Is the Gram Stain Useful in the Microbiologic Diagnosis of VAP? A Meta-analysis

John C. O’Horo,1 Deb Thompson,2 and Nasia Safdar3,4

1Department of Graduate Medical Education, Aurora Healthcare Metro, 2Infectious Disease Associates, Milwaukee; 3William S. Middleton Memorial Veterans Hospital, and 4Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison

In a meta-analysis examining respiratory specimen Gram stain for diagnosis of ventilator-associated pneumonia, absence of bacteria on Gram stain had a high negative predictive value, but a positive Gram stain correlated poorly with organisms recovered in culture.

Rapid and accurate diagnosis of ventilator-associated pneumonia (VAP) is a major challenge and no generally accepted gold standard exists for VAP diagnosis. We conducted a meta-analysis to examine the role of respiratory specimen Gram stain to diagnose VAP, and the correlation with final culture results. In 21 studies, pooled sensitivity of Gram stain for VAP was 0.79 (95% confidence interval [CI], 0.77–0.81; \( P < .0001 \)) and specificity was 0.75 (95% CI, 0.73–0.78; \( P < .0001 \)). Negative predictive value of Gram stain for a VAP prevalence of 20%–30% was 91%, suggesting that VAP is unlikely with a negative Gram stain but the positive predictive value of Gram stain was only 40%. Pooled kappa was 0.42 for gram-positive organisms and 0.34 for gram-negative organisms, suggesting fair concordance between organisms on Gram stain and recovery by culture. Therefore, a positive Gram stain should not be used to narrow anti-infective therapy until culture results become available.

INTRODUCTION

Ventilator-associated pneumonia (VAP) is a common serious intensive care unit (ICU) healthcare-associated infection [1]. In 2002, 250,205 cases of VAP were reported, with 35,967 associated deaths [2] and an attributable excess cost of $23,000 [3] to $39,000 [4] per infection. Appropriate early antibiotic therapy has been demonstrated to improve outcomes [5], but diagnosis of VAP is a major challenge as the diagnosis centers around a constellation of nonspecific signs/symptoms and radiographic imaging, often but not always coupled with Gram stain and culture of a respiratory specimen. No generally accepted gold standard exists for VAP diagnosis.

The microbiologic diagnosis of VAP requires a respiratory specimen. These include an endotracheal aspirate, or a quantitative or semiquantitative lower-respiratory specimen. Commonly used methods for the lower-respiratory tract specimen collection include the bronchoalveolar lavage (BAL), where the airway is visualized bronchoscopically and a sample is collected; the mini-BAL, where a sample is obtained via a blindly advanced catheter; the protected specimen brush (PSB), which brushes the lower airways through a bronchoscope channel; and the plugged telescoping catheter (PTC), a protected channel device for obtaining lower-respiratory-tract washings.

Invasive lower-respiratory-tract sampling has been found to decrease antibiotic use [6–8] when compared with qualitative endotracheal aspirates. Although data is limited, the available studies have not found 1 site or method of specimen collection to be superior in terms of influencing clinical outcomes [8,9] and thus either a qualitative endotracheal aspirate or a semiquantitative/quantitative lower-respiratory-tract specimen is considered acceptable for diagnosis of VAP [10].

While cultures of endotracheal aspirates or lower-respiratory-tract specimens are useful to confirm the
diagnosis of VAP and to tailor antibiotic therapy, the results are not available until 48–72 hours after collection of specimens. Appropriate initial antibiotic therapy is essential to optimize clinical outcomes [11], and judicious use of antibiotic is important to prevent complications of broad-spectrum antibiotic therapy [12]. Gram staining of respiratory specimens can provide rapid information, but the predictive value of this procedure for VAP and the concordance of Gram stain with subsequent culture results are unclear.

We undertook a meta-analysis to assess the utility of respiratory specimen Gram stain for diagnosis of VAP. Our primary objective was to determine the predictive value of respiratory specimen Gram stain for VAP in patients with signs and symptoms suggestive of VAP. Our secondary objective was to evaluate if the Gram stain can be used to guide empiric anti-infective therapy for a presumed VAP.

