Reviewer's report

Title: Antibiotics are prescribed inappropriately to adult pharyngitis patients and McIsaac modification of Centor score is the answer to reduce unnecessary antibiotic prescriptions in low socio-economic areas.

Version: 1 Date: 22 June 2012

Reviewer: Richelle Cooper

Reviewer's report:

General Comments

This is an interesting study on an important topic. The overprescription of antibiotics for URIs, and use of broad spectrum antibiotics when a narrow spectrum is appropriate. The context however is also important and not clear, and the question and interpretation seem to confuse primary treatment and goals of secondary prevention.

The main reason to treat GAS is to prevent rheumatic fever. The prevalence of rheumatic fever in industrialized countries has declined dramatically over time, but the disease remains a global health problem in developing countries. It is a disease that is more common in impoverished areas. The guidelines for treatment are based on industrialized regions, and except outbreaks in Salt Lake City, Utah, for most of the US this is not an issue, and in the Scandinavian countries not treating any GAS is probably also fine because the lack of rheumatic fever. However, in developing countries rheumatic fever is a problem and epidemiologic surveys using echo indicate prevalence of rheumatic heart disease is significantly underestimated. Rheumatic fever depends on the strain of GAS and host factors.

The CDC principles and western guidelines are based on prevalence of GAS (and rheumatic fever) and are most relevant for the industrialized countries, immunocompetent adults, without history of rheumatic fever or heart disease. It is not clear to me what the prevalence of rheumatic heart disease in Pakistan, and this is an important consideration in developing and discussing an appropriate strategy. It is not clear what treating guidelines are available or used in Pakistan, thus to compare or expect a Pakistani physician to read a paper that is in English and a guideline for practice in another country is not clear. The rationale should be provided. In addition this data is from tertiary care hospital, thus the reliability and accuracy of these estimates for Pakistan in general is not clear as this is a selected population.

With regards to treatment, primary treatment allows for the benefit of symptoms improvement, patients feel better 12-24 hours sooner, with abx assuming all patients are given antipyretics and other pain medications. To achieve this benefit, patients need to be treated early in the course of their illness and at the
time of MD assessment, before culture results are known. For secondary prevention, that is to treat to prevent RHD, one can wait for culture results. The current paper seems to confuse or fails to clarify these issues as it pertains to the definition of appropriate use of abx.

With regards to abx, a narrow spectrum abx is the right choice. But many ID physicians in the US advocate for cephalexin as first line, not second line as in this paper, and in the US erythromycin resistance is high and usually not the best choice for someone with a penicillin allergy.

Finally, the report is not as clear about objectives as it could be with an overemphasis on statistical testing and testing that seems inappropriate for the question asked. However, I admit the manner in which the question is stated leaves some confusion. Association is not correlation, and neither of these refer to accuracy. If prevalence is the outcome, and frequency of RX use and who it is used in, then simple descriptive statistics without statistical testing is appropriate.

Specific comments:

Major compulsory

1. Methods, page 5. Study population. This section describes the setting but is not clear about how patients were identified and approached. It is not clear if these are emergency department patients at the tertiary care center or urgent care patients, or what specifically was the site. The section is labeled study population, but it is also not just about the study population it describes the data collection. I would suggest reorganization and better use of subheadings.

2. Methods, study population. There are 2 groups of subjects in this study. The patients, and it is not clear if you excluded patients with a history of rheumatic fever, or heart disease. The second group of subjects are the physicians as this is really a study about MD diagnosis, and MD prescribing behavior. It is not clear if the patients and MDs were consented, the unit of analysis is the MD not the patient.


3. Methods, The time frame (month A, year 20BB to month C, year 20DD) is not reported. This might be useful in the section on study design.

4. Methods, data collection. It is not clear if the treating MD collected the clinical data and the culture or if this was done by a research assistant. Were the treating MDs aware that a culture would be performed, and were the patients contacted about results (and told to stop antibiotics if the culture was negative)?

5. Methods, page 6, definitions. The definition of inappropriate treatment and
scoring of the MD behavior is based on culture results, which seems to be an issue of who should or should not be treated for the secondary prevention. We have no information about the patients assessed with pharyngitis, but not considered for the study. The objectives never discuss what the loss function the authors will tolerate. That is, there is a trade off between sensitivity and specificity, even with a clinical tool, so we expect some under and overtreatment. What is an acceptable amount? The issue of which abx are appropriate is discussed in general comments.

6. Methods, page 7. Study outcomes and statistics. The outcomes are not stated and the analysis should address the a priori stated objectives but it seems there is an emphasis on statistical testing and use of tests that are not appropriate. The primary outcome is the prevalence of GABHS infection in this population, then secondary outcome is frequency of rx, then appropriateness of the rx. Separately there seems to be another outcome/analysis to assess the diagnostic accuracy of the modified Centor criteria. I see no reason to then look for association, chi-square or regression. The diagnostic performance with Sn, Sp, NPV, and PPV, describing the under and overtreatment (FN and FP) seems most important. In addition, as noted above the unit of analysis is the provider and thus the data by provider should be considered and reported, and confidence intervals adjusted for clustering on provider. The report would be improved by focusing on the descriptive data, rather than statistical testing.

7. Results para 1 should first describe enrollment. During X months of data collection, Y subjects with pharyngitis were evaluated, Z were thought to be bacterial pharyngitis and enrolled in the study? There were A number of treating MDs during that study time, and the median number of patients per subjects was B (range C-D), the treating MDs were attendings or residents and were emergency physicians or general practitioners or internists?

8. Limitations. The current study is missing a section on limitations. The authors need to realize that their critics will expound on the limitations of this study in forums where the authors may have limited ability to respond. Thus, the limitation section of this article provides a "golden opportunity" to present a realistic and rationale discussion of these issues. The authors need to discuss the potential for selection bias. The authors need to consider and discuss the differences and potential problems with applying criteria from an industrialized nation to a developing country where most would advocate a different approach secondary to higher incidence of RHD.

Minor essential/discretionary

9. The abstract background lists one aim, but I think the abstract may better reflect the paper if the objectives were clear as it sets up the methods, results and interpretation. The abstract methods are not clear about what the setting is for these patients, or a definition of appropriate. The results is not clear as the N for the subjects is not clear. The results should first report the number of patients, then the number positive GAS, then the number that received abx, then the number that received the various classes. That is, give the basic summary before
reporting validity or other statistical testing. Correlation should be reported with the \( r^2 \) and/or measures of association should present OR and 95\% CIs, not p-values. Validation of the clinical scoring should report the accuracy of the method to provide clarity of the number of false positives and false negatives associated with the strategy.

10. The results correlation is not clear and should not be a p-value. Rather report the number with culture positive treated and those culture negative treated, and stratify the table of demographics (provide data for the sample overall, then columns for those GABHS positive and for those GABHS negative). That is, revise table 1. In addition, in table 1 and the results stratifying by male and female is not necessary as I know of no data to suggest Centor criteria or GABHS is different by sex of the patient.

11. Table 4. Round off the Sn and Sp to appropriate significant figures and provide 95\% CIs, in addition provide the NPV and PPV.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests