Infant lung function: Criteria for selecting tidal flow-volume loops

Karen Eline Stensby Bains, Hrefna Katrín Gudmundsdóttir, Martin Färdig, Erik Amnö, Christine M. Jonassen, Björn Nordlund, Eva Maria Rehbinder, Håvard O. Skjerven, Corina Silvia Rueegg, Riyas Vettukattil, Karin C. Lødrup Carlsen

Infant lung function: Criteria for selecting tidal flow-volume loops

Karen Eline Stensby Bains, MD[1,2], Hrefna Katrín Gudmundsdóttir*, MD[1,2], Martin Färdig*, MSc[3,4], Erik Amnö, MD[3], Christine M. Jonassen, PhD[5,6], Björn Nordlund, RN, PhD[3,4], Eva Maria Rehbinder, MD, PhD[1,7], Håvard O. Skjerven, MD, PhD[1,2], Corina Silvia Rueegg, PhD[8], Riyas Vettukattil[1,2], PhD, Karin C. Lødrup Carlsen, MD, PhD[1,2].

*shared second authorship

Affiliations:

1. University of Oslo, Faculty of Medicine, Institute of Clinical Medicine, Oslo, Norway

2. Division of Paediatric and Adolescent Medicine, Oslo University Hospital, Oslo, Norway

3. Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden

4. Department of Women´s and Children´s Health, Karolinska Institutet, Stockholm, Sweden

5. Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway

6. Centre for Laboratory Medicine, Østfold Hospital Trust, Kalnes, Norway

7. Department of Dermatology, Oslo University Hospital, Oslo, Norway

8. Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway

Author contributions:

All authors have contributed substantially to the design and/or data collection and/or clinical follow-up of the PreventADALL study, revised the work critically for important intellectual content and approved the final version before submission.

Corresponding author:

Karen Eline Stensby Bains, MD.

Department of Pediatrics, Ullevål, Oslo University Hospital. Postboks 4956 Nydalen. 0424 Oslo, Norway.

E-mail: karen.eline.sh@gmail.com

Mobile: +4797745444.

Short title: Manual evaluation of infant lung function tests

Word count abstract: 249.

Word count manuscript (excluding figures and references): 3055.

No. of tables: 5.

No. of figures: 3.

Key words: Pulmonary function test, tidal flow-volume loops, ICC, reliability, infants, PreventADALL.

What's Known on This Subject: There is no clear consensus on criteria for manual inclusion or exclusion of flow-volume loops in tidal flow volume tests in awake infants.

Take-home message: A predefined set of inclusion and exclusion criteria provided a reliable and standardized procedure for manual selection of tidal flow-volume loops in infants that may be useful in clinical as well as research settings.

Abbreviations:

PreventADALL – Preventing Atopic Dermatitis and ALLergies

SOP – Standard Operating Procedure

TFV – Tidal flow-volume

t_{PTEF} – Peak tidal expiratory flow

t_E: time to expiration

t_{PTEF}/t_E – time to peak tidal expiratory flow to expiratory time

V_E: Expiratory volume

ICC – intra class correlation coefficient

V_T: tidal volume

V_T/kg: tidal volume per kilo

RR: Respiratory rate

Abstract

Tidal flow-volume (TFV) loops are commonly recorded in infants during sleep, due to the more regular breathing patterns compared to the awake state. Standardized deselection of loops outside pre-specified ranges are based on periods of regular breathing, while criteria and available software for visual evaluation of TFV loops are lacking. We aimed to determine the reliability of standardized criteria for manual selection of infant TFV loops.

Using a predefined set of criteria, three independent raters manually evaluated TFV loops among 57 randomly selected awake healthy 3-month-old infants with available TFV measurements in the Scandinavian PreventADALL study. The TFV loops were sampled using the Eco Medics Exhalyzer D. Criteria for selecting TFV loops included reproducible shape and volume with only one peak in tidal expiratory flow (PTEF), excluding loops with no clear or uneven flow towards PTEF. By intra class coefficient (ICC), the reliability of agreement between raters was determined for the time to PTEF to expiratory time (t_{PTEF}/t_E) and other TFV loop parameters.

Five infants had unsuccessful tests. Among the remaining 52 infants the raters selected a median of 25, 26 and 15 loops per test, respectively. The ICC (95% confidence intervals) were 0.97 (0.92, 0.98) for t_{PTEF}/t_E, 0.99 (0.99, 1,00) for respiratory rate, 0.98 (0.97, 0.99) for tidal volume/kg and 0.98 (0.97, 0.99) for expiratory volume, reflecting excellent agreement in all categories.

Manual TFV loops selection using standardized criteria provides a reliable alternative for lung function measures in awake infants with interrupted breathing cycles in a real-life setting.

Introduction

Infant lung function testing has been used to assess lung development, impact of environmental factors and to detect lung disease. By tracking through childhood and adolescence, lung function in infancy is a major predictor of adult lung function [1-3]. Measures of lung function in awake young children include tidal breath flow-volume loops [4], representing compound measurements of lung function, including size of airways, mechanical characteristics of the lung [5] and respiratory control [6]. During tidal flow-volume (TFV) loop sampling, abnormal patterns of breathing and airway obstruction may be exposed [7]. TFV loop measurements correlate with forced expiratory measurements [8, 9], and lower values of ratio of time to peak tidal expiratory flow to expiratory time (t_{PTEF}/t_E) in infancy are associated with chronic lung disease, wheeze in infancy and asthma later in life [10-14].