**METHODS**

From October 2011 to February 2012, a computerized search of PubMed (including MEDLINE) was performed using the combinations of the terms “ventilator associated pneumonia” or “VAP,” and “Gram stain,” “microscopic examination,” “endotracheal aspirate,” “BAL,” and “bronchoalveolar lavage.” Our search strategy and analysis complied with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [13].

Retrospective trials, randomized controlled trials, and prospective cohort studies were included. Studies using a multi-step algorithm to identify VAP where Gram stain was not the first step in the algorithm were excluded. Case reports, abstract-only, review articles, letters, and editorials were excluded. Pediatric studies were also excluded from the final analysis, as too few were found to allow a separate analysis.

All authors independently reviewed each report identified by the search strategy. Disagreements among reviewers regarding values or analysis assignments were resolved by discussion.

Data were extracted using a standardized form for each study reporting the location of the study, population of interest, inclusion and exclusion criteria, reference standard, method of obtaining sample for Gram stains, and correlation between Gram stain and reference standards. When numbers of true positive, true negatives, false positives, and false negatives were not included in a study, these were calculated using sensitivity, specificity, and total number of subjects.

We assessed study quality and assessed for potential bias using the recommendations described in *The Cochrane Handbook for Systematic Reviews of Interventions* [14].

Sensitivity and specificity for Gram stain were calculated from the data in each study. Pooled sensitivity and specificity for Gram stain were obtained with the use of the DerSimonian-Laird random effects model [15]. Sensitivity and specificity were also combined using the diagnostic odds ratio (DOR), a method proposed by Moses et al. [16], and plotted as a logit function on a summary receiver operating characteristic curve. This statistic can be interpreted as the ratio of the odds that a person who has VAP has a positive Gram stain to the odds that a person who does not have VAP has a positive Gram stain [16]. Confidence intervals of 95% were calculated for each statistic based on the F distribution method for the binomial proportion.

The correlation between Gram stain and subsequent culture results was measured using Cohen’s Kappa statistic (κ) [17]. κ is a robust tool for measuring correlation between observations, taking into account the variation due to chance. κ is defined as $\kappa = (p - p_e)/(1 - p_e)$, where $p$ is the observed proportion of agreement, and $p_e$ is the expected variation calculated as by a $\chi^2$ analysis. A $\kappa$ of <0.20 shows poor agreement, 0.21–0.30 fair, 0.31–0.60 moderate, 0.61–0.80 good, and 0.81–1.00 very good agreement [18].

