Measuring the Quality of Care for Group A Streptococcal Pharyngitis in 5 US Health Plans

Rita Mangione-Smith, MD, MPH; Marc N. Elliott, PhD; Lok Wong, MS; Laurie McDonald, MS; Joachim Roski, PhD

Background: There is a high degree of professional consensus that children diagnosed with pharyngitis should only receive antibiotics if they have a positive test for group A streptococcus (GAS).

Objectives: To develop and test the validity of a quality of care performance measure that examines GAS testing rates in children diagnosed with pharyngitis and prescribed an antibiotic.

Design: The measure developed examines the annual rate of GAS testing in children aged 2 to 18 years with an episode of pharyngitis who were prescribed antibiotics. The measure was tested for feasibility of implementation and validity in 5 health plans in the United States. Health plan administrative data were used to identify episodes of pharyngitis using International Classification of Diseases, Ninth Revision (ICD-9) codes 462, 463, and 034.0. Pharmacy data (National Drug Codes) were used to determine if antibiotics were prescribed for the pharyngitis episode. Laboratory claims data (Current Procedural Terminology codes) were used to determine whether a GAS test was performed. Rates of GAS testing in children with pharyngitis who received antibiotics were calculated for each health plan. Medical record abstractions were performed on a random sample (n=465) of cases to assess percent agreement with laboratory claims data for GAS testing. Sensitivity of the administrative data for accurately identifying when GAS tests were performed was also assessed.

Results: Of the 120,158 children aged 2 to 18 years who had at least 1 episode of pharyngitis during the measurement year, 51,172 (43%) received antibiotics. Group A streptococcal testing rates for patients who were prescribed antibiotics varied widely among the participating health plans (59%-83% of cases; P<.05). Percent agreement between administrative and medical records data for GAS tests was 86%. The sensitivity of the administrative data for accurately identifying when GAS tests were performed was 85%.

Conclusions: This quality measure is feasible to implement at the health plan level and validly assesses GAS testing rates using administrative data. The participating health plans are not performing GAS tests as indicated by current expert practice guidelines in a substantial proportion of cases. Improvements in adhering to these guidelines are warranted given the current levels of antibiotic overuse and antibiotic resistance nationally.

Arch Pediatr Adolesc Med. 2005;159:491-497
in-office test with reasonable sensitivity is available for confirmation of the clinical diagnosis. Both the American Academy of Pediatrics and the Infectious Diseases Society of America recommend that antibiotics should only be prescribed for children and adolescents (<18 years of age) who have a positive test for group A streptococcus (GAS). Because of this high degree of consensus, pharyngitis represents an excellent target condition for performance improvement in antibiotic prescribing practices.

One mechanism for underscoring the importance of specific medical practices is the public reporting of performance on selected quality of health care measures. The most widely used and accepted set of performance measures, the Health Plan Employer Data and Information Set (HEDIS), allows for measurement of the degree to which evidence-based medical practices are being implemented, allows purchasers and consumers to select health plans based on such information, and raises general awareness of the need to improve measured practices. All HEDIS measures are consistently and rigorously developed, tested, and rated according to the Desirable Attributes of HEDIS measures, and must go through the National Committee for Quality Assurance’s (NCQA) approval process. Some commonly reported HEDIS measures that have successfully gone through this process include the Comprehensive Diabetes Care Measure (which assesses hemoglobin A1c and lipid testing and control, performance of annual eye examinations, and monitoring for diabetic nephropathy), the Use of β-blockers After Acute Myocardial Infarction Measure, and the Use of Appropriate Medications for People With Asthma Measure.

A valid performance measure examining GAS testing rates in children diagnosed with pharyngitis and prescribed an antibiotic would shed light on the prevalence of inappropriate antibiotic prescribing for this condition and would potentially lead to decreased antibiotic use and antibiotic resistance in the community. The objective of the current study was to test the validity of a new quality of care performance measure developed to examine GAS testing rates for children and adolescents diagnosed with pharyngitis and prescribed an antibiotic. The results of this data analysis were ultimately used by the NCQA Committee on Performance Measurement to determine whether the measure met the standards required for inclusion in the HEDIS measurement set.

