BioMed Central

Research article

The inter-observer agreement of examining pre-school children with acute cough: a nested study

Alastair D Hay*[1], Andrew Wilson[2], Tom Fahey[3] and Tim J Peters[1]

Address: [1]Division of Primary Health Care, University of Bristol, Cotham House, Cotham Hill, Bristol, BS6, 6JL, UK, [2]Division of General Practice and Primary Health Care, University of Leicester, Leicester General Hospital, Gwendolen Road, Leicester, LE5 4PW, UK and [3]Tayside Centre for General Practice, University of Dundee, Kirsty Semple Way, Dundee, DD2 4AD, UK

Email: Alastair D Hay* - alastair.hay@bristol.ac.uk; Andrew Wilson - aw7@leicester.ac.uk; Tom Fahey - t.p.fahey@dundee.ac.uk; Tim J Peters - Tim.Peters@bristol.ac.uk

* Corresponding author

Abstract

Background: The presence of clinical signs have implications for diagnosis, prognosis and treatment. Therefore, the aim of this study was to examine the inter-observer agreement of clinical signs in pre-school children presenting to primary care.

Methods: A nested study comparing two clinical assessments within a prospective cohort of 256 pre-school children with acute cough recruited from eight general practices in Leicestershire, UK. We examined agreement (using kappa statistics) between unstandardised and standardised clinical assessments of tachypnoea, chest signs and fever.

Results: Kappa values were poor or fair for all clinical signs (range 0.12 to 0.39) with chest signs the most reliable.

Conclusions: Primary care clinicians should be aware that clinical signs may be unreliable when making diagnosis, prognosis and treatment decisions in pre-school children with cough. Future research should aim to further our understanding of how best to identify abnormal clinical signs.

Background

Cough is the most frequently managed problem in primary care and becomes increasingly common at the extremes of age [1,2]. Cough in pre-school children is usually due to simple, self limiting respiratory tract infection, but more severe causes need to be ruled out including pneumonia, bronchiolitis, pertussis, croup and asthma[2]. The presence of clinical signs may have diagnostic, prognostic, and treatment implications. The absence of tachypnoea has been shown to be most useful for ruling out pneumonia[3], and fever is associated with poor outcome in children with cough[4] and otitis media[5]. In a study of cough in adults, antibiotics were eight times more likely to be prescribed in patients with abnormal chest signs[6], and in another study 93% of adults presenting with the combination of cough and chest signs received antibiotics[7].

The reliability and accuracy of respiratory symptoms and signs have been assessed almost exclusively in secondary care[8], where relatively serious illness is more prevalent[9]. Given the diagnostic, prognostic and treatment implications of these clinical signs, we decided to examine the inter-observer agreement between a standardised and non-standardised clinical assessment in pre-school children presenting with acute cough in primary care. These were children already recruited to a cohort study investigating duration and complications of cough[4,10].

Methods

Practices and participants

Practice and participant recruitment have been described in detail elsewhere[10]. The Leicestershire Research Ethics Committee approved the study. To maximise the efficiency of child recruitment, practices with list sizes greater than 8000 were invited by letter to participate. Recruitment took place from November to April over two years between 1999 and 2001, at morning and evening surgeries rotated between practices. A researcher was located in the surgery during recruitment sessions to ensure all eligible children were invited to participate. These were children aged 0–4 years with a cough ≤ 28 days duration presenting to a General Practitioner (GP) or Nurse Practitioner (NP), without asthma (defined as recommended to be receiving preventive or regular reliever treatment) or any other chronic disease. Two observers examined each child.

Observer one

This was the GP or NP to whom the child presented. Our aim was not to alter the clinical assessment of observer one, but to ask the clinician to perform a routine, non-standardised, examination of the child. A standardised data collection sheet [see Additional file 1] included questions about respiratory rate, the presence of fever and chest signs, but only examined items were recorded. For respiratory rate and temperature, clinicians were asked to give a global opinion of abnormality. They were not required to count breaths per minute or use a thermometer, though they could record these data if they wished. Similarly, if the clinician auscultated the chest, they were able to record if abnormal signs (wheezes or crepitations) were present.

