ORIGINAL ARTICLE

Validity and reliability of acoustic analysis of respiratory sounds in infants

H E Elphick, G A Lancaster, A Solis, A Majumdar, R Gupta, R L Smyth

....

Objective: To investigate the validity and reliability of computerised acoustic analysis in the detection of abnormal respiratory noises in infants.

Methods: Blinded, prospective comparison of acoustic analysis with stethoscope examination. Validity and reliability of acoustic analysis were assessed by calculating the degree of observer agreement using the κ statistic with 95% confidence intervals (CI).

Results: 102 infants under 18 months were recruited. Convergent validity for agreement between stethoscope examination and acoustic analysis was poor for wheeze (κ = 0.07 (95% CI, −0.13 to 0.26)) and rattles (κ = 0.11 (−0.05 to 0.27)) and fair for crackles (κ = 0.36 (0.18 to 0.54)). Both the stethoscope and acoustic analysis distinguished well between sounds (discriminant validity). Agreement between observers for the presence of wheeze was poor for both stethoscope examination and acoustic analysis. Agreement for rattles was moderate for the stethoscope but poor for acoustic analysis. Agreement for crackles was moderate using both techniques. Within-observer reliability for all sounds using acoustic analysis was moderate to good.

Conclusions: The stethoscope is unreliable for assessing respiratory sounds in infants. This has important implications for its use as a diagnostic tool for lung disorders in infants, and confirms that it cannot be used as a gold standard. Because of the unreliability of the stethoscope, the validity of acoustic analysis could not be demonstrated, although it could discriminate between sounds well and showed good within-observer reliability. For acoustic analysis, targeted training and the development of computerised pattern recognition systems may improve reliability so that it can be used in clinical practice.

See end of article for authors' affiliations
....

Correspondence to:
Dr H E Elphick, Institute of Child Health, Royal Liverpool Children's Hospital, Eaton Road, Liverpool L12 2AP, UK; helphick@liv.ac.uk

Accepted
25 February 2004
....

The evaluation of findings using the clinical history and examination is an important component of clinical diagnosis. Despite this, few studies have assessed their precision and accuracy. Studies in adult subjects suggest that the level of agreement among independent observers for the presence of respiratory signs using a stethoscope is poor.[1–3] Inability to identify respiratory signs accurately has been shown to lead to an incorrect diagnosis by physicians on 28% of occasions.[4] In spite of its importance, there is no information in current reports on agreement between observers for examination of children and infants with a stethoscope.

Computerised acoustic analysis is a technique used to evaluate the acoustic properties of respiratory sounds. Guidelines on methods of recording and analysis of breath sounds have been published recently in an attempt to standardise terminology.[5] Although there were no references to infants in this document, this is a useful technique to apply to infants because it does not require any sedation and is non-invasive. The equipment is portable and measurements can therefore be made at the bedside.

No study has previously compared acoustic analysis with stethoscope examination in infants. This is a particularly important group in which to use terminology with accuracy, as the underlying respiratory diagnosis in this age group can be difficult to establish.[6] The present study was designed to compare the findings of acoustic analysis with those of the clinical examination. The aim was to investigate the validity and reliability of acoustic analysis in the detection of abnormal respiratory noises in infants.

METHODS

The study design was a blinded, prospective comparison of computerised acoustic analysis with stethoscope examination. Because of the poor agreement reported between observers for stethoscope examination in previous studies, this method could not be regarded as sufficiently reliable to be used as a gold (reference) standard against which to compare the acoustic analysis technique. Instead, we chose to measure agreement between the two methods.

The study received approval from the local paediatric research ethics committee, and informed consent was obtained from the parents of all participants.

