotics when the PSB or the BAL was performed, the third included patients who were receiving at least one antibiotic when the PSB or the BAL was performed, and the fourth included only patients for whom the diagnosis of nosocomial pneumonia was based on an autopsy. Thereafter, the author, whose mailing address was specified in the original article, was contacted to correct or complete information from the four tables (the rate of response was 62%).

**Display of the Data.** Data were recorded by one of the authors (JL) by using a spreadsheet program (Excel 4.0, 1992; Microsoft, Redmond, WA) and were verified at least twice before the analyses were performed.

**Quantitative Meta-Analysis.** Sensitivity and specificity for a given threshold of PSB and BAL, in comparison with the diagnostic criteria for pneumonia, were estimated for each study. In addition, we calculated the "global value" of a given threshold by dividing the sum of cases classified as true positives and true negatives (PSB or BAL vs. reference standard) with the total number of cases. For each specific test (e.g., bronchoscopic nonprotected BAL with quantitative culture), we also calculated an "unweighted pooled global value" (UPGV) by dividing the sum of true positives and true negatives with the total number of cases across all studies; in studies in which many thresholds for PSB or BAL were tested, only the cutoff with the best global value was used. For each UPGV, we also calculated the SD.

The UPGV can be criticized on many grounds; for example, it does not consider the number of cases and the internal variance of the different studies included in the meta-analysis. Moses et al. (95) suggested a new method to estimate the sensitivity and specificity of a test from cumulative evidence of pooled studies, by combining the results into a summary receiver operating characteristic (ROC) curve. The mathematical method is based on the following steps. The false-positive rate (FPR = 1 - specificity) of each study is converted to its logistic transform U, and the true-positive rate (TPR; or sensitivity) of each study is converted to its logistic transform V, after adding half to each cell of the contingency table (this
implies that there will be no zero cells and that there will not be any undefined transformation). For each study, we then calculated $D$ (difference) = $Y$ axis $V - U$, which is the log odds ratio of FPR, and $S$ (Sum) = $X$ axis = $V + U$. Then, each study’s point $(S_i, D_i)$ is plotted, and a regression line is fitted to these points, with $D$ as the dependent variable. Last, this best-fitted line is back-transformed to a summary ROC curve (95).

To determine which is the best test between PSB and BAL, summary ROC curves can be statistically compared by their area under the ROC curve. We believe that this method is inappropriate because a summary ROC curve is a reconstructed curve with estimated extremities. One can calculate a $Q$ value, which represents the intersection point of the summary ROC curve with a diagonal from the left upper corner to the right lower corner of the ROC space. Sensitivity and specificity are equal on this diagonal line. The more the summary ROC curve tends toward the left upper corner of the ROC space, the higher its $Q$ value. The highest $Q$ value represents the test with the best sensitivity and specificity. $Q$ values can be compared by using a statistic that has asymptotically a normal distribution (95).

RESULTS

From 1979 to January 1995, 291 articles dealing with PSB or BAL, for the diagnosis of pneumonia, were identified from the literature search; 95 were selected by the reviewers for closer scrutiny of inclusion and exclusion criteria, and 69 were excluded (Table 1). Overall agreement between the three readers for selection of studies was excellent (agreements of 99%, 98%, and 98%, respectively, and kappa scores of 0.925, 0.845, and 0.895, respectively). After screening, this metaanalysis included 18 prospective clinical studies (795 patients) on the validity of quantitative cultures (CFU/mL) of respiratory secretions collected with a PSB (CFU-PSB) to diagnose nosocomial pneumonia, 11 prospective clinical studies (447 patients) on the validity of quantitative cultures of a BAL (CFU-BAL), and 11 studies (766 patients) evaluating the percentage of infected cells in BAL (IC-BAL). The average age of patients included in the studies ranged from 42 to 63 yrs. All were mechanically ventilated during the previous ≥48 hrs (actually most were severely ill and had been ventilated for many days). Almost all patients were immunocompetent. These patients were considered for inclusion in 21 studies because a BNP was suspected (e.g., new infiltrates on lung x-rays, fever, purulent respiratory secretions, and leukocytosis), because an autopsy could be performed in two studies, or because the patients were considered at risk for a BNP. The ICU was medicosurgical in most instances (23 of 26 studies). Respiratory secretions were collected by fibroscopy in each instance.

