

# Patient–clinician agreement on signs and symptoms of ‘strep throat’: a MetroNet study

Jinping Xu, Kendra Schwartz, Joseph Monsur, Justin Northrup and Anne Victoria Neale

Xu J, Schwartz K, Monsur J, Northrup J and Neale AV. Patient–clinician agreement on signs and symptoms of ‘strep throat’: a MetroNet study. *Family Practice* 2004; **21**: 6pp.

**Background.** Despite substantial use of the telephone in health care, only a few studies have formally evaluated the appropriateness of telephone-based management for acute medical problems. The accuracy of patients’ report of signs and symptoms remains unknown.

**Objective.** We compared the agreement between patient self-assessment and clinician assessment on the typical signs and symptoms of group A  $\beta$ -haemolytic *Streptococcus* (GABHS) to investigate the potential difficulties of using patient self-report to triage sore throat patients.

**Methods.** In this cross-sectional study, each of 200 adult pharyngitis patients was instructed to examine him/herself and to record the symptoms and physical findings. Two clinicians independently interviewed and examined each patient and recorded their findings. Each patient then had a rapid GABHS antigen test, the results of which were blinded to both clinicians and patients. Each patient self-assessment was compared with the findings of each clinician, and the agreement and disagreement between them computed.

**Results.** We found varying levels of agreement ( $\kappa = -0.05$  to  $0.71$ ) between patients and clinicians on sore throat history and physical assessments. Importantly, there was fair to substantial agreement ( $\kappa = 0.20$ – $0.71$ ) on the key signs and symptoms used in GABHS clinical prediction rules. As expected, history items had the highest agreement ( $\kappa = 0.52$ – $0.71$ ). Patients were more likely than clinicians to report rather than deny a specific physical sign.

**Conclusion.** Adult sore throat patients may reliably report their symptoms, but may not be able to assess and report accurately on relevant physical signs of pharyngitis. Patients have a tendency to over-report physical signs. This study indicates the potential difficulties associated with telephone triage of sore throat patients, or other illnesses that require assessment of physical signs.

**Keywords.** Group A  $\beta$ -haemolytic *Streptococcus*, interobserver agreement, kappa coefficient, pharyngitis, sore throat, telephone triage.

## Introduction

Over 40 million adults with sore throat present annually at physician offices, urgent care centres and emergency departments,<sup>1</sup> accounting for 2.1% of ambulatory visits in the USA. After cough, sore throat is the second most common symptom for seeking care.<sup>2</sup> In healthy adult

patients, most sore throat complaints are due to uncomplicated and self-limited viral infections,<sup>2</sup> and are usually manageable by supportive measures for symptomatic relief. Only a small percentage (5–17%) of adult pharyngitis is caused by group A  $\beta$ -haemolytic *Streptococcus* (GABHS),<sup>2,3</sup> an organism justifying antibiotic treatment. Strategies for diagnosis and treatment of sore throat are directed primarily at identifying patients who require antibiotic therapy, and avoiding unnecessary and potentially deleterious treatment of those with viral infections.

Several clinical prediction rules have been validated to help clinicians assess the probability of GABHS throat infection.<sup>3,4</sup> Use of these prediction rules

---

Received 30 June 2004; Accepted 20 July 2004.

Department of Family Medicine, Wayne State University, Detroit, MI, USA. Correspondence to Jinping Xu, MD, MS, Division of Practice-based Research, Department of Family Medicine, Wayne State University, 101 E. Alexandrine, Detroit, MI 48201, USA; Email: jxu@med.wayne.edu

