Project title: PRImary care Streptococcal Management study (PRISM) (05/10/01) Rapid tests for streptococcal sore throat) protocol (version 4 10/09/08)

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[1], which is supported by recent ecological data.[2] 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[3]. 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 [4]:

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[4]).

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.[4 5 6;7] 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[8].

Which RADT and how to assess them? The MHRA[9] recently identified five RADTs marketed in the UK. Evidence from the previous literature suggests:

- **accuracy of RADTs** has mostly been evaluated in microbiological labs[9](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[10;11]; 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[12]; furthermore, the agreement of a test versus standard is not likely to exceed the comparison of the standard versus itself[13] – 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[14-16];

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[17];

- **comparison of ease of use** by practitioners have rarely been performed.[17]

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, clarity of reading),

Clinical scoring methods. Existing clinical scores also have most promise to be useful in practice [4 5 6] - the major candidate being the 'Centor' criteria, which has been operationalised in two recent primary care trials[18 22] 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[18 22]. 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.[18] 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[6;7])with a similar population used univariate analysis only and thus was over-inclusive);
- the criteria very probably have low specificity in primary care populations – 44% in a recent Canadian study[7] which would result in rather high rates of overall antibiotic use (46% of adults).[7]

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 5 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
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;

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 or less; and with some abnormality of examination of the throat – i.e. erythema and or pus - as in our previous studies in primary care[19]). Although there some evidence that those presenting acutely are more likely to have bacterial infection[20] and more likely to benefit from antibiotics it is important that the performance and use of tests reflect the generaliseable population presenting in primary care. Exclusion: other non infective causes of sore throat (e.g. apthous 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[9;21;22];
- one swab can also be used for more than one plate – providing identical numbers of colonies on up to 5 plates [9;23];
- two double swabs can be done in adults.[14]

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[9]), and also for one rapid test. Our piloting in 60 patients has also confirmed the 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 2x2 tables comparing RADTs with the results of the throat swab as the reference standard, calculating sensitivity, specificity, predictive values and likelihood ratios.

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[9] and the recent French Agency assessment[17]) b) ease of use[17] 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[9] (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 [17 9] - which superseded the Signify test and is now marketed by Unipath as IMI Test Pack 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[17]. 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) [17] and sensitivity. [17 9]
As the Strep A OIA Max test is not designed as a 'point of care' test it will be replaced with the Clearview Exact Test (Unipath).

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

	*Sensitivity (compared to throat swab) [9]	In vitro studies % detection of low bacterial counts (10^5 cfu/ml) [17]	Ease of use score (maximum 38) [17]
Streptatest	96% (from company)	75%	35
OSOM Ultra	91%	no data[17] (from company website: Mass. General Hosp. study =equivalent to Quickvue)	no data[17] (similar tests averaged 30+)
Clearview Exact	**95%**	**99%**	**No data**
Quickvue	87%	50%	34
Test Pack 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[24] than reported in the review[9]

Sample size. Most RADTs are very specific[9]. Sensitivity is the limiting factor, and the sensitivity from better performing previous studies is in the range of 80-90%.[9] 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,1460- 3920 allowing for the 5 RADTs, or 1500-4000 allowing for some leeway in the assumptions. If each adult provides four comparisons and each child two, then 400-1200 patients are required. Thus our minimum sample size is 438 and maximum 1176.

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)	3920	2780	1460
Number of patients required (assuming two double swabs in adults and one double swab in children)	1176	834	438

2) Clinical rule and clinical outcomes.

- **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[6]); 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
Clinical measures. Baseline proforma. In order to develop a clinical rule to predict bacterial infection, clinical features need to be compared with the results of the throat swab. GPs/Nurses will fill in the clinical details at baseline. This will consist of a single clinical sheet documenting baseline clinical data. We will collect data at presentation on temperature (using tempadot 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 4

point Likert scales (none, a slight problem, a moderately bad problem, a bad problem), and the presence of signs (pus, nodes, tender nodes, temperature) based on previous clinical scores[25 6 26 18]. The severity of symptoms not simply their presence is probably important: in piloting among 60 patients we have shown that the severity of symptoms rather than the presence of symptoms predicts bacterial infection which we documented by a four fold rise in antibody titres.

- **Other clinical outcomes.**

Symptom diary. Patients will fill a validated symptom dairy (as in phase 2). As in previous studies patients will be contacted when the research centre has received the consent form to check there are no problems filling in diaries.

If no diary is received after 2 - 3 weeks, one mailed reminder will be sent using a brief questionnaire containing the key diary items; if no diary or questionnaire has been returned a brief phone call will clarify the duration and severity of symptoms and whether antibiotic were used; this method has been shown to be both acceptable to patients and allows significantly less bias from low response rates[27].

