Eur Respir J 1998; 12: 944–953
DOI: 10.1183/09031936.98.12040944

European Respiratory Journal
ISSN 0903 - 1936

A collaborative study of infant respiratory function testing

I. Dundas*, C. Beardsmore**, T. Wellman**, J. Stocks*

A collaborative study of infant respiratory function testing. I. Dundas, C. Beardsmore, T. Wellman, J. Stocks. ©ERS Journals Ltd 1998.

ABSTRACT: The aims of this study were to compare inter-observer variability within and between two specialized infant lung function testing centres and to develop a strategy for performing and analysing infant respiratory function tests to facilitate future collaborative trials.

A protocol for data collection and analysis was developed using similar equipment and identical software. All raw data were exchanged on disk and analysed, blind to infant status. All data were cross-analysed by both centres to assess inter-observer variability. Outcome measures were functional residual capacity (FRC_{pleth}), airway resistance (R_{aw}) and maximal expiratory flow at FRC ($V'_{max,FRC}$). Subjects were recruited from the multicentre UK extracorporeal membrane oxygenation (ECMO) Trial and measured at around 1 yr of age. Forty-two infants attended the Institute of Child Health, London and 36 attended the Leicester Royal Infirmary. The proportion of infants treated with ECMO or conventional management at each centre was similar.

There were no significant differences between any of the outcome measures for infants tested at either centre. During a cross-analysis, the agreement between the two centres, within infant, was closer for $V'_{max,FRC}$ and FRC_{pleth} (within 10%) than for the more variable measurements of R_{aw} (within 20%).

A collaborative approach to trials with infant respiratory function as an outcome measure appears feasible, providing that close attention is paid to study design, and participants in such trials maintain a standard approach to data collection and analysis.
Eur Respir J 1998; 12: 944–953.

*Portex Anaesthesia, Intensive Therapy and Respiratory Medicine Unit, Institute of Child Health, London. **Dept of Child Health, Leicester Royal Infirmary, Leicester, UK.

Correspondence: I. Dundas
Portex Anaesthesia
Intensive Therapy and Respiratory Medicine Unit
Institute of Child Health
London WC1N 1EH
UK
Fax: 44 171 829 8634

Keywords: Clinical trial
infant
methods
outcome
respiratory function tests

Received: January 16 1998
Accepted after revision May 21 1998

This study was supported by The Wellcome Trust. J. Stocks is supported by SIMS Portex plc.

Infant respiratory function tests are time-consuming and relatively complex, but there is increasing interest in their use as outcome measures of interventions, or indicators of disease severity. However, accrual of adequate numbers of infants to studies based in one test centre or geographical area may take many years, and reporting of smaller numbers will result in insufficient power to evaluate the statistical significance of outcome measures. In contrast, large multicentre clinical trials may accrue more infants than can be reasonably measured at one centre. Collaboration between centres that perform these specialized tests would be the most realistic alternative, but a common methodological approach and compatible equipment and software are a prerequisite to pooling data. In 1994, the opportunity arose to undertake a respiratory follow up of all survivors of a national randomized trial of extracorporeal membrane oxygenation (ECMO), hereafter referred to as the main ECMO trial [1].

Mature neonates with reversible respiratory disease were eligible for the main ECMO trial, and the aim of the respiratory follow up was to compare the respiratory health and function at 1 yr in infants who were assigned to receive ECMO with that of similar infants who were assigned to conventional management (CM). In view of the wide geographical area involved, the number of infants to be measured, and the potential benefits of a collaborative study, it was decided that the respiratory measurements should be performed at two specialized centres in the UK, based at the Institute of Child Health, London (hereafter referred to as ICH) and at the Department of Child Health, Leicester Royal Infirmary (hereafter referred to as LEIC). Although the measurements of infant respiratory function were well established at both centres, the complexity of measurement procedure and methods of analysis led to considerable potential for differences in approach. Participation in this two-centre trial, which lasted from August 1994 to December 1996, offered the opportunity to address methodological issues, and to develop a standardized approach that could be applied within future multicentre trials.

The aims of this study were: 1) to develop a strategy for performing and analysing infant respiratory function tests to facilitate future multi-centre trials; and 2) to compare the inter-observer variability within and between two specialized infant lung function testing centres with respect to plethysmographic measurements of functional residual capacity (FRC_{pleth}) and airway resistance (R_{aw}), and assessments of maximal flow at FRC ($V'_{max,FRC}$), using the tidal rapid thoraco-abdominal compression technique (RTC), these being the main outcome variables of the respiratory follow-up [2].

Methods and subjects

All infants who were recruited into the main ECMO trial and survived to 1 yr were eligible for the respiratory

follow-up. Entry criteria for the main ECMO trial are published elsewhere [1]. One hundred and eighty-five infants were entered into the main ECMO trial, and 101 survived to become eligible for the respiratory follow-up at 1 yr.

Arrangements were made for clinical assessment and respiratory function testing at ICH or LEIC. Parents who elected to take part in the respiratory follow-up were given their choice of venue. Laboratory staff were blinded to the infants' management status (ECMO or CM), and parents were requested to withhold any information regarding neonatal history during their visit. On arrival in each laboratory, infants were dressed in a specially designed smock of high neck design to obscure any neck scars received during ECMO treatment. A detailed history was taken with reference to respiratory disease, and the infant was examined clinically by a staff member not directly involved in measurement and analysis. The respiratory questionnaire was designed to provide information concerning possible confounding variables such as smoking in the home or a family history of asthma. Baseline oxygen saturation was recorded using pulse oximetry, and the infant was weighed, wearing only the smock, before being sedated with triclofos sodium (100–150 mg·kg^{-1}) at ICH or a similar dose of chloral hydrate (80–140 mg·kg^{-1}) at LEIC. The infant's length was recorded to the nearest 0.1 cm as described previously [3].

The methodology for performing respiratory measurements at each test centre was already well established when the respiratory follow-up commenced [4, 5]. However, there were differences in equipment, software set-up and methodology between centres that could potentially influnce the approach to collection and analysis of data. These differences were examined prior to starting the respiratory follow-up and at regular intervals subsequently by arranging interlaboratory visits when equipment, infant measurements and analytical techniques were observed and compared, including cross-analysis of the same data. Following an interim analysis of results, 7 months after commencement of the study, minor amendments to the protocol, with respect to data collection, analysis and reporting results, were implemented as described below.

This study was approved by the local ethics committee at each centre. Written informed consent was obtained from one or both parents prior to measurements. Parents were usually present during the measurements.