improves GABHS diagnosis and decreases unnecessary antibiotic use. Few studies attempt to triage sore throat patients based on self-reported symptoms. Clancy *et al.*<sup>5</sup> demonstrated that adult sore throat patients could be stratified into higher and lower risks of strep throat by history alone. Primary care physicians often make preliminary evaluations based solely on patient report of signs and symptoms. This is especially true when patients telephone their physicians reporting symptoms such as a sore throat. The use of the telephone to deliver health care advice has increased considerably in recent years due to its improved access and convenience.<sup>6</sup> Telephone calls account for ~25% of all patient encounters in general practice in the USA.<sup>7</sup> Approximately 45% of calls are symptom-related and nearly half of these problems are managed exclusively over the telephone.<sup>8</sup> Despite the frequency of telephone contacts, the accuracy of patients' report of signs and symptoms remains unknown. We found no studies in the literature evaluating either satisfaction or clinical outcomes with telephone management of sore throat complaints. One study of general internists' attitudes toward, and practices associated with, telephone management of upper respiratory tract infection reported that only 40% of respondents were able to make important diagnostic distinctions about upper respiratory tract infections from a written vignette.<sup>9</sup>

In an earlier report from our study, Schwartz *et al.*<sup>10</sup> addressed the variability in clinician assessment of sore throat signs and symptoms, and noted that clinician agreement enhances the utility of pharyngitis clinical prediction rules. However, there are no studies on agreement between patients and clinicians regarding the key history and physical elements shared by these prediction rules (namely cough, fever, exudates and cervical adenopathy). This information is essential to assess the appropriate use of telephone triage with sore throat patients. If shown to be feasible and accurate, such telephone triage could be effective in reducing some unnecessary office visits.

We report on a study in which sore throat patients reported their signs and symptoms and also were independently assessed by two clinicians to determine agreement on certain history and physical elements. The study objective was to determine the agreement of patient self-assessment with clinician assessment. If patient-clinician agreement is high, a study of telephone triage of adult sore throat patients might be feasible.

## Methods

### *Setting and data collection*

Eight family practice centres and 116 medical professionals from central and eastern Michigan participated in this cross-sectional study. Clinicians and staff made an effort to recruit all patients age 18 and

older presenting with a 'sore throat' or 'strep throat' complaint during two sore throat seasons (December 1999 to April 2000 and September 2000 to March 2001). The Wayne State University Human Investigation Committee and other Human Subject Protection Committees as required by each individual family practice centre approved the protocol.

Each patient completed an assessment form that included history, physical and demographic items while s/he was waiting in the examination room to be seen by clinicians. The patient assessment form used lay language to define each sign and symptom. For example, we used the question 'Are the tonsils (a bump on each side of your throat behind your tongue) swollen or larger than usual?' to assess if swollen or enlarged tonsils were present. We used 'white spots on tonsils' instead of tonsillar exudate. To assist them in viewing their posterior pharynx, patients were provided with a simplified drawing of the oral pharynx anatomy, flashlights and mirrors, and asked to record their perception of redness, the presence of white spots and whether tonsils appeared swollen.

Each patient was then interviewed and examined independently by two clinicians because we were also interested in determining the inter-rater variability in clinician assessment of sore throat signs and symptoms.<sup>10</sup> A rapid antigen GABHS test was then performed; these results were for research purposes only and were not provided to clinicians or patients. If the clinician felt a diagnostic test was needed, the usual protocol of the office was followed, but the study test result remained confidential.

The QuickVue In-Line One-Step Strep A test (Quidel, San Diego, CA) was used for rapid detection of GABHS antigen. The manufacturer reports a sensitivity of 87% and specificity of 94% for the Quick Vue test compared with standard culture methods.

### *Data analysis*

*Agreement between patients and clinicians.* Patient-clinician agreement was estimated with the kappa coefficient, which corrects for chance agreement. Separate kappa statistics were computed to determine agreement on each history or physical exam item between the two dyads (patient-clinician 1 and patient-clinician 2). As general guidelines, a kappa of <0 indicates poor agreement, 0–0.2 represents slight agreement, 0.2–0.4 is fair agreement, 0.4–0.6 indicates moderate agreement, 0.6–0.8 shows substantial agreement, and 0.8–1.00 is almost perfect agreement.<sup>11</sup>

*Disagreement between patients and clinicians.* The patient-clinician disagreement was analysed using the McNemar statistic, a chi-square test of paired proportions based on the number of disagreements. This test indicates whether the disagreement is equally likely one way or the other (i.e. disagreement could take either of two forms: patient report of the presence of a symptom or sign and

clinician report of the absence of the same symptom or sign, defined here as patient ‘over-reporting’, or clinician report of symptom or sign and patient denial of symptom or sign, defined as patient ‘under-reporting’).