An assessment of heterogeneity was performed with the use of $I^2$ analysis, where 0% indicates low heterogeneity and 100%
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Subjects (N)</th>
<th>Population</th>
<th>Study Design</th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
<th>Method of Sample Collection</th>
<th>Reference Standard</th>
<th>Antibiotic Use</th>
<th>Prevalence, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marquette et al. 1994</td>
<td>France</td>
<td>72</td>
<td>Medical and surgical ICUs</td>
<td>Prospective cohort study</td>
<td>Clinically suspected pneumonia</td>
<td>None specified</td>
<td>BAL</td>
<td>Quantitative culture, showing &gt;1000 CFU/mL</td>
<td>41 received antibiotics in preceding</td>
<td>29.2</td>
</tr>
<tr>
<td>Sole-Violan et al. 1994</td>
<td>France</td>
<td>33</td>
<td>Medical ICU</td>
<td>Prospective cohort study</td>
<td>Clinically suspected pneumonia</td>
<td>None specified</td>
<td>BAL</td>
<td>Quantitative culture, showing &gt;1000 CFU/mL</td>
<td>Not specified</td>
<td>48.5</td>
</tr>
<tr>
<td>Allaouchiche et al. 1996</td>
<td>France</td>
<td>163</td>
<td>Medical-Surgical ICU</td>
<td>Prospective cohort study</td>
<td>Clinically suspected pneumonia</td>
<td>None specified</td>
<td>BAL</td>
<td>Quantitative culture, showing &gt;1000 CFU/mL</td>
<td>65/163 received antibiotics in previous 72 h</td>
<td>47.2</td>
</tr>
<tr>
<td>Torres et al. 1996</td>
<td>Spain</td>
<td>25</td>
<td>Adult respiratory ICU patients</td>
<td>Prospective cohort study</td>
<td>Patient died on mechanical ventilation</td>
<td>Immunocompromise or hematologic malignancy</td>
<td>BAL, protected BAL</td>
<td>Histologic diagnosis from bronchoscopic lung biopsy</td>
<td>17/25 subjects received antibiotics in previous 48 h</td>
<td>53.2</td>
</tr>
<tr>
<td>Papazian et al 1997</td>
<td>France</td>
<td>28</td>
<td>Adult medical ICU patients</td>
<td>Prospective cohort study</td>
<td>Patient died on mechanical ventilation</td>
<td>Immunocompromise, intubation &lt;72 h</td>
<td>BAL, mini-BAL, and blind bronchoalveolar sampling</td>
<td>Histologic examination from postmortem pneumonectomy</td>
<td>11/28 received antibiotics in 48 h preceding specimen collection</td>
<td>46.4</td>
</tr>
<tr>
<td>Prekates et al. 1998</td>
<td>Greece</td>
<td>75</td>
<td>Adult ICU patients</td>
<td>Prospective cohort study</td>
<td>Clinically suspected pneumonia</td>
<td>None specified</td>
<td>BAL</td>
<td>Clinical diagnosis of VAP</td>
<td>All received antibiotics in preceding 48 h</td>
<td>29.3</td>
</tr>
<tr>
<td>Allaouchiche et al. 1999</td>
<td>France</td>
<td>118</td>
<td>Adult ICU patients</td>
<td>Prospective cohort study</td>
<td>Clinically suspected pneumonia</td>
<td>None specified</td>
<td>BAL</td>
<td>Quantitative culture, showing &gt;1000 CFU/mL</td>
<td>61/118 subjects received antibiotics in previous 72 h</td>
<td>43.2</td>
</tr>
<tr>
<td>Duflo et al. 2001</td>
<td>France</td>
<td>116</td>
<td>Adult ICU patients</td>
<td>Prospective cohort study</td>
<td>Clinically suspected VAP</td>
<td>None specified</td>
<td>BAL</td>
<td>Quantitative culture</td>
<td>Not specified</td>
<td>44.0%</td>
</tr>
<tr>
<td>Timis et al. 2001</td>
<td>France</td>
<td>94</td>
<td>Adult ICU patients</td>
<td>Prospective cohort study</td>
<td>Clinically suspected VAP</td>
<td>Contraindication to BAL; no final diagnosis</td>
<td>BAL</td>
<td>Quantitative culture</td>
<td>Not specified</td>
<td>50.0</td>
</tr>
<tr>
<td>Aucar et al. 2003</td>
<td>United States</td>
<td>15</td>
<td>Surgical ICU patients</td>
<td>Prospective crossover case-controlled study</td>
<td>Suspected VAP</td>
<td>None specified</td>
<td>EA</td>
<td>Quantitative culture, showing &gt;1000 CFU/mL</td>
<td>10/15 subjects received antibiotics for longer than 72 h</td>
<td>44.4</td>
</tr>
<tr>
<td>Brasel et al. 2003</td>
<td>United States</td>
<td>35</td>
<td>Surgical ICU patients</td>
<td>Prospective cohort study</td>
<td>All ventilated patients</td>
<td>None specified</td>
<td>EA</td>
<td>Quantitative culture, showing&gt;10^6 CFU/mL</td>
<td>25/35 patients</td>
<td>36.2</td>
</tr>
<tr>
<td>Sirvent et al. 2003</td>
<td>Spain</td>
<td>72</td>
<td>Adult ICU patients</td>
<td>Prospective cohort study</td>
<td>Suspected pneumonia in ventilated patients</td>
<td>Immunocompromise or hematologic malignancy</td>
<td>Mini-BAL</td>
<td>Quantitative culture add the cutoff for positivity for each study</td>
<td>42 subjects received antibiotic therapy in previous 72 h</td>
<td>73.2</td>
</tr>
<tr>
<td>Fartoukh et al. 2003</td>
<td>France</td>
<td>79</td>
<td>Adult ICU patients</td>
<td>Prospective cohort study</td>
<td>Suspected pneumonia in ventilated patients</td>
<td>Immunocompromise or hematologic malignancy, poor oxygenation</td>
<td>BAL and PTC</td>
<td>Quantitative culture</td>
<td>46 subjects received antibiotics in preceding 48 h</td>
<td>50.6</td>
</tr>
</tbody>
</table>
### Table 1 continued.