**METHODS**

**DEVELOPMENT OF THE MEASURE**

Our research team at the University of California, Los Angeles, collaborated with members of the quality measure development group at NCQA to develop this performance measure. We first convened a panel of 7 experts: 1 representative from the Center for Disease Control and Prevention’s Respiratory Diseases Branch, 2 pediatric infectious disease specialists, 2 research experts in antibiotic overprescribing, and 2 health plan medical directors. Using NCQA’s Desirable Attributes of HEDIS measures as a guide, the group of experts developed the specifications for a quality of care performance measure that examines the rate of GAS testing in children aged 2 to 18 years diagnosed with pharyngitis and prescribed an antibiotic.

To be eligible for the measure, a child member would have to be covered by the plan’s commercial or Medicaid products, be between 2 and 18 years of age at the time of the encounter, have both medical and pharmacy benefits, and have an eligible visit for pharyngitis during the measurement period (July 1, 2000-June 30, 2001).

A pharyngitis visit (International Classification of Diseases, Ninth Revision [ICD-9] codes: 462 acute or unspecified pharyngitis, 463 acute tonsillitis, 034.0 streptococcal tonsillitis) was considered eligible for inclusion in the measure if (1) no additional diagnoses for the visit were assigned that might be treated appropriately with antibiotics (bacterial diagnoses); (2) no antibiotic prescription was issued during the 30 days prior to the visit date (which could represent an antibiotic refill for a prior diagnosis/visit rather than a new prescription linked to the current pharyngitis visit); (3) no additional outpatient visits with a bacterial diagnosis occurred during the 7 days before or the 7 days after the visit date; (4) the child member was continuously enrolled in the health plan during the 30 days prior to and the 7 days following the visit date with no gaps in enrollment during this period; and (5) an antibiotic was dispensed on or during the 7 days after the visit date (determined using National Drug Codes from pharmacy billing data); emergency department visits for pharyngitis were excluded to focus on the outpatient office-based management of pharyngitis where laboratory tests might be less available.

If a child had more than 1 episode of pharyngitis that met all of the inclusion criteria, the first eligible episode during the measurement period was selected for inclusion in the measure. The panel chose to include only the first eligible episode of pharyngitis for 2 main reasons. First, they were concerned that including repeated visits by 1 child to 1 provider would result in a loss of information due to clustering of prescribing and testing behavior within physician-patient dyads. Second, because this measure was being developed for potential inclusion in the HEDIS measurement set, the panel decided to be consistent with the majority of existing HEDIS measures where the individual patient is the unit of analysis rather than episodes of care.

We developed administrative data measure specifications to identify the eligible population and assess the management of pharyngitis cases. We defined the measurement period as a 12-month period starting July 1, 2000, and ending June 30, 2001. Using administrative data, health plans identified all outpatient visits (excluding emergency department visits) during this 12-month measurement period with a qualifying diagnosis of pharyngitis (462 acute or unspecified pharyngitis, 463 acute tonsillitis, 034.0 streptococcal tonsillitis). The health plan then identified the first such visit during the measurement period that met all of the inclusion criteria. Finally, using laboratory billing data (Current Procedural Terminology [CPT] codes), the health plans identified whether the member had a GAS test on or during the 3 days before or after the eligible pharyngitis visit.

**FEASIBILITY AND VALIDITY TESTING OF THE MEASURE**

The NCQA solicited health plans who participate in HEDIS to take part in testing the new performance measure. Five geographically dispersed and structurally diverse health plans agreed to test the newly developed specifications for the pharyngitis measure. These plans were located in the northeastern, southeastern, midwestern, northwestern, and southwestern regions.
of the United States. The types of health plans represented included: capitated health maintenance organizations, group model health maintenance organizations, and health maintenance organizations with preferred provider organization and independent practice association products. Three of the plans submitted commercial data only, while 2 plans submitted both commercial and Medicaid data. With this test, we evaluated the feasibility of implementing the measure at the health plan level and the validity of using administrative data to measure performance in this clinical area.