Observer two

This was one general practitioner (ADH), who performed a standardised clinical assessment within 30 minutes, before or after, observer one and was blind to the results of the other assessment. Data collected differed between children presenting in the first and second winters. In the first winter, we included a global assessment of the child's respiratory rate and auscultation of all respiratory zones of the chest. However, by the second winter, it became apparent that, in addition to the global assessment, we wanted a more accurate measure of temperature and respiratory rate [see Additional file 1]. We used a mercury thermometer placed in the axilla for five minutes and counted breaths over a 30 to 60 second period of settled behaviour[11].

Sample size

The sample size was determined by the primary research question, which was to quantify cough duration[10]. For this study, sample size is best considered through the precision attained in the agreement analyses as shown by the 95% confidence limits in Table 2.

Data entry and analysis

Data were single entered onto an Access database. No errors were found in 14 randomly selected cases. We used Stata version 7 to describe the clinical assessment data and generate chance adjusted (kappa) inter-observer agreement statistics[12]. Because kappa values decrease as the proportion of positive ratings become extreme, even when observers interpret signs consistently, we also calculated chance independent agreement values, or phi[13]. For the second winter data from observer two, the counted respiratory rates were converted into a binary variable using 40 breaths per minute as the upper limit of normal for children aged up to one year and 30 breaths per minute for children aged up to five years of age [14]. Similarly, measured temperatures were converted using an upper limit of normal of 37.5°C [11]. We did not compare the thermometer derived continuous measurements because of the small number of children in whom these data were available from both observers (23) and because we felt it was clinically more useful to dichotomise children into febrile or afebrile.

Table 1: Clinician and researcher clinical assessments

Variables	Observer one (un-standardised assessment)[a,c]	Observer two (standardised assessment)[c]
Breaths per minute counted	61/250 (24.4%)	95/116 (81.9%)[b]
Counted respiratory rate raised	15/61 (24.6%)	49/95 (51.6%)[b]
Raised respiratory rate (global opinion)	22/250 (8.8%)	12/218 (5.5%)[a]
Temperature recorded using thermometer	61/250 (24.4%)	103/116 (88.8%)[b]
Temperature recorded and raised (> 37.5°C)	6/61 (9.8%)	4/103 (3.9%)[b]
Fever (global opinion)	27/250 (10.8%)	Not examined.
Abnormal chest signs	53/246 (21.5%)	31/218[a] (14.2%)

[a] Data collected from both winters [b] Data collected on consecutive children for second winter only [c] Denominators vary due to missing data

Table 2: Inter-observer agreement values

Clinical sign	Number with complete data (%)	Observer one positive sign (%)	Observer two positive sign (%)	Kappa (chance corrected agreement with 95% CI)[b]	Phi (chance independent agreement)[c]
Raised respiratory rate (observer one opinion vs. observer two opinion)	214 (84%)	8.8%	5.5%	0.29 (0.16, 0.43)	0.54
Raised respiratory rate (observer one opinion vs. observer two counted rate)	93 (80%)a	8.8%	51.6%	0.12 (0.009, 0.23)	0.47
Fever (observer one opinion vs. observer two measured)	103 (89%)a	10.8%	3.9%	0.18 (0.005, 0.35)	0.42
Abnormal chest signs (observer one opinion vs. observer two opinion)	209 (82%)	21.5%	14.2%	0.39 (0.26, 0.53)	0.51

a Second winter data only, 116 children recruited. b Strength of agreement; < 0.2 poor, 0.2 − 0.4 fair, 0.41 − 0.6 moderate, 0.61 − 0.8 good, 0.81 − 1.0 very good.[15] c -1 perfect disagreement, 0 agreement no better than chance, +1 perfect agreement[13]

Results
Descriptive statistics
The cohort has been described in detail elsewhere[10]. We recruited 89% of eligible children presenting to 124 morning or evening surgeries at eight practices: two hundred and fifty six in total, 116 from the second winter. The two main reasons for not recruiting the 11% of eligible children were parental refusal and inability to read/write English. Sixty-one GPs and three NPs performed the role of observer one, and 96% of children were seen by a GP. Global assessment data from observer one were available in 98% of children for temperature and respiratory rate and 96% of children for chest signs. For observer two (ADH), data were available in 81% of children for respiratory rate, 85% for chest signs and 89% of children for temperature. Table 1 summarises the clinical data. For the first observer, one or more abnormal clinical findings were found in 80/241 (33%) of children with data complete for all three signs. Abnormal chest signs were found in 22%, fever in 11% and tachypnoea in 9%.

