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

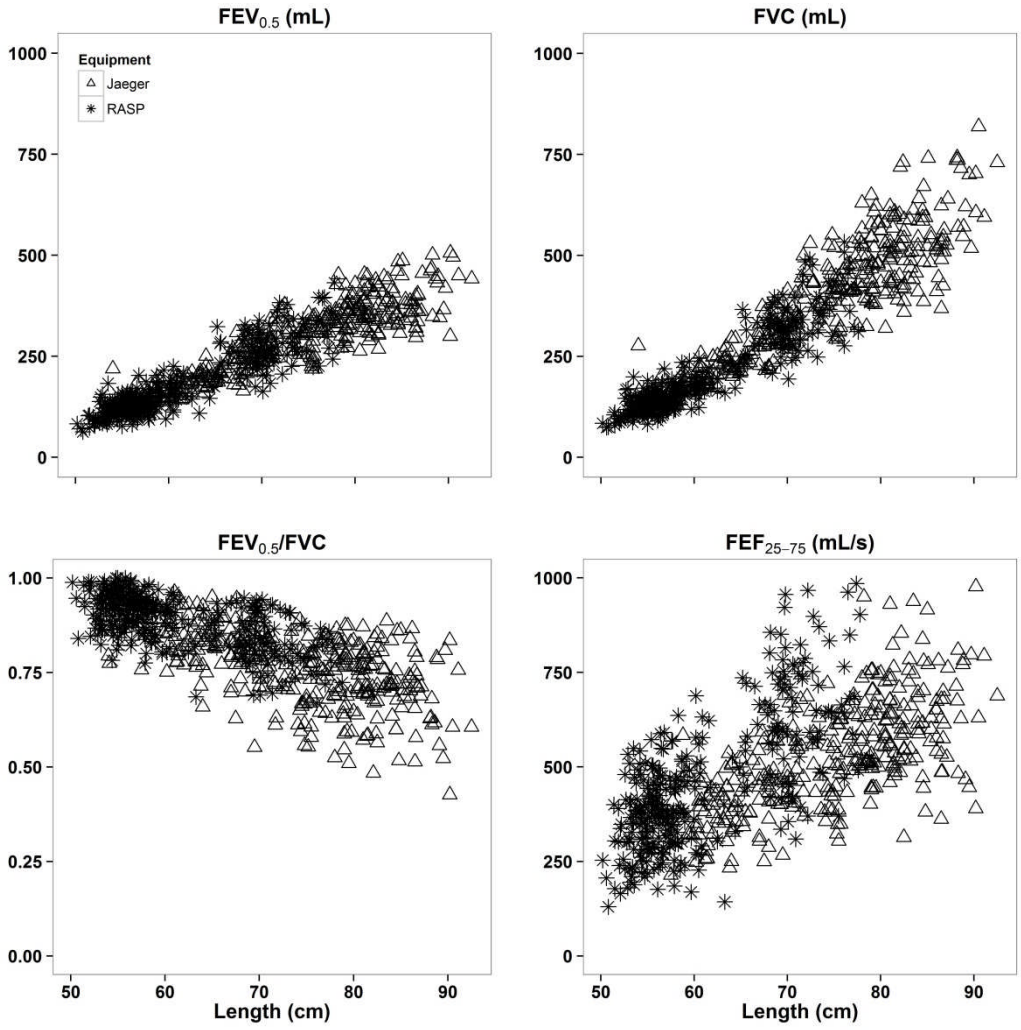
Data exclusions: To facilitate development of robust and reliable reference ranges for RVRTC outcomes based on sufficient sample size with relatively even spread over a wide range of age and body size in healthy infants, we excluded a few data points which were at the extremes of length and age range.

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.