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Subjects (N)</th>
<th>Population</th>
<th>Study Design</th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
<th>Method of Sample Collection</th>
<th>Reference Standard</th>
<th>Antibiotic Use</th>
<th>Prevalence, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mentec et al.</td>
<td>France</td>
<td>63</td>
<td>Adult ICU patients</td>
<td>Multicenter prospective trial</td>
<td>Suspected pneumonia in ventilated patients</td>
<td>None specified</td>
<td>BAL, PTC, and EA</td>
<td>PSB culture, showing &gt;1000 CFU/mL</td>
<td>37 received antibiotics at least 72 h prior</td>
<td>65.4</td>
</tr>
<tr>
<td>Brun-Buisson et al. 2005</td>
<td>France</td>
<td>68</td>
<td>Medical-Surgical ICU patients</td>
<td>Prospective cohort study</td>
<td>Clinically suspected pneumonia from CPIS</td>
<td>None specified</td>
<td>BAL and PTC</td>
<td>Quantitative culture, showing &gt;1000 CFU/mL or intracellular organisms</td>
<td>37 received antibiotic at least 72 h prior to specimen collection</td>
<td>51.5</td>
</tr>
<tr>
<td>Davis et al. 2005</td>
<td>United States</td>
<td>155</td>
<td>Trauma ICU patients</td>
<td>Retrospective chart review</td>
<td>Diagnosis of VAP on medical record</td>
<td>Incomplete record</td>
<td>BAL</td>
<td>Quantitative culture</td>
<td>Not reported</td>
<td>88.4</td>
</tr>
<tr>
<td>Kopelman et al. 2006</td>
<td>United States</td>
<td>223</td>
<td>Medical-Surgical ICU</td>
<td>Retrospective chart review</td>
<td>All patients for whom Gram stain and culture was obtained</td>
<td>None specified</td>
<td>BAL</td>
<td>Quantitative culture</td>
<td>Not specified</td>
<td>82.4</td>
</tr>
<tr>
<td>Veinstein et al 2006</td>
<td>France</td>
<td>78</td>
<td>Adult ICU patients</td>
<td>Multicenter prospective trial</td>
<td>Clinically suspected VAP</td>
<td>None specified</td>
<td>PTC</td>
<td>Clinical diagnosis of VAP</td>
<td>57 received antibiotics 48 h prior to specimen collection</td>
<td>47.4</td>
</tr>
<tr>
<td>Raghavedren et al. 2006</td>
<td>United States</td>
<td>124</td>
<td>Trauma ICU patients</td>
<td>Retrospective chart review</td>
<td>Clinically suspected VAP</td>
<td>None specified</td>
<td>BAL</td>
<td>Quantitative culture</td>
<td>Not specified</td>
<td>57.0</td>
</tr>
<tr>
<td>Albert et al. 2008</td>
<td>Canada</td>
<td>705</td>
<td>Adult ICU patients</td>
<td>Retrospective analysis of multicenter randomized control VAP trial</td>
<td>Clinically suspected VAP</td>
<td>Immunocompromise, using corticosteroids, MRSA pneumonia, and pseudomonas pneumonia</td>
<td>BAL or EA</td>
<td>Quantitative culture</td>
<td>443 subjects received antibiotics in 72 h prior to specimen collection</td>
<td>52.5</td>
</tr>
<tr>
<td>Goldberg et al. 2008</td>
<td>United States</td>
<td>229</td>
<td>Adult patients with suspected VAP</td>
<td>Prospective trial</td>
<td>Clinically suspected VAP</td>
<td>None specified</td>
<td>BAL</td>
<td>Quantitative culture</td>
<td>Not specified</td>
<td>23.0</td>
</tr>
</tbody>
</table>

