Accuracy of gram-stained smears as screening tests for Neisseria gonorrhoeae: a systematic review and meta-analysis

Mohamed Abdalla Ali Khalid
Laboratory Department, Kamal Medical Polyclinic. Riyadh 11912, Saudi Arabia

ABSTRACT

Background and Objective: A total of 86.9 million persons worldwide are infected with Neisseria gonorrhoeae (Ng). Although Gram-stained smears (GSS) provide a time- and cost-saving alternative to conventional laboratory tests, their global uptake partly depends on their performance. This study aimed to meta-analyze the diagnostic accuracy of GSS to screen for Ng. Materials and Methods: A literature search was conducted using the MEDLINE (1980 to 2020). Studies were included if they employed GSS to detect Ng in humans and compared the results with reference tests. Results: Eleven studies were reviewed and meta-analyzed and stratified by specimen type (Gram-stained urethral smears and Gram-stained endocervical, urethral swabs and urine smears) and reference test type (culture method or NAAT). Sensitivity was similarly high in GSS versus NAAT (93% [CI, 64% to 99%]) and GSS versus culture methods (87% [CI, 74% to 94%]), followed by Gram-stained urethral smears (97% [95% CI, 86% to 100%]) and Gram-stained endocervical, urethral swabs and urine smears (81% [CI, 67% to 90%]). Specificity was also high in GSS versus culture methods (98% [CI, 95% to 100%]) and GSS versus NAAT (94% [CI, 73% to 99%]), followed Gram-stained endocervical, urethral swabs and urine smears (98% [CI, 93% to 99%]) and Gram-stained urethral smears (96% [CI, 78% to 99%]). Conclusion: Data suggest that GSS have the highest accuracy when investigated against reference culture methods, and Gram-stained urethral smears have the highest accuracy, followed by Gram-stained endocervical, urethral swabs and urine smears. Given their accuracy, convenience, and quick turnaround time, GSS may be useful in expanding first-line screening Ng.

INTRODUCTION

Neisseria gonorrhoeae (Ng) is an etiologic agent of gonorrhea, one of the most common sexually transmitted diseases caused by bacteria,[1] with an estimated global annual incidence of 86.9 million in adults. Gonorrhea can present as urethritis in men, cervicitis, or urethritis in women, and in extragenital sites (pharynx, rectum, conjunctiva, and, rarely, systemically) in both sexes. [2,3] The ideal laboratory test for the detection of Ng should be sensitive, specific, easy to use, rapid, and affordable.[4] In the United States, the Center for
Disease Control and Prevention (CDC) recommends using culture and nucleic acid amplification tests (NAAT) for the diagnosis and detection of Neisseria gonorrhoeae (Ng) in genital, rectal, or pharyngeal secretions. These tests have high sensitivity and specificity, but they are not available until days after testing. This potentially leads to ongoing transmission due to treatment delay or loss to follow-up. A presumptive gonorrhea diagnosis can also be made based on light microscopic detection of the bacterium in Gram-stained smears (GSS). This enables immediate treatment, thus preventing ongoing transmission and loss to follow-up. Given the high absolute burden of high prevalence of Ng globally and the high costs of NAAT and culture methods, GSS testing can reduce the burden of cost considerably, thereby reducing gonorrhea prevalence. To address this knowledge gap, this study reviewed evidence on the diagnostic performance of GSS in screening for gonorrhea.

**METHODS**

This review focuses on the diagnostic accuracy variables (sensitivity, specificity, likelihood ratios [LRs], and diagnostic odds ratios [DORs]) of GSS that are used to screen for Ng in urethral swab andendocervical smears. The author evaluated the studies conducted worldwide in adults regardless of their risk profile in all study settings (laboratory- or field-based) and all study designs (cross-sectional studies and case control). The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for reporting synthesis were followed.