Tidal breathing measures have been obtained in awake and naturally sleeping infants and children, sometimes under sedation [15]. While it is possible to obtain lung function tests from awake preschool children actively participating, it has historically been challenging to assess lung function in infants without sedation or during natural sleep. Commonly, chloral hydrate has been used as a sedative [16] in in- and outpatient facilities [17]; a drug negatively affecting normal ventilation [18], associated with several cases of overdosing, respiratory depressions, cardiopulmonary arrests and fatal events [17]. As TFV loops often are more easily obtained during sleep, and breathing cycles are less likely to be interrupted, measurements during sleep has been preferred in infants and young children [8, 19]. However, measurements in the awake state may be advantageous as children are more likely to be awake than sleeping at clinical investigations. Furthermore, lung function in older children is measured in the awake state, and measures obtained in the awake compared to sleeping infants may be less influenced by external factors [20]. Associations between t_{PTEF}/t_E

and maternal smoking in-utero, and future asthma are observed in both awake [21, 22] and sleeping state [23, 24]. The clinical value of TFV measures on an individual level is debated [25, 26], partly due to the lack of reference values [27]. Guidelines for TFV measures are established [27], largely based on examination of sleeping or sedated children [19], while sources of variability and criteria for selection of loops are unclear. In the commonly used software, the only option for automatic selection of loops is by pre-defining a threshold for maximum deviation of millilitres from the median tidal volume (V_T) and valid minute ventilation range in liters per minute in a test.

The ATS/ERS guidelines state that automatic breath detection should be accompanied by a visual evaluation of the flow and volume signals, however there is no clear consensus on criteria for manual inclusion or exclusion of flow-volume loops in a test [27]. There is a need for a validated standard operating procedure (SOP), with clear criteria for inclusion and exclusion of TFV loops, for use in clinical as well as research setting, regardless of arousal state. Therefore, the aim of the present study was to determine the reliability of a predefined set of criteria for manual selection of TFV loops in infants.

Material and methods

Study subjects

Three independent raters evaluated TFV loop measures in 57 randomly selected infants with available lung function at three months of age, antenatally enrolled in the general population-based prospective mother-child birth cohort study; Preventing Atopic Dermatitis and ALLergies in children (PreventADALL) [28].

Two raters from Oslo university Hospital and one from Karolinska University Hospital had access to and evaluated TFV measures stored in a secure data server at the University of Oslo. All three raters were medical doctors with clinical experience from general paediatric medicine under training within paediatric pulmonology. Two raters (KESB and HKG) both performed infant lung function testing and evaluated the TFV loops, while the third (EA) participated in the loop evaluation and selection process only.

The PreventADALL study recruited 2697 pregnant women from Norway (Oslo University Hospital and Østfold Hospital Trust) and Sweden (the region of Stockholm) around 18 weeks pregnancy from December 2014 to October 2016, and their healthy infants born at or after gestational week 35.0. In the present study, the source population consisted of healthy, awake infants with lung function measurements obtained by study personnel in Oslo.

The PreventADALL study was approved by the Regional Committee for Medical and Health Research Ethics in South-Eastern Norway (2014/518) and Sweden (2014/2242-31-4). The study was registered at ClinicalTrials.gov (NCT02449850). Written informed consent was obtained from the pregnant women at enrolment and from parents at inclusion of the newborn infant.

Procedures

Trained study personnel measured lung function in awake infants at the follow-up examination at three months of age. Infants were calm and positioned supine in either a

stroller or on a firm pillow on caregiver's arm or lap, with head and neck in midline. The TFV loops were sampled using the Eco Medics Exhalyzer D® (Duernten, Switzerland) with ultrasonic flow meter attached to a CO_2 adapter and a dead space reducer (Set 1) with spirette (Spirette[TM], Eco Medics, Switzerland) and a tight-fitting face mask with inflated cuff covering nose and mouth to avoid air leaks. The equipment was calibrated daily for atmospheric pressure, temperature, and channel calibration, whereas flow calibration was executed between every subject. Analyses with the Spiroware® software version 3.2.1 were in line with international guidelines on infant lung function testing [15, 27]. A test run constituted of consecutive TFV loops to a maximum of 100 loops, as defined as cut-off in the software. Further details on lung function testing are outlined in online supplement 1.

A set of predefined criteria was outlined in a standard operating procedure (SOP) for manual selection of TFV loops, developed by the raters together with senior researchers in the field of tidal breathing measurements. Details of the SOP are given in the online supplement 2, with the main criteria illustrated in Figure 1 and 3. Briefly, the loops should be reproducible with fairly even shape and similar volumes with only one peak on expiratory flow, while allowing some normal variation as expected in a healthy child. Both consecutive and non-consecutive breaths were saved when deemed reproducible and with little deviation of volumes from mean tidal volume (V_T) during expiration or inspiration. Explicit criteria for exclusion of loops were no clear peak tidal expiratory flow (PTEF) or loops with an aborted or uneven flow towards PTEF at the beginning of the expiratory phase. Each lung function test was eventually rated into one of three pre-defined quality categories; successful, partly successful and not successful. A successful test was defined as a test with good reproducible quality and included preferably at least 10 loops. A partly successful test was defined as a test including fewer accepted loops or where reproducibility in loop shape or selected variables was uncertain, showing a greater variance in-between the concluded loops. Not successful tests

were of poor quality, with uncertainty whether the loops represented the infants' normal breathing or included no saved loops.

All three raters independently evaluated, selected, and rated all TFV loops sampled in 57 infants and recorded the observations electronically. The infants were randomly selected from a list of all infants attending the 3-month follow-up visit at Oslo University Hospital, using random sampling in SPSS. Each rater independently worked successively through the list to identify infants with lung function measures in the awake state and then evaluated the available loops to classify and qualify each test run. Thereafter, each rater independently deemed the test successful, partly successful or not successful, based upon a general evaluation of the test in relation to the criteria in the SOP (see Figure 1 and online supplement 2 for details). Samples were scrutinized and stored within a safe storage at the Services for Sensitive Data unit (TSD) at the University of Oslo [28].

