ORIGINAL ARTICLE

New reference ranges for interpreting forced expiratory manoeuvres in infants and implications for clinical interpretation: a multicentre collaboration

Sooky Lum,[1] Vassiliki Bountziouka,[1] Angela Wade,[2] Ah-Fong Hoo,[1,3] Jane Kirkby,[1,3] Antonio Moreno-Galdo,[4] Ines de Mir,[4] Olaia Sardon-Prado,[5,6] Paula Corcuera-Elosegui,[5] Joerg Mattes,[7,8] Luis Miguel Borrego,[9,10] Gwyneth Davies,[1,3] Janet Stocks[1]

▶ Additional material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/thoraxjnl-2015-207278).

For numbered affiliations see end of article.

Correspondence to
Dr Sooky Lum, Respiratory, Critical Care & Anaesthesia Section in IIIP Programme, UCL, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK;
s.lum@ucl.ac.uk

Received 7 May 2015
Revised 23 July 2015
Accepted 29 September 2015
Published Online First
2 November 2015

ABSTRACT

The raised volume rapid thoracoabdominal compression (RVRTC) technique is commonly used to obtain full forced expiratory manoeuvres from infants, but reference equations derived from 'in-house' equipment have been shown to be inappropriate for current commercially available devices.

Aim To explore the impact of equipment differences on RVRTC outcomes, derive robust equipment-specific RVRTC reference ranges and investigate their potential clinical impact on data interpretation.

Method RVRTC data from healthy subjects using Jaeger BabyBody or the 'Respiratory Analysis Software Program, RASP' systems were collated from four centres internationally. Data were excluded if gestational age <37 weeks or birth weight <2.5 kg. Reference equations for RVRTC outcomes were constructed using the LMS (lambda–mu–sigma) method, and compared with published equations using data from newborn screened infants with cystic fibrosis (CF).

Results RVRTC data from 429 healthy infants (50.3% boys; 88% white infants) on 639 occasions aged 4–118 weeks were available. When plotted against length, flows were significantly higher with RASP than Jaeger, requiring construction of separate equipment-specific regression equations. When comparing results derived from the new equations with those from widely used published equations based on different equipments, discrepancies in forced expiratory volumes and flows of up to 2.5 z-scores were observed, the magnitude of which increased with age. According to published equations, 25% of infants with CF fell below the 95% limits of normal for $FEV_{0.5}$, compared with only 10% when using the new equations.

Conclusions Use of equipment-specific prediction equations for RVRTC outcomes will enhance interpretation of infant lung function results; particularly during longitudinal follow-up.

Key messages

What is the key question?

- To what extent are results derived from the raised volume technique for assessing forced expiratory manoeuvres in infants misinterpreted due to use of inappropriate reference equations?

What is the bottom line?

- Equipment-specific reference equations are essential for accurate interpretation of lung function in early life since use of equations derived from different infant lung function devices may lead to significant misclassification of lung disease.

Why read on?

- The extent of bias that may be introduced by using reference ranges derived from different equipment emphasises the need to use equipment-specific reference ranges when interpreting infant lung function results to permit appropriate tracking of lung health across the early years.

Linked

▶ http://dx.doi.org/10.1136/thoraxjnl-2015-207911

CrossMark

To cite: Lum S, Bountziouka V, Wade A, *et al*. *Thorax* 2016;**71**:276–283.

INTRODUCTION

During the past 20 years, the raised volume rapid thoracoabdominal compression (RVRTC) technique for obtaining forced expiratory flow volume (FEFV) manoeuvres has been shown to discriminate clearly between healthy infants and those with lung disease.[1–8] However, replacement of 'in-house' infant lung function (ILF) equipment by commercial devices has complicated data interpretation. In a recent survey on the role of ILF tests in clinical practice, 77% of responders were using RVRTC, most of whom interpreted results using reference data published in 2000[9] prior to introduction of commercial RVRTC devices.[10] We have previously shown that prediction equations for RVRTC derived from 'in-house' equipment are inappropriate for commercially available devices, and although an interim correction factor was proposed,[11] this was only based on a limited dataset.

During the past 5 years, the increasing availability of RVRTC data from healthy infants studied using commercial equipment in various centres worldwide, together with development of sophisticated statistical modelling techniques has provided the opportunity to improve interpretation of ILF results by developing more robust equipment-specific reference equations.

The primary aims of this study were to:

1. investigate the extent to which published equations,[9] or a proposed interim correction factor[11] fit data collected using the only currently available commercial equipment for assessing RVRTC (Jaeger Masterscreen BabyBody; Carefusion, Hoechberg, Germany);
2. collate available RVRTC data from centres using Jaeger equipment and the same protocols and to derive equipment-specific reference ranges for RVRTC outcomes;
3. investigate the clinical implications of implementing the 'Jaeger-specific' reference equations in infants with cystic fibrosis (CF).

To facilitate interpretation of longitudinal lung function (LF) data throughout childhood and adolescence, our secondary aim was to apply similar methodology to develop separate reference equations for use in infants previously studied with 'in-house' equipment and the Respiratory Analysis Software Program (RASP), as used in various studies prior to 2002.

METHODS

An initial assessment of potentially available RVRTC data from healthy infants was undertaken by the American Thoracic Society/European Respiratory Society (ATS/ERS) Infant and Preschool LF testing Task Force in 2013. Collaborators were subsequently identified as those indicating willingness to participate once data collection for specific studies had been completed.

RVRTC data from healthy infants and children <2.5 years of age collected using the Jaeger BabyBody were available from four specialised paediatric centres in the UK, Spain, Portugal and Australia as described previously.[5 8 12–14] Similar data collected using the RASP system were also available from the UK centre.[3 15] Regrettably, although RVRTC data collection using the nSpire IPL system in healthy infants is currently in progress, investigators were not in a position to release these data at time of collation, and this device is no longer commercially available.

The reference population comprised healthy full-term (≥37 weeks' gestational age and birth weight ≥2.5 kg) infants without congenital abnormalities or respiratory compromise (ie, no current respiratory morbidity or history of respiratory illness requiring hospitalisation) recruited to epidemiological studies or as controls for clinical research.[3–5 8 12–15] A population of infants diagnosed with CF by newborn screening without significant comorbidity recruited to recent clinical research studies[13 16] was used to investigate the impact of using different equations to interpret results. Local research ethics committee approval was granted for each study (see online supplementary material for details) and written informed parental consent obtained for all infants.

Equipment and study protocol

LF tests were undertaken at least 3 weeks after any respiratory illness. Data were collected during quiet sleep, after oral sedation with chloral hydrate (50–100 mg/kg depending on age). RVRTC data collected using Jaeger were obtained using identical study protocols for data collection, analysis and quality control (QC).[17] RVRTC data using the 'in-house' RASP equipment and identical protocols were collated from studies undertaken in London prior to switching to the Jaeger equipment in 2002. RVRTC data collection and analyses were performed in accordance with ATS/ERS guidelines.[18] Researchers from all centres were trained by the ILF team (AFH, SL and JS) in London, with inter-laboratory visits to observe tests in progress and independent over-read of results to ensure QC.

Detailed descriptions of data collection and analysis using the Jaeger equipment have been published.[17] In brief, RVRTC was performed from an inflation pressure of 30 cm H_2O, the manoeuvre being repeated until three acceptable and repeatable FEFV curves were obtained. Forced expired volume in 0.5 s ($FEV_{0.5}$), FVC, $FEV_{0.5}$/FVC, forced expiratory flow when 75% FVC had been expired (FEF_{75}) and FEF between 25% and 75% FVC ($FEF_{25–75}$) were reported from the 'best' raised volume curve. The latter was defined as the technically acceptable FEFV curve with the highest sum of FVC and $FEV_{0.5}$.[18]

Statistical analysis

Descriptive characteristics are shown as mean (SD) or median (range) for continuous variables and as n (%) for categorical ones (IBM SPSS Statistics V.22). Multiple fractional polynomials[19] were used to identify the most suitable transformation of the independent variables (ie, height, weight, age) when modelling LF outcomes to achieve normality of the residuals. The 'nlme' package in R (V.3.1–117) was used to check the models and the assumptions for the residuals distribution using the selected multiple fractional polynomial transformations, taking into account repeated measurements in individuals nested within centres by applying a random intercept model. Reference equations for RVRTC outcomes were then constructed as described previously[20 21] using the LMS (lambda–mu–sigma) method[22] and the best polynomial combination. This method is an extension of regression analysis that includes three components: (1) skewness (lambda, L), which models the departure of variables from normality using a Box–Cox transformation; (2) median (μ, M) or predicted value; and (3) coefficient of variation (sigma, S), which models the spread of values around the median and adjusts for any non-uniform dispersion. The three quantities are allowed to change with length and/or age, to reflect changes in the distribution as children grow. The L, M and S coefficients are combined algebraically to convert individual observations to z-scores: $z\text{-score}=((\text{measurement}/M)^{L}-1)/(L\times S)$.[22] The LMS method was applied using the GAMLSS package in R.[23] Goodness of fit was assessed using the Schwarz Bayesian criterion, which compares consecutive models directly while adjusting for increased complexity to determine the simplest model with best fit.[24] See online supplementary material for further details. In addition to age, body size and sex, the potential impact of ethnicity and tobacco smoke exposure on RVRTC outcomes was also examined.