Notes review. Patients will be asked to fill in a FREEPOST card to return documenting reconsultation (for worsening of symptoms, presentation with new symptoms and/or complications); this data will also be assessed by a review of the GP's notes. Although notes will be the main way of determining adverse outcomes these cards will provide additional information since sometimes patients get worse (after they have stopped filling in their diaries) and or adverse events do not get recorded in notes. We will relate whether adverse outcomes occur to the presence or absence of Streptococcus, clinical presentation and management

These clinical outcomes will provide information to confirm the sample size calculation for the phase 2 trial, and will allow exploration of the predictive value of throat swabs/rapid tests and clinical data for symptom resolution and reconsultation/complications.

3) In vitro study of RADTs.

The number of organisms harvested crucially determine the performance of the RADT[12], and two well taken samples from the same individual will predict the other results with a sensitivity of 83%-91%.[14-16] 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[17], and performed very few tests for each RADT. Thus more comprehensive comparative data is needed.

As in the most recent in vitro study[17]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 [28] to clarify patients' and primary care professionals' understanding and concerns about the use of RADTs in both phase 1 and 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 . These will be the estimates of sensitivity and specificity of RADTs and the Centor criteria (or modified criteria) compared to throat swab; ease of use of RADTs; a qualitative model to understand patients' and GPs perceptions and to inform strategies for phase 2; estimates and predictors of symptom resolution and adverse clinical outcomes.

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:

a) **RADT using the best RADT from phase 1.** Depending on phase I results, 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. However we anticipate that it is likely that antibiotics 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.[7]

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.[27] 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[29] and recent data in lower RTIs[30]). It has been incorporated widely into routine practice in the UK since our 1997 trial[27] without any increase in complications of sore throat.[31]

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[9], 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 (ie. 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[32-34] and a lifestyle intervention trial.[35] The manualised approach maintains clear group differentiation despite individuals being randomised to different groups by the same health professional[32-35] and the data suggest that using such an approach the health professional and practice cluster effects are minimal.[32;36;37]

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 aged 3 years an over 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[19]). Most patients present within 5 days, but a smaller minority present with a longer duration of illness prior to seeing the doctor.[27] Since we wish this sample to be representative of those patients presenting to GPs[27], and prior duration predicts subsequent illness duration[38] (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, 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 [25 6 26] as in phase 1. The GPs will also be asked to rate the sore throat as Viral/bacterial.
Diary scores.[19;27] 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 phone 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[27;30;39] and are also more sensitive to change than criterion measures.[40] Temperature will be taken by patients and documented on a daily basis in the diary using tempadot thermometers as in our previous studies.[19 39]
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)[30], antibiotic use, and use of over the counter medicines (also see notes review)
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, supplemented by 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[3-5]. 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).
The medicalisation of illness. Patients' belief in the importance of seeing the doctor will be documented using 5 point Likert scales completed by patients[27] which we have shown to be reliable.[27] We will also document patients reconsultation behaviour by blinded notes review[19] (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 post code 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, side effects, economic data (see below) and any subsequent referrals. [19]

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) [27] 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[27] - then 254 patients per group are needed or 849 in total allowing for 10% loss to follow-up of notes.[39] 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 and beta=0.1), 242 per group, or 909 patients in total allowing for 20% loss to follow-up of diary information.[27;30] 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.[30] 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, and what effect size would be regarded as equivalent. 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[19] (which will involve 'censored' data due to variable follow-up time available). 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 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 specify *a priori* key subgroups as potential effect modifiers based on an updated literature review prior to analysis (currently we specify baseline symptom severity, and GPs' view of whether the infection is viral or bacterial). 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 proposed 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. 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.

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 effects	?	?	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. Thus 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 year) 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 of costs may prevail on long term costs.

The economic analysis will provide:

i) 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

ii) a systematic review of the literature on costs effectiveness of reduced antibiotic prescribing which would be included in the model 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. [19 39] 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 generaliseability of the results. Where unit costs are lacking they will be based on gross staff costs. 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. 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 [41] 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 [42]presenting combinations of clinical characteristics (cue profiles), including RADT results and/or clinical scores.
The Social Judgment Theory advocates a 'representative design' [43], 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 judgment 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[44] . 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[19], 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.[29;31]

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.[27;30]
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.[27;30;39]
Underperformance will be dealt with initially by letters from each regional coordinator to GPs, then visits to GPs (from the local champions in each Network, the overall study coordinator, and applicants and PI as necessary). How the problem is dealt with will depend on the particular issues rasied 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-March 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 – August 2007 prepare for phase II (including completing the economic modelling exercise and pilot recruitment);
August 2007-December 2007 Phase II pilot recruitment;
Jan 2008-April 2010 recruit patients

Jan 2010-June2010 follow-up, notes searching and data cleaning, development of vignette study and agreement from GPs to participate
June – September 2010 finalise data collection/analysis report writing, vignettes
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 3.3 years, supported by P/t secretarial staff ; lab technicians for 1 yr are needed to perform the in vitro studies and support microbiology ; 12 months of higher level RAs for economic analysis [1], and 1 year of P/T RA for the judgement analysis study are needed;
- microbiolology (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);
- support for transcribing qualitative data ;
- equipment (computers, tempadots);
- stationery for outcomes and also routine; also postage and phone;
- travel to practices, for meetings;
- qualitative consultancy (Dr Leydon); statistical consultancy (Dr Mullee).