For comparison of lung function measurement variables between manual and automatic selection of loops, system settings used for automatic selection of loops were set to standard from the manufacturer as described in online supplement 2. Results of software selected loops were reported electronically without any manual correction.

Definitions and outcomes

The primary outcome was the level of agreement among different raters of t_{PTEF}/t_E, and secondary outcomes were the agreement between the raters of t_{PTEF}/t_E categories <0.20, 0.20<0.25, and ≥0.25, as well as respiratory rate (RR), tidal volume per kilo (V_T/kg) and expiratory volume (V_E).

Statistical analysis

Descriptive statistics are presented in numbers and proportions for categorical variables and mean or median with standard deviation (SD) or minimum and maximum for continuous variables.

For the reliability analysis of the continuous variables, we calculated the intraclass correlation coefficient (ICC) with a two-way random-effects model [29] as we were evaluating a rater-based clinical assessment method where the three raters had similar characteristics, and we planned to generalize our reliability results to other raters. The analysis was based on the single-rater-type, with the reliability experiment in this paper comparing the actual rating of three independent raters. Absolute agreement for the outcomes was assessed by calculating ICC estimates and their 95% confidence intervals (95% CI) using IBM® SPSS® statistics version 25 (Chicago, IL, USA). ICC values above 0.90 indicate excellent reliability, values between 0.75 and 0.90 good reliability and values between 0.50 and 0.75 indicate moderate reliability [29]. The P0 (the null value of the ICC) was set to 0.6.

Agreement between raters for categorical variables are reported descriptively. We investigated whether the number of loops selected differed between raters using linear regression and calculating robust standard errors to adjust for the cluster "infant". This analysis was performed in STATA version 17.0 and including the 156 tests included in the main analysis of the paper. To have a statistical power of 80% to detect significant agreement exceeding 0.74, with an alpha level of 5% and three raters, the study required TFV loop tests from 53 subjects.

Results

The 57 infants (63% boys) had a mean (min, max) gestational age of 39.8 (36.6, 42.9) weeks and a mean weight at three months of age of 6.2 (4.6, 8.2) kg (Table 1). In five infants no TFV loops were saved after rater assessment by at least one rater and the lung function measurements of these infants were thus not included in the ICC analysis, see online supplement table S1 for details. The median number of loops saved by each rater per test among the 52 infants was 25, 26 and 15, respectively (Table 2A), while the software selected a median of 8 loops per test (Table 3B). In the tests concluded automatic by the software, there were examples of loops with aborted flows and double PTEF. Rater 2 selected on average 2.3 loops more than rater 1 (coeff 2.3, 95% CI 0.3,4.3 (p<0.001)) and rater 3 selected on average 7.6 loops less than rater 1 (coeff -7.6, 95% CI -9.4, -5.8 (p<0.001)).

The mean (SD) ratio of t_{PTEF}/t_E was 0.39 (0.08), 0.41 (0.08) and 0.39 (0.09) for the three raters (Table 3A), with an ICC of 0.97 (95% CI 0.92, 0.98; Figure 2A). The corresponding mean (SD) ratio of t_{PTEF}/t_E selected by the software was 0.52 (0.22; Table 3B).

The ICC for RR, V_T/kg and V_E was 0.99 (95% CI 0.99, 1.0), 0.98 (95% CI 0.97, 0.99) and 0.98 (95% CI 0.97, 0.99) respectively, as shown in figures 2B-2D. The ICC for t_{PTEF}/t_E, RR, V_T/kg and V_E for tests including 10 loops or more by each rater (n=37) was similar, as reported in online supplement Table S2.

None of the 52 infants had a t_{PTEF}/t_E ratio <0.20, while two infants had a ratio <0.25 as reported by either one or two of the raters. All other infants (96.2%) had a t_{PTEF}/t_E ratio ≥0.25 reported by all raters.

All three raters agreed on the quality category in 41 of the 57 infants (72%), with three of these deemed as not successful by all, see Table 5 for details. Selected TFV measurement

parameters for infants where one, two or three raters disagreed on the quality of the test are

listed in Table 4.

Discussion

The reliability of a set of predefined criteria for manual selection of TFV parameters among healthy 3-month-old infants was excellent, with an ICC of 0.97 (0.92, 0.98) for the t_{PTEF}/t_E ratio between three independent raters. Likewise, the ICC was higher than 0.90 for RR, V_T/kg and V_E.

To the best of our knowledge, this is the first study to validate a set of predefined criteria for individually evaluating TFV loops in awake infants, showing excellent agreement between three independent raters on the t_{PTEF}/t_E ratio, despite varying number of loops approved by each rater. In 25 sleeping infants in the first week of life Yuksel et al. [26] reported good interobserver repeatability between two observers of t_{PTEF}/t_E measured in a whole body plethysmograph, using the method of Bland and Altman.

While none of the infants had a t_{PTEF}/t_E <0.20, two infants were rated to have a t_{PTEF}/t_E between 0.20 and 0.25 by one or two raters, while all raters agreed on a t_{PTEF}/t_E ratio of 0.25 or higher in 52 of the 54 infants (96.3%). A cut-off value of t_{PTEF}/t_E ratio below 0.20 is associated with later bronchial obstruction [11, 30], whereas ratio values of 0.25 and higher have been regarded normal [8, 14, 26, 30, 31]. While low t_{PTEF}/t_E appears clinically relevant, the PreventADALL study is based on a normal population and we assumed, as observed, that the majority of included infants would have lung function values in the normal range. Therefore, we categorised the t_{PTEF}/t_E ratio into three categories of low (<0.20), marginal (0.20<0.25) and normal (≥0.25) values for comparison, in addition to the exact values included in the ICC analyses.

