# PRImary care Streptococcal Management study (PRISM): identifying clinical variables associated with Lancefield Group A beta-haemolytic streptococci and Lancefield non Group A streptococcal throat infections from two cohorts of patients presenting with acute sore throat



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID: | bmjopen-2013-003943 |
| Article Type: | Research |
| Date Submitted by the Author: | 04-Sep-2013 |
| Complete List of Authors: | Little, Paul; University of Southampton, Primary Care and Population Science;<br>Moore, Michael; University of Southampton Medical School, Primary Care Medical Group<br>Hobbs, Richard; University of Oxford, Dept of Primary Health Care<br>Mant, David; University of Oxford, Primary Health Care<br>McNulty, Cliodna; PHE,<br>Williamson, Ian; University of Southampton, Primary Care and Population Science<br>Cheng, Edith; University of Southampton, Primary Care and Population Science<br>Stuart, Beth; University of Southampton, Primary Care and Population Science<br>Kelly, Jo; University of Southampton, Primary Care and Population Sciences<br>Barnett, Jane; University of Southampton, Primary Care and Population Science<br>Mullee, Mark; University of Southampton, Primary Care and Population Science |
| Primary Subject Heading: | General practice / Family practice |
| Secondary Subject Heading: | Infectious diseases |
| Keywords: | Microbiology < BASIC SCIENCES, Diagnostic microbiology < INFECTIOUS DISEASES, PRIMARY CARE |
| | |

# PRImary care Streptococcal Management study (PRISM): identifying clinical variables associated with Lancefield Group A beta-haemolytic streptococci and Lancefield non Group A streptococcal throat infections from two cohorts of patients presenting with acute sore throat

Paul Little, Mike Moore, F.D.R Hobbs, David Mant, Cliodna McNulty, Ian Williamson , Edith Cheng, Beth Stuart, Joanne Kelly, Jane Barnett, and Mark Mullee on behalf of the PRISM investigators.

PRISM Investigators: University of Southampton: Paul Little, Ian Williamson, Mike Moore, Mark Mullee, Man Ying Edith Cheng, James Raftery, David Turner, Rafael Pinedo-Villanueva, Jo Kelly, Jane Barnett, Karen Middleton, Gerry Leydon; University of Oxford : Richard Hobbs, Richard McManus, David Mant, Paul Glasziou, Sue Smith, Diane Coulson; Health Protection Agency: Cliodna McNulty, Peter Hawtin

**Correspondence to Professor Little p.little@soton.ac.uk**
**Tel +44 2380 241050; fax +44 2380 701125**
**University of Southampton**
**Aldermoor Health Centre, Aldermoor close, Southampton UK**
**SO16 5ST**
**Word count 3020**

**Abstract.**

**Objective.** To assess the association between features of acute sore throat and the growth of streptococci from culturing a throat swab.

**Design.** Diagnostic cohort

**Setting.** UK General practices

**Participants.** Patients age 5 or over presenting with acute sore throat. Patients were recruited for a second cohort (cohort 2, n=517) consecutively after the first ( cohort 1, n=606) from similar practices.

**Main outcome.** Predictors of the presence of Lancefield A/C/G streptococci

**Results.** The clinical score developed from cohort1 had poor discrimination in cohort2 (bootstrapped estimate of area under ROC curve (AUC) 0.65), due to poor validity of the individual items in the second data set. Variables significant in multivariate analysis in both cohorts were: rapid attendance (prior duration 3 days or less; multivariate adjusted odds ratio 1.92 cohort,1.67 cohort2); fever in the last 24 hours (1.69,2.40); and doctor assessment of severity (severely inflamed pharynx/tonsils (2.28, 2.29). Absence of coryza or cough and purulent tonsils were significant in univariate analysis in both cohorts and in multivariate analysis in one cohort. A 5 item score based on Fever, Purulence, Attend rapidly (3 days or less), severely Inflamed tonsils, and No cough or coryza (acronym FeverPAIN) had moderate predictive value (bootstrapped area under ROC curve 0.73 cohort 1, 0.71 cohort 2) and identified a substantial number of participants at low risk of streptococcal infection (38% in cohort 1, 36% in cohort 2 scored <=1, associated with a streptococcal percentage of 13% and 18% respectively). A Centor score of <=1 identified 23% and 26% of participants, with streptococcal percentages of 10% and 28% respectively.

**Conclusion.** Items widely used to help identify streptococcal sore throat may not be the most consistent. A modified clinical scoring system (FeverPAIN) which requires further validation may be helpful clinically in identifying individuals who are unlikely to have major pathogenic streptococci.

**Article focus.** The aim was to assess which variables were most consistently associated with Lancefield groups A, C and G and to develop a clinical score to predict the presence of these streptococci among patients presenting in primary care with acute sore throat.

**Key messages**

- The study confirms that Lancefield groups C and G are not uncommon, and that items widely used to help identify presentations of streptococcal sore throat in primary care may not be the most reliable.
- A modified clinical scoring system (acronym FeverPAIN, a simple count of variables: Fever in the last 24 hours, Pus, Attend rapidly (<=3 days), severe Inflammation of tonsils, and No cough or coryza) may be helpful clinically in identifying individuals who are unlikely to have major pathogenic streptococci.

**Strengths and limitations.**

- This was one of the largest studies to date to develop a clinical score for streptococcal infection
- Two data sets were used to determine the most consistent variables
- Bootstrapping techniques were used to limit over-fitting, but the score requires further validation

**Background**

Antibiotic resistance is a major public health problem driven largely by antibiotic prescribing in primary care[1;2], and it is important to minimise antibiotic use in patients who will not benefit [3]. However, antibiotics are still prescribed in the majority of patients with acute sore throat, the commonest upper respiratory infection to present in primary care[4].