**Data Sources and Searches**

The search was conducted using MEDLINE (via PubMed) from 1980 to 2020. The last search was conducted on June 1, 2020. An example is MEDLINE search string (restricted to humans only): (“Cervical smear for gonococcal” [MeSH] OR “urethral smear for gonococcal” [MeSH] OR “diagnosis of gonorrhea” [MeSH]) AND (“Gram-stained smears for gonorrhea”).

**Study Selection**

This review includes both abstracts and full-text articles of studies conducted using adult humans, with sufficient raw data to recreate 2 × 2 diagnostic tables. Articles were not excluded on the basis of study location or study design. However, non-English articles, studies on the prevalence or accuracy of laboratory-based tests, and those missing relevant information on the index test (GSS) were excluded. Manufacturers’ reports were not included in this review because they usually provide inadequate details on study conduct, have overt conflicts of interest, often provide accuracy estimations without CIs, and exclude important methodological details in study design, patient populations, and samples. Figure 1 is a flowchart of the search. The author conducted the searches and screened the articles for eligibility. After the initial identification of all studies and deletion of duplicates, a preliminary screening of 937 articles was performed, based on the titles and abstracts. Of these, 26 were considered for full-text review, out of which 10 were included in the study.

**Data Extraction and Quality Assessment**

The author abstracted data using a prepiloted form and critiqued the quality of the studies. Data were extracted on the characteristics of the study population, including sampling strategies (purposive or consecutive random sampling), sample size, inclusion and exclusion criteria, specimen tested (endocervical and/or urethral swab), whether the test was a gram stain the reference standard, funding sources, and any reported conflicts of interest. The author also extracted raw data of true positive, true negative, false positive, and false negative results and items necessary for the assessment of the study quality.

In this study, the GSS method was considered, per CDC recommendations. First, swabs were rolled onto clean glass slides and smeared over an area of less than 1 cm². Second, the smears were heat-fixed and gram-stained. Finally, the stained smears were examined using a light microscope under oil immersion (1000× magnification) for gram-negative diplococci and their spatial relationship to polymorphonuclear leukocytes.

Culture techniques followed by confirmation of isolates by biochemical, enzymatic, serologic, or nucleic acid testing, for example, carbohydrate utilization, rapid enzyme substrate tests, serologic methods such as coagglutination, or fluorescent antibody tests, are considered reference standards on the basis of CDC recommendations.

In this study, the methodological and reporting qualities of studies were assessed using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2) tool. The QUADAS-2 checklist assesses potential bias in studies with respect to patient selection, index test, reference test, and patient flow. In assessing the quality of the studies, we also focused on the reference standards used and any reported conflict of interest.

**Data Synthesis and Analysis**

All statistical analyses were carried out using Intercooled Stata, version 15 (StataCorp, College Station, Texas, USA). For the meta-analysis of the estimates of accuracy, the researcher used the bivariate model, which assumes that the measures of sensitivity and specificity from a study are negatively correlated and that the logit transformations of sensitivity and specificity have a bivariate normal distribution. The sensitivity, specificity, positive LR, negative LR, and DOR were calculated. The LRs of a test inform the pretest probability of the disease and provide a post-test probability. A positive LR higher than 5 and negative LR less than 0.2 provide strong diagnostic evidence.

Before the meta-analysis, we stratified the studies into four subgroups based on the specimen tested and whether the reference test was a culture method or NAAT. Because the data were insufficient for all tests and all of the tests under investigation were GSS, we stratified evidence into two subgroups on the same
Khalid A: Accuracy of gram-stained smears as screening tests for Neisseria gonorrhoeae syndrome

Figure 1: Data extraction and quality assessment

Results

Characteristics of Studies

Author search returned 973 reports, of which 11 satisfied all of the inclusion criteria (Figure 1). Table 1 shows the study characteristics. A total of 11 studies were reviewed and analyzed.

Of the 11 studies, 3 (27%) were conducted in developing settings and 8 (73%) were conducted in developed settings. Sample sizes ranged from 95 to 27600 persons.