Details of equipment and in vitro *assessment*

Details of the equipment are summarized in table 1. Both plethysmographs were assessed using an identical test lung, made of copper tubing filled with copper wire, attached to a piston driven pump with a stroke volume of 7 mL. The volume of the test lung was approximately 200 mL (depending on the connectors used), and both centres obtained results to within 2% of the volume measured by water displacement. Following measurements using a face shaped out of putty or a nasal cast to simulate the space potentially occupied by an infant's nose and cheeks, the effective mask deadspace was considered to be 15 mL at ICH (*i.e.* 50% of the water displacement volume) and 7 mL at LEIC. The remaining difference in apparatus deadspace of 14 mL arose mainly from the design of the pneumotachograph (PNT: capillary at ICH and screen at LEIC), and partly from the slight differences in shutter configuration.

For measurements of $V'_{max,FRC}$ the deadspace and apparatus resistance presented to the infant was reduced, consisting of mask and PNT only (table 1). The inflatable jacket (Medical Engineering, Royal Postgraduate Medical School, Hammersmith Hospital, London) used for RTC measurements at LEIC was wrapped around infants with their arms within, and the entire anterior section was inflated. ICH infants wore a cummerbund-style adjustable jacket (Hannover Medical School, Germany) with their arms remaining outside.

Data collection. All data were displayed and recorded on an IBM-compatible computer, using interactive, operator controlled software (Respiratory Analysis Program (RASP), Physio Logic Ltd, Newbury, UK). Measurements were made only during quiet sleep when posture was stable, respiration regular and no eye movements observed [6]. Once asleep, the infant was wrapped in an adjustable in-flatable jacket, and a face mask and PNT were applied over the nose and mouth and made leakproof with therapeutic putty. Measurements of $V'_{max,FRC}$ were then performed as described previously [4, 7]. Jacket inflation pressures were applied in increments of 1 kPa until flow limitation was reached (no increase in flows with increasing applied pressure, for technically acceptable curves) up to a maximum of 10 kPa. Curves were acceptable if the time to peak jacket pressure was ð0.1 s, the interval between end inspiration and onset of jacket inflation was ð0.1 s and there was a rapid rise to peak expiratory flow with no evidence of glottic closure.

Reasons for rejection of data included: inspiration before FRC was reached, a change in end-expiratory volume following the manoeuvre, suggestive of a face mask leak,

Table 1. – Details of equipment at each centre

	ICH	LEIC
Infant plethysmograph	100 L custom built	85 L Jaeger
Pneumotachograph model	Fleisch "1" capillary type	Jaeger infant screen type
Effective deadspace: mask+PNT+shutter mL	41	19
Linear range of PNT	±500	±1000
Resistance: PNT and shutter at 100 mL·s^{-1} kPa·L^{-1}·s	0.5	0.23
Time constant* s	10–14	10–14
RTC measurements		
Deadspace face mask-PNT mL	27 (using Rendell-Baker size 2 mask)	16 (using Rendell-Baker size 2 mask)
Resistance: PNT at 100 mL·s^{-1} kPa·L^{-1}·s	0.20	0.18
Jacket model	Hannover arms out adjustable size	Hammersmith size 2–4 arms in

ICH: Institute of Child Health, London; LEIC: Department of Child Health, Leicester Royal Infirmary; PNT: pneumotachograph; RTC: rapid thoraco-abdmominal compression technique. *: combined mechanical and thermal time constant of the plethysmograph.

or evidence of expiratory braking during the forced expiratory manoeuvre. Whenever possible, the percentage of jacket pressure transmitted to the intrathoracic structures was assessed by performing an end inspiratory occlusion immediately prior to the jacket inflation and measuring the change in pressure at the airway opening during the subsequent jacket inflation relative to the applied jacket pressure [8].

The inflatable jacket was loosened before plethysmographic measurements were made. When measuring FRC_{pleth} at least five end inspiratory occlusions were performed, after a stable end expiratory level had been established, each being held for two to three respiratory efforts. Changes in plethysmographic and airway opening pressure were inspected and considered acceptable if no leaks were evident, and changes were in phase during airway occlusions. The infant was then allowed to rebreathe from a highly compliant 2 L bag containing warmed, humidified air (temperature at the mask ~37°C). The phase relationship between flow and plethysmographic signals was inspected and the temperature within the rebreathing bag adjusted if necessary, until a satisfactory pressure-flow loop was obtained [7, 9]. Three epochs of data, each of 18–36 s duration were collected, the rebreathing bag being emptied and refilled between each epoch.

The raw data from each infant, including calibration checks, were exchanged between centres by computer disk, soon after measurements were complete. Results were analysed, by both the test centre and the analysis centre, using the protocol agreed following an interim analysis. Minor adjustments to usual practice were made by both centres following this analysis, including standardization of sampling frequencies and the method by which end expiratory baselines were assessed, together with the recording of (rather than simply just performing) calibration checks and test occlusions with several pre- and post-manoeuvre breaths, the latter being to provide evidence in the data on disk regarding absence of any leaks [7, 10]. The aim of these adjustments was to minimize any inter-observer variability, and facilitate analysis by an operator who had not been present during the measurements, and therefore required evidence of quality control during data collection.

Data analysis

Interactive software (RASP), which recorded data in real time throughout the measurement period and allowed subsequent off-line analysis, was used by both centres. It was possible to customize both data collection and analysis to individual preference with respect, for example, to sampling frequency, number of breaths analysed, thresholds for event recognition, and methods of assessing end expiratory baselines. All data were stored on computer disk, including the original raw data channels (time, flow, volume and pressure signals), details of each analysed event, alongside the selected user options at time of analysis, and prevailing measurement conditions such as ambient pressure, apparatus dead space and calibrations at the time of data recording. Exchange of the raw data alone on disk allowed inspection of original signals and enabled reanalysis of results, blind to that of previous investigators. However, in the event of intercentre discrepancies in the calculated results, additional exchange of the analysis files facilitated rapid identification of the source of any bias.