Overview of the Methodologic Quality

Table 2 shows the percentage of included studies meeting the chosen quality criteria. The overall mean quality score was 42%, with a maximum score of 64% and a minimum of 19% (Table 3). Agreement between the three readers on the evaluation of the quality of each study was 71%. The kappa scores between the three pairs of readers were 0.49, 0.38, and 0.45, respectively. Table 3 also shows the diagnostic criteria used in the original articles to recognize cases of pneumonia; many articles used multiple reference standards, such as clinical data, radiologic data, blood culture, pleural culture, PSB culture, lung biopsy, or autopsy.
Table 3. Overall methodologic quality of the 26 included studies

Quantitative Meta-Analysis

Global Comparison of Tests. Table 4A, Table 4B, Table 4C and Table 4D shows the global value(s) of each study. The UPGVs were 0.83 ± 0.13 for CFU-PSB, 0.76 ± 0.16 for CFU-BAL, and 0.85 ± 0.08 for IC-BAL.

Table 4. Raw data of the 26 included studies

Table 4. Continued
Table 4. Raw data of the 26 included studies

Table 4. Continued

Figure 1 illustrates the summary ROC curves constructed from studies evaluating the diagnostic value of PSB and BAL. Each point, contributing to the summary ROC curve, represents one study. Some studies reported the sensitivity and specificity at many cutoff points (e.g., $10^3$, $10^4$, and $10^5$ CFU/mL); in such instances, we chose the cutoff point with the best global value. No significant differences were found between CFU-PSB and CFU-BAL or between CFU-BAL and IC-BAL.

**Figure 1.** Global evaluation of protected specimen brush (PSB) and bronchoalveolar lavage (BAL). Summary receiver operating characteristic (ROC) curves for PSB and BAL. Each point contributing to the summary ROC curve represents the best global value of one study. Summary ROC curves were constructed exclusively with studies reporting a high true-positive rate ($TPR, >50\%$) and a low false-positive rate ($FPR, <50\%$) (95); therefore, only the part of each summary ROC curve that lies inside the relevant region of the ROC space, i.e., the upper left quadrant, is illustrated. **Upper panel:** Results of studies on quantitative culture of PSB (CFU-PSB; Q value, 0.8608) and on quantitative culture of BAL (CFU-BAL; Q value, 0.8373); the difference between the Q values is not significant. **Lower panel:** Results of studies on quantitative culture of BAL (CFU-BAL; Q value, 0.8373) and on the percentage of infected cells collected in a BAL (IC-BAL; Q value, 0.8852); the difference between the Q values is not significant.

**Effect of Antibiotics.** The UPGV of CFU-PSB went from 90% in patients who did not receive antibiotics to 73% among those who did and likewise for CFU-BAL (89% to 72%) and IC-BAL (85% to 80%).
In Figure 2, summary ROC curves constructed with the results of patients who received previous antibiotics were compared with summary ROC curves of patients who did not. The use of antibiotics significantly decreased the predictive value of CFU-PSB, but it did not for CFU-BAL and IC-BAL.

**Figure 2.** Effect of antibiotics. Summary receiver operating characteristic (ROC) curves constructed to estimate the effect of previous antibiotics. Only the part of each summary ROC curve that lies inside the relevant region of the ROC space, i.e., true-positive rate (TPR) of >50% and false-positive rate (FPR) of <50%, is drawn. *Upper panel:* Results of studies on quantitative culture of protected specimen brush (PSB) (CFU-PSB), subdivided into two strata; i.e., in the first, patients who received antibiotics before the PSB (Q value, 0.6612), and in the second, patients who were not receiving antibiotics when the PSB was performed (Q value, 0.9120). The difference between the Q values is statistically significant (p = .0002). *Middle panel:* Results of studies on quantitative culture of bronchoalveolar lavage (BAL) (CFU-BAL); there is no statistical difference (p = .3928) between the Q values of those on previous antibiotics (Q value, 0.8388) and those who were not receiving antibiotics (Q value, 0.8870). *Lower panel:* Results of studies on the percentage of infected cells collected by a BAL (IC-BAL); there is no statistical difference (p = .8202) between the Q values of those on previous antibiotics (Q value, 0.8273) and those who were not receiving antibiotics (Q value, 0.8068).

**Autopsy as Criterion.** Part D of Table 4A, Table 4B, Table 4C and Table 4D shows that we have data on cases with autopsy in nine studies on CFU-PSB, in six on CFU-BAL, and in three on IC-BAL. Moses et al. (95) suggested to exclude studies that report a TPR of <50% or an FPR of >50%; using these criteria, eight studies on CFU-PSB were retained that included 136 patients (Q value, 0.96), five studies on CFU-BAL including 44 patients (Q value, 0.79), and three studies on IC-BAL including 39 patients (Q value not estimated). The Q value of two tests can be compared statistically only if the studies or the patients within the studies are independent; therefore, Q values were not compared statistically, because the only data still available for analysis were five studies on CFU-PSB, two on CFU-BAL, and none on IC-BAL.