Study Quality

Eight studies (73%) were cross-sectional (assessed using the QUADAS-2 checklist). Seven studies (64%) used a CDC-recommended reference standard (culture method), whereas the remaining four studies (36%) used molecular tests (NAAT, Gen-Probe, Ng-PCR Rotorgene system, or Xpert® CT/NG assay) as the reference standard. All the research groups administered the same reference test to all patients, thus avoiding partial or differential verification bias (Figure 2).

Five studies (45%) reported a financial relationship with or received funding, 8 (73%) omitted disclosure of conflicts of interest, 3 (27%) explicitly declared no conflict of interest, and 1 reported receiving tests in kind from manufacturers but no conflict of interest.
### Table 1: Characteristics of reviewed studies

<table>
<thead>
<tr>
<th>Author, Year (Reference)</th>
<th>Location</th>
<th>Sample size</th>
<th>Study design</th>
<th>Reference standard</th>
<th>Specimen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bhargava, 2017[11]</td>
<td>India</td>
<td>10,964</td>
<td>Note</td>
<td>Culture</td>
<td>Endocervical and urethral swab</td>
</tr>
<tr>
<td>Goodhart, 1982[14]</td>
<td>United States</td>
<td>401</td>
<td>Cross-sectional</td>
<td>Culture</td>
<td>Endocervical and urethral swab</td>
</tr>
<tr>
<td>Goh, 1985[17]</td>
<td>United States</td>
<td>27,600</td>
<td>Cross-sectional</td>
<td>Culture</td>
<td>Endocervical and urethral swab</td>
</tr>
<tr>
<td>Orellana, 2007[19]</td>
<td>Spain</td>
<td>491</td>
<td>Cross-sectional</td>
<td>Culture</td>
<td>urethral swab</td>
</tr>
<tr>
<td>Borg, 2017[21]</td>
<td>United kingdom</td>
<td>180</td>
<td>Retrospective audit</td>
<td>NAAT</td>
<td>urethral swab</td>
</tr>
<tr>
<td>Hun, 2017[12]</td>
<td>Malaysia</td>
<td>95</td>
<td>Cross-sectional</td>
<td>Xpert® CT/NG assay</td>
<td>urethral swab, urine</td>
</tr>
<tr>
<td>Hananta, 2017[13]</td>
<td>Indonesia</td>
<td>632</td>
<td>Post hoc, exploratory analysis</td>
<td>Ng-PCR Rotorgene System</td>
<td>Endocervical and urethral swab</td>
</tr>
</tbody>
</table>

**Figure 2: Quality assessment of diagnostic accuracy studies assessments**
Table 2: Sensitivity and specificity from each study

<table>
<thead>
<tr>
<th>Study</th>
<th>Sensitivity (95% CI), %</th>
<th>Specificity (95% CI), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bhargava/2017</td>
<td>95 (93–97)</td>
<td>99 (99–99)</td>
</tr>
<tr>
<td>Goodhart/1982</td>
<td>70 (63–76)</td>
<td>85 (79–89)</td>
</tr>
<tr>
<td>Tayor/2011</td>
<td>99 (93–100)</td>
<td>99 (97–100)</td>
</tr>
<tr>
<td>Goh /1985</td>
<td>90 (88–91)</td>
<td>98 (97–98)</td>
</tr>
<tr>
<td>D’ANGELO/1987</td>
<td>56 (42–69)</td>
<td>99 (98–100)</td>
</tr>
<tr>
<td>Orellana/2007</td>
<td>80 (61–92)</td>
<td>90 (87–93)</td>
</tr>
<tr>
<td>Bartelsman/2011</td>
<td>86 (83–87)</td>
<td>100 (100–100)</td>
</tr>
<tr>
<td>Borg/2017</td>
<td>91 (76–98)</td>
<td>64 (55–71)</td>
</tr>
<tr>
<td>JUCHAU/1995</td>
<td>100 (99–100)</td>
<td>100 (99–100)</td>
</tr>
<tr>
<td>Hun/2017</td>
<td>90 (74–98)</td>
<td>95 (87–99)</td>
</tr>
<tr>
<td>Hananta/2017</td>
<td>53 (43–62)</td>
<td>89 (86–92)</td>
</tr>
</tbody>
</table>