## Results

Two hundred patients presenting with a complaint of strep throat or sore throat participated in this study. Twenty-two subjects (11%) had a positive rapid GABHS antigen strep test, a prevalence that falls in the middle of the expected range (5–17%).<sup>2,3</sup> The majority of subjects were female (76%), educated beyond high school (67%), and with a mean age of 34 years (age ranged from 18 to 72 years). Thirty-nine percent were African-Americans.

### Agreement between patients and clinicians

Each patient’s self-assessment data were compared with each of the two clinicians’ assessment. Observed

agreement ( $P_o$ ), agreement expected by chance ( $P_e$ ), kappa coefficient and McNemar  $P$ -values are summarized in Table 1 for the patient–clinician 1 dyad, and in Table 2 for the patient–clinician 2 dyad. The results were similar between the two dyads (patient–clinician 1 and patient–clinician 2). The percentage of observed agreement between patient and clinician on history and physical findings ranged from 57 to 86%, with associated kappa coefficients between  $-0.05$  and  $0.71$ , respectively. There was fair to substantial agreement ( $\kappa = 0.20$ – $0.71$ ) for the four history and physical variables (history of cough, history of fever, tonsillar exudate and anterior cervical adenopathy) found in prospectively validated clinical prediction rules.<sup>3</sup> Both history elements (i.e. history of cough and history of fever) had moderate to substantial agreement ( $\kappa = 0.52$ – $0.71$ ) and the physical signs had fair (tonsillar exudate,  $\kappa = 0.33$  for dyad 1 and  $0.20$  for dyad 2) to moderate agreement (tender anterior cervical nodes,  $\kappa = 0.46$  for dyad 1 and  $0.38$  for dyad 2).

TABLE 1 Agreement of patient–clinician 1 dyad on sore throat signs and symptoms

Assessment variable	$n^a$	Patient reporting presence of variable (%)	$P_o$	$P_e$	$\kappa$ (95% CI)	Odds ratio ( $P$ -value)
History of cough <sup>b</sup>	174	102 (58.6)	0.86	0.53	0.71 (0.60 to 0.81)	0.14 ( $P < 0.001$ )
History of fever <sup>b</sup>	178	136 (76.4)	0.79	0.56	0.52 (0.39 to 0.64)	6.60 ( $P < 0.001$ )
Tender anterior cervical nodes <sup>b</sup>	194	103 (53.1)	0.73	0.49	0.46 (0.35 to 0.58)	4.89 ( $P < 0.001$ )
Tonsillar exudate <sup>b</sup>	181	55 (30.1)	0.75	0.63	0.33 (0.19 to 0.48)	3.09 ( $P = 0.001$ )
Enlarged anterior cervical nodes	191	124 (65)	0.61	0.47	0.26 (0.15 to 0.38)	4.77 ( $P < 0.001$ )
Enlarged tonsils	167	82 (49.1)	0.60	0.50	0.20 (0.07 to 0.34)	2.88 ( $P < 0.001$ )
Tender posterior cervical nodes	192	61 (31.8)	0.70	0.67	0.10 (0.00 to 0.20)	18.33 ( $P < 0.001$ )
Pharyngeal erythema	190	167 (87.9)	0.75	0.73	0.09 ( $-0.06$ to $0.25$ )	1.94 ( $P = 0.040$ )
Enlarged posterior cervical nodes	190	80 (42.1)	0.59	0.57	0.05 ( $-0.02$ to $0.12$ )	37.5 ( $P < 0.001$ )
Posterior pharyngeal exudate	185	52 (28.1)	0.66	0.68	$-0.05$ ( $-0.15$ to $0.05$ )	3.77 ( $P < 0.001$ )

The formulae for calculations of  $P_o$ ,  $P_e$  and Odds ratio are based on the  $2 \times 2$  table below.