Abbreviations: BAL, bronchoalveolar lavage; CFU, colony-forming units; CPIS, clinical pulmonary infection score; EA, endotracheal aspirate; h, hours; ICU, intensive care unit; PSB, protected specimen brush; PTC, plugged telescoping catheter; VAP, ventilator associated pneumonia.
high discordance between studies [19]. Subgroup analyses were conducted using meta-regression to determine what contribution individual factors such as reference standard and site of specimen (endotracheal aspirate vs lower-respiratory-tract specimen) has on the results, where $P$ values $< .05$ indicate a contribution to heterogeneity. One source of heterogeneity unique to diagnostic meta-analysis is threshold effect, occurring when studies use different thresholds to define a positive test. This is important for VAP where a quantitative lower-respiratory-tract specimen may be variably defined as positive based on the colony count. The presence of threshold effect is tested by calculating the Spearman’s coefficient between sensitivity and specificity, where values $<-0.5$ or $>0.5$ indicate possible threshold effect [20].

A summary measure of accuracy ($Q^*$) was calculated, which corresponds to the upper left-most point on the summary receiver operating characteristic curve where sensitivity equals specificity. This value can be between 0 and 1, with 1 indicating the highest sensitivity/specificity test. This value has been recommended over the area under the receiver operating characteristic curve region of greatest interest [16, 21]. In the presence of heterogeneity, the measure $Q^*$ may be better suited to comparing tests because it accounts for random thresholds, than are measures that do not adjust for the threshold effect, such as pooled sensitivity and specificity. Statistics were calculated manually and with MetaDisc software [22].

RESULTS

The search strategy yielded 645 distinct studies. After screening abstracts and reviewing 77 full-text articles, 24 studies were identified that met inclusion criteria. Reasons for exclusion and search strategy are described in the PRISMA flow diagram in Figure 1.

Table 1 summarizes the characteristics of the included studies. The most common criteria for collection of a respiratory specimen was suspicion of VAP by a treating physician, usually based on signs such as fever, leukocytosis, and/or increased arterial-alveolar gradient, though 1 included only patients who died on mechanical ventilation with suspected pneumonia [23]. Thirteen studies were conducted in Europe, and 8 in North America. Studies included a total of 3148 samples obtained from 2510 distinct patients. Seventeen studies obtained samples using BAL [24–39], 2 studies used mini-BAL [23, 29], 4 used endotracheal aspirate [30, 40–42], and 4 used protected telescoping catheters [30, 31, 36, 43]. For
a reference standard, 2 studies used histologic criteria [25, 26], 2 used expert consensus [27, 43], and the remainder used quantitative lower-respiratory-tract cultures coupled with clinical suspicion for VAP. The most common cutoff for positivity of lower-respiratory-tract cultures was $10 \times 10^3$ colony-forming units (CFU)/mL [24, 28–31, 37–40]. Albert et al. [41] Kopelman [33], and Raghavendran et al. [34] used a cutoff of $>10 \times 10^4$ CFU/mL. Brasel et al. [42], Goldberg et al [35], and Davis et al. [32] defined positivity as those with $>10 \times 10^5$ CFU/mL. Fartoukh et al. used differential cutoffs of $>10 \times 10^3$ for PTC samples, and $>10 \times 10^4$ for BAL samples [36].

The quality of samples and interobserver variability was noted to be high in 1 study using interobserver variability as a primary endpoint [44]. Three studies reported quality criteria, with 1 rejecting lower-respiratory-tract specimens with >1% squamous cells as contaminated [37], and another >1% ciliated cells [24]. Three studies attempted to decrease interobserver variability by using the same microbiologist to evaluate all specimens [24, 29]. All prospective studies stated that “standard” microbiologic technique was employed, though the fact that the studies span 16 years and 7 countries, it is possible that practice variability may exist between studies.

Figures 2 and 3 show the sensitivity and specificity of Gram stain for diagnosis of VAP compared to the reference standard in each study as well as the summary data. Pooled sensitivity was 0.79 (95% CI, .77–.81; $P < .0001$) and specificity was 0.74 (95% CI, .72–.76; $P < .0001$); however, each had high heterogeneity at 79.4% and 83.3%, respectively. Figure 4 summarizes the DOR of the studies. Pooled DOR was 16.44 (95% CI, 10.54–25.67; $P = .0000$), indicating good discriminatory value for the test.

