Relationship Between the Clinical Likelihood of Group A Streptococcal Pharyngitis and the Sensitivity of a Rapid Antigen-Detection Test in a Pediatric Practice

M. Bruce Edmonson, MD, MPH, and Kathryn R. Farwell, BS

ABSTRACT. Objective. The sensitivity of a rapid antigen-detection test (RADT) for group A streptococcal (GAS) pharyngitis is critical to whether the test is cost-effective and to whether a confirmatory throat culture is needed. We evaluated a second-generation RADT to determine if its sensitivity varies across the broad clinical spectrum of patients tested for GAS in pediatric outpatient practice.

Methods. We used laboratory logbooks from a single pediatric clinic to identify 1184 consecutive patient visits at which an RADT was performed. In a blinded chart review, we calculated McIsaac scores to separately estimate the pretest clinical likelihood of GAS pharyngitis for visits at which the RADT result was positive (n = 384) and for visits at which the result proved to be false-negative (n = 65). Positive RADT results were assumed to be true positives, and test sensitivity was estimated by dividing the number of positive results by the sum of positives and false-negatives.

Results. As the clinical likelihood of GAS increased, there were stepwise increases in RADT sensitivity (from 0.67 to 0.88). Sensitivity was low (0.73; 95% CI: 0.62–0.86) in patients clinically unlikely to have GAS (McIsaac score ≤2) and high (0.94; 95% CI: 0.89–0.99) in patients <15 years old who had tonsillar exudate and no cough. False-negative RADT results were associated with lighter growth of GAS than found on specimens obtained from a random sample of clinic patients who had only primary throat cultures ordered.

Conclusions. For pediatric patients who are clinically unlikely to have GAS pharyngitis, as indicated by a McIsaac score ≤2, the sensitivity of a second-generation RADT may drop below thresholds reported for cost-effectiveness. For children who have tonsillar exudate and no cough, the test may be sensitive enough to meet current pediatric practice guidelines for stand-alone testing. Pediatrics 2005;115:280–285; pharyngitis, streptococcal infections, rapid diagnostic tests.

ABBREVIATIONS. RADT, rapid antigen-detection test; GAS, group A streptococcus; TC, throat culture; CI, confidence interval; AOR, adjusted odds ratio.

Adequate sensitivity is critical to whether a rapid antigen-detection test (RADT) is cost-effective in the diagnosis and treatment of group A streptococcal (GAS) pharyngitis in children.2,3 If the RADT is used as a stand-alone test, its sensitivity must equal or exceed that of an office throat culture (TC).3,4 Evidence is mixed on whether RADT sensitivity does5,6 or does not7–10 meet this standard. More fundamentally, it is debated whether an office TC is sufficiently sensitive to serve as a reference standard for evaluation of rapid antigen tests.11–13

Those involved in the debate about RADT performance have generally failed to consider whether sensitivity might also depend on the clinical features of tested children. In adults, at least, a strong, direct relationship between RADT sensitivity and the clinical likelihood of GAS pharyngitis was observed in a recent emergency department study.14 Two other studies reached conflicting conclusions about whether such an effect occurred in children, but both studies evaluated relatively insensitive latex-agglutination assays.15,16 A third pediatric study did evaluate a more sensitive, second-generation RADT and found lower sensitivity in children who lacked certain clinical features of GAS pharyngitis, but the effect was modest and not statistically significant.17

It is biologically plausible that such a spectrum effect18 might occur in children tested for GAS pharyngitis. Children with mild pharyngitis are more likely to have sparse pharyngeal growth of GAS than children with classical signs and symptoms19,20 despite the fact that clinical features correlate poorly with development of antistreptococcal antibodies or risk of rheumatic fever.15,16,21 It is separately known that the sensitivity of rapid detection tests typically depends, in part, on the presence of adequate amounts of GAS antigen on the test swab.5–7,12,22–25

We were concerned that an otherwise sensitive, second-generation RADT might perform poorly in children with mild or nonclassical GAS pharyngitis and that indiscriminate use of the test might not be cost-effective. We designed a study to estimate the overall sensitivity of a second-generation RADT in a consecutive series of patients in a general pediatric practice and then to determine if its sensitivity varies in relation to the clinical likelihood of GAS pharyngitis. We also studied quantitative TCs in our patients to determine if false-negative RADT results might be associated with recovery of sparse amounts of GAS from pharyngeal swabs.
RESULTS
Our primary outcome was the association of RADT and TC results with GAS pharyngitis. We included 20725 patients with a primary TC and/or RADT result during the study period. Of these, 21547 patients had a chart record that was not lost to follow-up or a chart that could be accessed.