Inter-observer agreement
The number of children in whom inter-observer agreement was assessed is shown in Table 2. Kappa values were poor to fair for all clinical signs (range 0.12 to 0.39) with chest signs the most reliable[15]. Phi values showed less variation (range 0.42 to 0.51), with raised respiratory rate the most reliable.

Discussion
Summary of main results
This study shows that in usual practice, primary care clinicians found one or more abnormal sign in a third of preschool children with cough in primary care, and used a thermometer or formally counted the respiratory rate in a quarter. The inter-observer agreement between un-standardised and standardised assessments of these signs was at best fair.

Interpretation of results
Children presenting to primary care are seen earlier in the natural history of their condition than those presenting to secondary care, when signs are likely to be less subtle. Although we found similar levels of inter-observer agreement to studies in secondary care, it is disappointing that the kappa values were not higher. This may in part be explained by the low proportion with abnormal signs (as judged by either observer). This leads to paradoxically low kappa values[16,17]. We therefore also calculated phi values and, as would be expected, these showed less sensitivity to the proportion with positive signs. In general though, the level of agreement achieved calls into question the usefulness of signs in everyday clinical practice to assist diagnosis, prognosis and antibiotic treatment. For example, kappa values of ≥ 0.6 are recommended if symptoms or signs are to be used in clinical prediction rules[18]. In part, it may explain the wide variation seen in diagnostic labels used for respiratory tract infection in primary care[19]. However, it is possible that agreement might be improved if clinicians adopt a more standardised approach to assessment.

The second observer found a higher proportion of children with tachypnoea using counted respiratory rate compared with the global assessment. Previous research suggests that this may be because, in their global assessment of respiratory rate, clinicians adjust for other factors such as the child's general condition, presence of cyanosis, respiratory effort and accessory muscle use[3].

Where this fits in with other research
Notwithstanding the levels observed, our study has demonstrated similar inter-rater agreement to previous studies using higher levels of standardisation of examination in children and adults in secondary care. Studies of infants summarised in a review found inter-rater kappa values of 0.49 for respiratory retractions, 0.59 for accessory muscle use, 0.3 for crepitations and 0.29 for wheezing[3]. A study of adults found inter-rater kappas of 0.25 for tachypnoea,

0.51 for wheezes, 0.41 for crackles and 0.32 for bronchial breath sounds[20].

Limitations

While we have no reason to believe that the children recruited in the second winter differ systematically from those from the first winter, the lower number of children with measured temperature and counted respiratory rate from the second winter limits the precision of these estimates in our study. Respiratory rate can fluctuate quickly and it is possible that the 30 minutes maximum between clinical assessments explains some of the poor agreement. Our desire to compare usual clinical practice with a standardised assessment means we have not been able to assess the agreement of counted respiratory rate or thermometer measured temperature or further our understanding of how the clinicians identify abnormal clinical signs. We do not know from this study whether the standardised or non-standardised assessment is more accurate at predicting diagnosis or prognosis, nor have we assessed the intra-observer agreement of clinical signs. It is possible that the data collection form altered the clinical behaviour of observer one. This may have changed the number of children identified with abnormal signs, counted respiratory rate or thermometer-measured temperature. While we used mercury thermometry for the standardised assessment, we acknowledge its use in day-to-day practice is limited by the inconvenience of prolonged measurement time.