DOR had moderate heterogeneity ($I^2 = 69.9\%$). Investigation of sources of heterogeneity with meta-regression found no evidence of significant threshold effect ($P = .281$). Threshold effect refers to the variability in meta-analysis seen from different standards defining a test as positive, where 1 study may select a higher, more specific, threshold, while another a lower, more sensitive threshold. The plot of the sensitivity by specificity summary receiver operating curve (SROC) shows an inverted L-shape in this situation, as shown in Figure 5. The Q* value of 0.80 is consistent with a highly discriminatory test.

Subgroup Analyses
Antibiotic use at time of VAP specimen collection was reported in 18 studies [29–31, 36, 40, 41, 43], ranging in prevalence from

---

Figure 3. Pooled specificity of Gram stain for diagnosis of VAP. Abbreviations: BAL, bronchoalveolar lavage; Blind-PT, blind plugged telescoping catheter; CI, confidence interval; EA, endotracheal aspirate; Mini-B, mini bronchoalveolar lavage; PSB, protected specimen brush; PTC, plugged telescoping catheter; VAP, ventilator-associated pneumonia.
39% to 100%. Restricting to these studies, meta-regression found no significant contribution of the prevalence antibiotic use to heterogeneity ($P = .78$). Sensitivity and specificity for studies without antibiotic use was very similar to those with antibiotic use, at 0.82 (95% CI, .78–.86) and 0.70 (95% CI, .66–.75), respectively. It is possible that the Gram stain would be less impacted by antibiotic use because it does not distinguish between viable and dead microorganisms. Studies did not provide information about the length of antibiotic therapy, type of antibiotics being used, or timing of antibiotic administration relative to respiratory specimen collection.

There were 2 surgical and 2 trauma ICUs; the few studies did not allow for true subgroup analysis by type of ICU. When we grouped surgical and trauma ICUs together as “surgical,” meta-regression did not find a major contribution to heterogeneity ($P = .74$). Meta-regression also did not show a significant contribution of heterogeneity from either the method of collection (eg, BAL vs endotracheal aspirate) or choice of reference standard.

Although a DOR value of 16.44 indicates good discriminatory value for the Gram stain, the DOR has limitations because it does not take into account the prevalence of VAP. Positive predictive value (PPV) and negative predictive value (NPV) are more clinically meaningful and vary according to the prevalence of disease in a given population. In our included studies, prevalence of VAP ranged from 23% to 88%. Meta-regression showed, as expected, no effect of prevalence on sensitivity or specificity ($P = .71$), because sensitivity and specificity are largely considered independent of prevalence. Across the range of prevalence, the negative predictive value of Gram stain is high, showing that if a Gram stain is negative, the likelihood of not having VAP is high. Figure 6 demonstrates the PPV and NPV for Gram stain across different prevalence values, showing that there is a narrow prevalence range in which the test is clinically relevant. For a VAP prevalence between 20%–30% (which is a clinically common scenario), the NPV of Gram stain is over 90% and the PPV is 40%, indicating that only 40% of positive Gram stains indicate VAP, while VAP can be excluded in 90% of cases when a Gram stain is negative.

**Correlation Between Gram Stain Bacterial Morphology and Organisms Recovered in Culture**

Using the $\kappa$ statistic to examine the correlation between the presence of bacteria on Gram stain and organisms recovered in culture, we found a pooled $\kappa = 0.54$ (95% CI, .50–.58), indicating fair agreement. Individual-study $\kappa$ values ranged from 0 to 1.
to 0.95, summarized in Table 2. Eight studies provided data on concordance of initial Gram stain with final organism (e.g., culture showing a gram-positive [GP] organism in a specimen showing a GP Gram stain). One study [32] did not provide adequate information to calculate a $\kappa$ value, but noted concordance of 65% for GP organisms, and 74% for gram-negative (GN) organisms. In the remaining 7 studies [27, 28, 33–35, 41], pooled $\kappa$ was 0.40 (95% CI, .34–.46) and 0.30 (95% CI, .25–.36) for GP and GN organisms, respectively.