Patient	Clinician		
	Yes	No	
Yes	a	b	$g1$
No	c	d	$g2$
Totals	$f1$	$f2$	$n$

$P_o$  = proportion of observed agreement;  $(a + d)/n$ .

$P_e$  = proportion of agreement expected by chance;  $(f1g1 + f2g2)/n^2$ .

Odds ratio =  $b/c$ ; if the ratio is  $>1.0$ , it indicates patients over-reporting the assessment variable; if the ratio is  $<1.0$ , it indicates patients under-reporting the variable; if the ratio =  $1.0$ , it indicates patients reporting without a consistent pattern.

<sup>a</sup> Numbers reflect complete data for the dyads on each assessment variable.

<sup>b</sup> Component in GABHS clinical prediction rules<sup>3</sup>.

TABLE 2 Agreement of patient–clinician 2 dyad on sore throat signs and symptoms

Assessment variable	<i>n</i> <sup>a</sup>	Patient reporting presence of variable (%)	<i>P</i> <sub>o</sub>	<i>P</i> <sub>e</sub>	$\kappa$ (95% CI)	Odds ratio ( <i>P</i> -value)
History of cough <sup>b</sup>	175	99 (56.6)	0.81	0.52	0.61 (0.49 to 0.73)	0.50 ( <i>P</i> = 0.080)
History of fever <sup>b</sup>	177	138 (78)	0.81	0.55	0.57 (0.46 to 0.69)	$\infty^c$ ( <i>P</i> < 0.001)
Tender anterior cervical nodes <sup>b</sup>	192	102 (53.1)	0.68	0.49	0.38 (0.25 to 0.50)	3.07 ( <i>P</i> < 0.001)
Tonsillar exudate <sup>b</sup>	179	54 (30.2)	0.73	0.66	0.20 (0.07 to 0.34)	6.00 ( <i>P</i> < 0.001)
Enlarged anterior cervical nodes	189	124 (65.6)	0.57	0.49	0.15 (0.01 to 0.28)	2.42 ( <i>P</i> < 0.001)
Enlarged tonsils	164	82 (50)	0.60	0.50	0.20 (0.07 to 0.32)	5.00 ( <i>P</i> < 0.001)
Tender posterior cervical nodes	189	60 (31.7)	0.69	0.66	0.09 (−0.02 to 0.19)	13.50 ( <i>P</i> < 0.001)
Pharyngeal erythema	189	167 (88.4)	0.79	0.72	0.27 (0.11 to 0.44)	2.90 ( <i>P</i> = 0.003)
Enlarged posterior cervical nodes	187	78 (41.7)	0.59	0.57	−0.01 (−0.09 to 0.07)	9.13 ( <i>P</i> < 0.001)
Posterior pharyngeal exudate	185	53 (28.6)	0.68	0.69	−0.03 (−0.12 to 0.06)	6.38 ( <i>P</i> < 0.001)

The formulae for calculations of *P*<sub>o</sub>, *P*<sub>e</sub> and Odds ratio are based on the table below.

Patient	Clinician		
	Yes	No	
Yes	a	b	<i>g</i> <sub>1</sub>
No	c	d	<i>g</i> <sub>2</sub>
Totals	<i>f</i> <sub>1</sub>	<i>f</i> <sub>2</sub>	<i>n</i>

*P*<sub>o</sub> = proportion of observed agreement; (a + d)/*n*.

*P*<sub>e</sub> = proportion of agreement expected by chance; (*f*<sub>1</sub>*g*<sub>1</sub> + *f*<sub>2</sub>*g*<sub>2</sub>)/*n*<sup>2</sup>.