The ICC for t_{PTEF}/t_E was consistent across a varying number of tidal breathing loops, with the three raters approving a median of 15, 25 and 26 loops, where on average rater 2 saved more and rater 3 saved fewer loops per test than rater 1. The automatic selection of loops by the software in general resulted in fewer loops, with a smaller range, higher mean ratio of the t_{PTEF}/t_E, lower mean V_T/kg and V_E and similar mean RR as compared to the three raters. The large discrepancy also indicates that faulty loops such as double peaks and irregular shaped loops were not deselected by the automatic process. We therefore suggest that manually selected loops are more likely to be representative and of higher quality than are loops selected automatically. Stocks et al. suggested in 1994 that for infants older than six weeks of age 10 breath loops might be adequate, whereas in younger infants a t_{PTEF}/t_E based on the mean of 15-20 loops would be a closer estimate to their true value due to decreased within-subject variability with increased age [19]. A previous study has however documented tidal flow-volume loop indices based on only four loops selected from a preview of eight loops, due to data storage capacity at the time [32]. The ERS/ATS guidelines on pulmonary function testing for preschool children suggest that a reliable t_{PTEF}/t_E should be based on at least 10 loops [15] and these criteria are widely used for TFV measurements in infants as well. However, based on our results, including varying number of included loops by the raters, as shown in Table 2, we suggest that tests may be valid even with fewer than 10 loops.

We included both tests deemed successful, partly successful and not successful and found a consistent ICC of excellent agreement in both t_{PTEF}/t_E, as well as RR, V_T/kg and V_E. Defining the quality of a single test was based on the visual shape and reproducibility of the loops, after manual removal of loops with poor technical quality or without a well-defined PTEF. There are no clear-cut criteria for evaluation of the quality of the tests, however the criteria are outlined in the SOP for lung function analysis in the study provided in the online supplement

2. This can explain why the raters have deemed tests into different quality categories and saved different number of loops, where rater 3 on a general basis has saved fewer loops from the tests and in general deemed more tests partly successful compared to the other two raters. Despite these discrepancies, the ICC is high for the continuous variables.

The criteria for manual selection of TFV curves in a real-life setting provided sufficiently robust criteria for excellent agreement on the tests, supporting the usefulness of the criteria.

We are not aware of other studies that compare lung function variables from awake TFV measurements manually evaluated by several independent raters.

Strengths and limitations

The TFV tests were performed on healthy awake infants with characteristics reflecting a normal population under standardized circumstances by trained personnel (5, 25). The number of infants was pre-defined by power calculations to include sufficient number of tests. The study requirement of 53 infants to ensure a statistical power of 80% to detect significant agreement exceeding 0.74 was not met, however with 52 infants included in calculations resulting in an ICC above 0.90 for all variables, it is unlikely that including one more infant would affect the outcome.

There was little variance in the predefined categories of ratios of <0.20, $0.20<0.25$, and ≥0.25 defined to distinguish an assumed healthy infant from an infant with reduced lung function. The category based on t_{PTEF}/t_E $0.20<0.25$ is somewhat arbitrary but was predefined as being in the lower range of presumably normal TFV loops. The high ICC, reflecting excellent agreement between different raters, was evident for the continuous variables in all outcomes.

The present study provides a further step to standardise TFV measures in epidemiologic studies and clinical practise, in line with the need for further insight into lung function measurement techniques, allowing repeated measurements in awake young children [4]. It remains unclear if selecting loops by this method will be useful in the clinical practice, and further studies should be conducted to validate the use for long term care of patients.

Conclusion

Using a set of predefined selection criteria, manual selection of TFV loops from healthy awake three-month-old infants resulted in excellent agreement of TFV parameters between three independent raters. Our study provides a feasible and valid tool for selecting TFV measures in infants, that may particularly be useful in the absence of long sequences of regular breathing, such as in daily clinical practice.

Affiliations:

1. Division of Paediatric and Adolescent Medicine, Oslo University Hospital, Oslo, Norway

2. VID Specialized University, Oslo, Norway

3. Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden

4. Department of Women´s and Children´s Health, Karolinska Institutet, Stockholm, Sweden

References

1.   Stern DA, Morgan WJ, Wright AL, et al. Poor airway function in early infancy and lung function by age 22 years: a non-selective longitudinal cohort study. *Lancet* 2007: 370(9589): 758-764.

2.   Martinez FD. Early-Life Origins of Chronic Obstructive Pulmonary Disease. *N Engl J Med* 2016: 375(9): 871-878.

3.   Stocks J, Hislop A, Sonnappa S. Early lung development: lifelong effect on respiratory health and disease. *Lancet Respir Med* 2013: 1(9): 728-742.

4.   Moeller A, Carlsen KH, Sly PD, et al. Monitoring asthma in childhood: lung function, bronchial responsiveness and inflammation. *European respiratory review : an official journal of the European Respiratory Society* 2015: 24(136): 204-215.

5.   van der Ent CK, Brackel HJ, van der Laag J, et al. Tidal breathing analysis as a measure of airway obstruction in children three years of age and older. *American journal of respiratory and critical care medicine* 1996: 153(4 Pt 1): 1253-1258.

6.   Seddon PC, Davis GM, Coates AL. Do tidal expiratory flow patterns reflect lung mechanics in infants? *American journal of respiratory and critical care medicine* 1996: 153(4 Pt 1): 1248-1252.