To examine the potential clinical implications of using different equations, RVRTC data from infants with CF diagnosed by newborn screening that had been collected using the Jaeger equipment were expressed as z-scores using the new equipment-specific prediction equations and results compared with those derived from published reference equations ('Jones' equations)[9] and after applying a previously proposed interim correction factor.[11]

RESULTS

Healthy infants

RVRTC data from 431 healthy term infants on 653 test occasions were collated from the four centres. RASP data were collected between 1997 and 2002, and Jaeger data between 2001 and 2014. After 14 exclusions (figure 1), data were available from 429 infants (50% boys; 88% white infants) on 639 test occasions (age: 3.8–117.8 weeks; weight: 3.0–14.8 kg; length: 50.2–92.5 cm). Group characteristics of the reference population according to equipment and centre are summarised in table 1.

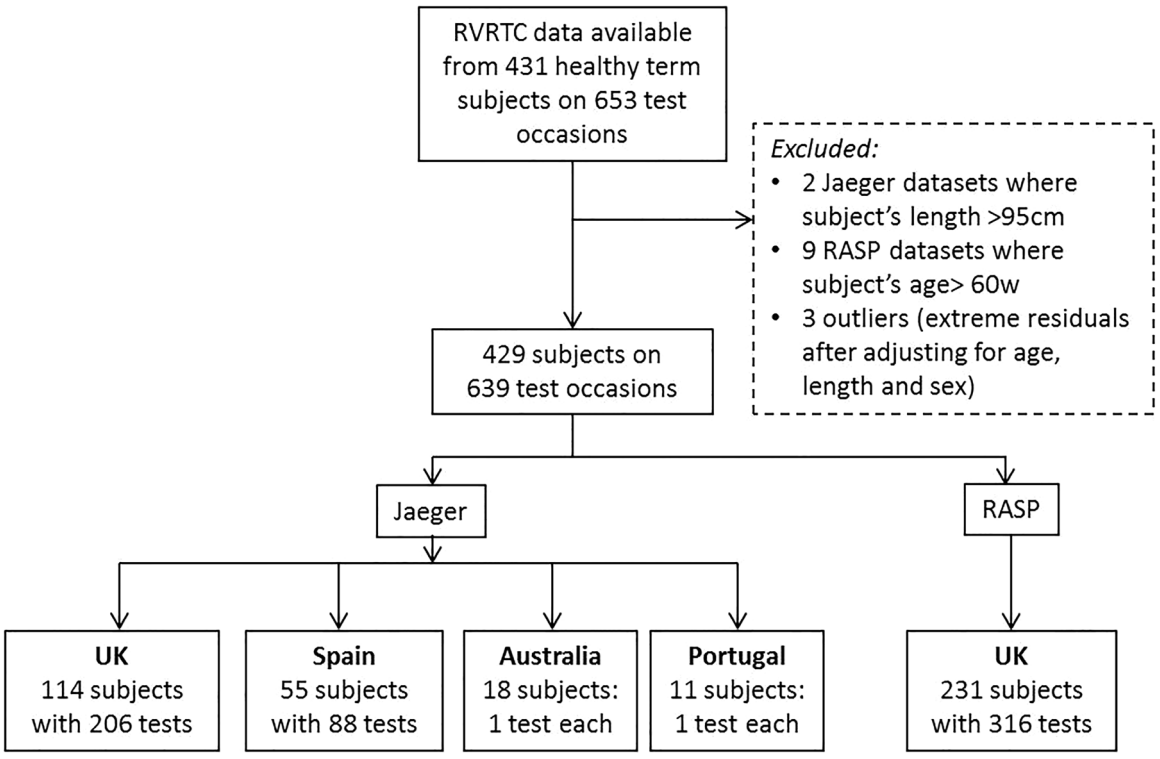


Figure 1 Flowchart illustrating collated FEFV data from four centres. To ensure prediction equations were not biased by distribution, the range of collated data used for deriving the prediction equations were limited to subjects with length <95 cm for Jaeger data and age <60 w for RASP data. FEFV, forced expiratory flow volume; RVRTC, raised volume rapid thoracoabdominal compression; RASP, Respiratory Analysis Software Program.

Jaeger RVRTC data plotted against length showed a positive association between LF and growth with good overlay between centres (figure 2).

Extent to which RVRTC data fit published equations

When comparing RVRTC data between equipment (RASP vs Jaeger), the spread for LF outcomes, especially FEF_{25-75}, was higher for RASP data and appeared to increase with growth (see online supplementary figure S1). After adjusting for age, height and sex, the Jones equations[9] provided a reasonable *mean* fit for the RASP data from healthy term infants and young children but, with the exception of $zFEV_{0.5}$/FVC, the *spread* of data was wider than expected (mean (SD) $zFEV_{0.5}$: −0.24 (1.32); zFVC: −0.38 (1.19); $zFEF_{25-75}$: −0.03 (1.38) z-scores) such that the lower limit of normal would be inappropriate. The Jones equations[9] were not appropriate for Jaeger data, with LF from healthy infants being significantly lower than expected by an average (SD) of 0.61 (1.09) z-scores for $FEV_{0.5}$, 0.36 (1.09) z-scores for FVC, 0.31 (1.18) for $FEV_{0.5}$ /FVC and 1.13 (1.05) z-scores for FEF_{25-75}. The overestimation of abnormalities among these healthy subjects increased with length (figure 3A) and age (see online supplementary figure S2A).

New equipment-specific RVRTC reference equations

The Jaeger-specific RVRTC reference equations for infants and young children aged between 4 and 118 weeks, with lengths

Table 1 Group characteristics of reference population according to centre and equipment

Equipment	Jaeger					RASP
centre	London, UK	Newcastle, Australia	Barcelona-Donostia, Spain	Lisbon, Portugal	ALL	London, UK
Subjects, n	114	18	55	11	198	231
Boys	45%	61%	44%	82%	48%	52%
White infants	76%	83%	100%	100%	84%	91%
Test occasions, n	206	18	88	11	323	316
Postnatal age* (w)	50.7 (8–112)	56.4 (15–112)	73.2 (21–104)	57.9 (45–118)	52.2 (8–118)	7.7 (4–59)
Length* (cm)	75.5 (54.1–92.5)	76.8 (57.2–85.9)	79.0 (54–86.8)	78.0 (71.5–84.5)	75.9 (54–92.5)	57.9 (50.2–78.8)
z-Length†	0.76 (0.94)	−0.24 (1.29)	0.10 (1.17)	−0.04 (1.3)	0.50 (1.10)	0.25 (1.04)
z-Weight†	0.32 (0.92)	0.20 (1.38)	0.14 (0.95)	−0.02 (0.90)	0.25 (0.96)	−0.13 (0.92)
Smoking in pregnancy	12%	22%	7%	18%	12%	36%
Postnatal smoking exposure	15%	67%	15%	36%	18%	44%

*Median (range).
†Mean (SD) according to UK-WHO reference range for term infants.[25]
RASP, Respiratory Analysis Software Program; z, z-scores.

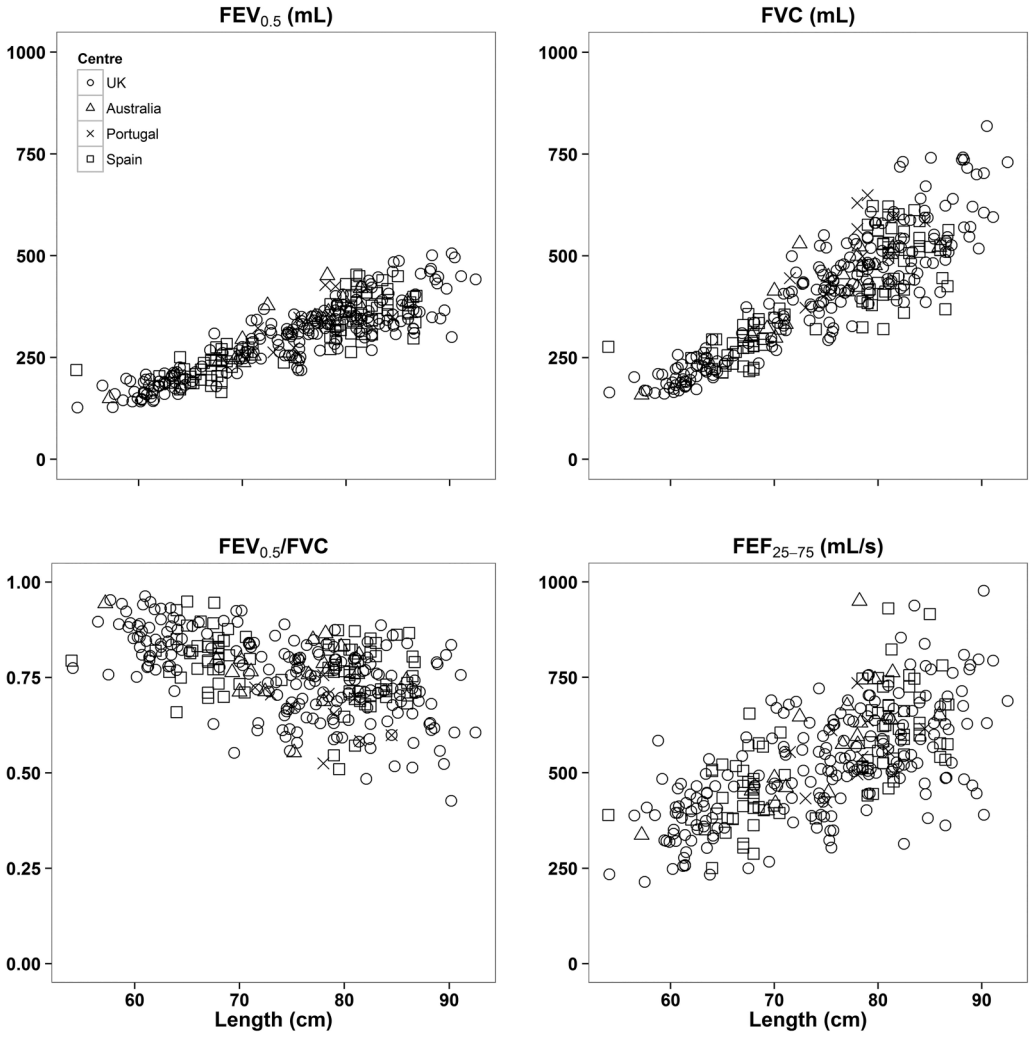


Figure 2 Forced expiratory flow volume (FEFV) data using the Jaeger equipment according to centre. As can be seen, there was good overlay between centres. Further details are presented in the online supplementary figure S4.