Based on the inherent intra-subject variability of each of the respiratory parameters, the aim of the inter-observer analysis of each infant respiratory function measurement was to report values analysed by each centre to within 10% of each other for FRC_{pleth}, $V'_{max,FRC}$, tidal volume (V_T), and respiratory rate (RR), and 20% for the more variable measures of R_{aw} [5, 11], using the test centre result as the numerator and the analysis centre result as the denominator. Where agreement was not within these agreed limits, the analysis files were examined and amended where appropriate. Following an interim analysis, it also became apparent that, when a considerable amount of data had been collected, selection of the optimal data for analysing R_{aw} with respect to achievement of body temperature barometric pressure and saturated with water conditions (BTPS) and minimal drift was difficult without some reference to quality control at the time of data collection. It was agreed, therefore, that the test centre should indicate the best epochs of data from which to select data for analysis prior to sending the data to the test centre for reanalysis.

RR and V_T were reported as the mean of 25 breaths, collected in five separate epochs, each consisting of five breaths immediately preceding the first five partial forced expiratory manoeuvres. FRC_{pleth} was reported as the mean of all technically acceptable occlusions, where the infant was occluded at end inspiration or within 10% of the start of expiration. The calculated value from each occlusion represented the mean of one to three complete respiratory efforts, using both inspiratory and expiratory excursions. Data samples within 5% of peak airway opening pressure (P_{ao}) (*i.e.* the end expiratory plateau) were excluded, thus limiting analysis of the relationship between plethysmographic volume and P_{ao} to periods of rapid change. The subsequent correction to FRC was performed by subtracting the volume occluded above end expiratory volume and the apparatus dead space from the total occluded gas volume. If a degree of glottic closure was present, it was permissible to cautiously "truncate" the measured effort, measuring over a shorter period, *e.g.* between 95 to 20% peak to trough thereby excluding portions where changes in P_{ao} did not reflect changes in mean alveolar pressure. A minimum of three occlusions were initially required to report results, but this requirement was relaxed following the interim analysis to allow reporting of otherwise good quality data, providing that at least two technically satisfactory measurements within 10% each other were obtained.

Data for analysis of R_{aw} were considered to be technically acceptable when flow and plethysmographic pressure (P_{pleth}) were in phase on a time-series axis, and closed at points of zero flow when viewed on an X–Y plot. Default strategies for reporting values of R_{aw} using the RASP software allowed selection of data on a breath-by-breath basis. R_{aw} was reported as a mean of 5–7 breaths at 50% maximum tidal flows during the beginning of inspiration and end of expiration for each selected breath. Although measurements of R_{aw} were made only after thermal equilibrium was reached, signals could be adversely affected by changing atmospheric conditions, which at times led to

drift even when adequate time had been allowed for stable conditions to be met.

$V'_{max,FRC}$ values were reported both as the highest, and as the mean of the four highest, values of technically acceptable data. The maximal and mean jacket pressures (P_J) were noted, together with the pressure during the best manoeuvre (optimal P_J), % P_J transmission [8], and the total number of manoeuvres performed.

Where technically satisfactory measurements were not obtained or data were not collected, the reasons for failure were noted.

Analysis of within-infant, intercentre differences. Infant details and results were double-entered on an Excel spreadsheet (version 5.0, Microsoft Corporation, Redmond, WA, USA) and checked for data transcription errors, by both the test centre and the analysis centre independently, before comparing results between centres. The aim was to minimize differences in measured parameters that were not due to biological variability alone, for example data transcription and entry errors, data selection bias or differing analysis strategies.

Reporting of results

A short clinical report, based on the results obtained from the test centre, was sent to the referring paediatrician and infant's family doctor as soon as possible after testing.

The definitive results with respect to comparisons of respiratory function between the ECMO and CM groups, which are described elsewhere [2, 12], were reported from the analysis centre, *i.e.* the centre that did not test the infant, and therefore had no contact with the family. This ensured complete blinding to management status. The management status at trial entry of the whole group of infants was not revealed until both data collection and analysis was complete, when background details of surviving infants, including a description of their status at trial entry and management (ECMO or CM), were obtained from the ECMO trial data coordinator.

Statistical methods

Comparison of infants studied in the two different centres. For comparisons between the two groups of infants studied at ICH and LEIC, unpaired t-tests were used to examine background details and respiratory function results where data were normally distributed. The Mann-Whitney U-test was used to assess differences between groups of non-normally distributed data. A p-value <0.05, and 95% confidence intervals (CI) of the differences between groups or paired data not encompassing zero, were considered significant.

Comparison of within-subject, intercentre differences. Intercentre, within-subject differences, which would reflect variation in analytical techniques, were evaluated according to the method of BLAND and ALTMAN [13], by calculating the mean and standard deviation (SD) of the difference between pairs of measurements (ICH - LEIC). Results were displayed graphically by plotting the differences between pairs against their mean, together with the group mean difference and limits of agreement (±2SD of the within pair differences).

Results

Demographic results

Of the 101 infants who survived to 1 yr, two were withdrawn from the main trial by their parents, eight infants were excluded from respiratory follow-up for primarily medical reasons: five due to congenital heart disease, one due to gross developmental impairment and two who were still in hospital at approximately 1 yr of age. Eight infants did not attend due to parental refusal, two because of the parents' inability to attend before the child was 18 months old, two due to repeated (more than three occasions) cancellations following onset of respiratory tract infections <3 weeks before their appointment and one due to adverse social circumstances. Thus, respiratory follow-up was possible in 78 (77%) of all survivors.

Between August 1994 and December 1996, of the 78 infants who took part in the respiratory follow-up with their parents, 42 attended ICH and 36 attended LEIC. Those who participated in the respiratory follow-up were similar to the whole group of survivors with respect to gestational age, birth weight, sex distribution, primary diagnosis, age and severity of disease at trial entry [1, 2, 12] (data not shown). Details of the infants attending each centre are summarized in table 2. The age, weight and length of infants attending each test centre were similar, as were birth weight and gestational age. The reported prevalence of maternal tobacco-smoking during pregnancy and

Table 2. – Infant characteristics according to test centre

	ICH n=42	LEIC n=36	95% CI of the group mean difference (ICH - LEIC)
Age months	13.6 (1.8)	13.9 (2.0)	-1.1, 0.6
Weight kg	9.9 (1.5)	10.5 (1.6)	-1.2, 0.2
Length cm	78.1 (3.5)	78.4 (3.6)	-2.0, 1.2
Birthweight g	3378 (645)	3521 (468)	-400, 116
Gestation weeks	39.3 (2.4)	39.6 (2.0)	-1.4, 0.7
Male n (%)	24 (57)	24 (67)	-31, 12
Caucasian n (%)	32 (76)	28 (78)	-20, 17
Smoking during pregnancy n (%)	7 (17)	4 (11)	-10, 21
Current maternal smoking n (%)	13 (31)	12 (33)	-23, 18
Any smoke exposure n (%)	29 (69)	17 (47)	-0.1, 44
Current respiratory medicine n (%)	10 (24)	7 (19)	-14, 23
Symptomatic at testing n (%)	3 (7)	9 (25)	-34, -2*
Recent URTI n (%)	7 (17)	19 (53)	-57, -15***
Normal medical exam n (%)	29 (69)	29 (81)	-31, 8