Management of acute sore throat is often based on features associated with Lancefield Group A beta-haemolytic streptococci (GABHS), and clinical scores to predict GABHS have some promise to be useful [5 6 7] including the simple 'Centor'criteria - 3 out of 4 of pus, cervical nodes, a history of fever and no history of cough. These criteria are widely advocated in clinical practice guidance [8-12] have some validation in large routinely collected data sets[13], and are probably reasonably calibrated[14]. However concern has been raised about their use in low prevalence settings such a primary care[14] and these criteria have low specificity[9] leading to high rates of overall antibiotic use[9]. Furthermore small studies in typical primary care settings have suggested other features might be useful in refining the criteria - such as shorter prior duration, severity of pain, and muscles ache[15 7]. The issue of which variables most strongly predict streptococcal infections is therefore still not settled.

We previously reported evidence that group C and G streptococci present in a similar manner to Group A[16], and found that some of the variables which comprise very commonly used clinical prediction rules (such as purulence) might not be significant, and other variables not commonly used might be important (such as speed of presentation, severity of inflammation). This suggests confirmation is needed regarding which variables are important and the need to assess a wide range of potential variables in different data sets.

We compare findings from a new cohort with the original cohort[16] regarding the predictors of the presence of pathogenic streptococci including Group A, C and G in throat swab cultures from patients presenting with sore throat in primary care.

## Methods

This study was designed to assess both the validation characteristics of five widely available rapid streptococcal antigen tests (not reported here), and to assess which clinical variables were associated with streptococcal infection. The inclusion criteria, clinical data collection and the collection and transport of swabs have been described previously[16], but will be summarised.

**Inclusion:** The target group were patients aged 5 and over presenting to primary care clinicians with acute sore throat (< 2 weeks) where the sore throat was the predominant clinical feature (or where the clinician felt that the pharyngitis was driving the illness presentation) and with an abnormality on examination (erythema or pus of the throat - similar to a previous study in primary care[17]). Exclusions were: where the clinician judged there were other causes of sore throat (e.g. apthous ulceration, candida, drugs), or unable to consent (e.g. dementia, uncontrolled psychosis).

**Clinical Data**: Following informed consent baseline clinical data were collected by the health professional. The clinical proforma collected information on age; gender; current smoking status; past history of quinsy; data on symptom severity for the symptoms of sore throat, difficulty swallowing, fever, cough, coryza ('runny nose') headache, muscle ache, abdominal pain, diarrhoea vomiting, earache (each symptom was rated 0=no problem 1= slight problem 2=moderately bad problem 3=severe problem); and examination for oral temperature using Tempadot thermometers[18], the severity of tonsillar and pharyngeal inflammation, the presence of cervical glands, tonsillar exudate, fetor and palatal oedema. Patients then completed a daily symptom diary until symptom resolution (not reported here).

**Throat swabs.** A throat swab was sent to a central laboratory, where culture and sensitivity were performed for all significant pathogens in line with National Standard Operating Proceedures[19;20]. Mean time between specimen collection and receipt at laboratory was 2.9 days (data incomplete for 13 samples). A swab was inoculated onto a blood agar plate and staph/strep agar plate (E&O Laboratories Ltd, Bonnybridge, Scotland) and spread for single colonies. Plates were incubated anaerobically for 48 hours[19;20]. Plates were read after 24 hours incubation and negative cultures reincubated for an additional 24 hours. Suspected beta-haemolytic streptococcal isolates were identified via visual analysis of colony morphology and Lancefield grouping (PathoDx Strep Grouping Kit, Oxoid, UK), in accordance with the National Standard Operating Procedures[19;20]. Antibiotic sensitivities were conducted using disc diffusion techniques. [21].

**Sample size.** In order to determine the association of clinical variables with streptococcal infection, assuming that at least one third of individuals would have streptococci (based on our first data set), and that variables in the streptococcal group were found in 30-80% of individuals, then to detect a variable with an odds ratio of 2 required 407 individuals with complete results.

**Analysis. Primary analysis**. Our original intention was to use a traditional 'sequential'approach to the development and validation of clinical scores - to develop them in one data set, and due to the problem of over-fitting in one data set, to then validate in another data set. However, some variables

included in score from the first data set did not perform well in the second data set (severity of sore throat, cervical glands) and some variables not included in the first score were significant in the second data set (fever, pus). This poor consistency resulted in very poor discriminatory performance of the first clinical score when used in the second data set. Since one data set was clearly was insufficient to identify consistent variables we used both data sets to identify variables, and used bootstrapping to overcome the problem of over-fitting. We have shown previously from the first data set that patients with group C and G beta-haemolytic strains presented with similar clinical features to individuals with Group A beta-haemolytic strains. We first assessed the independent clinical features associated with combined group A , C and G streptococci in the both data sets. Clinical variables were included in a logistic regression model to assess their association with the presence of Lancefield group A,C, and G streptococci. Forward selection was used: variables were included if significant at the 10% level and retained in multivariate analysis if they remained significant at the 5% level. Missing variables were not imputed. Continuous variables were dichotomised using previous cut-offs (age 10 or less; prior duration longer than the median of 3 days[7]). Given the much higher asymptomatic carriage rates of streptococci in children[22] we did not include age in the final multivariate models. Clinical scoring systems or clinical prediction rules are most likely to be useful if they are simple to remember and use, which suggests few variables should be used - preferably using a simple count of the predictive variables. We estimated the increase in area under the ROC curve starting with the most predictive variables, with the aim of maximising the area under the curve without including unnecessary variables, and generated a basic model using variables that were significant in multivariate analysis in both data sets. However a clinical score using very few variables will potentially limit the grading of risk (since there will be fewer categories) and variable performance of one item in different cohorts will unduly affect validity. Hence we also generated a score to include variables that were significant in univariate analysis in both data sets and multivariate analysis in at least one of the data sets .