Participants

We recruited 102 subjects from inpatients in the general paediatric wards at Alder Hey Children's Hospital, Liverpool, with noisy breathing arising from lower respiratory tract illness. Infants of either sex under the age of 18 months were included. There were no exclusion criteria. Clinical information was not shared with any of the investigators before the study measurements took place. On the days of recruitment a consecutive sample was taken, regardless of the clinical condition of the patient, and all participants included were subjected to both measurements, except for two who were not recorded by acoustic analysis—one because of breakdown of recording equipment and one because of acute deterioration in clinical status. Each child was examined using a stethoscope by two experienced clinicians, and the findings on each infant were recorded independently of each other and independently from the acoustic analysis findings. Blinding was ensured by asking each clinician to complete evaluation forms, which were then placed into a coded envelope which was not opened until the acoustic analysis had been completed. No treatments were given between tests, which were done either concurrently or immediately sequentially.
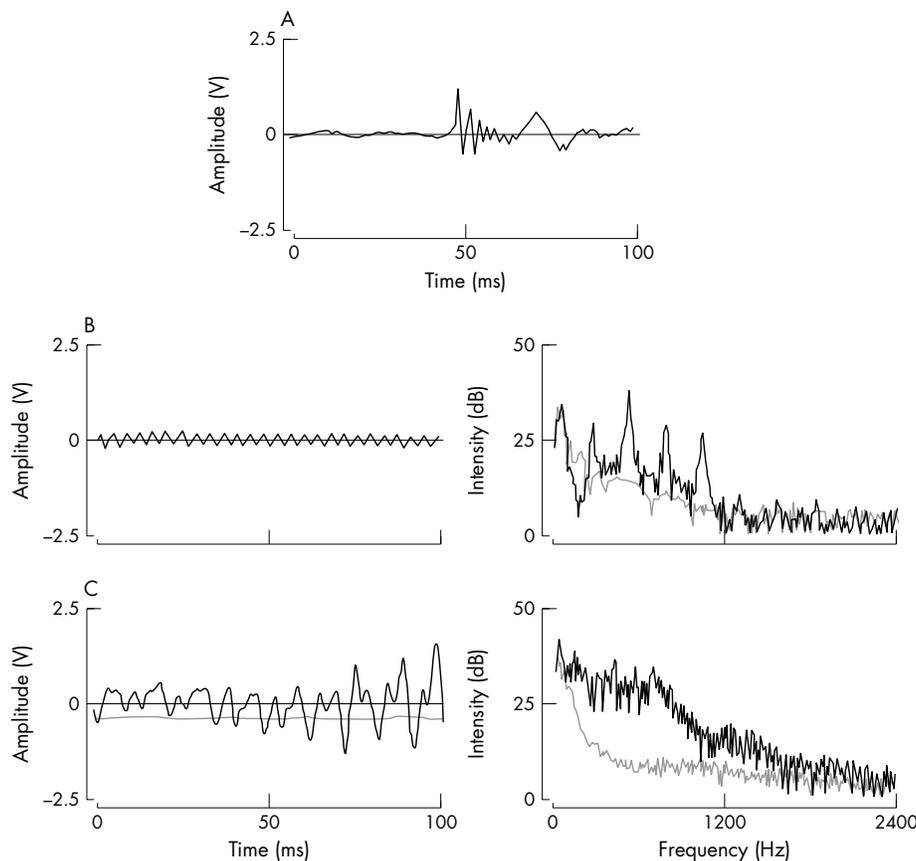
Figure 1 Acoustic characteristics of respiratory noises in infants. (A) Waveform pattern representing crackles. (B) Waveform and power spectrum patterns representing wheeze. (C) Waveform and power spectrum patterns representing rattles.

Stethoscope examination

The clinicians (AM and RG) were specialist registrars with more than six years' experience in paediatrics as well as having experience in respiratory paediatrics. Definitions of

Table 1 Patient characteristics

Age (months) (mean (SD))	5.1 (3.3)
Weight (kg) (mean (SD))	6.8 (3.5)
Per cent male	69
Diagnosis (n)	
Bronchiolitis	
RSV	51/88
Adenovirus	1/88
Virus unknown	36/88
Required oxygen	66/88
Required nasogastric/IV feeding	49/88
Required IV antibiotics	9/88
Required bronchodilators	22/88
Average length hospital stay	5.7 days
Other diagnoses (n)	
Pneumonia/atelectasis	3
Cerebral palsy	2
BPD	1
Asthma	3
VATER anomaly	1
Cystic fibrosis	1
Whooping cough	1

BPD, bronchopulmonary dysplasia; IV, intravenous; RSV, respiratory syncytial virus; VATER, vertebral/anal/tracheal/esophageal/renal syndrome.

the respiratory noises to be assessed were agreed on before the start of the study. Further than this, we did not undertake to train the clinicians to high levels of agreement, so that estimates of reproducibility were reflective of the everyday clinical environment. Each examination took place for one minute during a time when the infant was quiet and settled. The right upper zone of the anterior chest was the only site examined, so that a direct comparison could be made with acoustic analysis. The clinicians were given an evaluation form which required assessment of the presence or absence of wheeze, rattles, and crackles.