Odds ratio = b/c; if the ratio is >1.0, it indicates patients over-reporting the assessment variable; if the ratio is <1.0, it indicates patients under-reporting the variable; if the ratio = 1.0, it indicates patients reporting without a consistent pattern.

<sup>a</sup> Numbers reflect complete data for the dyads on each assessment variable.

<sup>b</sup> Component in GABHS clinical prediction rules<sup>3</sup>.

<sup>c</sup> 34/0 (i.e. b = 34, c = 0).

### Disagreement between patients and clinicians

The McNemar test was used to analyse the disagreement between patient and clinician. The McNemar statistic revealed that when there was disagreement between patient and clinician, the disagreement was more likely to take the form of the patient reporting the presence of a physical sign, while the clinician reported an absence of the same sign. For example, there were 49 cases of disagreement in the patient–clinician 2 dyad regarding the presence of tonsillar exudate. In 42 of these 49 disagreements, the patient reported the presence of exudate, while the clinician did not find exudate. In the remaining seven cases, the patients denied exudate while the clinicians reported exudate present. This pattern of patient ‘over-reporting’ was found for all the eight clinical physical signs assessed (Tables 1 and 2).

## Discussion

We found inconsistent agreement between patient self-assessment and clinician assessment on symptoms and

signs in adult patients with sore throats. Using the clinician assessment as the gold standard, patients accurately reported their symptoms, such as presence or absence of cough or fever. However, they less accurately assessed and reported relevant physical signs of strep pharyngitis, such as tonsillar exudate, cervical adenopathy and pharyngeal erythema. These findings are probably due to the very nature of these physical signs. Although a clear and simplified drawing of the oral pharynx anatomy was provided with simple definitions of terms to assist patients in their self-examination, the distinction between normal mucosa, which are red, and abnormal redness may not be apparent. In two clinician inter-observer reliability studies, both Schwartz *et al.*<sup>10</sup> and Donner-Banzhoff *et al.*<sup>12</sup> found that practitioners agreement on physical findings related to sore throat ranged from no agreement to moderate agreement ( $\kappa$  = −0.03 to 0.53). Our findings on patient–clinician agreement highlight strengths and weaknesses in patients’ ability to conduct self-examination for sore throat. The kappa value of the key signs found in most

GABHS prediction rules (i.e. tonsillar exudate and anterior cervical adenopathy) were higher than other signs assessed in the study; however, they were still lower than the clinician–clinician agreement reported by Schwartz *et al.*<sup>10</sup> As expected, the kappa values of history items were very similar to clinician–clinician agreement, as these are more patient-dependent.

Using the clinician assessment as the gold standard, patients had a tendency to over-report sore throat physical signs when patient and clinician disagreed across all the eight signs assessed in the study. There are plausible explanations for patient overestimation of signs. Patient concern or anxiety about their symptoms may lead to over-reporting if doubt is present about a specific physical sign. Another possibility is that patients may exaggerate their illness in order to obtain antibiotics. This finding indicates the potential difficulties associated with telephone triage of sore throat patients, or any other illness that requires assessment of physical signs.

Increased use of the telephone in health care represents just one of the developments in information technology that are likely to alter the way in which clinical care is delivered in the future.<sup>13</sup> Telephone consultations are found to be safe alternatives in the triage of requests for same day appointments and after-hours care.<sup>14,15</sup> Patients' satisfaction with telephone consultation is high.<sup>14,15</sup> It is often stated that 80% of diagnoses are made from the history,<sup>6</sup> and one might expect that an appreciable proportion of consultations could take place by telephone. This could help patients, who save travel time and costs and do not need to arrange childcare, or miss work, even if it may not save time for health professionals.<sup>6</sup>