MEDICAL RECORD ABSTRACTIONS

Each plan abstracted medical records for a random sample of the eligible episodes they identified to verify and validate information obtained from administrative data. Using a medical record abstraction form designed by the research team, health plan personnel abstracted approximately 40 to 60 medical records that indicated that a GAS test had been performed for the eligible visit and 40 to 60 records that indicated that no GAS test had been performed, as specified by the administrative data. The abstractors were asked to verify if the medical record had evidence of the following data: (1) the pharyngitis diagnosis and the date of the visit identified using administrative data; (2) evidence of an antibiotic prescription on or during the 7 days after the visit date; (3) any exclusionary bacterial diagnoses assigned on the visit date, a prescription for antibiotics issued during the 30 days prior to the visit date; (4) any additional visits during the 7-day period before and after the visit date for which a bacterial diagnosis was assigned; and (5) evidence of GAS testing on or during the 3 days before or after the visit date. These data were used as a gold standard to verify the positive predictive validity of the administrative data concerning the diagnosis assigned on the eligible visit date, the positive predictive validity of the health plan’s pharmacy data for capturing antibiotics prescribed, the presence of clinical exclusions, and the sensitivity of the health plan’s laboratory claims data for capturing GAS tests performed during the episode. Four of the 5 participating health plans agreed to perform the medical record abstractions.

ANALYSIS

We compared the rates of GAS testing for the 5 participating health plans using logistic regression analysis, entering dummy variables for each health plan. We did this to test whether or not there was meaningful variation in performance on this measure among the 5 plans. We added 2 covariates, child age and visit occurrence during GAS season (November-March), to the model to permit case-mix adjustment of antibiotic prescribing and GAS testing rates. We calculated exact, binomial 95% confidence intervals (CIs) for all reported proportions. In 4 of the 5 participating health plans, we used medical records data to examine the positive predictive validity of the administrative data for determining the diagnosis assigned, ie, how often the administrative data indicated that a child was diagnosed with pharyngitis and if the medical record diagnosis was the same. Similarly, we used medical records data to assess the positive predictive validity of the pharmacy data for determining whether a drug was prescribed at a given visit. In 3 of the 5 plans, we used medical records data to validate the sensitivity of laboratory claims data for determining whether or not a GAS test was performed at a given visit. For GAS tests, we were most interested in examining the false-negative rate, ie, how often there were cases for which the medical record documented evidence of a GAS test being performed but the administrative laboratory data did not.

All study procedures were reviewed and exempted by the General Campus Institutional Review Board at University of California, Los Angeles.

RESULTS

The population-based rate for pharyngitis diagnoses for the 5 participating health plans during the measurement year was 216 per 1000 enrolled members aged 2 to 18 years. Pooling the data across the 5 participating health plans, there were 310 314 incident cases of pharyngitis during the measurement year. Of these cases, 129 631 were excluded because they met 1 or more of
the 4 exclusion criteria (Figure). Antibiotics were prescribed in 43% of the remaining cases. Thus, 51,172 children aged 2 to 18 years had at least 1 eligible episode of pharyngitis during the measurement year. A GAS test was performed in 74% of these cases (Figure).

The antibiotic prescribing rates (Table 1) for pharyngitis varied widely among the participating health plans (range = 9%-61%; P < .05). The GAS testing rate (Table 1) also varied widely by health plan for children diagnosed with pharyngitis and prescribed an antibiotic (range = 59%-83%; P < .05). In some cases, plan B used plan-specific procedure codes for GAS testing in their administrative data rather than using the standard CPT codes for these tests. As a result, plan B had a falsely low rate of performance on this measure after running the administrative data specifications on their sample of 427 encounters (19%; 95% CI, 15%-23%). Their data are not included in Table 1 because of this coding discrepancy.

Antibiotic prescribing rates and GAS testing rates that were case-mix adjusted for child age and GAS season of the visit varied little by health plan. Although the case-mix variables were predictive at the individual level, they varied little at the health plan level, resulting in almost no net adjustment of rates across plans. Using the medical record as the gold standard for diagnosis assigned, the positive predictive validity of the administrative data for correctly identifying pharyngitis diagnoses was 88% (range, 81%-98%) (Table 2). Thus the administrative data indicated a diagnosis of pharyngitis in 2% to 19% of cases whereas the medical record did not (false-positive rate). Medical records data also indicated a secondary (competing) bacterial diagnosis for the index episode of care in 10% of cases whereas the administrative data did not.