**DISCUSSION**

This meta-analysis addressed the following four questions: a) What was the quality of the studies evaluating CFU-PSB, CFU-BAL, and IC-BAL as diagnostic tools for BNP? b) Which test seemed more valid to diagnose BNP in mechanically ventilated adults, CFU-PSB, CFU-BAL, or IC-BAL? c) What was the effect of antibiotics on the validity of PSB and BAL? and d) Were there enough autopsies reported in the literature to use autopsy as the criterion for PSB and BAL?

**Quality of the Studies.** By using explicit criteria, an overall mean quality score of 42% was estimated for the studies included in this meta-analysis (Table 2). Weaknesses were noted in the protocol design, mostly with respect to the reporting and handling of withdrawals (eight points), the use of another reference standard than autopsy (ten points), the blinding procedures (11 points), and assessment of reproducibility of PSB or BAL (four points). Nevertheless, we consider that the quality of most studies included in the meta-analysis was good, because the study protocols were adequate and the statistical
analyses were appropriate.

Which Is the Best Test? In a previous meta-analysis, Cook et al. (93) reported that CFU-PSB and CFU-BAL are both reliable tests to diagnose nosocomial pneumonia. However, the analytical method used did not consider jointly the sensitivity and the specificity of PSB and BAL; it was the best method available at that time, but such analysis can underestimate test accuracy (91). The model used in the present meta-analysis simultaneously considers sensitivity and specificity by comparing the summary ROC curve of each test, thus allowing a better comparison. The results suggest that CFU-BAL and IC-BAL seem as valid as CFU-PSB to diagnose nosocomial pneumonia.

Baker et al. (96) published a review about decision-making in BNP. They studied the validity of the bacterial index that was suggested as a diagnostic marker of nosocomial pneumonia by Johanson et al (71). This index is calculated by adding the log of the concentration in CFU of all organisms isolated from a sample of respiratory secretions. The bacterial index of a sample in which two germs are isolated in concentrations of $10^2$ and $10^3$ CFU/mL will be the same as the index of a sample with one germ at $10^5$ CFU/mL; the bacterial index is 5 in both instances, but there are 1,100 CFU/mL in the former, whereas there are 100,000 CFU/mL in the latter. Therefore, we agree with Marquette et al. (97), who wrote that this index is mathematically incorrect, and this may diminish our confidence with this criterion as a diagnostic marker of BNP.

Baker et al. (96) also looked at the discriminative power of CFU-PSB and CFU-BAL, and they found that it was quite similar; but they did not address the question of the effect of previous antibiotics on the accuracy of these tests.

Effect of Antibiotics. The administration of antibiotics has been shown experimentally to markedly decrease the lobar bacterial burden of lungs with bronchopneumonia and even to increase the number of sterile lungs (50, 65, 71). Nevertheless, it is still considered that the effects of previous antibiotherapy on quantitative culture of respiratory secretions remain unknown in human beings (98). Torres et al. (99) and Souweine et al. (100) reported that previous administration of antibiotics decreased the positive and negative predictive values of CFU-PSB, CFU-BAL, and IC-BAL, but the number of cases was too small to make a statistical comparison between these tests. We found that receiving previous antibiotics decreased the validity of CFU-PSB, CFU-BAL, and IC-BAL; however, by using summary ROC curves, we found that the effect of previous antibiotics on the validity of CFU-PSB was statistically significant, but it was not for CFU-BAL and IC-BAL. That previous use of antibiotics decreases the validity that CFU-PSB could be explained in the following way: PSB brushes only a small part of infected bronchus, but BAL, wedged or not, washes thousands of bronchioli and alveoli; the effect of previous use of antibiotics should be more important if only a small part of the lung is sampled, and if only few bacteria are collected. Another possibility could be that studies on BAL simply looked at patients who had been taking antibiotics for a long period of time, but studies on PSB looked at patients who had been taking antibiotics for a short period of time. Because the study by Souweine et al. (100) was not available when we were planning the meta-analysis, we did not ask the authors of the original articles how long the previous antibiotics were taken by the patients they studied. We asked them, however, if antibiotics were given at least in the last 24 hrs before the sample was collected, and two strata were considered, one for those who were receiving antibiotics and another for those who were not. Moreover, PSB and BAL were performed during the same bronchoscopy in nine (13, 14, 17, 24, 42, 48, 50, 101, 102) of the 16 studies on PSB that were retained in the meta-analysis. Our results clearly demonstrate that CFU-BAL and
IC-BAL are more resistant to the effects of antibiotics than is CFU-PSB. Souweine et al. (100) suggested that the validity of CFU-PSB in patients receiving antibiotics can be improved by lowering the cutoff point (e.g., $10^2$ rather than $10^3$ CFU/mL for PSB and $10^3$ rather than $10^4$ CFU/mL for BAL), but more data are needed to implement this recommendation.