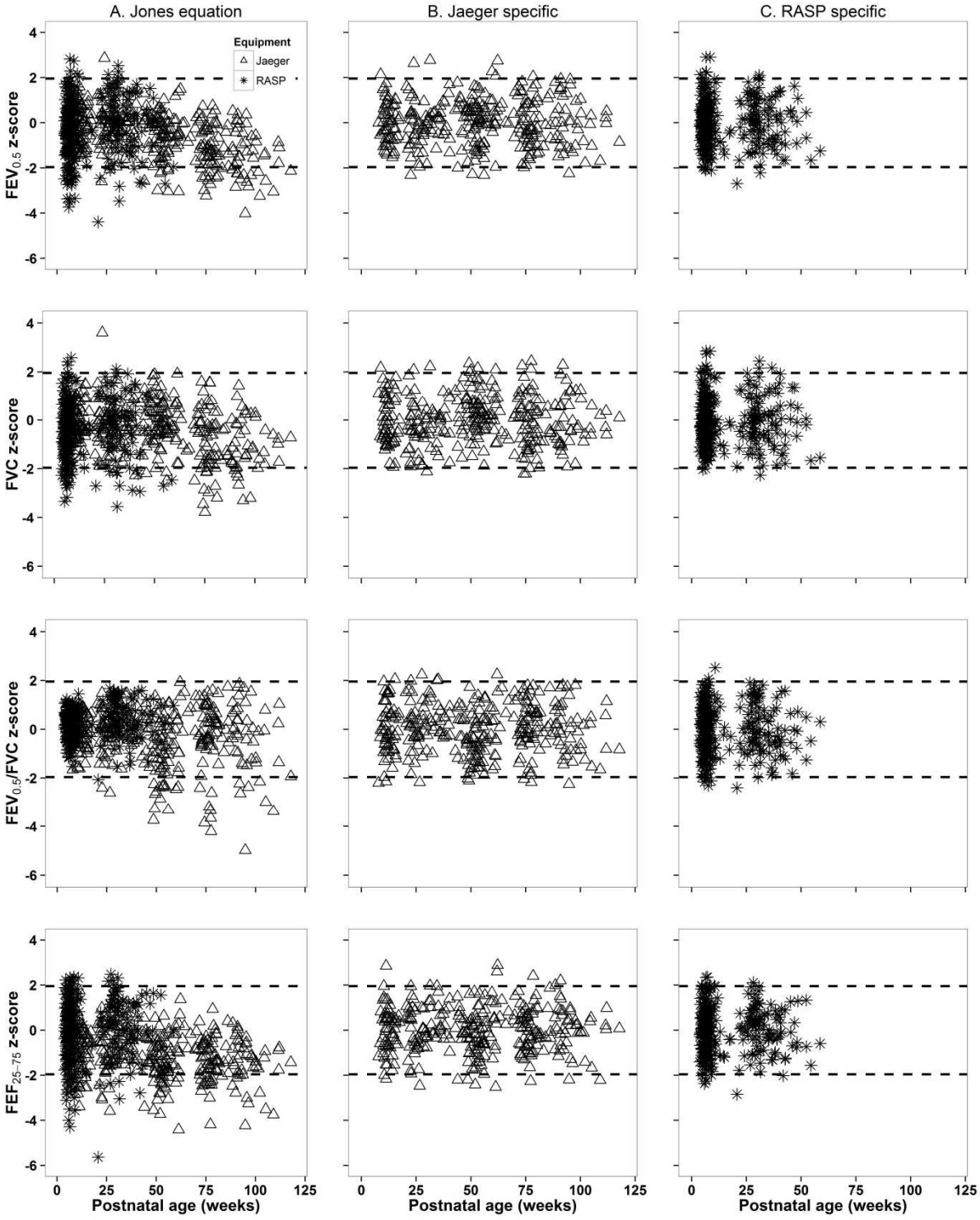
Table S1 RASP RVRTC prediction equations

	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.

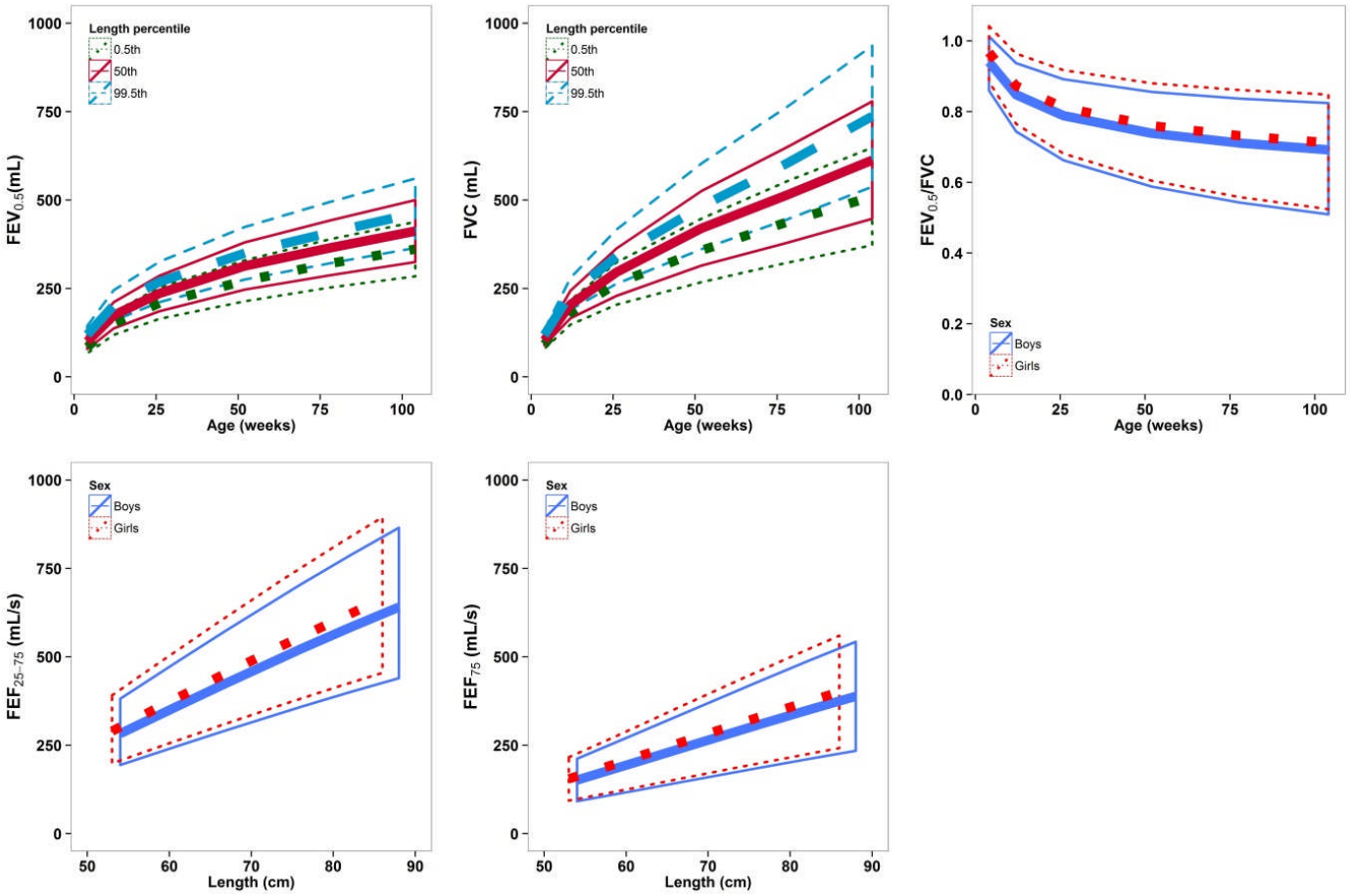
Figure S2 RVRTC data from healthy infants and young children plotted against age according to Jones et al[8] and the new equipment-specific equations