Table 3. Results of Meta-analysis, by specimen and reference standard subgroup

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Pooled Sensitivity (95% CI), %</th>
<th>Pooled Specificity (95% CI), %</th>
<th>Positive LR (95% CI), %</th>
<th>Negative LR (95% CI), %</th>
<th>DOR (95% CI),%</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSS verse culture methods</td>
<td>87 (74–94)</td>
<td>98 (95–100)</td>
<td>55.9 (16–196)</td>
<td>0.13 (0.06–0.28)</td>
<td>417 (78–2226)</td>
</tr>
<tr>
<td>GSS verse NAAT</td>
<td>93 (64–99)</td>
<td>94 (73–99)</td>
<td>16.2 (2.7–96)</td>
<td>0.07 (0.01–0.52)</td>
<td>225 (6–7842)</td>
</tr>
<tr>
<td>Gram-stained urethral smears</td>
<td>97 (86–100)</td>
<td>96 (78–99)</td>
<td>25.9 (3.7–180.7)</td>
<td>0.03 (0.00–0.17)</td>
<td>901 (24–33445)</td>
</tr>
<tr>
<td>Gram-stained endocervical, urethral swabs and urine smears</td>
<td>81 (67–90)</td>
<td>98 (93–99)</td>
<td>40.8 (11.5–143.8)</td>
<td>0.19 (0.10–0.36)</td>
<td>215 (41–1126)</td>
</tr>
</tbody>
</table>

DOR: diagnostic odds ratio; LR: likelihood ratio; GSS: Gram-stained smears; NAAT: Nucleic acid amplification tests

Results Pooled by Subgroup
Table 2 reports estimates of sensitivity and specificity from each study. Table 3 pooled estimates for each subgroup.

Diagnostic performance of GSS versus culture methods
The tests investigated in this subgroup were GSS compared with the reference culture method. Among the seven data points, the pooled sensitivity was 87% (95% CI, 74%–94%) and the pooled specificity was 98% (CI, 95%–100%). The positive LR was (55.9% [CI, 16%–196%]), negative LR (0.13 [CI, 0.06%–0.28%]), and DOR (417 [CI, 78%–2226%]).

Diagnostic performance of GSS versus NAAT
The tests in this subgroup were GSS and NAAT (Gen-Probe, Ng-PCR Rotor gene system, and Xpert® CT/NG assay). Among the four data points, the pooled sensitivity was 93% (CI, 64%–99 %) and the pooled specificity was 94% (CI, 73%–99%). The positive LR was 16.2 (CI, 2.7%–96%), the negative LR was 0.07 (CI, 0.01%–0.52%), and the DOR was 225 (CI, 6%–7842%).

Diagnostic performance of Gram-stained urethral smears
Among the four data points, the pooled sensitivity was 97% (CI, 86%–100%) and the pooled specificity was 96% (CI, 78%–99%). The DOR for this subgroup was 901(CI, 24%–33445%). The positive LR was 25.9 (3.7%–180.7%), and a low negative LR was (0.03 [CI, 0.00% to 0.17%]), indicating high accuracy for urethral specimens.
Diagnostic performance of Gram-stained endocervical, urethral swabs and urine smears

Among the six datapoints, the pooled sensitivity was 81% (CI, 67%–90%) and the pooled specificity was 98% (CI, 93%–99%). The positive LR (40.8 [CI, 11.5%–143.8%]), negative LR (0.19 [CI, 0.10%–0.36%]), and DOR (215 [CI, 41%–1126%]), indicated reduced accuracy for specimens other than urethral specimens.

