




Comparison of different analysis algorithms to calculate multiple-breath washout outcomes

Pinelopi Anagnostopoulou^{1,4}, Nadja Kranz^{1,4}, Jeremias Wolfensberger¹, Marisa Guidi¹, Sylvia Nyilas¹, Cordula Koerner-Rettberg², Sophie Yammine¹, Florian Singer ^{1,3} and Philipp Latzin¹

Affiliations: ¹Pediatric Respiratory Medicine, Dept of Pediatrics, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland. ²Dept of Pediatric Pulmonology, University Children's Hospital of Ruhr University Bochum at St Josef-Hospital, Bochum, Germany. ³University Children's Hospital Zurich, Zurich, Switzerland. ⁴Both authors contributed equally.

Correspondence: Philipp Latzin, Inselspital, Bern University Hospital, 3010 Bern, Switzerland.
E-mail: philipp.latzin@insel.ch

ABSTRACT Lung clearance index (LCI) is the main outcome of the multiple-breath washout (MBW) test. Current recommendations for LCI acquisition are based on low-grade evidence. The aim of this study was to challenge those recommendations using alternative methods for LCI analysis.

Nitrogen MBW measurements from school-aged children, 20 healthy controls, 20 with cystic fibrosis (CF) and 17 with primary ciliary dyskinesia (PCD), were analysed using 1) current algorithms (standard), 2) three alternative algorithms to detect with higher precision the end of MBW testing and 3) two alternative algorithms to determine exhaled tracer gas concentrations. LCI values, intra-test repeatability, and ability to discriminate between health and lung disease were compared between these methods.

The analysis methods strongly influenced LCI (mean±SD overall differences (%) between standard and alternative analysis methods: $-4.9\pm 5.7\%$; range: $-66-19\%$), but did not improve intra-test variability. Discrimination between health and disease was comparable as areas under the receiver operator curves were not greater than that from standard analysis.

This study supports current recommendations for LCI calculation in children. Alternative methods influence LCI estimates and hamper comparability between MBW setups. Alternative algorithms, whenever used, should be carefully reported.



@ERSpublications

Lung clearance index values are strongly affected by the algorithms used for the analysis
<http://ow.ly/h2Rs30ktPiN>

Cite this article as: Anagnostopoulou P, Kranz N, Wolfensberger J, *et al.* Comparison of different analysis algorithms to calculate multiple-breath washout outcomes. *ERJ Open Res* 2018; 4: 00021-2017 [<https://doi.org/10.1183/23120541.00021-2017>].



Introduction

Multiple-breath washout (MBW) is an increasingly acknowledged lung function test in patients with chronic lung disease [1, 2]. In principle, MBW measures the efficiency of tracer gas clearance from the lungs across multiple breaths. Lung clearance index (LCI) is the most commonly used MBW outcome. LCI estimates global ventilation inhomogeneity, a biomarker of central and peripheral airway obstruction, for example in cystic fibrosis (CF) [3, 4] and primary ciliary dyskinesia (PCD) [5, 6].

Yet MBW methodology is complex and requires signal processing and breath-by-breath analyses. A series of mathematical algorithms are needed to transform the acquired signals into meaningful outcomes. Systematic research on MBW analysis is scarce and it remains unclear what would constitute the most sensible MBW analysis process. Software algorithms and LCI values differ largely between similar setups [1, 7, 8]. This is, partially, the reason for between-centre variability in LCI values, which limits their use in multicentre comparisons and underlines the usefulness of current reference values [1].

The current consensus statement provides several analysis recommendations [9], which are based on low-grade evidence and often lack systematic validation. LCI is the number of functional residual capacity (FRC) lung turnovers needed to reduce alveolar tracer-gas concentration to the cut-off of 2.5% (1/40th) of its starting concentration, calculated as the ratio of cumulative expiratory volume (CEV) to FRC (CEV/FRC). However, it is recommended to determine it in the first of three consecutive breaths below 2.5%. Thus, unlike the definition, by nature LCI is not calculated directly at 2.5% but at varying concentrations below this cut-off, depending on chance, breathing pattern and other factors. The effect of this inaccuracy on outcome measures is unclear [9].