Because any new model developed from a single data set may be over-fitted, bootstrapped estimates are provided for the area under the ROC curve for internal validation for the new model (see Table 2)[23]. For the Centor criteria (an established model) non boot-strapped estimates are provided.

**Calibration**. We assessed calibration of the scores by assessing the differences between observed and expected percentages of streptococci using the Chi squares test.

**Secondary analyses.** We also present the results of alternative analyses a**) the sequential approach** of generate a prediction rule in one data set and validate it in the second, or b**) the combined approach:** to use a combined data set for greater power.

We also explored other approaches to variable selection such as using the criterion of being significant in univariate analysis in both data sets (which resulted in the same variable selection), explored the impact of variable omission and substitution, and assessed the discrimination comparing

the model having the exact logistic coefficients for each variable with the simple clinical score (i.e. a score which comprised a simple count of the variables).

**Results.**

The recruitment rate was estimated during the recruitment of the first cohort: the median recruitment (the number of patients/months recruiting) was 4.7 patients per month – close to the expected rate from national data[24] and that patients from higher recruiting doctors (higher than the median - average 11.8 patients per month) compared with lower recruiting doctors (an average 2.6 patients per month) had very similar number of features that predict streptococcal infections (respectively a mean of 3.3 features and 3.4 features using the streptococcal score developed from the first data set), suggesting little or no recruitment bias based on clinical characteristics in these practices . Both cohorts used very similar practices: for the first data set we recruited patients from 15 practices, for the second data set 12 of these 15 practices participated.

Patients were recruited from January 2007 until October 2008 (96% of patients were recruited after Jan 2008 when the first data set was completed). All 517 patients recruited in the second data period had some useable data, and complete data was available in 460 patients. In the second data set pathogenic streptococci were found in 207 (40%), mainly A (143), C (30), and G (20) with some B (9) D (2) and F (3) . These are very similar figures to the first data set (ref BJGP paper).

**Primary analyses**

The independent variables associated with Lancefield group A, C or G streptococci in the second data set are shown in Table 1 , with the uni- and multivariate odds ratios also reported from the first data set for ease of comparison. The clinical features predicting the presence of group A, C or G beta-haemolytic streptococci significantly in multivariate analysis in both data sets were rapid attendance (a short prior illness duration of 3 days or less; multivariate adjusted odds ratio in the first data set 1.92; 1.67 in the second data set), fever in the last 24 hours (odds ratios 1.69 and 2.40 respectively) and doctor assessment of severity of inflammation (severely inflamed tonsils: 2.28; 2.29). Additional variables significant in univariate analysis in both data sets and significant in multivariate analysis in at least one of the data sets were items suggesting a purely pharyngeal illness (the absence of coryza and the absence of cough), purulent tonsils, and muscle aches. 'Absence of coryza' performed only marginally better than 'absence of cough' in the two data sets, so based on the similarity of these items and their performance, the helpful concept for clinicians of a purely oro-pharyngeal illness (i.e. when both cough and coryza are absent) and the prior extensive use of 'absence of cough' in the Centor criteria, the consensus among the study team was to use the combined variable 'absence of cough or coryza' which also performed marginally better than either alone.

Table 2 shows the incremental performance in terms of area under the ROC curve (AUC) as successive variables are added to the models in both data sets. There is modest improvement in AUC after the first three variables are added, and no improvement when the 6th variable (muscle aches) is added. However if a basic score (model 3) is used the grading of risk at lower scores is crude as few patients can be categorised as lower risk: only 19% of the first data set and 22% of the second data set score 0 and respectively 15% and 22% of these groups have streptococci (see Appendix 1 for full table).

A Centor score of <=1 was identified among 23% in the first cohort and 26% of the second cohort and streptococcal percentages were isolated in 10% and 28% of these groups respectively (see Table 3). By comparison the extended five point FeverPAIN (model 5 from Table2) provides a finer grading of risk and significantly more patients in both cohorts can be categorised as low risk of streptococcal infection (<20% chance of streptococci, see Table 3): using the modified FeverPAIN score there were more than 30% of patients scoring <=1 (first data set 38%; second data set 36%) and fewer of these patients (13%, and 18%) respectively had streptococci, shown graphically in figure 1.

**Calibration.** FeverPAIN calibrated well in both data sets, with no significant differences between the percentages of observed and predicted presence of streptococci. The calibration of the Centor criteria was good in the first data set but poor in the second data set with significant differences between observed and predicted at low scores (See Appendix 2).

The 'sequential' approach of developing the first score then testing it in the second data set demonstrated poor performance of the first score in the second data set. The approach of using a combined data to set to provide more power generated an 8 item unwieldy score, and obscured the major differences in performance between the data sets (See Appendix 3 for details).

**Discussion.**

This study provides evidence to confirm that streptococcal sore throats are currently common in primary care and that Lancefield groups C and G make up a quarter of streptococcal sore throats. The study also confirms that the best predictors of streptococcal infection may not include some of the features traditionally used, and that traditional scoring systems may have limited clinical utility in identifying individuals who have a low likelihood of streptococcal infection i.e. who do not need to have antibiotics.

