BMJ Open

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

Paul Little,[1] Michael Moore,[1] F D R Hobbs,[2] David Mant,[2] Cliodna McNulty,[3] Ian Williamson,[1] Edith Cheng,[1] Beth Stuart,[1] Joanne Kelly,[1] Jane Barnett,[1] Mark Mullee,[1] on behalf of the PRISM investigators

To cite: Little P, Moore M, Hobbs FDR, *et al*. PRImary care Streptococcal Management (PRISM) study: identifying clinical variables associated with Lancefield group A β-haemolytic streptococci and Lancefield non-Group A streptococcal throat infections from two cohorts of patients presenting with an acute sore throat. *BMJ Open* 2013;**3**:e003943. doi:10.1136/bmjopen-2013-003943

▸ Prepublication history and additional material for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2013-003943).

Received 4 September 2013
Accepted 3 October 2013

For numbered affiliations see end of article.

Correspondence to
Professor Paul Little;
p.little@soton.ac.uk

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 aged 5 or over presenting with an 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 cohort 1 had poor discrimination in cohort 2 (bootstrapped estimate of area under the receiver operator characteristic (ROC) curve (0.65), due to the 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 OR 1.92 cohort, 1.67 cohort 2); fever in the last 24 h (1.69, 2.40); and doctor assessment of severity (severely inflamed pharynx/tonsils (2.28, 2.29)). The absence of coryza or cough and purulent tonsils were significant in univariate analysis in both cohorts and in multivariate analysis in one cohort. A five-item score based on Fever, Purulence, Attend rapidly (3 days or less), severely Inflamed tonsils and No cough or coryza (FeverPAIN) had moderate predictive value (bootstrapped area under the 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.

Conclusions: 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 clinically helpful in identifying individuals who are unlikely to have major pathogenic streptococci.

ARTICLE SUMMARY

Strengths and limitations of this study

- This was one of the largest studies until now 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 overfitting, 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 β-haemolytic streptococci (GABHS), and clinical scores to predict GABHS have

shown promise to be useful[5–7] including the simple 'Centor'criteria—3 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 as 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 that other features might be useful in refining the criteria—such as shorter prior duration, severity of pain and muscle ache.[7 15] 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 and severity of inflammation). This suggests that 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 groups A, C and G in throat swab cultures from patients presenting with sore throat in primary care.

METHODS

This study was designed to assess not only the validation characteristics of five widely available rapid streptococcal antigen tests (not reported here), but also 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 years and over presenting to primary care clinicians with an 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 (eg, aphthous ulceration, candida and drugs) or unable to consent (eg, dementia and 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; 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] The mean time between specimen collection and receipt at the laboratory was 2.9 days (data incomplete for 13 samples). A swab was inoculated onto a blood agar plate and a staph/strep agar plate (E&O Laboratories Ltd, Bonnybridge, Scotland) and spread for single colonies. Plates were incubated anaerobically for 48 h.[19 20] Plates were read after 24 h incubation and negative cultures reincubated for an additional 24 h. Suspected β-haemolytic streptococcal isolates were identified by a 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 OR 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 overfitting in one data set, to then validate in another data set. However, some variables included in the 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 and 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 insufficient to identify consistent variables, we used both data sets to identify variables and used bootstrapping to overcome the problem of overfitting. We have shown previously from the first data set that

patients with group C and G β-haemolytic strains presented with similar clinical features to individuals with group A β-haemolytic strains. We first assessed the independent clinical features associated with combined group A, C and G streptococci in 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 that few variables should be used—preferably using a simple count of the predictive variables. We estimated the increase in area under the receiver operator characteristic (ROC) curve starting with the most predictive variables, with the aim of maximising the area under the curve (AUC) 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 overfitted, 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 bootstrapped estimates are provided.

Calibration

We assessed calibration of the scores by assessing the differences between the observed and expected percentages of streptococci using the χ^2 test.

Secondary analyses

We also presented the results of alternative analyses (A) the sequential approach of generating a prediction rule in one data set and validating it in the second or (B) the combined approach: using 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), exploring the impact of variable omission and substitution, and assessing the discrimination comparing the model having the exact logistic coefficients for each variable with the simple clinical score (ie, 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/month—close to the expected rate from national data.[24] Patients from higher recruiting doctors (higher than the median—average 11.8 patients/month) compared with lower recruiting doctors (an average 2.6 patients/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, and 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 January 2008 when the first data set was completed). All 517 patients recruited in the second data period had some usable data, and complete data were available in 460 patients. In the second data set, pathogenic streptococci were found in 207 patients (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.[16]

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 univariate and multivariate ORs also reported from the first data set for ease of comparison. The clinical features predicting the significant presence of group A, C or G β-haemolytic streptococci in multivariate analysis in both data sets were rapid attendance (a short prior illness duration of 3 days or less; multivariate adjusted OR in the first data set 1.92; 1.67 in the second data set), fever in the last 24 h (ORs 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 as well as 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 oropharyngeal illness (ie, 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

Table 1 Second data set: clinical variables in patients with Lancefield groups A, C and G compared with patients having no growth of Lancefield groups C, G or A streptococci, with ORs (95% CI)

	Second data set				First data set	
	With streptococci	No streptococci	Univariate OR	*Multivariate OR	Univariate OR	Multivariate OR
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 of 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 h)	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.
Estimates of the univariate and multivariate analyses in the first data set are also shown for comparison.
When assessing the combined variable 'absence of cough or coryza', the individual items are omitted.