Telephone management may be appropriate for many acute disorders, e.g. respiratory tract infections, musculoskeletal problems, such as back pain, and common symptoms such as headache and fever.<sup>9,16</sup> However, only a few studies have formally evaluated the appropriateness of telephone-based management for an acute medical problem such as uncomplicated urinary tract infection (UTI) in women.<sup>17,18</sup> Such studies showed that telephone-based care is safe and cost-effective in management of uncomplicated UTI in women, a common illness diagnosed mainly by symptoms. For example, in women who present with one or more symptoms of UTI, the probability of infection is ~50%.<sup>19</sup> Specific combinations of symptoms (e.g. dysuria and frequency without vaginal discharge or irritation) raise the probability of UTI to >90%, effectively ruling in the diagnosis based on history alone.<sup>19</sup> In contrast, GABHS pharyngitis diagnosis in primary care adult patients is uncommon (~10%), and it requires not only symptoms but also signs obtained by physical examination, which makes it less appropriate for telephone-based management since it is difficult effectively to rule in or rule out the diagnosis of GABHS based on symptoms alone.

A key concern in telephone consultation is whether telephone management is appropriate. Mistakes in

telephone diagnosis and triage can have severe consequences. Both the benefits and the limits of telephone medicine compared with face to face consultation need to be determined, so that practitioners and patients can use it as effectively and safely as possible. Telephone communication clearly has limitations as a clinical diagnostic tool, including absence of visual clues and inability to confirm the diagnosis with a clinical examination. More research is needed to support the evidence-based practice of telemedicine.

One limitation to this study that may reduce its generalizability to other clinical situations is that all the patients were recruited in family practice offices affiliated with residency training programmes. The patient population may not be similar to those of non-training offices; however, the signs and symptoms of sore throat should be similar among two such populations. Another limitation when making inferences about telephone triage of sore throat patients is that the patient self-assessment data were obtained by self-report in the office instead of by telephone interview. There appears to be good agreement between telephone and face to face interviews in the psychiatric literature,<sup>20,21</sup> but few studies have compared the agreement between telephone interview and self-assessment of symptoms. Recently, Allen-Davis *et al.*<sup>22</sup> reported on a telephone triage study of vulvovaginal complaints and found that eliciting patients' sexual activity over the telephone yielded answers that were significantly different from those self-reported in the office. We agree with the author's speculation that there are some areas of medicine in general, and gynaecology in particular, wherein patients are reluctant to share information, especially the intimate or possibly embarrassing details. Although we cannot know from our study whether telephone-collected data would have differed significantly from in-office self-assessment, we believe that it is unlikely because of the nature of the questions related to sore throat diagnosis. Finally, in a low prevalence situation, such as strep throat in adults, kappa coefficients are known to underestimate the inter-observer agreement.<sup>10,23</sup>

In conclusion, we found that there is moderate to substantial agreement among patients and clinicians on the key history components, and fair to moderate agreement on the key physical examination components of validated sore throat clinical prediction rules. However, patients have a tendency to over-report sore throat physical findings, which indicates the potential difficulties associated with telephone triage of sore throat patients, or other illnesses for which the diagnosis is based in large part on the assessment of physical signs.

## Acknowledgements

This paper was presented at Michigan Family Practice Research Day XXV, East Lansing, Michigan, May 23, 2002.

## Declaration

**Funding:** This study was funded in part by the Michigan Consortium for Family Practice Research. MetroNet is a metropolitan Detroit practice-based research network.  
**Ethical approval:** This study had IRB approval.  
**Conflicts of interest:** None.