**DISCUSSION**

Appropriate early therapy for VAP is essential to improve clinical outcomes and reduce morbidity. A potential way to accomplish early therapy may be to use the Gram stain of a respiratory specimen as a guide to assessing likelihood of VAP. Our results indicate that while a negative Gram stain suggests a very low likelihood of VAP, a positive Gram stain is not very specific for VAP. Moreover, there was only fair agreement between organisms seen on the Gram stain and the organism responsible for VAP recovered on culture. It is likely that the presence of bacteria on the Gram stain of a respiratory specimen indicates aspiration of oropharyngeal organisms has occurred [45]. As not all of the aspirated organisms are associated with developing VAP, this may account for the discrepant results. Although we were not able to examine this in depth, it is likely that culture results would be affected more than the Gram stain by prior antibiotic therapy, thus contributing to discrepant results between Gram stain and culture.

Our findings have important implications for antimicrobial therapy of VAP. First, empiric therapy for presumed VAP should comprise antibiotics with a broad spectrum of activity directed against the common pathogens implicated in VAP [10]. Presence of a single type of organism on Gram stain does not allow narrowing of initial therapy. Although the concordance between Gram stain and culture was higher for GP organisms than GN organisms, the overall correlation was only moderate. Second, a negative Gram stain was highly predictive of absence of VAP. Our data suggests that, in the absence of a high clinical pretest probability for VAP, a negative Gram stain in a clinically stable patient with suspicion for infection should prompt a search for alternative sites of presumed infection. Our study design and data did not permit assessment of withholding of antibiotic therapy in patients with negative
Gram stain and a low pretest probability of VAP, but future studies should examine this possibility.

Our findings are in keeping with and extend the results of previous studies on this subject. In a systematic review, Rea-Neto et al. [46] reported that the presence of bacteria in Gram stains of BAL specimens had a sensitivity of 44%–90% and specificity of 49%–100%. However, a meta-analysis and further quantitative exploration of the diagnostic yield of a Gram stain was not performed in this study.

Our results should be interpreted in the context of the limitations of this analysis. Antibiotic therapy at the time the respiratory specimen is obtained may negatively impact the diagnostic performance of the Gram stain. While we undertook a subgroup analysis to examine this possibility (which did not suggest a major decrease in yield of Gram stain), our analysis was constrained by lack of detail regarding the type and spectrum of antibiotic therapy, and it is very likely that prior or ongoing antibiotic therapy would alter the yield of the Gram stain and culture of a respiratory specimen. We also found a high degree of heterogeneity in our results. The clinical heterogeneity likely is a result of differing patient populations, varying methods in obtaining respiratory specimens, and variability in processing specimen for Gram stain, although the statistical heterogeneity could not be fully explained by these factors. Although we included all methods for obtaining lower-respiratory specimens, numbers of samples collected via mini-BAL, EA, and PTC were not large enough to allow for meaningful comparison. Further, the lack of an established gold standard means the individual reference standards may over- or underestimate the true prevalence of disease. For example, if in a particular study, VAP is defined by a positive lower-respiratory-tract culture, antibiotic therapy may cause VAP to be underdiagnosed, thus leading to misclassification bias and underestimation of disease prevalence.

These limitations notwithstanding, our study summarizes the available literature and places into context the conflicting results from previous studies, some of which have showed very good concordance between Gram stain and culture [20, 37]. The preponderance of data summarized in our analyses, however, suggests a limited degree of agreement between Gram stain and culture. Therefore, de-escalation or modification of therapy should continue to occur at the 48- to 72-hour mark when culture data become available. There continues to be a need for better diagnostic methods for VAP in critically ill patients.

In conclusion, a positive Gram stain is only moderately specific for VAP, but a negative Gram stain suggests VAP is unlikely. Ultimately, VAP remains a clinical diagnosis, requiring...
the consideration of multiple laboratory, physical examination findings, and radiographic features. No single factor can confirm or rule out a diagnosis of VAP. Future research into whether the Gram stain can be incorporated into an antimicrobial stewardship program decision support process is needed. Studies that specifically examine the impact on Gram stain and culture of prior or concurrent antibiotic therapy at the time the respiratory specimen is obtained are also necessary. Finally, research studies on less invasive methods of respiratory specimen collection (such as mini-BAL) would be of value to determine if the performance of Gram stain of a respiratory specimen obtained by that route is similar to that obtained by more established methods.

**Note**

**Potential conflicts of interest.** All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

**References**