Acoustic analysis

Recordings of the infant's respiratory sounds were undertaken by HE using a contact sensor (Siemens EMT 25C) placed over the right upper zone of the infant's chest, anteriorly, attached to the skin using a double sided adhesive ring. The recording on each occasion lasted for approximately one minute of tidal breathing, with the baby quiet or asleep. Recordings were made in a quiet room with background noise kept to a minimum. Specialised equipment for sound recording and analysis was used (RALE, Respiratory Acoustics Laboratory Environment). The signals were filtered, amplified, and transmitted to an IBM compatible personal computer. The sound was analysed using a fast Fourier transformation technique. A waveform (amplitude/time), a power spectrum (intensity/frequency), and a sonogram (frequency/time, with intensity as a colour scale), were displayed on the computer screen and each of these visual patterns could be assessed for the identification of wheeze, rattles, and crackles.

Classification of sounds

The classical wheeze is a high pitched whistling sound with a musical quality, characterised acoustically by a sinusoidal waveform with distinct peaks in the power spectrum display.[7] Crackles are defined as intermittent sounds of short duration with a characteristic sharp, sudden deflection followed by rapidly dampened wave deflections.[8] Rattles (referred to in some parts of the United Kingdom as "ruttles")[9] are much lower in pitch than a wheeze, with a continuous non-musical quality, and may reflect excessive secretions in the airways (fig 1). Rattles are often mistaken for wheeze,[10] but in acoustic terms, this noise is quite distinct, with an irregular non-sinusoidal waveform and diffuse peaks in the power spectrum.

Sample size calculation

Sample size was calculated using Nquery Advisor v4.0.[11] For the validity comparison between stethoscope examination and acoustic analysis, for a sample size of 96 a two sided 95% confidence interval (CI) for the κ statistic will extend approximately 0.2 from the observed value of κ, assuming the true value of κ to be between 0.1 and 0.4 and the proportion of successes to be 0.5.

Data analysis

The degree of observer agreement was represented using the κ statistic with 95% CI using the StatsDirect statistical package.[12] The κ statistic can range from −1 to +1. A negative value implies that the proportion of agreement made by chance is greater than the proportion of observed agreement—that is, the agreement is less than would have been expected by chance. Positive values ranging from 0 to <0.2 indicate poor agreement, >0.2 to 0.4 fair agreement, >0.4 to 0.6 moderate agreement, >0.6 to 0.8 good agreement, and >0.8 to 1 very good agreement.[13]

Validity

Construct validity was assessed by determining convergent and discriminant validity (n = 100). Convergent validity shows the agreement between the two different methods in measuring the same sound. Discriminant validity is represented by the ability of the test to discriminate between different sounds.

Reliability

Separate blind evaluations by different observers (between-observer agreement) and repeat assessments of the same patient by the same observer (within-observer agreement) were carried out. These were done in four consecutive subsets of participants.

- Between-observer agreement for the stethoscope was assessed by comparing the findings of clinicians AM and RG (n = 36). This analysis was included to compare our findings with previous studies.
- Between-observer agreement for the acoustic analysis was assessed by comparing the results of analysers HE and AS, of a subset of the breath sound recordings (n = 20).
- Within-observer agreement for the acoustic analysis instrument was assessed by re-recording each patient 5 minutes after the initial recording (n = 58). All of these recordings were analysed by the same analyser (HE).
- Within-observer agreement for repeat analysis was assessed by analysis of the same recording by the same analyser (HE) on two separate occasions, two months apart (n = 19).

RESULTS

In all, 102 infants were recruited, achieving our calculated sample size requirement. Acoustic and stethoscope data were available for 100 infants. Table 1 outlines the clinical characteristics of the patient group.

Validity

Convergent and discriminant validity are shown using 2×2 tables (table 2).