For definition of groups, see table 1. Values are given as mean and SD. CI: confidence interval; URTI: upper respratory tract infection. *: p∂0.05; ***: p∂0.001

Table 3. – Number and percentage of successful measurements at each centre

	Test centre		
	ICH n=42	LEIC n=36	95% CI (ICH - LEIC)
FRC_{pleth}	39 (93)	33 (92)	-10, 13
R_{aw}	35 (83)	21 (58)	5, 45*
$V'_{max,FRC}$	37 (88)	34 (94)	-13, 9
Tidal breathing parameters	41 (98)	36 (100)	-7, 3

For definition of groups, see table 1. Values are shown as n (%). FRC_{pleth}: functional residual capacity by plethysmography; R_{aw}: airway resistance; $V'_{max,FRC}$: maximum expiratory flow at functional residual capacity; CI: confidence interval. *: $p<0.05$.

postnatally was similar at each centre, although a greater proportion of infants attending ICH were reported to be exposed to tobacco smoke from any source (parental and/or other regular exposures) at the time of follow-up. More infants who attended LEIC were symptomatic at testing (respiratory symptoms included rhinorrhea, coryzal illness, cough and wheeze) and had suffered recent upper respiratory tract infections. The majority of infants were, however, coded as normal following examination by a paediatrician blind to management status.

Causes of failure

The percentage of successful measurements for each outcome parameter, grouped by test centre is summarized in table 3. Tidal parameters were reported for all LEIC infants, and all but one ICH infant (who woke before measurements of $V'_{max,FRC}$ were attempted). Most infants (93%) had their resting lung volume (FRC_{pleth}) successfully measured, in similar proportions at each test centre. However, a significantly smaller proportion of infants measured in Leicester had successful R_{aw} results reported (p=0.014). For infants attending ICH, the failure to report FRC_{pleth} or R_{aw} was usually due to waking before measurements were complete (n=5). In a further three infants, R_{aw} could not be reported for technical reasons, signals being adversely affected by weather or building work in the area. In contrast, only one infant in LEIC woke during plethysmographic measurements of R_{aw}, but FRC_{pleth} was unsuccessful in three infants (one technical failure, two attempts invalidated by glottic closure during expiratory efforts), and R_{aw} failed in 14 infants. The latter was due to 10 technical failures, subsequent to the introduction of a new amplifier partway through the respiratory follow-up (seven occasions) or insufficient humidification in the rebreathing bag (three occasions), and one failure for physiological reasons (glottis closure). The remaining three measurements failed because FRC_{pleth} was unsuccessful, as noted previously, so that the occlusion data could not be used to calibrate R_{aw} data. Successful measurements of $V'_{max,FRC}$ were achieved in 37 (88%) infants attending ICH and 34 (94%) from LEIC. At ICH, one infant woke, in three others there was persistent early inspiration during each manoeuvre, and in the remaining infant, suboptimal jacket pressures were applied, such that flow limitation was not demonstrated. In the two LEIC infants in whom measurements were unsuccessful, the jackets were of a poor fit, resulting in late application of thoraco-abdominal pressures in all manoeuvres, thereby failing to meet quality control criteria with respect to jacket inflation time.

Comparison of respiratory function in infants studied in the two centres

There were no significant differences for any of the outcome measures, *i.e.* FRC_{pleth}, R_{aw} or $V'_{max,FRC}$ between the infants studied at the two centres (table 4) The group mean (range) FRC_{pleth} was 284 (160–450) mL for infants attending ICH, and 281 (145–492) mL for those attending LEIC. The number of FRC_{pleth} manoeuvres reported by each centre following the cross-centre analysis, expressed as median (range) was 4 (2–9) for ICH analysis of LEIC data and 7 (2–12) for LEIC analysis of ICH data. However, only one infant from each centre had their lung volume reported on only two manoeuvres. When corrected for weight, the group mean (range) FRC_{pleth} of infants attending ICH was 29 (17–46) $mL \cdot kg^{-1}$ and that at LEIC 27 (15–49) $mL \cdot kg^{-1}$. Although airway resistance tended to be higher amongst those infants attending ICH, these differences did not reach significance (95% CI (ICH - LEIC) for initial inspiratory R_{aw} ($R_{aw,II}$) and late expiratory R_{aw} ($R_{aw,EE}$) 0.14, 0.77 and -0.15, 1.12 $kPa \cdot L^{-1} \cdot s$ respectively).

$V'_{max,FRC}$, whether expressed as the highest value obtained or as the mean of the four highest technically acceptable manoeuvres, was similar for infants attending both test centres. The mean (range) $V'_{max,FRC}$ was 171 (48–415) $mL \cdot s^{-1}$ at ICH and 145 (27–320) $mL \cdot s^{-1}$ at LEIC, with similar values for the highest technically acceptable $V'_{max,FRC}$ when expressed as mean (table 4). However,

Table 4. – Comparison of within-subject respiratory function results analysed by each centre

Test centre	ICH n=42		LEIC n=36		Mean (95% CI) (Analysis centre - test centre) n=78
Analysis centre	ICH	LEIC	ICH	LEIC	
FRC_{pleth}	286 (58)	284 (59)	281 (73)	277 (67)	1.11 (-0.53, 2.75)
$R_{aw,II}$ $kPa \cdot L^{-1} \cdot s$	2.19 (0.79)	2.30 (0.89)	1.98 (0.76)	1.86 (0.84)	0.04 (-0.02, 0.10)
$R_{aw,EE}$ $kPa \cdot L^{-1} \cdot s$	2.91 (1.22)	2.93 (1.18)	2.45 (1.12)	2.32 (1.16)	-0.02 (-0.12, 0.08)
Mean $V'_{max,FRC}$ $mL \cdot s^{-1}$	168 (73)	171 (73)	145 (75)	146 (75)	0.91 (-0.94, 2.77)
Best $V'_{max,FRC}$ $mL \cdot s^{-1}$	178 (77)	176 (80)	158 (82)	156 (80)	1.43 (-1.42, 4.28)
V_T mL	94 (17)	94 (17)	83 (16)	83 (16)	0.04 (-0.57, 0.64)
RR min^{-1}	33 (7)	33 (8)	33 (10)	33 (10)	-0.13 (-0.39, 0.14)

For definition of groups, see table 1. Values are given as mean (SD). For definition of abbreviations, see table 3; $R_{aw,II}$: airway resistance during initial inspiration; $R_{aw,EE}$: airway resistance during late expiration; V_T: tidal volume; RR: respiratory rate.

despite the similarity of results between the two centres, the applied jacket pressures were significantly higher at ICH than at LEIC; mean (SD) 4.8 (1.7) kPa and 2.6 (0.9) kPa, respectively ($p<0.001$), reflecting the lower percentage of jacket pressure transmitted to the infant at ICH (29% *versus* 58% at LEIC). The median number of manoeuvres was 18 (range 8–40) at each centre.