Study Setting
All study subjects were selected from patients who visited a single pediatric clinic in Madison, Wisconsin, operated by the University of Wisconsin Hospital and Clinics. This free-standing facility serves as both a primary care practice site and an after-hours/weekend urgent care clinic. The clinic is staffed by academic generalist pediatricians, a nurse practitioner, medical students, pediatric residents, and community pediatricians. Approximately 77% of the 21547 patient visits to the clinic in 2002 were covered by various capitated or contracted-care insurance arrangements, 13% by other commercial insurance, 7% by Medicaid/General Assistance, and 3% by no insurance.

Study Design
This is a retrospective, cross-sectional comparison of laboratory records and clinical chart data on consecutive patients (<24 years old) who visited the clinic between January 2000 and May 2002 and had a diagnostic test to detect pharyngeal GAS. According to clinic protocol, clinicians were free to order an on-site RADT or to send a primary TC specimen to the hospital laboratory. According to discussions with clinic staff, factors favoring use of an RADT (instead of a primary TC) included anticipated obstacles to telephone follow-up and treatment, after-hours care, clinical judgment that GAS was likely and that treatment should not be delayed, and physician preference. Clinical practice was not to order a diagnostic test (TC or RADT) for GAS cases in which patients were currently taking a systemic antimicrobial agent. A back-up TC specimen was routinely sent to the hospital laboratory whenever an RADT result was negative. Clinic staff maintained handwritten logbooks to record the results of each RADT and TC performed in the clinic.

For the primary analysis, we restricted the sampling frame to all patients who had an RADT performed. We then used test results from clinic logbooks to select subjects according to the following scheme: (1) all patients who had a positive RADT result; (2) all patients who had a false-negative RADT result (back-up TC positive for GAS), and (3) a random sample of patients who had a true-negative RADT result (back-up TC negative). Small numbers of tested patients were excluded from the sample because either chart records could not be located or a back-up TC had not been performed despite a negative RADT result (see Fig 1).

A second analysis was designed to determine if a false-negative RADT result was associated with light growth of GAS on TC. We compared 2 groups: (1) case patients (described above) who had a false-negative RADT result and (2) unmatched controls randomly sampled from patients who had a positive primary TC during the study period.

We used patient identifiers to link logbook and hospital-laboratory data, assigned a randomly generated, unique study number to each selected subject, and entered test results into a microcomputer database (Epi-Info 2000, Centers for Disease Control and Prevention, Atlanta, GA). Next, we obtained and photocopied all chart data (telephone notes and handwritten and transcribed visit notes) relevant to each subject's visit. To permit an unbiased abstraction of chart information, we then modified the photocopied notes by blacking-out patient identifiers (name, medical record number, visit date) and any information about visit outcome (RADT/TC result, diagnosis, or treatment). Finally, we abstracted clinical data according to a prospectively designed format that included age, gender, visit date (month/year), patient/parent-reported history (sore throat, temperature >38°C, cough, runny nose, diagnosis of GAS pharyngitis within the previous month, and school/household exposure to GAS), and physical findings (temperature >38°C, tonsillar swelling, tonsillar exudate, anterior cervical adenopathy (tender versus nontender). We interpreted absence of an explicit chart notation about any particular symptom (eg, cough) to mean that it did not occur and contradictory notations to mean that it did occur. Contradictory information about physical findings was resolved in favor of the attending physician note.

The pretest likelihood of GAS pharyngitis was the primary exposure of interest in this study and was estimated according to a clinical sore-throat-scoring scheme developed by McIsaac:26,27 This score has been prospectively validated in a mixed population of children and adults and is computed by first recording a point for each of the following: history of or measured temperature >38°C, absence of cough, tender anterior cervical adenopathy, tonsillar swelling or exudate, and age <15 years. According to the method of McIsaac, points were summed to a maximum allowable score of 4.