Table 2 Validity of acoustic analysis v stethoscope

A Convergent validity: acoustic analysis v stethoscope

	Wheeze			Rattles			Crackles		
	Yes	No	Total	Yes	No	Total	Yes	No	Total
Yes	38	27	65	45	35	80	21	24	45
No	18	17	35	8	12	20	6	49	55
Total	56	44	100	53	47	100	27	73	100
	κ=0.07 (−0.13 to 0.26)			κ=0.11 (−0.05 to 0.27)			κ=0.36 (0.18 to 0.54)		

B Discriminant validity: stethoscope

	Wheeze v rattles			Wheeze v crackles			Rattles v crackles		
	Yes	No	Total	Yes	No	Total	Yes	No	Total
Yes	30	26	56	12	42	54	8	44	52
No	22	22	44	15	31	46	19	29	48
Total	52	48	100	27	73	100	27	73	100
	κ=0.01 (−0.18 to 0.21)			κ=−0.1 (−0.27 to 0.07)			κ=−0.24 (−0.4 to 0.07)		

C Discriminant validity: acoustic analysis

	Wheeze v rattles			Wheeze v crackles			Rattles v crackles		
	Yes	No	Total	Yes	No	Total	Yes	No	Total
Yes	55	10	65	27	38	65	32	50	82
No	24	11	35	15	20	35	10	8	18
Total	79	21	100	42	58	100	42	58	100
	κ=0.16 (−0.02 to 0.33)			κ=−0.03 (−0.19 to 0.16)			κ=−0.13 (−0.2 to 0.05)		

2×2 tables demonstrating convergent and discriminant validity; κ (95% confidence interval) given for each comparison.

Table 3 Reliability of acoustic analysis v stethoscope

	Wheeze			Rattles			Crackles		
	Yes	No	Total	Yes	No	Total	Yes	No	Total
A *Between-observer agreement for stethoscope*									
Yes	13	14	27	10	4	14	3	2	5
No	2	7	9	4	18	22	3	28	31
Total	15	21	36	14	22	36	6	30	36
	κ = 0.18 (−0.08 to 0.44)			κ = 0.53 (0.21 to 0.86)			κ = 0.46 (0.14 to 0.79)		
B *Between-observer agreement for acoustic analysis*									
Yes	4	7	11	7	6	13	4	4	8
No	1	8	9	2	5	7	1	11	12
Total	5	15	20	9	11	20	5	15	20
	κ = 0.24 (−0.12 to 0.60)			κ = 0.22 (−0.18 to 0.63)			κ = 0.44 (0.03 to 0.86)		
C *Within-observer agreement for acoustic analysis (instrument)*									
Yes	29	9	38	44	4	48	15	7	22
No	3	17	20	3	7	10	4	32	36
Total	32	26	58	47	11	58	19	39	58
	κ = 0.57 (0.32 to 0.82)			κ = 0.59 (0.34 to 0.85)			κ = 0.56 (0.30 to 0.82)		
D *Within-observer agreement for acoustic analysis (repeat analysis)*									
Yes	9	1	10	16	0	16	6	1	7
No	1	8	9	1	2	3	1	11	12
Total	10	9	19	17	2	19	7	12	19
	κ = 0.79 (0.43 to 1.24)			κ = 0.77 (0.33 to 1.20)			κ = 0.77 (0.32 to 1.22)		

2×2 tables showing for how many infants there was agreement between the two measures; κ (95% confidence interval) given for each comparison. Rows and columns are defined as (A) and (B) observer 1 v observer 2; (C) and (D) occasion 1 v occasion 2.

Convergent validity
Convergent validity is shown in table 2A. Of 100 infants, the number for whom there was agreement for identification of wheeze was 55 (κ = 0.07 (95% CI, −0.13 to 0.26)), for rattles it was 57 (κ = 0.11 (−0.05 to 0.27)), and for crackles it was 70 (κ = 0.36 (95% CI, 0.18 to 0.54)). The κ scores indicate poor agreement, except for crackles, for which agreement was fair.

Discriminant validity
To demonstrate discriminant validity, there should ideally be total disagreement between the techniques, and κ values for agreement should be negative. In 100 infants, for wheeze and rattles the stethoscope (table 2B) was able to distinguish between the two noises in 48 (κ = 0.01 (−0.18 to 0.21)), compared with acoustic analysis (table 2C), which was able to distinguish between wheeze and rattles in 34 (κ = 0.16 (−0.02 to 0.33)). For wheeze v crackles, the stethoscope was able to distinguish between the noises in 57 infants (κ = −0.1 (–0.27 to 0.07)) and acoustic analysis in 53 (κ = −0.03 (−0.19 to 0.16)). For rattles v crackles, the stethoscope was able to distinguish between the noises in 63 infants (κ = −0.24 (−0.4 to 0.07)) and acoustic analysis in 60 (κ = −0.13 (−0.2 to 0.05)). The negative values for κ imply that agreement is less than would have been expected by chance.