Tidal breathing parameters were not selected outcome measures, but were recorded and analysed to provide additional evidence for quality control and baseline information regarding respiratory function. The respiratory rate was not significantly different between centres, the mean (range) was 33 (19–53) min^{-1} at ICH and 33 (24–74) min^{-1} at LEIC. However, the group mean V_T was larger for infants attending ICH than LEIC; mean (range) 94 (64–134) and 83 (52–121) mL respectively, 95% CI; 3–18 mL ($p< 0.05$). These differences persisted when corrected for weight.

Comparison of within-subject respiratory function results analysed by each centre

In addition to comparing results from infants studied at the two measurement centres, the within-infant inter-observer variability between centres was also assessed. As stated previously, the aim was to obtain values within 10% of each other for tidal breathing parameters, $V'_{max,FRC}$ and FRC_{pleth}, and within 20% for R_{aw}. The results of the cross-analysis of each centre's data are shown in table 4.

FRC_{pleth}. FRC_{pleth} analyses were generally within 10% agreement. On initial comparison of the results from the two centres, discrepancies >10% occurred in 12 infants. In four of these infants, this was due to data collection in which it had not been possible to report values based on at least three occlusions, either due to the infant waking early (one infant), or insufficient collection of end inspiratory occlusion data (three infants). Following the revision of guidelines and reanalysis by each centre, the definitive results reported were all within the agreed limits, although it was necessary to report FRC_{pleth} as the mean of only two manoeuvres in two infants. In the remaining eight infants, discrepancies were due to data transcription errors (two infants), different epochs of data being selected for analysis (one infant) and differing analysis strategies, which were mainly related to "truncating" glottic expiratory efforts by varying amounts (five infants). All were readily resolved after critical inspection of data by each centre, so that the final agreement for FRC_{pleth} was within 10% all infants. The between-centre agreement for FRC_{pleth} is shown in figure 1.

Airway resistance. For R_{aw} analyses, there were initially discrepancies of >20% in 10 infants. These were due to operator errors in selecting analysis options within the software that did not adhere to the agreed protocol (three infants) or selection of different epochs of data for analysis and reporting (seven infants), *i.e.* biological variability. Of the infants in whom R_{aw} measurements were not reported, one of three technical failures at ICH (not meeting quality control criteria) was failed by LEIC alone, whereas the remaining two were failed by both centres. Similarly, four of 14 technical failures at LEIC were failed by

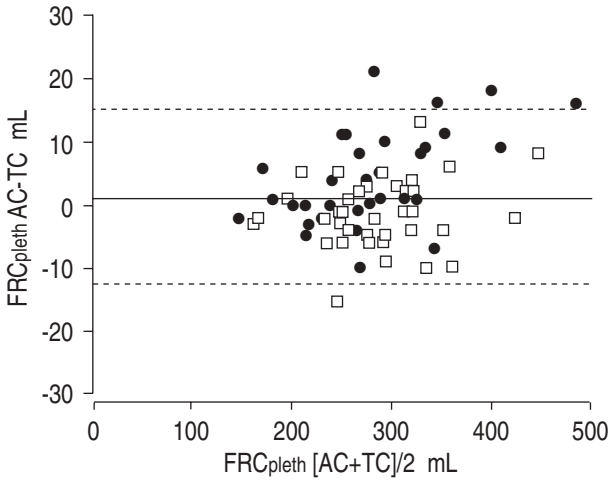

Fig. 1. – Scatter plot, according to the method of BLAND and ALTMAN [13], of the differences between functional residual capacity (FRC), analysed at each centre, and their mean. AC: analysis centre; TC: test centre; pleth: plethysmography. ——: mean between-centre difference; - - - -: 95% limits of agreement of the mean difference. Infants tested at Institute of Child Health (□); and Leicester Royal Infirmary (■).

ICH alone and the remainder by both centres. After inspection and amendment of the analysed data by each centre, agreement was improved to within 20% in all but three of the 56 infants. The between-centre agreement for R_{aw} calculated during initial inspiration is displayed in figure 2.

$V'_{max,FRC}$. Following the interim analysis, when a common volume baseline filtering strategy was adopted by both centres, and it was agreed to establish the FRC level by regression through the end expiratory points preceding the RTC manoeuvre, the agreement was generally good for $V'_{max,FRC}$ analyses. Initially, analyses of $V'_{max,FRC}$ were outside the agreed limits of 10% for six infants. One was due to a data transcription error, two to the selection of different epochs of data from which the best four manoeuvres were reported, one was attributed to volume baseline filtering strategies, whereas data from two infants

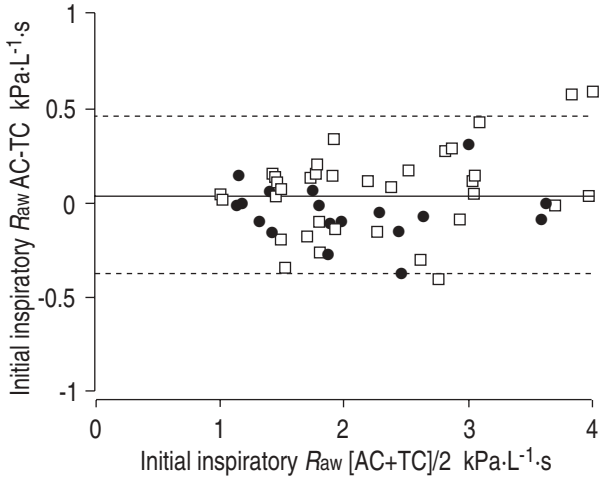

Fig. 2. – Scatter plot, according to the method of BLAND and ALTMAN [13], of the differences between initial inspiratory airway resistance (R_{aw}), analysed at each centre, and their mean. AC: analysis centre; TC: test centre. ——: mean. - - - -: 95% limits of agreement of the mean difference. Infants tested at Institute of Child Health (□); and Leicester Royal Infirmary (■).