Collection/Processing of Throat Swabs
During the first 11 months of the study period, pharyngeal specimens were obtained with individual swabs: one for RADT use and a second swab for TC to back-up any negative RADT result. For the remainder of the study period, specimens were obtained with double swabs (Culturette II, Becton-Dickinson, Cockeysville, MD). The throat-swabbing technique was not stan-

Fig 1. Results of RADTs and TCs performed on clinic patients, January 2000–May 2002. Subjects below the dotted line were selected for chart audits.
dardized, although providers were aware of recommendations regarding preferred sampling sites (each tonsil/tonsillar bed and posterior pharynx, avoiding the tongue/soft palate28). RADT specimens were processed immediately on-site in the clinic laboratory, and back-up swabs were sent to the hospital laboratory.

RADT and Reference-Standard Procedures

The RADT was conducted with CARDS QS Strep A (Quidel Corp, San Diego, CA), a lateral-flow immunoassay that uses chemochromogenic indicators to detect GAS. The test was conducted by medical assistants and phlebotomists in an on-site clinic laboratory. Each test run included its own positive and negative control and, as indicated by manufacturer instructions regarding interpretation, color changes in the control windows or test well, was repeated on a new specimen whenever a result was judged to be invalid. Microbiologists from the hospital central laboratory periodically conducted proficiency testing of clinic staff and reviewed RADT log sheets.

Back-up TCs were plated within 1/2 hour of receipt by the hospital laboratory on 5% sheep blood agar, incubated anaerobically at 35 to 37°C, held overnight, and read each of the following 2 mornings. Subculture plates (blood agar) were inoculated for selected β-hemolytic colonies. β-Hemolytic colonies were confirmed as GAS by latex agglutination (PATHO Dx, Remel, Lenexa, KS) or disk detection of pyrrolidonyl peptidase (PYR Disk, Remel). GAS growth was routinely quantified on a scale of 1+ (few) to 5+ (many), with 1+ and 2+ defined according to number of colonies (1–20 and >20, respectively) in the first quadrant. 3+ defined as colonies filling the first quadrant, 4+ as colonies filling at least the first and second quadrants, and 5+ as growth extending into the fourth quadrant. Blinding of culture microbiologists to RADT results was assured, because the clinic and laboratory were physically separated and labeling of culture plates did not indicate whether the source specimen was a primary throat swab or a back-up (RADT-negative) swab.

Statistical Analysis

We could not calculate RADT sensitivity directly, because positive test results in clinic patients were not routinely verified by TC. However, the RADT used in this study29 and other second-generation tests17,20,31,32 have such high specificity (ranging from 0.93 to 0.99) that we assumed, for the purposes of this study, that all positive RADT results were true positive, overall RADT sensitivity was 0.86 (384/449 [95% CI: 0.83–0.89]) and overall GAS prevalence among patients tested with an RADT was 0.38 (449/1184 [95% CI: 0.35–0.41]).

Clinic providers were free to order a TC for any patient, and a chart audit based on a random sample (n = 100) of the 2530 patients who had primary TC indicated that these patients were clinically less likely to have GAS (31.0% had a McIsaac score ≤2 vs 20.2% of patients tested with RADT [P = .01]). The prevalence of GAS, 0.26 (659/2530 [95% CI: 0.24–0.28]), was lower in patients who had a primary TC ordered than in patients who had an RADT.

Table 1 shows RADT results stratified by clinical characteristics of study patients. Univariate comparisons show that test sensitivity increased (P < .05) as patient age decreased (in 5-year increments), when sore throat was a symptom, when no cough was reported, when a patient had recently completed treatment for GAS pharyngitis, or when tonsillar exudate or swelling were noted on physical examination. We failed to find a statistically significant variation in RADT sensitivity according to patient gender, year of visit, season of year, history of runny nose, fever, tender anterior cervical nodes, or history of recent household/school exposure to a reported case of GAS pharyngitis.

Table 2 shows a direct, stepwise relationship between the clinical likelihood of GAS (as measured by McIsaac score) and estimated RADT sensitivity (range: 0.67–0.88; P = .03 for trend). For the 240 (20.2%) patients who had a McIsaac score ≤2, test sensitivity was 0.73 (95% CI: 0.62–0.86). A total of 444 tests (240 RADT + 204 back-up TC) was performed for these 240 patients, and 13 (26.5%) of the 49 GAS infections in this group were not detected until results of back-up cultures became available.