Again using the medical record as the gold standard for antibiotics prescribed, the positive predictive validity of the administrative data for correctly identifying whether or not an antibiotic was prescribed on the visit date was 88% (range, 72%-95%) (Table 2).

The overall rate of agreement between the administrative data and medical records data on whether a GAS test was performed was 86% (range, 82%-91%) (Table 3). The sensitivity of the administrative data for accurately identifying GAS tests documented in the medical record was 85%. In 10% of cases overall, the medical record indicated a GAS test was performed that was not detected with the administrative laboratory data (false-negative rate). In 4% of cases, evidence for a GAS test was identified using administrative data, but could not be confirmed in the medical record.

As noted in the footnotes of Table 1 and Table 3, we were not able to accurately analyze plan B’s performance on this measure secondary to a plan-specific cod-
ing discrepancy for GAS testing. Using administrative laboratory data, we found that health plan B had a significantly lower rate of obtaining GAS tests when compared with all other plans (19%, P<.05). In discussing these results with the project team at health plan B, we determined that they use plan-specific procedure codes for GAS tests, which are frequently not converted to traditional CPT codes in their databases. Thus, when they ran the specifications for this measure, many of their GAS tests were not captured. In fact, the concordance rate between GAS tests detected in their medical records vs their administrative data was only 49%. Medical records data for plan B (n=124) indicated that GAS tests were performed in 92% of the pharyngitis cases for which antibiotics were prescribed. Using their administrative data to score the measure using these same 124 cases, their score was 50%, thus underscoring plan B's use of plan-specific procedure codes for GAS testing rather than standard CPT codes. Since performance on this measure using either medical records or administrative data was similar and highly correlated in all other plans, it is assumed that plan B's true performance is around 92%.

The results of this measure validation study, which examined rates of GAS testing in children with pharyngitis in 5 United States health plans, indicate that there is a substantial degree of variability in performance on the measure developed and significant room for improvement in 2 (plans A and C) of the 5 participating health plans (Table 1). The level of performance indicates that clinical guidelines addressing testing for GAS infection prior to treating pharyngitis with an antibiotic are not followed in 25% of cases on average (Table 1). This rate is likely to be an underestimate of the true degree of poor performance since test results are not always followed up, and children with negative GAS tests are sometimes still prescribed an antibiotic. Among children aged 2 to 18 years, we expect approximately 25% to 30% of pharyngitis cases to be secondary to GAS and thus appropriately treated with an antibiotic. Using the National Ambulatory Medical Care Survey data, McCaig et al found that between 1999 and 2000, 69% of children 15 years or younger diagnosed with pharyngitis received an antibiotic. In contrast, the current analysis revealed that during the 1 year study period, on average, 43% of children diagnosed with pharyngitis received antibiotics. Of these children, on average, only 75% were tested for GAS. This rate should be 100%. Although prescribing rates in the current study were substantially lower than those reported by McCaig et al, there is still sizable room for improvement in the management of this condition given the expected rate of GAS test positivity.

If a health plan had a large population of children aged 5 to 10 years who are more prone to GAS infections by virtue of their age, the plan might have a higher than expected antibiotic prescribing rate for pharyngitis (>25%-30% of cases) despite appropriate testing. Alternatively, if the majority of a health plan's eligible cases occurred during the GAS season (November-March), they might also have a higher than expected antibiotic prescribing rate for pharyngitis. However, seasonal adjustment and age adjustment lead to similar findings among health plans.

Reliance on administrative data for quality of care performance measurement is limited by potential inaccuracies of diagnostic coding in these databases. Thus we performed a medical records validation study to examine the degree to which diagnostic inaccuracy might have influenced performance on this measure. We found that the administrative data had a high positive predictive validity for correctly identifying pharyngitis cases. In only 12% of cases, the administrative data indicated a diagnosis of pharyngitis that could not be verified in the medical record on the date of service. Similarly, in their medical records validation study of administrative data from 7 health plans in Colorado, Maselli et al found that in 17% of cases, administrative data indicated a diagnosis of pharyngitis that could not be confirmed in the medical record.

