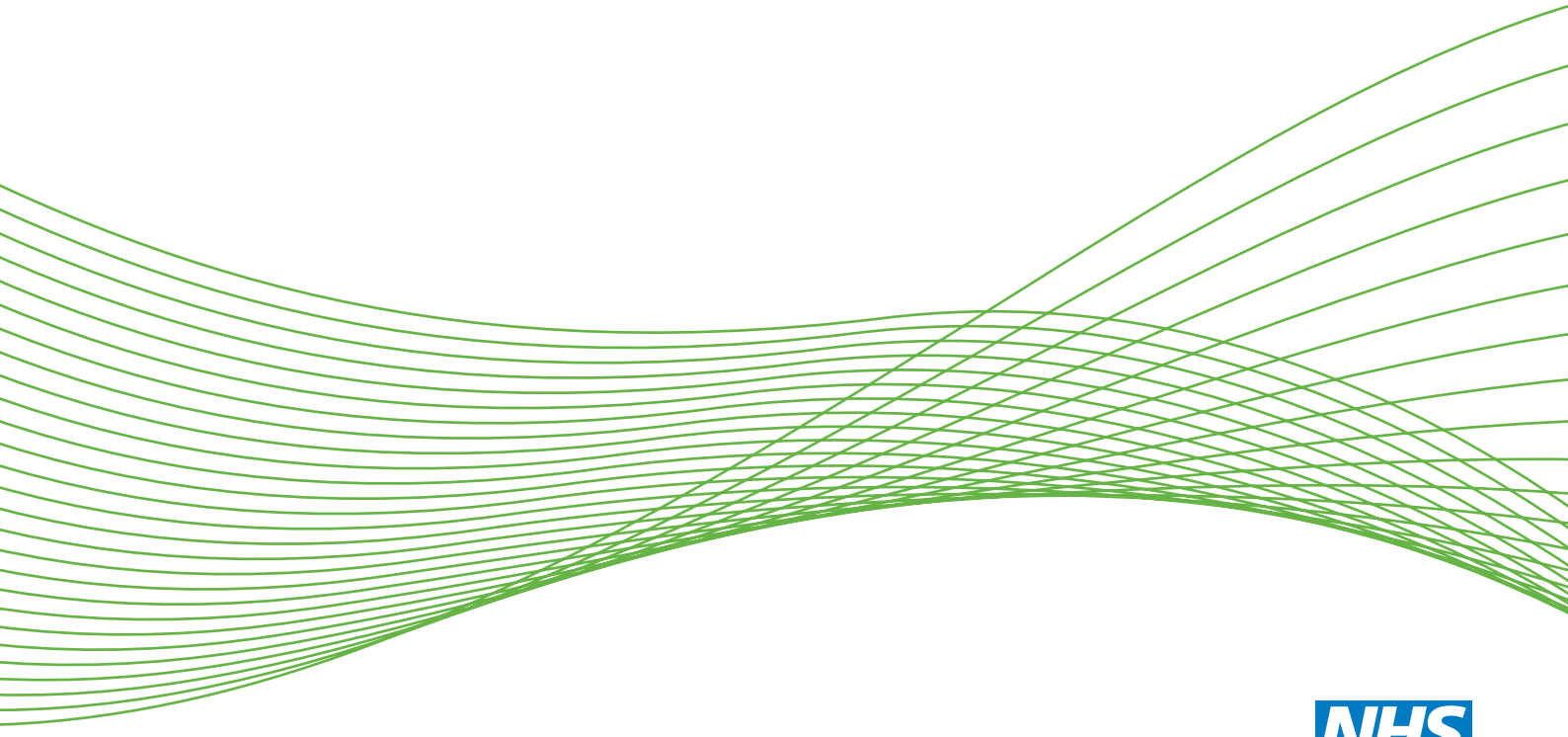
HEALTH TECHNOLOGY ASSESSMENT

VOLUME 18 ISSUE 6 JANUARY 2014
ISSN 1366-5278

PRImary care Streptococcal Management (PRISM) study: in vitro study, diagnostic cohorts and a pragmatic adaptive randomised controlled trial with nested qualitative study and cost-effectiveness study

Paul Little, FD Richard Hobbs, Michael Moore, David Mant, Ian Williamson, Cliodna McNulty, Gemma Lasseter, MY Edith Cheng, Geraldine Leydon, Lisa McDermott, David Turner, Rafael Pinedo-Villanueva, James Raftery, Paul Glasziou and Mark Mullee on behalf of the PRISM investigators



DOI 10.3310/hta18060

NHS
National Institute for Health Research

PRImary care Streptococcal Management (PRISM) study: in vitro study, diagnostic cohorts and a pragmatic adaptive randomised controlled trial with nested qualitative study and cost-effectiveness study

Paul Little,[1]* FD Richard Hobbs,[2] Michael Moore,[1] David Mant,[3] Ian Williamson,[1] Cliodna McNulty,[4] Gemma Lasseter,[4] MY Edith Cheng,[1] Geraldine Leydon,[1] Lisa McDermott,[1] David Turner,[1] Rafael Pinedo-Villanueva,[1] James Raftery,[1] Paul Glasziou[3] and Mark Mullee[1] on behalf of the PRISM investigators

[1]Primary Care and Population Sciences Unit, University of Southampton, Southampton, UK
[2]Primary Care Clinical Sciences, University of Birmingham, Birmingham, UK
[3]Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
[4]Public Health England, Primary Care Unit, Microbiology Laboratory, Gloucestershire Royal Hospital, Gloucester, UK

*Corresponding author

Declared competing interests of authors: Paul Little is a member of the NIHR Journals Library Board. James Raftery is a member of the HTA Editorial Board and NIHR Journals Library Editorial Group.

Published January 2014
DOI: 10.3310/hta18060

This report should be referenced as follows:

Little P, Hobbs FDR, Moore M, Mant D, Williamson I, McNulty C, *et al.* PRImary care Streptococcal Management (PRISM) study: in vitro study, diagnostic cohorts and a pragmatic adaptive randomised controlled trial with nested qualitative study and cost-effectiveness study. *Health Technol Assess* 2014;**18**(6).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents®*/ Clinical Medicine.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Five-year impact factor: 5.804

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: www.hta.ac.uk/

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 05/10/01. The contractual start date was in October 2006. The draft report began editorial review in August 2012 and was accepted for publication in January 2013. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded 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. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Editor-in-Chief of *Health Technology Assessment* and NIHR Journals Library

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the HTA Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andree Le May Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

Professor Aileen Clarke Professor of Health Sciences, Warwick Medical School, University of Warwick, UK

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Peter Davidson Director of NETSCC, HTA, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Professor Elaine McColl Director, Newcastle Clinical Trials Unit, Institute of Health and Society, Newcastle University, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Honorary Professor, Business School, Winchester University and Medical School, University of Warwick, UK

Professor Jane Norman Professor of Maternal and Fetal Health, University of Edinburgh, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professorial Research Associate, University College London, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: nihredit@southampton.ac.uk

Abstract

PRImary care Streptococcal Management (PRISM) study: in vitro study, diagnostic cohorts and a pragmatic adaptive randomised controlled trial with nested qualitative study and cost-effectiveness study

Paul Little,[1]* FD Richard Hobbs,[2] Michael Moore,[1] David Mant,[3] Ian Williamson,[1] Cliodna McNulty,[4] Gemma Lasseter,[4] MY Edith Cheng,[1] Geraldine Leydon,[1] Lisa McDermott,[1] David Turner,[1] Rafael Pinedo-Villanueva,[1] James Raftery,[1] Paul Glasziou[3] and Mark Mullee[1] on behalf of the PRISM investigators

[1]Primary Care and Population Sciences Unit, University of Southampton, Southampton, UK
[2]Primary Care Clinical Sciences, University of Birmingham, Birmingham, UK
[3]Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
[4]Public Health England, Primary Care Unit, Microbiology Laboratory, Gloucestershire Royal Hospital, Gloucester, UK

*Corresponding author p.little@soton.ac.uk

Background: Antibiotics are still prescribed to most patients attending primary care with acute sore throat, despite evidence that there is modest benefit overall from antibiotics. Targeting antibiotics using either clinical scoring methods or rapid antigen detection tests (RADTs) could help. However, there is debate about which groups of streptococci are important (particularly Lancefield groups C and G), and uncertainty about the variables that most clearly predict the presence of streptococci.

Objective: This study aimed to compare clinical scores or RADTs with delayed antibiotic prescribing.

Design: The study comprised a RADT in vitro study; two diagnostic cohorts to develop streptococcal scores (score 1; score 2); and, finally, an open pragmatic randomised controlled trial with nested qualitative and cost-effectiveness studies.

Setting: The setting was UK primary care general practices.

Participants: Participants were patients aged ≥ 3 years with acute sore throat.

Interventions: An internet program randomised patients to targeted antibiotic use according to (1) delayed antibiotics (control group), (2) clinical score or (3) RADT used according to clinical score.

Main outcome measures: The main outcome measures were self-reported antibiotic use and symptom duration and severity on seven-point Likert scales (primary outcome: mean sore throat/difficulty swallowing score in the first 2–4 days).

Results: The IMI TestPack Plus Strep A (Inverness Medical, Bedford, UK) was sensitive, specific and easy to use. Lancefield group A/C/G streptococci were found in 40% of cohort 2 and 34% of cohort 1. A five-point score predicting the presence of A/C/G streptococci [FeverPAIN: Fever; Purulence; Attend rapidly (≤ 3 days); severe Inflammation; and No cough or coryza] had moderate predictive value (bootstrapped estimates of area under receiver operating characteristic curve: 0.73 cohort 1, 0.71

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

cohort 2) and identified a substantial number of participants at low risk of streptococcal infection. In total, 38% of cohort 1 and 36% of cohort 2 scored ≤ 1 for FeverPAIN, associated with streptococcal percentages of 13% and 18%, respectively. In an adaptive trial design, the preliminary score (score 1; $n = 1129$) was replaced by FeverPAIN ($n = 631$). For score 1, there were no significant differences between groups. For FeverPAIN, symptom severity was documented in 80% of patients, and was lower in the clinical score group than in the delayed prescribing group (–0.33; 95% confidence interval –0.64 to –0.02; $p = 0.039$; equivalent to one in three rating sore throat a slight rather than moderately bad problem), and a similar reduction was observed for the RADT group (–0.30; –0.61 to 0.00; $p = 0.053$). Moderately bad or worse symptoms resolved significantly faster (30%) in the clinical score group (hazard ratio 1.30; 1.03 to 1.63) but not the RADT group (1.11; 0.88 to 1.40). In the delayed group, 75/164 (46%) used antibiotics, and 29% fewer used antibiotics in the clinical score group (risk ratio 0.71; 0.50 to 0.95; $p = 0.018$) and 27% fewer in the RADT group (0.73; 0.52 to 0.98; $p = 0.033$). No significant differences in complications or reconsultations were found. The clinical score group dominated both other groups for both the cost/quality-adjusted life-years and cost/change in symptom severity analyses, being both less costly and more effective, and cost-effectiveness acceptability curves indicated the clinical score to be the most likely to be cost-effective from an NHS perspective. Patients were positive about RADTs. Health professionals' concerns about test validity, the time the test took and medicalising self-limiting illness lessened after using the tests. For both RADTs and clinical scores, there were tensions with established clinical experience.

Conclusions: Targeting antibiotics using a clinical score (FeverPAIN) efficiently improves symptoms and reduces antibiotic use. RADTs used in combination with FeverPAIN provide no clear advantages over FeverPAIN alone, and RADTs are unlikely to be incorporated into practice until health professionals' concerns are met and they have experience of using them. Clinical scores also face barriers related to clinicians' perceptions of their utility in the face of experience. This study has demonstrated the limitation of using one data set to develop a clinical score. FeverPAIN, derived from two data sets, appears to be valid and its use improves outcomes, but diagnostic studies to confirm the validity of FeverPAIN in other data sets and settings are needed. Experienced clinicians need to identify barriers to the use of clinical scoring methods. Implementation studies that address perceived barriers in the use of FeverPAIN are needed.

Trial registration: Current Controlled Trials ISRCTN32027234.

Source of funding: This project was funded by the NIHR Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 18, No. 6. See the NIHR Journals Library website for further project information.

Contents

List of tables xiii

List of figures xv

List of abbreviations xvii

Plain English summary xix

Scientific summary xxiii

Chapter 1 In vitro evaluation of five rapid antigen detection tests for group A beta-haemolytic streptococcal sore throat infections **1**
Abstract 1
Background 1
Objectives 1
Method 1
Results 1
Conclusion 1
Background 1
Method 2
Rapid antigen detection kits 2
Sample size for kit sensitivity testing 3
Sensitivity testing: ability to detect group A beta-haemolytic streptococcus 3
Specificity testing: assessment of false-positives due to commensal flora 3
Ease of use of rapid antigen detection tests 3
Swab type 3
Results 3
Sensitivity results: ability to detect group A beta-haemolytic streptococcus 3
Specificity 6
Swab type 6
Ease of use/storage, cost and instruction clarity (see Tables 2 *and* 3*): suitability of rapid antigen detection tests for general practice* 7
Clarity of kit instructions and price per test 8
Discussion 8
Strengths and limitations 8
Main findings and comparison with existing literature 9
Swab type 9
Conclusion 9

Chapter 2 The incidence and clinical variables associated with streptococcal throat infections **11**
Abstract 11
Background 11
Objective 11
Design 11
Setting 11
Patients 11
Methods 11

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Results 11
Conclusion 11
Background 12
Methods 12
Inclusion 12
Clinical data 12
Throat swabs 13
Sample size 13
Analysis 13
Results 13
Recruitment bias 14
Discussion 19
Strengths and limitations of the study 19
Main findings and comparison with existing literature 20
Conclusion 20

Chapter 3 Two diagnostic cohorts to identify clinical variables associated with Lancefield group A beta-haemolytic streptococci and Lancefield non-group A streptococcal throat infections **21**
Abstract 21
Background 21
Objective 21
Design 21
Setting 21
Patients 21
Results 21
Conclusions 22
Background 22
Methods 22
Sample size 22
Analysis 22
Results 23
Discussion 26
Strengths and limitations of the study 26
Main findings in the context of previous literature 26
Clinical utility 27
Conclusion 27

Chapter 4 Randomised controlled trial of a clinical score and rapid antigen detection test for sore throats **29**
Abstract 29
Objective 29
Design 29
Setting 29
Patients 29
Intervention 29
Outcomes 29
Results 29
Conclusion 29
Background 30
Methods 30
Development of clinical scores 30
Trial recruitment 31

Intervention groups 31
Data collection 32
Analysis 33
Results 34
Evidence of differential effectiveness of the first and second parts of the trial 34
Baseline table 35
Main findings 35
Selection and attrition bias 37
Discussion 38
Strengths and potential limitations 38
Main results in context of previous literature 38
Conclusion 39

Chapter 5 A qualitative study of general practitioner, nurse practitioner and patient views about the use of rapid streptococcus antigen detection tests in primary care: 'swamped with sore throats?' 41
Abstract 41
Background 41
Objective 41
Method 41
Results 41
Conclusions 41
Background 41
Objective 42
Methods 42
Participants and procedure 42
The interviews 42
Analysis 43
Findings 43
Participants 43
Themes 43
Discussion 49
Strengths and limitations 49
Main findings 49
Comparison with existing literature 50
Implications for clinical practice 50
Conclusion 50

Chapter 6 Health economic analysis of the randomised controlled trial 51
Abstract 51
Aims 51
Methods 51
Results 51
Conclusion 51
Introduction 51
Methods 52
Data sources for economic evaluation 52
Analysis of costs 52
Participants 54
Outcome data used 54
Analysis 55

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Results 55
Cost-effectiveness analysis 57
Cost–utility analysis 59
Results of cost–utility study 60
Discussion 62
Statement of main findings 62
Strengths and potential limitations 62
Relation to existing literature 63
Conclusion 63

Acknowledgements **65**

References **67**

Appendix 1 Preparation of group A beta-haemolytic streptococcus and commensal stock cultures **73**

Appendix 2 Manufacture's swab recommendations for five rapid antigen detection tests **75**

Appendix 3 Detailed sensitivity results **77**

Appendix 4 Clinical variables in patients with group A, C or G beta-haemolytic streptococci compared with patients with no growth of C, G or A beta-haemolytic streptococci using more levels for variables **79**

Appendix 5 **81**

Appendix 6 Data for score 1 **83**

Appendix 7 Health professional testing of rapid streptococcal antigen detection test kits for ease of use **85**

Appendix 8 Protocol **87**

List of tables

TABLE 1 Strains used for in vitro RADT investigation 4

TABLE 2 General ease-of-use classification: for five RADTs for the detection GABHS, based on general characteristics and the reader's opinion 5

TABLE 3 Classification of five RADTs base on price per test and instruction insert clarity (April 2008 prices) 8

TABLE 4 Clinical variables in patients with GABHS (n=136) compared with all other patients 15

TABLE 5 Clinical variables in patients with Lancefield groups C and G beta-haemolytic streptococci and among patients who have no growth of group A, C or G beta-haemolytic streptococci 16

TABLE 6 Clinical variables in patients with non-GABHS (C and G) and GABHS compared with patients who have no growth of C, G or A streptococci 18

TABLE 7 Second data set: clinical variables in patients with Lancefield group A, C and G compared with patients with no growth of Lancefield C, G or A streptococci, with odds ratios (95% CI) 24

TABLE 8 Sequential area under the ROC curve values as successive variables added (p-values given for comparison with previous model unless specified) 25

TABLE 9 Number of individuals with Lancefield group A, C or G streptococci (%) at each level of clinical scores, and the total number of individuals at each level (and per cent of the total sample) 25

TABLE 10 Baseline clinical assessment and prescribing strategy used by the GP at the baseline consultation 35

TABLE 11 Symptom severity, antibiotic use, intention to consult in the future (moderately likely or more likely) and reconsultations with sore throat 36

TABLE 12 Practice demographics 42

TABLE 13 Themes identified in analysis 44

TABLE 14 Types of contacts and unit costs used 53

TABLE 15 Resource use and cost by study group 56

TABLE 16 Cost per point change in symptom score results (means and 95% CIs) 57

TABLE 17 Cost–utility outcomes data (95% CI) 59

TABLE 18 Cost–utility results for FeverPAIN using adjusted QALY data for 14-day period (means and 95% CIs) 60

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 19 Cost-utility results for FeverPAIN using adjusted QALY data for 28-day period (means and 95% CIs) 60

TABLE 20 Symptom severity, antibiotic use, intention to consult in the future (moderately likely or more likely) and reconsultations with sore throat for score 1. Results are risk ratios (95% CIs) or mean differences (95% CIs) 83

TABLE 21 Strategy used by clinicians when using score 1 84

List of figures

FIGURE 1 The five RADTs evaluated. Dipsticks, left to right: OSOM, QuickVue Dipstick and Streptatest. Cassettes, top to bottom: IMI TestPack Plus and Clearview. Showing positive reactions for all 2

FIGURE 2 Sensitivity (with 95% CI) for five RADTs, in relation to swab type used; all GABHS strains and concentrations 6

FIGURE 3 Sensitivity (with 95% CI) for each swab type; all RADTs, GABHS strains and GABHS concentrations 6

FIGURE 4 Using 'best' swab type, RADT sensitivity at each GABHS concentration; combined results for all GABHS strains 7

FIGURE 5 Percentage of symptoms and signs at presentation for patients infected with GABHS, group C or G streptococci or no group A, C or G streptococci 17

FIGURE 6 Log-univariate likelihood ratios [log of likelihood ratio for a negative test (log-LR–) and log of likelihood ratio for a positive test (log-LR+)] for individual symptoms and signs at presentation for patients with group A, C or G beta-haemolytic streptococci 19

FIGURE 7 Comparison of low scores ($\leq$1) for Centor criteria and FeverPAIN 26

FIGURE 8 The CONSORT (CONsolidated Standards Of Reporting Trials) flow diagram for the second phase of the trial (using FeverPAIN). a, Estimates or those not assessed or declined based on physician report 34

FIGURE 9 Daily VASs for the 257 participants used in the cost–utility analysis 55

FIGURE 10 Cost-effectiveness acceptability curve for cost-effectiveness study 58

FIGURE 11 Scatterplot showing FeverPAIN vs. delayed prescribing group for cost per point change in symptom score 58

FIGURE 12 Scatterplot showing RADT vs. delayed prescribing group for cost per point change in symptom score 58

FIGURE 13 Scatterplot showing RADT vs. clinical score group for cost per point change in symptom score 59

FIGURE 14 Cost-effectiveness acceptability curve for 14-day QALY difference 60

FIGURE 15 Cost-effectiveness acceptability curve for 28-day QALY difference 61

FIGURE 16 Scatterplot showing clinical score group vs. delayed prescribing group for cost/QALY analysis for 14-day QALY difference 61

FIGURE 17 Scatterplot showing RADT group vs. delayed prescribing group for cost/QALY analysis for 14-day QALY difference 61

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

FIGURE 18 Scatterplot showing RADT group vs. clinical score group for cost/QALY analysis 62

FIGURE 19 The CONSORT (CONsolidated Standards Of Reporting Trials) trial flow diagram for first phase of the trial (score 1) 84

List of abbreviations

A&E	accident and emergency
BHI	brain-heart infusion
CBA	Columbia agar with horse blood
CEAC	cost-effectiveness acceptability curve
CFU	colony-forming unit
CI	confidence interval
GABHS	group A beta-haemolytic streptococcus
GP	general practitioner
HCP	health-care practitioner
HRQoL	health-related quality of life
MRC	Medical Research Council
NCTC	National Collection of Type Cultures
NP	nurse practitioner
PSSRU	Personal Social Services Research Unit
QALY	quality-adjusted life-year
RADT	rapid antigen detection test
ROC	receiver operating characteristic
SD	standard deviation
VAS	visual analogue score

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Plain English summary

General background

The overuse of antibiotics in general practices, mostly for illnesses such as sore throat, chest infections and ear infections, is potentially a big problem for us all for several reasons. First, antibiotic overuse increases the risk of antibiotic resistance, whereby bacteria become resistant to antibiotics and are no longer killed by antibiotics. This could potentially lead to serious infections as a result of 'superbugs' becoming untreatable both now and for future generations. Antibiotics commonly cause side effects such as allergic reactions, diarrhoea and skin rashes. Using them also increases people's belief in them – because they think it is the antibiotics that helped them get better, when in fact they would have got better in the same time anyway. This leads people to think that they need to come back the next time they get an infection – so it 'medicalises' illnesses, uses NHS resources and also exposes patients to unnecessary antibiotics.

Background – the context for sore throats

Antibiotics are still prescribed for most patients with a sore throat attending their general practitioner (GP) or nurse in primary care. This is despite the best available evidence, which suggests there is a modest benefit overall from antibiotics. One approach to tackle this is to target antibiotics better, using a simple 'clinical score' – whereby doctors or nurses prescribe according to particular symptoms and examination findings. Another approach is to use rapid antigen detection tests (RADTs), which are very commonly used in many countries. To use a RADT, a swab is taken from the throat, and the RADT gives a quick answer as to whether the most important bacteria are present or not. The particular type of bacteria that RADTs pick up is a common type of streptococcus bacteria – called Lancefield group A haemolytic streptococcus (GABHS). This bacterium can cause both a sore throat and more serious illnesses.

However, there are problems with using either a clinical score or a RADT:

- There is debate about which RADT should be used and how.
- It is unclear whether other bacteria (other than GABHS) are important, particularly streptococci from other groups – Lancefield groups C and G. RADTs will not pick up these other bacteria.
- For a clinical score, it is not clear which symptoms and examination findings most clearly tell us whether bacteria are present.
- There have also been very few good studies that compare RADTs with clinical scores, or with other approaches, such as delayed antibiotic prescribing. Delayed prescribing is where the patient is advised to use an antibiotic after several days if symptoms are not starting to settle.

The PRImary care Streptococcal Management (PRISM) study was made up of several substudies that tackled these issues:

Laboratory study

If rapid antigen tests are to be used for patients in everyday practice, they have to be accurate, easy to use, inexpensive and potentially widely available. Several such tests are available, and in the first study five RADTs were tested in the laboratory with different types and concentrations of bacteria. One of the best of these was the IMI test, which was both one of the most accurate and found to be relatively easy to use.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Clinical study – developing a clinical score

Two large groups of patients (606 in the first group, 517 in the second) came to see the doctor or nurse with a sore throat and agreed to take part. Their symptoms and signs were documented and a throat swab was sent to the laboratory to see if bacteria were present. The results showed that patients who had Lancefield groups C or G bacteria had the same kind of illness as those with group A strains. It was also possible to develop a useful clinical score to help pick up the main types of bacteria (A, C or G) based on a simple count of five items. The five items make up the acronym FeverPAIN:

- **F**ever during the last 24 hours
- **P**us (white spots) on the tonsils
- coming quickly to see the doctor within 3 days (**A**ttend rapidly)
- very **I**nflamed tonsils
- and **N**o cough or runny nose.

Trial of clinical scores and rapid antigen detection tests

The trial compared three ways of managing sore throat among 1760 patients who came to see their doctor:

1. Delayed antibiotic prescribing group (the control group).
2. Clinical score group: the score was worked out and antibiotics were advised for high scores. No antibiotics were advised for low scores, and delayed antibiotics for those in the middle. The first clinical score that was developed (score 1; $n = 1129$) was replaced by a more valid score (FeverPAIN; $n = 631$) as the trial went on.
3. RADT group: the clinical score was also worked out. For low and middle scores, the plan was similar to that used in the clinical score group. A RADT was used for those with high scores, and, if the result was positive, antibiotics were advised and, if the result was negative, no antibiotics were given.

The study found that using the clinical score (FeverPAIN) improved control of symptoms, and both the clinical score and the RADT reduced antibiotic use. Moderately bad or worse symptoms resolved significantly faster (30% faster) in the clinical score group but not in the RADT group (11% faster).

Health economic analysis

If RADTs were to be used more widely, it would be important to show that using them is a cost-effective use of time and money for the health service. The study showed that using RADTs was probably more expensive and less cost-effective than using the clinical score.

Qualitative study

Face-to-face and telephone interviews were done with 51 people – GPs, nurse practitioners and patients from general practices across Hampshire, Oxfordshire and the West Midlands. Patients and nurses were very positive about using clinical scores and RADTs. Doctors had a number of concerns about both RADTs and clinical scores that would need to be addressed before widespread implementation would work – particularly related to the perceived usefulness of clinical scores in the face of clinical experience and intuition.

Conclusions

There are RADTs that are not expensive, easy to use and are potentially widely available for use in primary care. Although they will detect GABHS, RADTs are not designed to detect other strains such as Lancefield C or G strains. Lancefield C or G strains commonly cause streptococcal sore throats, and patients have a similar illness to those who have A strains. A five-item score (acronym FeverPAIN) to predict streptococcal infection is likely to be valid but further validation is preferable. When antibiotics are targeted using a clinical score (FeverPAIN), this improves control of symptoms, reduces antibiotic use and is very cost-effective. Using a RADT in addition to using the clinical score provides no clear benefits for patients over using the clinical score alone. RADT use is also more costly, probably less cost-effective and faces several barriers from clinicians. To implement the use of clinical scores more widely in everyday practice will require addressing the issues doctors have.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Scientific summary

Background

The overuse of antibiotics in primary care not only increases the risk of antibiotic resistance but exposes patients to side effects, and medicalises what are mostly self-limiting illnesses. Antibiotics are still prescribed for most patients attending primary care with acute sore throat, despite evidence from systematic reviews that there is modest benefit overall from antibiotics. Approaches to targeting antibiotics could facilitate more appropriate use of antibiotics, either targeting antibiotics using clinical scoring methods or using rapid antigen detection tests (RADTs). RADTs are very commonly used in many countries and are designed to detect the major bacterial pathogen Lancefield group A beta-haemolytic streptococcus (GABHS). However, there is debate about the importance of other major groups of streptococci (particularly Lancefield groups C and G). Furthermore, there is uncertainty about the variables that most clearly predict the presence of streptococci, and about the most appropriate RADTs to use in primary care. There is also very little robust trial evidence comparing management alternatives.

Objectives

1. *In vitro study*: to assess in vitro validity and ease of use of several commercially available RADTs to detect GABHS and to explore the impact of using commercially available swabs instead of the swabs provided with the kits.
2. *Clinical diagnostic study*: to assess the incidence and clinical variables associated with streptococcal infections and develop a clinical score to help target antibiotic use.
3. *Randomised controlled trial*: to compare the targeting of antibiotics using a clinical score, or, alternatively, using a clinical score combined with a RADT, with empirical delayed antibiotic prescribing.
4. *Qualitative study*: to explore patients' and health-care professionals' (HCP) views of clinical scores and RADTs.
5. *Cost-effectiveness analysis*: to assess resource use and the health-related quality of life associated with clinical scores and RADTs and to show whether these can represent an efficient use of NHS resources.

Methods

1. *In vitro study*: different concentrations and strains of GABHS and non-GABHS were assessed with OSOM® Ultra Strep A (Bio-Stat Limited, Stockport, UK), QuickVue® Dipstick Strep A test (TK Diagnostic, Oxford, UK), Streptatest® (DECTRA PHARM, Strasbourg, France), Clearview® Exact (Inverness Medical Professional Diagnostics, Bedford, UK) and IMI TestPack® Plus Strep A (Inverness Medical, Bedford, UK). Each kit was rated for ease of use. Test kit swabs were also compared with commercially available swabs.
2. *Clinical diagnostic study*: the variables significantly associated with the presence of pathogenic streptococci from throat swabs were assessed among patients aged ≥ 5 years presenting with acute sore throat in two cohorts of patients. Logistic regression was used to identify significant variables to incorporate in clinical prediction rules, and bootstrapping was used to estimate the area under the receiver operating characteristic (ROC) curve. Patients were recruited for a second cohort (cohort 2, $n = 517$) consecutively after the first (cohort 1, $n = 606$) from similar practices.
3. *Randomised controlled trial*: 1760 patients aged ≥ 3 years with acute sore throat were individually randomised using a web-based program to one of three structured approaches targeting antibiotic use according to (1) delayed antibiotic prescribing (control), (2) clinical score or (3) RADT use determined by a clinical score. The main outcomes were symptom severity, symptom duration and antibiotic use

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

(all of which were documented in > 80% of participants). The trial was not initially designed as an adaptive trial, but revision of the trial design and modification of the intervention were needed once it became clear that the first clinical score that was developed was not likely to be valid. The adaptive design was implemented once the change in design had been agreed with the funder.

4. *Qualitative study*: Semi-structured face-to-face and telephone interviews were conducted with general practitioners (GPs) and nurse practitioners (NPs) from general practices across Hampshire, Oxfordshire and the West Midlands.
5. *Cost-effectiveness analysis*: A cost–utility study [cost/quality-adjusted life-year (QALY)] and a cost-effectiveness study (cost/change in symptom severity) were carried out as part of the randomised controlled trial. Resource use data were obtained from GP case notes and from study clinicians. QALYs were estimated by means of EQ5D scores obtained from the 14-day diary.

Results

1. *In vitro* study: The IMI test was the easiest to use. Sensitivity increased with higher streptococcal concentration: at high concentrations sensitivity ranged from 62% [Clearview; 95% confidence interval (CI) 51% to 72%] to 95% (OSOM and IMI; 95% CI 88% to 98%). All tests were specific (100%). Most kits performed well independent of what swab was used, but Clearview was much more sensitive with polyester swabs than the kit swabs and rayon swabs.
2. *Clinical diagnostic study*: A, C or G beta-haemolytic streptococci were found in 40% of participants in cohort 2 and 34% in cohort 1. There was variation in the items that were significant in multivariate analysis in both cohorts. The clinical features predicting the presence of these streptococci in multivariate analysis in both cohorts were as follows: short prior duration of illness (attend rapidly in ≤ 3 days; multivariate-adjusted odds ratio 1.92 cohort 1, 1.67 cohort 2); fever in the last 24 hours (1.69, 2.40); and doctor's assessment of severity (severely inflamed pharynx/tonsils) (2.28, 2.29). Absence of coryza or cough and purulent tonsils were also significant predictive variables in univariate analysis in both cohorts and in multivariate analysis in at least one cohort. A five-item score based on Fever, Purulence, Attend rapidly (≤ 3 days), severe Inflammation and No cough or coryza (acronym FeverPAIN) had moderate predictive value (bootstrapped estimates of area under ROC curve 0.73 cohort 1, 0.71 cohort 2) and performed well in identifying a substantial number of participants at low risk of streptococcal infection (38% in cohort 1 and 36% in cohort 2 scored ≤ 1, associated with streptococcal percentages of 13% and 18%, respectively). A Centor score of ≤ 1 identified 23% and 26% of participants with streptococcal percentages 10% and 28%, respectively.
3. *Randomised controlled trial*: A preliminary score to predict streptococcal infection (score 1; $n = 1129$) was replaced by a more valid score (FeverPAIN, $n = 631$) in an adaptive trial design. There were no significant differences between groups for score 1, and it performed significantly less well than FeverPAIN for the key outcomes. For FeverPAIN, symptom severity was documented in 80% of patients [delayed 168/207 (81%); clinical score 168/211 (80%); RADT 166/213 (78%)]. Severity was lower in the clinical score group than in the delayed prescribing group (–0.33; 95% CI –0.64 to –0.02; $p = 0.039$; equivalent to one in three rating sore throat a slight rather than moderately bad problem), and a similar reduction was observed for the RADT group (–0.30; –0.61 to –0.00; $p = 0.053$). Moderately bad or worse symptoms resolved significantly faster (30%) in the clinical score group (hazard ratio 1.30; 95% CI 1.03 to 1.63) but not in the RADT group (1.11; 95% CI 0.88 to 1.40). In the delayed group, 75/164 (46%) used antibiotics, and 29% fewer used antibiotics in the clinical score group (risk ratio 0.71; 95% CI 0.50 to 0.95; $p = 0.018$) and 27% fewer in the RADT group (0.73; 95% CI 0.52 to 0.98; $p = 0.033$). No significant differences in complications or reconsultations were found.
4. *Qualitative study*: Fifty-one participants took part in the study. Of these, 42 were HCPs (29 GPs and 13 NPs) and nine were patients. HCPs could see a positive role for RADTs in terms of reassurance, as an educational tool for patients and for aiding inexperienced practitioners, but also had major concerns about RADT use in clinical practice. A particular concern was the tension and possible disconnect with clinical experience and intuition – a concern which was also raised about clinical scores. Other issues included the validity of the tests (the role of other bacteria, and carrier states), the issues of time and

resource use, and the potential for medicalisation of self-limiting illness. In contrast, however, experience of using RADTs over time seemed to make some participants more positive about using the tests. Moreover, patients were much more positive about the place of RADTs in providing reassurance and in limiting their antibiotic use.

5. *Cost-effectiveness analysis*: As score 1 had not been shown to be effective but FeverPAIN appeared to be effective, the cost-effectiveness results are only presented for the second part of the trial, when FeverPAIN was used. There were 499 individuals who had both symptom severity and cost data from case notes review. Costs for the initial visit and for the 1-month follow-up were similar at £51, £44 and £52 for the delayed, clinical score and RADT groups, respectively. The clinical score group dominated both other groups for both the cost/QALY and cost/change in symptom severity analyses, being both less costly and more effective. Cost-effectiveness acceptability curves indicated the clinical score method to be the most likely to be cost-effective in both cases.