Table 2 Sequential area under the receiver operator characteristic (ROC) curve as successive variables are added (p values given for comparison with previous model unless specified)

Model	1 Severely inflamed tonsils	2 +Short duration	3 +Fever in the last 24 h	4 +Pus	5 +No cough or coryza	6 +Muscle aches	Centor 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)

coryza' which also performed marginally better than either alone.

Table 2 shows the incremental performance in terms of area under the ROC curve 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 sixth 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 15% and 22%, respectively, of these groups have streptococci (see online supplementary appendix 1 for the 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 score—Fever, Purulence, Attend rapidly (3 days or less), severely Inflamed tonsils and No cough or coryza (FeverPAIN; model 5 from table 2) provides a finer grading of risk and significantly more patients in both cohorts can be categorised as being at 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 the percentages of observed and predicted presence of streptococci at low scores (see online supplementary appendix 2).

The 'sequential' approach of developing the first score and 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 set to provide more power generated an eight-item unwieldy score and obscured the major differences in performance between the data sets (see online supplementary 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 that is, 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 have explored 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 of developing and validating 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

Table 3 Number of individuals with Lancefield groups 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). <: Please check table haeding italic words.>

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%)

Two clinical scores are shown: (1) a modified streptococcal score (model 5; 1 point each for: Fever during the last 24 h, Purulent tonsils, Attend rapidly (3 days or less), very Inflamed throat and No cough or coryza (FeverPAIN)) and (2) for comparison the Centor score (1 point each for pus, fever in the last 24 h, cervical glands and the absence of cough).

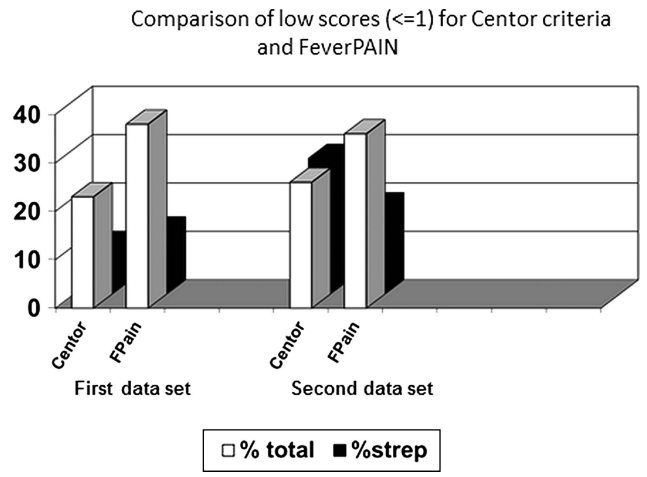

Figure 1 Comparison of low scores (≤1) for Centor criteria and Fever, Purulent tonsils, Attend rapidly (3 days or less), very Inflamed throat and No cough or coryza (FeverPAIN).

discriminatory performance of the first score when used in the second data set suggest 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 eight-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,[26] 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.[27] 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 bootstrapped estimates of the area under the ROC curve to limit overfitting, nevertheless the proposed model should have further validation.[15]

Main findings in the context of previous literature

Group A β-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.[28] 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 we 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 (ie, symptoms developing rapidly resulting in a 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 the 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 was 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 group A, C and G streptococci requires further validation, but should enable clinicians to 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.

Author affiliations
[1]Primary Care and Population Science Unit, University of Southampton, Southampton, UK
[2]Department of Primary Health Care, University of Oxford, Oxford, UK
[3]PHE, Gloucester, UK

Acknowledgements The authors are grateful to all the patients and Health Care Professions who have contributed their time and effort and helpful insights to make PRImary care Streptococcal Management (PRISM) study.

Collaborators PRISM Investigators—University of Southampton: PL, IW, MM, MM, EC, James Raftery, David Turner, Rafael Pinedo-Villanueva, JK, JB, Karen Middleton, Gerry Leydon; University of Oxford : Richard Hobbs, Richard McManus, DM, Paul Glasziou, Sue Smith, Diane Coulson; Health Protection Agency: CMN, Peter Hawtin.