In 10% of cases overall, visits identified using administrative data as having no competing bacterial diagnosis assigned on the episode date, had a secondary bacterial diagnosis according to the medical record. This might be problematic because this performance measure assumes that the antibiotic is prescribed for the pharyngitis diagnosis. In cases for which a secondary bacterial diagnosis was not coded for the visit, it may have been appropriate for the physician to prescribe an antibiotic for that condition, (eg, acute otitis media, and not pharyngitis), negating the need for a GAS test. This would suggest that the benchmark for this measure should more accurately be placed around 90% rather than 100%, ie, a health plan's performance would be considered very good on this measure if they performed GAS testing in 90% of children diagnosed with pharyngitis and prescribed an antibiotic. Using 90% as the benchmark, only 2 of the health plans participating in the validation study (plans D and E) demonstrated good performance on the measure (Table 1). If adopted as a HEDIS measure, it is likely that health plans would have a strong incentive to improve the accuracy of coding for pharyngitis visits with secondary bacterial diagnoses and that the benchmark could eventually be set closer to 100%.

About 7% (29/427) of antibiotic prescriptions identified in the health plans' pharmacy databases could not be validated in the medical record. Maselli et al similarly found that in 10% of cases, antibiotic prescriptions identified in pharmacy databases could not be confirmed in medical records. In the current study, for 76% (22/29) of these cases the antibiotic prescription was filled 1 to 3 days after the index visit. In 90% (20/22) of the cases where there was a delayed prescription, a throat culture had been done. Based on these results, a majority of these cases likely represent prescriptions phoned in to a pharmacy after a culture was found to be positive. These delayed prescriptions may not be well documented in the medical record.

Health plan B had an exceptionally low antibiotic prescribing rate (9%) for pharyngitis (Table 1). However, we do not believe this is an erroneous result, as the concordance rate between antibiotic prescriptions identifi-
ified using administrative data vs medical records data was 90% in this health plan. This particular plan also had exceptionally low antibiotic prescribing rates for upper respiratory infections (2%) and bronchitis (4%) during the same time period. This health plan’s low antibiotic prescribing rates may represent the effects of quality improvement efforts focused on decreasing inappropriate antibiotic prescribing for upper respiratory infections, which may have led to a spillover effect on pharyngitis prescribing patterns.17

Using administrative laboratory data to identify when GAS testing was performed is limited by the fact that only GAS tests resulting in a claim were counted. Thus, the measure potentially underestimates the number of GAS tests performed. Using the medical record as the gold standard to determine if GAS testing was performed, we found that in 10% of cases, GAS test was documented in the medical record but was not identified in the laboratory claims data. If this measure is adopted into the HEDIS measurement set, it will be imperative that NCQA instruct plans in advance regarding the CPT codes in use. It is likely that health plans would have a strong incentive to convert any plan-specific codes to standard CPT codes to avoid high false-negative rates for appropriate GAS testing and the consequent appearance of having poor performance on this measure.

Based on the results of this validation study, 3 main revisions were made to the measure. First, the expert panel concluded that comprehensively including all potentially relevant secondary bacterial diagnoses was not feasible. Thus they decided to only include visits with a single diagnosis of pharyngitis. However, benchmark performance is still expected to be 90% rather than 100% because the medical records validation study indicated that in 10% of cases, the administrative data indicated a single diagnosis of pharyngitis while the medical records data indicated a secondary bacterial diagnosis. Second, the panel narrowed the length of time for capturing antibiotic prescriptions from 7 days to 3 days following the eligible episode of pharyngitis. This was done because 94% of the prescriptions issued occurred during the 3 days following the episode date. This revision further avoids counting antibiotic prescriptions that may not have been linked to the eligible episode of care. Finally, the panel changed the eligible age range from 2 to 18 years of age to 2 to 16 years during the measurement year. The panel felt that adolescents older than 16 years could be treated similarly to adults18 and clinical criteria could be used to diagnose GAS pharyngitis rather than a laboratory test.