7.   Pijnenburg MW, Baraldi E, Brand PL, et al. Monitoring asthma in children. *Eur Respir J* 2015: 45(4): 906-925.

8.   Hevroni A, Goldman A, Blank-Brachfeld M, et al. Use of tidal breathing curves for evaluating expiratory airway obstruction in infants. *J Asthma* 2018: 55(12): 1331-1337.

9.   Morris MJ, Lane DJ. Tidal expiratory flow patterns in airflow obstruction. *Thorax* 1981: 36(2): 135-142.

10.   Schmalisch G, Wilitzki S, Wauer RR. Differences in tidal breathing between infants with chronic lung diseases and healthy controls. *BMC Pediatr* 2005: 5: 36.

11.   Haland G, Carlsen KC, Sandvik L, et al. Reduced lung function at birth and the risk of asthma at 10 years of age. *N Engl J Med* 2006: 355(16): 1682-1689.

12.   Lodrup Carlsen KC, Carlsen KH, Nafstad P, et al. Perinatal risk factors for recurrent wheeze in early life. *Pediatr Allergy Immunol* 1999: 10(2): 89-95.

13.   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(3): 700-705.

14.   Young S, Arnott J, Le Souef PN, et al. Flow limitation during tidal expiration in symptom-free infants and the subsequent development of asthma. *The Journal of pediatrics* 1994: 124(5 Pt 1): 681-688.

15.   Beydon N, Davis SD, Lombardi E, et al. An official American Thoracic Society/European Respiratory Society statement: pulmonary function testing in preschool children. *Am J Respir Crit Care Med* 2007: 175(12): 1304-1345.

16.   Lesnick BL, Davis SD. Infant pulmonary function testing: overview of technology and practical considerations--new current procedural terminology codes effective 2010. *Chest* 2011: 139(5): 1197-1202.

17.   Paparella SF. Chloral Hydrate: Safety Risks Still Worth Mentioning. *Journal of emergency nursing: JEN : official publication of the Emergency Department Nurses Association* 2018: 44(1): 81-83.

18.   Aldrete JA, Itkin IH. Effects of chloral hydrate on the respiration of nonasthmatic and asthmatic patients. *The Journal of allergy* 1969: 43(6): 342-348.

19.   Stocks J, Dezateux CA, Jackson EA, et al. Analysis of tidal breathing parameters in infancy: how variable is TPTEF:TE? *American journal of respiratory and critical care medicine* 1994: 150(5 Pt 1): 1347-1354.

20.   Carlsen KCL, Magnus P, Carlsen K-H. Lung function by tidal breathing in awake healthy newborns and infants. *European Respiratory Journal* 1994: 7: 1600-1668.

21.   Carlsen KCL, Jaakkola JJK, Nafstad P, et al. In utero exposure to cigarette smoking influences lung function at birth. *European Respiratory Journal* 1997: 10: 1774-1779.

22.   Haland G, Carlsen KC, Sandvik L, et al. Reduced lung function at birth and the risk of asthma at 10 years of age. *N Engl J Med* 2006: 355(16): 1682-1689.

23.   Stick SM, Burton PR, Gurrin L, et al. Effects of maternal smoking during pregnancy and a family history of asthma on respiratory function in newborn infants. *Lancet* 1996: 348(9034): 1060-1064.

24.   Turner S, Fielding S, Mullane D, et al. A longitudinal study of lung function from 1 month to 18 years of age. *Thorax* 2014: 69(11): 1015-1020.

25.   Carlsen KC. Tidal Breathing Measurements. *In:* Hammer J, Eber E, eds. Paediatric Pulmonary Function Testing. Karger, Basel, 2005; pp. 10-19.

26.   Yuksel B, Greenough A, Giffin F, et al. Tidal breathing parameters in the first week of life and subsequent cough and wheeze. *Thorax* 1996: 51(8): 815-818.

27.   Bates JH, Schmalisch G, Filbrun D, et al. Tidal breath analysis for infant pulmonary function testing. ERS/ATS Task Force on Standards for Infant Respiratory Function Testing. European Respiratory Society/American Thoracic Society. *Eur Respir J* 2000: 16(6): 1180-1192.

28.   Lodrup Carlsen KC, Rehbinder EM, Skjerven HO, et al. Preventing Atopic Dermatitis and ALLergies in Children-the PreventADALL study. *Allergy* 2018: 73(10): 2063-2070.

29.   Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016: 15(2): 155-163.

30.   Lodrup Carlsen KC, Carlsen KH, Nafstad P, et al. Perinatal risk factors for recurrent wheeze in early life. *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology* 1999: 10(2): 89-95.

31.   Martinez FD, Morgan WJ, Wright AL, et al. Diminished lung function as a predisposing factor for wheezing respiratory illness in infants. *The New England journal of medicine* 1988: 319(17): 1112-1117.

32.   Lodrup Carlsen KC, Stenzler A, Carlsen KH. Determinants of tidal flow volume loop indices in neonates and children with and without asthma. *Pediatr Pulmonol* 1997: 24(6): 391-396.

Tables and figures

Table 1 Characteristics for the 57 infants with tidal flow-volume measures assessed by three independent raters.

Boys, n (%)	36 (63.2)
	Mean (min, max)
Gestational age at birth (weeks)	39.8 (36.6, 42.9)
Birthweight (kg)	3.6 (2.8, 4.9)
Birth length (cm)	51 (48.0, 56.0)
Weight at 3 months (kg)	6.2 (4.57, 8.18)
Length at 3 months (cm)	62 (56.5, 67.9)

Table 2 The median (min, max) number of loops per rater for a) the 52 tests that all three raters found appropriate for further analyses, and b) tests including at least 10 loops.