between 54 and 92 cm are given in table 2. The mean (SD) z-scores for all RVRTC outcomes in healthy controls approximated 0 (1) with 95% of data falling within±1.96 z-scores (figure 3B; see online supplementary figure S2B). Fitted centiles with the corresponding limits of normal are shown in online supplementary figure S3. No significant associations between Jaeger RVRTC outcomes and either ethnicity or tobacco smoke exposure were identified. When modelling was limited to data from white subjects or those not exposed to tobacco smoking, prediction equations were similar to those derived from all data (data not shown). Comparisons of data from the different centres are shown in the online supplementary figure S4.

Prediction equations for RASP data, applicable to infants aged 4–59 weeks and 50–79 cm are presented in the online supplementary table S1 (see also figure 3C and online supplementary figure S2C). For infants studied using RASP, in whom there was much higher tobacco smoke exposure (see online supplementary table S2), lower flows were observed among those exposed.

Clinical implications of using different RVRTC reference equations when interpreting data from infants with CF

Jaeger RVRTC data were also available from 100 (51% boys) newborn screened infants and young children with CF, who were studied in London on 233 test occasions between 2009 and 2014 (median (range) age: 50.4 w (8–111). The impact of using inappropriate reference equations for interpreting results is clearly illustrated in figure 4. Reliance on the 'Jones' reference equations[9] would have led to 57/233 (24.5%) tests in CF infants being classified as having 'abnormal' $FEV_{0.5}$ (≤1.96 z-scores). However, application of the new Jaeger-specific equations indicated that 58% (33/57) of CF infants with abnormal results according to Jones equations were misclassified. Similarly, of the 41 (18%) and 83 (36%) of CF infants identified as having diminished FVC and $FEF_{25–75}$ respectively according to Jones' equations, 61% (25/41) and 69% (57/83) would have been misclassified. When using the Jones equations, abnormalities in FVC and $FEV_{0.5}$ were increasingly overestimated with age (figure 5A). Thus both the prevalence and severity of LF abnormalities in CF infants were grossly overestimated when using the Jones equations. While application of the interim correction factor to the Jones equations[11] tended to slightly overestimate abnormalities in younger infants, they appeared adequate for older subjects (figure 5B).

DISCUSSION

In this study, we have shown conclusively that published reference data[9] for RVRTC outcomes which were based on data collected using 'in-house' systems are inappropriate for data

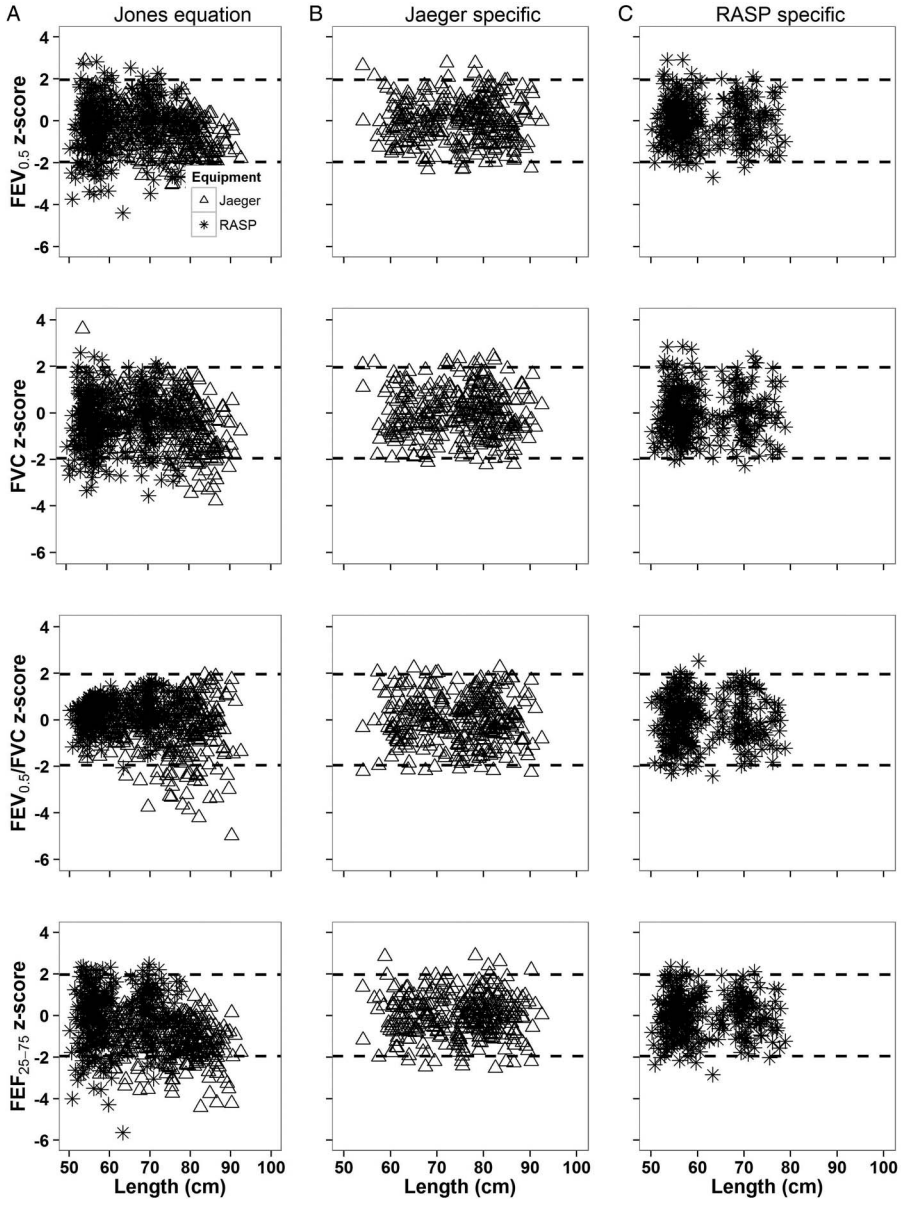


Figure 3 Forced expiratory flow volume (FEFV) data from healthy infants according to published and new equipment-specific prediction equations. Within a normal population, one would expect 95% of a healthy population to fall within ±1.96 SD or z-scores of predicted values, with only 2.5% below the lower limit of normal (LLN: ≤−1.96 z).[28] While this is true for the new equipment-specific equations shown in panels B and C, when the Jones equations[9] are applied to results from healthy infants studied using either the Respiratory Analysis Software Program (RASP) system or Jaeger equipment (A), up to 15% fell below the LLN depending on outcome.

collected using other equipment, whether 'in-house' or commercially available. Since RVRTC is being increasingly used worldwide with Jaeger BabyBody being the only commercial equipment currently available, it is essential to have appropriate and robust equipment-specific reference equations for interpreting data. FEFV data from healthy infants collected using Jaeger equipment were significantly lower than those predicted by the Jones equations, resulting in an overestimation of abnormalities in both health and disease, the magnitude of which increased with age. While adequate for older infants and toddlers, interim efforts to apply an adjustment for equipment-specific differences[11] also overestimated abnormality in younger infants. Through multicentre collaboration and using sophisticated statistical modelling techniques, we have now developed improved equipment-specific RVRTC reference equations which will enhance interpretation of infant LF results over the first 2 years of life.

Strengths and limitations

The multicentre collaboration which provided the largest dataset of RVRTC results to date using the same methodology and equipment is one of the main strengths of this study. In addition, the training provided by experienced investigators in London, with independent QC and over-read of data from other centres ensured a high degree of QC and reliability, minimising the chance that the lower flows and volumes from Jaeger were simply due to failure to achieve flow limitation or poor QC.

A potential limitation of the study is the small number of data from two of the sites and the possible bias that could be introduced due to the fact that the Institute of Child Health in London not only provided the majority of data and developed standardised QC criteria but also trained investigators from other centres. Nevertheless, availability of the recently published ILF testing manual[17] with a step-by-step guide to ILF assessments using Jaeger equipment should facilitate future quality and consistency of such assessments, allowing these new equations to be used reliably to interpret RVRTC data from other centres using the same equipment and QC. There is an urgent need for users of nSpire infant equipment to undertake a similar exercise.