In a multivariate analysis restricted to the 449 patients who had GAS, 3 factors were independently associated (P < .05) with increased likelihood of a positive (vs a false-negative) RADT result: (1) tonsillar exudate (adjusted odds ratio [AOR]: 3.7; 95% CI: 1.7–7.7); (2) age <15 years (AOR: 3.1; 95% CI: 1.1–8.8); and (3) lack of cough (AOR: 2.4; 95% CI: 1.4–4.2). Based on weighted distributions of these 3 parameters in our sample, we estimated that 180 (15.2%) of 1184 tested patients had all 3 of these factors; GAS prevalence (0.55) and RADT sensitivity (0.94; 95% CI: 0.89–0.99) were particularly high in this group. In contrast, only an estimated 33 (2.8%) of 1184 tested subjects were ≥15 years old, had a cough, and lacked tonsillar exudate; point estimates of GAS prevalence (0.12) and RADT sensitivity (0.50) were particularly low in this group, but these estimates are imprecise because they are based on only 4 cases of GAS (2 of which were detected by RADT) in this small group of patients.

In a subgroup comparison of patients <3 years old (n = 109) versus patients 3 to 4 years old (n = 155), we found that the youngest patients were more likely to have a McIsaac score ≤2 (22.0% vs 7.1%; P < .001) and less likely to have GAS (prevalence: 0.12 vs 0.43; P < .001). Although a precise evaluation of test performance in children <3 years old was constrained by the small numbers who had GAS, we
found that test sensitivity was 0.85 (95% CI: 0.54–0.98) in this group. We failed to find a statistically significant variation in sensitivity according to McIsaac score within the group of patients <3 years old.

Table 3 shows semiquantitative GAS TC results from the 65 patients who had a false-negative RADT result. Compared with growth from a random sample of control patients who had a positive primary TC, lighter growth of GAS was strongly associated with having a false-negative RADT result ($P < .001$, Wilcoxon rank-sum test). Although the case and control subjects were drawn from different sampling frames, the subjects themselves had similar ages (median: 7 vs 8 years, respectively) and McIsaac scores (20.0% of each group had scores ≥2).

**DISCUSSION**

Results of this study demonstrate substantial variation in the sensitivity of an RADT across the broad clinical spectrum of patients tested for GAS pharyngitis in a general pediatric clinic. Test sensitivity was directly related to the overall clinical likelihood of GAS and was particularly enhanced in patients who...
had pharyngeal exudate. We also found that false-negative test results were associated with lighter-than-expected growth of pharyngeal GAS on back-up TC, a finding that confirms previous laboratory results and lends biological plausibility to our clinical observations.

Although we collected study data retrospectively, we did use a previously validated measure to estimate the clinical likelihood of GAS pharyngitis and calculated likelihood scores independently of test results. Clinic procedures also provide reasonable assurance that rapid antigen tests and reference-standard TCs were interpreted independently of each other and without reference to clinical information about patients. We evaluated only 1 second-generation RADT and cannot directly apply our results to other rapid GAS immunoassays, but we do note that our estimate of overall test sensitivity was comparable to levels reported previously for the same product and for a commonly used optical immunoassay (summarized by Webb). Our sensitivity estimates were based on data from a clinic at which providers used clinical judgment to decide whether to order an RADT, but our findings should still be applicable to other settings as long as case mix (as measured by McIsaac score) is taken into account. In general, nonselective application of this RADT would likely increase the proportion of tested patients who have low McIsaac scores and, in turn, reduce overall test sensitivity.

One limitation of our retrospective study design is that providers did not necessarily record history and physical findings before seeing RADT results. It is possible that signs and symptoms of GAS pharyngitis were falsely ascribed to some patients once their positive RADT test result became available. We did observe, however, that RADT sensitivity was related to patient age and history of recent GAS pharyngitis, 2 factors that were unlikely to have been influenced by recall bias.