**Strengths and limitations of the study.**

These data sets are some of the largest from a typical primary care setting to have assessed the importance of the range of streptococci, and to explore the range of potential clinical predictors of streptococcal infection. There were few missing data (less than 5% for any analysis), and little evidence of recruitment bias either in recruitment rates or clinical characteristics. The conventional approach to develop and validate a diagnostic model is to develop it in one data set and test it in another. However, the variability of performance of variables in these data sets and the poor discriminatory performance of the first score when used in the second data set suggests that such an approach is unlikely to provide the most reliable method of variable selection for a clinical prediction rule, supported by similar findings in the development of clinical prediction rules for other acute infections[25]. This suggests that the choice of variables to include in clinical prediction models preferably should be based on multiple cohorts at different times and/or different settings. The alternative approach of combining data sets to increase power generated an 8 variable score with improved discrimination but is unwieldy for clinical purposes. The combined data set also hid the considerable variability between data sets in performance of both individual variables and also of the first score. Further support for the poor clinical utility of the first score comes from the trial (reference trial paper here) which demonstrates that using the first score does not significantly improve outcomes, similar to a previous trial of the Centor criteria which also demonstrated no impact on antibiotic use[26]. Over and above the most basic model (short prior duration, severe inflammation, fever) the choice of additional variables to include (pus and 'absence of cough and coryza') was determined by consensus, including a consideration of the strength of prior evidence, but omission of key variables or substitution did not have major effects on the discrimination. Although we have provided boot-strapped estimates of the area under the ROC curve to limit over-fitting, nevertheless the proposed model should have further validation.[15]

**Main findings in the context of previous literature.**

Group A beta-haemolytic streptococci have dominated previous literature due to their association with major non suppurative adverse outcomes - particularly Rheumatic fever and Glomerulonephritis[5]. Hence the clinical predictors of group A infection [5 6 7] - especially pus, cervical nodes, a history of

fever and no history of cough - have been widely used in clinical guidelines[8;10;22]. Trials using these as inclusion criteria may have larger effect sizes for antibiotics than trials using less selected patients – although the validity of historical comparisons is questionable[27]. We were unable to confirm the importance of cervical glands as a predictor of streptococcal infection in the second data set, and in the first data set were unable to confirm the importance of purulence[7;9]. From these two data sets the features that may be most important are the speed of presentation (i.e. symptoms developing rapidly resulting in short prior duration of illness), the severity of tonsillar inflammation, and fever. These variables have been identified in studies from typical primary care settings [7;15] but previous studies have been limited by lack of multivariate analysis or limited power.

**Clinical utility.** Scoring systems are most helpful clinically for reducing antibiotic use if they identify as large a group as possible of individuals unlikely to have streptococcus. From these data sets the Centor criteria are likely to identify relatively few such individuals who do not have streptococci: only 23% in the first data set and 26% in the second data set had a score <=1 and in the second data set the percentage of streptococci were high (28%). A low count (<=1) using a modified score (FeverPAIN) identified more than 35% of patients in both data sets who are unlikely to have streptococci (between 13 and 18%).

**Conclusion.**

Items traditionally used to help identify presentations of streptococcal sore throat in primary care may not be reliable. Conventional clinical scoring systems may not be very helpful clinically in identifying individuals who are unlikely to have major pathogenic streptococci. A modified clinical rule developed for targeting Lancefield groups A,C and G streptococci requires further validation, but should enable clinicians to both target those at high risk of streptococcal infections and identify more than one third of those presenting with sore throat as being at low (<20%) risk of streptococcal infection.

**Figure Legend:**

**Figure 1**. Comparison of low scores (<=1) for Centor criteria and FeverPAIN.

**Competing interest**

There are no competing interests.

All authors have completed the Unified Competing Interest form at http://www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work ; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years , no other relationships or activities that could appear to have influenced the submitted work.

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences to exploit all subsidiary rights, as set out in our licence ( http://resources.bmj.com/bmj/authors/checklists-forms/licence-for-publication)

**Data sharing**
The Authors are happy to share data and collaborate with other investigators as appropriate (for example in larger merged individual patient data studies).

**Funding**
This study was funded by the HTA Programme (study number 05/10/01) and approved by a Multicentre Research Ethics Committee (number 06/MRE06/17).

**Contributorship statement**
Razia Meer-Baloch (senior trial manager), day to day coordination of the Birmingham study centre, and commented on drafts of the papers.
Edith Cheng. (Study Statistician, University of Southampton) developed the protocol, contributed to quantitative analysis, and drafting of the paper.
Paul Glasziou (GP and Professor of Primary care, University of Oxford) developed the protocol for funding, contributed to management of the clinical studies, commented on the paper.
F.D.R. Hobbs (GP and Professor of Primary Care, University of Birmingham) developed the protocol for funding, contributed to management of all studies, supervised the Birmingham study centre and contributed to the drafting of the paper.
Jo Kelly and Jane Barnett (senior trial managers, University of Southampton) developed the protocol, provided day to day overall management of the study, coordinated recruitment in the lead study centre and coordination of other centres, commented on drafts of the paper.
Gerry Leydon. (Social Scientist; Principal Research Fellow, University of Southampton), developed the protocol for funding, contributed to management , and commented on drafts of the paper.
Paul Little (GP and Professor of Primary Care Research, University of Southampton) had the original idea for the protocol, led protocol development and the funding application, supervised the running of the lead study centre and coordination of centres, contributed to the analysis, led the drafting of the paper.
David Mant (Professor of General Practice (now Emeritus Professor of General Practice), University of Oxford) developed the protocol for funding, supervised the running of clinical studies in the Oxford centre and contributed to the analysis and the drafting of the paper Richard McManus (GP and Professor of Primary Care, University of Birmingham) developed the protocol for funding, contributed to management of all studies, supervised the Birmingham Network and contributed to the drafting of the paper.
sLisa McDermott. (Social Scientist; Research Assistant, University of Southampton), developed the protocol, and commented on drafts of the paper.

Cliodna McNulty (consultant microbiologist for the HPA) and Gemma Lasseter developed the protocol and contributed to the management and write up of the study.
Peter Hawtin (consultant microbiologist for the HPA) developed the protocol for funding, contributed to the design and running of the in invitro and diagnostic phases of the study.
Karen Middleton (data manager University of Southampton). Provided administrative support, developed data management protocols, coordinated data entry, and commented on drafts of the paper.
Michael Moore (GP and Reader in Primary Care, University of Southampton), developed the protocol for funding, contributed to the management of the study, contributed to the analysis and to the drafting of the paper .
Mark Mullee (Lead Study Statistician, Director Research Design Service, University of Southampton) developed the protocol for funding, contributed to study management, supervised data management, led the quantitative analysis and contributed to the drafting of the paper
Ian Williamson (GP and Senior Lecturer in Primary Care, University of Southampton), developed the protocol for funding, contributed to the management of the study and drafting of the paper.
Sue Smith, Mary Selwood, and Diane Coulson (trial managers, University of Oxford) provided day to day coordination of the Oxford study centre and commented on drafts of the paper.
We are grateful to all the patients and Health Care Professions who have contributed their time and effort and helpful insights to make PRISM possible.