Equipment-specific differences and prediction equations

As reported previously, possible reasons for the lower flows observed when using Jaeger rather than 'in-house' equipment

Table 2 Jaeger rapid thoracoabdominal compression (RVRTC) prediction equations

	Jaeger
$FEV_{0.5}$	
M	$\exp(8.8873 - 25.349/\sqrt{\text{Length}} - 1.668/\sqrt{\text{Age}})$
L	1
S	0.1296
FVC	
M	$\exp(4.6391 + 0.023 \times \text{Length} - 2.496/\sqrt{\text{Age}})$
L	1
S	$\exp(-1.6217 - 1.839/\sqrt{\text{Age}})$
$FEV_{0.5}$/FVC	
M	exp(0.0977−0.0942 × LN(Age)−0.0285×Sex)
L	2.380
S	exp(−3.4316+0.3038 × LN(Age))
FEF_{25-75}	
M	exp(7.8253−114.29/Length−0.064 × Sex)
L	0.672005
S	0.2027
FEF_{75}	
M	exp(7.5205−131.29/Length−0.0662 × Sex)
L	1
S	0.2417

These reference equations are only valid for subjects between 4 and 118 weeks of age, and 54–92 cm in length. Age, age in weeks; L, lambda (skewness); length, length in cm; LN, natural logarithm; M, μ (median); S, sigma (coefficient of variation); sex, girl=0 and boy=1.

include differences in internal algorithms relating to BTPS (Body Temperature and Pressure, Saturated) and volume drift corrections.[11] It is therefore not surprising that the Jones equations[9] are inappropriate for interpreting data from current commercial equipment. The overestimation of predicted values according to Jones such that Jaeger results were underestimated, resulted in a progressive overestimation of abnormalities with age (figure 5A). These results emphasise the importance of examining whether potential lung function reference equations are appropriate across the entire age or height range studied and not simply as a group mean. We previously developed an interim correction factor that could be applied to the Jones equations in an attempt to address observed equipment-related differences.[11] However, due to the relatively few healthy infants studied with the Jaeger device at the time, sample size was unsuitable for complex modelling and hence did not completely solve the problem. Thus although application of the interim correction factors improved the fit for Jaeger data from older infants, they still tended to overestimate abnormalities in younger infants (figure 5B). Reasons for discrepancies in this age group could also relate to the limited number of very young infants included within the original Jones dataset.

Neither the Jones nor the new Jaeger equations should be applied to data collected using the RASP system, prediction equations for which are presented in online supplementary table S1. Although use of such equations will be limited to the relatively small number of centres which used this equipment and software in the past,[3] their availability will facilitate improved interpretation of longitudinal studies, such as that currently being undertaken in clinically diagnosed children with CF,[26 27] who are now being followed up during adolescence (http://www.ucl.ac.uk/london-cystic-fibrosis/).

Reference population

As reported previously,[9 11] length and age are major determinants of infant LF, with sex also contributing to the $FEV_{0.5}$/FVC ratio and $FEF_{\%}$, indicating relatively smaller airway calibre in relation to lung size in boys compared with girls during the first 2 years of life. Ethnic differences in LF in older subjects have been well documented.[28] However, within the current reference population, ethnicity was not shown to be associated with RVRTC outcomes, possibly due to the limited data from those of non-white, European descent. Results from such subjects must therefore be interpreted cautiously until further ethnic-specific data are available.

In contrast to the approach taken by Jones *et al*,[9] we did not include 'smoking status' in the regression models so that the potential impact of tobacco-smoke exposure can be examined separately. Following successive government strategies to tackle smoking in the UK,[29] smoking prevalence in London mothers has fallen over the past decade from ~40% in the 1990s[15 30] to around 16% currently.[8] The lack of association between tobacco smoke exposure and LF in infants studied using the Jaeger equipment, probably reflects the recent low exposure rates. By contrast, during earlier assessments using the RASP device where a much higher proportion of infants were exposed to household smoking,[15 30] $FEV_{0.5}$/FVC and FEF_{25-75} were both significantly lower when compared with those not exposed (see online supplementary table S2).

Clinical implications

The clinical importance of appropriate reference equations for interpreting LF data in older children and adults is well recognised.[31 32] Its relevance to infants and young children is particularly timely given the importance of accurately interpreting outcome in diseases such as CF in the first few years of life in an era when the potential for early disease-modifying treatments is becoming a reality. Understanding which infants have abnormal LF or how change over time occurs within individuals is only possible with appropriate reference data to track lung health accurately. Our results clearly illustrate the potential consequences of using inappropriate reference equations. Despite the well-recognised difficulties,[33] we have always recruited and

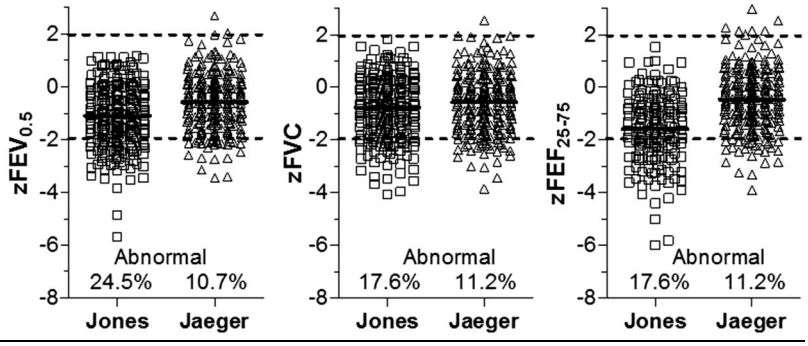


Figure 4 Comparison of forced expiratory flow volume data from newborn screened infants with cystic fibrosis (CF) collected using the Jaeger device when interpreted according to Jones *et al* and the new Jaeger-specific equations. Data were available from 100 infants with CF on 233 occasions. Solid lines denote the mean value for the group and dotted lines denote the upper and lower limit of the normal range (as defined by±1.96 z-scores).

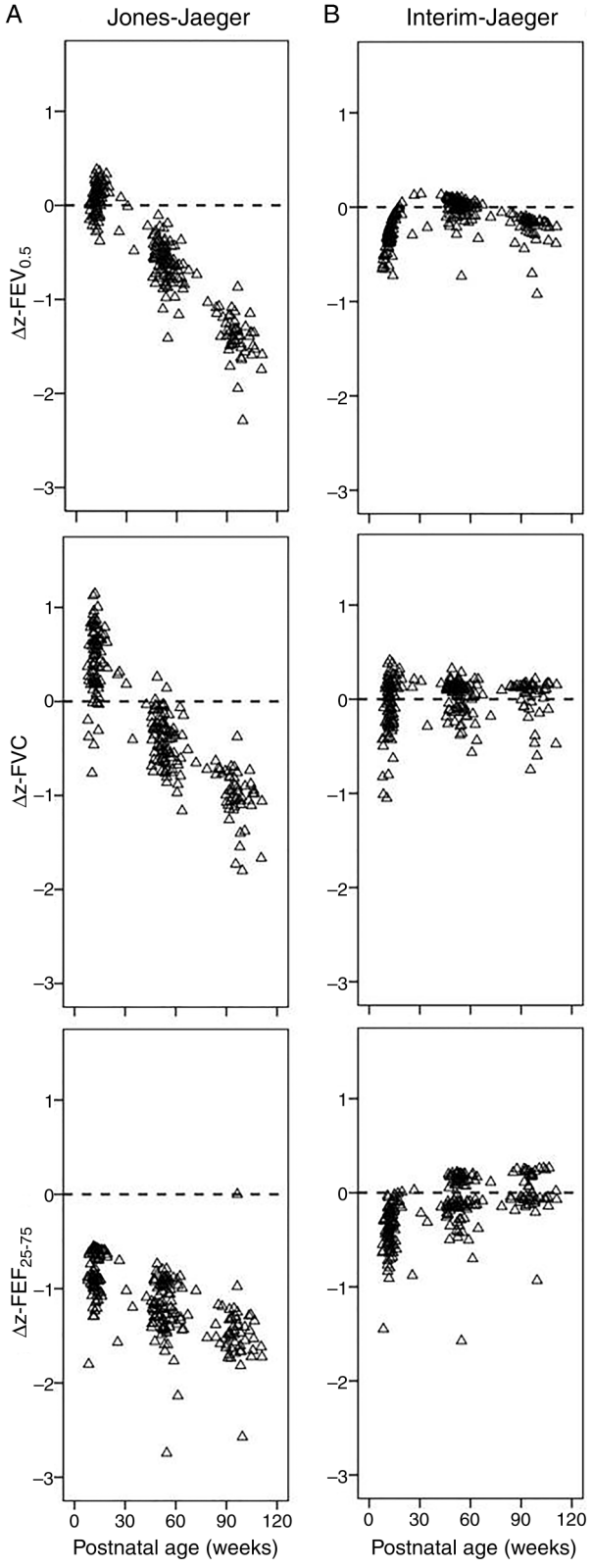


Figure 5 Discrepancies in calculated forced expiratory flow volume (FEFV) z-scores between existing and new reference equations according to subjects' age. Graphs are plotted as the z-score difference (Δz) in FEFV outcomes in each infant between results derived from (A) Jones[9] minus current Jaeger-specific and (B) Interim correction factor applied to Jones[11] minus new Jaeger-specific reference equations. When using the Jones equations, the relative underestimation of lung function is progressive with age, whereas the interim correction factor[11] adjusts adequately for infants above ~15 weeks of age but tends to underestimate results in younger infants.

prospectively assessed healthy controls for clinical research projects in order to strengthen interpretation of results. The importance of such an approach is illustrated by our current findings, since without such controls we would have misinterpreted both the prevalence and magnitude of abnormalities among new-born screened infants with CF during the first 2 years of life, as well as the change over time. Despite the shift in z-scores between published and new equipment-specific equations, the magnitude of the *difference* in z-scores between CF infants and healthy controls remained similar whether previously published[9 11] or current reference ranges were applied. This confirms recent reports that LF in newborn screened CF infants remains stable or improves during the first 2 years of life, with considerably smaller deficits than previously observed prior to the introduction of newborn screening.[13 16] By contrast, had we relied solely on the original Jones equations without a contemporaneous control group, we would have falsely concluded that LF deteriorates with age in newborn screened infants, with mean $FEV_{0.5}$ approximating −2 z-scores during the second year of life (figure 5A). We suggest that such scenarios may contribute the significant 'abnormalities' in infant LF outcomes reported by others in the newborn screened CF literature.[7] Continued use of inappropriate reference equations for longitudinal follow up would result in significant overdiagnosis of abnormal LF, and overestimation of the rate of decline in LF during the first 2 years of life.