Another way to obviate the problem of previous antibiotic use would be to interrupt its usage before PSB or BAL is performed. It is true that the length of time without antibiotics modulates the number of organisms in quantitative cultures of PSB (74). Stopping antibiotics 24 or 48 hrs before PSB or BAL has been advocated, but the validity of such recommendation is not based on hard data. Actually, the time lag needed to neutralize the effects of previous antibiotics on the validity of CFU-PSB, CFU-BAL, and IC-BAL is still unknown, but some data suggest that it may be $\sim 24$ hrs (100).

Recently, the effect of previous antibiotic therapy on the validity of CFU-PSB, CFU-BAL, and IC-BAL was assessed by Timsit et al (46). They found no significant differences for these three tests. This study is not included in our meta-analysis because it was published after January 1995, when we began to collect data. Some studies (12, 14, 22, 81) included in this meta-analysis reported that previous antibiotics did not change PSB or BAL results; however, other studies reported a decreased accuracy for CFU-PSB (30), for CFU-BAL (102), and for IC-BAL (42, 65, 81). This meta-analysis reports that CFU-PSB is less resistant to previous antibiotics than CFU-BAL and IC-BAL, but it does not explain this finding. We believe that our conclusion with respect to CFU-PSB holds true. However, only further studies will allow estimation of the effects of recent vs. current antibiotics on CFU-PSB, CFU-BAL, and IC-BAL and verification of whether CFU-BAL and IC-BAL are more resistant to previous antibiotics because a BAL washes a larger area of the lungs and the airways than a PSB.

We calculated UPGVs and summary ROC curves; it was noteworthy that the results were not exactly the same for the effect of antibiotics. UPGVs suggested that antibiotics decreased the diagnostic value of CFU-PSB (from 90% to 73%), CFU-BAL (from 89% to 72%), and IC-BAL (from 85% to 80%). Summary ROC curves suggested that antibiotics significantly decreased the validity of CFU-PSB (Q value went from 0.9120 to 0.6612) but not that of CFU-BAL (Q values of 0.8870 and 0.8388) or that of IC-BAL (Q values of 0.8068 and 0.8273). The mathematical model suggested by Moses et al. (95) should be better than the UPGV to compare the validity of two tests, because a ROC curve is the best way to capture the trade-off between the sensitivity and the specificity and because a summary ROC curve considers the results of each individual study.

**Autopsy as Reference Standard.** Autopsy is still considered the criterion for the diagnosis of nosocomial pneumonia by many authors (37, 50, 82, 83, 103, 104). In this meta-analysis, we considered it inappropriate to statistically compare Q values of studies in which nosocomial pneumonia was diagnosed by an autopsy, because it was impossible to consider the effect of antibiotics in this group of patients and because only two studies on CFU-BAL and none on IC-BAL could be retained for such analysis. Some articles in which autopsy was used as the criterion have been published since January 1995, when we began this meta-analysis. Chastre et al. (105) studied many lung specimens per patient in 20 patients; the sensitivity and the specificity of CFU-PSB (cutoff, $\geq 10^3$ CFU/mL) were 82% and 89%, respectively, 91% and 78% for CFU-BAL ($\geq 10^4$ CFU/mL), and 91% and 89% for IC-BAL ($\geq 5\%$). Marquette et al. (97) studied 28 patients, including 19 cases of pneumonia; the sensitivity and the specificity of CFU-PSB ($\geq 10^3$ CFU/mL) were 55% and 88%, respectively, 47% and 100% for CFU-BAL ($\geq 10^4$ CFU/mL), and
37% and 100% for IC-BAL, regardless of the percentage of intracellular bacteria; the global value of these tests was quite similar (68%, 64%, and 57%, respectively). Papazian et al. (9) studied 38 patients, including 18 cases of pneumonia; the sensitivity and the specificity were 33% and 95%, respectively, for CFU-PSB (≥10³ CFU/mL) and 61% and 70% for CFU-BAL (≥10³ CFU/mL); the global value of these tests was exactly the same (66%), but the area under the ROC curve of CFU-BAL was greater than the area of CFU-PSB (this was not analyzed statistically). The results of the studies by Marquette et al. (97) and by Papazian et al. (9), in which data are reported per patient rather than per lung specimen, support the results of our meta-analysis to the effect that CFU-BAL and IC-BAL are at least as reliable as CFU-PSB, if not better.