Another limitation of our study is that we used a single TC as a reference standard. TCs were plated in a hospital laboratory, incubated anaerobically, and enhanced by selective use of subcultures, but estimates of overall RADT sensitivity probably would have been lower if we had used a more sensitive reference standard such as double culture, Todd-Hewitt broth, or polymerase chain reaction. Because false-negative RADT results were jointly associated with clinically mild illness and sparse colony counts of GAS, however, application of a more sensitive reference standard probably would have increased the magnitude of the observed spectrum effect.

Although specimens from clinic patients who had a negative RADT results were routinely tested by TC, positive RADT results were not confirmed. Test sensitivity was estimated by assuming, for the purpose of this study, that false-positives were so uncommon that they did not occur. Such an assumption, if true, virtually eliminates the possibility that study results were distorted by verification bias. If a small number of false-positives did occur but occurred randomly and independently of clinical likelihood of GAS pharyngitis, the net effect would be that we slightly overestimated the overall sensitivity of the RADT.

During analysis of the data, we were still concerned about the validity of our sensitivity estimates for the subgroup of patients who had a history of recently treated GAS pharyngitis. We estimated that RADT sensitivity for these patients was 0.98 (95% CI: 0.94–1.00), a level so high that we speculated, based on the suggestion of others, that nonviable GAS antigen might persist in the pharynx and produce misleading (false-positive) RADT results. In a recent study, however, a second-generation RADT proved to be both highly specific (0.96) and sensitive (0.91) in patients who had a clinical relapse after completion of treatment for GAS pharyngitis. These results were consistent with our findings and help validate our method of estimating test sensitivity.

Two recent pediatric studies concluded that an RADT without back-up culture is a more cost-effective strategy than either a primary TC or an RADT with a back-up culture. Neither study seems to have modeled the economic impact of a possible interaction between test sensitivity and pretest likelihood of GAS. In 1-way sensitivity analyses, 1 of these studies found that a primary TC was the economically favored strategy when either GAS prevalence dropped below 0.20 (assuming test sensitivity to be 0.89) or test sensitivity dropped fell below 0.82 (assuming GAS prevalence to be 0.29). For our patients who had a McIsaac score ≤2, the GAS prevalence (0.20) and test sensitivity (0.74) were both at or below the levels at which any RADT strategy was favored over a primary TC.

These findings and our study results, taken together, suggest that the cost-effectiveness of pediatric RADT strategies needs to be reevaluated, particularly for children who are clinically unlikely to have GAS pharyngitis. Based on our results, it seems that RADT testing is particularly inefficient for patients with low McIsaac scores, for whom rapid testing is both unlikely to yield a positive result and very likely to incur the high marginal costs associated with back-up TCs mandated by current guidelines.

**TABLE 3. Growth of GAS on TCs From Patients Who Had a False-Negative RADT Result Compared to Growth From Patients Who Had a Positive Primary TC**

<table>
<thead>
<tr>
<th>Test Result</th>
<th>Growth of GAS on Throat Culture*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1+</td>
</tr>
<tr>
<td>False-negative RADT</td>
<td></td>
</tr>
<tr>
<td>(n = 65)</td>
<td></td>
</tr>
<tr>
<td>Positive Primary TC</td>
<td></td>
</tr>
<tr>
<td>(n = 65)†</td>
<td></td>
</tr>
</tbody>
</table>

* Semiquantitative growth scored 1+ to 5+, ranging from scant growth (1–20 colonies in the first quadrant of the culture plate) to heavy growth (colonies present on all 4 quadrants) (see “Methods”).

† Results are based on a random sample (n = 65) of 659 positive cultures reported during the study period for patient visits at which only a primary TC was performed.
than in children, it may be justifiable to rely on a clinical scoring scheme to avoid laboratory tests altogether for low-risk patients and to order a single test (RADT or TC) for those at higher risk. Based on the results of our study, we believe that RADT performance and cost-effectiveness may be compromised for those pediatric patients who have low (≤2) McIsaac scores. At the other end of the clinical spectrum, a second-generation RADT may be sufficiently sensitive to rely on stand-alone testing for children who have exudative pharyngitis and no cough. Although these findings should be prospectively confirmed and subjected to a formal decision analysis, our results suggest that selective use of RADTs and back-up TCs for pediatric patients might be more cost-effective than the nonselective strategies previously studied.

REFERENCES