CONCLUSIONS AND FUTURE DIRECTIONS

We have shown that published reference ranges derived from 'in-house' equipment are inappropriate for interpreting RVRTC data collected using other equipment, including the only commercially available system currently available. We present new robust equipment-specific RVRTC prediction equations to improve interpretation of data previously collected using the RASP system as well as current data obtained with the Jaeger device, although further work is required to ascertain whether these are equally applicable to infants who are not of white European descent. As in older subjects, even when adequate reference equations are available, assessment and interpretation of LF in infants should always be undertaken within an appropriate clinical context. Our equipment-specific equations for infant LF testing will improve the ability to track lung health from early life. This should, in turn, ultimately improve our understanding of the evolution of respiratory morbidity in diseases such as CF, and enable an appropriate evaluation of the utility of infant LF testing as an outcome measure in clinical trials.

Author affiliations
[1]Respiratory, Critical Care & Anaesthesia section in IIIP Programme, UCL, Institute of Child Health, London, UK
[2]Department of Clinical Epidemiology, Nutrition and Biostatistics Section, UCL, Institute of Child Health, UK
[3]Respiratory Medicine, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK
[4]Paediatric Pulmonology Unit, Hospital Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain
[5]Division of Paediatric Respiratory Medicine, Donostia University Hospital, San Sebastian, Spain
[6]Department of Paediatrics, University of the Basque Country, San Sebastian, Spain
[7]School of Medicine and Public Health, University of Newcastle, Newcastle, Australia
[8]Department of Paediatric Respiratory & Sleep Medicine, John Hunter Children's Hospital, Newcastle, Australia
[9]Centro de Alergia, CUF Descobertas Hospital, Lisbon, Portugal
[10]CEDOC, NOVA Medical School, Lisbon, Portugal

Acknowledgements We thank the infants and parents participating in the research studies. We also would like to thank Iris Goetz, Georg Hulskamp, Henrik

Ljungberg, Padmaja Subbarao, Teresa Bandeira, Joanne Chittenden, Amit Gupta, Alicia Bolton, The Thanh Diem Nguyen, Lena Thia, Deeba Ahmed, Jane Chudleigh, Joanne Miles, Lucy Brennan, Biarta Rhys-Jones, Alison Rowland, Robyn Hankins, Majella Maher and Bruce Whitehead for their help in recruitment and data collection.

Contributors SL and JS were responsible for the conception and design of this study; JS, AM-G, LMB and JM were responsible for supervision of studies within their respective laboratories; A-FH provided technical training, supervision and audit of data collection and analyses; AF-H, SL, JK, IdM, OS-P, JM and PC-E recruited the infants and undertook all data collection and analyses; SL collated all data from participating centres and together with VB and AW performed statistical analyses and modelling; SL, GD and JS drafted the manuscript; all remaining authors revised and approved the manuscript before submission.

Funding This study is supported by grants from the Dunhill Medical Trust, UK; British Lung Foundation; Cystic Fibrosis Trust; Special Trustees: Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK; Smiths Medical Ltd, UK; Hunter Medical Research Institute, Newcastle, Australia; Hunter Children's Research Foundation, Newcastle, Australia and the Spanish team had a non-restricted grant from AbbVie.

Competing interests None declared.

Ethics approval London studies: East London and the City Health Authority Research Ethics Committee (REC) (P/97/250); UCL Institute of Child Health/Great Ormond Street Hospital REC (96EB23; 05/Q0508/141); North Thames Multi-centre REC (09/H071/314); Australian study: Hunter New England Health Human REC/09/07/15/5.04); Barcelona-Donostia study: Hospital Donostia: "Comité Ético de Investigación Clínica del Área Sanitaria de Gipuzkoa". Approval date: 22 September 2010; Hospital Vall d'Hebron: "Comité Ético de Investigación Clínica del Hospital Universitario Vall d'Hebron". Approval date: 14 September 2010; Lisbon study: Lisbon REC (Approval date: December 2005).

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Turner DJ, Stick SM, Lesouef KL, *et al*. A new technique to generate and assess forced expiration from raised lung volume in infants. *Am J Respir Crit Care Med* 1995;151:1441–50.
2 Feher A, Castile R, Kisling J, *et al*. Flow limitation in normal infants: a new method for forced expiratory maneuvers from raised lung volumes. *J Appl Physiol (1985)* 1996;80:2019–25.
3 Ranganathan SC, Stocks J, Dezateux C, *et al*. The evolution of airway function in early childhood following clinical diagnosis of cystic fibrosis. *Am J Respir Crit Care Med* 2004;169:928–33.
4 Lum S,Gustafsson P, Ljungberg H, *et al*. Early detection of cystic fibrosis lung disease: multiple-breath washout versus raised volume tests. *Thorax* 2007;62:341–7.
5 Borrego LM, Stocks J, Leiria-Pinto P, *et al*. Lung function and clinical risk factors for asthma in infants and young children with recurrent wheeze. *Thorax* 2009;64:203–9.
6 Filbrun AG, Popova AP, Linn MJ, *et al*. Longitudinal measures of lung function in infants with bronchopulmonary dysplasia. *Pediatr Pulmonol* 2011;46:369–75.
7 Pillarisetti N, Williamson E, Linnane B, *et al*. Infection, inflammation, and lung function decline in infants with cystic fibrosis. *Am J Respir Crit Care Med* 2011;184:75–81.
8 Hoo AF, Gupta A, Lum S, *et al*. Impact of ethnicity and extreme prematurity on infant pulmonary function. *Pediatr Pulmonol* 2014;49:679–87.
9 Jones M, Castile R, Davis S, *et al*. Forced expiratory flows and volumes in infants. Normative data and lung growth. *Am J Respir Crit Care Med* 2000;161(2 Pt 1):353–9.
10 Peterson-Carmichael SL, Rosenfeld M, Ascher SB, *et al*. Survey of clinical infant lung function testing practices. *Pediatr Pulmonol* 2014;49:126–31.
11 Lum S, Hoo AF, Hulskamp G, *et al*. Potential misinterpretation of infant lung function unless prospective healthy controls are studied. *Pediatr Pulmonol* 2010;45:906–13.
12 Mattes J, Hankin R, Hilton J, *et al*. Frequent persistent wheeze in infancy may be associated with impaired FEV 0.5 and FVC values. *Am J Respir Crit Care Med* 2011;183:A1890.
13 Nguyen TT, Thia LP, Hoo AF, *et al*. Evolution of lung function during the first year of life in newborn screened cystic fibrosis infants. *Thorax* 2014;69:910–17.
14 De Mir I, Sardon-Prado O, Corcuera-Elosegui P, *et al*. Impaired lung function in "healthy" preterm infants: Interpretation using new equipment-specific reference equations. *Eur Respir J* 2014;44(Suppl 58):1151.
15 Dezateux C, Lum S, Hoo AF, *et al*. Low birth weight for gestation and airway function in infancy: exploring the fetal origins hypothesis. *Thorax* 2004;59:60–6.
16 Thia L, Hoo AF, Brennan L, *et al*. Stable lung function is maintained over 2 years in newborn screened (NBS) CF infants. *Eur Respir J* 2013;42(Suppl 57):1072s.
17 Hoo AF, Lum S, Mattes J, *et al*. *Manual of infant lung function tests*. London: UCL Discovery, 2014.
18 American Thoracic Society; European Respiratory Society. ATS/ERS statement: raised volume forced expirations in infants: guidelines for current practice. *Am J Respir Crit Care Med* 2005;172:1463–71.
19 Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Stat* 1994;43:429–67.
20 Cole TJ, Stanojevic S, Stocks J, *et al*. Age- and size-related reference ranges: a case study of spirometry through childhood and adulthood. *Stat Med* 2009;28:880–98.
21 Stanojevic S, Wade A, Cole TJ, *et al*. Spirometry Centile Charts for Young Caucasian Children: The Asthma UK Collaborative Initiative *Am J Respir Crit Care Med* 2009;180:547–52.
22 Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992;11:1305–19.
23 Rigby RA, Stasinopoulos DM. Generalised additive models for location, scale and shape. *Appl Stat* 2005;54(Part 3):507–54.
24 Schwarz GE. Estimating the dimension of a model. *Ann Statist* 1978;6:461–4.
25 Wright CM, Williams AF, Elliman D, *et al*. Using the new UK-WHO growth charts. *BMJ* 2010;340:c1140.
26 Kozlowska WJ, Bush A, Wade A, *et al*. Lung function from infancy to the preschool years after clinical diagnosis of cystic fibrosis. *Am J Respir Crit Care Med* 2008;178:42–9.
27 Aurora P, Stanojevic S, Wade A, *et al*. Lung clearance index at 4 years predicts subsequent lung function in children with cystic fibrosis. *Am J Respir Crit Care Med* 2011;183:752–8.
28 Quanjer PH, Stanojevic S, Cole TJ, *et al*. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012;40:1324–43.
29 ASH (Action on smoking and health). Annual Report 2014, ASH, 2014.
30 Hoo AF, Henschen M, Dezateux C, *et al*. Respiratory function among preterm infants whose mothers smoked during pregnancy. *Am J Respir Crit Care Med* 1998;158:700–5.
31 Pellegrino R, Viegi G, Brusasco V, *et al*. Interpretative strategies for lung function tests. *Eur Respir J* 2005;26:948–68.
32 Stanojevic S, Wade A, Stocks J. Reference values for lung function: past, present and future. *Eur Respir J* 2010;36:12–9.
33 Stocks J, Modi N, Tepper R. Need for healthy control subjects when assessing lung function in infants with respiratory disease. *Am J Respir Crit Care Med* 2010;182:1340–2.