One could question the view that autopsy represents the ideal criterion. The appropriateness of studying a diagnostic test only among dead patients is debated for different reasons (93). For example, nosocomial pneumonia detected in dead patients is probably different from that in patients who are alive, the former being more severe than the latter. In addition, a dying patient may develop a BNP after first being suspected of having it and after undergoing bronchoscopy; in addition, the results of the post mortem tissue examination may be unreliable if the time between death and autopsy is long. Considering the limitations mentioned above, one can conclude that autopsy cannot be considered a perfect reference standard for the diagnosis of BNP.

Lung biopsy was considered a potential criterion for a while (12), but new data suggest that the rate of false results of lung biopsy is 25%, compared with systematic autopsy with multiple tissue specimens collected all over both lungs (82, 83, 97). This lack of reliability of lung biopsy is probably related to the fact that the infectious process of nosocomial pneumonia is focally distributed, although it is widespread in the lungs (83). Therefore, as Cook et al. (93), we conclude that there is no perfect reference standard for nosocomial pneumonia.

Validity of This Meta-Analysis. As Sacks et al. (106) wrote, a major issue in pooling data is whether the results of the separate trials can be meaningfully combined. In the studies included in this meta-analysis, the patients, the maneuvers (PSB and BAL), and the tests (quantitative cultures and intracellular bacteria) were quite similar. However, many original articles used multiple reference standards (clinical, radiologic, microbiological, and pathologic; Table 3); one may question the validity of these criteria and that they were multiple. The validity of the diagnosis of BNP should be better if the criteria are applied a posteriori, considering all data available during the ICU stay, rather than a priori when the diagnosis of BNP was suspected. In the studies included in this meta-analysis, when the diagnosis of BNP was based on clinical and laboratory criteria, it was ascertained a posteriori; this approach, i.e., an a posteriori rather than an a priori decision for the establishment of a definite diagnosis, strengthens the validity of these studies and of this meta-analysis. Also a new test, or new criterion, can be validated by verifying whether it agrees more closely than another test with a consensus value derived from other reference standards for the disease in question (107). This method works well if the number of reference standards contributing to the consensus is large, which was the case in the original articles included in the meta-analysis. We considered appropriate, although not ideal, diagnosis of BNP established a posteriori and based on combined diagnostic criteria. Of course, the absence of a perfect reference standard for the diagnosis of BNP is a limitation of this meta-analysis, because the reference standards were multiple and the combination of tests not necessarily uniform. However, it is well recognized that most physicians base their diagnosis of BNP on clinical criteria and simple noninvasive methods, such as blood or pleural
culture (108). Our analysis only reflects current practice and should be considered as valid as the articles it analyzed. Eventually, with more studies available using uniform reference criteria, analysis in homogeneous subgroups with respect to diagnosis could be performed.

In this meta-analysis, studies that reported a TPR of <50% or an FPR of >50% were not considered for the summary ROC curve construction. One can ask whether this was appropriate. Moses et al. (95) suggested to exclude values of TPR so low and values of FPR so large that the test would not be used; the rationale was to avoid "model-dependent extrapolation from irrelevant regions of ROC space." Accordingly, we excluded studies with such extreme results.

Heterogeneity between the analyzed studies may be another concern. For example, the tests, PSB and BAL, were not performed exactly the same way in the different studies, and the reference standard to diagnose a nosocomial pneumonia was not exactly the same. Nevertheless, a significant heterogeneity of patient population between studies is unlikely because the studies were quite similar. For example, all studies were prospective, their internal validity was good, cases of community-acquired pneumonia were excluded, all patients were intubated adults, and all PSB and BAL were bronchoscopic; moreover, we excluded all studies that reported a TPR of <50% or an FPR of >50%, and this may be another way to decrease the risk of overdispersion.