Reliability
The 2×2 classification tables for each comparison together with the κ statistics are shown in table 3.

Between-observer agreement for stethoscope
These data are shown in table 3A. Of 36 infants, the number for whom there was agreement was 20 (κ = 0.18 (95% CI, −0.08 to 0.44)). This κ score represents poor agreement. The number of infants for whom there was agreement for rattles was 28 (κ = 0.53 (0.21 to 0.86)) and for crackles it was 31 (κ = 0.46 (0.14 to 0.79)). These κ scores represent moderate agreement.

Between-observer agreement for acoustic analysis
These data are shown in table 3B. Of 20 infants, the number for whom there was agreement for wheeze was 12 (κ = 0.24

(−0.12 to 0.60)), for rattles it was 7 (κ = 0.22 (−0.18 to 0.63)), and for crackles it was 15 (κ = 0.44 (0.03 to 0.86)). Agreement was therefore poor for wheeze and rattles and moderate for crackles.

Within-observer agreement for acoustic analysis (instrument)
These data are shown in table 3C. Of 58 infants, the number for whom there was agreement for wheeze was 46 (κ = 0.57 (0.32 to 0.82)), for rattles it was 51 (κ = 0.59 (0.34 to 0.85)), and for crackles it was 47 (κ = 0.56 (0.30 to 0.82)). The κ scores indicate moderate agreement for each comparison.

Within-observer agreement for acoustic analysis (repeat analysis)
These data are shown in table 3D. Of 19 infants, the number for whom there was agreement for wheeze was 17 (κ = 0.79 (0.43 to 1.24)), for rattles it was 18 (κ = 0.77 (0.33 to 1.2)), and for crackles it was 17 (κ = 0.77 (0.32 to 1.22)). All these κ scores represent good agreement.

DISCUSSION
This study is the first to compare acoustic analysis with stethoscope examination in infants. One of the main findings was that the reliability of stethoscope examination of respiratory sounds in infants between two experienced observers was poor to moderate. Between-observer reliability using computerised acoustic analysis was also poor to moderate, but this may have reflected the relative inexperience of the second analyst used in the study. Within-observer reliability for all sounds using acoustic analysis was moderate to good. A perfect correlation would never occur, as changes in the subject's clinical state—even as a result of a cough—may take place between recordings, but our findings do reflect good stability of the technique. Validity of computerised acoustic analysis was poor to fair, although it was difficult to draw clear conclusions owing to the unreliability of the stethoscope to which it was being compared.

A strength of this study was that rigorous methodology was used, in line with the Cochrane Methods Group recommendations for studies of diagnostic accuracy.[14] In this

way most potential biases were overcome. A sample size was calculated for the comparison between the techniques and this sample was achieved. A problem with our study was that smaller sample sizes for the subsets were used for the reliability comparisons. These smaller subsets were unavoidable because of time and personnel restraints, and resulted in wider confidence intervals for these comparisons.

The κ values for agreement using the stethoscope in our study were similar to or lower than those reported previously in studies in adults,[2–4] and also lower than the value in a previous study in which a comparison was made between computer analysis and subjective assessment.[15] All the previous studies assessed only the presence or absence of one noise, wheeze. Ours is the only study so far to have investigated in detail the distinction between three different sounds, which—while reflecting clinical practice more accurately—added to the complexity of the study and probably reduced the apparent reliability of the two measurements.

Most of the infants included in this study had bronchiolitis, which is characteristically associated with a combination of more than one respiratory sound. Each participant therefore is likely to have had more than one noise identified during the minute-long study period. Other studies have investigated infants with bronchiolitis and recurrent wheeze using computerised acoustic analysis.[16 17] ''Complex repetitive waveforms'' were described in some of these infants. These are likely to be similar to those noises recorded as ''rattles'' in our study. The presence of different sounds and the inconsistency in nomenclature add confusion to the evaluation of an already challenging group of individuals in terms of their age and ability to cooperate.[18]

In clinical practice, the stethoscope is widely used in the evaluation of young children with respiratory symptoms, given the paucity of more objective diagnostic alternatives. One of the most common diseases of childhood, asthma, is usually diagnosed on the basis of a history of recurrent wheeze, and epidemiological studies of the natural history and treatment of asthma are based on cohorts characterised by the presence or absence of wheeze. In bronchiolitis, oxygen requirement and the presence of crackles on examination of the chest have been shown to be the only clinical predictors of severity of the illness.[19] The poor level of agreement for stethoscope findings in our study raises concern over the reliability of this method for the evaluating wheezing illness in this age group, and for the initial assessment of infants with bronchiolitis.