**Table 1. Second data set: clinical variables in patients with Lancefield Group A, C and G compared with patients having no growth of Lancefield C,G or A streptococci, with odds ratios (95% CI).** Estimates of the uni- and multi-variate analysis in the first data set are also shown for comparison

| | Second Data set | | | | First Data set | |
|---|---|---|---|---|---|---|
| | With Streptococci | No streptococci | Univariate Odds Ratio | *Multivariate Odds Ratio | Univariate Odds ratio | Multivariate Odds ratio |
| Prior duration <=3 days | 102/176 (58%) | 126/308 (41%) | 1.99 (1.37 to 2.90) | 1.67 (1.10 to 2.54) | 2.64 (1.82 to 3.82) | 1.92 (1.26 to 2.92) |
| Cervical glands | 150/188 (80%) | 245/318 (77%) | 1.18 (0.76 to 1.83) | 1.20 (0.67 to 2.16) | 4.27 (2.41 to 7.57) | 2.93 (1.55 to 5.52) |
| Severely inflamed tonsils | 38/167 (23%) | 23/294 (8%) | 3.47 (1.99 to 6.07) | 2.29 (1.23 to 4.26) | 3.62 (2.32 to 5.64) | 2.28 (1.39 to 3.74) |
| Absence runny nose (coryza) | 149/193 (77%) | 197/323 (61%) | 1.58 (1.22 to 2.05) | 1.91 (1.21 to 3.00) | 2.17 (1.48 to 3.17) | 1.55 (0.99 to 2.41) |
| Age group <=10 years | 12/176 (7%) | 18/308 (6%) | 1.18 (0.55 to 2.51) | 0.80 (0.35 to 1.83) | 2.54 (1.50 to 4.29) | 1.95 (1.05 to 3.62) |
| Very bad sore throat | 167/193 (87%) | 283/323 (88%) | 0.91 (0.53 to 1.54) | 1.08 (0.44 to 2.68) | 4.16 (1.75 to 9.87) | 3.31 (1.24 to 8.83) |
| Absence of cough | 127/193 (66%) | 167/324 (52%) | 1.81 (1.25 to 2.61) | 1.11 (0.70 to 1.75) | 3.83 (2.35 to 6.25) | 2.73 (1.56 to 4.76) |
| Purulent tonsils | 98/192 (51%) | 93/323 (29%) | 2.58 (1.78 to 3.74) | 1.75 (1.13 to 2.72) | 2.51 (1.75 to 3.60) | 1.06 (0.67 to 1.66) |
| Fever (last 24 hours) | 137/193 (71%) | 168/324 (52%) | 2.27 (1.55 to 3.32) | 2.40 (1.52 to 3.77) | 2.80 (1.86 to 4.21) | 1.69 (1.05 to 2.71) |
| Muscle aches | 111/176 (63%) | 150/307 (49%) | 1.79 (1.22 to 2.61) | 1.31 (0.85 to 2.01) | 2.02 (1.39 to 2.94) | 2.20 (1.41 to 3.42) |
| Headache | 128/193 (66%) | 200/323 (62%) | 1.21 (0.83 to 1.76) | 1.15 (0.72 to 1.84) | 2.00 (1.36 to 2.96) | 1.41 (0.89 to 2.25) |
| Absence of cough or coryza | 110/193 (57%) | 137/323 (42%) | 1.80 (1.25 to 2.58) | 1.36 (0.89 to 2.08) | 2.66 (1.85 to 3.81) | 2.45 (1.62 to 3.68) |

*all multivariate estimates adjusted for other significant predictors in each data set.

When assessing the combined variable 'absence of cough or coryza' the individual items are omitted

**Table2. Sequential area under ROC curve as successive variables added (p values given for comparison with previous model unless specified)**

| | 1 | 2 | 3 | 4 | 5 | 6 | Centor |
|---|---|---|---|---|---|---|---|
| Model | Severly inflamed tonsils | + short duration | +fever last 24h | + pus | +no cough or coryza | +muscle aches | Vs model 5 |
| Second data set (p) | 0.575 (0.538, 0.612) | 0.644 (0.596, 0.692) (p=0.006) | 0.689 (0.634, 0.738) (p=0.003) | 0.702 (0.651, 0.748) (p=0.104) | 0.713 (0.661, 0.758) (p=0.803) | 0.708 (0.668 , 0.762) (p=0.334) | 0.650 (0.600, 0.700) (p=0.123) |
| First data set (p) | 0.602 (0.562, 0.641) | 0.676 (0.631, 0.720) (p=<0.001) | 0.706 (0.660, 0.751) (p=0.017) | 0.713 (0.665, 0.761) (p=0.597) | 0.735 (0.690, 0.779) (p=0.025) | 0.738 (0.705, 0.791) (p=0.143) | 0.716 (0.674, 0.758) (p=0.291) |

**Table 3. Number of individual with Lancefield Group A,C, or G streptococci (%) at each level of clinical scores, and the total number of individuals at each level (and % of the total sample).** Two clinical scores are shown: 1) a modified streptococcal score (model5)(5 point score, acronym FeverPAIN: 1 point each for: Fever during the last 24 hours, Purulent tonsils, Attend rapidly (3 days or less), very Inflamed throat, and No cough or coryza) and 2) for comparison the Centor score (1 point each for pus, fever in the last 24 hours, cervical glands and the absence of cough),