## References

- <sup>1</sup> Cypress B. *Office Visits for the Respiratory System*. Hyattsville (MD): US Department of Health, Education and Welfare; 1979.
- <sup>2</sup> Bisno AL. Acute pharyngitis. *N Engl J Med* 2001; **344**: 205–211.
- <sup>3</sup> Centor RM, Witherspoon JM, Dalton HP, Brody CE, Link K. The diagnosis of strep throat in adults in the emergency room. *Med Decis Making* 1981; **1**: 239–246.
- <sup>4</sup> McIsaac WJ, Goel V, To T, Low DE. The validity of a sore throat score in family practice. *Can Med Assoc J* 2000; **163**: 811–815.
- <sup>5</sup> Clancy CM, Centor RM, Campbell MS, Dalton HP. Rational decision making based on history: adult sore throats. *J Gen Intern Med* 1988; **3**: 213–217.
- <sup>6</sup> Toon PD. Using telephones in primary care. *Br Med J* 2002; **324**: 1230–1231.
- <sup>7</sup> *Medical Practice in the United States*. Princeton (NJ): Robert Wood Johnson Foundation; 1981.
- <sup>8</sup> Johnson BE, Johnson CA. Telephone medicine: a general internal medicine experience. *J Gen Intern Med* 1990; **5**: 234–239.
- <sup>9</sup> Jepson S, Holbrook JH, Hale D, Lyon J. Management of upper respiratory tract infections by telephone. *West J Med* 1994; **160**: 529–533.
- <sup>10</sup> Schwartz K, Monsur J, Northrup J, West P, Neale AV. Pharyngitis clinical prediction rules: effect of interobserver agreement: a MetroNet study. *J Clin Epidemiol* 2004; **57**: 142–146.
- <sup>11</sup> Stoler MH, Schiffman M. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *J Am Med Assoc* 2001; **285**: 1500–1505.
- <sup>12</sup> Donner-Banzhoff N, Beck C, Meyer F, Werner JA, Baum E. Clinical findings in patients presenting with sore throat. A study on inter-observer reliability. *Fam Pract* 2002; **19**: 466–468.
- <sup>13</sup> Car J, Sheikh A. Telephone consultations. *Br Med J* 2003; **326**: 966–969.
- <sup>14</sup> McKinstry B, Walker J, Campbell C, Heaney D, Wyke S. Telephone consultations to manage requests for same-day appointments: a randomised controlled trial in two practices. *Br J Gen Pract* 2002; **52**: 306–310.
- <sup>15</sup> Lattimer V, George S, Thompson F *et al*. Safety and effectiveness of nurse telephone consultation in out of hours primary care: randomised controlled trial. The South Wiltshire Out of Hours Project (SWOOP) Group. *Br Med J* 1998; **317**: 1054–1059.
- <sup>16</sup> Manniche C, Asmussen K, Lauritsen B, Vinterberg H, Kreiner S, Jordan A. Low Back Pain Rating Scale: validation of a tool for assessment of low back pain. *Pain* 1994; **57**: 317–326.
- <sup>17</sup> Barry HC, Hickner J, Ebell MH, Ettenhofer T. A randomized controlled trial of telephone management of suspected urinary tract infections in women. *J Fam Pract* 2001; **50**: 589–594.
- <sup>18</sup> Saint S, Scholes D, Fihn SD, Farrell RG, Stamm WE. The effectiveness of a clinical practice guideline for the management of presumed uncomplicated urinary tract infection in women. *Am J Med* 1999; **106**: 636–641.
- <sup>19</sup> Bent S, Nallamotheu BK, Simel DL, Fihn SD, Saint S. Does this woman have an acute uncomplicated urinary tract infection? *J Am Med Assoc* 2002; **287**: 2701–2710.
- <sup>20</sup> Revicki DA, Tohen M, Gyulai L *et al*. Telephone versus in-person clinical and health status assessment interviews in patients with bipolar disorder. *Harv Rev Psychiatry* 1997; **5**: 75–81.
- <sup>21</sup> Rohde P, Lewinsohn PM, Seeley JR. Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. *Am J Psychiatry* 1997; **154**: 1593–1598.
- <sup>22</sup> Allen-Davis JT, Beck A, Parker R, Ellis JL, Polley D. Assessment of vulvovaginal complaints: accuracy of telephone triage and in-office diagnosis. *Obstet Gynecol* 2002; **99**: 18–22.
- <sup>23</sup> Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990; **43**: 551–558.