Conclusions

1. *In vitro study*: The IMI TestPack was suitable for use with high sensitivity, specificity and ease of use.
2. *Clinical diagnostic study*: Non-group A strains commonly cause streptococcal sore throats, and present with similar symptomatic clinical features to group A streptococci. The variables that are predictive of streptococcal infection vary between cohorts, and thus the conventional approach of a single development cohort and then a subsequent validation cohort may not identify the optimal variables to test in a clinical prediction rule. From the two cohorts, a five-item score (acronym FeverPAIN) is likely to be valid (Fever during the last 24 hours, Purulent tonsils, Attend rapidly ($\leq$ 3 days), very Inflamed pharynx, No cough/coryza), but further validation is required.
3. *Randomised controlled trial*: Targeting antibiotic using a clinical score (FeverPAIN) improves control of symptoms and reduces antibiotic use. A rapid antigen test combined with a clinical score provides similar benefits for antibiotic use, but no clear advantages over using a clinical scoring method alone.
4. *Qualitative study*: It is unlikely that RADTs will have a comfortable place in clinical practice in the near future until health professionals' concerns are met, and they have direct experience of using them. The routine use of clinical scoring systems for acute upper respiratory tract illness also faces barriers related to clinicians' perceptions of their utility in the face of clinician experience and intuition.
5. *Cost-effectiveness study*: Using a clinical score appears to be an efficient use of health-care resources compared with either delayed antibiotic prescribing or the use of a RADT combined with a clinical score.

Overall conclusion

Rapid antigen detection tests that are inexpensive, accurate and easy to use are potentially widely available for use in primary care. Although they will detect Lancefield group A streptococci, they are not designed to detect non-group A strains that commonly cause streptococcal sore throats, and present with a similar illness to group A streptococci. The variables that are predictive of streptococcal infection vary between different samples, but from the two cohorts of patients a five-item score (acronym FeverPAIN) is likely to be valid, but further validation is required. Targeting antibiotic using a clinical score (FeverPAIN) improves control of symptoms, reduces antibiotic use and is very cost-effective. Using a rapid antigen test in addition to using the clinical score provides no clear benefits for patients over using the clinical score alone, and is more costly, less cost-effective and faces several barriers from clinicians. The implementation of clinical scoring methods in everyday practice will require health professionals' issues related to the perceived utility of clinical scores in the face of clinical experience and intuition to be addressed.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Suggestions for further research

This study has demonstrated the limitation of using one data set to develop a clinical score. FeverPAIN, derived from two data sets, appears to be valid and its use improves outcomes, but diagnostic studies to confirm the validity of FeverPAIN in other data sets and settings are needed.

Experienced clinicians need to identify barriers to the use of clinical scoring methods. Implementation studies that address perceived barriers in the use of FeverPAIN are needed.

Trial registration

This trial is registered as ISRCTN32027234.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1 In vitro evaluation of five rapid antigen detection tests for group A beta-haemolytic streptococcal sore throat infections

G Lasseter, C McNulty, FDR Hobbs, D Mant and P Little on behalf of the PRISM Investigators.

Abstract

Background

Using accurate and easy-to-use rapid antigen detection tests (RADTs) to identify group A beta-haemolytic streptococcus (GABHS) sore throat infections could reduce unnecessary prescription of antibiotics. Although widely used in Finland, France and the United States, uncertainty regarding RADTs' sensitivity and ability to impact on prescribing decisions has resulted in limited uptake in the UK.

Objectives

We aimed to evaluate current RADTs available in the UK in controlled, parallel in vitro trials, in order to eliminate the spectrum bias and variability in sampling associated with many clinically based studies.

Method

We compared the ease of use of five UK RADTs and their ability to detect different concentrations and strains of GABHS: OSOM® Ultra Strep A (Bio-Stat Limited, Stockport, UK), QuickVue® Dipstick Strep A test (TK Diagnostic, Oxford, UK), Streptatest® (DECTRA PHARM, Strasbourg, France), Clearview® Exact (Inverness Medical Professional Diagnostics, Bedford, UK) and IMI TestPack® Plus Strep A (Inverness Medical, Bedford, UK). We also measured whether the RADTs falsely identified common throat commensals as GABHS. All kits were tested with single-tipped polyester swabs, rayon swabs and the kit swabs provided by the manufacturer.

Results

The IMI TestPack was the easiest RADT to use. The ability to detect all positive GABHS (the sensitivity of the RADTs) varied considerably between kits from 62% [95% confidence interval (CI) 51% to 72%] for Clearview to 95% (95% CI 88% to 98%) for the OSOM and IMI TestPack at the highest GABHS concentration. None of the RADTs gave any false-positive results with commensal flora – they were 100% specific. For most of the kits, the supplied swab performed well with the exception of the Clearview pack, which performed much better with a polyester swab.

Conclusion

The IMI TestPack is suitable for use in primary care, as it had high sensitivity, specificity and was the easiest to use. If Clearview is used, a polyester swab rather than the manufacturer's swab is preferable.

Background

Most acute sore throats resolve in 1 week and are primarily viral in aetiology,[1] but, still, the majority of patients who present to UK general practitioners (GPs) are prescribed antibiotics.[2] Current therapeutic and diagnostic strategies aim to identify GABHS, which cause 5–10% of adult sore throats.

Microbiological diagnosis of GABHS sore throat infections is routinely performed by throat-swab culture. Unfortunately, culture delays results for at least 18–72 hours and, therefore, treatment must be postponed awaiting results or based on clinical characteristics present at patient consultation.[1]

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

The use of rapid point-of-care tests for GABHS offers an alternative to culture or empirical antibiotics. RADTs provide results while a patient waits in the GP's surgery and decisions regarding treatment can be based on objective evidence. RADTs have the potential to reduce unnecessary or delayed antibiotic treatment, eliminate laboratory involvement and reduce overall consultation times.

Although there is no international consensus on RADT use, these kits have been widely adopted in Finland, France and the United States. The UK Clinical Knowledge Summaries, which provide the main online guidance for GPs, do not encourage the use of RADTs, citing the limited evidence for any impact on prescribing decisions (http://www.cks.nhs.uk/sore_throat_acute).

The reliability of RADTs is variable, and often inadequate when compared with carefully performed culture.[3] Numerous clinical studies have reported the percentage of GABHS infections that RADTs can detect, which varies from 48.0% to 98.9%.[3] These discrepancies can be attributed to differences in study populations (spectrum bias), sampling techniques, RADT kits, culture methods and variations attributable to the personnel performing the tests. To ensure that RADTs are evaluated objectively, a standardised in vitro method using known concentrations of GABHS would remove the inherent biases associated with these clinical studies.[4]

The majority of near-patient RADTs for the detection of GABHS are sold by manufacturers with kit swabs provided. It is widely assumed that swab type has no impact on RADT performance, but this may be unjustified, as the kit swabs provided are specifically manufactured and validated for each RADT. Manufacturers recommend specific swab types to use in conjunction with their RADTs, safeguarding kit sensitivity and specificity. Nonetheless, clinical validation studies routinely disregard these recommendations by using a variety of swab types.

The purpose of this study was to evaluate in vitro the ease of use and accuracy (sensitivity and specificity) of the five most commonly used RADTs in Europe, and explore the implications of using different swab types.

Method

Rapid antigen detection kits

The kits tested represented the most commonly used kits in Europe, the most widely available kits in the UK and those that performed reasonably in a previous study:[4] Clearview Exact Test, IMI TestPack Plus Strep A, OSOM Ultra Strep A, QuickVue Dipstick Strep A test and Streptatest (*Figure 1*).

FIGURE 1 The five RADTs evaluated. Dipsticks, left to right: OSOM, QuickVue Dipstick and Streptatest. Cassettes, top to bottom: IMI TestPack Plus and Clearview. Showing positive reactions for all.

Sample size for kit sensitivity testing

Previous studies have shown that some RADTs are able to detect between 80% and 90% of GABHS infections.[5] Assuming that the best RADT in this study would achieve a sensitivity of 85–95%, and to estimate with 95% confidence that the sensitivity of a RADT was within ± 5% (i.e. to be confident that the sensitivity was not < 80%, which would be less useful clinically), we estimated that 1460–3920 samples were required for all five RADTs.

Sensitivity testing: ability to detect group A beta-haemolytic streptococcus

Four *Streptococcus pyogenes* strains were used that are associated with clinical sore throat infections from the National Collection of Type Cultures (NCTC) (Health Protection Agency, Colindale, UK) (*Table 1*). Streptococcal dilutions ranging from 10^2 up to 10^8 colony-forming units (CFU)/ml were tested against RADTs in duplicate (for detailed culture methods, see *Appendix 1*). To mimic conditions used to undertake RADTs in a GP surgery, each streptococcal dilution was administered as a 100-µl aliquot on to a swab and these swabs were then used to perform the RADTs (as per manufacturers' instructions).

These trials showed that all RADTs detected GABHS at 10^7 CFU/ml, whereas some kits failed to detect streptococci at concentrations < 10^6 CFU/ml. These tests were duplicated to ensure accuracy. Subsequently, each *S. pyogenes* stock solution was adjusted to achieve four dilutions of GABHS within this range: 2.5×10^6, 5×10^6, 7.5×10^6 and 10×10^6 CFU/ml. Each RADT was tested 20 times using the swab technique outlined above at each dilution, with four GABHS strains (320 tests per RADT). The final result for each RADT investigation was interpreted in conjunction with the manufacturer's instructions: positive, slightly positive, negative or invalid (see *Figure 1*).

Specificity testing: assessment of false-positives due to commensal flora

A panel of 23 commensal control strains normally found in the throat were obtained from the NCTC, the American Type Culture Collection and the National Collection of Pathogenic Fungi (see *Table 1*). Strains were evaluated in seven test solutions, which included several organisms from similar or identical genera. The concentrations of these organisms were adjusted to represent the upper limits often found in clinical samples, with each organism equal to 10^7 CFU/ml.[6–9]

Each RADT was tested 10 times with seven commensal groups (70 tests per RADT).

Ease of use of rapid antigen detection tests

A biomedical scientist and consultant microbiologist evaluated 10 characteristics of each RADT (*Table 2*) – these included packaging, shelf-life, test procedure, controls, interpretation and timing of results. RADT instruction inserts were assessed on six main features: layout, font size, general clarity, stand-alone clarity of visual instructions, clarity of instructions for determining final results and inclusion of all pertinent information (see *Table 2* for scoring). Each test kit could score a maximum total of 22 points. Price per test (2008) was calculated excluding company discounts for bulk purchases.

Swab type

The manufacturers' instructions for the RADTs evaluated specifically recommended the use of either polyester or rayon swabs (see *Appendix 2*). Consequently, all kits were tested in vitro with single-tipped polyester swabs (Ref: 170C, Copan Diagnostic, Barloworld Scientific, Staffordshire, UK), rayon swabs (Ref: 141C, Copan Diagnostic) and the kit swabs provided by the manufacturer.

Results

Sensitivity results: ability to detect group A beta-haemolytic streptococcus

All the test kits performed better with increasing concentration of GABHS (*Figure 2*). All RADTs were positive at the highest concentration of GABHS of 10×10^6 CFU/ml, whereas at lower concentrations of GABHS the kits varied in their ability to give positive results. At a GABHS concentration of 10×10^6 CFU/ml,

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 1 Strains used for in vitro RADT investigation

Strains	Group	Catalogue number		
		ATCC	NCTC	NCPF
Streptococcus pyogenes	GABHS		12,696, 8312, 8308, 10,867	
Streptococcus dysgalactiae subsp. *equisimilis*	Group 1	35,666		
Streptococcus anginosus	Group 1	33,397		
Streptococcus intermedius	Group 1		11,324	
Streptococcus constellatus	Group 1		11,325	
Streptococcus mitis	Group 1		12,261	
Streptococcus mutans	Group 1		10,449	
Streptococcus sanguinis	Group 1		10,904	
Streptococcus salivarius	Group 1		11,389	
Streptococcus pneumoniae	Group 1	33,400		
Staphylococcus aureus	Group 2	31,153	8178	
Staphylococcus epidermidis	Group 2	14,990	11,047	
Haemophilus influenzae	Group 3	9332		
Haemophilus influenzae type b	Group 3		7279	
Haemophilus parainfluenzae	Group 3		11,607	
Moraxella catarrhalis	Group 4		3622	
Neisseria pharyngis var. *flavus*	Group 4		4591	
Neisseria pharyngis var. *siccus*	Group 4		4590	
Escherichia coli	Group 5		10,418	
Pseudomonas aeruginosa	Group 5		10,662	
Candida albicans	Group 6			3091
Bacteroides melaninogenicus	Group 7	15,930	11,321	
Fusobacterium nucleatum	Group 7		11,326	
Veillonella parvula	Group 7		11,809	

ATCC, American Type Culture Collection; NCPF, National Collection of Pathogenic Fungi.

the OSOM and IMI TestPack detected 95% of the test samples as positive, whereas Streptatest detected 79% (95% CI 67% to 85%), QuickVue 70% (95% CI 59% to 80%) and Clearview 62% (95% CI 51% to 72%). The Streptatest, QuickVue and Clearview were all at least 15% less sensitive than OSOM and IMI TestPack at all concentrations of GABHS. Detailed results are available in *Appendix 3*.

GABHS strain type caused only minor fluctuations in RADT sensitivity; strain 8312 was associated with an overall sensitivity of 41% (95% CI 38% to 44%), strain 8308 a sensitivity of 48% (95% CI 45% to 51%), strain 12,969 a sensitivity of 49% (95% CI 46% to 52%) and strain 10,867 a sensitivity of 51% (95% CI 48% to 54%).

TABLE 2 General ease-of-use classification: for five RADTs for the detection GABHS, based on general characteristics and the reader's opinion

Characteristics of the test	Score	Clearview Exact Test[a]	IMI TestPack Plus[a]	OSOM Ultra	QuickVue Dipstick	Streptatest
Shelf-life						
> 12 months	2	2	2	1	2	2
≤ 12 months	1					
Number of tests per box						
> 25 units	2	1	1	2	1	1
≤ 25 units	1					
Storage volume (one box)						
< 0.003 m^3	2	1	1	2	2	2
> 0.003 m^3	1					
Steps needed						
Additional extraction step	1	2	1	2	2	2
None	2					
End point of test						
Timed 5 minutes	1	1	2	1	1	1
Visual end point	2					
Reading stability of end point						
Read at 5 minutes only	1	2	2	1	1	2
Read at 5 minutes; valid ≤ 10 minutes	2					
Timer provided						
Not needed	3	1	3	1	1	2
Yes	2					
No	1					
Internal controls						
Positive and negative control	2	1	2	1	1	1
Negative control only	1					
Opinion of readers						
Ease of performance						
Very easy	2	2	2	1	1	1
Easy	1					
Ease of interpreting results						
Very easy	2	2	2	1	1	1
Easy	1					
Score (max. total = 22)		15	18	13	14	15

max., maximum.
a Cassettes. All other RADTs were dipsticks.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

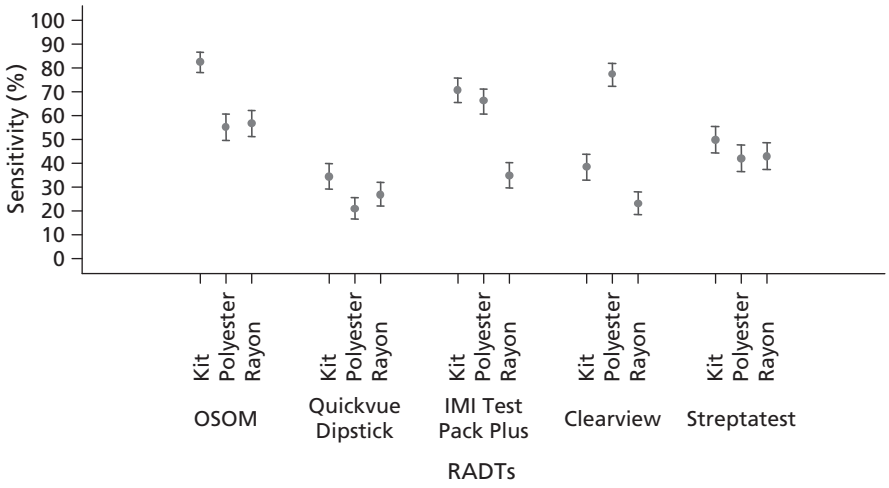


FIGURE 2 Sensitivity (with 95% CI) for five RADTs, in relation to swab type used; all GABHS strains and concentrations.

Specificity

None of the RADTs gave any false-positives when tested 10 times with each commensal group. Thus, the specificity of each RADT was 100% (95% CI 72% to 100%).

Swab type

Combining the results from all RADTs demonstrated the impact of swab type: kit swabs were associated with the highest sensitivity results overall at 55% (95% CI 53% to 57%), compared with polyester at 52% (95% CI 50% to 54%) or rayon at 37% (95% CI 35% to 39%).

Figure 2 presents the compiled sensitivity results for all GABHS strain types and GABHS concentrations, and demonstrates the most important two-way interaction between RADTs and swab type. The kit swabs provided produced the highest sensitivity results with the OSOM, QuickVue Dipstick, IMI TestPack Plus and Streptatest kits. However, Clearview sensitivity was notably better with polyester swabs at 77% (95% CI 67% to 87%); in comparison, the kits swabs provided had a sensitivity of only 38% (95% CI 28% to 48%) and rayon swabs of 23% (95% CI 13% to 33%).

Figure 3 shows the effect of swab type on RADTs when the influences of strain type and GABHS concentration are removed. Overall, RADT sensitivity was best when using the kit swabs provided at 54%. This is in comparison with polyester swabs at 52% and rayon swabs at 37%.

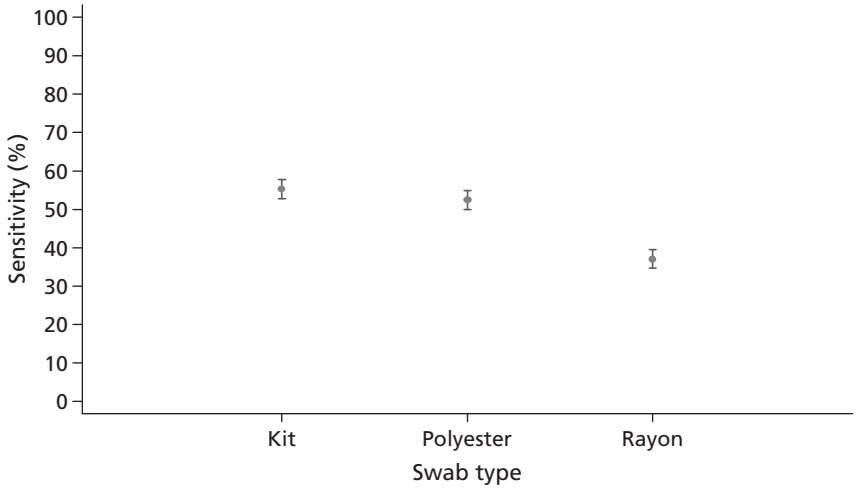


FIGURE 3 Sensitivity (with 95% CI) for each swab type; all RADTs, GABHS strains and GABHS concentrations.

Figure 4 demonstrates RADT sensitivities at all GABHS concentrations, minimising the effect of the least important variable – GABHS strain. At 7.5×10^6 CFU/ml the OSOM/kit swab combination achieved a sensitivity of 95% (95% CI 87% to 100%), which was matched only by Clearview using polyester swabs. However, Clearview and polyester swabs achieved 100% (95% CI 96% to 100%) sensitivity at 10×10^6 CFU/ml, whereas the sensitivity of OSOM and kit swabs remained at 95% at this concentration. However, it is worth noting that the sensitivity of Clearview when using the kit swabs provided was considerably reduced, reaching a maximum sensitivity of only 62% (95% CI 51% to 72%) at 10×10^6 CFU/ml, which is lower than any other RADT.

Ease of use/storage, cost and instruction clarity (see Tables 2 *and* 3*): suitability of rapid antigen detection tests for general practice*

As the study evaluated a selection of cassette and dipstick formats, all the kits had different characteristics, with each RADT having a variety of strengths and weaknesses.

The OSOM Ultra kit had a shorter shelf-life than all the others and contained the most tests per kit ($n = 50$). This may be important if surgeries are undertaking less than one test per week, as the tests may not be used before the expiry date.

The Clearview and IMI TestPack kits were provided in larger boxes than those of the other RADTs. This could present a storage problem for GP surgeries, particularly those practices with minimal storage space or those storing numerous RADTs kits.

The IMI TestPack needed an additional sample extraction step, not required for the other RADTs, which was simply the addition of another extraction reagent. However, this additional step was straightforward, adding only seconds on to the total test procedure.

The IMI TestPack had an end point that could be read at between 5 and 10 minutes, with the availability of the final results confirmed by a novel 'end-of-assay window' (see *Figure 1*). The other four kits had timed end points at 5 minutes, and two of the four kit instructions stated it was essential to read the results at 5 minutes. The IMI TestPack was deemed the easiest format for routine clinical practice because of the flexibility in reading the end point.

Overall, the cassette formats were preferred to the dipsticks, as they were easier to handle and discard. All test reagents are absorbed into the cassette test device, thus eliminating the hazard of test reagent spillages, a clear advantage for any point-of-care testing kit. Overall, the IMI TestPack achieved the highest score for general ease of use, totalling 18 out of 22 points (see *Table 2*). Streptatest and Clearview both

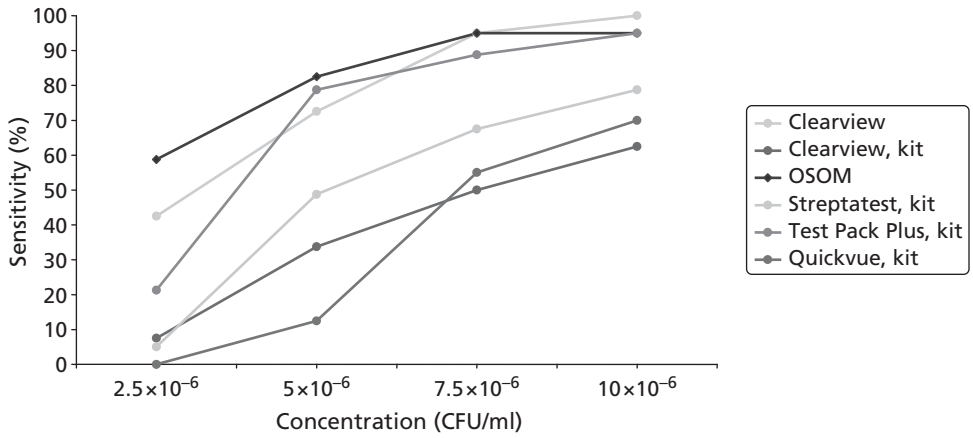


FIGURE 4 Using 'best' swab type, RADT sensitivity at each GABHS concentration; combined results for all GABHS strains.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

scored 15 points, QuickVue Dipstick 14 points and OSOM 13 points. Overall, the two cassette formats achieved higher scores than the dipsticks.

Clarity of kit instructions and price per test

The instruction clarity of the RADTs is outlined in *Table 3*, along with price per test. Notably there was no standard format used by the RADT manufacturers and, consequently, the quality of the kit instructions varied between RADTs. Streptatest was felt to have the best instructions overall: clear and simple, with a logical layout. The IMI TestPack was a close second, losing marks on small font size and the inclusion of too much information.

Overall, OSOM was the most expensive and judged to have the poorest instructions. Streptatest was the cheapest kit, with the best instruction clarity. The IMI TestPack performed well overall, marred only by price per test. Clearview was the second cheapest RADT, and had an average instruction manual. The difference in price per test was more than threefold: Streptatest was the cheapest at £1.38 and OSOM the most expensive at £3.46.

Discussion

In this in vitro study, the specificity of RADTs kits was very good. Sensitivity varied with streptococcal concentration and two RADTs were more sensitive. Kits also varied considerably in their ease of use and expense.

Strengths and limitations

Clinical throat swabs are notoriously imprecise samples: two simultaneous swabs often vary in the number of GABHS collected.[10,11] With no standard method for throat swab collection, swab material, transport medium and environmental factors can all affect GABHS collection and survival, and ultimately RADT performance. Our in vitro study removed these biases by using precise GABHS concentrations and directly comparing the performance of each RADT.

The GABHS strains used for sensitivity testing were chosen because of their association with sore throat episodes and included two mucoid strains. Testing more strains of GABHS may have altered our sensitivity findings; however, this seems unlikely, given that the four very different strains included in this study had little impact on RADT sensitivity.

This study was undertaken by a biomedical scientist with laboratory training, as we wanted to minimise any user bias in our method. Previous research has shown that the professional skills of the person performing the RADT can affect its sensitivity. Consequently, the accuracy of these tests may be affected when performed by untrained personnel in a general practice setting.

TABLE 3 Classification of five RADTs base on price per test and instruction insert clarity (April 2008 prices)

RADT	Price per test (inclusive of VAT; £)	Instruction clarity
Streptatest	1.38	Excellent
Clearview	1.87	Good
QuickVue Dipstick	2.35	Poor
IMI TestPack Plus	3.15	Very good
OSOM	3.46	Very poor

A significant proportion of streptococcal sore throats are group C or G streptococci which probably present with similar clinical features to group A streptococci.[12,13] Investigating sore throat infections with highly specific RADTs will result in cases of group C and G infections being undiagnosed and untreated. The significance and clinical relevance of this issue will be investigated further in the diagnostic studies (see *Chapters 2* and *3*).

Main findings and comparison with existing literature

The OSOM and IMI TestPack were 15% more sensitive (detected 15% more positive samples) than Streptatest, QuickVue and Clearview at all concentrations of GABHS. The OSOM kit detected more strains of GABHS at lower concentrations than the IMI TestPack. However, the OSOM kit was the most expensive kit and had poor instructions. Overall, the simple and practical IMI TestPack outperformed the other kits for ease of use, demonstrating the kit's suitability for the general practice setting. None of the RADTs gave any false-positive results with the commensal flora [specificity 100% (95% CI 72% to 100%)]. Variation in GABHS strain type had little effect on RADT sensitivity, whereas increasing streptococcal concentrations improved the sensitivities of all kits.

Only one similar in vitro study has been identified, a French-language paper by Charlier-Bret *et al.* published in 2004.[4] Our findings were in line with this study, confirming that RADT sensitivity typically falls between 10^5 and 10^7 CFU/ml. Overall, Charlier-Bret *et al.* evaluated four of the five kits assessed in our study, reporting the most sensitive RADT as the IMI TestPack, followed in decreasing sensitivity by Streptatest, QuickVue Dipstick and Clearview. The OSOM kit was not evaluated.

The cost of each RADT is likely to be a barrier for its use in the health systems such as UK primary care, in which doctors working in primary care do not pay for laboratory diagnostic tests. If health-care providers wish to encourage the use of RADTs in primary care in similar environments, they are likely to need to reimburse physicians for their use. This may well be a cost-effective policy if RADTs can be shown to reduce antibiotic prescribing more effectively than the alternative strategies.

Swab type

A previous report by Bourbeau[14] noted that swab fibre composition, swab tip preparation, swab tip characterisation (fibre, foam, flocked), shaft type (hollow, solid) and transport medium can impact on swab performance. In light of this information, and because the clinical validation part of the PRISM study used rayon swabs for validation of all kits, despite three of the manufacturers recommending polyester swabs (see *Appendix 2*), users of RADT should note that the 'best' swab type is not always that provided with the RADT kit or recommended by the manufacturer.

Conclusion

The OSOM and kit swab combination was the most consistently sensitive RADT but has some disadvantages. This study provides sufficient evidence that the IMI TestPack is suitable for use in primary care (and the most suitable for use in the trial – see *Chapter 4*), as it had high sensitivity and specificity and was easy to use. Our results showed that RADT sensitivity altered considerably depending on the swab type used. The kit swabs provided by the manufacturer with each RADT gave the most sensitive results for all RADTs except Clearview, which performed better with polyester swabs. Future clinical trials should choose swab types carefully, and manufacturers should demonstrate that their swab types provide the optimal results, as the sensitivity of RADTs can be significantly impaired by swab type.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Chapter 2 The incidence and clinical variables associated with streptococcal throat infections

P Little, FDR Hobbs, D Mant, C McNulty and M Mullee on behalf of the PRISM investigators.

Abstract

Background
Management of acute sore throat is often based on features associated with GABHS, but the features that best predict GABHS require clarification. Non-group A streptococcal strains share major similarities with group A strains, but their clinical presentation and incidence has not been clarified.

Objective
The aim of this study was to assess the incidence and clinical features associated with streptococcal infections.

Design
This study comprised a prospective diagnostic cohort.

Setting
The setting was UK primary care.

Patients
The patients included in the study were aged ≥ 5 years and presented with acute sore throat and clinical signs in the pharynx (acute pharyngitis).

Methods
The presence of pathogenic streptococci isolated from throat swabs was documented.

Results
Pathogenic streptococci were found among 204/597 patients (34%; 95% CI 31% to 38%). Of these, 33% (68/204) were non-group A streptococci – mostly C (29), G (18) and B (17); and rarely D (3) and *S. pneumoniae* (1). Patients presented with similar features whether the streptococci found were group A or non group-A. The features that best predicted A, C or G beta-haemolytic streptococci were the patients' self-reported assessment of severity (odds ratio for a bad sore throat 3.31; 95% CI 1.24 to 8.83); absence of a bad cough (2.73; 95% CI 1.56 to 4.76), absence of coryza (1.54; 95% CI 0.99 to 2.41); muscle aches rated moderately bad or worse (2.20; 95% CI 1.41 to 3.42); and clinicians' assessment of severity (severely inflamed tonsils 2.28; 95% CI 1.39 to 3.74).

Conclusion
Non-group A strains commonly cause streptococcal sore throats, and present with similar symptoms and clinical features to group A streptococci. The best features to predict streptococcal sore throat presenting in primary care deserve re-examining.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Background

Acute sore throat is one of the commonest presentations in clinical practice and most patients are treated with antibiotics despite a Cochrane review suggesting modest symptomatic benefit.[15] A reasonable strategy in reducing the public health threat of antibiotic resistance is to limit antibiotic use to the minority of individuals with streptococcal infections who are more likely to benefit[13,16,17] and avoid treatment in those unlikely to benefit. GABHS is the most frequent major bacterial pathogen in pharyngitis.

Rapid streptococcal antigen tests to detect GABHS are used widely in developed countries to target treatment, as GABHS is the most common major pathogen,[1] and the use of rapid tests may help practitioners to reduce prescribing.[18,19] Much less emphasis has hitherto been placed on other Lancefield groups (particularly C and G),[20] and there are no rapid antigen tests available for detecting these groups. The incidence of rheumatic fever, which is probably not caused by group C and G streptococci, has dramatically declined in developed countries[15] and antibiotic treatment to prevent rheumatic fever is an extremely inefficient use of health-care resources.[17] The major virulence factors among group A streptococci are shared by group C and G streptococci – particularly the M proteins, peptidase, hyaluronic capsule and streptokinase,[12,20] and similar numbers of cases of streptococcal septicaemia due to C and G and to A streptococci are regularly reported.[21] However, such complications are rare, and hence the major benefit of targeting streptococcal infections – which may apply to Groups C, G and A if clinical presentation is similar – is likely to be in limiting antibiotic treatment to individuals who will benefit from more rapid symptom resolution[15] and a shorter infective period.[16] A small study in two Norwegian practices suggested that Lancefield groups C and G presentations were similar to those of group A,[12] which is supported by Tiemstra *et al.*;[13] conversely, a substantial study concluded that their clinical presentations were different.[22] Therefore, both the relevance of C and G streptococci in symptomatic presentation and the clinical predictors require clarification. We report new data on the epidemiology of pathogens that cause pharyngitis and the predictors of the presence of pathogenic streptococci for patients presenting with pharyngitis in primary care.

Methods

Inclusion

Health professionals in general practices in the south and central areas of England recruited adults or children aged ≥ 5 years presenting with acute sore throat (< 2 weeks), when the sore throat was the predominant clinical feature (or when the clinician felt that the pharyngitis was driving the illness presentation), and with an abnormality on examination of the throat (erythema with or without pus and anterior cervical glands) – similar to a previous study in primary care.[23] Exclusion criteria were as follows: other non-infective causes of sore throat (e.g. aphthous ulceration, candida, drugs) or unable to consent (e.g. dementia, uncontrolled psychosis).

Clinical data

Following informed consent, baseline clinical data were collected by the recruiting health professional.[24–26] The case report form collected information on age, gender, current smoking status and past history of quinsy[27] as well as 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 by the patient as follows: 0 = no problem, 1 = slight problem, 2 = moderately bad problem, 3 = severe problem). The doctor or nurse documented examination findings for oral temperature using Tempa•DOT™ thermometers (3M, St Paul, MN, USA)[28] the severity of tonsillar and pharyngeal inflammation, and the presence of cervical glands, tonsillar exudate, fetor and palatal oedema.[24–26]

Throat swabs

At the training session in each practice, clinicians were instructed in standard study procedures, including how to take a throat swab. Swabs were taken by the clinician and sent to a central laboratory for culture of and sensitivity testing to all significant pathogens, in line with national standard operating procedures.[29,30] Mean time between specimen collection and receipt at laboratory was 2.9 days (data incomplete for 13 samples). The swabs were inoculated on to 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.[29,30] Plates were read after 24 hours of 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), in accordance with the national standard operating procedures.[29,30] Antibiotic sensitivities were conducted using disc diffusion techniques.[31]

Sample size

In order to determine the predictive value of clinical variables, we estimated that a subgroup of 139 patients with a clinical presentation not associated with streptococcal infection would provide estimates of a negative predictive value of 90% with 95% CIs of ± 5%, and a subgroup of 93 individuals with streptococcal infection would provide estimates of a positive predictive value of 60% with 95% CIs of ± 10%. We estimated that a sample size of 455 patients would be sufficient to detect an odds ratio of 2 (assuming $\alpha = 0.05$ and $\beta = 0.2$) for variables with a prevalence of 20–65% among patients without streptococcal infection.

Analysis

Clinical variables were included in a logistic regression model to assess their association with the presence of 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, with no evidence of collinearity. All variables significant in univariate analysis were checked again in the final model. Cases with missing data for a particular analysis were excluded. For variables with several levels (e.g. sore throat), to facilitate use in a simple clinical score (i.e. ease of implementation), a cut-off was normally made at or near an odds ratio of 2. Continuous variables were dichotomised using previous cut-offs (age $\leq$ 10 years; prior duration longer than the median of 3 days).[25] For duration, there was a progressive reduced likelihood of infection with group A streptococci. With longer prior duration, however, we dichotomised at the median for ease of implementation. We also present a version of the final model with more categories for each variables (i.e. not dichotomised, using ordered categorical variables). Such a model could potentially be used with computerised practices to document more precisely risk of streptococcal infection (*Appendix 4*). Although for non-group A streptococci to date we assumed Lancefield groups B and D and also *Pneumococcus* were not to be counted as significant pharyngeal pathogens (*Tables 2* and *3* and *Appendix 4*), given the ongoing debates about this issue,[13] we also present the multivariate analysis when these streptococci are included as potentially significant pathogens (see legends to *Tables 2* and *3*). Given the higher asymptomatic carriage rates of streptococci in children, we did not include age in the final multivariate models.