Contributors Razia Meer-Baloch (senior trial manager) was involved in day-to-day coordination of the Birmingham study centre, and commented on

drafts of the manuscripts. EC developed the protocol, and contributed to the quantitative analysis and drafting of the manuscript. Paul Glazioudeveloped the protocol for funding, contributed to the management of the clinical studiesand commented on the manuscript. FDRH developed the protocol for funding, contributed to the management of all studies, supervised the Birmingham study centre and contributed to the drafting of the manuscript. JK and JB developed the protocol, provided day-to-day overall management of the study, coordinated recruitment in the lead study centre and coordination of other centresand commented on drafts of the manuscript. Gerry Leydon developed the protocol for funding, contributed to management and commented on drafts of the manuscript. PL had the original idea for the protocol, led the protocol development and the funding application, supervised the running of the lead study centre and coordination of centres, contributed to the analysis and led the drafting of themanuscript. DM 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 manuscript. Richard McManus developed the protocol for funding, contributed to the management of all studies, supervised the Birmingham Network and contributed to the drafting of themanuscript. sLisa McDermott developed the protocol and commented on drafts of the manuscript. CM and Gemma Lasseter developed the protocol and contributed to the management and write-up of the study. Peter Hawtin developed the protocol for funding, contributed to the design and running of the in vitro and diagnostic phases of the study. Karen Middletonprovided administrative support, developed data management protocols, coordinated data entry and commented on drafts of the manuscript. MM developed the protocol for fundingand contributed not only to the management of the studybut also to the analysis and drafting of the manuscript. Mark Mullee developed the protocol for funding, contributed to the management of the study, supervised data management, led the quantitative analysis and contributed to the drafting of the manuscript, IW developed the protocol for funding and also contributed to the management of the study and the drafting of the manuscript. Sue Smith, Mary Selwood and Diane Coulson provided day-to-day coordination of the Oxford study centre and commented on drafts of the manuscript.possible.

Funding This project was funded by the National Institute for Health Research Heath Technology Assessment (HTA) Programme (project number 05/10/01) and will be published in full in the Health Technology Assessment journal. Further information is available at www.nets.nihr.ac.uk/projects/hta/051001. This report presents independent research commissioned by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

Competing interests None.

Ethics approval This study was approved by a Multicentre Research Ethics Committee (number 06/MRE06/17).

Patient consent Obtained.

Provenance and peer review Not commissioned; internally peer reviewed.

Data sharing statement The authors are happy to share data and collaborate with other investigators as appropriate (eg, in larger merged individual patient data studies).

REFERENCES

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;365:579–87.
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: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.
5. Del Mar C. Managing sore throat: a literature review I: making the diagnosis. *Med J Aust* 1992;156:572–5.
6. Dagnelie C, Bartelink M, Van Der Graaf Y, *et al*. Towards better diagnosis of throat infections with GABHS in general practice. *Br J Gen Pract* 1998;48:959–62.
7. Dobbs F. A scoring system for predicting group A streptococcal throat infection. *Br J Gen Pract* 1996;46:461–4.
8. NICE guideline development group. Prescribing of antibiotics for self-limiting respiratory tract infections in adults and children in primary care 2008. http://www nice org uk/Guidance/CG69
9. McIsaac W, Kellner J, Aufricht P, *et al*. Empirical validation of guidelines for the management of pharyngitis in children and adults . *JAMA* 2004;291:1587–95.
10. Cooper R, Hoffman J, Bartlett J, *et al*. Principles of appropriate antibiotic use for acute pharyngitis in adults. *Ann Intern Med* 2001;134:509–17.
11. Principles of Appropriate Antibiotic Use for Acute Pharyngitis in Adults 2001. http://www acponlineorg/clinical_information/guidelines
12. Vincent M, Celestin N, Hussain A. Pharyngitis. *Am Fam Phys* 2004;69:1465–70.
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;172:847–52.
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 Med* 2011;9:67.
15. Lindbaek M, Hoiby E, Steinsholt I, *et al*. Clinical symptoms and signs in sore throat patients with large colony variant β-haemolytic streptococci groups C or G versus group A. *Br J Gen Pract* 2005;55:615–19.
16. Little P, Hobbs R, Mant D, *et al*. Incidence and clinical variables associated with streptococcal throat infections: a prospective diagnostic cohort study. *Br J Gen Pract* 2012;62: e787–94.
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. *BMJ* 1997;315:350–2.
18. Rogers M. A viable alternative to the glass/mercury thermometer. *Paediatr Nurs* 1992;4.
19. National Standard Method: investigation of throat swabs. BSOP 2008;9: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. http://www.hpa-standardmethods.org.uk/documents/bsopid/pdf/bsopid4.pdf
21. Methods for Antimicrobial Susceptibility Testing. British Society for Antimicrobial Chemotherapy 2009. 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 (Suppl 1):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 edn. London: HMSO, 1994.
25. Little P, Turner S, Rumsby K, *et al*. Dipsticks and diagnostic algorithms in urinary tract infection: development and validation, randomised trial, economic analysis, observational cohort and qualitative study. *Health Technol Assess* 2009;13: 1–96.
26. Little P, Hobbs FDR, Moore M, *et al*. Clinical score and rapid antigen detection test to guide antibiotic use for sore throats: randomised controlled trial of PRISM (primary care streptococcal management). *BMJ* 2013;347:f5806.
27. 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;53:666–71.
28. Spinks A, Glasziou P, Del Mar C. Antibiotics for sore throat. *Cochrane Database Syst Rev* 2006; doi: 10.1002/14651858.CD000023.pub3.