The potential for publication bias deserves a comment. Including only published studies might enhance chances of a statistically significant effect, because trials with negative results are less likely to be published. There are two solutions to this problem: the first consists in searching for all the unpublished studies and including them in the meta-analysis; and the second is to estimate the publication bias. We looked for unpublished results, but only one thesis was found. Methods have been developed to deal with publication bias in clinical trials (109, 110), but their applicability to diagnostic test assessment has not been explored (91). Also, it should be emphasized that randomized clinical trials and the study of a diagnostic test are two different kinds of studies. The literature shows that authors are less willing to publish negative randomized studies, but there is no such thing as a negative accuracy for a test; i.e., a test is more or less accurate, but its accuracy cannot be negative. Considering all these points, the publication bias was not addressed in this meta-analysis. Moreover, it appears to us that the summary ROC curves of both PSB and BAL are so good that the reliability of these tests would probably remain fair, even if a publication bias existed.

This meta-analysis does not address the question of the best threshold of the tests studied. The cutoff chosen varied frequently in the studies that we found in the literature, and many studies presented results according to several thresholds. Therefore, we were confronted with choices with respect to the test thresholds and decided to use the threshold with the best global value in each study for three reasons. First, the threshold of a given test can change from one study to another, because there are shifts in the patient populations and variability between sites, observers, and techniques (95); choosing the threshold with the best global value should decrease the effects of these caveats. Second, the model suggested by Moses et al. (95), to compare the summary ROC curve of different tests, can be used, although the test threshold defining a positive outcome differs from study to study. Third, it seemed to us that the best way to make an objective decision was to choose the threshold that gave the best global value in each study, an approach that did not leave room for subjective decisions with respect to the choice of threshold; so, instead of comparing studies according to the same arbitrary threshold, we compared the threshold with the best global value of each study. Nevertheless, a clinician may want to know the best threshold of each
The answer is probably that the best threshold changes from one hospital to another and from one patient population to another. However, Table 4A, Table 4B, Table 4C and Table 4D shows that the variation in threshold with the best global value was quite small for PSB and for BAL (10^3 or 10^4 CFU/mL for both tests).

**CONCLUSIONS AND RECOMMENDATION**

Presently available prospective studies allow us to conclude that the accuracy of CFU-PSB, CFU-BAL, and IC-BAL is not significantly different. However, it seems that previous use of antibiotics markedly decreases the accuracy of CFU-PSB ($p = .0002$), whereas it does not for CFU-BAL and IC-BAL; more studies are needed to confirm the latter finding and to explain why CFU-BAL and IC-BAL are more resistant to previous antibiotics than CFU-PSB.

Considering the results of this meta-analysis, which reflects the data available in the literature, we recommend BAL over PSB. With BAL, one can estimate IC-BAL as well as perform a culture of lower respiratory tract secretions. Immediate diagnosis of nosocomial pneumonia is possible with the IC-BAL, and culture allows the identification of the causative germ(s) in the following days.

The results of this meta-analysis and data available in the literature suggest that CFU-PSB, CFU-BAL, and IC-BAL may be valid tests to diagnose nosocomial pneumonia in ventilated adults. However, it did not address the question of their usefulness, which is still a matter of debate (15, 35). Bronchoscopic PSB and BAL are costly, and they can cause significant complications. Moreover, they would be useful only if the additional information obtained from these techniques increased significantly the previous probability of BNP, based on the usual clinical, radiographic, and commonly obtained laboratory data. The decision to perform bronchoscopic PSB or BAL should consider both their risks, such as pneumothorax or bleeding, and benefits, such as a decrease in mortality rate or in the length of stay in the ICU. Only an outcome analysis could clarify whether PSB and BAL are useful in the clinical setting.

**REFERENCES**


15. Chastre J, Fagon JY: Invasive diagnostic testing should be routinely used to manage ventilated patients with suspected pneumonia. *Am J Respir Crit Care Med* 1994; 150:570-574


Supported, in part, by the Fonds de la Recherche en Santé du Québec and the Fondation Sainte-Justine.


Address requests for reprints to: Jacques Lacroix, MD, FRCP, Sainte-Justine Hospital, 3175 Côte Sainte-Catherine, Montreal, Quebec, Canada H3T 1C5. E-mail: sip@point-net.com

Crit Care Med 1999 November;27(11):2548-2560

Copyright © 1999 Lippincott Williams & Wilkins

All rights reserved
Use the pop-up window to navigate search results and find highlighted terms in this document.

☐ Hide SearchPLUS
☐ Disable Highlights