Results

In total, 70 GPs and practice nurses in south and central England recruited 606 patients from March 2007 until January 2008. Recruitment took a year because of the limited duration of recruitment in many practices – the median time spent recruiting was 3 months. However, the median recruitment rate (the number of patients/months recruiting) was 4.7 patients per month – close to the expected rate from national data.[32] Sixty-seven out of 605 (11%) patients were under age 10 years, 106/604 (18%) were smokers and 109/605 (32%) were male.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Of the 606 patients recruited, 592 had microbiology results and 567 had useable baseline clinical data.

Pathogenic streptococci were found in 202 patients (34%): of these, 136 had GABHS, and 66/202 (33%) had non-Lancefield group A streptococci – mainly groups C (27), G (18) and B (17), but also D (3) and *Pneumococcus* (1).

Patients who had GABHS strains were more likely than all other patients to have a short duration of illness (≤ 3 days), anterior cervical glands, be aged < 10 years, have a moderately bad or worse sore throat, have moderately bad or worse muscle aches, have had fever during the last 24 hours and not have a bad cough (*Table 4*). Although purulent tonsils were predictive in univariate analysis, they did not independently predict the presence of GABHS in this cohort (see *Table 4*).

Patients with group C and G beta-haemolytic strains presented with similar clinical features to individuals with GABHS strains (*Table 5* and *Figure 5*) – with the exception of age, as children were very unlikely to have C and G strains and more likely to have GABHS strains. Many of the features associated with GABHS strains were associated with the presence of C and G streptococci.

The independent clinical features associated with combined group A, C and G streptococci were as follows: rapid attendance (prior duration ≤ 3 days), moderately bad or worse muscle aches, moderately bad or worse sore throat, the absence of a bad cough, severely inflamed tonsils, age < 10 years, fever during the last 24 hours and anterior cervical glands (*Table 6*). The absence of a 'runny nose' (coryza) was also very close to significance in multivariate analysis ($p = 0.054$). There were too few patients with group B infections to assess the strain with confidence but many had similar features in terms of a severe sore throat (17/17; 100%), purulent tonsils (10/17; 59%), cervical glands (13/16; 81%), short prior duration (9/17; 53%), the absence of a bad cough (14/17; 82%) and no runny nose (13/17; 76%); however, the rate of fever in these patients (10/17; 59%) was similar to that observed in those patients from whom no streptococci were isolated and only 3/17 (18%) had severely inflamed tonsils. A fuller model – in which the variables in *Table 6* are not dichotomised – is presented in *Appendix 4*, which supports the overall findings from the simpler models, with the exception of coryza (which is no longer significant).

There is considerable variation as to how well each indicator performs in helping to rule in or rule out the presence of streptococci (*Figure 6*).

Recruitment bias

Comparing patients of higher-recruiting doctors (higher than the median – average 11.8 patients per month) with patients of lower-recruiting doctors (an average of 2.6 patients per month), there was no difference in the number of features that predicted streptococcal infections in multivariate analysis (see *Table 6* for significant features, respectively a mean of 3.3 features and 3.4 features), suggesting little or no recruitment bias based on clinical characteristics.

TABLE 4 Clinical variables in patients with GABHS (n=136) compared with all other patients

Symptom or sign	Patients with GABHS [n (%)]	Patients with no GABHS [n (%)]	Univariate odds ratio (95% CI)	p-value	Multivariate odds ratio[a] (CI)	p-value
Prior duration ≤ 3 days	94/135 (70)	213/454 (47)	2.59 (1.72 to 3.91)	< 0.001	1.92 (1.23 to 3.01)	0.004
Cervical glands	121/133 (91)	332/449 (74)	3.55 (1.89 to 6.67)	< 0.001	2.63 (1.32 to 5.23)	0.006
Severely inflamed tonsils	39/132 (30)	62/442 (14)	2.57 (1.62 to 4.07)	< 0.001	1.63 (0.98 to 2.69)	0.059
Absence of runny nose	96/135 (71)	257/454 (57)	1.89 (1.24 to 2.86)	0.003	1.29 (0.81 to 2.05)	0.284
Age group ≤ 10 years	32/136 (24)	32/455 (7)	4.07 (2.38 to 6.94)	< 0.001	3.49 (1.89 to 6.43)	< 0.001
Sore throat (moderately bad or worse)	131/136 (96)	402/454 (89)	3.39 (1.33 to 8.66)	0.011	3.26 (1.11 to 9.53)	0.031
Absence of moderately bad cough	123/136 (90)	304/455 (67)	4.70 (2.57 to 8.60)	< 0.001	4.02 (2.13 to 7.57)	< 0.001
Purulent tonsils	70/135 (52)	151/454 (33)	2.16 (1.46 to 3.19)	< 0.001	1.23 (0.79 to 1.91)	0.352
Fever (during last 24 hours)	109/136 (80)	269/455 (59)	2.79 (1.76 to 4.43)	< 0.001	1.82 (1.09 to 3.02)	0.021
Muscle aches (moderately bad)	52/135 (39)	116/454 (26)	1.83 (1.22 to 2.74)	0.004	1.85 (1.18 to 2.91)	0.008
Headache	100/136 (74)	278/454 (61)	1.76 (1.15 to 2.69)	0.009	1.28 (0.79 to 2.07)	0.318

a Multivariate model controlled for prior duration, cervical glands, severe sore throat, absence of cough, bad muscle aches and fever.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 5 Clinical variables in patients with Lancefield groups C and G beta-haemolytic streptococci and among patients who have no growth of group A, C or G beta-haemolytic streptococci

Symptom or sign	Patients with non-group A strain [*n* (%)]	Patients with no pathogenic streptococci [*n* (%)]	Univariate odds ratio (95% CI)	*p*-value	Multivariate odds ratio (95% CI)	*p*-value
Prior duration ≤ 3 days	29/45 (64)	184/409 (45)	2.22 (1.17 to 4.21)	0.015	1.74 (0.88 to 3.42)	0.110
Cervical glands	42/45 (93)	290/404 (72)	5.50 (1.67 to 18.11)	0.005	4.28 (1.27 to 14.40)	0.019
Severely inflamed tonsils	17/45 (38)	45/397 (11)	4.75 (2.41 to 9.35)	< 0.001	3.66 (1.80 to 7.44)	<0.001
Absence of runny nose	34/45 (76)	223/409 (55)	2.58 (1.27 to 5.23)	0.009	2.20 (1.06 to 4.60)	0.035
Age group ≤ 10 years	0/45 (0)	32/410 (8)	0.00 (N/A)	N/A	N/A	N/A
Sore throat (moderately bad or worse)	44/45 (98)	358/409 (88)	6.27 (0.84 to 46.50)	0.073	4.05 (0.53 to 30.90)	0.178
Absence of moderately bad cough	36/45 (80)	268/410 (65)	2.12 (0.99 to 4.52)	0.052	1.40 (0.61 to 3.21)	0.430
Purulent tonsils	25/45 (56)	126/409 (31)	2.81 (1.50 to 5.24)	0.001	1.57 (0.76 to 3.23)	0.225
Fever (during last 24 hours)	34/45 (76)	235/410 (57)	2.30 (1.14 to 4.67)	0.021	1.61 (0.75 to 3.43)	0.219
Muscle aches (moderately bad)	18/45 (40)	98/409 (24)	2.12 (1.12 to 4.00)	0.021	2.36 (1.20 to 4.65)	0.013
Headache	35/45 (78)	243/409 (59)	2.39 (1.15 to 4.96)	0.019	1.79 (0.82 to 3.92)	0.147

N/A, not applicable.

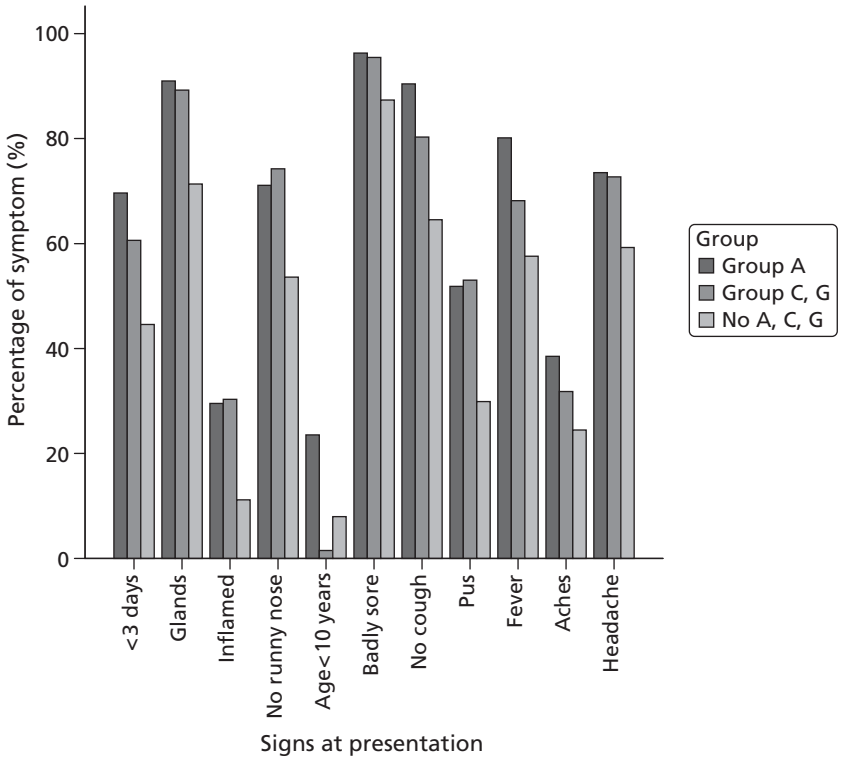


FIGURE 5 Percentage of symptoms and signs at presentation for patients infected with GABHS, group C or G streptococci or no group A, C or G streptococci.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 6 Clinical variables in patients with non-GABHS (C and G) and GABHS compared with patients who have no growth of C, G or A streptococci

Symptom or sign	Patients with group A, C or G streptococci [*n* (%)]	Patients with no group A, C or G streptococci [*n* (%)]	Univariate odds ratio (95% CI)	*p*-value	Multivariate odds ratio[a] (95% CI)	*p*-value
Prior duration ≤ 3 days	123/180 (68)	184/409 (45)	2.64 (1.82 to 3.82)	< 0.001	1.92 (1.26 to 2.92)	0.002
Cervical glands	163/178 (92)	290/404 (72)	4.27 (2.41 to 7.57)	< 0.001	2.93 (1.55 to 5.52)	0.001
Severely inflamed tonsils	56/177 (32)	45/397 (11)	3.62 (2.32 to 5.64)	< 0.001	2.28 (1.39 to 3.74)	0.001
Absence of runny nose	130/180 (72)	223/409 (55)	2.17 (1.48 to 3.17)	< 0.001	1.55 (0.99 to 2.41)	0.054
Age group ≤ 10 years	32/181 (18)	32/410 (8)	2.54 (1.50 to 4.29)	0.001	1.95 (1.05 to 3.62)	0.033
Sore throat (moderately bad or worse)	175/181 (97)	358/409 (88)	4.16 (1.75 to 9.87)	0.001	3.31 (1.24 to 8.83)	0.017
Absence of moderately bad cough	159/181 (88)	268/410 (65)	3.83 (2.35 to 6.25)	< 0.001	2.73 (1.56 to 4.76)	< 0.001
Purulent tonsils	95/180 (53)	126/409 (31)	2.51 (1.75 to 3.60)	< 0.001	1.06 (0.67 to 1.66)	0.814
Fever (during last 24 hours)	143/181 (79)	235/410 (57)	2.80 (1.86 to 4.21)	< 0.001	1.69 (1.05 to 2.71)	0.030
Muscle aches (moderately bad)	70/180 (39)	98/409 (24)	2.02 (1.39 to 2.94)	< 0.001	2.20 (1.41 to 3.42)	< 0.001
Headache	135/181 (75)	243/409 (59)	2.00 (1.36 to 2.96)	< 0.001	1.41 (0.89 to 2.25)	0.143

a All multivariate estimates adjusted for prior duration, cervical glands, severity of sore throat, severity of inflammation, the absence of cough, the absence of coryza, muscle aches and fever.

If other streptococci are included (B and D), then the significant predictors are short prior duration, cervical glands, severity of sore throat, severity of inflammations, absence of cough, absence of coryza and muscle aches.

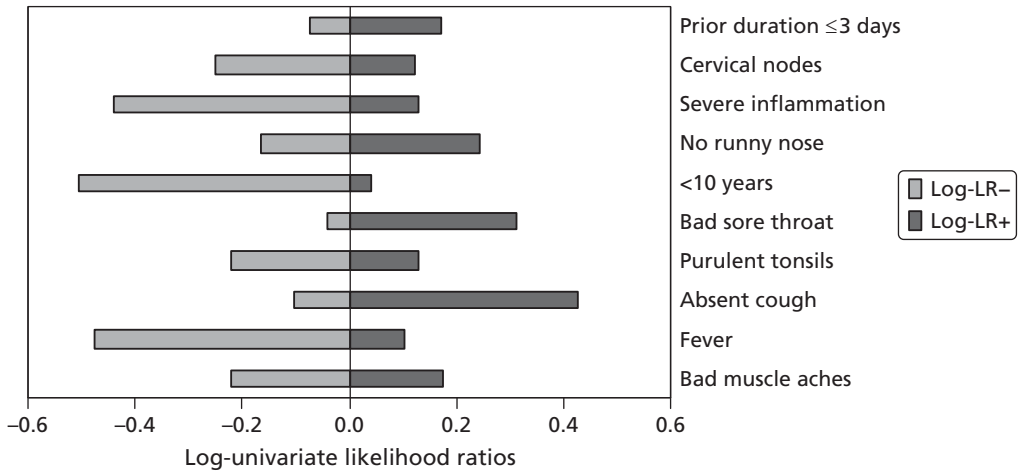


FIGURE 6 Log-univariate likelihood ratios [log of likelihood ratio for a negative test (log-LR–) and log of likelihood ratio for a positive test (log-LR+)] for individual symptoms and signs at presentation for patients with group A, C or G beta-haemolytic streptococci.

Discussion

This study documents that non-GABHS strains are a common cause of streptococcal sore throat in primary care and have similar symptomatic presentations to GABHS and that the best predictors of streptococcal infection may not include some of the features traditionally used.

Strengths and limitations of the study

The power to detect variables associated with groups C and G streptococci was limited; however, this is one of the largest studies to assess a broad range of clinical variables, and we found similar features to a smaller study reported previously.[11] Missing data were minimal (< 5% for any analysis), and, although consecutive recruitment of cases was difficult to enforce in practice, very little evidence of recruitment bias was found when recruitment rates were compared with expected rates of recruitment from national samples. Selection bias is a potentially important issue among low-recruiting doctors, but we found no evidence of clinical differences between patients of higher- and lower-recruiting doctors. Overall, fewer children than expected from historical data sets[2] were recruited, which probably reflects the reluctance of parents and/or GPs to expose children to a throat swab, but, as we elected not to include age per se in the model, the impact of this should be slight. The time between taking a swab and receipt at the laboratory was slightly longer than expected, but Lancefield groups A, C and G are not particularly sensitive to transport conditions, and we found relatively high percentages of streptococci compared with previous literature. As there was no indication from sentinel practices or microbiology laboratories of a streptococcal epidemic, the higher streptococcal percentages and more florid clinical signs than previous studies in a similar geographical area[33] may indicate changing consultation thresholds. Although the study indicates that clinical features not traditionally incorporated in making diagnostic assessments may possibly not be important, further data sets are needed before recommending a key variable set. The way variables are operationalised may also be important: McIsaac *et al.*[34,35] use tonsillar swelling or exudate, Centor *et al.*[24] just exudate and we chose exudate, as swelling is not necessarily an acute feature. We used intermediate cut points when indicated rather than the extremes of each scale; however, as the judgement of intermediate points may be more variable, using the extremes (none or very) may be more reliable for developing a clinical prediction rule.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Main findings and comparison with existing literature

Traditionally clinicians have been predominantly interested in GABHS because of their association with major non-suppurative adverse outcomes – particularly rheumatic fever.[1] Therefore, the clinical predictors of GABHS infection[1,25,36] – especially pus, cervical nodes, a history of fever and no history of cough[37] – have been widely used in clinical guidelines.[16,17,38] Historical comparisons tentatively suggest that these variables may identify a group of patients who are more likely to benefit from antibiotics.[15] We confirmed the importance of cervical glands and the absence of a bad cough and of fever.[25,34] However, this study documents that, although the feature of purulence is associated with the presence of GABHS in univariate analysis, in this data set it is not independently predictive – and other features may be important, particularly the severity of both the sore throat and the inflammation, the prior duration (reflecting a more rapid, severe onset), muscle aches and possibly the absence of coryza. Some of these features were identified previously in studies in typical primary-care settings,[12,25] but previous studies were limited by a lack of multivariate analysis or limited power.

The clinical presentation of infection with group C and G streptococci suggests strongly that not only are these presentations unlikely to be due to commensal carriage, but they are causing a similar clinical syndrome to GABHS. This supports those studies that observed similar symptomatic presentation.[12,13] If group C and G streptococci are clinically important, then rapid streptococcal antigen tests (which are targeted at GABHS only) will miss a significant proportion of streptococcal infection.

Conclusion

Group C and G streptococcal infections present with symptomatic illness in a similar manner to GABHS. RADTs, which are widely used in many developed countries to detect GABHS, will miss these organisms. The best features to predict streptococcal sore throat presenting in primary care also deserve revisiting, as features not commonly used in diagnosis (e.g. rapid presentation, severity of inflammation) may be useful clinically.

Chapter 3 Two diagnostic cohorts to identify clinical variables associated with Lancefield group A beta-haemolytic streptococci and Lancefield non-group A streptococcal throat infections

P Little, M Moore, FDR Hobbs, D Mant, C McNulty, I Williamson, MYE Cheng and M Mullee on behalf of the PRISM investigators.

Abstract

Background

Clinical variables traditionally associated with pathogenic streptococci – such as the presence of cervical glands – and clinical scores derived from these variables are commonly used to minimise antibiotic use for individuals with a low likelihood of pathogenic streptococci.

Objective

The aim of this study was to assess the association between features of pharyngitis and the growth of Lancefield group A, C or G streptococci from culture of a throat swab in two cohorts.

Design

This was a diagnostic cohort study.

Setting

The setting was general practices in the UK.

Patients

Patients included in the study were aged ≥ 5 years and presented with acute pharyngitis. Patients were recruited for a second cohort (cohort 2, $n = 517$) consecutively after the first (cohort 1, $n = 606$) from similar practices.

Results

A, C or G beta-haemolytic streptococci were found in 40% of participants in cohort 2 (compared with 34% in cohort 1). The clinical features that predicted the presence of these streptococci in multivariate analysis in both cohorts were as follows: rapid attendance (short prior duration of ≤ 3 days; multivariate-adjusted odds ratio 1.92 cohort 1, 1.67 cohort 2); fever in the last 24 hours (1.69, 2.40); and severe inflammation as assessed by doctor [severely inflamed pharynx/tonsils (2.28, 2.29)]. Absence of coryza or cough and purulent tonsils were also significant predictive variables in univariate analysis in both cohorts and in multivariate analysis in at least one cohort. A five-item score, based on Fever, Purulence, Attend rapidly (≤ 3 days), severe Inflammation and No cough or coryza (acronym FeverPAIN), had moderate predictive value [bootstrapped estimates of area under receiver-operating-characteristic (ROC) curve: 0.73 cohort 1, 0.71 cohort 2] and performed well in identifying a substantial number of participants at low risk of streptococcal infection (38% in cohort 1, 36% in cohort 2 scored ≤ 1, associated with streptococcal percentages of 13% and 18%, respectively). A Centor score of ≤ 1 identified 23% and 26% of participants, with streptococcal percentages of 10% and 28%, respectively.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Conclusions
Items widely used to help identify presentations of streptococcal sore throat in primary care may not be the most valid. 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.

Background

Antibiotic resistance is a major public health problem, driven largely by antibiotic prescribing in primary care,[39] and it is imperative to minimise antibiotic use in patients who will not benefit from it.[40] However, antibiotics are still prescribed to the majority of patients with acute sore throat, the most common upper respiratory tract infection to present in primary care.[2]

The management of acute sore throat is often based on features associated with GABHS, and clinical scores to predict GABHS have some promise to be useful,[1,25,36] including the simple Centor criteria – three out of four of pus, cervical nodes, a history of fever and no history of cough – which are widely advocated in clinical practice guidance.[16,17,34,41,42] However, these criteria have low specificity,[34] leading to high rates of overall antibiotic use.[34] 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.[12,25] 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 GABHS (see *Chapter 2*), and found that some of the variables that constitute 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), suggesting the clear need to assess a wide range of potential variables in different data sets.

We compare findings from a new cohort with the original cohort (see *Chapter 2*) 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

The inclusion criteria, the clinical data collection, and the collection and transport of swabs are as described in *Chapter 2*.

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 infection (based on our first data set), and that variables in the streptococcal group were found in 30–80% of individuals, the detection of a variable with an odds ratio of 2 required 407 individuals with complete results.

Analysis
As we found previously from the first data set that patients with group C and G beta-haemolytic strains presented with similar clinical features to individuals with GABHS, we 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 $\leq$ 10 years; prior duration longer than the median of 3 days).[25] Given the higher asymptomatic carriage rates of streptococci in children, 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 (as there will be fewer categories) and variable performance of one item in different cohorts will unduly affect reliability. Therefore, we also generated an expanded 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 8*).[43] For the Centor criteria (an established model), non-bootstrapped estimates are provided.

Results

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 useable data, and complete data were available for 460 patients. In the second data set, pathogenic streptococci were found in 207 patients (40%) – mainly A (143), C (30) and G (20), but some B (9), D (2) and F (3). These are very similar figures to the first data set (see *Chapter 2*).

The independent variables associated with Lancefield group A, C or G streptococci in the second data set are shown in *Table 7*, with the univariate 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 pharynx/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 oropharyngeal 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'.

Table 8 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 area under the curve 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 at low 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 5* for full table).

A Centor score of ≤ 1 was identified among 23% of the first cohort and 26% of the second cohort and streptococcal percentages were isolated in 10% and 28% of these groups, respectively (*Table 9*). By comparison, the extended five-point FeverPAIN (model 5 from *Table 8*) provides a finer grading of risk and significantly more patients in both cohorts can be categorised as at low risk of streptococcal infection with FeverPAIN (< 20% chance of streptococci, see *Table 9*): using the modified FeverPAIN score > 30% of patients scored ≤ 1 (first data set 38%; second data set 36%) and fewer of these patients (13% and 18%, respectively) had streptococci. This is shown graphically in *Figure 7*.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 7 Second data set: clinical variables in patients with Lancefield group A, C and G compared with patients with no growth of Lancefield C, G or A streptococci, with odds ratios (95% CI)

	Second data set				First data set	
Symptom or sign	With streptococci [*n* (%)]	No streptococci [*n* (%)]	Univariate odds ratio	Multivariate odds ratio[a]	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 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 ≤ 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)

a All multivariate estimates adjusted for other significant predictors in each data set.

Estimates of the uni- 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 8 Sequential area under the ROC curve values as successive variables added (*p*-values given for comparison with previous model unless specified)

	Model						
	1	2	3	4	5	6	
Data set	Very inflamed tonsils	+ Short duration	+ Fever last 24 hours	+ Pus	+ No cough or coryza	+ Muscle aches	Centor vs. model 5
Second data set (*p*)	0.575	0.646 (0.006)	0.689 (0.003)	0.702 (0.104)	0.713 (0.803)	0.708 (0.334)	0.650 (0.123)
First data set (*p*)	0.602	0.676 (< 0.001)	0.706 (0.017)	0.713 (0.597)	0.735 (0.025)	0.738 (0.143)	0.716 (0.291)

TABLE 9 Number of individuals with Lancefield group A, C or G streptococci (%) at each level of clinical scores, and the total number of individuals at each level (and per cent of the total sample)[a]

	Score					
Clinical score	0	1	2	3	4	Total
First data set						
FeverPAIN						
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						
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)

a For example, taking the second column, as regards FeverPAIN there were 63 individuals with score 0, which represents 11% of the sample, and, of those 63 individuals, seven (11%) had Lancefield group A, C or G streptococci.

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

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

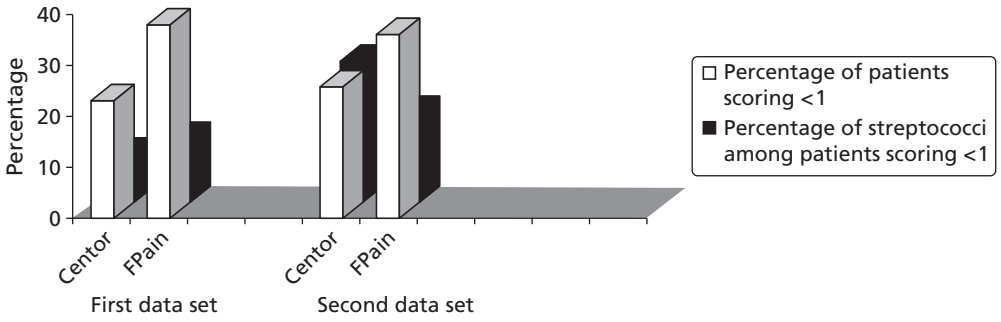


FIGURE 7 Comparison of low scores (≤ 1) for Centor criteria and FeverPAIN.

Discussion

This study provides evidence to confirm that streptococcal sore throats are common in primary care, as 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 individuals 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 (< 5% for any analysis), and little evidence of recruitment bias, either in recruitment rates or in 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 the performance of variables in these data sets – particularly striking for some variables such as cervical glands – suggests that such an approach is unlikely to provide the most valid method of variable selection for a clinical prediction rule, which is supported by similar findings in the development of clinical prediction rules for other acute infections.[44] The reason for this variation is unclear, but it may reflect varying infective agents, populations or clinician factors (e.g. reliability of history taking or examination). The variability suggests that the choice of variables to include in clinical prediction models should be based either on very large single cohorts or on multiple cohorts at different times and/or different settings. 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. Although we have provided bootstrapped estimates of the area under the ROC curve to limit overfitting, nevertheless the proposed model should have further validation.[12]

Main findings in the context of previous literature

GABHS have dominated previous literature because of their association with major non-suppurative adverse outcomes – particularly rheumatic fever and glomerulonephritis.[1] Therefore, the clinical predictors of GABHS[1,25,36] – especially pus, cervical nodes, a history of fever and no history of cough – have been widely used in clinical guidelines.[5,16,17] 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.[15] 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.[25,34] 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 inflammation and fever. These variables have been identified in studies from typical primary-care settings,[12,25] but previous studies have been limited by a 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, of these, in the second data set the percentage of patients with streptococci was high (28%). A low count (≤ 1) using a modified score (Fever SPIN) identified $> 35\%$ of patients in both data sets as 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 valid. 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.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Chapter 4 Randomised controlled trial of a clinical score and rapid antigen detection test for sore throats

P Little, M Moore, FDR Hobbs, D Mant, C McNulty, I Williamson, MYE Cheng, P Glasziou and M Mullee on behalf of the PRISM investigators.

Abstract

Objective
The aim of the study was to compare clinical scores and RADTs with delayed antibiotic prescribing.

Design
This was an open, adaptive, pragmatic, parallel-group randomised controlled trial.

Setting
The setting was UK primary care.

Patients
Patients included in the study were aged ≥ 3 years and had acute sore throat.

Intervention
An internet program randomised patients to targeted antibiotic use according to (1) delayed antibiotics (control group), (2) clinical score or (3) RADT used according to clinical score.

Outcomes
The outcomes were as follows: self-reported antibiotic use, symptom duration and severity on seven-point Likert scales (primary outcome was mean sore throat and difficulty swallowing score for the 2–4 days following the consultation primary outcome).

Results
A preliminary score to predict streptococcal infection (score 1; $n = 1129$) was replaced by a more valid score [score 2; $n = 631$; features: Fever during previous 24 hours, Purulence, Attend rapidly (≤ 3 days), very Inflamed tonsils, No cough/coryza (FeverPAIN)]. For score 1, there were no significant differences between groups. For FeverPAIN, symptom severity was documented in 80% of patients [delayed 168/207 (81%); clinical score 168/211 (80%); RADT 166/213 (78%)]. Severity was lower in the clinical score group than in the delayed prescribing group (–0.33; 95% CI –0.64 to –0.02; $p = 0.039$; equivalent to one in three rating sore throat a slight rather than moderately bad problem), and a similar reduction was observed for the RADT group (–0.30; –0.61 to –0.00; $p = 0.053$). Moderately bad or worse symptoms resolved significantly faster (30%) in the clinical score group (hazard ratio 1.30; 95% CI 1.03 to 1.63), but not in the RADT group (1.11; 0.88 to 1.40). In the delayed group, 75/164 (46%) used antibiotics, and 29% fewer used antibiotics in the clinical score group (risk ratio 0.71; 0.50 to 0.95; $p = 0.018$) and 27% fewer for the RADT group (0.73; 0.52 to 0.98; $p = 0.033$). No significant differences in complications or reconsultations were found.

Conclusion
Targeting antibiotics for acute sore throat using a clinical score improves symptoms and reduces antibiotic use. RADTs used according to a clinical score provide similar benefits, but no clear advantages over a clinical score alone.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Background

Sore throat is one of the unusual respiratory infections for which there are several reasonable diagnostic strategies: RADTs are one of the most common near-patient tests in clinical use internationally, and clinical scores to predict streptococcal infection are also widely used and advocated.[17] Using clinical scores or rapid tests has the potential to better target antibiotics, prevent progression of the illness and complications, improve symptom control and reduce overall antibiotic use compared with empirical management strategies such as delayed prescribing or no offer of antibiotics.[45] However, there is a paucity of evidence for clinical scores for most of these outcomes: recent evidence from a small Canadian trial suggests that clinical scoring methods may not help modify antibiotic prescribing, and that RADTs can significantly help in limiting the use of antibiotics, but no important patient outcomes, such as symptom control or progression of illness, were reported.[19] Further evidence is needed to confirm whether the use of RADTs can modify antibiotic use and patient outcomes.

Our earlier studies, the first in vitro and diagnostic phases of this project, provided the best evidence for choosing a RADT that is valid and widely available and developed alternative clinical scores to predict streptococcal infections. We report here the second phase of the project, which aimed to compare three strategies for limiting and/or targeting antibiotic use for sore throat: delayed prescribing, the use of a clinical streptococcal score (to predict streptococcal infection) and the targeted use of RADTs.

Methods

The trial used an adaptive design: the first part used the clinical score developed from the earliest diagnostic data from phase 1 (score 1), and the second part used a modified clinical score which included a new diagnostic cohort from phase 1 (score 2: FeverPAIN). Score 1 was used at the start of the trial, and, when FeverPAIN became available – following agreement with both the funders and the ethics committee to an adaptive design – score 2 was used in the second part of the trial.

Development of clinical scores

The clinical score development has been reported in *Chapter 3*. In brief, two diagnostic cohorts prior to this trial cohort were used to develop clinical scores to predict streptococcal infection.

Score 1

The first diagnostic cohort documented that Lancefield group C and G streptococci presented with very similar clinical features to GABHS. We developed a clinical score (score 1) that ranged from 0 to 6 and was based on a simple count of the variables that independently predicted the presence of A, C and G streptococci: rapid attendance (short prior duration of ≤ 3 days), moderately bad or worse muscle aches, moderately bad or worse sore throat, the absence of a bad cough and severely inflamed tonsils and anterior cervical glands.

Score 2

The second diagnostic cohort did not confirm that all the above variables were significantly associated with streptococcal infection and also identified new variables, which suggests that a single cohort may not be a valid enough basis to identify variables in order to generate a clinical score.

A modified score (score 2) was generated based on a simple count of five variables that were significant in univariate analysis in both cohorts and also significant in multivariate analysis in one of the cohorts: Fever during the last 24 hours, Purulent tonsils, Attend rapidly (prior duration ≤ 3 days), very Inflamed tonsils and No cough or coryza (i.e. a purely pharyngeal illness) – giving the acronym FeverPAIN.

Trial recruitment

Patients presenting with acute sore throat were recruited by health professionals in general practices in south and central England (mainly GPs but also triage practice nurses).

Inclusion

Adults, and children aged ≥ 3 years, presenting with acute sore throat were recruited (with ≤ 2 weeks of sore throat, and with some abnormality on examination of the throat – i.e. erythema and/or pus – as in our previous studies in primary care).[23]

Exclusion

Exclusion criteria were as follows: other non-infective causes of sore throat (e.g. aphthous ulceration, candida, drugs), unable to consent (e.g. dementia, uncontrolled psychosis).

Baseline clinical measures

The recruiting health professional completed clinical details at baseline on temperature (using Tempa•DOT thermometers), the presence and severity of baseline symptoms (sore throat, difficulty swallowing, fever during the illness, runny nose, cough, feeling unwell, diarrhoea, vomiting, headache, muscles ache, abdominal pain, sleep disturbance, interference with normal activities) on four-point Likert scales (none, a slight problem, a moderately bad problem, a bad problem) and the presence of the following signs: pus, nodes, tender nodes and temperature.[24–26,37] Clinicians were asked to complete non-recruitment logs, but, owing to time pressures in acute clinics, this was often not done. Clinicians documented the most common reasons why patients were not approached and why they declined in an end-of-study questionnaire.

Randomisation

Following the baseline assessment, patients were individually randomised using a web-based computer randomisation service – with permuted block sizes of 3, 6, 9 and 12 also randomly chosen – to one of three groups (see below). Originally, the protocol included stratification by clinician belief in the likelihood of bacterial infection, but following discussion with the funder this was judged to be unnecessary.

Intervention groups

The aim of the trial was to compare methods of targeting antibiotics, by comparing a RADT or a clinical score with the empirical strategy of patient choice in the use of a delayed prescription.

(a) *Delayed antibiotics (control)* A prescription was prepared and left in reception with advice to the patient to collect the prescription after 3–5 days if symptoms were not starting to settle, or sooner if symptoms became significantly worse.[33] The rationale for using delayed prescribing is that it is safe, it should, according to previous data, result in rates of antibiotic use similar to those when there is no initial offer of antibiotics, it changes belief in antibiotics as effectively as not prescribing and it modifies reconsultation more effectively than not prescribing (based on our previous studies both of sore throat[46] and lower respiratory tract infections[47]). It has been incorporated widely into routine practice in the UK since our 1997 trial[33] without any increase in complications of sore throat.[48]

(b) *Clinical score* Antibiotics were not offered to those with very low scores (0 or 1). For high scores (≥ 4, 63% streptococci), immediate antibiotics were offered, and for intermediate scores (2 or 3, 39% streptococci) delayed antibiotics were offered.