A			**B**		
Rater	Tests (n)	Number of loops, median (min, max)	Rater	Tests (n)	Number of loops, median (min, max)
1	52	25 (3, 53)	1	45	30 (10, 53)
2	52	26 (3, 67)	2	44	29 (11, 67)
3	52	15 (2, 52)	3	37	24 (10, 52)

Table 3 Median (min, max) number of loops and mean (SD) t_{PTEF}/t_E, RR, V_T/kg and V_E per rater for a) infants where data from tests were saved by all three raters (n=52), and b) all infants where data were saved by one rater and data selected by the software.

A

Rater	Infants (n)	Median (min, max) number of loops	t_{PTEF}/t_E	RR	V_T/kg	V_E
1	52	25 (3, 53)	0.39 (0.08)	67 (13.4)	6.4 (2.1)	39.1 (12.2)
2	52	26 (3, 67)	0.41 (0.08)	66 (13.1)	6,7 (2.1)	40.2 (12.2)
3	52	15 (2, 52)	0.39 (0.09)	65 (13.6)	6.9 (2.1)	41.5 (12.0)

B

Rater	Infants (n)	Median (min, max) number of loops	t_{PTEF}/t_E	RR	V_T/kg	V_E
1	54	23 (3, 53)	0.39 (0.08)	67 (13.3)	6.3 (2.2)	38.4 (12.5)
2	56	25 (1, 67)	0.43 (0.11)	66 (13.0)	6.4 (2.2)	38.9 (12.9)
3	52	15 (2, 52)	0.39 (0.09)	65 (13.6)	6.9 (2.1)	41.5 (12.0)
Spiroware®	56	8 (1, 44)	0.52 (0.22)	64 (14.4)	5.4 (2.0)	29.3 (10.2)

Abbreviations:

t_{PTEF}/t_E: ratio of time to peak tidal expiratory flow to expiratory time

RR: respiratory rate

V_T/kg: tidal volume per kilogram

V_E: expiratory volume

Table 4 An overview of all infants with rater disagreement on quality of the test.

Infant	Rater	Quality of test	Loops before selection	Loops after selection	t_{PTEF}/t_E	V_T/kg (ml)
1	1	Successful		30	0.44	4.6
	2	Successful	60	22	0.49	5.4
	3	Partly successful		13	0.47	5.6
6	1	Partly successful		5	0.44	6.6
	2	Partly successful	44	9	0.55	6.9
	3	Not successful		3	0.45	8.3
8	1	Partly successful		8	0.40	2.4
	2	Partly successful	32	8	0.37	2.4
	3	Not successful		0	-	-
9	1	Partly successful		3	0.46	7.4
	2	Partly successful	38	5	0.53	6.5
	3	Not successful		2	0.51	7.7
13	1	Partly successful		6	0.38	11.6
	2	Successful	60	15	0.44	12.1
	3	Partly successful		7	0.39	11.8
15	1	Successful		19	0.54	7.3
	2	Successful	60	26	0.57	7.4
	3	Partly successful		8	0.52	7.4
22	1	Partly successful		5	0.31	4.9
	2	Partly successful	60	7	0.35	5.1
	3	Not successful		4	0.25	5.4
24	1	Successful		12	0.41	6.7
	2	Successful	60	20	0.46	8.4
	3	Partly successful		5	0.37	7.4
35	1	Successful		18	0.45	4.9
	2	Partly successful	60	6	0.45	6.1
	3	Not successful		3	0.51	7.1
39	1	Partly successful		3	0.44	3.3
	2	Partly successful	14	4	0.47	3.8
	3	Not successful		0	-	-
40	1	Successful		11	0.39	5.9
	2	Partly successful	38	11	0.47	6.4
	3	Not successful		3	0.47	6.0
41	1	Successful		11	0.31	5.5
	2	Successful	40	13	0.36	5.2
	3	Partly successful		7	0.29	6.1
47	1	Partly successful		5	0.26	4.4
	2	Partly successful	42	4	0.32	5.3
	3	Not successful		3	0.26	4.6
48	1	Successful		12	0.60	6.9
	2	Successful	50	11	0.59	7.4
	3	Partly successful		5	0.60	7.5
51	1	Successful		10	0.38	2.9
	2	Partly successful	56	7	0.51	4.6
	3	Not successful		4	0.39	5.7
53	1	Partly successful		3	0.34	1.6
	2	Partly successful	31	3	0.34	1.6
	3	Not successful		2	0.32	1.7

Abbreviations:

t_{PTEF}/t_E: time to peak tidal expiratory flow to expiratory time

V_T/kg: tidal volume per kilogram

Table 5 Conclusion by raters on quality of tests from all 57 infants

Conclusion on quality	Successful test	Partly successful test	Not successful test	Total number of infants
3 raters agreed	37	1	3	41
2 raters agreed	5	8	0	13
No agreement	4*	8*	10*	3

*) concluded by one rater

Figure 1 Criteria for inclusion and exclusion of loops in a test.