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Clinical score | 0 | 1 | 2 | 3 | 4+ | Total |
| **First data set** | | | | | | | |
| | FeverPAIN score | | | | | | |
| Streptococci | | 7 (11%) | 21 (14%) | 45 (30%) | 40 (39%) | 62 (62%) | 175 (31%) |
| Total | | 63 (11%) | 155 (27%) | 149 (26%) | 103 (18%) | 100 (17%) | 570 (100%) |
| | Centor score | | | | | | |
| Streptococci | | 3 (7%) | 10 (11%) | 45 (23%) | 65 (43%) | 55 (57%) | 178 (31%) |
| *Total* | | *45 (8%)* | *88 (15%)* | *199 (34%)* | *152 (26%)* | *97 (17%)* | *581 (100%)* |
| | | | | | | | |
| **Second data set** | | | | | | | |
| | FeverPAIN score | | | | | | |
| Streptococci | | 9 (19%) | 22 (18%) | 46 (35%) | 41 (48%) | 49 (65%) | 167 (36%) |
| Total | | 48 (10%) | 121 (26%) | 130 (28%) | 86 (19%) | 75 (16%) | 460 (100%) |
| | Centor score | | | | | | |
| Streptococci | | 0 (0) | 36 (32%) | 36 (23%) | 69 (50%) | 47 (58%) | 188 (37%) |
| *Total* | | 15(3%) | 114 (23%) | 157 (31%) | 138 (27%) | 81 (16%) | 505 (100%) |

Reference List

(1) Goossens H, Ferech M, Vander Stichele R, et al. **Outpatient antibiotic use in Europe and association with resistance: a cross-national database study.** *Lancet* 2005; 365:579-587.

(2) Costelloe C, Metcalfe C, Lovering A, et al. Effect of antibiotic prescribing in primary care on antimicrobial resistance in individual patients: systematic review and meta-analysis. *BMJ* 2010; 340:doi: http://dx.doi.org/10.1136/bmj.c2096.

(3) SMAC. Standing Medical Advisory Committee (SMAC) report: The Path of Least Resistance. *Occasional Report* 1998.

(4) Petersen I, Johnson A, Islam A, et al. Protective effect of antibiotics against serious complications of common respiratory tract infections: retrospective cohort study with the UK General Practice Research Database. *BMJ* 2007; 335:982 (10 November), doi:10.1136/bmj.39345.405243.BE.

(5) Del Mar C. Managing sore throat: a literature review I: Making the diagnosis. *Med J Austr* 1992; 156:572-575.

(6) Dagnelie C, Bartelink M, Van Der Graaf Y, et al. Towards better diagnosis of throat infections with GABHS in general practice. *B J G P* 1998; 48:959-962.

(7) Dobbs F. A scoring system for predicting Group A streptococcal throat infection. *B J G P* 1996; 46:461-464.

(8) NICE guideline development group. Prescribing of antibiotics for self-limiting respiratory tract infections in adults and children in primary care. *http://www nice org uk/Guidance/CG69* 2008.

(9) McIsaac W, Kellner J, Aufricht P, et al. **Empirical Validation of Guidelines for the Management of Pharyngitis in Children and Adults** . *JAMA* 2004; 291:1587-1595.

(10) Cooper R, Hoffman J, Bartlett J, et al. Principles of appropriate antibiotic use for acute pharyngitis in adults. *Ann Intern Med* 2001; 134:509-517.

(11) Principles of Appropriate Antibiotic Use for Acute Pharyngitis in Adults. *http://www acponline org/clinical_information/guidelines* 2001.

(12) Vincent M, Celestin N, Hussain A. Pharyngitis. *Am Fam Phys* 2004; 69:1465-1470.

(13) Fine A, Nizet V, Mandl K. Large-scale validation of the Centor and McIsaac scores to predict group A streptococcal pharyngitis. *Arch Intern Med* 2012; 172(11): doi 10.1001/archinternmed.2012.95:847-852.

(14) Aalbers J, O'Brien K, Chan W-S, et al. Predicting streptococcal pharyngitis in adults in primary care: a systematic review of the diagnostic accuracy of symptoms and signs and validation of the Centor score. *BMC Medicine* 2011; 9:67 doi:10.1186/1741-7015-9-67.

(15) Lindbaek M, Hoiby E, Steinsholt I, et al. Clinical symptoms and signs in sore throat patients with large colony variant ß-haemolytic streptococci groups C or G versus group A. *BJGP* 2005; 55:615-619.

(16) Little P, Hobbs R, Mant D, et al. Incidence and clinical variables associated with streptococcal throat infections:a prospective diagnostic cohort study. *BJGP* 2012; 62(604):e787-94. doi: 10.3399/bjgp12X658322.

(17) Little PS, Gould C, Williamson.I., et al. Reattendance and complications in a randomised trial of prescribing strategies for sore throat: the medicalising effect of prescribing antibiotics. *B M J* 1997; 315:350-352.

(18) Rogers M. A viable alternative to the glass/mercury thermometer. *Paed Nurs* 1992; 4(9).

(19) National Standard Method: Investigation of throat swabs. *BSOP* 2008; 9(8):http://www.hpa-standardmethods.org.uk/documents/bsop/pdf/bsop9.pdf .

(20) National Standard Methods. Identification of Streptococcus species, Entrococcus species and morphologically similar organisms. *Health Protection Agency* 2007; ID4 (2.1); Available from: http://www.hpa-standardmethods.org.uk/documents/bsopid/pdf/bsopid4.pdf.

(21) Methods for Antimicrobial Susceptibility Testing. *British Society for Antimicrobial Chemotherapy* 2009; Available from http://www.bsac.org.uk/_db/_documents/Version_8_-_January_2009.pdf.