(c) *RADT group* Health economic modelling of the phase 1 results was used to estimate the most efficient use of RADTs. The modelling indicated that RADTs would be best targeted at those with intermediate and high clinical scores, for whom prescription of antibiotics was most likely. Thus, using score 2 (FeverPAIN), those with low clinical scores (0 or 1) were not offered antibiotics or a rapid test (< 20% streptococci from the first phase); those with a clinical score of 2 (33% streptococci) were offered a delayed prescription and those with higher scores (≥ 3, 55% streptococci) had a rapid test on surgery premises, and following the test patients with negative results were not offered antibiotics. All patients – as in the other groups – were advised to use analgesia (regular paracetamol and/or ibuprofen). The original intention was that the choice of RADT (the IMI test pack) would be based on

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

both a clinical study and an in vitro study. However, there was sufficient evidence of validation characteristics from in vitro studies, which provided better evidence of validity by avoiding sampling biases. Ease of use was another important consideration and was assessed during the in vitro study and also confirmed by a small panel of GPs and nurses (see *Appendix 7* at the end of the report). Taking the swab and performing the test takes approximately 5–7 minutes to get an answer.

As this was a pragmatic trial, clinicians were asked to use the intended strategy when this could be agreed with the patient, but clinicians were given flexibility to negotiate other strategies, as would happen in practice. Thus, for example, not all those randomised to delayed prescribing were given a delayed prescription.

Data collection

Patients were blind to the precise details of the groups being tested, but the pragmatic interventions, and the key outcomes (reported symptom severity and duration), made full blinding impossible. Data collection by the research team (telephone or notes review) was blind to group as far as possible, but details of patient management were available in the notes. No changes in planned outcome measures were made following trial commencement.

Daily diary

Symptoms

Patients completed a symptom diary each night until symptoms resolved or for up to 14 nights.[23,33] Each symptom was scored 0 = no problem to 6 = as bad as it could be: sore throat, difficulty swallowing, feeling unwell, fevers, sleep disturbance. A telephone call from the research assistant in the first few days aimed to resolve any problems the patient may have had filling out the diary. Temperature was taken by patients and documented on a daily basis in the diary using Tempa•DOT thermometers, as in previous studies.[23,49] If a diary was not received after 3 weeks, a brief questionnaire was sent to elicit key outcomes, and then a telephone call was made if the brief questionnaire was not received. The diary comprised information on the following:

- *Symptom severity* A two-item score (sore throat, difficulty swallowing) was chosen as the main symptomatic outcome for symptom severity, as it is more reliable than either item alone and is internally reliable (Cronbach's $\alpha = 0.92$).
- *Duration of illness* The diary permitted documentation of the duration of illness, particularly the duration of more significant illness (illness rated moderately bad or worse),[47] which is more likely to be significant for both patients and health professionals in deciding management strategies.
- *Antibiotic use* Patients reported whether antibiotics were used, as we previously showed that self-report from diary data agreed well with whether delayed prescriptions were collected.[3]
- *Side effects* Diarrhoea and skin rash were documented in the diary, and also from the notes review (see below).
- *Medicalising beliefs* Patients' belief in the importance of seeing the doctor for future episodes was documented using Likert scales completed by patients,[33] which we have shown to be reliable.[33]

Notes review

During the available follow-up time (which varied from 1 month to 2 years), all of the patients' notes were reviewed to document returns, time to return, reasons for returns, complications, side effects, economic data (see below) and any subsequent referrals.[23]

Sample size calculations

We used the NQUERY multiple group sample size programme for the three groups for all calculations.

Primary outcome: symptom severity and duration

The time when a RADT is most likely to help patients is when the inflammation due to bacterial infection is likely to be at its greatest in the first few days after seeing the doctor. We assumed the minimum effect size for the symptom severity score was a 0.33 standardised effect size [i.e. 0.33 standard deviation (SD)] on days 2–4, when patients rate their sore throat at its worst. We estimated that to detect a 0.33 standardised effect size difference between the RADT group and the other groups (assuming both control groups are 0.33 SD higher than the RADT group) required a minimum of 134 per group (for $\alpha = 0.05$, $\beta = 0.2$) or 495 in total, allowing for a 20% loss to follow-up (which was the target for the second phase of the trial). For $\alpha = 0.01$ and $\beta = 0.1$, 242 per group, or 909 patients in total, were needed, allowing for 20% loss to follow-up of diary information.[33,47] A standardised effect size of 0.33 is classified as a small effect size; 0.33 SD we estimated was equivalent to approximately half of the patients rating sore throat a mild rather than moderately bad problem,[47] or a difference between groups of the duration of symptoms rated moderately bad or worse of 1–2 days.

'Medicalising' effect of using RADT ($\alpha = 0.05$, $\beta = 0.2$)

A possible proxy for 'medicalising' behaviour is the change in beliefs about the need to see doctors in future episodes. Assuming a 15% differences between groups (57% in delayed and clinical score groups and 72% in RADT group),[33] 152 patients per group were needed. To assess the medicalising effect on reattendance, we assumed RADTs might change subsequent attendance by 11% (RADT 38%, clinical score 27%, delayed prescribing 27%) – as observed with the medicalising effect of antibiotic prescribing strategies in a previous trial over a similar follow-up period[33] – which would require 254 patients per group or 849 in total, allowing for 10% loss to follow-up of notes.[49]

Analysis

The analysis plan was finalised by the trial management group prior to performing the analysis. Analysis of covariance was performed for the main continuous outcomes (diary scores) and Cox regression for the duration of symptoms rated moderately bad or worse. Logistic regression was used for dichotomised outcomes [e.g. belief in the need to see the doctor for future episodes and return to the surgery (controlling for follow-up time)]. Odds ratios were converted to risk ratios using the method of Zhang and Yu.[50] The models controlled for baseline severity (a strong predictor of all outcomes), and for potential confounders if appropriate (in this case, fever during the last 24 hours). The primary analysis was an intention-to-treat analysis based on complete data sets – given the problem of imputing modest differences for rapidly changing symptomatic outcomes when no outcome data are available. Although a per protocol analysis was initially considered, given the pragmatic nature of the study and the leeway given to GPs to negotiate management, it was difficult to operationalise what per protocol might mean, so a per protocol analysis was not performed. The clinical and demographic characteristics of those not followed up were compared to assess, respectively, possible selection and non-response bias.
Non-differential selection bias to the trial was assessed by a comparison of the clinical features in the current trial with those in the previous diagnostic study, and also with a parallel observational cohort among a wide range of practices which used the same clinical proforma [the MRC (Medical Research Council) DESCARTE (decision rule for severe symptoms and complications of acute red throat in everyday practice) study, Little *et al.*, University of Southampton, which recruited nearly 12,000 patients]. No interim analysis was performed, and no subgroups specified in advance. To assess the differences between the two phases of the trial, interaction terms were used in the models, and, as significant interactions were found, the data from both phases were presented separately. The study team agreed in advance that, if there were significant differences between score 1 and score 2, score 2 results would be presented separately as the main results (i.e. using score 2, FeverPAIN), with the detailed results from the first part of the trial being documented in *Appendix 6*.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Results

Patients presenting in primary care were recruited from 23 October 2008 until 18 April 2011 from 48 practices overall, 46 of which recruited for the first part of the trial and 21 for the second part of the trial (see *Figure 8* for score 2; *Appendix 6*, *Figure 19*, for score 1). These practices recruited 1760 patients (1129 in the first part of the trial and 631 in the second part of the trial).

Evidence of differential effectiveness of the first and second parts of the trial

Improvement in symptom scores in the first 3 days following the index consultation was significantly greater with the second clinical score than with the first score (interaction term –0.38, 95% CI –0.76 to –0.01; $p = 0.043$), and less with the rapid test group (interaction term –0.18, –0.55 to 0.19; $p = 0.345$). There was also a significantly larger effect on symptom resolution when using the second score (interaction term hazard ratio 1.35, 95% CI 1.01 to 1.79; $p = 0.043$), but little difference when the rapid test was used (interaction term hazard ratio 1.01, 0.76 to 1.35; $p = 0.93$). Similarly, there was a significantly greater effect on antibiotic use when using the second score (interaction term odds ratio 0.43, 95% CI 0.24 to 0.76; $p = 0.004$) and a lesser effect for the rapid test group (0.74, 0.42 to 1.32; $p = 0.31$). Thus, the results of the trial when using the second score (FeverPAIN) are presented in the text as the main findings, and the results of the first score shown in *Appendix 6* – which show no significant differences for any outcome despite good compliance with the intended strategies in each group (see below).

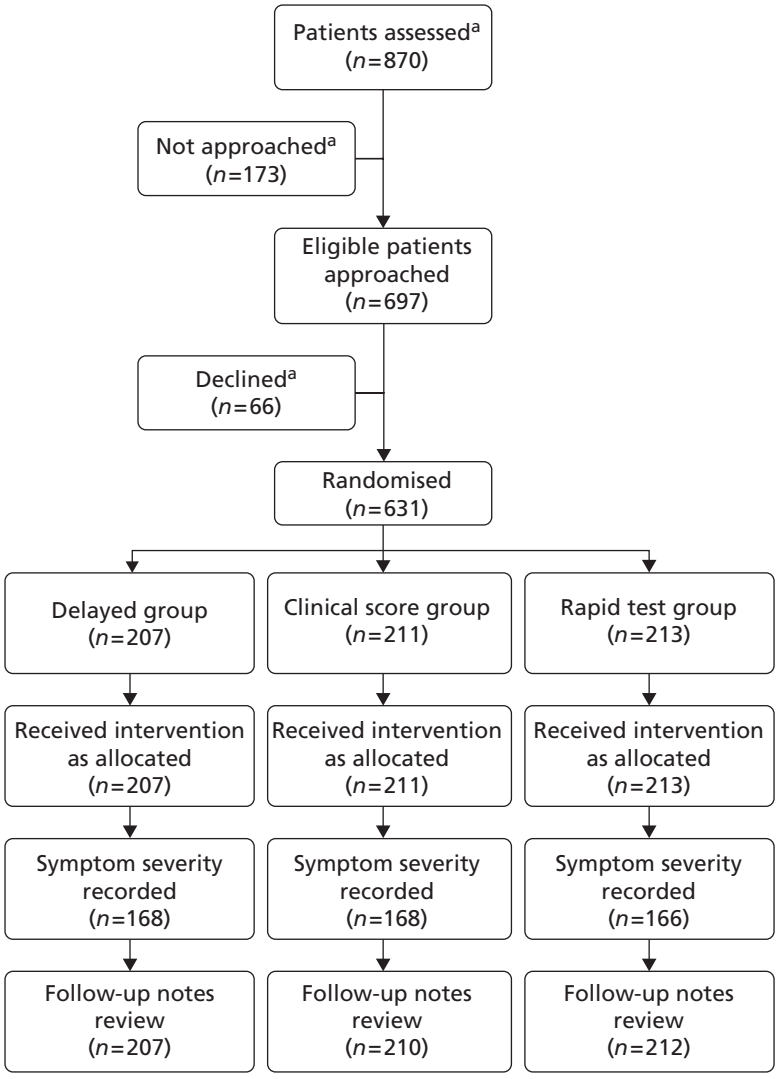


FIGURE 8 The CONSORT (CONsolidated Standards Of Reporting Trials) flow diagram for the second phase of the trial (using FeverPAIN). a, Estimates or those not assessed or declined based on physician report.

Baseline table

Group characteristics

Most baseline characteristics of the groups were similar (*Table 10*), except that fever reported in the last 24 hours was more common in the clinical score group, and, as fever did modestly change the estimates, all results are controlled for fever in addition to baseline severity of sore throat and difficulty swallowing. Female patients were slightly less common in the clinical score group, but including gender in the model made no difference to the estimates, so results are presented without controlling for gender.

Compliance with prescribing strategy

Table 10 also shows the prescribing strategy used at the baseline consultation, and demonstrates that groups were well differentiated. Compliance with the intended strategy in the clinical score and rapid test groups depended on the clinical score for each patient, and reasonable compliance with the intended strategy was demonstrated: overall, the intended strategy was undertaken in 83% of consultations (520/629) – 79% (162/205) in the delayed group, 85% (179/211) in the clinical score group and 179/213 (84%) in the rapid test group. When delayed prescribing was advised, the advice of the doctors in each group was similar: 129/162 (80%) in the delayed prescribing group, 73/89 (82%) in the clinical score group and 37/48 (77%) of patients in the rapid test group were advised to wait 5 days. More than 90% of patients in each group were advised to wait at least 3 days. Compliance with the intended strategies was also good for score 1 (see *Appendix 6*, *Table 21*).

Main findings

Symptom severity (the mean score of soreness and difficulty swallowing in days 2–4)

The mean rating of sore throat and difficulty swallowing in days 2–4 was slightly > 3 in the delayed prescribing (control) group (3 being a moderately bad problem) (*Table 11*). Both the clinical score and the rapid test groups had better symptom control [about one-third of a point; this is equivalent to one person

TABLE 10 Baseline clinical assessment and prescribing strategy used by the GP at the baseline consultation

Data taken from CRF	Group 1: delayed prescription (control)	Group 2: Clinical score only	Group 3: RADT plus clinical score
Clinical assessment			
Mean severity of sore throat/difficulty swallowing on a four-point Likert scale	3.2 (SD 0.72)	3.2 (SD 0.72)	3.2 (SD 0.71)
Prior duration (days)	4.9 (SD 3.9)	4.5 (SD 3.8)	5.0 (SD 4.6)
Age (years)	29 (SD 16)	31 (SD 17)	29 (SD 17)
Female gender (*n*)	139/207 (67%)	127/210 (60%)	138/212 (65%)
Smoker (*n*)	31/207 (15%)	39/210 (19%)	44/212 (21%)
Fever in last 24 hours (*n*)	111/207 (54%)	135/211 (64%)	112/212 (53%)
Temperature (°C)	37.0 (SD 0.72)	36.9 (SD 0.70)	36.9 (SD 0.69)
Pus on tonsils (*n*)	54/206 (26%)	56/210 (27%)	53/212 (25%)
Strategy used by clinician			
No offer of antibiotics (*n*)	21/207 (10%)	87/211 (41%)	126/213 (59%)
Immediate antibiotics (*n*)	21/207 (10%)	33/211 (16%)	38/213 (18%)
Delayed antibiotics (*n*)	164/207 (79%)	91/211 (43%)	48/213 (23%)

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 11 Symptom severity, antibiotic use, intention to consult in the future (moderately likely or more likely) and reconsultations with sore throat

Outcome	Delayed antibiotics (control)	Clinical score	RADT
Mean severity of sore throat and difficulty swallowing on days 2–4 (seven-point scale: 0 = no problem; 6 = as bad as it could be)			
Crude mean	3.11 (SD 1.49)	2.88 (SD 1.52)	2.83 (SD 1.62)
Adjusted mean difference[a]		–0.33 (95% CI –0.64 to –0.02; $p = 0.039$)	–0.30 (95% CI –0.61 to 0.004; $p = 0.053$)
Duration of symptoms rated moderately bad or worse			
Median duration (interquartile range; days)	5 (3 to 7)	4 (2 to 6)	4 (2 to 7)
Hazard ratio[a]	1.00	1.30 (95% CI 1.03 to 1.63; $p = 0.028$)	1.11 (95% CI 0.88 to 1.40; $p = 0.372$)
Antibiotic use			
Crude percentage [*n* (%)]	75/164 (46)	60/161 (37)	58/164 (35)
Risk ratio[a]	1.00	0.71 (95% CI 0.50 to 0.95; $p = 0.018$)	0.73 (95% CI 0.52 to 0.98; $p = 0.033$)
Belief in the need to see the doctor in future episodes (slightly likely or less)			
Crude percentage [*n* (%)]	62/163 (38)	54/155 (35)	64/161 (40)
Risk ratio[a]		0.97 (95% CI 0.71 to 1.27; $p = 0.847$)	1.03 (95% CI 0.76 to 1.32; $p = 0.855$)
Return within 1 month with sore throat			
Crude percentage [*n* (%)]	17/207 (8)	17/210 (8)	13/212 (6)
Risk ratio[a]	1.00	0.91 (95% CI 0.47 to 1.72; $p = 0.777$)	0.74 (95% CI 0.36 to 1.47; $p = 0.397$)
Return after 1 month with sore throat (mean follow-up 0.73 years)			
Crude percentage [*n* (%)]	31/207 (15)	26/210 (12)	34/211 (16)
Risk ratio[a]	1.00	0.79 (95% CI 0.47 to 1.29; $p = 0.353$)	1.06 (0.66 to 1.63; $p = 0.813$)

a All models controlled for baseline symptom severity (of sore throat and difficulty swallowing) and fever during the previous 24 hours.
Model for return within 1 month also controlled for prior antibiotic use; model for returns after 1 month additionally controlled for prior attendance with sore throat, and follow-up duration.

in three rating sore throat and difficulty swallowing a slight rather than a moderately bad problem (see *Table 11*), although the rapid test group just failed to reach significance].

Duration of moderately bad symptoms

In the delayed prescribing (control) group, symptoms rated moderately bad or worse lasted a median of 5.0 days. Resolution was significantly faster (30%) in the clinical score group [risk ratio 1.30 (95% CI 1.03 to 1.63)] – equivalent to saving a day of moderately bad symptoms. Resolution was 11% faster in the rapid test group but not significantly (see *Table 11*).

Use of antibiotics

Of the delayed prescribing group, 46% reported using antibiotics. The other two groups had even lower antibiotic usage: 37% in the clinical score group and 35% in the RADT group – an estimated 29% relative reduction (risk ratio 0.71) and 27% relative reduction (risk ratio 0.73), respectively (see *Table 11*).

Belief in the need to see the doctor in future

There was a trivial difference in belief in the need to see the doctor when either treated as a continuous variable or dichotomised (see *Table 11*).

Return to the surgery

There were no significant differences in return to the surgery during the following month or the subsequent follow-up.

Complications

There were no suppurative complications (otitis media, sinusitis, quinsy or cellulitis) in any group in either phase of the trial. Fewer than 1% of patients returned with either skin rash or diarrhoea within a month of the index consultation in any group for FeverPAIN (delayed prescribing 0/207, clinical score 2/210, rapid test 1/211), and there were similar findings for score 1 (delayed prescribing 5/374; clinical score 0/380; rapid test 1/359).

Selection and attrition bias

There was no evidence of clinical selection bias when comparing the patients in the two parts of the trial: score 2 and score 1 patients were similar for mean severity of sore throat and difficulty swallowing at baseline (3.20 vs. 3.12, respectively), as well as mean prior duration of illness (4.77 vs. 4.41 days), purulence [163/628 (26%) vs. 294/1124 (26%)], fever [358/630 (57%) vs. 645/1126 (57%)], mean age (29 vs. 29 years) and female gender [404/629 (64%) vs. 732/1128 (65%)]. Follow-up for diary information was also similar [502/631 (80%) vs. 901/1129 (80%)].

Based on clinician report, most patients who were assessed and approached agreed, which suggests the major issue for selection bias is likely to be which patients clinicians assessed and approached. The most common reasons given for not approaching patients were too little time to consent and recruit (87% of health professionals rated this the most common reason), followed by patients being too unwell (10%). Similarly, the most common reasons given for patients declining were having too little time to complete study materials (75%) or too little time to consent and be recruited (16%). The characteristics of patients recruited by GPs at high-recruiting surgeries [> 80 patients per year, contributing 49% (309/631) of score 2 patients] were similar to those of patients from lower-recruiting surgeries (mean sore/difficulty swallowing scores at baseline 3.2 vs. 3.2), although the mean number of streptococcal score features was slightly higher for those from high-recruiting surgeries (1.9 vs. 1.6, respectively). There was some evidence of modest clinical selection bias found in the trial population when it was compared with the previously described diagnostic studies. Although the diagnostic study used to generate FeverPAIN was similar for fever (59%), mean temperature (36.9 °C) and mean sore/difficulty swallowing score (diagnostic 3.1, trial 3.2), the baseline mean number of streptococcal score features was slightly lower in the trial population (diagnostic 2.1 vs. trial 1.7, SD 1.3). Furthermore, the purely observational cohort (MRC DESCARTE) had very similar patient characteristics to the diagnostic study (mean sore/difficulty swallowing score 3.0; mean number of streptococcal features 2.0). Given the similarity of patient-rated severity in all these studies, the slightly lower number of streptococcal features in the trial population may represent some recruitment bias in some practices based on recruiting fewer patients with more florid clinical signs. However, if practices that recruited patients with lower mean streptococcal scores (< 1.33) are excluded (leaving 448 patients), the baseline streptococcal scores for the trial are similar (2.1 diagnostic vs. 2.0 trial), and sore throat and difficulty swallowing symptoms are significantly better for the clinical score group (interaction term –0.60, 95% CI –1.01 to –0.20; $p = 0.004$), and also better for the rapid test group, but not significantly (interaction term –0.26, –0.67 to 0.16; $p = 0.223$). If this group of practices is then selected (i.e. excluding practices with low streptococcal features), the mean severity of sore throat and difficulty swallowing is

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

improved by half a point in the clinical score group (–0.51, –0.87 to –0.15; $p = 0.005$), which is equivalent to one patient in two rating sore throat a slight rather than a moderately bad problem, with a similar effect in the rapid test group (–0.38, –0.75 to –0.01; $p = 0.042$) (see *Table 11* for comparison). Thus, the estimates for the trial are probably conservative.

Discussion

To our knowledge, this is only trial to date to compare the impact of using a RADT and clinical scoring methods on both symptom control and antibiotic use for acute sore throat, which is the most common upper respiratory tract infection managed in clinical practice. It demonstrates that there are likely to be improvements in both symptom control and antibiotic use compared with empirical delayed prescribing when using either a clinical scoring method or a RADT. These results also suggest that combining the RADT with the clinical score confers no additional benefit over using the clinical score alone.

Strengths and potential limitations

Type II error (power)

We were constrained to score 2 (FeverPAIN) results for our main findings, which reduced the power to assess dichotomous outcomes, although the trial was adequately powered for symptomatic outcomes.

Type I error (chance)

Symptomatic outcomes and antibiotic use changed significantly in the same direction in both the clinical score and rapid test groups, which suggests that these results are unlikely to be due to chance.

Confounding

Although the overall trial population (both scores) was well balanced, the groups for score 1 and score 2 were slightly unbalanced (in opposite directions) for fever reported in the last 24 hours. However, the estimates controlled for this in analyses, and in other respects trial groups were similar.

Selection bias

An important limitation of all trials is that health professionals may use their judgement in selecting trial patients. Time pressures were the main reason given for not recruiting patients, but there was some evidence of modest clinical selection: the baseline severity scores and the number of streptococcal features were slightly lower in the trial than in the prior observational studies. However, differences were modest (mean streptococcal score at baseline 15% lower), and excluding practices with low scores did not alter the inferences and, if anything, increased the estimates – suggesting that the trial results are likely to be conservative.

Compliance

This was a pragmatic trial, so, although clinicians were encouraged to use the intended strategy, there was variation as happens in negotiation with patients in everyday practice; nevertheless, compliance with the intended intervention was good. In particular, the results for score 2 cannot be explained by greater compliance of clinicians over time, as score 1 compliance was as good, if not slightly better than for score 2.

Main results in context of previous literature

Using either RADTs or a clinical score (FeverPAIN) is likely to improve symptom control moderately compared with an empirical management strategy of delayed prescribing. As the score requires two examination components, it will be more difficult to use in telephone triage or out-of-hours contexts; it is likely to be difficult for patients to assess the severity of inflammation, although it might be feasible for patients to report the presence of pus. The effect on symptoms of the score in the first few days after seeing the doctor is to make a difference of one person in three rating sore throat and difficulty

swallowing a slight problem rather than a moderately bad problem. Similarly, antibiotic use was lower in both of the intervention groups. These results also suggest there is likely to be little additional benefit either for symptom management or antibiotic use when using a strategy for rapid antigen testing targeted by clinical score when compared with using a clinical scoring method alone. This might be because the diagnostic advantage in RADTs that identify GABHS is in part matched by the disadvantage of not identifying group C and G streptococci, which provide similar symptom burden to GABHS. The additional disadvantages of RADTs are the costs of the test (although most RADTs cost less than £5, and the one chosen for the current study was approximately £3) and the staff time taken to get a result (5 minutes). A previous small trial of RADTs[19] demonstrated that using the Centor score with or without rapid tests on its own did not modify antibiotic use, but this trial unfortunately did not report symptomatic outcomes. The difference between these trials may be that the current trial used a slightly more valid score: we have shown that individual items from the Centor score and score 1 did not perform optimally in our two prior diagnostic cohorts in terms of identifying patients with a low likelihood of streptococcal infection.

Based on the results of the previous diagnostic study, the limited performance of score 1 in ruling out streptococcal infection might help explain the differences in antibiotic use, but it is less clear why symptom management should be significantly better with FeverPAIN than with score 1. It may be that the particular combinations of more florid symptoms in FeverPAIN (e.g. fever, pus) are more important in determining symptom burden, and/or better in determining symptom response to antibiotics due to either microbiological or patient factors (e.g. the differential nature of organisms on the surface and in the crypts of the tonsils, or the relation between symptoms, the immune response and prognosis).[1,51] Furthermore, as our qualitative data suggest, doctors prefer clinical intuition to rigid use of clinical scores, it is also possible that FeverPAIN has greater clinical face validity for clinicians or patients. This could perhaps facilitate stronger advocacy and potentially better adherence to the proposed prescribing strategy – although our relatively crude data concerning what health professionals did (but not how they did it) suggest health professionals complied reasonably well with the proposed prescribing strategies.

The rate of antibiotic use reported in the current trial with delayed prescribing (> 40%) is significantly higher than previous research,[33] but the number of features associated with streptococcal infection is also higher (e.g. 15% had purulent tonsils in the previous trial compared with 26% in this trial), so it may be that more patients with milder sore throats are now self-caring rather than consulting their GP compared with 15 years ago. This trial also recruited in slightly different settings: the first trial recruited predominantly in deprived inner-city settings, whereas the current trial had a wider range of practices.

We were unable to demonstrate any difference in belief in the need to see the doctor, or reconsultations either within a month or with longer follow-up (i.e. no apparent 'medicalising' effect of the RADT strategy). This contrasts with a dramatic medicalising effect of prescribing antibiotics.[23,33] However, the lack of an effect with a RADT in a trial setting in which RADTs are not used routinely may mean that it is difficult to demonstrate medicalisation in the short term. Longer-term implementation studies or possibly international comparison studies may be needed.

Conclusion

Compared with empirical delayed antibiotic prescribing for acute sore throat, a clinical score improves both symptoms and antibiotic use. Using the clinical score combined with targeted use of a RADT provides similar benefits, but no clear advantages compared with using a clinical score alone.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Chapter 5 A qualitative study of general practitioner, nurse practitioner and patient views about the use of rapid streptococcus antigen detection tests in primary care: 'swamped with sore throats?'

G Leydon, L McDermott, M Moore, I Williamson, FDR Hobbs and P Little on behalf of the PRISM investigators.

Abstract

Background
The use of clinical scoring methods and RADTs could help target antibiotics at those who are most likely to benefit from them and help reduce antibiotic use, but there has been little exploration of what the key issues might be for patients and health-care practitioners (HCPs).

Objective
The aim of this study is to explore patient and HCPs' views of clinical scores and RADTs.

Method
This was a qualitative study carried out in the primary-care setting. Semi-structured face-to-face and telephone interviews were conducted with GPs and nurse practitioners (NPs) from general practices across Hampshire, Oxfordshire and the West Midlands.

Results
In total, 51 participants took part in the study. Of these, 42 were HCPs (29 GPs and 13 NPs) and nine were patients. HCPs could see a positive role for RADTs in terms of reassurance, as an educational tool for patients and for aiding inexperienced practitioners, but also had major concerns about RADT use in clinical practice. Particular concerns included the validity of the tests (the role of other bacteria, and carrier states), the tension and possible disconnect with clinical assessment/intuition, the issues of time and resource use, and the potential for medicalisation of self-limiting illness. In contrast, however, experience of using RADTs over time seemed to make some participants more positive about using the tests. Moreover, patients were much more positive about the place of RADTs in providing reassurance and in limiting their antibiotic use.

Conclusions
It is unlikely that RADTs will have a (comfortable) place in clinical practice in the near future until health professionals' concerns are met, and they have direct experience of using them. The routine use of clinical scoring systems for acute upper respiratory tract illness also faces important barriers related to clinicians' perceptions of their utility in the face of clinician experience and intuition.

Background

At present, RADTs are widely used internationally. However, there are concerns about the use of RADTs based on anecdotal evidence that suggests that the procedure may be time-consuming and increase patient expectation for the test, and thus result in increased levels of patients consulting with a sore throat.[52] Limited qualitative work has been carried out regarding patients and practitioners' use of

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

point-of-care tests for respiratory infections. Qualitative work with both patients and HCPs across Europe suggests both patients and HCPs may be willing to use CRP (C reactive protein) as a point-of-care test in lower respiratory tract infection, but major issues for HCPs are questionable test performance, problems interpreting results, a detraction from clinical reasoning, costs, time and patients not wanting, or demanding, the tests.[53] We are unaware of any qualitative research that has documented and explored HCPs' or patients' perceptions of RADTs.

This study provided a nested qualitative component to the PRISM trial, which aimed to provide insight into the use of RADTs and clinical scoring methods.

Objective

The study aimed to explore HCPs' (NP and GP) and patients' experiences of and attitudes towards RADTs.

Methods

Participants and procedure

Participants were identified and recruited (by JB, JK, KM) from practices across the south and south-east of England, aiming for a mix of inner-city and rural locations (*Table 12*). HCPs were eligible for inclusion if their practice had consented to take part in the PRISM trial. NPs were eligible for inclusion if they had responsibilities that would involve using a RADT in practice. Patients were eligible for inclusion if they had experienced a RADT during a consultation.

The interviews

Trained interviewers (TL, RC, HH) conducted both face-to-face and telephone interviews (according to participants' preference) with each lasting approximately half an hour. All interviews were audio recorded and transcribed verbatim in preparation for analysis.

Qualitative interviews provided the best method for gathering insights into participants' individual experiences of using RADTs, and their views about the RADTs. The interview guide was amended before the final 11 interviews were conducted in order to include a question that asked about experiences of

TABLE 12 Practice demographics

Registered practice population size	
Mean	4049
Range	3272–16,596
IMD score	
Mean	11.55
Range	2.70–36.92
England IMD score	
Average	21.7
Lowest	2.6
Highest	68.9

IMD, Index of Multiple Deprivation.

using clinical scoring methods. A semi-structured interview guide included key topic areas, while also providing flexibility to explore unanticipated issues.

Analysis

Inductive thematic analysis[54] was conducted on all transcripts to determine HCPs and patients' experiences and views of RADTs in practice. The first phase involved 'immersion' in the data, that is transcripts were read and reread a number of times in order for the researchers to become as familiar as possible with the data. The transcripts were then read again and meaningful sections of each transcript were systematically assigned a code to reflect issues which were represented within them. Each code label referred directly to the operationalisation of the theme discussed. This process was repeated after initial coding was complete and themes that consistently occurred within the data were identified and labelled. The codes and definitions were refined during a continuing process, which involved themes being linked, grouped, moved, relabelled, added and removed to produce a set of themes and subthemes that adequately fitted and thoroughly explained the data. Two qualitative researchers (LM and GL) iteratively developed the themes through discussion and adjustment when appropriate, before reaching 'inter-rater' agreement on the final themes and descriptions.

Findings

Participants

In total, 51 participants took part in the study. Of these, 42 were HCPs (29 GPs and 13 NPs) and nine were patients. GPs/NPs were recruited from 15 practices across Hampshire, Oxfordshire and the West Midlands. Registered practice population size varied from 3272 to 16,596. Surgeries were recruited from both inner-city and rural locations, with the Index of Multiple Deprivation score ranging from 2.7 to 36.9.

Themes

The analysis identified five main themes, and a total of 18 subthemes (*Table 13*).

In the following sections we describe each theme in turn, using exemplary quotations for illustrative purposes.

Practicalities

When invited to discuss their practical experiences of the rapid tests, HCPs' views highlighted five key subthemes, all relating to the practicalities of RADTs.

Reduces antibiotic prescribing

In terms of positive practical feedback, a few practitioners in the trial reported a reduction in their antibiotic prescribing rates for sore throat as a result of using the RADTs:

> *It [use of antibiotic] has substantially decreased since I started the trial.*
>
> *Int. 43, GP*

Easy to use

Despite concerns about the tests increasing consultation lengths (reported in a later section), HCPs reported that the tests were easy to use once they were familiar with them:

> *It was not that it was a particularly difficult test, it was just the first time you use something you need to know what to do.*
>
> *Int. 29, GP*

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 13 Themes identified in analysis

Theme	Subtheme
(1) Practicalities	Benefits: • Reduces antibiotic prescribing • Easy to use • Useful for inexperienced staff Concerns: • Time concerns • Cost concerns
(2) Accuracy of diagnosis	Benefits: • Confirmation of diagnosis Concerns: • Conflict of opinion • Identification of carriers • Alternative bacteria not detected
(3) Patient outcomes	Benefits: • Reassurance for patient • Education tool for patient Concerns: • Medicalisation concerns
(4) Experience of using clinical scores	Benefits: • Useful for inexperienced practitioners Concerns: • Unnecessary for experienced practitioners • Time-consuming
(5) RADT trial participation	Trial participation increases positive views Negative perceived views of non-trial practitioners Positive views of trial practitioners
(6) Patient views	Reassured by test No change in GP attendance Preference for no antibiotics

Useful for inexperienced practitioners

In terms of the practicalities or perhaps skill of appropriate prescribing, some of the more experienced interviewees suggested that the tests could support prescribing decisions for new or inexperienced staff members:

> *If you're in a teaching or training practice, then it's useful, to get the clinicians to learn.*
>
> *Int. 48, GP*

Time concerns

Perhaps unsurprisingly, given the time pressures and short consultation times in general practice and the challenges of incorporating new practices within them, time concerns were commonly reported. In short, interviewees reported concerns that the RADTs took up too much time during a consultation. However, HCPs did acknowledge that, as the RADT use in this instance formed part of a trial, some of the additional time involved in explaining the study had added considerably to the burden of the test and, outside of a trial context, use of the test may not be as time-consuming:

> *It's the additional time that's needed to do it, is the major problem.*
>
> *Int. 47, GP*

> *Usually the longest bit is the patient reading through all the information about the test.*
>
> *Int. 30, GP*

Practitioners who took part in the trial also reported that their concerns relating to time constraints were reduced once they had started using the test and were clearer on what to do:

> *. . . my first thought when the girls [trial researchers] came to talk to us about it was, oh gosh, how long is it going to take? And then once you realize that, if you were doing it regularly, you could do it much quicker and that's not a problem.*
>
> *Int. 25, GP*

Cost concerns

Practitioners were consistently also concerned about the cost implications of using the tests. In particular, participants were concerned that RADTs would increase the expenditure of the surgery:

> *I don't think we would use them, if there was a cost issue. Why do I say that? Because all our other pathology is free.*
>
> *Int. 26, GP*

Accuracy of diagnosis

One of the key potential benefits of using RADTs in practice is the promise of improving the accuracy of diagnosis. Four subthemes related to the accuracy of diagnosis when using the RADT were identified, and each is taken in turn.