DISCUSSION

The present meta-analysis suggests that GSS has the highest accuracy when compared to reference culture methods; however, its accuracy is relatively less than that of NAAT (Gen-Probe, Ng-PCR Rotor gene system, and Xpert® CT/NG assay). Additionally, GSS showed the highest accuracy in the urethral compared to other specimens, such as endocervical, urethral swabs, and urine specimens. However, all subgroups showed high positive LRs, low negative LRs, and high DORs, and the best LRs and DORs were reported for GSS compared to culture methods, followed by those of GSS compared to NAAT. In contrast, the best LRs and DORs were reported for Gram-stained urethral smears, followed by those of Gram-stained endocervical, urethral swabs, and urine smears. Given the convenience of GSS and their rapid turnaround time, these results show great potential for expanded first-line screening for Ng infection and demonstrate the utility of GSS in gonorrhea screening of at-risk populations.

The high positive and low negative LRs found in each subgroup, especially those that tested urethral specimens and those obtained after comparing GSS to reference methods, also imply that GSS can meaningfully inform the posttest probability of infection. The pooled accuracies of these subgroups have implications for their use in clinical and nonclinical outreach settings. For example, Gram-stained endocervical, urethral swabs, and urine smearsshowed a slightly higher false-negative rate than Gram-stained urethral smears. The false-negative rate is of particular concern in high-risk groups, in which a high rate is more likely to lead to an undetected infection. In such scenarios, timely confirmatory testing could resolve the preliminary screening results. However, the convenience and rapid turnaround time of GSS and their ease of use compensate for their slightly lower sensitivity. In summary, the GSS can be safely integrated into expanded screening initiatives as a first-line screening test.

The results of this study should be interpreted with caution. First, reference standards were found to influence the accuracy of GSS. When the ideal culture reference standard was used, specificity was higher than when a NAAT reference standard was used. In contrast, the sensitivity was high when a NAAT reference standard was used. Only seven of the included studies used the culture reference standard to ascertain true disease status. Misclassification by reference standards is known to influence the measured sensitivity and specificity of the index tests.

Accuracy estimates from studies that used imperfect reference standards to ascertain true disease status may have been artificially inflated (as more sensitive than culture reference standard) or lowered (does not allow for testing of antimicrobial susceptibility) because of misclassification by the reference standard. Standardization of reference standards is required for future diagnostic accuracy studies.

Second, the effect of antimicrobial susceptibility testing on diagnostic accuracy is worth further consideration. Antimicrobial susceptibility testing leads to the expansion of surveillance of antimicrobial resistance and treatment failures, and promotes responsible antimicrobial use and stewardship.

Third, the index tests included in this meta-analysis detected intercellular diplococcus Ng and therefore could not detect infection within 2–10 days. If the clinical suspicion of a positive GSS result is high, further testing is required. In the case of a possible false-negative result, further screening with another conventional laboratory-based test could be considered, depending on available resources. More research is needed to determine how to effectively link screening with further linkages and follow-up, especially in hard-to-reach populations and low-resource settings.

Finally, evidence on GSS will be of greater use to policymakers and guideline developers if outcomes are documented beyond accuracy. These include patient-centered outcomes and operational research outcomes, such as acceptability, preference, feasibility, and impact. Future research on the cost-effectiveness of GSS in different settings, populations, and contexts is warranted to make informed decisions on this test and on testing strategies.

GSS offers many advantages: a fast turnaround time, declaration of results at the point of care with the potential for affecting clinical management, early detection of undiagnosed cases of gonorrhea, and high intra- and inter-observer agreement or concordance, and reduction in the test cost. Given the lack of global evidence, this review comes closer to independently assessing the role of GSS for widespread use in the field by synthesizing all available data on their accuracy and providing further evidence on the benefits of GSS.

In this study, GSS is found to be accurate and suitable for screening initiatives. As a result of their accuracy and the urgent need to increase gonorrhea screening in marginalized and at-risk populations, this test could play a substantial role in expanded screening initiatives, which would eventually impact the control of gonorrhea infection at the population level.

Funding
None.

Ethics approval and consent to participate
Not applicable.

Conflicts of interest
The authors declare no conflict of interest.
Khalid A: Accuracy of gram-stained smears as screening tests for Neisseria gonorrhoea syndrome

REFERENCES