(22) Pelucchi C, Grigoryan L, Galeone C, et al. **Guideline for the management of acute sore throat.** *Clin Microbiol Infect* 2012; 18 s1:1-28.

(23) Harrel FE. Regression Modeling Strategies. With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer Verlag; 2001.

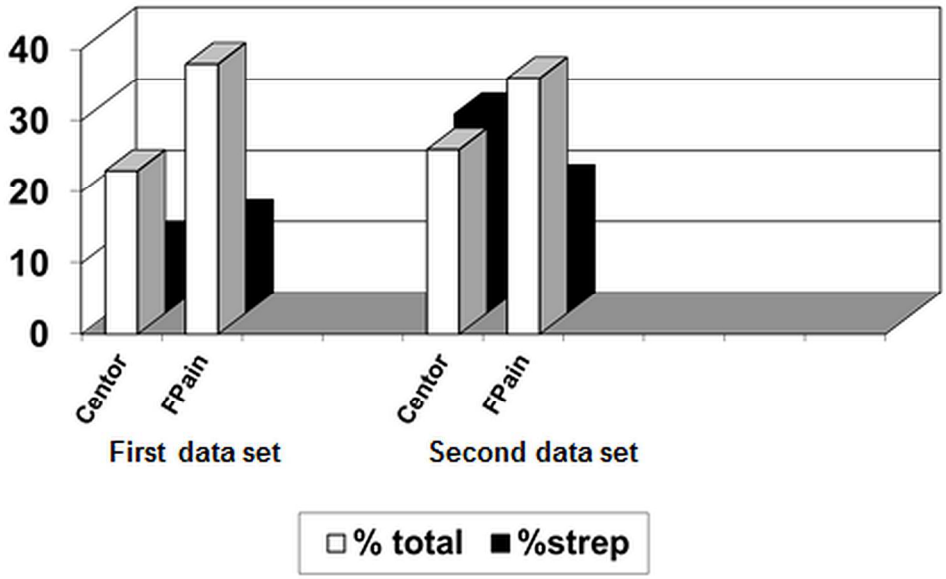
(24) HMSO, OPCS. Morbidity statistics from general practice: Fourth National study 1991. 1 ed. London: HMSO; 1994.

(25) Little P, Turner S, Rumsby K, et al. Urinary tract infection: development and validation, randomised trial, economic analysis, observational cohort and qualitative study. *Health Technol Assess* 2009; 13: No 19.

(26) Worrall G, Hutchinson J, Sherman G, et al. Diagnosing streptococcal sore throat in adults a randomized controlled trial of in-office aids. *Can Fam Physician* 2007; 53:666-671.

(27) Spinks A, Glasziou P, Del Mar C. Antibiotics for sore throat. *The Cochrane Library* 2011; DOI: 10.1002/14651858.CD000023.pub3.

Figure 1. Comparison of low scores (<=1) for Centor criteria and FeverPAIN



**Appendix 1.**

Number and percentage of Lancefield group A,C, or G streptococci according to each level of 3 point 'basic' score (Model 3) including the variables significant in multivariate analysis in both data sets (1 point each for short prior duration, fever in the last 24 hours, and severely inflamed tonsils). The total number at each level, and percentage of the total sample are also shown.[1]

| Score | 0 | 1 | 2 | 3 | Total |
|---|---|---|---|---|---|
| First data set | | | | | |
| Streptococci n (%) | 16 (15%) | 43 (19%) | 80 (43%) | 37 (67%) | 176 (31%) |
| *Total n (%)* | *109 (19%)* | *221(39%)* | *187(33%)* | *55 (10%)* | *572(100%)* |
| | | | | | |
| Second data set | | | | | |
| Streptococci n (%) | 22 (22%) | 46 (25%) | 81 (52%) | 18 (82%) | 167 (36%) |
| Total n (%) | 100 (22%) | 183 (40%) | 156 (34%) | 22 (5%) | 461 (100%) |
| | | | | | |

[1]For example taking the first column at the top of the table: there were 109 individuals with score 0 which represents 19% of the sample, and of those 109 individuals, 16(15%) had Lancefield Group A,C or G streptococci

## Appendix 2. Calibration of Centor criteria and FeverPAIN

**First data set**

| FeverPain | N of cga events/patients | Proportion Observed (95% CI) | Proportion Predicted (95% CI) |
|---|---|---|---|
| 0 | 7/63 | 0.1111 (0.0327, 0.1895) | 0.0898 (0.0609, 0.1305) |
| 1 | 21/155 | 0.1355 (0.0813, 0.1897) | 0.1602 (0.1246, 0.2034) |
| 2 | 45/149 | 0.3020 (0.2279, 0.3761) | 0.2692 (0.2310, 0.3111) |
| 3 | 40/103 | 0.3883 (0.2936, 0.4831) | 0.4158 (0.3656, 0.4677) |
| 4 | 43/75 | 0.5733 (0.4604, 0.6863) | 0.5789 (0.5029, 0.6514) |
| 5 | 19/25 | 0.7600 (0.5888, 0.9312) | 0.7265 (0.6319, 0.8043) |

Chi squared test statistic = 2.33, p=0.6745 – no significant difference between the observed and predicted

| Centor | N with cga (C, G,or A streptococci) events/patients | Proportion Observed (95% CI) | Proportion Predicted (95% CI) |
|---|---|---|---|
| 0 | 3/45 | 0.0667 (0.0001, 0.1405) | 0.0605 (0.0363, 0.0992) |
| 1 | 10/88 | 0.1136 (0.0468, 0.1805) | 0.1229 (0.0887, 0.1679) |
| 2 | 45/199 | 0.2261 (0.1677, 0.2845) | 0.2337 (0.1959, 0.2763) |
| 3 | 65/152 | 0.4276 (0.3486, 0.5067) | 0.3989 (0.3518, 0.4480) |
| 4 | 55/97 | 0.5670 (0.4677, 0.6663) | 0.5909 (0.5117, 0.6656) |