Confirmation of diagnosis

When congruent with a HCP's diagnosis or management disposal, the RADT result could work to reassure the HCP and confirm their opinion of the diagnosis:

> *It's been reassuring and it reinforces my credibility of making a clinical judgement*
>
> *Int. 27, GP*

Conflict of opinion

In contrast, HCPs reported that they would find their treatment decisions difficult when there was a conflict between the RADT results and their opinion of the most likely diagnosis:

> *If it's a negative test and that person's got clinical signs of a bacterial sore throat . . . you then have to say, err, what do I do here?*
>
> *Int. 43, GP*

> *I haven't yet evolved, because it's not available to me . . . a way of handling a positive test that I actually don't want to treat.*
>
> *Int. 42, GP*

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

This resonates with other supported decisions in general practice, such as the use of severity measures for the diagnosis and management of depression, for which GPs have described a careful weighing of the measured evidence versus their GP intuition based on experience.[55]

Identification of carriers

Practitioners reported concern that the RADT would produce a positive test result for patients who were only carriers for GABHS, and could result in an unnecessary prescription of antibiotics:

> *. . . I think they're quite limited really because, because of the fact that streptococcal, you know, that strep is carried so widely anyway . . . .*
>
> *Int. 15, GP*

Failure to detect alternative bacteria

Practitioners were also concerned about the fact that the RADT cannot detect all forms of bacteria:

> *This test just checks for streptococcal, it doesn't check for the other bacteria.*
>
> *Int. 46, GP*

> *I mean it doesn't cover everything does it? It just covers some of the streps doesn't it. . . . don't know how many it covers.*
>
> *Int. 29, GP*

Patient outcomes

Practitioners were invited to reflect on the impact the RADTs might have on patient outcomes. Three subthemes referring to patient-related outcomes were identified; again, each is taken in turn.

Reassurance for the patient

Linked to views that a test score can enhance the credibility of a treatment disposal, HCPs described how the RADT might quickly provide reassurance for patients in the consultation:

> *I think . . . they find it very reassuring, as it's so rapid.*
>
> *Int. 28, GP*

> *You've got your answer. . . . and you can reassure the patient that you've got the answer, rather than what we tend to say is "well, this may or may not be".*
>
> *Int. 51, NP*

Education tool for patient

In a related fashion, interviewees described how they had used the RADT as a tool to assist in patient education. More particularly, such tests could be used to explain why antibiotics were not always necessary for sore throat:

> *It [RADT] helps us to explain to patients why we're not prescribing for a condition that, hopefully, is going to be self-limiting.*
>
> *Int. 47, GP*

> *You are also able to enter into the discussion with them about the, you know, the supposed benefits of antibiotics, the fact they won't work for everything and it allows you to open this door or viral versus bacterial.*
>
> *Int. 14, GP*

Medicalisation concerns

Practitioners reported concern that the increased availability of RADTs could lead to patients viewing all sore throats as more serious and thereby increase consultations because of demand for the test:

> *I think people would be, oh, and when I went to the GP surgery they did this special test and I think . . . if you were a surgery who were offering it you'd be swamped with sore throats, and if you were a surgery who weren't offering it, you'd be swamped with, when my friend up the road gets a special test. So, yeah, I think it might medicalise . . . something that will go away on its own anyway.*
>
> *Int. 04, GP*

> *The problems come when people then start to think they have to have it done every time.*
>
> *Int. 27, GP*

Experience of using clinical scores

The last 11 interviewees were asked to discuss their views and experiences of the use of the clinical score. None of the HCPs reported using the clinical scoring method prior to taking part in the trial. Although there were some positive perceptions of the scoring, the majority of comments and views on the scores were negative. Three subthemes relating to experiences of using the clinical scores were identified and these are described below.

Useful for inexperienced practitioners

Although most practitioners felt that the scoring was unnecessary for experienced members of staff, it was consistently reported that the clinical score did have a 'place within primary care' as a useful educational tool for inexperienced practitioners:

> *If you're a teaching or training practice, then it's useful to get the clinicians to learn and think about this in a much more systematic manner.*
>
> *Int. 48, GP*

Unnecessary for experienced practitioners

By contrast, the most common complaint regarding the clinical scores was that they were unnecessary for an experienced HCP. GPs felt that the questions that form part of the score should be asked when taking the history and the additional documentation of these was unnecessary. However, it was noted that NPs did not support this view and felt that the scores would be a useful tool:

> *One doesn't want to complicate incredibly minor medicine with tools and scanning on bits of paper.*
>
> *Int. 44, GP*

> *There's scores for everything at the moment and I think sometimes, you know, if I'm honest, you don't need a doctor to be scoring things.*
>
> *Int. 47, GP*

Time-consuming

Some practitioners felt that the clinical scoring was too time-consuming to fit into usual practice. In particular, it was the time taken to document the score, as opposed to conducting the score itself, that was viewed as a problem.

> *It's very time consuming, you know, you're whizzing through a busy surgery . . . one would hope one knew the right questions to ask.*
>
> *Int. 44, GP*

Rapid test trial participation

All participants were invited to discuss their views and opinions following participation in the RADT trial. Three subthemes were identified.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Trial participation increases positive views

Practitioners described how their views about RADTs had changed positively since being in the trial:

> *. . . my first thought when the girls [trial researchers] came to talk to us about it was, oh gosh, how long is it going to take? And then once you realize that, if you were doing it regularly, you could do it much quicker and that's not a problem*
>
> *Int. 25, GP*

Negative perceived views of non-trial practitioners

Practitioners suggested that GPs and NPs who had not been part of the PRISM trial would be reluctant to use the RADT:

> *My worry is that if you try to sell it to people who have not been part of the PRISM trial, they won't do the test at all.*
>
> *Int. 26, GP*

Positive views of trial practitioners

In contrast, practitioners who had taken part in the trial reported that in general they would use the RADT in general practice, if available:

> *. . . by doing the rapid test, you are [likely] to have less consultations and repeat consultations and that again gives me a personal benefit*
>
> *Int. 28, GP*

Patient views

Three subthemes relating to patient experiences with the RADT emerged.

Reassured by test

Patients described feeling reassured with their diagnosis and the treatment option that had been provided because of having had the RADT, which was congruent with the HCPs' views of the tests providing reassurance:

> *. . . it was quite reassuring being the patient and having the right test for the right medicine . . . .*
>
> *Int. 32, patient*

No change in general practitioner attendance

Patients reported that the provision of the RADT would not influence their decision to attend the GP's surgery for a sore throat:

> *I would still see how it went before I sought medical advice . . . I wouldn't just run straight to the doctors. I would wait, see how it developed*
>
> *Int. 29, patient*

Therefore, patients' views were in contrast to GPs', who described concerns that the routine use of the RADTs would encourage increased patient attendance.

Preference for no antibiotics

Patients reported that they would prefer not to take antibiotics unless needed, and they perceived the RADT as assisting in supporting their no-antibiotic preference, ruling out medication when it is not clinically indicated:

> *If they need to be prescribed and they can help you, then that's fine but if they're not really going to make any difference then I don't see the point.*
>
> *Int. 02, patient*

Discussion

The study identified three key areas surrounding HCPs' views of the use of RADTs: practicalities relating to the tests; the accuracy of diagnosis provided by the RADT; and patient-related outcomes. Participation in the trial was also identified as strongly influencing attitudes towards the use of RADTs. In addition, patients' views of the RADT relating to future GP attendance, antibiotic expectation and reassurance were also identified.

Strengths and limitations

One of the main strengths of the study is its attempt to include both HCPs and patients, to capture a dual view of the RADT in practice. As both groups had used the RADT, the study was able to directly examine the experience of conducting the RADT as a HCP and the experience of receiving the test as a patient. This process was able to reveal differences between HCPs' perceptions of patient attitudes towards the RADT and actual patient views. The sample size was relatively large for a qualitative study; however, as the patient sample size was small, further work could usefully focus on this group to confirm and provide additional insight to the present findings. The NP sample was smaller than the GP sample, but consensus in most views was clear within the NP sample and NPs reported similar views of the RADTs to the GP sample. The only clear difference between the samples was identified in a view that was widely held by GPs but not NPs: that the RADT may be 'unnecessary for experienced practitioners'.

Nesting qualitative interviews into the PRISM trial provided in-depth views, which help interpret the results obtained from the main trial. In terms of limitations, three researchers conducted the interviews and the level of exploration varied. Nonetheless, all interviews explored the key questions/topics and this permitted comparative analysis across the data corpus. Transparency of the analysis process and agreement on themes across the team ensured that final themes were robust.

Main findings

Overall, the interviews suggest that taking part in the PRISM trial had influenced the HCPs' attitudes towards the use of RADTs, in that the tests were now viewed more positively than before participation in the trial. Positive views towards the RADTs related to patient outcomes, such as the test results providing reassurance to the patient that they do not need antibiotics and also providing an educational tool to indicate why antibiotics may not always be necessary. HCPs also felt that the test could provide a useful confirmation of diagnosis, which reassured them and supported their prescribing decisions. In relation to the practicalities of conducting the RADT, HCPs reported that it was easy to use and possibly a useful tool for less experienced staff members and that they felt it had reduced their antibiotic prescribing for sore throat – hence, potentially useful for experienced and less experienced staff.

There were also some negative views and concerns relating to the RADT. HCPs reported concerns regarding the additional costs to the practice and time taken to conduct the RADTs. However, the interviews also suggested that concerns relating to time were most likely to be reduced after the HCP had conducted the tests a few times and had become familiar with the procedure. Concerns were also raised in relation to the accuracy of diagnosis in terms of the possible identification of GABHS carriers who may be unnecessarily prescribed antibiotics, and the possible presence of alternative forms of bacteria that the RADT could not detect. HCPs were concerned with the difficulty in making a prescribing decision when the RADT result differed from their own opinion of diagnosis; however, it was also reported that the RADT could reassure them if they were in doubt over the diagnosis. Finally, there was some concern that the use of RADTs could increase the 'medicalisation' of sore throat and increase patient attendance to the GP for this condition. However, patient interviews suggested that tests would be unlikely to increase their GP attendance, and that patients valued the use of the RADT, reporting that they felt reassured by it.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Comparison with existing literature

It was reported that HCPs viewed the RADTs positively as a tool that could provide a confirmation of their own diagnosis and, therefore, support their prescribing decision. However, HCPs reported a negative view of the RADT if it presented evidence for a diagnosis that differed from their own opinion. Similar findings were also reported in a study that used a computer-based clinical intervention to assist with antibiotic prescribing decisions for respiratory tract infection.[56] GPs reported acceptance of use of the system if it was perceived as a tool that could support their own decisions. However, if the system were perceived as a method of enforcing a prescribing decision, the GPs viewed it negatively and did not wish to use it during consultations. In a similar vein, a study exploring the use of severity measures for the diagnosis of depression described the measures as beneficial when they confirmed a GP diagnosis and treatment disposal, but as of limited value when the score was not congruent with the GP's view founded on clinical experience and intuition.[57]

Practitioners also reported concern that use of the RADT could increase patient pressure for tests and GP consultations for sore throat. However, patients reported that, although they felt reassured by the tests, they would be unlikely to increase their GP attendance for the condition or their expectation for antibiotics. These findings are analogous to perceptions about prescriptions: GPs often perceive a pressure from patients to prescribe, investigate and refer despite the fact that patients do not report such expectations.[58] More recent findings from a qualitative study of patients attending with suspected urinary tract infection also identified patients' reluctance to rely on antibiotics to ameliorate symptoms, whereas GP interviews highlighted belief in patient pressure or desire for antibiotic prescriptions,[44,59] thus suggesting an ongoing disconnection between HCPs' views of patients' expectations and patients' expectations.

Implications for clinical practice

Practitioners reported that being in the PRISM trial had resulted in them viewing RADTs more positively than before their participation. It was also reported that GPs who had not taken part in the trial were likely to view RADTs more negatively than their participating counterparts. This suggests that offering GPs the opportunity to pilot or trial the tests before agreeing to implement them in practice could enhance how the RADTs are subsequently embedded into routine primary care. Practitioners who discussed the clinical scoring method widely held the view that it was time-consuming, unnecessary and beneficial only for inexperienced staff. These perceptions may hinder acceptance of the score by experienced clinicians in the future and would need to be addressed in future initiatives aimed at incorporating these tests into practice.

Patients' views of using the RADTs were generally positive, which suggests that patients would not object to their use in primary care. A larger sample of patients would be useful to further test and confirm the acceptability of the RADTs. HCPs were concerned that the RADT would increase consultations, but patients reported that their likelihood of attending the GP would, in fact, remain the same. This finding could be further explored with patients and, if confirmed, could usefully be presented to GPs to reduce their concerns about using RADTs in relation to the effects on patient attendance.

Conclusion

It is unlikely that RADTs will have a place in clinical practice in the near future, until HCPs' concerns are addressed, and, importantly, until they have direct experience of using them in real consultations. The routine use of clinical scoring systems for acute upper respiratory tract illness also faces important barriers related to clinicians' perception of their utility in the face of their clinical experience and intuition.

Chapter 6 Health economic analysis of the randomised controlled trial

D Turner, R Pinedo-Villanueva, J Raftery, R Hobbs and P Little on behalf of the PRISM investigators.

Abstract

Aims

The aim of this study was to assess resource use and health-related quality of life (HRQoL) associated with clinical scores and RADTs and to show whether these can represent an efficient use of NHS resources.

Methods

A cost-effectiveness study (cost/change in symptom severity) and a cost–utility study [cost/quality-adjusted life-year (QALY)] were carried out as part of the randomised controlled trial. Resource use data were obtained from GP case notes and from study clinicians. QALYs were estimated by means of EQ5D scores obtained from the 14-day diary.

Results

The cost-effectiveness results presented are for FeverPAIN compared with control and RADT. In total, 498 individuals had both symptom severity and cost data from case notes review. Costs for the initial visit and for the 1-month follow-up were similar at £51, £44 and £49 for the delayed prescribing, FeverPAIN and RADT groups, respectively. FeverPAIN dominated both other groups for the cost/change in symptom severity analyses, being both less costly and more effective. Cost-effectiveness acceptability curves (CEACs) indicated the clinical score group to be the most likely to be cost-effective. The cost–utility analysis showed considerable uncertainty around change in QALYs, but FeverPAIN appeared to be the most likely to be cost-effective, particularly for lower values of QALY.

Conclusion

FeverPAIN is a more efficient use of health-care resources than the other approaches.

Introduction

We are aware of no randomised trials of RADTs in which an economic evaluation has been performed. There have been modelling studies in children[60] and adults,[61] but these studies did not collect resource use or quality of life data, and the studies came to opposite conclusions.

Economic evaluations explore whether any particular intervention represents a good use of scarce health-care resources. They do this by measuring both the costs and the benefits generated by any particular intervention(s). This obviously applies to interventions with high resource implications, such as high-cost drugs and complex therapies. However, the same principle also applies to interventions that are much simpler and have low resource implications. The total impact of these low-resource interventions may also be important if they are for common conditions that are responsible for large numbers of consultations in primary care.

The work presented here examines two low-cost interventions. FeverPAIN, a clinical scoring algorithm, and a near-patient test designed to better manage the care of individuals with acute sore throat. Both were expected to have low resource implications, particularly the FeverPAIN algorithm. Those patients offered the RADT had this in addition to the clinical score. Both of these interventions could have implications for the use of other health-care resources and for the health of individuals receiving them.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

The work presented here reports on an economic analysis conducted alongside the clinical trial reported previously in this report. The aim of this economic analysis was to assess the resource and health impacts of both of the above interventions, compared with a control group comprising the strategy of delayed antibiotics, to assess whether either of these interventions would represent an efficient use of health-care resources.

Methods

Data sources for economic evaluation

A number of sources were used to derive data for the economic evaluation. Information regarding the initial (randomisation) contact was entered directly into the website by study practitioners. This included duration, whether a RADT had been performed and prescription of antibiotics. Information on health-care resource use was obtained by a review of practice notes, carried out by a study researcher who visited each participating practice, using a data collection proforma. Resource use data collection was limited to upper respiratory tract-related contacts. Data collected included the following: GP and NP visits; antibiotics; practice visits for complications of infections and antibiotic complications; and hospital admissions related to infections. Data were collected in three time periods: first, the 12 months prior to recruitment (antibiotics; GP and NP visits); second, the 4 weeks following recruitment (GP and NP visits; antibiotics; visits for complications; secondary-care resource use); and, third, resource use in the period from 4 weeks after recruitment up to the end of data collection.

The final important source of health economics data was the 14-day diary completed by study participants. This included the EQ5D[62] completed at baseline and at 14 days. This was used to calculate QALYs. If individuals did not return this diary, then data on symptoms were obtained from a short questionnaire or a telephone interview. However, these sources did not include the EQ5D.

Analysis of costs

Recruitment visit and cost of intervention

As the interventions evaluated in this study could be expected to have implications for the duration of the initial consultation, we measured this duration as part of the study. As noted above, practitioners entered the length of contact into the study database, which was costed using pound per minute of GP time.[63] The costs associated with the clinical score intervention itself were limited to any differences in length of consultation required for this intervention. The costs associated with the clinical score plus near-patient test group comprised the additional time required to provide the intervention as well as the cost of the near-patient test itself, of £3.25 per test, obtained via correspondence with the manufacturer as £65 for 20 tests. However, this test was used only if the clinical score indicated sufficient severity (FeverPAIN ≥ 3), so this cost of £3.25 per test was applied only to the subset of participants in the RADT group who actually had the test.

Case notes review

In addition to the resources needed to directly provide the interventions (practitioner time and cost of test), the study groups may have had differences in use of health-care resources associated with acute sore throat. To examine this we collected resource use data directly from general practice patient notes. Data were obtained for the previous 12 months to test the equivalence of the three groups in terms of baseline resource use for respiratory tract infections. Data were collected for the 12-month follow-up period to test whether there were any long-term differences in costs. However, we found no statistically significant differences between any groups for these two time periods. For this reason, the analysis presented here uses resource use in the recruitment visit as well as in the 4-week follow-up period after this visit. The reviewer identified if the participant had seen a GP or a NP in the relevant time period. They also identified if any symptoms or diagnoses had been reported. These symptoms and/or diagnoses were identified only if a visit had been made. There was a small discrepancy between these two types of data, with some cases

in which symptoms/diagnosis had indicated a visit and no data were available on the type of contact. We assumed that a visit would be recorded if either condition applied, that is a diagnosis or symptom had been reported or a visit to a GP or NP had been noted. Reported contacts with either a GP or a NP were costed using unit costs from the PSSRU (Personal Social Services Research Unit), a widely used source of unit cost data.[63] In those cases in which diagnosis or symptom data indicated that a contact had been made but it was not clear who had been seen, a cost for either a GP or a NP contact was randomly assigned. This was done in the same ratio as that observed for those contacts for which type of contact was specified: 85% of reported contacts in the follow-up period were with a GP. This method was used as data were obtained by a researcher from practice notes. Therefore, we had no a priori reason to suppose that missing data were related to any patient characteristic.

For non-complicated visits to the practice, the notes review indicated whether antibiotics had been prescribed, but no information was obtained on the type of antibiotic used. We had no reason to believe that choice of antibiotic varied by arm. We assumed a cost equal to 28 tablets of amoxicillin.[64] The costs of all drugs prescribed also included a prescribing fee of £1.26 (NHS Drug Tariff: www.ppa.org.uk/edt/December_2012/mindex.htm). In addition, the case notes review also identified visits resulting from complications, either of the illness or from treatment. Complications of the illness included the following: otitis media, quinsy/peritonsillar cellulitis; cellulitis; sinusitis; scarlet fever; and pneumonia. Complications of treatment reported in the case notes review included the following: rash and diarrhoea. In these cases, more information was collected: including who was seen, where they were seen, type of antibiotic and whether a hospital stay had occurred. Again, any community care contacts were costed using PSSRU unit costs.[63] However, secondary-care contacts, such as inpatient stays, visits to accident and emergency (A&E), and visits to hospital doctors were costed using NHS reference costs.[65] The types of resource use included in this study and the unit costs that were used to cost these are described in *Table 14*.

In the event that an individual suffered a complication, a named antibiotic was used and the appropriate cost was used. For a visit to A&E, a cost was obtained from NHS reference costs using a weighted average of all costs for A&E services leading to admission. One individual had a short inpatient stay, which was costed at £473, based on a weighted average of respiratory-related short-stay episodes. All costs used were in 2010/11 UK pounds.

TABLE 14 Types of contacts and unit costs used

Cost item	Value (£)	Source
Visit to GP	31	PSSRU[63]
GP home visit	104	PSSRU[63]
Per minute of GP time	2.67	PSSRU[63]
Visit to NP	13	PSSRU[63]
Antibiotics (amoxicillin)	1.29	BNF[64]
Antibiotics (other)	1.35 to 2.04	BNF[64]
Prescribing cost	1.26	NHS Drug Tariff
Visit to A&E	108	NHS reference costs[65]
Overnight hospital stay	473	NHS reference costs[65]

BNF, *British National Formulary*.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Participants

Two clinical scoring algorithms were used in the clinical study. Score 1 was used at the start of the study and was used with 1129 participants. This was replaced as planned by FeverPAIN in the second part of the study (for 613 participants). The results of the clinical trial indicated no statistically significant results for score 1, but did find statistically significant results for FeverPAIN compared with delayed antibiotics (see *Chapter 4*). We confined our economic analysis to FeverPAIN, as this was the score tested in the clinical trial.

Outcome data used

Outcomes in the economic analysis were the symptom severity score (primary outcome measure in trial) and QALYs based on EQ5D.[62] EQ5D was obtained from the patient-completed diary. Values were obtained at baseline and at 14 days after recruitment. However, as EQ5D scores were obtained from the 14-day diary, we did not have EQ5D values for the end of the 28-day follow-up period. We assumed that health obtained at the end of the 14-day period persisted to the end of the study period, that is the last value obtained was carried forward for 14 days. However, we also calculated QALYs for the 14-day period to test the effect of this assumption on generated results. These were scored using the standard UK tariff (www.euroqol.org). Values obtained were converted into QALYs using 'area under the curve', that is values obtained were multiplied by duration spent in those health states. The estimates of QALY used in the analysis presented here are differences from baseline EQ5D, calculated by estimating 'area under the curve', then subtracting the area that would be generated if baseline EQ5D had been maintained for 28 days. The same method was used for the 14-day QALY estimates, but the baseline QALY for the 14-day period was subtracted. The estimated QALY gain was compared with baseline values of EQ5D.

This 'area under the curve' method assumes a linear relationship between values at different time points; that is, if the EQ5D score changes between baseline and 14 days, a straight line would be drawn between the two points (i.e. HRQoL changes at a uniform rate between the two time points). There is no guarantee that HRQoL would change in this manner, particularly with an acute illness. A number of alternative patterns of recovery would be possible. Two are considered here. First, individuals recover quickly in the first few days, with only minor improvements in health for the rest of the 14-day period. Second, individuals have poor health for the majority of the 14-day period and then recover quickly. Linear interpolation between the two EQ5D scores would understate the QALYs generated in the first case and overstate them in the second. It was not clear at the design stage how individuals' health would change and what the effect of the profile of health change might be. It was not considered feasible to ask participants to complete the EQ5D daily without dramatically complicating and lengthening the short diary, possibly reducing completion rates. We included a page in the patient diary with a number of visual analogue scores (VAS). These were modified slightly in order to fit VASs for 14 days over two pages in the diary. However, the scales and the description of end points were not changed. Individuals were asked to complete these until they felt better and these scores were used to describe the profile of recovery from acute sore throat.

As VASs were completed only until recovery, we lacked scores for the full 14 days for most respondents. Confining analysis to complete data for each day would mean that only cases that took longer to recover would appear in the estimates for the later days. This would bias estimates downwards, and give a misleading picture of, the profile of potential recovery. Therefore, we assumed that the last value given when the individual had deemed themselves to be 'fully recovered' could be carried forward to represent the score they would have given in the remaining time period. These were calculated for individuals for whom we had both baseline and 14-day EQ5D, that is those individuals who we were able to calculate QALY scores for. Examination of the health profiles obtained from this process indicated that change in health did not differ markedly from a straight line recovery (*Figure 9*), although there did appear to be an increased rate of recovery in the first 8 days compared with the next 6. Linear interpolation between EQ5D scores at 0 and 14 days was used in our analysis.

The primary outcome measure was the symptom severity score used in the main clinical study. This was the mean rating of sore throat and difficulty of swallowing for days 2–4. A higher score indicates worse symptoms.

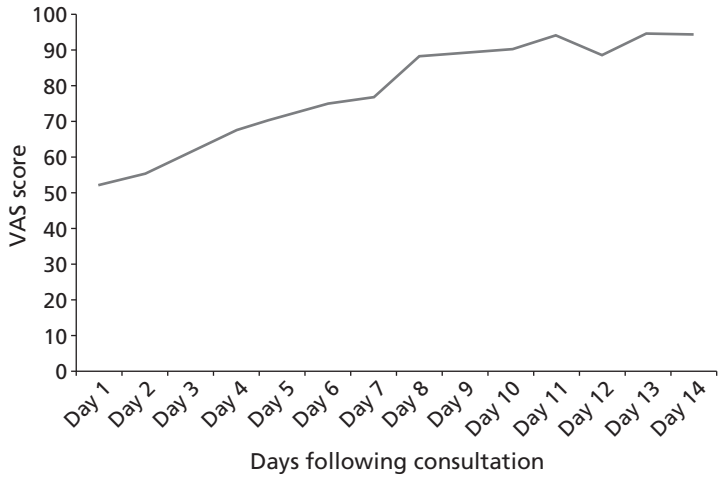


FIGURE 9 Daily VASs for the 257 participants used in the cost–utility analysis.

Analysis

We present results for both a cost-effectiveness study using cost per point change in symptom severity score and a cost–utility analysis based on incremental cost per QALY. The time frame of the analysis is for 1 month after randomisation and the perspective that of the NHS. Costs included those associated with the initial visit as well as those in the 1-month follow-up. Data were analysed in PASW v. 18 (SPSS, Quarry Bay, Hong Kong), Stata v. 11.2 (StataCorp LP, College Station, TX, USA) and Microsoft Excel (Microsoft Corporation, Redmond, WA, USA). Mean and 95% CIs were generated for use cost variables. For outcome variables (both symptom score and QALYs), mean values were estimated using regression equations which allowed us to control for baseline characteristics (fever and baseline symptoms). Mean QALY gain at 14 and 28 days (with 95% CI) for each arm are adjusted by baseline sore throat/swallowing clinical score and fever. To do this, we estimated a linear model for the QALY gain controlling for the clinical score and fever, replicating the method used in the clinical study. We then predicted mean values and 95% CIs. Analyses were performed using the *regress* and *adjust* commands in Stata v. 11.2. Bootstrapping using 5000 samples was used to generate CEACs. These show the probability that each intervention is cost-effective at different values placed upon the outcome variable. The regression methods previously described were used to adjust for baseline characteristics when producing these bootstrapped estimates. These present results that allow for the uncertainty arising from trial data. Bootstrapping was also used to generate scatterplots on the cost-effectiveness plane. These also show the uncertainty inherent in study data.

Results

Individuals were included in the economic analyses if they had data for the relevant outcome measure and cost results from the case notes review. Two outcome measures were used in our analysis. First, we used the clinical symptom score used in the clinical study. This comprised a cost-effectiveness study. Second, we used the EQ5D to derive QALYs; this comprised a cost–utility study. Included in the cost-effectiveness study were 498 individuals for whom we had both cost and clinical symptom scores, with 168, 167 and 163 individuals included in the delayed prescribing, FeverPAIN and RADT groups, respectively. However, as individuals only had an EQ5D score if they returned the diary and also completed the EQ5D question in the diary, we only had EQ5D scores for 274 individuals at baseline (55%) and 262 individuals at 14-day follow-up (53%). There were 257 (52%) patients for whom we had complete EQ5D data for both baseline and follow–up and for whom an estimate of QALY could be calculated. This gave 80, 85 and 92 cases in the three study groups (delayed, FeverPAIN and RADT, respectively). For this reason, the cost–utility study was carried out on a smaller sample than the cost-effectiveness study.

Cost results were obtained for these three groups and are presented in *Table 15*. These results concern the 498 individuals who constitute the cost-effectiveness study. Mean cost was modest in each group at £44

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 15 Resource use and cost by study group

Resource item	Delayed ($n = 168$)			FeverPAIN ($n = 167$)			RADT ($n = 163$)		
	Total number	Mean number	Mean cost (£)	Total number	Mean number	Mean cost (£)	Total number	Mean number	Mean cost (£)
Recruitment visit (minutes)	2418	14.39	38.43	2350	14.07	37.57	2585	15.86	42.34
NPT	–	–	–	–	–	–	41	0.25	0.82
Immediate antibiotics	14	0.08	0.21	24	0.14	0.37	29	0.18	0.45
Delayed antibiotics	136	0.81	2.06	70	0.42	1.07	38	0.23	0.59
Total antibiotics	150	0.89	2.28	94	0.56	1.44	67	0.41	1.05
Total cost of recruitment visit			*40.70*			*39.01*			*44.21*
GP visits in 4-week follow-up	29	0.17	5.35	25	0.15	4.64	24	0.15	4.56
NP visits in follow-up	9	0.05	0.70	2	0.01	0.16	–	–	–
Antibiotics in 4-week follow-up	17	0.10	0.26	11	0.07	0.17	13	0.08	0.20
Total cost of practice visits in 4-week follow-up			*6.31*			*4.96*			*4.77*
Complications									
Otitis media	–	–	–				1	0.01	0.08
Quinsy	2	0.01	4.28						
Cellulitis	–	–	–						
Sinusitis	–	–	–						
Pneumonia	–	–	–						
Complication of treatment				1	0.01	0.19	1	0.01	0.19
Total cost of 4-week follow-up			10.58			5.15			5.05
Total NHS cost (95% CI)			*51.29 (43.34 to 59.23)*			*44.16 (41.32 to 46.99)*			*49.26 (46.03 to 52.49)*

NPT, near-patient test.

to £51 for the 1-month period, the majority of which is attributable to the first (recruitment) visit. GP-reported duration of contact was similar between the delayed and clinical score groups, but was slightly longer in the RADT group at 15.9 minutes. This difference, and the cost of the test, results in a slightly higher cost for the recruitment visit, £44 compared with around £40 in the other two groups. *t*-Tests indicated statistically significant differences between costs for the FeverPAIN group and the RADT group ($p < 0.05$).

Costs in the 4-week follow-up period were low, with only around 15% of individuals having reconsultations. Complications were low with only three complications of illness and two complications of treatment reported. Quinsy was associated with one short hospital stay (one night) and one A&E visit. Both cases occurred in the delayed prescribing group.

The prescribing of antibiotics differed between groups. We did not have data from case notes review or clinician report on whether delayed prescribing prescriptions had been dispensed. Both the clinical score group and the RADT group appeared to be associated with lower use of antibiotics. Costs were calculated for all prescribing cases as if they had been dispensed. This could overstate prescribing costs, but this effect would be small because of the low cost of antibiotics. We use *t*-tests to examine differences between total costs. These showed a statistically significant difference in costs between the delayed prescribing and the clinical score groups.

Cost-effectiveness analysis

Table 16 shows the results based on the symptom score for the 498 individuals included in the cost-effectiveness study. Means for symptom score were adjusted for differences in baseline symptoms and fever, as detailed earlier. In terms of point estimates, the FeverPAIN group dominated the other two groups, being more clinically effective (having a lower symptom score) and less costly. However, for the comparison of FeverPAIN against RADT, it can be seen that the point estimates of symptoms are very similar with overlapping CIs. Therefore, it is important to consider uncertainty around these results. This can be seen in the CEAC (*Figure 10*). The value of a point change in the symptom score was varied between £0 and £500 in this analysis. For this entire range, the clinical score group was the most likely to be cost-effective. This was most marked at lower values of a point change on the symptom score, as the FeverPAIN approach was associated with lower costs. Although estimated symptoms were similar between FeverPAIN and RADT, FeverPAIN had a higher probability of being the most cost-effective option at each value of a point change in symptoms because it had lower costs.

Scatterplots for the cost-effectiveness analysis are given in *Figures 11–13*. In these scatterplots, a negative value of incremental effectiveness indicates a better outcome, as lower symptom scores are preferable. *Figure 11* shows that the FeverPAIN group have lower symptom scores than the delayed group for most simulations and, for the vast majority of simulations, the costs of FeverPAIN are lower than those of delayed prescribing. *Figure 12* shows that the RADT group appears to generate lower symptom scores than the delayed prescribing group. However, costs were very similar and are distributed fairly evenly over the *x*-axis. Finally, *Figure 13* shows that the RADT appears to show similar effectiveness, but at increased costs, to the clinical score group.

TABLE 16 Cost per point change in symptom score results (means and 95% CIs)

Treatment group	Symptom	Total costs (£)	Incremental analysis
Delayed prescribing	3.15 (2.93 to 3.37)	51.30 (43.30 to 59.20)	Dominated
FeverPAIN	2.83 (2.61 to 3.05)	44.20 (41.30 to 47.00)	
RADT	2.84 (2.62 to 3.07)	49.30 (46.00 to 52.50)	Dominated

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

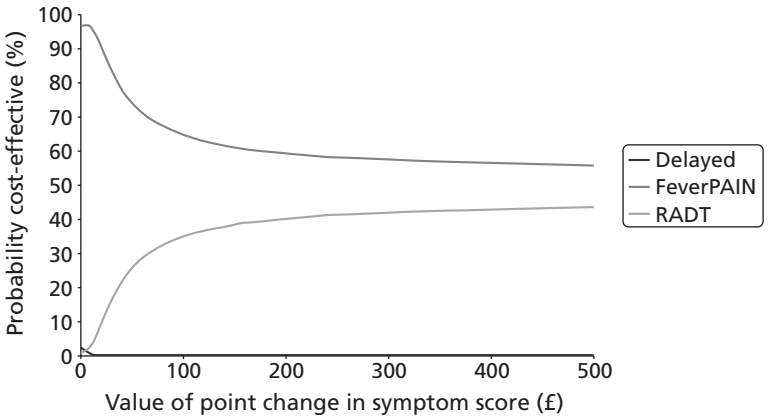


FIGURE 10 Cost-effectiveness acceptability curve for cost-effectiveness study.