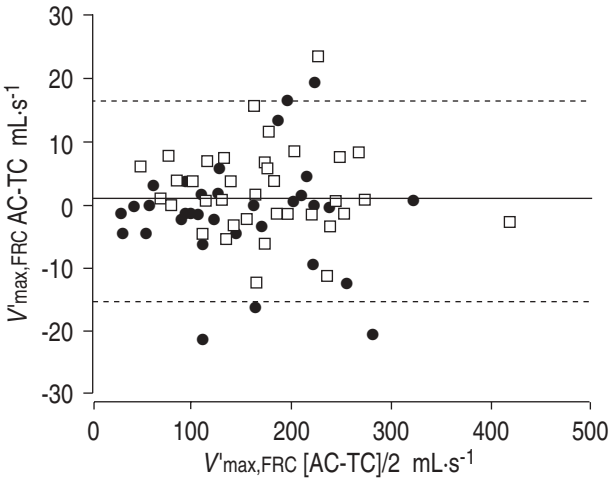

Fig. 3. – Scatter plot, according to the method of BLAND and ALTMAN [13], of the differences between maximum expiratory flow at functional residual capacity, analysed at each centre, and their mean. AC: analysis centre; TC: test centre. ——: mean. - - - - : 95% limits of agreement of the mean difference. Infants tested at Institute of Child Health (□); and Leicester Royal Infirmary (■).

required recalibration to adjust a zero offset of the flow signal before a good agreement could be achieved. Following inspection and amendment of the analysed data by each centre, three infants remained outside the agreed limits. Two of these had very low, flow-limited values of $V'_{max,FRC}$, so that although the discrepancy between laboratories exceeded 10%, absolute values from the two centres were within 6 mL·s^{-1} of each other. The poor between-centre agreement in the remaining infant (20%) arose from the inclusion of data by one centre, but not the other, of one manoeuvre where instability of breathing patterns made establishment of the FRC level very difficult to determine. The between-centre agreement for within-subject values of $V'_{max,FRC}$ is displayed in figure 3.

For tidal breathing parameters, between-centre differences >10% were seen in three infants, all of which were due to selection of different epochs of data (biological variability). A closer agreement was obtained by selecting alternative epochs of quiet breathing for analysis.

Discussion

One of the major advantages of the respiratory follow-up study design was that the analysis of respiratory function in survivors recruited from the ECMO trial could be made by operators who were fully blinded to their management status. However, before combining data from each centre, it was necessary to ensure that there was no bias either within or between laboratories that might confound interpretation of the results. It was necessary, therefore, to assess the agreement between centres with respect to both the background and respiratory function data.

Population

Infants and their families attended the respiratory follow-up from all over the UK. The majority of infants attending each test centre were of Caucasian ethnicity. Amongst the remaining ethnic groups, there were five infants of Afro-Caribbean origin, all of whom attended ICH, whereas more Asian infants attended LEIC than ICH. However, when examined according to management group (ECMO *versus* CM) there were similar proportions of infants from each ethnic background within each group [12].

The prevalence of maternal smoking during pregnancy was similar at each centre, but lower than that reported elsewhere [14, 15]. In particular, of those attending ICH, the proportion of infants reported to be exposed to intrauterine tobacco was 17% (seven of 42) compared with 43% of infants in a recent East London-based prospective study [16] who attended ICH for similar measurements during an overlapping time period. This suggests that there may have been under-reporting of maternal smoking during pregnancy by mothers of the respiratory follow-up population at each test centre. The suggestion is reinforced by the increase in maternal smoking at the time of the laboratory visit at around 1 yr, when 31 and 33% (ICH and LEIC, respectively) of infants were reported to be exposed: thereby approaching other reported prevalences [14, 15].

Although efforts were made to avoid testing infants within 3 weeks of onset of upper respiratory tract infection (URTI) symptoms at ICH and within 4 weeks at LEIC, this was not always possible, particularly amongst infants who suffered recurrent respiratory infections, or when long-distance travel arrangements had been made. The proportion of infants who had had recent symptoms of URTI was greater at LEIC, which may have reflected the long symptom-free time period stipulated at that centre. However, a greater proportion of infants were also symptomatic at testing in LEIC (25% *versus* 7%). Local laboratory practice with respect to booking and, if necessary, postponing appointments would also have influenced these findings. There was no evident correlation between recent URTI symptoms and reported values of R_{aw}, with the 18 infants who had symptoms of recent URTI and successful measurements of R_{aw} having values that ranged 1.10–3.68 kPa·L^{-1}·s^{-1} during initial inspiration, the phase of respiration most likely to be adversely affected by such infections. In future, greater standardization with respect to the duration of a symptom-free period should be adhered to. However, in clinical follow-up studies such as this, it may be preferable to emphasize the absence of current symptoms, rather than a prolonged symptom-free period, since the latter is difficult to achieve in a population prone to recurrent infections.

Evaluation of methodology

To our knowledge, this is the first report of a multicentre study using established methods of infant lung function testing where results have been included as outcome measures. Within the field of infant respiratory function assessment, considerable variability exists with respect to equipment and techniques. However, efforts have recently been directed towards standardizing the nomenclature [17, 18], measurement conditions [19] and assessments of respiratory function in infants [7, 20].

One of the unique features of the respiratory follow-up was the opportunity to critically evaluate variability within infants for technically complex and operator controlled

analysis, which was only partly automated, such that selection of data and analysis strategy could influence results markedly. Ideally, an assessment of analysis variability would have preceded this study, but limited time and funding precluded this approach. Operator selection of the most suitable data for analysis, the relatively low signal-to-noise ratio of respiratory signals in infants, and the inherent biological variability of certain parameters all increase within and between-subject variability in this age group. Consequently, a careful selection of a limited number of relevant outcome measures, together with recruitment of adequate numbers of infants, is essential to achieve an adequate power of study [7].

The aim of the study design was to reduce both systematic and random errors and prevent measurement and analysis bias affecting reported results. Statistical assessment of differences between the two groups of infants studied at ICH and LEIC aimed to detect any sampling differences, for example in background characteristics, management status, disease severity, and any measurement bias arising from differences in apparatus or technique. The statistical assessment of the whole group of infants, analysed by each centre, in contrast, aimed to highlight the biological variability together with the inter-observer variability in the analytical approach and data selection.