Our study shows how unreliable the stethoscope is in the assessment of respiratory sounds in infants. This has important implications for its use as a diagnostic tool for lung disorders in infants and confirms that it cannot be used as a gold standard. Because of the unreliability of the stethoscope, the validity of acoustic analysis could not be demonstrated, although it was able to discriminate between sounds well and showed good within-observer reliability.

Acoustic analysis does have a possible role, provided that validity and reliability can be improved. In addition to specific and targeted training, the development of simpler and lower cost instruments and computerised pattern recognition systems may reduce variability in the interpretation of respiratory sounds in infants.

. . . .

Authors' affiliations
H E Elphick, R L Smyth, Institute of Child Health, Royal Liverpool Children's Hospital, Liverpool, UK
G A Lancaster, Centre for Medical Statistics and Health Evaluation, University of Liverpool
A Solis, A Majumdar, R Gupta, Respiratory Unit, Royal Liverpool Children's Hospital

REFERENCES
1 **Godfrey S**, Edwards RHT, Campbell EJM, et al. Repeatability of physical signs in airway obstruction. Thorax 1969;**24**:4–9.
2 **Smyllie HC**, Blendis LM, Armitage P. Observer disagreement in physical signs in the respiratory system. Lancet 1965;**2**:412–13.
3 **Brooks D**, Thomas J. Interrater reliability of auscultation of breath sounds among physical therapists. Phys Ther 1995;**75**:1082–8.
4 **Spiteri MA**, Cook DG, Clarke SW. Reliability of eliciting physical signs in examination of the chest. Lancet 1988;**i**:873–5.
5 **Sovijarvi ARA**, Vanderschoot J, Earis JE. Computerised Respiratory Sound Analysis (CORSA): recommended standards for terms and techniques. ERS Task Force Report. Eur Resp Rev 2000;**10**:595–649.
6 **Silverman M**. Out of the mouths of babes and sucklings: lessons from early childhood asthma. Thorax 1993;**48**:1200–4.
7 **Pasterkamp H**, Carson C, Daien D, et al. Digital respirosonography: new images of lung sounds. Chest 1989;**96**:1405–12.
8 **Pasterkamp H**, Kraman SS, Wodicka GR. State of the art. Respiratory sounds: advances beyond the stethoscope. Am J Respir Crit Care Med 1997;**156**:974–87.
9 **Elphick HE**, Ritson S, Rodgers H, et al. When a ''wheeze'' is not a wheeze: acoustic analysis of breath sounds in infants. Eur Respir J 2000;**16**:593–7.
10 **Cane RC**, McKenzie SM. What do parents of wheezy children understand by ''Wheeze''? Arch Dis Child 2000;**82**:327–32.
11 **Elashoff JD**. Nquery Advisor version 4.0 User's guide. Los Angeles, CA, 2000.
12 StatsDirect Ltd, 2002. StatsDirect statistical software. Web address: http://www.statsdirect.co.uk.
13 **Landis R**, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;**33**:159–74.
14 **Cochrane Methods Group**. Systematic review of screening and diagnostic tests: recommended methods, updated 6 June 1996. Available at http://www.cochrane.org/cochrane/sadtdoc1.htm.
15 **Pasterkamp H**, Wiebicke W, Fenton R. Subjective assessment vs computer analysis of wheezing in asthma. Chest 1987;**91**:376–81.
16 **Tal A**, Sanchez I, Pasterkamp H. Respirosonography in infants with acute bronchiolitis. Am J Dis Child 1991;**145**:1405–10.
17 **Sanchez I**, Navarro H, Bertrand P, et al. Acoustic analysis of wheezing in infants with acute bronchial obstruction. A follow-up study. Rev Med Chil 2002;**130**:760–7.
18 **Postiaux G**, Lens E. Pulmonary stethacoustic nomenclature: why not a worldwide consensus? Rev Mal Respir 1999;**16**:1075–90.
19 **Mulholland EK**, Olinsky A, Shann FA. Clinical findings and severity of acute bronchiolitis. Lancet 1990;**335**:1259–61.