New reference ranges for interpreting forced expiratory manoeuvres in infants and implications for clinical interpretation: a multi-centre collaboration

Sooky Lum, Vassiliki Bountziouka, Angela Wade, Ah-Fong Hoo, Jane Kirkby, Antonio Moreno-Galdo, Ines de Mir, Olaia Sardon-Prado, Paula Corcuera-Elosegui, Joerg Mattes, Luis Miguel Borrego, Gwyneth Davies, Janet Stocks

Online supplement

This supplement contains additional details to compliment the main manuscript

Additional information

Details of ethics committee approval for each study (approval ID):

London studies: East London and the City Health Authority Research Ethics Committee (REC) (P/97/250); UCL Institute of Child Health/Great Ormond Street Hospital REC (96EB23; 05/Q0508/141); North Thames Multi-centre REC (09/H071/314);

Australian study: Hunter New England Health Human REC (09/07/15/5.04)

Barcelona-Donostia study: Hospital Donostia: "Comité Ético de Investigación Clínica del Área Sanitaria de Gipuzkoa". Approval date: 22 September 2010; Hospital Vall d'Hebron: "Comité Ético de Investigación Clínica del Hospital Universitario Vall d'Hebron". Approval date: 14 September 2010

Lisbon study: Lisbon REC (Approval date: December 2005)

Statistical methods

Descriptive characteristics are shown as mean (SD) or median (range) for the continuous variables and as n (%) for the categorical ones. Multiple fractional polynomials[1] whereby a combination of integer or fractional power terms are fitted to produce a polynomial equation, were used to identify the most suitable transformation in any combination of the independent variables (i.e. height, weight, age) when modelling lung function outcomes to achieve normality of the residuals. R package "mfp" (Multivariable Fractional Polynomials) was used for this purpose (original by Gareth Ambler and modified by Axel Benner (2014), R package version 1.5.0.). The "nlme" package in R (version 3.1-117) was used to check the models and the distributions of residuals using selected mfp transformations taking into account repeated measures of individuals nested within centres by applying a random intercept model. Test occasions (i.e. repeat assessments within individuals) were not included as main effects in the fixed part of the equation as not all four centres had repeated measurements. Random slope models were not tested.

Reference equations for raised volume rapid thoraco-abdominal compression (RVRTC) outcomes were then constructed as described [2,3] with the LMS (lambda-mu-sigma) method[4] fitting the best polynomial combination as indicated previously, using the GAMLSS package in R.[5]. This method is an extension of the regression analysis that includes three components: 1) skewness (lambda, L), which models the departure of variables from normality using a Box-Cox transformation; 2) median (mu, M) or predicted value; and 3) coefficient of variation (sigma, S), which models the spread of values around the median and adjusts for any non-uniform dispersion. The three quantities are allowed to change with length and/or age, to reflect changes in the distribution as children grow. The L, M and S coefficients are combined algebraically to convert individual observations to z-scores: z-score = $((\text{measurement}/M)^{L} - 1)/(L \times S)$.[4] Residual plots from multilevel models were used to check the skewness of the distribution. When no skewness was indicated, L was fixed at 1. Indication of skewness was present for $FEV_{0.5}$/FVC and FEF_{25-75}. This was not found to be dependent on either age or length (transformed as previously indicated from polynomials) using the LMS method. Spread dependency on age and length (transformed as previously indicated from polynomials) was tested in the LMS method.

Normality of the residuals was tested using histograms and Q-Q plots, while a plot of the residuals vs the fitted values from each model was used to check the assumption of homoscedasticity. Goodness of fit was assessed using the Schwarz Bayesian criterion, which compares consecutive models directly while adjusting for increased complexity to determine the simplest model with best fit.[6] Since RASP data were only available from London, and the distribution of age and body size differed in the two datasets, the decision was made to derive separate prediction equations for Jaeger and RASP equipment.

Modelling was performed using R v.3.1.0 incorporating packages as given previously. IBM SPSS Statistics v.22 was used for data inspection, distribution and descriptive statistics.

RESULTS

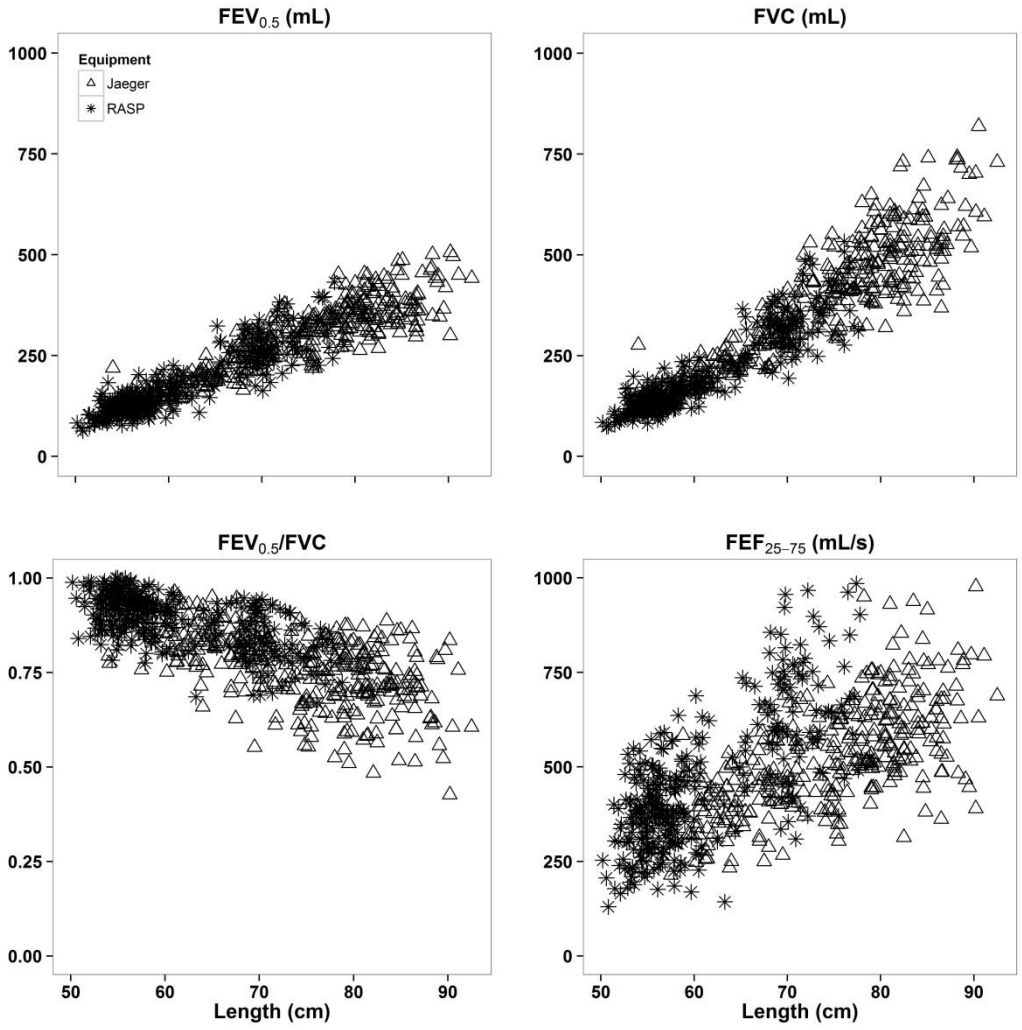
Data exclusions: To facilitate development of robust and reliable reference ranges for RVRTC outcomes based on sufficient sample size with relatively even spread over a wide range of age and body size in healthy infants, we excluded a few data points which were at the extremes of length and age range.

Despite being recruited from the same area of London and measured by the same team of respiratory physiologists, the children studied using Jaeger equipment were significantly taller and heavier than those tested a decade earlier using RASP. A thorough investigation of background details for the two cohorts revealed no specific cause for this difference which may simply be attributable to chance due to the relatively small sample size and the fact that differences of up to 0.5 z-scores can occur by chance when comparing populations with less than 300 subjects.[7]

RASP vs. Jaeger®

When plotted according to equipment, there was relatively good overlay for FVC and $FEV_{0.5}$ among younger children but RASP FEF_{25-75} were significantly higher than Jaeger® data (Figure S1). Initial attempts to model all the data by including equipment as an independent variable were not successful in achieving adequate fit. As the distribution of RASP and Jaeger® data were markedly different, with more RASP data being available from younger infants, equipment-specific reference equations were derived separately for RASP and Jaeger® data. Equations for the more widely available Jaeger® equipment have been presented in the main manuscript. The following section focusses on equations for interpreting RASP data, which are of relevance to laboratories which have previously collected data with this device and are therefore particularly relevant for ongoing longitudinal follow up studies into later childhood.