Conclusions

Primary care clinicians should be aware that clinical signs may be unreliable when making diagnosis, prognosis and treatment decisions in pre-school children with cough. Future research should aim to further our understanding of how best to identify abnormal clinical signs and examine the inter- and intra-observer agreement of standardised clinical assessments.

Competing interests

None declared.

Authors' contributions

AH and AW conceived the idea for the study and AH analysed the data. AH drafted the paper with subsequent contributions from all the authors. AH is the guarantor.

Additional material

Additional File 1

This was the data collection form used for the clinical data recorded by observer one (sections G and H, pages 5 and 6) and observer two (section P, page 10).
Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2296-5-4-S1.doc]

References

1. McCormick A, Fleming D, Charlton J: **Morbidity statistics from general practice.** *Fourth national study 1991–1992. London, HMSO* 1995, **4:**.
2. Okkes M, Oskam SK, Lamberts H: **The Probability of Specific Diagnoses for Patients Presenting with Common Symptoms to Dutch Family Physicians.** *J Fam Pract* 2002, **51:**31-6.
3. Margolis P, Gadomski A: **Does this infant have pneumonia?** *JAMA* 1998, **279:**308-13.
4. Hay AD, Fahey T, Peters TJ, Wilson AD: **Predicting complications from acute cough in pre-school children in primary care: a prospective cohort study.** *Br J Gen Pract* 2004, **54:**9-14.
5. Little P, Gould C, Moore M, Warner G, Dunleavey J, Williamson I, Del Mar C, Doust J: **Predictors of poor outcome and benefits from antibiotics in children with acute otitis media: pragmatic randomised trial.** *BMJ* 2002, **325:**22.
6. Holmes WF, Macfarlane JT, Macfarlane RM, Hubbard R: **Symptoms, signs, and prescribing for acute lower respiratory tract illness.** *Br J Gen Pract* 2001, **51:**177-81.
7. Howie JG: **Diagnosis — the Achilles heel?** *J R Coll Gen Pract* 1972, **22:**310-5.
8. Elmore JG, Feinstein AR: **A bibliography of publications on observer variability (final installment).** *J Clin Epidemiol* 1992, **45:**567-80.
9. Lozano JM, Steinhoff M, Ruiz JG, Mesa ML, Martinez N, Dussan B: **Clinical predictors of acute radiological pneumonia and hypoxaemia at high altitude.** *Arch Dis Child* 1994, **71:**323-7.
10. Hay AD, Wilson AD, Fahey T, Peters TJ: **The natural history of cough in pre-school children: a prospective cohort study.** *Fam Pract* 2003, **20:**696-705.
11. Swash M: **Hutchison's Clinical Methods.** *London: Balliere Tindall* 1989.
12. Stata Corporation: **Stata statistical software. (7.0).** *College Station, TX* 2001.
13. **Users' Guide to the Medical Literature.** *Chicago: American Medical Association* 2001.
14. **Advanced Life Support Group. Recognition of the seriously ill child.** In *Advanced Paediatric Life Support, the Practical Approach* Edited by: *Mackway-Jones K, Molyneux E, Phillips B, Wieteska S. London: BMJ Publishing Group*; 1994:13.
15. Altman DG: **Practical Statistics for Medical Research.** *London: Chapman and Hall* 1997.
16. Feinstein AR, Cicchetti DV: **High agreement but low kappa: I. The problems of two paradoxes.** *J Clin Epidemiol* 1990, **43:**543-9.
17. Cicchetti DV, Feinstein AR: **High agreement but low kappa: II. Resolving the paradoxes.** *J Clin Epidemiol* 1990, **43:**551-8.
18. Laupacis A, Sekar N, Stiell IG: **Clinical prediction rules. A review and suggested modifications of methodological standards.** *JAMA* 1997, **277:**488-94.
19. Howie JG, Richardson IM, Gill G, Durno D: **Respiratory illness and antibiotic use in general practice.** *J R Coll Gen Pract* 1971, **21:**657-63.
20. Spiteri MA, Cook DG, Clarke SW: **Reliability of eliciting physical signs in examination of the chest.** *Lancet* 1988, **1:**873-5.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2296/5/4/prepub