	Inclusion criteria	Exclusion criteria
Loop shape	Even	Aborted flow towards PTEF
Loop shape	Reproducible	No clear single PTEF
Volumetrics	Little deviation in flow, V_T, t_E, t_I	Large volume or flow deviation from mean
RR	Reproducible RR	Abnormally high RR

Abbreviations:

V_T: Tidal volume

t_E: Expiratory time

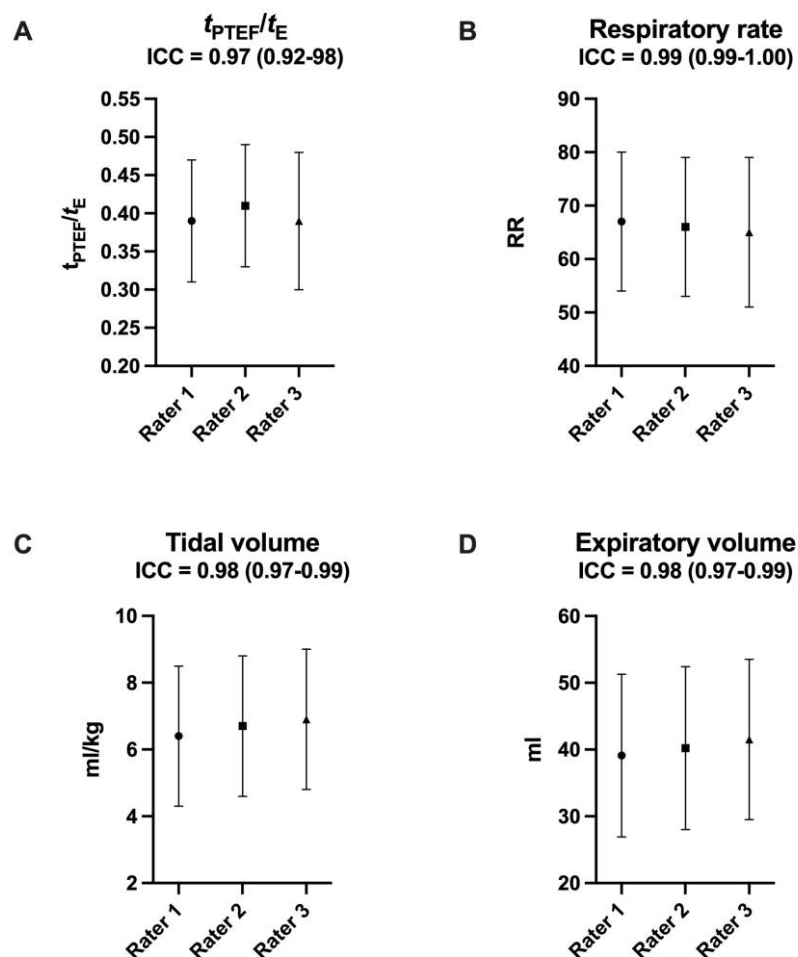t_I: Inspiration time

RR: Respiratory rate

PTEF: Peak tidal expiratory flow

t_{PTEF}: time to peak tidal expiratory flow

Figure 2 shows rater agreement with the mean ICC (95% CI) in heading of each figure and bars for each rater shows mean value with SD for a) t_{PTEF}/t_E, b) respiratory rate, c) tidal volume and d) expiratory volume, and rater bars showing individual mean with SD.

Abbreviations:

$t_{PTEF}/t_{E:}$ ratio of time to peak tidal expiratory flow to expiratory time

RR: Respiratory rate

ICC: intra class correlation

Figure 3 Tidal flow-volume loops. In a) examples of excluded loops (red) due to deviating volume (left), notch on expiratory flow (middle) and two expiratory flow peaks (right), and in b) a TFV test before (left) and after (right) manual selection of loops.
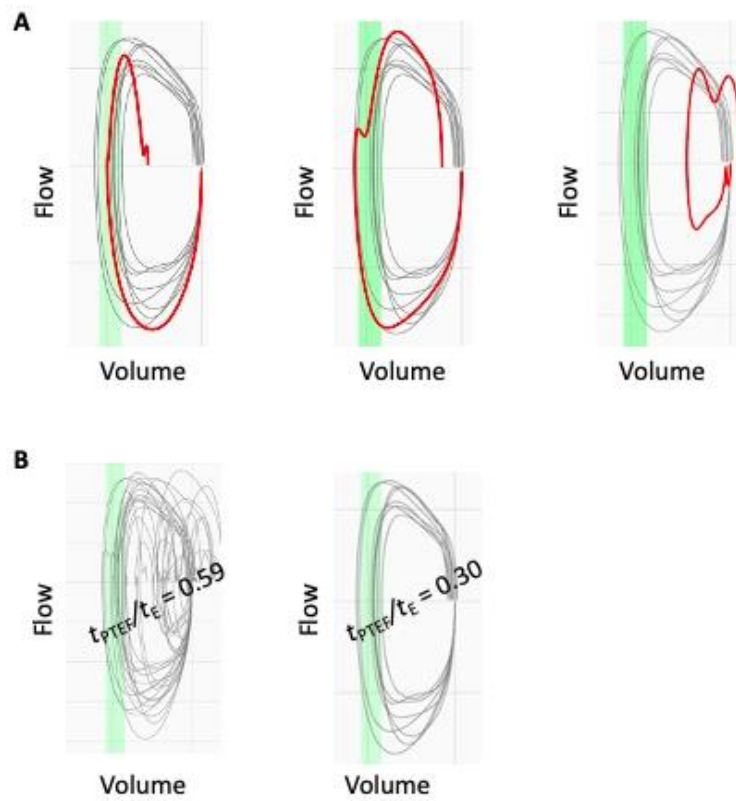
Abbreviations:

t_{PTEF}/t_E: ratio of time to peak tidal expiratory flow to expiratory time

Online supplementary material

Method supplementary 1

Standard operating procedure for TFV loop measurements

The equipment consisted of the lung function machine Exhalyzer® D, infant face mask (size 2), dead space reducer set 1 (DSR1) for babies up to 15 kilograms, bacteria filter, Spirette, calibration syringe 100 milliliters for babies up to 15 kilograms, nafion tube for gas sampling, flow head, CO2 adapter with capnostat sensor, SF6 gas tube, and disinfectant. The study personnel were qualified to perform lung function measurements after individual training. For safety considerations, the Exhalyzer® D and the study personnel's hands were disinfected and non-single use equipment was sterilized after each measure. The required equipment was assembled, the Exhalyzer® D connected to the air and oxygen through the central gas supply, calibrated for atmospheric pressure and room temperature (maximum deviation: temperature ± 0.5 °C, atmospheric pressure ± 5 hPa), flow (maximum deviation: inspiratory and expiratory flow 0.99%), channel (maximum deviation: ± 20 milliliter/second from the desired flow of 200 milliliter/second) and tracer gas prior to lung function measurements, according to the manufacturer's instructions. Prior to measurements, name, gender and date of birth was recorded. Length and weight of the baby was measured after the lung function testing, to ensure calm babies, and later merged with lung function variables. Lung function was measured while the infants were held in a semi-declined or declined position in parent's arms or lap lying on a firm pillow or lying flat in stroller to ensure similar positioning of the child regardless of placement. Essential requirements were that the infants had to be calm, preferably not fed right before the measure, lying comfortable, and were able to breathe freely and unrestrictedly. The mask was held with a stable grip with the whole hand around the mask placed over the mouth and nose of the infant, using fingers to control minimal leakage from the mask. When the infant breathed calmly and evenly into the mask, the tidal breath flow-volume loop (TFV) loop measurement was started. A series of at least 10 consecutive breaths was aspired. Arousal state, time to last feeding prior to measure, recently used inhalations or other medications and deviations from the standard operating procedure were documented.

Method supplementary 2

Standard operating procedure for quality assessment of TFV loop measurements

Spiroware® software version 3.2.1 was used for lung function analysis. For automatic selection of loops by the software, the automatic breath acceptance criteria were set to exclude curves $\pm$ 10% from median V_T and the valid minute ventilation range was set to minimum 0.5 and maximum 2.5 l/min as default by the manufacturer.

The following TFV loop parameters were included for manual analysis: time to peak tidal expiratory flow divided by expiratry time (t_{PTEF}/t_E) expiratory time (t_E), time to peak tidal expiratory flow (t_{PTEF}), expiratory volume (V_E), tidal volume (V_T), respiratory rate (RR) and volume at peak expiratory flow (V_{PTEF}).

When analyzing a test run, all breaths were visualized and selected parameters as mentioned above displayed on the screen for inspection for each individual breath. Firstly, all loops that were technically unacceptable (uneven shape with no clear peak tidal expiratory flow (PTEF) or double PTEF, kinks, notches, aborted or uneven flow towards PTEF, incomplete inspiration or expiration phase, deviant shapes or volumes) were removed.

Secondly, each breath was evaluated, with particular focus on the expiratory phase. The breaths ought to have a reproducible shape, only one PTEF and justly be at the same place on the time-flow loop as the rest of the selected curves, fairly even shapes with reproducible ratios and similar volumes, little deviation of volumes in inspiratory and expiratory phase from mean V_T, minor deviation of flow in inspiration and expiration from the mean of remaining curves, while still allowing normal variation, as the breathing pattern of healthy children may vary. Ideally consecutive breaths of equal shape should be saved when there were few (<10) loops, but also non-consecutive breaths could be saved when reproducible. Loops with abnormally high RR were excluded, with the aim of preferably reproducible RR. With all uncertainty, the three investigators erred on the side of normality.

Explicit criteria for exclusion of loops were loops with kink, notches, aborted or uneven flow towards PTEF and no clear PTEF or double PTEF.

Finally, the measurements were concluded, saved and reported. Upon report, the rater reported the quality of the test: "Successful" for tests that are of good, reproducible quality and preferably include at least 10 curves. "Partly successful" for tests that include few curves or when reproducibility in variables is uncertain. "Not successful" for tests of bad quality or uncertainty around whether the test represents the way the child breathes, or when no loops are saved.

Table S1 Tests deemed unsuitable for analysis by at least one of the three raters and thus excluded from the ICC analyses.

Infant	Rater	Quality of test	Initial number of loops	Loops after selection	t_{PTEF}/t_E	V_T/kg (ml)
3	1	Not successful		0	-	-
	2	Not successful	36	1	0.84	3.1
	3	Not successful		0	-	-
7	1	Not successful		0	-	-
	2	Not successful	24	0	-	-
	3	Not successful		0	-	-
8	1	Partly successful		8	0.40	2.4
	2	Partly successful	32	8	0.37	2.4
	3	Not successful		0	-	-
25	1	Not successful		0	-	-
	2	Not successful	33	2	0.72	3.6
	3	Not successful		0	-	-
39	1	Partly successful		3	0.44	3.3
	2	Partly successful	14	4	0.47	3.8
	3	Not successful		0	-	-

Abbreviations:

t_{PTEF}/t_E: ratio of time to peak tidal expiratory flow to expiratory time

V_T/kg: Tidal volume per kilogram

Table S2 ICC for tests by all raters

Included loops (range)	Infants (n)	t_{PTEF}/t_E	RR	V_T/kg	V_E
2-67	52	0.97 (0.92-0.98)	0.99 (0.99-1.0)	0.98 (0.97-0.99)	0.98 (0.97-0.99)
10-67[†]	37	0.98 (0.95-0.99)	0.99 (0.99-1.0)	0.99 (0.98-1.0)	0.99 (0.97-0.99)

[†]Tests including 10 loops or more

Abbreviations:

t_{PTEF}/t_E: ratio of time to peak tidal expiratory flow to expiratory time

RR: respiratory rate

V_T/kg: tidal volume per kilogram

V_E: expiratory volume