Reference List

(1) Del Mar C, Glaziou P. Antibiotics for the symptoms and complications of sore throat. In: The Cochrane Data base of systematic reviews. The Cochrane Library 1998.

(2) 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; 52:187-193.

(3) 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; 26:268-274.

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

(5) Dagnelie C, Bartelink M, Van Der Graaf Y, Goessens W, De Melker R. Towards better diagnosis of throat infections with GABHS in general practice. B J G P 1998; 48:959-962.

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

(7) 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; 291:1587-1595.

(8) Kolmos H, Little P. Should general practitioners perform diagnostic tests on patients before prescribing antibiotics? BMJ 1999; 318:799-802.

[1] The economic analysis would be carried out under the supervision of JR, with the help of the team in WIHRD/SHTAC who do such work for NICE. Whether this work would rely on hiring an RA (which would be actively explored) or by contracting with existing staff is left open.

(9) Dean L, Perry K. Group A streptococcus rapid antigen detection kits - a review of the evaluation literature. 04123 ed. London: DOH; 2005.

(10) 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; 105:586-591.

(11) Hobbs R, Delaney B, Fitzmaurice D. A Systematic Review of Near Patient Testing in Primary Care. NHS National R+D Health Technology Assessment Programme 96 A.D..

(12) Schwartz R. Evaluation of rapid streptococcal detection tests. Ped Inf Dis J 1997; 16:1099-1100.

(13) 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. B J G P 2000; 50:43-45.

(14) 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; 7:765-769.

(15) Battle C, Glasgow L. Reliability of bacteriologic identification of beta-hemolytic streptococci in private offices. Am J Dis Child 1971; 122:134-136.

(16) Breese B, Disney F. The accuracy of diagnosis of beta streptococcal infections on clinical grounds. J Pediatr 1954; 44:670-673.

(17) 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; 52:438-443.

(18) 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. B M J 2000; 320:150-154.

(19) 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. B M J 1997; 315:350-352.

(20) 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. B M J 1999; 319:736-737.

(21) 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. Ped Inf Dis J 1987; 6:762-764.

(22) 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-e670.

(23) Anhalt J, Heiter B, Naumovitz D, Bourbeau P. Comparison of three methods for detection of Group A Streptococci in throat swabs. J Clin Micro 1992; 30:2135-2138.

(24) Nerbrand C, Jasir A, Schalen C. Are current rapid detection tests for Group A Streptococci sensitive enough? Evaluation of 2 commercial kits. Scand J Infectious Dis 2002; 34:797-799.

(25) Centor RM, Witherspoon JM, Dalton HP. The diagnosis of strep throat in the emergency room. Med Decis Making 1981; 1:239-246.

(26) Breese B. A simple scorecard for the tentative diagnosis of streptococcal pharyngitis. Am J Dis Child 1977; 131:514-517.

(27) Little PS, Williamson I, Warner G, Gould C, Gantley M, Kinmonth AL. An open randomised trial of prescribing strategies for sore throat. B M J 1997; 314:722-727.

(28) Glaser B, Strauss A. The Discovery of grounded theory: strategies for qualitative research. New York: Aldine; 1967.

(29) Little P. Delayed prescribing of antibiotics for upper respiratory tract infection. BMJ 2005; 331:301-302.

(30) 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; 293:3029-3035.

(31) 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; 331:328-329.

(32) 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; 54:189-195.

(33) Yardley L, Beech S, Zander L, Evans T, Weinman J. A randomized controlled trial of exercise therapy for dizziness and vertigo in primary care. B J G P 1998; 48:1434-1435.

(34) 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.

(35) 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; 328:1054-1060.

(36) Little PS, Williamson I, Warner G, Gould C, Kinmonth AL, Gantley M. An open randomised trial of prescribing strategies for sore throat. B M J 1997; 314:722-727.

(37) 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; 322:336-342.

(38) 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; 319:736-737.

(39) 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; 322:336-342.

(40) Watson L, Little P, Williamson I, Moore M, Warner G. Validation study of a diary for use in acute lower respiratory tract infection. Family Practice 2001; 18:553-554.

(41) Wigton R. Social Judgment Theory and Medical Judgment. Thinking and Reasoning 1996; 2 (2/3):175-190.

(42) Cooksey R. The methodology of Social Judgment Theory. Thinking and Reasoning 1996; 2(2/3):141-173.

(43) Goldstein W. Social Judgment Theory: Applying and Extending Brunswik's Probabilistic Functionalism. In: Koehler D, Harvey N, editors. Handbook of Judgment & Decision Making. 1 ed. Oxford: Blackwell; 2004. 37-61.

(44) 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; 18:149-182.