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

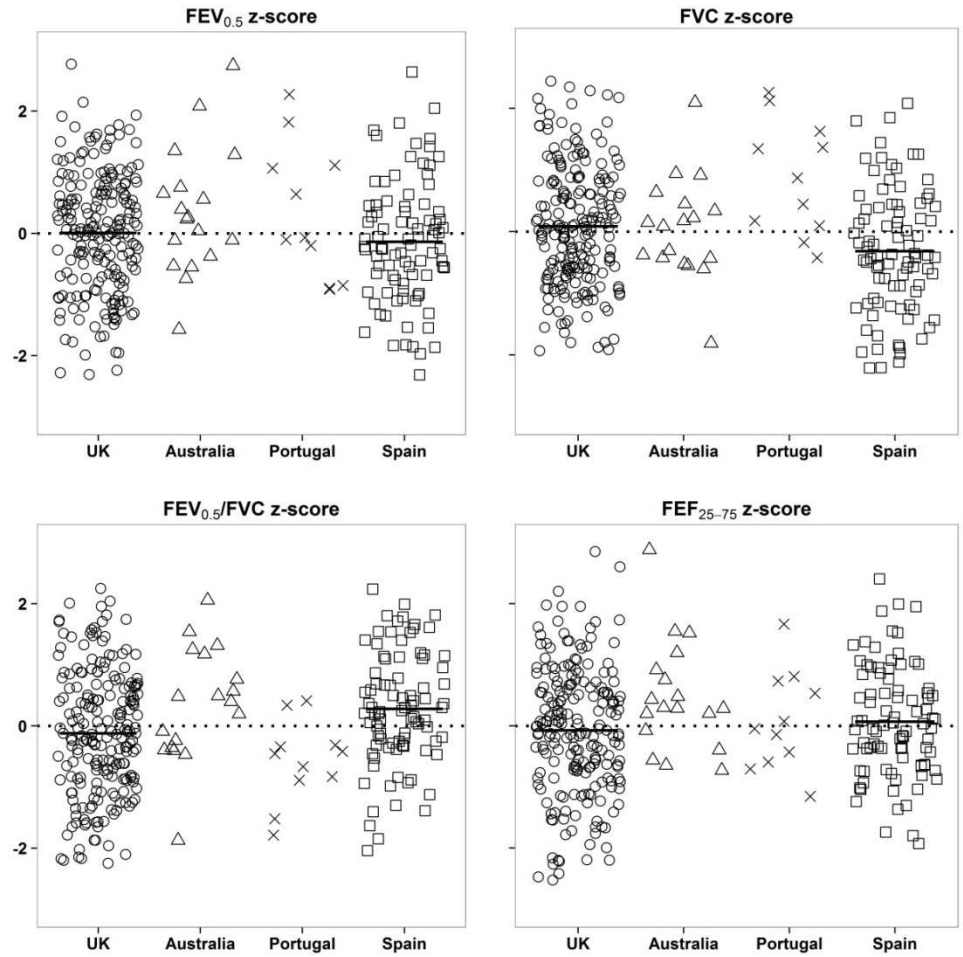


Fitted centiles for $FEV_{0.5}$, FVC and $FEV_{0.5}$/FVC were plotted against age as the spread (S) for these outcomes is age dependent. As the median (M) for $FEV_{0.5}$ and FVC is age and length dependent, the fitted centiles are shown for infants who may be of average length for age (50th percentile [**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.