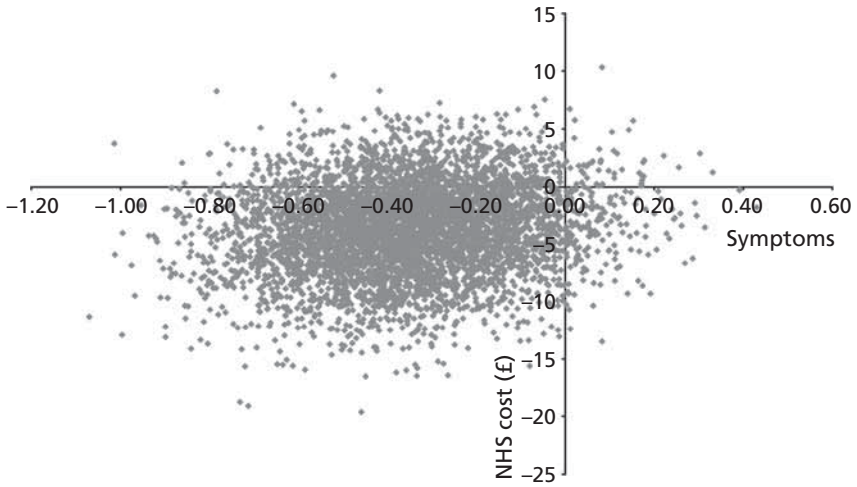


FIGURE 11 Scatterplot showing FeverPAIN vs. delayed prescribing group for cost per point change in symptom score.

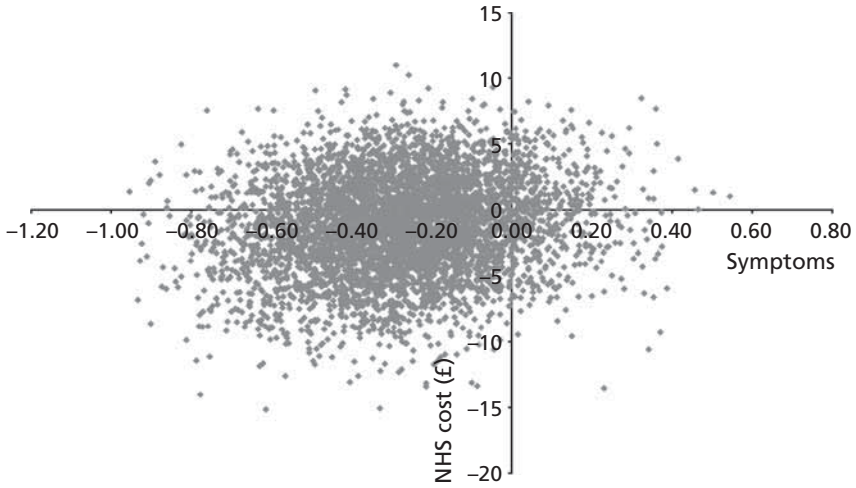


FIGURE 12 Scatterplot showing RADT vs. delayed prescribing group for cost per point change in symptom score.

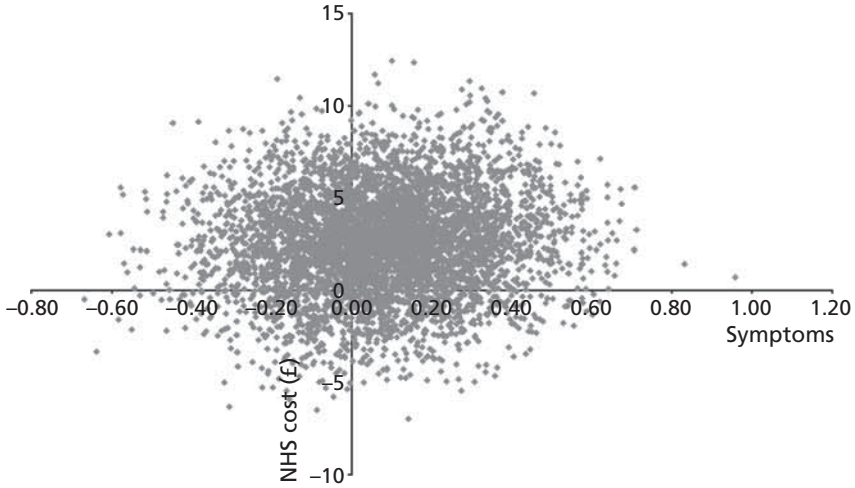


FIGURE 13 Scatterplot showing RADT vs. clinical score group for cost per point change in symptom score.

Cost–utility analysis

Adjustment for baseline for estimated QALY gain

We used regression analysis on the QALY measure used to test any associations between QALY gain and baseline characteristics. This replicated methods used in the clinical study on the clinical outcome measures. However, in this case this regression analysis was carried out on only the 257 individuals in the complete case analysis. As in the clinical study, we found statistically significant associations between QALY gain and baseline measures of fever and severity score.

Bootstrapping to produce CEACs also included adjustment by baseline symptom severity and fever.

The results for the complete case analysis for the outcome measures used in the cost–utility analysis can be seen in *Table 17*. EQ5D scores for both baseline and 14 days were obtained from the patient-completed diaries. EQ5D scores were very similar in all three groups. It can also be seen that there is a marked increase in EQ5D score in all groups between baseline and the 14-day follow-up. This indicates that the individuals' sore throats were having a marked effect on their self-reported EQ5D score. The estimated 14- and 28-day QALY gains compared with baseline are also shown in this table. We show results adjusted for baseline characteristics. As stated earlier, a linear relationship is assumed between the baseline and 14-day EQ5D scores to generate QALY gain at 14 days. The QALY gain is, therefore, the area of a triangle formed by the difference between the 14-day EQ5D and the baseline EQ5D multiplied by 14 days. As we assumed that the 14-day EQ5D score is carried forward to estimate QALYs gained for 28 days, the QALY gain generated in the 14- to 28-day period is represented by a rectangle formed by the difference between the EQ5D scores multiplied by this 14-day period. Therefore, the QALY gain in the second 14-day period is twice that of the first. The QALY gain estimated in the total recruitment to 28-day period is, therefore, three times that estimated in the 14-day period.

TABLE 17 Cost–utility outcomes data (95% CI)

Measure	Control (*n* = 80)	FeverPAIN (*n* = 85)	RADT (*n* = 92)
EQ5D – Baseline	0.63 (0.57 to 0.69)	0.61 (0.54 to 0.68)	0.65 (0.59 to 0.72)
EQ5D – 14 Days	0.92 (0.89 to 0.95)	0.94 (0.92 to 0.97)	0.94 (0.92 to 0.96)
14-day QALY	0.0057 (0.0044 to 0.0070)	0.0058 (0.0045 to 0.0071)	0.00584 (0.0046 to 0.0071)
28-day QALY	0.0171 (0.0131 to 0.021)	0.0174 (0.0135 to 0.0213)	0.0175 (0.0138 to 0.0212)

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Results of cost–utility study

The results presented are for the 257 individuals included in the cost–utility complete case analysis. *Table 18* shows the results of the cost per QALY analysis for the 14-day results. The QALY difference from baseline ranged from 0.0057 to 0.0058. There were no clear differences in generated QALYs among the three groups, as can be seen from the wide and overlapping CIs. The costs per patient were very similar to those generated for the larger data set of 498. However, as there was considerable uncertainty in these data, no clear conclusion should be drawn from this. The 28-day results are presented in *Table 19* and these are similar (although larger in magnitude) with no clear differences among groups in QALY scores. As the QALY gain is marginally higher in the RADT group than in the FeverPAIN group, RADT generates additional QALYs at between £74,286 to £24,528 per QALY, depending on which QALY measure is used.

To take account of uncertainty, CEACs were generated for these results (*Figures 14* and *15*). These both show considerable uncertainty. This is largely generated by uncertainty around the QALY estimate. At a value of £30,000 per QALY, the probabilities that the three groups are cost-effective are 25%, 40% and 35%, for the delayed prescribing, FeverPAIN and RADT groups, respectively, for the 14-day QALY gain. For the 28-day QALY gain, the same values are 28%, 38% and 35%. As can be seen from the scatterplots (*Figures 16–18*), the distribution of estimated incremental QALYs is fairly evenly distributed around the

TABLE 18 Cost–utility results for FeverPAIN using adjusted QALY data for 14-day period (means and 95% CIs)

Treatment group	QALY	Total costs (£)	Incremental effect	Incremental costs (£)	ICER (£)
Delayed prescribing	0.0057 (0.0044 to 0.007)	49.70 (43.30 to 56.00)	Dominated		
FeverPAIN	0.0058 (0.0045 to 0.0071)	45.90 (41.50 to 50.20)			
RADT	0.00584 (0.0046 to 0.0071)	48.50 (45.00 to 52.00)	0.000035	2.60	74,286

TABLE 19 Cost-utility results for FeverPAIN using adjusted QALY data for 28-day period (means and 95% CIs)

Treatment group	QALY	Total costs (£)	Incremental effect	Incremental costs (£)	ICER (£)
Delayed prescribing	0.0171 (0.0131 to 0.0211)	49.70 (43.30 to 56.00)	Dominated		
FeverPAIN	0.01741 (0.0135 to 0.0213)	45.90 (41.50 to 50.20)			
RADT	0.01752 (0.0138 to 0.0212)	48.50 (45.00 to 52.00)	0.000106	2.60	24,528

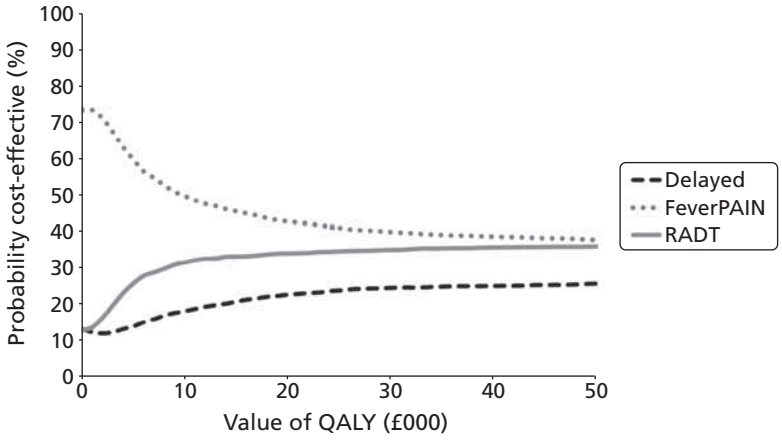


FIGURE 14 Cost-effectiveness acceptability curve for 14-day QALY difference.

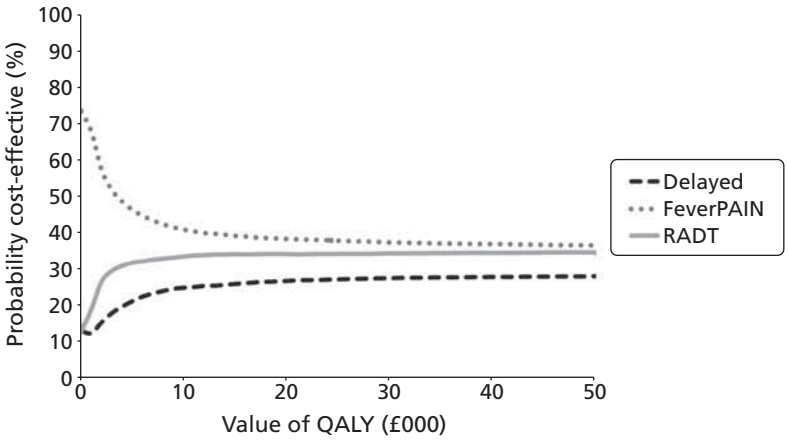


FIGURE 15 Cost-effectiveness acceptability curve for 28-day QALY difference.

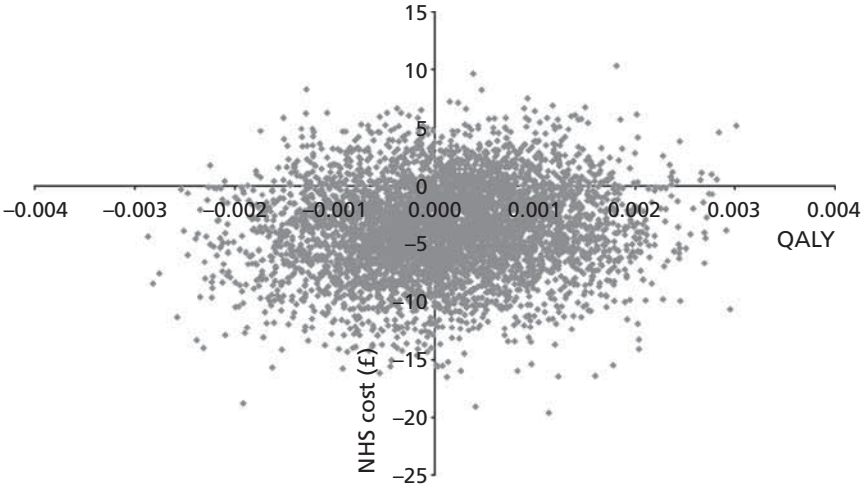


FIGURE 16 Scatterplot showing clinical score group vs. delayed prescribing group for cost/QALY analysis for 14-day QALY difference.

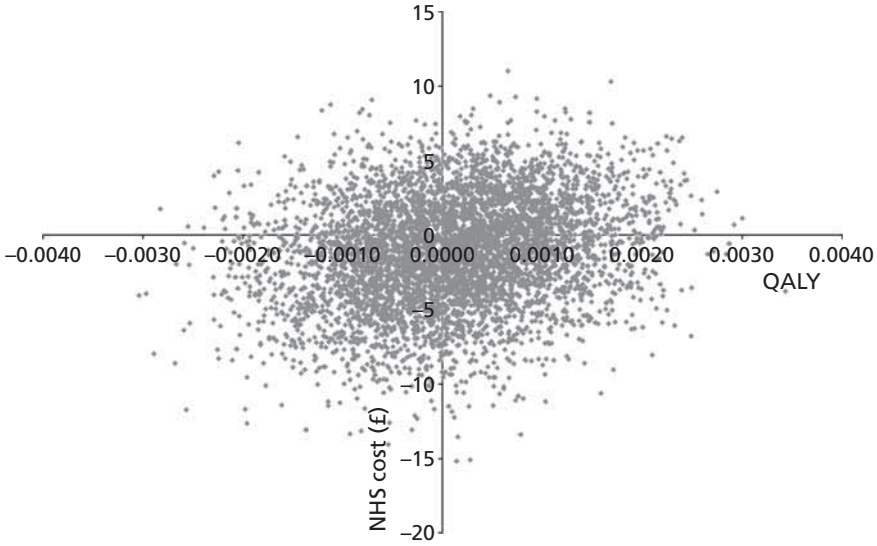


FIGURE 17 Scatterplot showing RADT group vs. delayed prescribing group for cost/QALY analysis for 14-day QALY difference.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

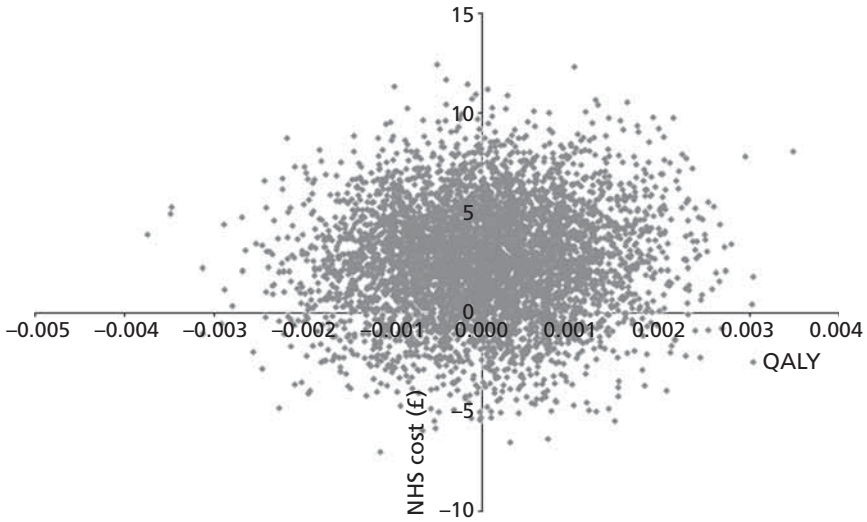


FIGURE 18 Scatterplot showing RADT group vs. clinical score group for cost/QALY analysis.

y-axis. These scatterplots are for the 14-day QALY analysis. The 28-day analysis plots are very similar, apart from the magnitude of the QALY difference, so are not shown here.

Discussion

Statement of main findings

The cost-effectiveness analysis showed that the clinical scoring algorithm used in the FeverPAIN group was effective in reducing symptoms. Costs in all three groups were similar and modest, but the FeverPAIN score group was associated with lower costs. This indicates that the use of the FeverPAIN score may be an efficient use of health-care resources. However, for the cost–utility analysis (cost/QALY) there was a less clear message, as differences in QALYs generated were very small with wide CIs, and therefore there was no statistically significant differences between any groups. The CEACs for the cost–utility study indicate that FeverPAIN is the most likely to be cost-effective over all values of a QALY examined; however, they also indicate considerable uncertainty. The cost/QALY offers weak support for the conclusions of the cost-effectiveness analysis.

Strengths and potential limitations

A strength of the current work is that it used a clinical measure of symptoms as well as the QALY. The former was plausibly more sensitive to changes in participants' symptoms of sore throat. Its administration ensured it picked up changes day to day. Another strength concerns the costs obtained from a case note review being comprehensive. A weakness of the current study is that the 14-day diary had EQ5D data only at two time points (0 and 14 days) and for 52% of all participants. We did not find any statistically significant differences in terms of QALY differences for participants. There could be a number of reasons for this. First, the smaller data set for which we had QALY score may not be representative of the larger group of individuals in the cost-effectiveness study. Second, it may be that many individuals would have recovered before the 14-day QALY was asked and thus their health may have returned to normal. For this reason, the 14-day EQ5D score and hence the QALY difference at 14 days may not be well correlated with changes in symptom scores. It may also be the case that the EQ5D is not sensitive in assessing HRQoL related to change in sore throat-related symptoms. The true impact of RADTs in practice may not be fully captured in our estimation of costs: we estimated an additional 4.5 minutes in consultation time on average when a test was performed, which may be an underestimate of the time taken to do a RADT – the pack and reagents have to be prepared, a throat swab taken, the antigen extracted from the swab, reagents mixed, the resulting mixture applied to the test pack, and then for a negative result there is a 5- to 10-minute wait (see *Appendix 7*). The reason for this possible underestimate is unclear, but clinicians may have been able to be more rapid with trial paperwork when they knew they had to do a

RADT. The time taken to feed back the results may also not have been captured. Furthermore, the 'costs' of the stress of a longer consultation, and the knock-on effect for other consultations, are not included in costs. Thus, it is likely that RADTs are, if anything, less cost-effective than we have estimated.

The estimates of both clinical effectiveness and cost-effectiveness may be modified by alternative algorithms for using RADTs (e.g. using a RADT at lower FeverPAIN scores). However, a more widespread use of RADTs would probably not be acceptable to clinicians in the UK based on our qualitative work, and the targeting of RADTs in the current study was based on modelling to determine the most efficient use of the tests in identifying patients with streptococci, so more widespread use of RADTs is likely to be a less efficient policy.

Relation to existing literature

Webb[60] used decision analysis for a paediatric population and compared immediate treatment, culture, RADT and RADT and culture and concluded that empirical treating was most cost-effective. Neuner *et al.*[61] modelled five strategies (no treatment; immediate treatment; culture and delayed antibiotics until results; RADT with culture for negative results; and RADT alone) and, in contrast to Webb, found empirical treatment to be the least cost-effective, but all others to have a similar QALY gain, and culture to be the least expensive so the most cost-effective. Neither of these studies collected prospective data from a trial and neither performed direct measurement of quality of life or the change in quality of life over time.

Conclusion

The FeverPAIN algorithm enabled an efficient use of health-care resources compared with the other two groups based on changes in symptoms, the primary outcome. As it appears to be more clinically effective and less costly than delayed prescribing and less costly than RADT, it would appear reasonable to prefer it to both alternatives on economic grounds. The cost per QALY analysis gave a less clear message, but did not contradict the cost-effectiveness analysis.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Acknowledgements

We are grateful to all the patients and HCPs who have contributed their time and effort and helpful insights to make PRISM possible.

Contributions of authors

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 co-ordination of centres, contributed to the analysis of *Chapters 2–6*, led the drafting of the diagnostic and trial chapters, contributed to the drafting of other chapters and co-ordinated drafting of the report.

FD Richard Hobbs (GP and Professor of Primary Care, University of Birmingham) developed the protocol for funding, contributed to the management of all studies, supervised the Birmingham study centre and contributed to the drafting of all chapters.

Michael Moore (GP and Reader in Primary Care, University of Southampton) developed the protocol for funding, contributed to the management of the clinical studies, and contributed to the analysis of the diagnostic studies and to the drafting of *Chapters 3–6*.

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 of the diagnostic studies and trial and to the drafting of all chapters except *Chapters 5* and *6*.

Ian Williamson (GP and Senior Lecturer in Primary Care, University of Southampton) developed the protocol for funding and contributed to the management of the clinical studies and to drafting of *Chapters 3–6*.

Cliodna McNulty (Consultant Microbiologist for the Public Health England) developed the protocol for funding, led the design and supervised the running of the in vitro study and microbiological aspects of the studies, and contributed to the management and write up of the diagnostic studies and the trial.

Gemma Lasseter (Biomedical Laboratory Scientist, Public Health England) developed the protocol, performed the in vitro study and lead the drafting of *Chapter 1*.

Man Ying Edith Cheng (Study Statistician, University of Southampton) developed the protocol and contributed to the quantitative analysis of the diagnostic studies and trial and to the drafting of *Chapters 2–4*.

Gerry Leydon (Social Scientist and Principal Research Fellow, University of Southampton) developed the protocol for funding, contributed to the management of the studies, and led the design, supervision, analysis and write up of the qualitative study, and commented on drafts of other chapters.

Lisa McDermott (Social Scientist and Research Assistant, University of Southampton) developed the protocol for funding and contributed to the design, supervision, analysis and write up of the qualitative study.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

David Turner, **Rafael Pinedo-Villanueva** and **James Raftery** (Health Economics, University of Southampton) developed the protocol for funding and led the health economic analysis as well as the drafting of *Chapter 6*.

Paul Glasziou (GP and Professor of Primary Care, University of Oxford) developed the protocol for funding, contributed to the management of the clinical studies, commented on all chapters, and contributed to analysis and drafting of *Chapters 2–4*.

Mark Mullee (Lead Study Statistician, Director of Research Design Service, University of Southampton) developed the protocol for funding, contributed to study management, supervised data management, led the quantitative analysis of the diagnostic studies and trial, and contributed to the drafting of *Chapters 2–4*.

Contributions of other PRISM investigators

Razia Meer-Baloch (Senior Trial Manager) co-ordinates the day-to-day running of the Birmingham study centre and commented on draft papers.

Jo Kelly and **Jane Barnett** (Senior Trial Managers, University of Southampton) developed the protocol, provided day-to-day overall management of the studies, co-ordinated recruitment in the lead study centre and co-ordinated other centres, and commented on drafts of all chapters.

Tessa Lambton, **Rebecca Cooper** and **Hugo Henderson** (all medical students, University of Southampton) contributed to the design, data collection, analysis and write up of the qualitative study.

Peter Hawtin (Consultant Microbiologist for Public Health England) developed the protocol for funding and contributed to the design and running of the in vitro study and the processing of samples for the clinical studies.

Karen Middleton (Data Manager, University of Southampton) provided administrative support for all studies, developed data management protocols, co-ordinated data entry and commented on drafts of the chapters.

Sue Smith, **Mary Selwood** and **Diane Coulson** (Trial Managers, University of Oxford) provided day-to-day co-ordination of the Oxford study centre and commented on drafts of the papers.

Publication

Little P, Hobbs FDR, Moore M, Mant D, Williamson I, McNulty C, *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. doi:10.1136/bmj.f5806.

References

1. Del Mar C. Managing sore throat: a literature review I: Making the diagnosis. *Med J Austr* 1992;**156**:572–5.
2. Petersen I, Johnson A, Islam A, Duckworth G, Livermore D, Hayward A. 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. http://dx.doi.org/10.1136/bmj.39345.405243.BE
3. Dean L, Perry K. *Group A Streptococcus Rapid Antigen Detection Kits – A Review of the Evaluation Literature*. MHRA report no. 04123 ed. London: DOH; 2005.
4. Charlier-Bret N, Boucher B, Poyart C, Quesne G, Bingen E, Doit C *et al.* Rapid antigen detection tests for diagnosis of group A stretococcal pharyngits: comparative evaluation of sensitivity and practicabiltiy of 16 in vitro diagnostic medical devices performed in July 2002 by the French health products safety agency (Afssaps) as part of its market control mission. *Pathobiologie* 2004;**52**:438–43.
5. Pelucchi C, Grigoryan L, Galeone C, Esposito S, Huovinen P, Little P, *et al.* Guideline for the management of acute sore throat. *Clin Microbiol Infect* 2012;**18**(Suppl. 1):1–28. http://dx.doi.org/10.1111/j.1469-0691.2012.03766.x
6. Brodsky L, Nagy M, Volk M, Stanievich J, Moore L. The relationship of tonsil bacterial concentration to surface and core cultures in chronic tonsillar disease in children. *Int J Ped Otorhinolaryngol* 1991;**21**:33–9. http://dx.doi.org/10.1016/0165-5876(91)90057-I
7. Brodsky L, Moore L, Stanievich J, Ogra P. The immunology of tonsils in children: the effect of bacterial load on the presence of B- and T-cell subsets. *Laryngoscope* 1988;**98**:93–8.
8. Kuhn J, Brook I, Church L, Bianchi D, Thompson D. Quantitative bacteriology of tonsils removed from children with tonsillitis hypertrophy and recurrent tonsillitis with and without hypertrophy. *Ann Otol Rhinol Laryngol* 1995;**104**:646–52.
9. Ozawa A, Sawamura S, Ozawa A, Sawamura S. Microbial ecology and tonsillar infection. *Acta Otolaryngol Suppl* 1988;**454**:178–84. http://dx.doi.org/10.3109/00016488809125023
10. Gerber MA. Culturing throat swabs. The end of an era? *J Pediatr* 1985;**107**:85–8. http://dx.doi.org/10.1016/S0022-3476(85)80620-X
11. Gerber M. Diagnosis of group A beta-hemolytic streptococcal pharyngitis: use of antigen detection tests. *Diagn Microbiol Infect Dis* 1986;**4**:5–15.
12. Lindbaek M, Hoiby E, Steinsholt I, Hjortdahl P. Clinical symptoms and signs in sore throat patients with large colony variant β-haemolytic streptococci groups C or G versus group A. *BJGP* 2005;**55**:615–19.
13. Tiemstra J, Miranda R. Role of non-group A streptococci in acute pharyngitis. *J Am Board Fam Med* 2009;**22**:663–9. http://dx.doi.org/10.3122/jabfm.2009.06.090035
14. Bourbeau P. Just a swab you say? Balderdash! *Clin Microbiol Newslett* 2005;**27**:19–23. http://dx.doi.org/10.1016/j.clinmicnews.2005.01.003
15. Spinks A, Glaziou PP, Del Mar CB. Antibiotics for sore throat. *Cochrane Database Syst Rev* 2006; Issue 4: Art No. CD000023.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

16. Cooper R, Hoffman J, Bartlett J, Besser R, Gonzales R, Hickner J. Principles of appropriate antibiotic use for acute pharyngitis in adults. *Ann Intern Med* 2001;**134**:509–17. http://dx.doi.org/10.7326/0003-4819-134-6-200103200-00019

17. NICE Guideline Development Group. Prescribing of antibiotics for self-limiting respiratory tract infections in adults and children in primary care. URL: www.nice.org.uk/Guidance/CG69 (accessed July 2008).

18. Llor C, Madurell J, Balague-Corbella M, Gomez M, Cots J. Impact on antibiotic prescription of rapid antigen detection testing in acute pharyngitis in adults: a randomised clinical trial. *BJGP* 2011;**61**:e244–51. http://dx.doi.org/10.3399/bjgp11X572436

19. Worrall G, Hutchinson J, Sherman G, Griffiths J. Diagnosing streptococcal sore throat in adults a randomized controlled trial of in-office aids. *Can Fam Physician* 2007;**53**:666–71.

20. Efstratiou A. Pyogenic streptococci of Lancefield groups C and G as pathogens in man. *J Appl Micorbiol Symp Ser* 1997;**83**:72–9S. http://dx.doi.org/10.1046/j.1365-2672.83.s1.8.x

21. Health Protection Agency. Health protection report: streptococcal bacteraemias. *HPA Weekly Report* 2011;**4**(46).

22. Meier C, Centor RM, Graham L, Dalton HP. Clinical and microbiological evidence for endemic pharyngitis among adults due to group C streptococci. *Arch Intern Med* 1990;**150**:825–9. http://dx.doi.org/10.1001/archinte.150.4.825

23. Little PS, Gould C, Williamson I, Warner G, Gantley M, Kinmonth 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. http://dx.doi.org/10.1136/bmj.315.7104.350

24. Centor RM, Witherspoon JM, Dalton HP. The diagnosis of strep throat in the emergency room. *Med Decis Making* 1981;**1**:239–46. http://dx.doi.org/10.1177/0272989X8100100304

25. Dobbs F. A scoring system for predicting group A streptococcal throat infection. *BJGP* 1996;**46**:461–4.

26. Breese B. A simple scorecard for the tentative diagnosis of streptococcal pharyngitis. *Am J Dis Child* 1977;**131**:514–17. http://dx.doi.org/10.1001/archpedi.1977.02120180028003

27. Dunn N, Lane D, Everitt H, Little P. Use of antibiotics for sore throat and incidence of quinsy. *BJGP* 2007;**57**:45–9.

28. Rogers M. A viable alternative to the glass/mercury thermometer. *Paed Nurs* 1992;**4**:8–11.

29. National Standard Method. Investigation of throat swabs. *BSOP* 2008;**9**. URL: www.hemltd.ru/publications/sections/Normativ/foreign/samples/medicine/NHS010/article.pdf (accessed 31 December 2009).

30. National Standard Methods. Identification of *Streptococcus* species, *Entrococcus* species and morphologically similar organisms. *Health Protection Agency*; 2007. URL: www.hpa-standardmethods.org.uk/documents/bsopid/pdf/bsopid4.pdf (accessed 17 September 2007).

31. British Society for Antimicrobial Chemotherapy. Methods for antimicrobial susceptibility testing. British Society for Antimicrobial Chemotherapy; 2009. URL: www.bsac.org.uk/_db/_documents/Version_8_-_January_2009.pdf (accessed January 2009).

32. Her Majesty's Stationery Office (HMSO), Office of Population Censuses and Surveys (OPCS). *Morbidity Statistics from General Practice: Fourth National Study 1991–1992*. 1st edn. London: HMSO; 1994.

33. Little PS, Williamson I, Warner G, Gould C, Gantley M, Kinmonth AL. An open randomised trial of prescribing strategies for sore throat. *BMJ* 1997;**314**:722–7. http://dx.doi.org/10.1136/bmj.314.7082.722

34. McIsaac W, Kellner J, Aufricht P, Vanjaka A, Low D. Empirical validation of guidelines for the management of pharyngitis in children and adults. *JAMA* 2004;**291**:1587–95. http://dx.doi.org/10.1001/jama.291.13.1587

35. McIsaac W, Goel V, To T, Low D. The validity of a sore throat score in family practice. *CMAJ* 2000;**163**:811–15.

36. Dagnelie C, Bartelink M, Van Der Graaf Y, Goessens W, De Melker R. Towards better diagnosis of throat infections with GABHS in general practice. *BJGP* 1998;**48**:959–62.

37. Zwart S, Sachs A, Ruijs G, Hoes A, DeMelker R. Penicillin for acute sore throat: randomised double blind trial of seven days versus three days treatment or placebo in adults. *BMJ* 2000;**320**:150–4. http://dx.doi.org/10.1136/bmj.320.7228.150

38. Snow V, Mottur-Pilson C, Cooper R, Hoffman J. Principles of appropriate antibiotic use for acute pharyngitis in adults. *Ann Intern Med* 2001;**134**:506–8. http://dx.doi.org/10.7326/0003-4819-134-6-200103200-00018

39. Goossens H, Ferech M, Vander Stichele R, Elseviers M, ESAC project group. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet* 2005;**365**:579–87. http://dx.doi.org/10.1016/S0140-6736(05)17907-0

40. Standing Medical Advisory Committee (SMAC). *The path of least resistance*. Department of Health; 1998.

41. American College of Physicians. *Principles of appropriate antibiotic use for acute pharyngitis in adults*. American College of Physicians; 2001. URL: www.acponline.org/clinical_information/guidelines (accessed March 2001).

42. Vincent M, Celestin N, Hussain A. Pharyngitis. *Am Fam Phys* 2004;**69**:1465–70. http://dx.doi.org/10.1097/00006534-196401000-00031

43. Harrel FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York, NY: Springer Verlag; 2001.

44. Little P, Turner S, Rumsby K, Warner G, Moore M, Lowes JA, *et al.* Urinary tract infection: development and validation, randomised trial, economic analysis, observational cohort and qualitative study. *Health Technol Assess* 2009;**13**(19).

45. Schwarz B, Marcy M, Phillips W, Gerber M, Dowell S. Pharyngitis – principles of judicious use of antimicrobial agents. *Pediatrics* 1998;**101**:171–4.

46. Little P. Delayed prescribing of antibiotics for upper respiratory tract infection. *BMJ* 2005;**331**:301–2. http://dx.doi.org/10.1136/bmj.331.7512.301

47. Little P, Rumsby K, Kelly J, Watson L, Moore M, Warner G, *et al.* Information leaflet and antibiotic prescribing strategies for acute lower respiratory tract infection: a randomised controlled trial. *JAMA* 2005;**293**:3029–35.

48. Sharland M, Kendall H, Yeates D, Randall A, Hughes G, Glasziou P, *et al.* Antibiotic prescribing in general practice and hospital admissions for peritonsillar abscess, mastoiditis, and rheumatic fever in children: time trend analysis. *BMJ* 2005;**331**:328–9. http://dx.doi.org/10.1136/bmj.38503.706887.AE1

49. Little P, Gould C, Williamson I, Moore M, Warner G, Dunleavey J. A pragmatic randomised controlled trial of two prescribing strategies for acute otitis media. *BMJ* 2001;**322**:336–42. http://dx.doi.org/10.1136/bmj.322.7282.336

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

50. Zhang J, Yu K. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA* 1998;**280**:1690–1. http://dx.doi.org/10.1001/jama.280.19.1690

51. Little P, Gould C, Williamson I, Warner G, Gantley M, Kinmonth A. Clinical and psychosocial predictors of illness duration from a randomised controlled trial of prescribing strategies for sore throat. *BMJ* 1999;**319**:736–7. http://dx.doi.org/10.1136/bmj.319.7212.736

52. Sheeler R, Little P. Rapid streptococcal testing for sore throat and antibiotic resistance. *Clin Microbiol Infect* 2006;**12**(Suppl. 9):3–7. http://dx.doi.org/10.1111/j.1469-0691.2006.01656.x

53. Wood F, Brookes-Howell L, Hook K, Cooper L, Verheij T, Goossens H, *et al.* A multi-country qualitative study of clinicians' and patients' views on point of care tests for lower respiratory tract infection. *Fam Pract* 2011;**28**:661–9. http://dx.doi.org/10.1093/fampra/cmr031

54. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2012;**3**:77–101.