Figure S1 RVRTC outcomes according to equipment



While there was reasonable overlay for $FEV_{0.5}$ and FVC among younger infants, for any given length, FEF_{25-75} data were considerably lower for Jaeger® data than RASP, necessitating the use of separate reference equations. Prediction equations for RASP data are presented in Table S1.

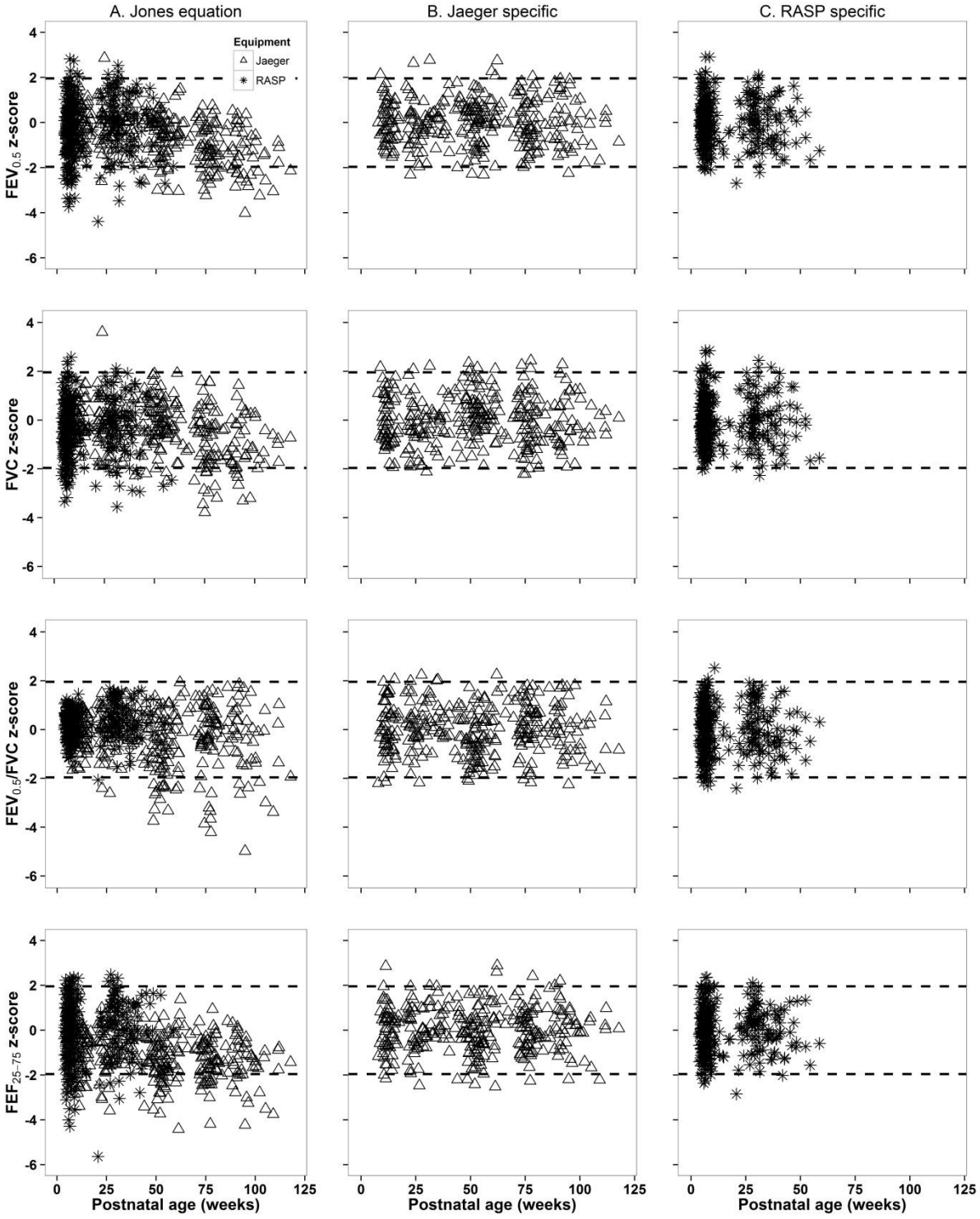
Table S1 RASP RVRTC prediction equations

	RASP
$FEV_{0.5}$	
M	exp(5.7153-3794/(Length^2)+0.1892*LN(Age))
L	1
S	0.1776
FVC	
M	exp(6.9725-136.49/Length+0.2185*LN(Age))
L	1
S	0.1760
$FEV_{0.5}$/FVC	
M	exp(0.0678-0.0036*Length+0.296/Age)
L	3.3441095
S	exp(-3.4119+0.2366*LN(Age))
FEF_{25-75}	
M	exp(4.0381+0.0034*Length-0.1057*Sex)
L	0.9032
S	0.2642
FEF_{75}	
M	exp(6.9172–5077/(Length^2)-0.1577*Sex)
L	0.6822
S	0.317

Abbreviations: L: lambda (skewness); M: mu (mean); S: sigma (coefficient of variation); Length: Length in cm; Age: Age in weeks; LN: natural logarithm. Sex: Girl = 0; Boy=1; These reference equations are only valid for subjects between 4-59 weeks of age and 50-79 cm in length.

A comparison of RVRTC data from healthy infants and young children according to published and new equipment-specific equations is presented in Figure S2. It can be seen that while many healthy infants fall outside the 95% 'normal range' (i.e. 13% for $FEV_{0.5}$, 10% for FVC and 17% for FEF_{25-75}) and could therefore be misclassified as 'abnormal' when using the Jones equations,[8] once the new equipment-specific equations are applied, 95% subjects fall within the normal range.

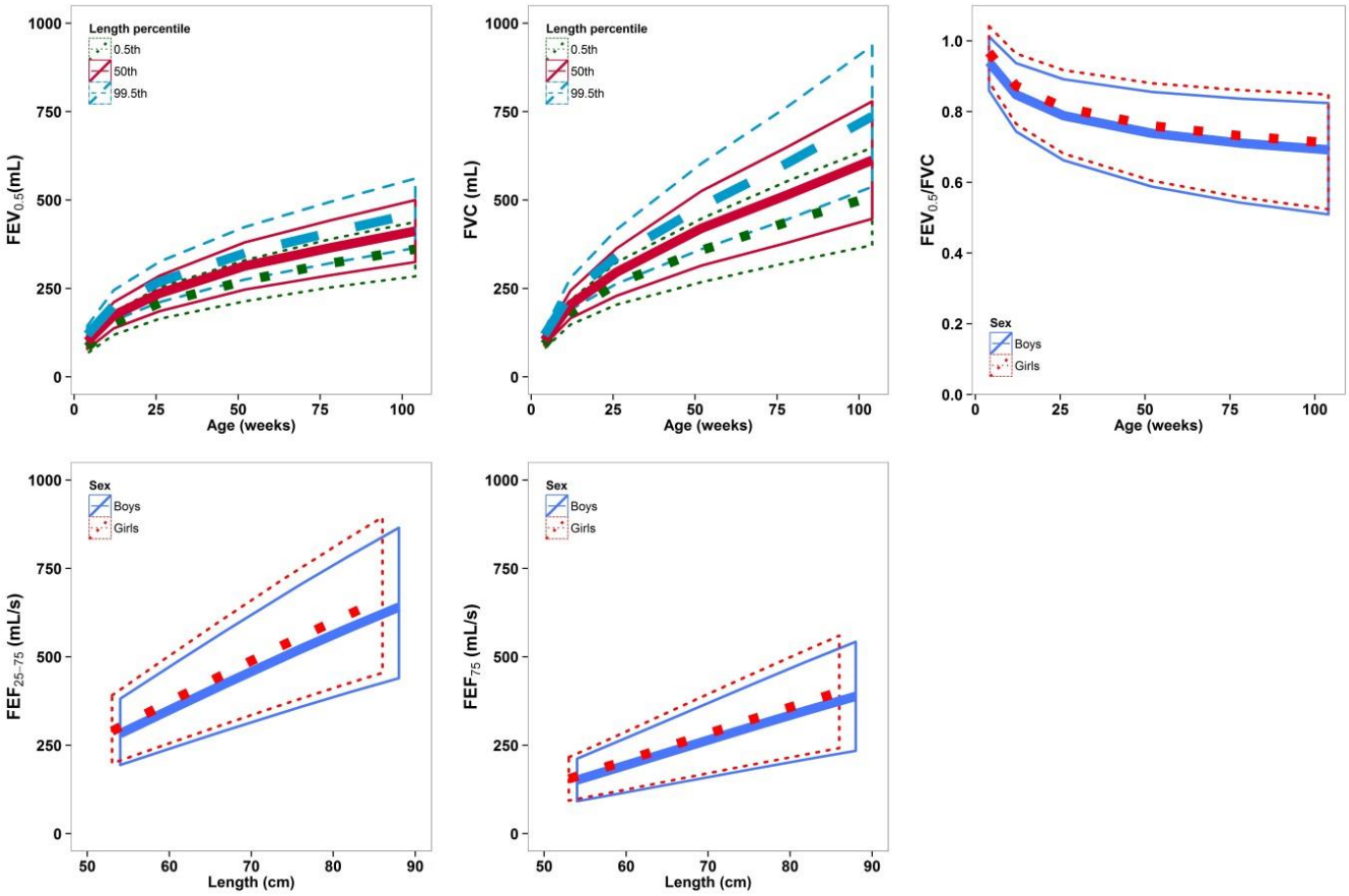
Figure S2 RVRTC data from healthy infants and young children plotted against age according to Jones et al[8] and the new equipment-specific equations



The dashed horizontal lines denote the upper and lower limit of normality. Within a healthy population, provided appropriate reference equations are applied, 95% of results should lie within ±1.96 z-scores of the predicted range.