The Committee on Performance Measurement at NCQA is the final decision-making body regarding inclusion of quality measures in the HEDIS measurement set. In addition to the changes made by the expert panel, the Committee on Performance Measurement revised the final measure specifications to include emergency department visits for pharyngitis to obtain a more complete representation of how this condition is managed in the outpatient setting. They also deleted the requirement that no additional outpatient visits with a bacterial diagnosis occur during the 7 days before or after the pharyngitis visit date. This was done because in the validation study, less than 1% of pharyngitis visits were excluded for this reason (Figure).

LIMITATIONS

Our validation study has several limitations. Because we did not sample upper respiratory infection cases without a diagnosis of pharyngitis in the administrative data, we cannot assess the sensitivity of the administrative data for detecting all pharyngitis diagnoses noted in medical records. It is certainly possible that some cases of pharyngitis are assigned ICD-9 codes other than the ones used to identify these cases in the current study. To our knowledge, there are no studies that have evaluated the sensitivity of administrative data for identifying pharyngitis diagnoses specified in the medical record. Investigations examining the sensitivity of administrative data for identifying various other diagnoses have found sensitivities in the range of 57% to 93% with the majority reporting values greater than 80%.19-22 For the current study, we can only comment on the correctness or positive predictive validity of the administrative data for identifying pharyngitis diagnoses using the medical record as the gold standard for diagnosis. However, given that the population-based rate of this diagnosis in our administrative data (216 pharyngitis cases per 1000 enrolled members aged 2-18 years) is similar to the national population-based rate of pharyngitis in children and adolescents (140 pharyngitis cases per 1000 population aged 0-15 years in 2000),1 administrative data likely have reasonable sensitivity for identifying cases with this diagnosis.

When comparing administrative data cases with medical records, we only sampled pharyngitis cases for which an antibiotic was prescribed, thus negating our ability to assess the sensitivity of the administrative data for detecting antibiotics prescribed using the medical record as the gold standard for this information. Thus, we can only comment on the concordance of the medical records data with the administrative data for antibiotic prescriptions issued.

Because emergency department visits were not counted when assessing whether any additional visits with bacterial diagnoses had occurred during the 7 days before or after the eligible pharyngitis visit, it is possible that antibiotics prescribed at an emergency department visit during this time frame could have erroneously been linked to the pharyngitis episode.

CONCLUSIONS

Based on the results of this validation study, it is clear that a considerable degree of antibiotic prescribing for children with pharyngitis in the absence of GAS testing continues to occur in diverse medical care settings throughout the United States. This study showed this newly developed measure to be a valid, although imperfect, measurement tool for assessing whether or not GAS testing is performed when antibiotics are prescribed for pediatric pharyngitis. Based on the validation test findings, the revised version of the measure, Appropriate Testing in Children with Pharyngitis, was implemented in the
HEDIS measurement set in 2004. More than 90% of the nation’s health plans report on HEDIS measures. Implementation of this measure should help to focus health plan quality improvement efforts nationwide on the critical public health issue of inappropriate antibiotic prescribing, and more specifically, on using GAS test results to guide antibiotic management decisions for pharyngitis in children.

Accepted for Publication: February 7, 2005.

Correspondence: Rita Mangione-Smith, MD, MPH, University of California, Los Angeles, Department of Pediatrics, 10833 Le Conte Ave, Los Angeles, CA 90095-1752 (ritams@ucla.edu).

Funding/Support: This study was conducted under funding from the Centers for Disease Control and Prevention through a cooperative agreement with the Association of American Medical Colleges, Grant MM-0163-02/02.

Acknowledgment: We thank the members of the NCQA Appropriate Antibiotic Use Technical Subgroup for their invaluable contributions to this project. We are fortunate to be able to draw on the experience, knowledge and expertise of the following individuals: Richard Besser, MD, National Center for Infectious Diseases, Centers for Disease Control and Prevention; Jonathan A. Finkelstein, MD MPH, Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston, Mass; S. Michael Marcy, MD, Kaiser Foundation Hospital, Panorama City, Calif; Samuel Ho, MD, PacificCare Health Systems, Santa Ana, Calif; Arch Mainous, PhD, Department of Family Medicine, Medical University of South Carolina, Charleston; Robert Scalettar MD, MPH, Anthem Blue Cross Blue Shield, North Haven, Conn; Ellen R. Wald, MD, Department of Pediatrics, University of Pittsburgh, Pittsburgh, Pa.

REFERENCES