55. Leydon GM, Dowrick CF, McBride A, Burgess H, Howe AC, Moore M, *et al.* Questionnaire severity measures for depression: a threat to the doctor-patient relationship? *BJGP* 2011;**61**(583):117–23. http://dx.doi.org/10.3399/bjgp11X556236

56. McDermott L, Yardley L, Little P, Ashworth M, Gulliford M. eCRT research team: developing a computer delivered, theory based intervention for guideline implementation in general practice. *BMC Fam Pract* 2010;**11**:90.

57. Dowrick C, Leydon G, McBride A, Howe A, Burgess H, Clarke P, *et al.* Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ* 2009;**338**:b663. http://dx.doi.org/10.1136/bmj.b663

58. Little P, Dorward M, Warner G, Moore M, Stephens K, Senior J. Importance of patient pressure and perceived pressure and perceived medical need for investigations, referral, and prescribing in primary care: nested observational study. *BMJ* 2004;**328**:444. http://dx.doi.org/10.1136/bmj.38013.644086.7C

59. Leydon G, Turner S, Smith H, Little P. The journey from self-care to GP care: a qualitative interview study of women presenting with symptoms of urinary tract infection. *BJGP* 2009;**59**:219–25. http://dx.doi.org/10.3399/bjgp09X453459

60. Webb K. Does culture confirmation of high-sensitivity rapid streptococcal tests make sense? A medical decision analysis. *Pediatrics* 1998;**101**:E2. http://dx.doi.org/10.1542/peds.101.2.e2

61. Neuner J, Hamel M, Phillips R, Bona K, Aronson M. Diagnosis and management of adults with pharyngitis. A cost-effectiveness analysis. *Ann Intern Med* 2003;**139**:113–22. http://dx.doi.org/10.7326/0003-4819-139-2-200307150-00011

62. The Euroqol Group. EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208.

63. Curtis L. *The Unit Costs of Healthcare*. Canterbury, UK: Personal Social Services Research Unit; 2011.

64. Joint Formulary Committee. *British National Formulary*. BMA and Royal Pharmaceutical Society; 2010.

65. NHS. *2010–2011 reference costs publication*. London: Department of Health; 2012.

66. Little P, Watson L, Morgan S, Williamson I. Antibiotic prescribing and admissions with major suppurative complications of respiratory tract infections: a data linkage study. *BJGP* 2002;**52**:187–93.

67. Ashworth M, Latinovic J, Charlton K, Cox K, Rowlands G, Gulliford M. Why has antibiotic prescribing for respiratory illness declined in primary care? A longitudinal study using the General Practice Research Database. *J Public Health* 2004;**26**:268–74. http://dx.doi.org/10.1093/pubmed/fdh160

68. Kolmos H, Little P. Should general practitioners perform diagnostic tests on patients before prescribing antibiotics? *BMJ* 1999;**318**:799–802. http://dx.doi.org/10.1136/bmj.318.7186.799

69. Poses R, Cebul R, Collins M, Fager S. The importance of disease prevalence in transporting clinical prediction rules: the case of streptococcal pharyngitis. *Ann Intern Med* 1986;**105**:586–91. http://dx.doi.org/10.7326/0003-4819-105-4-586

70. Hobbs R, Delaney B, Fitzmaurice D. A systematic review of near patient testing in primary care. *Health Technol Assess* 1997;**1**(5).

71. Schwartz R. Evaluation of rapid streptococcal detection tests. *Ped Inf Dis J* 1997;**16**:1099–100. http://dx.doi.org/10.1097/00006454-199711000-00028

72. Little P, Barnett J, Kinmonth AL, Margetts B, Gabbay J, Thompson R, *et al.* Can dietary assessment in general practice target patients with unhealthy diets. *BJGP* 2000;**50**:43–5.

73. Lewey S, White C, Lieberman MM, Morales E. Evaluation of the throat culture as a follow-up for an initially negative enzyme immunosorbent assay rapid streptococcal antigen detection test. *Ped Inf Dis J* 1988;**7**:765–9. http://dx.doi.org/10.1097/00006454-198811000-00004

74. Battle C, Glasgow L. Reliability of bacteriologic identification of beta-hemolytic streptococci in private offices. *Am J Dis Child* 1971;**122**:134–6.

75. Breese B, Disney F. The accuracy of diagnosis of beta streptococcal infections on clinical grounds. *J Pediatr* 1954;**44**:670–3. http://dx.doi.org/10.1016/S0022-3476(54)80008-4

76. Bodino J, Lopez E, Rubeglio E, de Giavedoni C. Evaluation of a rapid test for group A *Streptococcus* at a physician's office and hospital laboratory in Buenos Aires, Argentina. *Pediatr Infect Dis J* 1987;**6**:762–4.

77. Gieseker K, Roe M, MacKenzie T, Todd J. Evaluating the American Academy of Pediatrics diagnostic standard for *Streptococcus pyogenes* pharyngitis: backup culture versus repeat rapid antigen testing (and response to this article). *Pediatrics* 2003;**111**:e666–70.

78. Anhalt J, Heiter B, Naumovitz D, Bourbeau P. Comparison of three methods for detection of group A streptococci in throat swabs. *J Clin Microbiol* 1992;**30**:2135–8.

79. Nerbrand C, Jasir A, Schalen C. Are current rapid detection tests for group A streptococci sensitive enough? Evaluation of 2 commercial kits. *Scand J Infect Dis* 2002;**34**:797–9. http://dx.doi.org/10.1080/0036554021000026953

80. Glaser B, Strauss A. The discovery of grounded theory: strategies for qualitative research. New York, NY: Aldine; 1967.

81. Little P, Dorward M, Gralton S, Hammerton L, Pillinger J, White P, *et al.* A randomised controlled trial of three pragmatic approaches to initiate increased physical activity in sedentary patients with risk factors for cardiovascular disease. *BJGP* 2004;**54**:189–95.

82. Yardley L, Beech S, Zander L, Evans T, Weinman J. A randomized controlled trial of exercise therapy for dizziness and vertigo in primary care. *BJGP* 1998;**48**:1434–5.

83. Yardley L, Donovan-Hall M, Smith H, Walsh B, Mullee M, Bronstein A. Effectiveness of primary care vestibular rehabilitation for vestibular dizziness. *Ann Int Med* 2004;**141**:598–605. http://dx.doi.org/10.7326/0003-4819-141-8-200410190-00007

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

84. Little P, Kelly J, Barnett J, Dorward M, Warm D, Margetts B. Randomised controlled factorial trial of dietary advice for patients with a single high blood pressure reading in primary care. *BMJ* 2004;**328**:1054–60. http://dx.doi.org/10.1136/bmj.38037.435972.EE

85. Little PS, Williamson I, Warner G, Gould C, Kinmonth AL, Gantley M. An open randomised trial of prescribing strategies for sore throat. *BMJ* 1997;**314**:722–7. http://dx.doi.org/10.1136/bmj.314.7082.722

86. Little P, Gould C, Williamson I, Warner G, Gantley M, Kinmonth A. Clinical and psychological predictors of illness duration from randomised controlled trial of prescribing strategies for sore throat. *BMJ* 1999;**319**:736–7.

87. Watson L, Little P, Williamson I, Moore M, Warner G. Validation study of a diary for use in acute lower respiratory tract infection. *Fam Pract* 2001;**18**:553–4. http://dx.doi.org/10.1093/fampra/18.5.553

88. Wigton R. Social judgment theory and medical judgment. *Think Reasoning* 1996;**2**:175–90. http://dx.doi.org/10.1080/135467896394492

89. Cooksey R. The methodology of social judgment theory. *Think Reasoning* 1996;**2**:141–73. http://dx.doi.org/10.1080/135467896394483

90. Goldstein W. Social judgment theory: applying and extending Brunswik's probabilistic functionalism. In Koehler D, Harvey N, editors. *Handbook of Judgment & Decision Making*. 1st edn. Oxford, UK: Blackwell; 2004. pp. 37–61.

91. Gonzales-Vallejo C, Sorum P, Stewart T, Chessare J, Mumpower J. Physicians' diagnostic judgements and treatment decisions for acute otitis media in children. *Med Decis Making* 1998;**18**:149–82.

Appendix 1 Preparation of group A beta-haemolytic streptococcus and commensal stock cultures

Each *S. pyogenes* strain was grown on Columbia agar with horse blood (CBA) (Oxoid, Hampshire, UK) for 18 hours in 5% CO_2 at 37 °C. From the resulting cultures, one colony was inoculated into 10 ml of brain–heart infusion (BHI) broth (Oxoid) and incubated for 18 hours in 5% CO_2 at 37 °C. The resulting suspension was diluted by transferring 1 ml of it into 100 ml of stock BHI broth. This solution was incubated at 37 °C without shaking until an optical density of 1.0 was reached at 600 nm, as measured with a Sanyo-Gallenkamp SP50 spectrophotometer (Integrated Services, NJ, USA). To accurately determine the CFU per millilitre of the solution, any clumped bacteria were separated. Each solution was centrifuged for 10 minutes at 3000 r.p.m., the supernatant was discarded and the resulting pellet resuspended in 100 ml of sterile saline (Oxoid). This suspension was vortexed to resuspend the bacteria, producing a stock solution with an approximate concentration of 5×10^9 CFU/ml. Performing 1/10 serial dilutions into additional sterile saline produced a range of suspensions from 10^2 to 10^8 CFU/ml. Exact bacterial concentrations were determined by spreading 100-µl aliquots from each dilution on to duplicate CBA plates (Oxoid). After incubation for 18–24 hours in 5% CO_2 at 37 °C, the colonies formed were counted and the bacterial concentration of each suspension calculated. Aliquots of 50 ml stock solution were prepared in sufficient numbers to carry out the total evaluation. Solutions were stored for a maximum of 4 months at –20 °C (Refrigerator Global 48F; ESTA, Vibocold A/S, Viborg, Denmark).

The culture method outlined above for sensitivity testing was repeated for the commensal organisms. The method was modified for the anaerobic commensals by substituting the BHI broth for Schaedler anaerobe broth (Oxoid) and incubation in anaerobic conditions (MACS MG500 Anaerobic Workstation, Don Whitley Scientific Ltd, Shipley, UK).

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Appendix 2 Manufacture's swab recommendations for five rapid antigen detection tests

Kit	Swab type	Additional suggestions
Clearview Exact Test	Polyester	Use swabs such as those provided
		Plastic shafts only, such as those provided
		Do not use calcium alginate, cotton-tipped or wooden-shaft swabs
OSOM Ultra Strep A	Rayon	Only use swabs provided
		Do not use swabs from other suppliers – not validated
QuickVue Dipstick Strep A Test	Rayon	Any rayon-tipped swabs, on solid-plastic shafts
		Do not use calcium alginate, cotton-tipped, hollow or wooden-shaft swabs
Streptatest	Polyester	Only use swabs provided
IMI TestPack Plus Strep A	Polyester	Polyester-tipped swabs only
		Do not use calcium alginate, cotton-tipped, hollow or wooden-shaft swabs

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Appendix 3 Detailed sensitivity results

The OSOM with its kit swabs was the most sensitive kit at GABHS concentrations of 2.5×10^6 CFU/ml and 5×10^6 CFU/ml: sensitivity was 59% (95% CI 48% to 69%) and 82% (95% CI 73% to 89%), respectively (see *Figure 2*). At a concentration of 7.5×10^6 CFU/ml, the OSOM/kit swab combination achieved a sensitivity of 95% (95% CI 88% to 98%), matched only by Clearview with polyester swabs. However, Clearview and polyester swabs achieved 100% (95% CI 96% to 100%) sensitivity at 10×10^6 CFU/ml, whereas the sensitivity of OSOM and kit swabs remained at 95% at this concentration. The IMI TestPack with its kit swabs achieved a sensitivity of 21% (95% CI 14% to 31%) at 2.5×10^6 CFU/ml, yet its sensitivity improved at higher GABHS concentrations, matching the OSOM/kit swab 95% sensitivity at 10×10^6 CFU/ml. The sensitivities of QuickVue and Streptatest kits were poor overall, with the best performances achieved with the integral kit swabs: maximum sensitivities at 10×10^6 CFU/ml of 70% (95% CI 59% to 80%) and 79% (95% CI 67% to 85%), respectively. Notably, the sensitivity of Clearview when using the kit swabs provided was considerably reduced, reaching a maximum sensitivity of only 62% (95% CI 51% to 72%) at 10×10^6 CFU/ml.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Appendix 4 Clinical variables in patients with group A, C or G beta-haemolytic streptococci compared with patients with no growth of C, G or A beta-haemolytic streptococci using more levels for variables

Symptom or sign	CGA [*n* (%)]	No CGA [*n* (%)]	Univariate OR (95% CI)	*p*-value	Multivariate OR (95% CI)	*p*-value
Prior duration						
5+ days	32/180 (18)	181/409 (44)	1.0		1.0	
3–4 days	74/180 (41)	131/409 (32)	3.2 (2.0 to 5.1)	< 0.001	2.0 (1.2 to 3.5)	0.010
1–2 days	74/180 (41)	97/409 (24)	4.3 (2.7 to 7.0)	< 0.001	3.6 (2.0 to 6.5)	< 0.001
Glands						
None	15/178 (8)	114/404 (28)	1.0		1.0	
Small	58/178 (33)	125/404 (31)	3.5 (1.9 to 6.6)	< 0.001	2.4 (1.2 to 5.0)	0.019
Medium	93/178 (52)	145/404 (36)	4.9 (2.7 to 8.9)	< 0.001	3.0 (1.5 to 6.1)	0.002
Large	12/178 (7)	20/404 (5)	4.6 (1.9 to 11.2)	0.001	1.2 (0.4 to 3.6)	0.768
Tonsils inflamed						
None	21/177 (12)	109/397 (27)	1.0		1.0	
Slight	29/177 (16)	110/397 (28)	1.4 (0.7 to 2.5)	0.322	1.2 (0.6 to 2.4)	0.670
Moderately bad	71/177 (40)	133/397 (34)	2.8 (1.6 to 4.8)	< 0.001	1.5 (0.8 to 2.9)	0.187
Severe	56/177 (32)	45/397 (11)	6.5 (3.5 to 11.9)	< 0.001	3.3 (1.6 to 7.0)	0.002
Runny nose						
Severe	4/180 (2)	17/409 (4)	1.0		1.0	
Moderately bad	12/180 (7)	63/409 (15)	0.81 (0.2 to 2.8)	0.741	0.9 (0.2 to 3.9)	0.904
Slight	34/180 (19)	106/409 (26)	1.4 (0.4 to 4.3)	0.599	1.1 (0.3 to 4.4)	0.853
None	130/180 (72)	223/409 (55)	2.5 (0.8 to 7.5)	0.109	1.6 (0..4 to 5.8)	0.506
Age group						
21+ years	93/181 (51)	265/411 (64)	1.0		1.0	
11–20 years	56/181 (31)	114/411 (28)	1.4 (0.9 to 2.1)	0.097	1.1 (0.7 to 1.8)	0.741
≤ 10 years	32/181 (18)	32/411 (8)	2.8 (1.7 to 4.9)	< 0.001	2.4 (1.2 to 4.7)	0.015
Sore throat						
Slight	6/181 (3)	51/409 (12)	1.0		1.0	
Moderately bad	80/181 (44)	199/409 (49)	3.4 (1.4 to 8.3)	0.006	3.0 (1.1 to 8.5)	0.033
Severe	95/181 (53)	159/409 (39)	5.1 (2.1 to 12.3)	< 0.001	3.8 (1.3 to 10.9)	0.012

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

APPENDIX 4

Symptom or sign	CGA [*n* (%)]	No CGA [*n* (%)]	Univariate OR (95% CI)	*p*-value	Multivariate OR (95% CI)	*p*-value
Cough						
Severe	1/181 (1)	34/410 (8)	1.0		1.0	
Moderately bad	21/181 (12)	108/410 (26)	6.6 (0.9 to 51.0)	0.070	3.7 (0.5 to 30.0)	0.225
Slight	42/181 (23)	94/410 (23)	15.0 (2.0 to 115.0)	0.008	6.6 (0.8 to 53.0)	0.077
None	117/181 (65)	174/410 (42)	23.0 (3.0 to 169.0)	0.002	8.6 (1.1 to 68.0)	0.041
Pus on tonsils	95/180 (53)	126/409 (31)	2.5 (1.8 to 3.6)	< 0.001	0.79 (0.5 to 1.3)	0.365
Fever (24 hours)						
None	38/181 (21)	175/410 (43)	1.0		1.0	
Slight	59/181 (33)	114/410 (28)	2.4 (1.5 to 3.8)	< 0.001	1.7 (1.0 to 2.9)	0.073
Moderately bad	63/181 (35)	88/410 (21)	3.3 (2.0 to 5.3)	< 0.001	1.5 (0.8 to 2.7)	0.184
Severe	21/181 (12)	33/410 (8)	2.9 (1.5 to 5.6)	0.001	1.0 (0.4 to 2.2)	0.946
Muscle aches						
None	61/180 (34)	206/409 (50)	1.0		1.0	
Slight	49/180 (27)	105/409 (26)	1.6 (1.0 to 2.5)	0.044	1.5 (0.9 to 2.5)	0.166
Moderately bad	50/180 (28)	75/409 (18)	2.3 (1.4 to 3.6)	0.001	2.6 (1.5 to 4.6)	0.001
Severe	20/180 (11)	23/409 (6)	2.9 (1.5 to 5.7)	0.001	3.7 (1.6 to 8.8)	0.002
Headache						
None	46/181 (25)	166/409 (41)	1.0		1.0	
Slight	55/181 (30)	127/409 (31)	1.6 (1.0 to 2.5)	0.054	1.2 (0.7 to 2.2)	0.484
Moderately bad	59/181 (33)	85/409 (21)	2.5 (1.6 to 4.0)	< 0.001	1.7 (0.9 to 3.2)	0.079
Severe	21/181 (12)	31/409 (8)	2.4 (1.3 to 4.7)	0.006	1.3 (0.6 to 3.0)	0.538

CGA; growth of A, C or G beta-haemolytic streptococci; OR, odds ratio.

Appendix 5

Number (%) of patients with Lancefield group A, C or G streptococci according to each level of a three-point 'basic' score (model 3), including the variables significant in multivariate analysis in both data sets (one point each for short prior duration, fever in the last 24 hours and severely inflamed tonsils). The total number at each level, and the percentage of the total sample in brackets, is also shown.

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

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Appendix 6 Data for score 1

TABLE 20 Symptom severity, antibiotic use, intention to consult in the future (moderately likely or more likely) and reconsultations with sore throat for score 1. Results are risk ratios (95% CIs) or mean differences (95% CIs)

Measure	Delayed prescribing (control)	Clinical score	RADT
Mean severity of sore throat and difficulty swallowing in the 2–4 days after seeing the doctor (0 = no problem, 6 = as bad as it could be)			
Crude mean	2.95 (SD 1.44)	3.05 (SD 1.49)	2.83 (SD 1.50)
Mean difference[a]		0.06 (–0.15 to 0.28; $p = 0.560$)	–0.12 (–0.34 to 0.10; $p = 0.270$)
Duration of symptoms rated moderately bad or worse			
Hazard ratio	1.00	0.95 (0.80 to 1.13; $p = 0.543$)	1.10 (0.92 to 1.31; $p = 0.282$)
Antibiotic use			
Crude percentage [*n* (%)]	111/284 (39)	137/294 (47)	98/281 (35)
Risk ratio[a]	1.00	1.20 (0.99 to 1.42; $p = 0.059$)	0.88 (0.69 to 1.09; $p = 0.265$)
Belief in the need to see the doctor in future episodes			
Crude percentage [*n* (%)]	91/278 (33)	79/285 (28)	76/273 (28)
Risk ratio[a]		0.85 (0.64 to 1.09; $p = 0.205$)	0.86 (0.65 to 1.10; $p = 0.248$)
Return to the surgery within 1 month with sore throat			
Crude percentage [*n* (%)]	43/374 (11)	34/380 (9)	46/359 (13)
Risk ratio[a]	1.00	0.76 (0.49 to 1.16; $p = 0.205$)	1.11 (0.74 to 1.62; $p = 0.618$)
Return to the surgery after 1 month with sore throat (mean follow-up 0.73 years)			
Crude percentage [*n* (%)]	75/374 (20)	84/380 (22)	69/359 (19)
Risk ratio[a]	1.00	1.10 (0.83 to 1.44; $p = 0.488$)	0.95 (0.70 to 1.27; $p = 0.728$)

a All models controlled for baseline severity of sore throat and difficulty swallowing and fever during the previous 24 hours.
Model for return within 1 month also controlled for prior antibiotic use. Model for returns after 1 month additionally controlled for prior attendance with sore throat, and follow-up duration.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

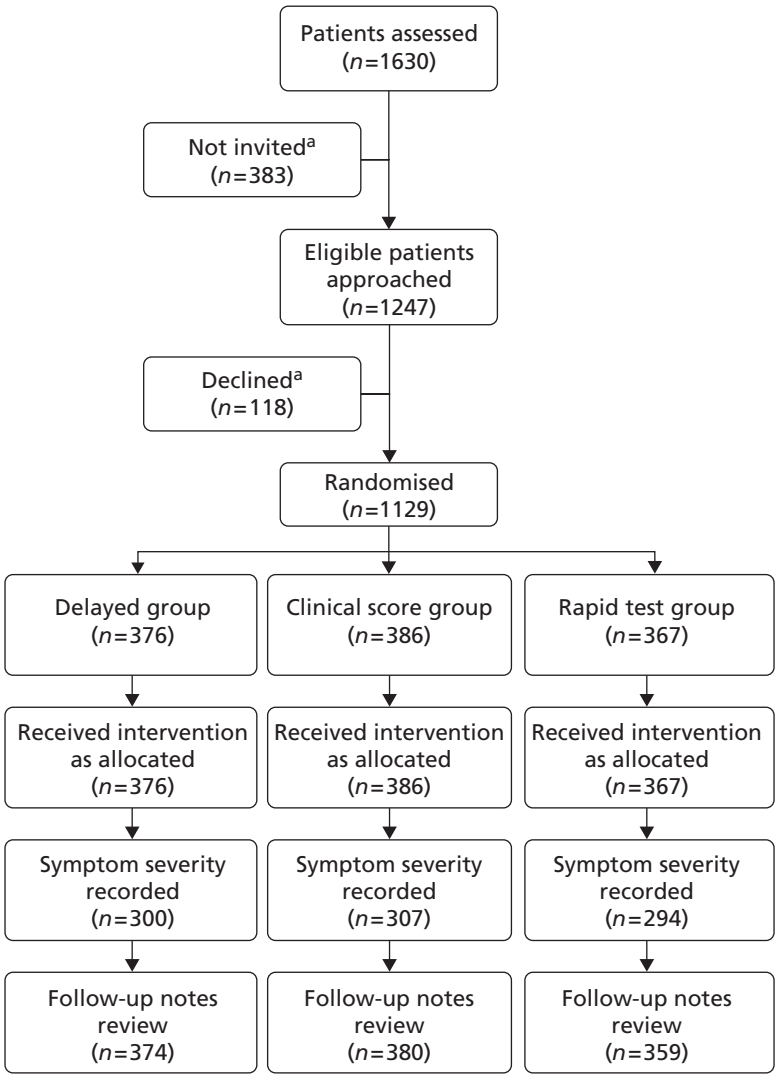


FIGURE 19 The CONSORT (CONsolidated Standards Of Reporting Trials) trial flow diagram for first phase of the trial (score 1).

Compliance with intended strategy for score 1

As with score 2, there was good compliance with the intended strategy in each group: overall, in 88% (982/119) of consultations the intended strategy was adhered to, and, when delayed prescribing was advised, 468/612 (76%) of patients were advised to wait at least 5 days.

TABLE 21 Strategy used by clinicians when using score 1

Strategy used by clinician	Control (delayed prescription) [*n* (%)]	Clinical score [*n* (%)]	Rapid test [*n* (%)]
No offer of antibiotics	30/376 (8)	122/385 (32)	200/367 (55)
Immediate antibiotics	20/376 (5)	74/385(19)	53/367 (14)
Delayed antibiotics	326/376 (87)	189/385 (49)	114/367 (31)

Appendix 7 Health professional testing of rapid streptococcal antigen detection test kits for ease of use

A panel of 10 health practitioners (GPs and NPs) did further testing of ease of use of the five RADTs kits. Each kit was tested four times consecutively by each participant, who were given test bacterial solutions to use as in the in vitro study (*Chapter 1*). The order of testing was balanced between individuals. The time taken to do the test was recorded after the first use and also after the last use, when practitioners were more practised. After each reading, the practitioners rated ease of reading the result and ease of use on seven-point scales (1 = very easy, 2 = easy, 3 = moderately easy, 4 = neither easy nor difficult, 5 = moderately difficult, 6 = difficult and 7 = very difficult).

All tests performed well. The time taken to perform the tests on the fourth repetition was documented and was around 5 minutes for all tests (Clearview 4.6 minutes, OSOM 6.0 minutes, IMI TestPack 4.7 minutes, Quickview 5.5 minutes, Streptatest 4.9 minutes). The time in practice to use the test would probably be nearer 8–10 minutes, as it would need to include the time taken to explain the test to the patient, prepare the test pack and reagents, take the swab and feed back the results. In terms of ease of use, on the fourth repetition the tests were all rated on average easy or moderately easy to use (mean ratings, respectively, 2.4, 1.6, 2.8, 2.6 and 2.0) and easy to interpret (mean ratings, respectively, 1.8, 1.5, 1.9, 1.5 and 1.8).

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Appendix 8 Protocol

Project title: PRImary care streptococcal management study (PRISM) (05/10/01) rapid tests for streptococcal sore throat) (Protocol version 1 7/7/06)

Planned investigation

Our Research Objectives are:

1. to assess which RADT is the most accurate in predicting the presence of group A streptococcus by throat swab in a clinical sample from primary care;
2. to estimate the error from sampling bias by performing parallel standardised in vitro studies;
3. to assess the validity of a scoring system based on the throat swab as the reference standard (such as the Centor criteria) in a UK population;
4. to assess the effectiveness and cost-effectiveness of rapid tests when compared to clinical scoring rules and delayed antibiotic prescription;
5. to explore the effect of additional benefit from the RADT use on GP diagnostic prediction accuracy and treatment decisions.

Existing research.

Overview

Antibiotic resistance is a major threat to public health: the key to reducing the risk from antibiotic resistance is to reduce use for those patients who will not benefit. Equally it is important for patients and society – particularly in terms of sickness absence – not to deny health benefit to those who will suffer severe or prolonged symptoms, and where possible to find effective alternatives to antibiotics. Sore throat is the commonest URTI managed in primary care, and primary care is where the majority of antibiotics are prescribed i.e. where the battle to improving targeting of antibiotics must be won.

This project, in providing key information about the most effective way of targeting antibiotics, has the potential to make a significant impact in improving clinical diagnosis for everyday clinical practice for patients with URTI. URTI is the only respiratory infection where there are good range of diagnostic alternatives (including both clinical scores and near patient tests), although there have been very few randomised trials of diagnostic methods. The impact is likely to be in improving the short term health – by minimising unnecessarily severe or prolonged symptoms and reducing side effects from unnecessary antibiotics; minimising the long term public health risks of inappropriate antibiotics; and providing a model for different management strategies for other RTIs.

Antibiotics

A Cochrane review suggests modest symptomatic benefit of antibiotics for sore throat and that antibiotics prevent complications,[15] which is supported by recent ecological data.[66] The solution is not to increase antibiotic use indiscriminately but to better identify individuals who are likely to benefit.

Better targeting?

Sore throat is one of the rare respiratory infections where there are several reasonable diagnostic alternatives. Showing which of these help best in managing symptoms and minimising inappropriate antibiotic prescribing will be invaluable for the management of sore throat – where antibiotics are still prescribed in 50–60% of patients.[67] Demonstrating what works in URTI may also be a model for what is possible for other RTIs. The available methods of diagnosis have been systematically reviewed:[1]

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Throat swab

This is the standard diagnostic method and the traditional clinical 'gold standard', but results take days, it increases costs, may miss significant infection (due to the organisms in the tonsillar crypts being different to those on the surface) and is not specific (due to prior 'carriage' of organisms in the pharynx[1]).

Near patient tests (NPTs)

Rapid streptococcal antigen detection tests (RADTs) are a practical alternative to throat swabs in managing sore throat. Although RADTs (and swabs) have the same limitations as throat swabs – i.e. they cannot differentiate infection from prior carriage, and may miss infection in the tonsillar crypts – they nevertheless have the potential to halve antibiotic prescribing for acute sore throat, and significantly improve targeting of antibiotics.[1,25,34,36] The potential disadvantage of rapid tests is that they may foster the belief that patients need to see their doctor in order to have the test in order for decisions to made about treatment i.e. potentially 'medicalising' URTI.[68]

Which RADT and how to assess them?

The MHRA[3] recently identified 5 RADTs marketed in the UK. Evidence from the previous literature suggests:

- **accuracy of RADTs** has mostly been evaluated in microbiological labs[3](67% of studies); the fewer clinical studies have rarely compared all the commonly available tests in the UK, nor among typical UK primary care populations – which is particularly important in view of 'spectrum bias' when not using primary care populations;[69,70] thus a clinical sample in the intended setting is important to assess overall performance characteristics of RADTs;
- **sampling bias** has rarely been fully addressed: the key issue determining performance of RADTs is the number of organisms harvested since the performance is strongly related to the number of colonies growing on the agar plate;[71] furthermore, the agreement of a test versus standard is not likely to exceed the comparison of the standard versus itself[72] – and in the case of throat swabs well taken throat swabs taken from the same individual only achieve 83%–91% sensitivity when compared to each other;[73–75] finally, since the sensitivities from previous studies comparing RADTs to throat swabs suggest that the better RADTs are likely to be in the above range (i.e. 85%+), sampling bias is likely to provide the main error in any clinical validation study of RADTs, and a clinical validation study alone may well not provide definitive answers regarding the best RADT to use; to get round this problem the performance in standardised conditions i.e. in vitro performance can also be estimated;[4]
- **comparison of ease of use** by practitioners have rarely been performed.[4]

Any assessment of an RADTs therefore has to assess performance in clinical settings (which includes the performance of the test in the intended setting and sampling error), performance excluding sampling error (by assessing performance in standardised conditions), and ease of use (time taken, ease of use and clarity of reading).

Clinical scoring methods

Existing clinical scores also have most promise to be useful in practice[1,25,36] – the major candidate being the 'Centor' criteria, which has been operationalised in two recent primary care trials[23,37] as 3 out of 4 of pus, cervical nodes, a history of fever and no history of cough.

Is it plausible that benefit exists from using clinical scores?

Even with the key limitations of validity identified above, preliminary indirect evidence indicates that existing scores may predict benefit. Clinical scoring methods which predict bacterial infection not only have the potential to predict symptomatic benefit from antibiotics, but also are likely to predict an increased risk for complications.[23,37] The trials included in the systematic review suggest 8–12 hours symptomatic benefit from antibiotics, whereas in two trials among selected patients with 3 out of the 4 'Centor' criteria,

1–2 days mean benefit was documented.[37] However this is indirect historical comparison: the patients in the systematic review were not necessarily comparable, and historical comparisons are notoriously unreliable. Other problems with the Centor criteria are:

- that there has been no robust validation in a typical UK population (i.e. the issue of spectrum bias since studies did not use typical primary care populations; the study in Ireland (Dobbs[25,34])with a similar population used univariate analysis only and thus was overinclusive);
- the criteria very probably have low specificity in primary care populations – 44% in a recent Canadian study[34] which would result in rather high rates of overall antibiotic use (46% of adults).[34]

This would suggest that using the Centor criteria alone will not significantly improve antibiotic targeting; that a modified approach should be considered based on validation using UK data if the Centor criteria are to be used. Once the best RADT and best clinical scores have been decided, this does not necessarily translate into predicting benefit for patients. It is thus crucial to test the performance of the best performing RADT and clinical score in a pragmatic trial against each other and against other treatment strategies, especially since the RADT strategy is likely to increase costs and may have 'medicalising' consequences.

MRC DESCARTE study

Since the submission of this application our group has also been funded by the MRC to undertake the DESCARTE study (DEcision rule for the Symptoms and Complications of Acute Red Throat in Everyday practice). This is a large cohort study and is very simple – using a one-page tick box clinical proforma only, and then subsequent documentation of adverse events. The great advantage to PRISM of the overlap between the studies is that we will be in a unique position of being able to compare the characteristics of patients and outcomes in PRISM with a wider clinical population to assess issues of generaliseability and spectrum bias. Furthermore both studies will provide important and overlapping scientific information; the results of DESCARTE (which will tell us about 'at risk' groups of patients) will be used in assessing the management and symptomatic outcomes of such 'at risk' patients in Phase II of PRISM; conversely PRISM will allow assessment of the potential for rapid tests and clinical scores to target antibiotics to individuals at risk of adverse events in the DESCARTE data set. Providing data to address this issue (i.e. the potential for the PRISM strategies to modify adverse events) will provide an additional dimension to PRISM that would not be possible unless DESCARTE was funded, and is made possible by compatible clinical proformas. We will be able to include the potential for the PRISM strategies to modify adverse events in our modelling exercise at the end of Phase II. Thus there are significant scientific advantages in the overlap between studies and little direct competition between these studies.

Research methods

Design summary

This study is in two phases:

Phase I is a validation and development phase and will include five components:

1. a clinical study to determine the ease of use and overall the performance in clinical settings of the five currently available RADTs using the throat swab as the reference standard;
2. nested data from the same sample will be used to assess whether the a scoring system based on the throat swab as a reference standard (such as the Centor criteria) requires modification, and to explore the feasibility of developing a clinical scoring rule based on immunological evidence of bacterial infection as the reference standard;
3. in vitro studies to assess the performance of RADTs in standardised conditions and thus assess the issue of sampling bias when using RADTs;
4. a qualitative study to explore patients and GPs' perceptions about the use of RADTs;

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Phase II. This trial will compare management using a) the best RADT defined from phase 1 compared with b) a clinical scoring rule (a Centor-like criteria based on predicting the results of throat swabs) and c) with the empirical strategy of delayed antibiotic prescription. Phase II will include a cost consequences analysis, which along with a review of the longer term effects of reduced antibiotic resistance will feed into a simple cost effectiveness model.

Methods

Phase 1

1) Clinical study of RADTs

Inclusion: Adults/children aged 5 and over presenting with acute sore throat (< 2 weeks; and with some abnormality of examination of the throat – i.e. erythema and or pus – as in our previous studies in primary care).[23] Although there some evidence that those presenting acutely are more likely to have bacterial infection[51] and more likely to benefit from antibiotics it is important that the performance and use of tests reflect the generalisable population presenting in primary care. Exclusion: other non infective causes of sore throat (e.g. aphthous ulceration, candida, drugs), unable to consent (e.g. dementia, uncontrolled psychosis).