Careful assessment and evaluation of both equipment and measurement and analysis techniques, with subsequent amendments, led to improved between-centre agreement. With both laboratories also concurrently involved in other studies, limitations were, however, placed on the extent to which adaptations could be made to existing equipment and protocols.

In vitro assessment of both plethysmographs, using the same copper test lung, showed that lung volume was measured accurately at both centres. The effective dead space that the infant breathed through was greater at ICH, mainly due to the different design of PNT and shutter configuration, but also due to the different volume subtracted for effective dead space of the face mask. The latter may have been due to the application of larger quantities of therapeutic putty at LEIC. Differences in tidal volume between centres may have been attributable to either dead space or design of PNT. In a small group of infants, tidal volume and end tidal CO_2 were measured using PNTs of both designs on the same measurement occasion (data not shown). CO_2 values were similar, but V_T was larger using a Fleisch "1" compared with a Hans Rudolph 100 PNT. This presumably reflects the fact that any small increase in effective dead space will be compensated for by an increased V_T to bring end tidal CO_2 back to normal values. Despite the effect on tidal volume, the PNT design did not appear to influence between centre interpretation of the selected outcome measures of respiratory function and lung volume. However, for future collaborative studies, we would try to use equipment that was of as low a dead space and as similar as possible.

The linearity and resistance of measurement equipment used at each centre was assessed and found to be adequate over the range of flows and pressures used in these measurements. Had the study involved healthy infants with correspondingly higher expiratory flows during forced expiration measurements, a PNT with greater range may well have been required. Future studies will be able to use a new generation of screen PNTs, which have a considerably lower dead space but a greater linear range.

Prior to each study, known signals, including the functional time constant (*i.e.* combined mechanical and thermal) of each plethysmograph were recorded and checked at each centre, and these data were exchanged between centres. The weight and length of each infant were recorded, using regularly calibrated equipment, thereby avoiding any systematic errors in anthropometry at or between centres. The influence of jacket style during measures of forced expiration at each centre on the range of pressures used to obtain $V'_{max,FRC}$ was addressed. The Hammersmith jacket used at LEIC was almost twice as efficient at transmitting pressure to the pleura when compared with the Hannover jacket used at ICH. Consequently, higher jacket inflation pressures were used at ICH. Despite these methodological differences, the mean and wide range of forced expiratory flows obtained at each centre suggested that the driving pressures (the product of jacket inflation pressure and jacket efficiency plus elastic recoil of thoracic structures) used at each centre were similar [8].

The success rate of measurements at each test centre was similar, with the exception of R_{aw}, where the increased failure rate at LEIC was mainly attributable to untimely problems with a replacement amplifier which adversely affected the quality of the plethysmographic pressure signals in a number of infants. The lack of adequate commercially available equipment for infant respiratory function tests, with many centres relying on custom built systems, contributes to the problems of access to essential spare parts. An additional component of study design that is a prerequisite for future similar trials is thus the need for equipment to be standardized during the trial period, and to have ready access to replacement parts, so that no disruption to data collection occurs during critical time frames.

Success rates of infant respiratory function parameters are rarely reported, but compared with a younger group of infants measured at ICH [5, 21], in whom 85% of measurements of R_{aw} were successful, a similar rate of 83% was reported in the ICH respiratory follow-up population. A significant number of plethysmographic measurements at ICH were lost when the infants woke before completion of the protocol, despite increasing the usual dose of sedation from 100 to 150 mg·kg^{-1} triclofos sodium (equivalent to 100 mg·kg^{-1} chloral hydrate) when necessary. This was still somewhat lower than maximum doses used at LEIC, which may explain the fewer failures due to early awakening at the latter. A number of infants recruited from the ECMO trial were >1 yr of age, taking longer to settle into quiet sleep and waking earlier. The time available for measurements during sedated sleep is short, and this limitation should be borne in mind when designing future protocols that include infant respiratory function. However, with similar numbers of infants attending each centre, and similar overall failure rates at both, predominance of one centre when reporting and interpreting results was avoided.

During the course of this study, the difficulties in imposing appropriate criteria to maintain stringent quality control within clinical trials became apparent, particularly when attempts were being made to follow up an entire cohort, as in this study. In such circumstances, the individual data from each infant are unique and cannot be simply replaced by recruiting additional subjects. These difficulties led to our decision to relax standard criteria, which

normally specify a minimum of three technically acceptable measures before reporting FRC_{pleth}, to allow the mean of two such measures to be reported, providing that they were within 10%. Controversy remains over the number of measurements that should be used to report the mean for the various respiratory parameters commonly measured in infants [7]. Further objective evaluation is needed, so that the time required for data collection can be minimized in infants without compromising accuracy.

The process of exchanging raw data on computer disks between centres was a valuable opportunity to evaluate approaches to analysis critically using software that although identical at each centre, could be readily customized to user preference with respect to collection, display, calibration, filtering and reporting of results. Exchanging data on disk between centres highlighted the need for meticulous attention to quality control during both calibration and data collection. The effect, for example, of different strategies of drift correction and determination of volume baselines prior to analysis of $V'_{max,FRC}$ (regression over five breaths *versus* selection of the most representative end expiratory level from a single breath over this period) on within-infant between-centre analysis, was of initial concern although readily standardized.

Approximately 10% of analyses needed to be re-examined when the cross-analysis between centres showed a poor agreement. Although, in several cases, differences were due to the selection of data, there were also four apparent random errors during analysis and reporting of the results. We speculate that most of these errors would not have been detected had exchange of data and comparison of analyses not taken place, thus emphasizing the need to develop software where such transmission errors are minimized.

The design of the respiratory follow-up served to increase confidence in the respiratory follow-up findings and was readily incorporated within the existing data collection and analysis protocol at each test centre. During the current study, prior commitment to other ongoing studies at each centre compromised full standardization, but in future similar trials, even greater consistency could be achieved. This will require the development of improved software and measurement equipment and adherence to mutually agreed international standards, which are currently evolving (European Respiratory Society/American Thoracic Society Task Force on Infant Respiratory Function). Some of the essential features required for successful design of future multicentre trials in this field are summarized in table 5.