Chi-squared test statistic = 0.92, p=0.8218 – no` significant difference between observed and predicted

**Second data set**

| FeverPain | N with cga events/patients | Proportion Observed (95% CI) | Proportion Predicted (95% CI) |
|---|---|---|---|
| 0 | 9/48 | 0.1875 (0.0756, 0.2994) | 0.1341 (0.0914, 0.1924) |
| 1 | 22/121 | 0.1818 (0.1126, 0.4366) | 0.2194 (0.1732, 0.2738) |
| 2 | 46/130 | 0.3538 (0.3703, 0.5832) | 0.3377 (0.2930, 0.3856) |
| 3 | 41/86 | 0.4767 (0.3703, 0.5832) | 0.4806 (0.4223, 0.5395) |
| 4 | 38/61 | 0.6230 (0.5000, 0.7459) | 0.6268 (0.5415, 0.7048) |
| 5 | 11/14 | 0.7857 (0.5621, 1.000) | 0.7529 (0.6494, 0.8337) |

Chi squared test statistic = 2.42, p=0.6597 – no significant difference between the observed and predicted

| Centor | N with cga events/patients | Proportion Observed (95% CI) | Proportion Predicted (95% CI) |
|---|---|---|---|
| 0 | 0/15 | 0.00 | 0.1436 (0.0938, 0.2137) |
| 1 | 36/114 | 0.3158 (0.2299,0.4017) | 0.2216 (0.1709, 0.2824) |
| 2 | 36/157 | 0.2293 (0.1632, 0.2954) | 0.3260 (0.2827, 0.3724) |
| 3 | 69/138 | 0.5000 (0.4161, 0.5839) | 0.4509 (0.3993, 0.5036) |

| Centor | N with cga events/patients | Proportion Observed (95% CI) | Proportion Predicted (95% CI) |
|---|---|---|---|
| 4 | 47/81 | 0.5802 (0.4718, 0.6887) | 0.5823 (0.4991, 0.6613) |

Chi-squared test statistic = 16.39, p=0.0009 – significant difference between observed and predicted

**Appendix 3. Secondary analyses.**

- **Sequential approach**. This approach uses the first clinical score developed in the first data set, and validates it in the second. The score for the first data set had an Area under the ROC curve (AUC) of 0.759 (95% confidence intervals 0.719 to 0.800) in the first data set and 0.651 (0.600 to 0.702) in the second data set due to the poor performance of the constituent variable in the second data set.
- **Combined datasets.** In the combined data multivariate analysis resulted in all the FeverPAIN variables being significant and also muscle aches, cervical glands and very bad sore throat (i.e. 8 variables to be used in an extended score). This score has an AUC of 0.740 (0.708, 0.773), compared with 0.713 (95% CI 0.681, 0.745) for FeverPAIN, 0.683(0.649, 0.717) for Centor, and 0.710 (0.679 to 0.741) for the first clinical score. The AUC for FeverPAIN is significantly ($p<0.05$) better than for Centor, and the extended score is significantly better than FeverPAIN.

**Exploring the omission of rapid attendance or substitution with muscle aches.** Given concerns that the variable rapid attendance might be less generaliseable to other health care contexts we explored the implications for discrimination of both excluding rapid attendance (FeverPIN), or replacing the A of attendance with A for muscle aches (i.e.FeverPaIN with 'a' for aches) since muscle aches also fulfilled the criterion for inclusion in the score: discrimination was a little lower for FeverPIN (second data set AUC 0.698 (0.649, 0.746), first data set 0.713 (0.668, 0.758)) and similar for Fever PaIN (second data set 0.703 (0.654, 0.751); first data set 0.728 (0.683, 0.773) ).

**The use of a simple score compared with the exact coefficients.** The AUC does not alter much comparing the model with the precise logistic coefficients with the model of the simple rounded score. For FeverPAIN in the first data set the AUCs were 0.735 (0.691, 0.779) vs 0.726 (0.682, 0.770) respectively and the second data set 0.713 (0.661, 0.757) vs 0.700 (0.650, 0.748).

Reporting based on STARD initiative

| | |
|---|---|
| 1. Title, abstract and key words | Page 1-2 |
| 2. Research question or aims | Page 3 |
| 3. Describe participants, inclusion criteria | Page 4 |
| 4. Recruitment mechanisms | Page 4 |
| 5. Participant sampling (e.g. consecutive) | Page 4 |
| 6. Data collection (prospective or retrospective) | Page 4 |
| 7. Describe reference standard and its rationale | Page 4 |
| 8. Describe technical specifications | Page 4 |
| 9. Definition and rationale for cut-off points of index test and standard | Page 4 |
| 10. Describe the number training of staff performing tests and standard | Page 4 |
| 11. Were the readers of the index test blinded ? | Page 4 |
| 12. Describe the methods for calculating or comparing measures, and describing uncertainty | Pages 4-6 |
| 13. Describe methods for calculating reproducibility (if done) | N/A |
| 14.Report when the study was done | Page 7 |
| 15. Report clinical and demographic details | Page 7 |
| 16. Report how many participants did not undergo the index test/standard | Page 7 |
| 17 Report time intervals between index test and standard | Page 4 |
| 18. Report severity of disease in those with and without target condition | Tables1-3 |
| 19. Cross tabulation of index test by results of standard | Table 1 |
| 20. Report any adverse events (test or | N/A |

| | |
|---|---|
| standard) | |
| 21. Estimates of diagnostic accuracy and of uncertainty (CIs) | Table 1 |
| 22. Report how indeterminate results missing responses and outliers were handled | Page 5 |
| 23. report estimates of diagnostic accuracy between readers | N/A |
| 24. report estimates of reproducibility if done | N/A |
| 25. Discuss clinical applicability of study findings | Pages 9-10 |