Throat swabs Despite the theoretical potential for either overgrowth, failed growth, or poor operator performance of tests, evidence suggests that:

- the performance of a test based on a swab done in the doctors practice are equivalent to the results from the same swab in the laboratory;[3,76,77]
- one swab can also be used for more than one plate – providing identical numbers of colonies on up to 5 plates;[3,78]
- two double swabs can be done in adults.[73]

Thus any clinical validation study can take advantage of using two double swabs in each adult, using the same swab for both RADT and culture, and can minimise practice disruption by letting laboratory staff perform the tests. In adults two double throat swabs will be taken (allowing four tests for each adult), but in children only one double swab is likely to be acceptable.

Based on the above evidence each swab will be sent to a central laboratory (which was shown to be feasible in piloting). Each swab will be used for both conventional microbiology (culture and sensitivity – using Todd-Hewitt broth, which provides the best reference standard in this context),[3] and also for one rapid test. Our piloting in 60 patients has also confirmed that using the same swab for one RADT and the culture is both feasible, and minimises sampling variation.

Analysis The rationale of the rapid test is to replace the need for a throat swab, and obtain the same clinical information as the throat swab, but in a much more timely manner. Thus the primary analysis of the accuracy of the RADTs will be the analysis of 2 × 2 tables comparing RADTs with the results of the throat swab as the reference standard, calculating sensitivity, specificity, predictive values and likelihood ratios. Secondary analysis will also allow us to explore the relationship between RADTs and evidence of bacterial infection using a rise in immunological titres as the reference standard (see below).

The criteria for choosing RADTs are a) acceptable sensitivity (> 80%) from previous in vitro or clinical studies (based on the systematic review by the MHRA[3] and the recent French Agency assessment)[4] b) ease of use[4] c) availability and EU 'CE' marking. In terms of availability in the UK and CE marking, a recent review this year by the MHRA identified 5 tests as being available and marketed in the UK[3] (Signify Strep A (Abbott); Directigen 1, 2, 3, (Beckton/Dickinson);OSOM Ultra Strep A (Genzyme); QuickVue in line (Quidel) Strep A OIA MAX (Thermo Biostar)). However, since the MHRA review Directigen 123 and Signify Strep A are no longer available. Instead of Abbott Signify strep A we propose using the better performing Abbott

Test pack plus Strep A[3,4] – which superseded the Signify test and is now marketed by Unipath as IMI TestPack plus Strep A. The Directigen test we propose replacing with Streptatest (Dectrapharm, Strasbourg) which is available for UK use, CE marked, and performed very well in the French Health Products Safety Agency tests.[4] Streptatest performed very well both the in vitro studies (in the top 3), and was also rated as easiest to use of all 16 tests compared. All our proposed tests performed acceptably for ease of use (range 24–35 out of a maximum 38, with the OIA max test performing worst and the Streptatest best) [4] and sensitivity.[3,4]

Summary of performance of proposed rapid tests for Phase I clinical study

	*Sensitivity (compared to throat swab)[3]	In vitro studies % detection of low bacterial counts (10^5 cfu/ml)[4]	Ease of use score (maximum 38)[4]
Streptatest	96% (from company)	75%	35
OSOM Ultra	91%	no data[4] (from company website: Mass. General Hosp. study = equivalent to QuickVue)	no data[4] (similar tests averaged 30+)
Strep A OIA Max	92%	50%	24
QuickVue	87%	50%	34
TestPack Plus	89%	100%	28

*The studies were mostly not based in primary care and for the few clinical studies that have been performed in typical primary-care settings the sensitivities are lower[79] than reported in the review.[3]

Sample size Most RADTs are very specific.[3] Sensitivity is the limiting factor, and the sensitivity from better performing previous studies is in the range of 80–90%.[3] However we will be comparing a rapid test versus the results from the same throat swab which should provide higher sensitivities. Assuming 25% of the sample have streptococcus (based on our piloting) and a sensitivity of 85%–95% for the best RADT to estimate, with 95% confidence, sensitivity to within +/–5% (i.e. to be confident that the sensitivity is not less than 80% which would be less useful clinically) then 73 to 196 samples with streptococcus are required for each RADT (see table below), or 292–784 in total, 1,460- 3,920 allowing for the 5 RADTs, or 1,500–4,000 allowing for some leeway in the assumptions. If each adult provides four comparisons and each child two, then 400–1,200 patients are required. Thus our minimum sample size is 438 and maximum 1,176.

Sample size to estimate sensitivity with 95% confidence intervals of +/–5%

	Sensitivity		
	85%	90%	95%
Sample with streptococcus present	196	139	73
Sample for each RADT (assuming 25% have streptococcus)	784	556	292
Total number of tests (for 5 RADTs)	3,920	2,780	1,460
Number of patients required (assuming two double swabs in adults and one double swab in children)	1,176	834	438

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

2) Development of clinical rule

Confirming the validity of the Centor criteria in a modern UK primary care population This study provides an opportunity to modify a clinical rule (such as Centor) which aims to predict the results of throat swabs. The rationale for this is that there have been no UK studies which have independently predicted the presence of Streptococcus (the Dobbs score used univariate analysis only, hence was over inclusive and rather unwieldy);[25] furthermore multivariate analysis of our pilot data in 120 individuals suggest the key independent variables to predict the presence of streptococcus were a history of fever at any time in the illness, a history of fever in the last 24 hours, the absence of cough, anterior cervical glands and muscle aches i.e. different from Centor. There are no additional costs to this aspect of the study since we will be documenting clinical details and taking throat swabs anyway. All patients in this phase will be offered a delayed prescription (antibiotics to be used after 3–5 days collected from the GP practice reception if symptoms worsen or not starting to settle[33]). We have used this approach in piloting which was acceptable to patients and clinicians.

3) In vitro study of RADTs

The number of organisms harvested crucially determine the performance of the RADT,[71] and two well taken samples from the same individual will predict the other results with a sensitivity of 83%–91%.[73–75] Variation in RADT performance eliminating sampling bias will be assessed by in parallel in vitro studies (using different antigen loads comparable to the data coming from clinical studies, and also controls). The most recent and comprehensive of in vitro studies compared 16 tests but only included 4 of the 5 tests proposed,[4] and performed very few tests for each RADT. Thus, more comprehensive comparative data are needed.

As in the most recent in vitro study[4] we will compare the performance of RADTs using 4 strains of group A beta-haemolytic streptococci and a control strain (a Group C streptococcus, or Moraxella), in three dilutions each (10^5, 10^6, and 10^7 Colony forming units per ml) corresponding to the range of growths normally seen from throat swabs in the community. We will perform 20 tests for each RADT at each dilution and for each strain (i.e. 200 tests for each RADT) which will provide similar power to the clinical study.

Assessing ease of use of RADTs by clinicians Most of the in vitro tests will be carried out by laboratory personnel who will rate ease of use, but we will also arrange a panel of 10 GPs and practice nurses – who will perform the tests in clinical practice – to perform four tests with each RADT, and the order of RADTs will be randomised. Several tests are required so that each clinician gets used to doing each test competently. For each test the clinician will document time to do the test, time to get the result (seconds), overall ease of use (on a five-point Likert scale – very easy, easy, neither easy nor difficult, difficult, very difficult) and clarity of result (clear, unclear). After using all the RADTs each clinician will be asked to rank the RADTs in order of their overall preference.

4) Qualitative study

The qualitative study will be based on grounded theory methodology[80] to clarify patients' and primary care professionals' understanding and concerns about the use of RADTs in both phase 1 and phase 2. Grounded theory provides a framework for guiding decisions about sampling, data collection and analysis. It is furthermore a powerful method for discovering, understanding and developing explanatory concepts. A maximum variety sample of 15–20 patients and 15–20 GPs and nurses will be recruited. The purposive sample of patients will include both sexes, a range of ages, people from different socio-economic classes and minority ethnic groups. In keeping with grounded theory we will undertake further theoretical sampling to: a) extend information on the concepts identified, b) contrast/confirm/challenge the data already collected and c) to fill in any missing gaps in the data. An interview guide will be designed to reflect the study's objectives and existing literature in the area. The interview guide will be flexible and responsive to the respondent's narrative. This means that the results of earlier interviews will inform the

kinds of questions asked at subsequent ones, in an iterative fashion. However, we expect that the questions asked will explore the following general areas: the experience of sore throat or managing sore throat, decision-making or perceived decision making about consulting the general practitioner, perceived outcomes of consultations, decision-making about RADTs, and the consequences of treatment. In this way, the interview will follow the respondents' agenda as far as possible, while remaining relevant to the issues of the use of RADTs. Open ended face-to-face, in-depth interviews will be carried out and audiotaped for transcription; field notes will be kept and memoranda written to aid with analysis. Transcripts will be analysed systematically through constant comparison analysis. Each interview will be analysed to identify primary issues and categories, a process known as open coding. These categories will then be compared with others within the transcript and across other transcripts, as well as to categories and concepts within the existing literature. The next stage, axial coding, will cross-link the concepts to generate new meanings and concepts. The final stage, selective coding, will cross-link the concepts to generate themes that will represent the most theoretically abstract unit of analysis from which theoretical explanations can be generated. The qualitative research will provide explanatory models clarifying understandings and concerns around the use of RADTs for both patients and doctors, sensitive to the context within which they experience and manage the illness. The model will inform the implementation of the randomised controlled trial – as well as the treatment and management of sore throat and the use of RADTs more generally. The qualitative work will start in phase 1 but continue in phase 2. The purpose of the qualitative work in the trial phase is twofold a) to help understand issues surrounding the process and experience of intervention early on in the trial to be able to modify trial procedures and trial documentation in the feasibility phase b) to understand attitudes and experiences of the health-care professionals and patients involved in the trial that can help to explain the quantitative outcomes.

Outcomes from Phase I

The aim of Phase I is to provide sufficient outputs to inform the randomised trial. We anticipate the outputs will be the estimates of sensitivity and specificity of RADTs and the Centor criteria (or modified criteria); ease of use of RADTs; a qualitative model to understand patients' and GPs perceptions and to inform strategies for phase 2.

Phase 2

Planned interventions

Patients will be individually randomised using a web based randomisation service – with permuted block size of 3, 6, 9, and 12 being randomly chosen – to three groups, stratified by physician belief in the likelihood of streptococcal infection (rated as either a probable bacterial infection or probably not a bacterial infection).

(a) **RADT using the best RADT from phase 1.** Depending on Phase I results and our modelling exercise, we provisionally assume the most efficient use of RADTs will be by targeting them to those with intermediate clinical scores. Thus those with low clinical scores (e.g. 0/1 Centor) will be unlikely to have bacterial infection; those with very high clinical scores (e.g. 4 Centor) much more likely to have bacterial infection. Only patients with a positive result from the RADT will be offered antibiotics. All patients – as in the other groups – will be advised to use analgesia (regular paracetamol and/or ibuprofen).

(b) **A clinical scoring rule** – either the Centor criteria, or the clinical rule developed from Phase I (whichever performs best from Phase I). The precise strategy to use the clinical score will be based on the results of Phase I and the modelling exercise. However we anticipate that it is likely that antibiotic will not be offered at all to those with very low scores (e.g. Centor 1), for high scores (e.g. Centor 4) then immediate antibiotics will be advised unless symptoms settle rapidly within 48 hours, and for intermediate scores delayed antibiotics will be offered (see c). Such a modified use of the criteria is necessary since using a single cut-off leading to antibiotic use (e.g. 3+ Centor) from recent data nearly would mean 50% of patients are likely to need antibiotics.[34]

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

(c) **The empirical strategy of delayed prescribing** (prescription to be collected from reception after 3–5 days if symptoms are not starting to settle, or sooner if symptoms get significantly worse). Based on previous work this is likely to result in 25% taking antibiotics.[33] The rationale for delayed prescribing is that it is safe – and arguably safer than not prescribing at all since it provides a back up for unwell individual or those deteriorating, and changes belief and reconsultation behaviour as effectively or possibly more effectively than not prescribing (based on both our previous study in sore throat[46] and recent data in lower RTIs[47]). It has been incorporated widely into routine practice in the UK since our 1997 trial[33] without any increase in complications of sore throat.[48]

Arguments for individual versus cluster randomisation.

Overview We will make a final decision on the method of randomisation by undertaking pilot work in phase 1 and before the main recruitment for phase 2. We set out the arguments for and against the two possible methods (individual or cluster) below. There is no doubt that individual randomisation is optimal and it is our preferred method (the proposed design and sample size is therefore currently based on this approach) and we have implemented individual randomisation before in several previous trials. However, we recognise that this is logistically more difficult to implement than a cluster design in a study of this scale. If individual randomisation proves logistically impossible, we will adopt a cluster design but with rotation of intervention within practices in different seasons to try to minimise bias. Obviously any logistic benefit of cluster randomisation will have to be traded-off against the increased sample size required. We propose patient based randomisation (i.e. individual) rather than practice based (i.e. cluster) randomisation based on the following arguments and practical experience:

Cluster randomisation This works well when there are no differential pressures on recruitment between groups. For this trial a cluster randomised trial (i.e. randomising by practice) would very probably result in differential recruitment bias between groups. This study may well change such perceptions of clinicians, but we have to go from where we are: RADTs are used widely in Europe and the USA, but not currently in much use in the UK,[3] and the RADT group is likely in the current climate to be less attractive since it is more time-consuming, involves some minor discomfort for patients (and in some, gagging and vomiting). Practices in the UK randomised to use RADTs alone would therefore very probably recruit less well, and with different patients in RADT and control groups. The issue of differential recruitment in cluster trials is not theoretical – it occurred recently in the MRC UKBEAM trial both in terms of differential numbers and differential characteristics of patients – where there were considerably fewer incentives for differential recruitment than the current proposed trial – and the cluster design element of the trial had to be abandoned. The other disadvantage of cluster randomisation is the additional design effect requiring inflation of the sample size.

Individual randomisation The potential disadvantage of individual randomisation (i.e. by patient within practices) relates to concerns about group differentiation: however each group will be administered using a standardised manualised approach. Our group now has extensive experience in the successful completion of 13 such trials
including several recent behavioural trials, antibiotic prescribing strategy trials very similar to the current trial proposed[81–83] and a lifestyle intervention trial.[84] The manualised approach maintains clear group differentiation despite individuals being randomised to different groups by the same health professional[81–84] and the data suggest that using such an approach the health professional and practice cluster effects are minimal.[49,81,85]

Thus our aim will be to use individual randomisation and only if this proved unfeasible in the feasibility phase then use practice-based randomisation.

Planned inclusion criteria Previously well subjects with acute illness (2 weeks or less), presenting with sore throat as the main symptom, with an abnormal examination of the pharynx (similar criteria to previous studies of sore throat in this group).[23] Most patients present within 5 days, but a smaller minority

present with a longer duration of illness prior to seeing the doctor.[33] Since we wish this sample to be representative of those patients presenting to GPs,[33] and prior duration predicts subsequent illness duration[86] (i.e. an important group to help) we do not wish to exclude those with longer prior duration of illness.

Exclusion criteria Quinsy, previous rheumatic fever and glomerulonephritis. Serious chronic disorders where antibiotics are needed (e.g. cystic fibrosis, valvular heart disease), or mental health problems (e.g. learning difficulties – unable to complete outcome measures).

Informed consent Parents will have to sign a consent form on behalf of children. Some of the older children e.g. the 4 and 5 year olds who have some language will be able to understand the patient information leaflet, and will be encouraged to sign or mark the consent form. GCP (Good Clinical Practice) training will be provided to all participating practices, with particular emphasis on the complexities of randomising children in clinical trials

Proposed time period for retention of relevant trial documentation

Trial documentation will be kept for 15 years.

Proposed outcomes/data collection

Clinical data Baseline clinical data will be collected[24–26] as in phase 1.

Diary scores.[23,33] Each symptom is scored 0 = no problem to 6 = as bad as it could be: sore throat, difficulty swallowing, feeling unwell, fevers, sleep disturbance) which patients fill out on all days until their symptoms have resolved. A telephone call from the research assistant (RA) in the first few days resolves any problems the patient may have filling out the diary. We have chose the two item score (sore throat, difficulty swallowing) as the main outcome as it is more reliable than either item alone and is internally reliable (Cronbach's alpha = 0.92); these simple diaries have been used in several of our studies[33,47,49] and are also more sensitive to change than criterion measures.[87] Temperature will be taken by patients and documented on a daily basis in the diary using tempadot thermometers as in our previous studies.[23,49]

Antibiotic use It is vital to document antibiotic use since rapid tests may achieve the same symptomatic benefit as the other strategies, but with the advantage of reduced antibiotic use. Our proposed method is self-report, using a box at the front reception for delayed prescriptions, and also documenting prescribing information from notes. One alternative is to trace prescriptions using stamps; this would require GPs to use stamped prescriptions and possibly make them less likely to recruit given the ease of mislaying pads/stamps (most GPs now print prescriptions); also cashing a prescription does not mean it is used.[3] It is also not feasible nor sensible to use more invasive methods of documenting antibiotic use (e.g. 'smart' containers, urinary antibiotic estimation etc.) since these are liable to artificially alter antibiotic compliance, and thus potentially modify symptomatic outcomes. Thus our main outcome for antibiotic use is self report (i.e. to give patients 'permission' to say whether they used antibiotics or not) backed by the evidence from unused delayed prescriptions and notes review- as we have documented in our previous studies.[3,4,6] We have previously showed that self report from the diaries agreed well with whether delayed prescriptions were collected,[3] that self report correlated with weighed bottles for paracetamol use[4] – supported by another study in our group which has compared self report and weighing (Prof. Mant personal communication).

Side effects It is also important to document side effects of antibiotics since rapid tests may achieve the same symptomatic benefit as the other strategies but with the advantage of reduced side effects by minimising antibiotic use. Diarrhoea and skin rash will be documented in the diary, and also – where these are serious enough to contact the doctor – from the notes review (see below).

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Duration of illness The diaries will also allow us to document duration of illness (until very little/no problem), the duration of moderately bad illness (until rated less than a moderately bad problem),[47] antibiotic use, and use of over the counter medicines (also see notes review).

The medicalisation of illness Patients' belief in the importance of seeing the doctor will be documented using 5 point Likert scales completed by patients[33] which we have shown to be reliable.[33] We will also document patients reconsultation behaviour by blinded notes review[23] (see below).

Time We will document time taken in the consultation (on the same sheets as the clinical sheets). Patients will document time off work and or time to resume normal activities in the diaries.

Socio-demographic data Age, gender, household income, social deprivation indices based on postcode will be recorded.

Notes review During the available follow-up time (which will vary from 1 month to 2 years) all patient's notes will be reviewed to document returns, time to return, reasons for returns, complications, economic data (see below) and any subsequent referrals.[23]

Sample size Phase II. 'Medicalising' effect of using RADT. (alpha = 0.05, beta = 0.2)

This is the limiting sample size calculation although not the primary outcome. A good proxy for 'medicalising' behaviour is the change in beliefs about the need to see doctors in future episodes – assuming there are 15% differences between groups (22% were observed in our previous trial)[33] then only 152 patients per group are needed (see table below). However a harder behavioural outcome is preferable: to assess the medicalising effect on reattendance behaviour; if we assume that using RADTs may change subsequent attendance by 11% (RADT 38%, clinical score 27%, delayed prescribing 27%) – as observed in the medicalising effect of prescribing strategies in a previous trial over a similar follow-up period[33] – then 254 patients per group are needed or 849 in total allowing for 10% loss to follow-up of notes.[49] An alpha of 0.01 and beta 0.1 would require an unfeasibly large sample (460 per group or >1500 patients in total) or another 2 centres for the trial.

Primary outcome: symptom severity (alpha = 0.01, beta = 0.1)

Diary score The time when an RADT is most likely to help patients is when the inflammation due to bacterial infection is at its greatest in the first few days after seeing the doctor. We assume the minimum effect size for the symptoms severity score is a 0.33 standardised effect size (i.e. 0.33 SD) on days 2–5 is when patients rate their sore throat at its worst. To detect a 0.33 standardised effect size difference between the RADT group and control groups (assuming both control groups are 0.33 SD higher than the RADT group) requires a minimum of 134 per group (for alpha = 0.05, beta = 0.2) but preferably for (for alpha = 0.01, beta = 0.1), 242 per group, or 909 patients in total allowing for 20% loss to follow-up of diary information.[33,47] A standardised effect size of 0.33 is classified as a small effect size, (0.33 SD is equivalent to half patients rating sore throat a mild rather than moderately bad problem, or duration of sore throat one days difference), and in this context this order of effect size was judged to be the smallest worthy of treatment by general practitioners.[47] A much smaller effect size (e.g. 0.25 standardised effect size) would result in an unfeasibly large sample (see table below). We will explore in the phase 1 (including the qualitative work) whether such an effect size (i.e. 0.33 SD) is regarded by patients as being the minimal worth treating. An alpha of 0.01 is preferable since it allows for type I error – with 2 comparisons between RADT and the other two groups, and also for multiple outcomes (severity/duration), and a beta of 0.1 will help ensure that we do not miss an effect in our primary outcome.

Sample size in each group[1] –range of options for Phase II trial (our proposed sample sizes for each outcome are in bold)

	Difference between RADT group and other two groups	Alpha 0.01 Beta = 0.1	Alpha 0.01 Beta = 0.2	Alpha 0.05 beta = 0.2
Standardised effect size (continuous outcomes e.g. symptom severity, duration, time to first return for sore throat)	0.5 (i.e. RADT 0.5 lower than other 2 groups)	107	85	59
	0.33	**242**	193	134
	0.25	420	335	233
'Medicalisation' and antibiotic use:				
Proportions: behaviour (return to surgery); antibiotic use	11% (38%, 27%, 27%)	460	366	**254**
	15% (42%, 27%, 27%)	253	202	140
	20% (47%, 27%, 27%)	146	117	81
Proportions: beliefs (in the need to see the doctor; belief in antibiotics)	15% (57%, 57%, 72%)	274	219	**152**
	20% (57%, 57%, 77%)	152	121	84

1. We used the NQUERY multiple group sample size programme for three groups and assumed both clinical score and control groups had similar figures; if the control group fares worse than the clinical score group (i.e. the spread of observations is wider) fewer numbers will be needed in each group. The numbers in each cell are the numbers with complete data required in each group.

Statistics analysis, type and frequency

We will perform analysis of covariance for the main continuous outcomes (diary scores, symptom duration). Log-rank tests and Cox regression will be used to assess the time to return to the surgery with a new episode of sore throat as we used in our previous trial[23] (which will involve 'censored' data due to variable follow-up time available). We anticipate logistic regression will be used for dichotomised outcomes (e.g. belief in the need to see the doctor, belief in antibiotics) and Poisson regression for incidence rates (e.g. rates of return to the surgery with sore throat – which more closely follow a Poisson distribution rather than a normal distribution). The models will control for stratification, and for confounders if appropriate (although randomisation should ensure that confounders are balanced between groups). We will present the results with 95% confidence intervals. The primary analysis will be an intention to treat analysis based on finding the differences proposed. We will also perform secondary analyses: a) a per protocol analysis, and also b) an equivalence analysis if appropriate (having established in our qualitative work what patients regard as equivalent). The clinical and demographic characteristics of a) eligible patients not consenting, and also b) those not followed up, will be compared to assess respectively possible selection and non-response bias.

Our key presentation of the data will be of the main symptomatic outcomes alongside the data on antibiotic use and side effects from antibiotics.

Economic analysis

The type of economic analysis is informed by the likely directions of changes in both costs and benefits are summarised in *Table 1*. This shows that cost effectiveness may not be an issue, and if it is, only short term cost differences will be captured in the trial. Long term cost effects of reduced antibiotic resistance can only be estimated from the literature, and are likely to be highly uncertain.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 1 Likely changes in costs and health effects due to RADTs (or rule)

	Cost effects		Health effects	
	Short term	Long term	Short term	Long term
RADTs (or rule)	Up (cost of tests, + medicalisation)	Up (medicalisation)	Unchanged or improved	Unchanged or improved
Antibiotic use	Down	Down (less costs due to AB resistance)	Up (fewer side effects)	Up (Less AB resistance)
Net	?	?	Up	Up

The table indicated that net cost effects are uncertain but benefits likely to increase, modestly in short term, perhaps more in longer run, but subject to uncertainty. If costs were reduced, then RADTs would be dominant (benefits up, costs down). Only if costs increase is cost effectiveness an issue.

Short term costs depend on whether RADTs or a rule is preferred, as the costs of the latter would be low. If RADTs were preferred, their increased could lead to reduced cost per test over time (whether short or longer term). Reduced antibiotic use would reduce costs immediately. Medicalisation in the short term would increase costs. Net short term net costs could move up or down.

Longer term costs depend on the balance between the increased costs of medicalisation (long term) and the possibly reduced costs due to reduced AB resistance. As neither of these are likely to be established with any certainty in the trial (follow up 1 or 2 years?), various assumptions will have to be made and tested in sensitivity analysis. As costs in the future would be reduced to net present values by discounting, their timing would also be important. Overall, it seems possible that a rough balance may prevail on long-term costs.

The economic analysis will provide:

1. a cost consequences analysis plus a simple cost effectiveness model to be used for sensitivity analysis linking costs to each of the main outcomes in the trial; and
2. a review of the literature on costs and benefits of reduced antibiotic prescribing which would be included in the model if and as appropriate. Further work might then be deemed worthwhile or not, an issue which would be discussed with NCCHTA and a case made for further resources if appropriate.

The cost of intervention and follow-up related service use, including the time taken to train staff, surgery attendance, admissions and referrals, will be collected in the trial. The major increment in health service costs associated with advice to use antibiotics and/or RADTs are likely to be due to the time taken in the index consultation, and the effect on subsequent consultation behaviour.[23,49] Resource use data will be collected by notes review, GP and nurse documentation (e.g. of consultation time) and patient self-report. Although our primary analysis will be from the health service perspective we are also interested in the personal costs of managing sore throat. Thus during piloting and in qualitative work we will explore the range of resource use – e.g. pharmacy use, transport to pharmacy and to surgery, time taken off work/ school etc. – to make sure we are not missing important costs. During this phase we will also explore streamlined computerised methods of collecting NHS data on resource use from GP notes since most practices are likely to be computerised. We will model the potential long-term economic costs and health benefits by extrapolating the trial's results based on assumptions about behaviour change among GPs and patients. The potential impact of any new generation rapid test will be included in post trial modelling.

Unit costs will be based on national rather than local unit costs wherever possible to aid generalisability of the results. The first phase of analysis will be to perform a cost-consequence analysis where the health

service costs and consequences of the different strategies are compared – symptom duration, symptom severity, quality of life in the immediate episode, antibiotic use, side effects of antibiotics (diarrhoea, rash). A simple model will be constructed to estimate incremental cost per moderately bad sore throat prevented, per day with sore throat, with and without side effects. These data will be collected either as part of Study 1.

The implications of reduced antibiotic use and changed antibiotic resistance will be derived from a systematic review of relevant studies. While this element is difficult to quantify, ignoring it would be to set these key effects to zero.

These longer terms changes in costs and benefits will be included in the model. Sensitivity analysis will explore plausible scenarios and the scope for further more detailed work.

Assessment of the potential effect on clinical behaviour

This sub-study will allow us to understand some of the key issues in applying the trial evidence in practice. A Judgement Analysis (JA) study based on social judgement theory[88] will be used at the end of Phase II to estimate the impact on GP behaviour of the trial results (availability of RADT information and/or clinical scores) in their assessment of patients. We will present the study results to both participating GPs and a further sample of 'naïve' GPs. We will construct a series of vignettes[89] presenting combinations of clinical characteristics (cue profiles), including RADT results and/or clinical scores.

The Social Judgement Theory advocates a 'representative design',[90] therefore vignettes should be representative of patients that doctors deal with in their practice. To this effect, cue profiles will be constructed from the patients participating in phase 1. Out of these, a certain number of cue profiles will be randomly selected. The number will depend on the number of cues that we decide to include – as a rule-of-thumb, 5–6 cues require a minimum of 30 cue profiles, but feasibility (stamina and patience of GPs) argues for a limited number of cues and cue profiles. Vignettes will be presented to GPs who will be asked to estimate the likelihood of each patient having streptococcus (on a 0–100 VAS), and to decide whether they would prescribe antibiotics or not.

The judgement and treatment policies of each GP will then be modelled as separate regression equations. These equations will show which cues each GP actually used in assessing likelihood of infection and making treatment decisions (a cue is considered used if its regression co-efficient is significant). We hypothesise that GPs who took part in the main trial are more likely to use the RADT results or the clinical scores than 'naïve' GPs both to assess likelihood of streptococcus infection and to decide about treatment.[91] To assess consistency of cue use and judgments/decisions, the same vignettes will be presented to the GPs a week later in a different order. Development and preparation of the vignettes will commence during the last 9 months of the study, but the final version of the vignettes can only be sent to the GPs once we have the results of the trial. We will therefore analyse the JA study in the penultimate month of the proposed study period allowing a month for write up.

Risk and benefits for trial participants

Since the study will use existing widely practiced strategies (but used in an ad hoc manner) there should be no risk to participants. All participants will be advised to return to the doctor if their symptoms are worsening. Major complications are unlikely in this sample,[23] and given structured advice to patients – probably better information than is normally available in routine care – the standard of care they receive is likely to be higher than routine practice. We have chosen the control group to be delayed prescription which provides both a way of minimising antibiotic use, minimising the medicalisation of illness, reducing return rates to the surgery, and a safety net for patients: it is not associated with any higher risk of complications.[46,48]

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Independent supervision

We will nominate a Chairman and two independent members, one of whom will be a statistician, to form a trial steering committee (TSC) which will meet early in the feasibility phase, and thereafter annually unless there are problems in which case more frequent meetings will be arranged.

Recruitment

The current research climate in primary care makes robust recruitment essential. To ensure recruitment we will book nurse sessions in advance with adequate reimbursement (for both opportunistic referrals, and the invited community sample), and very conservatively assume that 1 : 3–4 sessions will be receive a referral, of these 1 : 3–4 parents agree, and that up to two full winters may be needed for Phase I, and two for Phase II. Three centres are needed (Southampton; Birmingham and Oxford), and Oxford's experience of recruiting children for such studies will be invaluable. In recent years in the difficult research climate in primary care our three centres have shown that they can recruit both adults and children in the numbers required for the successful completion of this study. We are keenly aware of the issue of feasibility, and agree that for consultations for some GPs on some days there may not be time – which is why we have assumed conservatively that recruitment may be as little as 1/2 of what we expect in one winter, and have gone further to allow two winters for both phases. We will explore the issue of any possible bias due to differential recruitment rate among GPs in analysis, but previous studies have suggested little evidence of recruitment bias.[33,47]

The clinical proformas, and taking of the two double throat swabs, provides the main work for GPs/nurses. The clinical proformas are very simple, and they have already been piloted, and so has the taking of throat swabs. One of the arguments for the feasibility study is to be able to confirm recruitment rates, potentially overcome any issues of feasibility, and as necessary widen the net of GPs (and in the worst case scenario of course stop the study, and thus minimise risk to the NCCHTA).

The GPs we are recruiting will all be part of local Networks. Our group has a wide experience of recruitment and retention of GPs and patients to primary care trials. A number of strategies can be utilised to maintain GP recruitment including recruitment holidays, electronic reminders, Network newsletters, and incentives – so GPs will not be allowed to 'forget' the study, (also see trial management).

Will recruitment and logistic organisation for PRISM overlap with the MRC DESCARTE study? The proposed study for this application (PRISM) is more intensive than the DESCARTE study – with clinical sampling. Therefore PRISM requires more support than the very simple DESCARTE study. Our approach will be to target research practices for PRISM, with local nurses and the RA in each centre providing support to groups of practices, i.e. we will be targeting particular practices to perform PRISM. In reality it will not be 'either' PRISM 'or' DESCARTE: there will be no competition between the studies since we are using the same baseline clinical proforma for both studies. Thus although the operation and data management of the studies will be completely independent, the PRISM data can contribute data to the data set for DESCARTE and the PRISM patient information leaflet will consent patients to the documentation of adverse events (i.e. the main outcome in DESCARTE; we would have consented patients to this anyway even if DESCARTE had not been funded).

Trial management

The Trial management group will meet 2 monthly initially, then 6 monthly if progress is good, or more often as needed. The study team will pay close attention to both recruitment retention and performance of GPs/practices.

Some GPs will certainly stop recruiting; as with all our studies we will maintain recruitment of GPs throughout the study period.[33,47,49]

Underperformance will be dealt with initially by letters from each regional co-ordinator to GPs, then visits to GPs (from the local champions in each Network, the overall study co-ordinator, and applicants and PI as

necessary). How the problem is dealt with will depend on the particular issues raised e.g. clarity of the proforma or how to document clinical features (which can be clarified), how most efficiently to recruit (examples of good practice can be provided from the other Networks) etc. If performance in one year is poor and remains poor and cannot be improved in the last resort we will transfer the funds to the better performing practices in addition to continuing to recruit practices.

The new UKPCRN arrangements will also hopefully help in managing Network performance.

Project timetable

Oct 2006-Mar 2007: Recruit practices, and feasibility phase (pilot, train nurses; obtain ethics and RM+G approval for all sites; perform qualitative work to explore patients perceptions);

Feb 2007-Dec 2007: Recruit patients for Phase I (1 winter);

Jun 2007-Aug 2007: Prepare for Phase II (including completing the economic modelling exercise and pilot recruitment);

Aug 2007-Dec 2007: Phase II pilot recruitment;

Jan 2008-Apr 2010: Recruit patients;

Jan 2010-Jun 2010: Follow-up, notes searching and data cleaning, development of vignette study and agreement from GPs to participate;

Jun – Sep 2010: Finalise data collection/analysis report writing, vignettes sent to GPs.

Consumers

There is no national consumer group associated with the management of sore throat, but we will invite 2 lay members from one of our practices to join the trial management group (as we have done in a similar validation study funded by the NCCHTA of urinary dipsticks). The perspective of consumers will also be explored during the qualitative phase, and their perceptions incorporated into the trial materials and procedures.

Justification of Support. We need:

- staff: 1 trial manager is needed for 4 years; the trial manager will have overall responsibility for day to day running of the trial and will be supported by a part time secretary who will also manage the data bases; the model of the trial manager also running one centre with secretarial support has worked in our multicentre MRC ATEAM trial; an RA is needed in the Birmingham and Oxford for 3 years, supported by P/t secretarial staff; lab technicians for 1 year are needed to perform the in vitro studies and support microbiology; 12 months of higher level RAs for economic analysis, and 1 year of P/T RA for the judgement analysis study are needed;
- microbiology (throat swabs and antibody titres);
- GP and nurse panel time: these costs may be negotiable as part of support for science;
- data management (web based randomisation service; data entry);
- support for recruitment (£40 per patient research costs to allow booking of time in advance): we may be able to negotiate these costs as part of support for science;
- support for transcribing qualitative data;
- equipment (computers, tempadots);
- stationery for outcomes and also routine; also postage and telephone;
- travel to practices, for meetings;
- qualitative consultancy (Dr Leydon); statistical consultancy (Dr Mullee).

© Queen's Printer and Controller of HMSO 2014. This work was produced by Little *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

Published by the NIHR Journals Library