The selective use of specialized, automated and validated software potentially could speed data analysis and reduce within-subject and inter-observer variability [22]. Unfortunately, reproducibility does not equate to accuracy, and arbitrary exclusions to achieve a low within-subject variability may not provide a faithful representation of respiratory function, especially when a marked biological variability is present, as may occur in certain disease states. Some compromise is, therefore, needed, whereby obvious artefacts are excluded while still retaining physiologically meaningful data. While a fully automated system could ensure virtually identical results on reanalysis in a multicentre setting that would be operator-independent, infant signals, with their high variability and low signal-to-noise ratio, often require an interactive operator input for correct interpretation. Consequently, it is unlikely that assessments of respiratory function in infants will ever become as routine as those in older subjects, and that operators with extensive training and experience will always be required if successful measurements are to be achieved.

Table 5. – Methodological basis for mulitcentre trials

Comparison, assessment and standardization of equipment
Pilot study to assess accrual rate and validate protocol
Power of study calculation, based on the number of outcome measures, knowledge of within-subject variability of each parameter and estimated sample size
Frequent inter-laboratory visits during trial period
Identical measurement and analysis protocol using similar software
Exchange of raw data and cross-analysis of a subset of the data
Examination of population characteristics for between-centre bias.

Conclusions

In conclusion, a multicentre approach to trials with parameters of infant respiratory function as outcome measures appears feasible, provided that close attention is paid to study design, and participants in such trials maintain a standard approach to data collection and analysis. Collaboration of this nature, despite requiring considerable time and effort, is of great benefit with respect to quality control within, as well as between, specialized infant lung function testing centres. Furthermore, there is increasing recognition of the need for multicentre studies of infant respiratory function if we are to achieve sufficient statistical power to address clinically relevant questions regarding respiratory disease during early childhood over a realistic time period.

Acknowledgements: The authors thank the Wellcome Trust for financial support for the respiratory follow-up and D. Elbourne, A. Johnson, A. Wrotchford, K. Enock and C. Harris of the National Perinatal Epidemiology Unit for their help and advice. Thanks also to R. Tasker, D. Macrae, A. Goldman, O. Crawford, A. Fenton, S. Bohin and K. Makkonen for their help with clinical assessment of infants. Finally, the authors are very grateful to the infants and their parents who took part in this study.

References

1. UK Collaborative ECMO Trial Group, Field D, Davis C, Elbourne D, Grant A, Johnson A, *et al*. UK collaborative randomised trial of neonatal extracorporeal membrane oxygenation. *Lancet* 1996; 348: 75–82.
2. Dundas I, Beardsmore CS, Enoch K, Poole K, Stocks J. Respiratory function at 1 year in survivors of UK ECMO Trial. *Am J Respir Crit Care Med* 1997; 155: A237 Abstract.
3. Gaultier C, Fletcher ME, Beardsmore C, England S, Motoyama E. Respiratory function measurements in infants: measurement conditions. *Eur Respir J* 1995; 8: 1057–1066.

4. Beardsmore CS, Thompson JR, Williams A, *et al.* Pulmonary function in infants with cystic fibrosis: the effect of antibiotic treatment. *Arch Dis Child* 1994; 71: 133–137.
5. Dundas I, Dezateux CA, Fletcher ME, Jackson EA, Stocks J. Comparison of single-breath and plethysmographic measurements of resistance in infancy. *Am J Respir Crit Care Med* 1995; 151: 1451–1458.
6. Prechtl HFR. The behavioural states of the newborn infant (a review). *Brain Res* 1974; 76: 185–212.
7. Stocks J, Sly PD, Tepper RS, Morgan WJ. Infant Respiratory Function Testing. New York, Wiley, 1996.
8. Stick S, Turner D, LeSouëf PN. Transmission of pressure across the chest wall during the rapid thoracic compression technique in infants. *J Appl Physiol* 1994; 76: 1411–1416.
9. Stocks J, Levy NM, Godfrey S. A new apparatus for the accurate measurement of airway resistance in infancy. *J Appl Physiol* 1977; 43: 155–159.
10. Stocks J, Nothen U, Sutherland P, Hatch D, Helms P. Improved accuracy of the occlusion technique for assessing total respiratory compliance in infants. *Pediatr Pulmonol* 1987; 3: 71–77.
11. Dezateux CA, Stocks J, Dundas I, Jackson EA, Fletcher ME. The relationship between tPTEF:tE and specific airways conductance in infancy. *Pediatr Pulmonol* 1994; 18: 299–307.
12. Makkonen K, Wrotchford A, Dundas I, Beardsmore CS. Respiratory morbidity of survivors of the UK ECMO trial: a one year follow up. *Eur Respir J* 1997; 10: 305s Abstract.
13. Bland JM, Altman DJ. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 307–310.
14. Dezateux C, Delves HT, Stocks J, Wade A, Pilgrim L, Costeloe K. Urinary antimony in infancy. *Arch Dis Child* 1997; 76: 432–436.
15. Trigg CJ, Bennett JB, Tooley M, Sibbald B, D'Souza MF, Davies RJ. A general practice based survey of bronchial hyperresponsiveness and its relation to symptoms, sex, age, atopy, and smoking. *Thorax* 1990; 45: 866–872.
16. Dezateux C, Stocks J, Dundas I, Pilgrim L, Fletcher M. Premorbid infant respiratory function and wheezing in the first year of life: a population based study in London. *Eur Respir J* 1996; 9: 466s Abstract.
17. Quanjer PH, Sly PD, Stocks J. Uniform symbols, abbreviations and units in pediatric pulmonary function testing. *Pediatr Pulmonol* 1997; 24: 2–11.
18. Quanjer PH, Sly PD, Stocks J. Respiratory function measurements in infants: symbols, abbreviations and units. Report of working group ERS/ATS statement. *Eur Respir J* 1995; 8: 1039–1056.
19. Gaultier C, Fletcher M, Beardsmore C, Motoyama E. Respiratory function measurements in infants: measurement conditions. *Am J Respir Crit Care Med* 1995; 151: 2058–2064.
20. American Thoracic Society/European Respiratory Society. Respiratory mechanics in infants: physiologic evaluation in health and disease. *Am Rev Respir Dis* 1993; 147: 474–496.
21. Dezateux C, Fletcher ME, Dundas I, Stocks J. Infant respiratory function after RSV-proven bronchiolitis. *Am J Respir Crit Care Med* 1997; 155: 1349–1355.
22. van der Ent CK, Brackel HJL, Mulder P, Bogaard JM. Improvement of tidal breathing pattern analysis in children with asthma by on-line automatic data processing. *Eur Respir J* 1996; 9: 1306–1313.