Based on the new Jaeger® equations, fitted centiles with the corresponding upper and lower limits for the RVRTC outcomes are illustrated in Figure S3.

Figure S3 Fitted centiles for RVRTC outcomes based on the new Jaeger equations

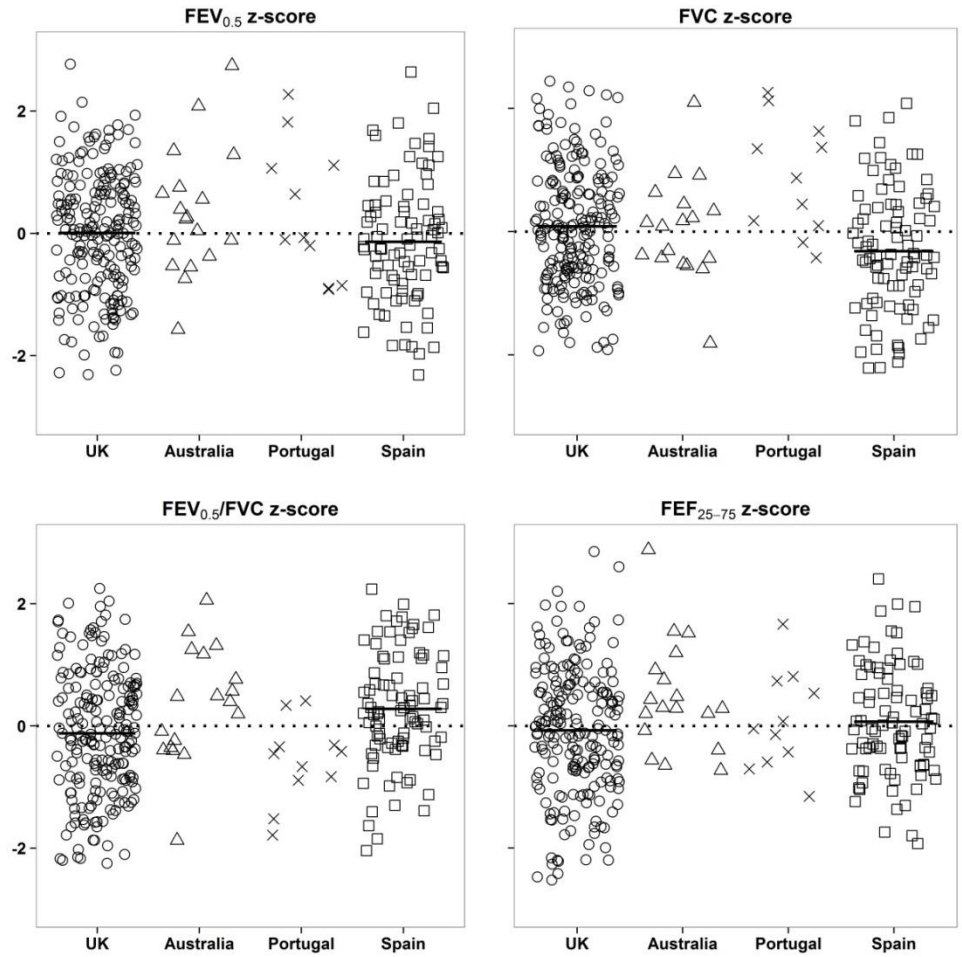


Fitted centiles for $FEV_{0.5}$, FVC and $FEV_{0.5}$/FVC were plotted against age as the spread (S) for these outcomes is age dependent. As the median (M) for $FEV_{0.5}$ and FVC is age and length dependent, the fitted centiles are shown for infants who may be of average length for age (50th percentile [**RED solid lines**]), short (0.5th percentile [**GREEN dotted lines**]) or tall (99.5th percentile [**BLUE dashed lines**]) for age. Bold lines indicate median values while thinner lines indicate upper and lower limit of the respective median (5th and 95th centile: ±1.65 SD). For example: at 1 year of age, for an infant whose length is at the 50th percentile, predicted average (95%CI) for FVC is 420 mL (315; 526); whereas the corresponding FVC would be 358 mL (268; 448) for a short-for-age infant (0.5th centile for length) and 482 mL (361; 604) for one who is tall-for-age (99.5th percentile for length). $FEV_{0.5}$/FVC and $FEF_{\%}$, are also sex dependent, thus the fitted centiles are plotted for boys (**BLUE solid lines**) and girls (**RED dotted lines**) vs. length as the spread (S) was not age dependent.

Comparison of Jaeger® results between centres

A comparison of RVRTC outcomes from the four centres collected using the Jaeger Babybody and expressed as z-scores using the new equipment-specific reference equations is shown in Figure S4. Although the limited number of infants from Portugal and Australia precluded any formal analysis, the majority of individual observations from all centres fell within ±2 z-scores. The distribution of $FEV_{0.5}$ results was very similar in the larger UK and Spanish datasets, but slightly lower FVC (and hence higher $FEV_{0.5}$/FVC and FEF_{25-75}) were observed among the Spanish infants. Despite this being the largest collation of RVRTC data from healthy infants to date, sample size was still relatively small to undertake inter-centre comparisons with any confidence, since differences of up to 0.5 z-scores can occur by chance within the same dataset due to sampling error when there are less than 300 per group (i.e. 150 boys and 150 girls).[7]

Figure S4 Comparison of RVRTC outcomes collected using the Jaeger® Babybody and expressed as z-scores using the new equipment-specific reference equations



The horizontal line within the UK and Spanish data denotes the mean value for the group.

Impact of smoking status on RVRTC outcomes

For infants studied using RASP, lower flows were observed among the 44% exposed to tobacco smoke, as reported previously (Table S2).[9,10] By contrast, no significant associations were identified between RVRTC outcomes and tobacco smoke exposure in infants studied using the Jaeger® device, probably reflecting the low exposure (18%) within this group.

Table S2 Impact of tobacco smoke exposure on healthy infants

	RASP			Jaeger®		
	Smoking exposure YES	**Smoking exposure No**	**Mean (95%CI) difference (Yes-No)**	**Smoking exposure YES**	**Smoking exposure No**	**Mean (95%CI) difference (Yes-No)**
n	140 (44%)	176 (56%)		59 (18%)	262 (82%)	
$zFEV_{0.5}$	-0.07 (1.05)	0.06 (0.96)	-0.13 (-0.36; 0.09)	0.15 (1.04)	-0.04 (0.99)	0.19(-0.09; 0.48)
zFVC	-0.03 (1.05)	0.03 (0.97)	-0.06 (-0.28; 0.16)	0.01 (1.05)	-0.01 (1.0)	0.02 (-0.27; 0.30)
$zFEV_{0.5}$/FVC	-0.14 (1.00)	0.11 (0.99)	**-0.26 (-0.48; -0.04)***	0.16 (0.93)	-0.04 (1.02)	0.20 (-0.08; 0.48)
$zFEF_{25-75}$	-0.15 (0.99)	0.12 (0.99)	**-0.28 (-0.50; -0.05)***	0.16 (0.90)	-0.04 (1.02)	0.21 (-0.08; 0.49)

Bold fonts indicate significant differences between smoking exposure groups. * p<0.05.

REFERENCES

1. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. Appl Stat 1994;**43**:429-67.
2. Stanojevic S, Wade A, Stocks J, et al. Reference ranges for spirometry across all ages: a new approach. Am J Respir Crit Care Med 2008;**177**(3):253-60.
3. Cole TJ, Stanojevic S, Stocks J, et al. Age- and size-related reference ranges: a case study of spirometry through childhood and adulthood. Stat Med 2009;**28**(5):880-98.
4. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992;**11**(10):1305-19.
5. Rigby RA, Stasinopoulos DM. Generalised additive models for location, scale and shape. Appl Stat 2005;**54**(Part 3):507-54.
6. Schwarz GE. Estimating the dimension of a model. Ann Statist 1978;**6**(2):461-64.
7. Quanjer PH, Stocks J, Cole TJ, et al. Influence of secular trends and sample size on reference equations for lung function tests. Eur Respir J 2011;**37**(3):658-64.
8. Jones M, Castile R, Davis S, et al. Forced expiratory flows and volumes in infants. Normative data and lung growth. Am J Respir Crit Care Med 2000;**161**(2 Pt 1):353-9.
9. Dezateux CA, Lum S, Hoo AF, et al. Low birthweight for gestation and airway function in infancy: exploring the fetal origins hypothesis. Thorax 2004;**59**(1):60-66.
10. Ranganathan SC, Stocks J, Dezateux C, et al. The evolution of airway function in early childhood following clinical diagnosis of cystic fibrosis. Am J Respir Crit Care Med 2004;**169**(8):928-33.