Another challenging issue related to the MBW test is the acquisition of the tracer gas concentration per breath. Most of the studies across the paediatric population use the end-tidal tracer gas concentration per breath [9]. This may be susceptible to low signal-to-noise ratio at the end of the test, introducing an error of unknown size. Alternatively, gas concentrations can be estimated from various portions of each washout breath. Several MBW systems use the mean concentration per breath instead, either as a default or as an optional setting [9–11], and one study has shown significant differences in LCI calculated using the mean or the end-tidal nitrogen concentration [11].

In this study we hypothesised that 1) algorithms that can calculate with higher precision the LCI at the cut-off of 2.5%, and 2) algorithms that estimate the tracer gas concentration in different parts of the breath will increase the robustness of the analysis, by reducing the intra-subject variability in LCI values. Therefore, we used 1) three alternative algorithms to detect with higher precision the end of MBW testing and 2) two alternate algorithms to measure exhaled tracer gas concentrations, and compared the MBW results with those derived from the recommended analysis method [9]. Primary outcomes were changes in LCI values, intra-test repeatability, and ability to discriminate between healthy children and children with CF or PCD lung disease. Secondary outcomes were changes in lung volumes determining LCI, *i.e.* CEV and FRC.

Methods

Study design

This is a retrospective analysis of prospectively collected data. Nitrogen (N_2)MBW measurements were obtained in school-aged healthy children and children with CF or PCD. All participants were free from acute respiratory disease for at least 2 weeks prior to testing. For healthy controls, additional exclusion criteria were asthma or other respiratory disease, history of prematurity, and bone, neuromuscular or cardiac disease that could affect lung function. Measurements were performed in the University Hospital of Bern, Bern, Switzerland for healthy controls and children with CF, and in the University Children's Hospital of Ruhr, Bochum, Germany for children with PCD. The study was approved by the Ethics Committees of the Canton of Bern, Switzerland and of the Ruhr University of Bochum, Germany. We obtained written informed consent from parents or participants older than 18 years. Some data from this cohort have been recently published [12, 13].

N_2 MBW measurements

Each child performed 3–4 N_2 MBW according to the current consensus statement [9] using the ultrasonic flowmeter (Exhalyzer D; Eco Medics AG, Duernten, Switzerland) and the corresponding software (Spiroware 3.1.6; Eco Medics AG). During the test children were sitting upright, wearing a nose clip and breathing tidally through a snorkel mouthpiece, as previously described [14].

Standard analysis of the data

We analysed the data with custom-made software (LungSim, Version 4.8.5; NM GmbH, Thalwil, Switzerland, which is based on Matlab (The Mathworks Inc., Natick, MA, USA)) [15] using raw N_2 MBW

signals (A-files, Spiroware 3.1.6). Calibration, body temperature and pressure saturated correction, and signal synchronisation were performed automatically. Re-inspired nitrogen was always subtracted to obtain net nitrogen volume using the post-gas sampling point method. The main output parameters were LCI, FRC and CEV. The software calculated LCI_{standard} according to current recommendations, *i.e.* end-tidal nitrogen concentration (C_{et}) defined as the average value between 95% and 98% of expired volume and LCI as the ratio of CEV to FRC (CEV/FRC) at the first of three consecutive breaths below the cut-off of 2.5% (1/40th) [9] (figure 1a).

Alternative analysis methods

We used LungSim to apply novel and currently used methods.

Linear interpolation analysis

This new algorithm was used with the aim to calculate precisely the LCI at 2.5% with a focus on the end of the test. In order to calculate MBW outcomes directly at the 2.5% cut-off, we interpolated linearly the C_{et} of the breath where LCI_{standard} was calculated, and the C_{et} of the previous breath (figure 1b). CEV_{linear} , FRC_{linear} and LCI_{linear} were then calculated at the 2.5% cut-off (a detailed description of the method can be found in the online supplementary material).

Fitting-curve analysis

This new algorithm was used with the aim to calculate precisely the LCI at 2.5% taking the whole washout curve into account. We used the C_{et} per breath to fit a curve using a least squares regression model (figure 1c). $CEV_{\text{fit-curve}}$, $FRC_{\text{fit-curve}}$ and $LCI_{\text{fit-curve}}$ were then calculated at the time point that the fitting curve crossed the 2.5% cut-off (see the online supplementary material).

Analysis at the first breath below the 2.5% cut-off

This new algorithm was used mostly with the aim of challenging the standard method. MBW outcomes derived from the first breath below the 2.5% cut-off, independent of whether this was followed by two consecutive breaths below this cut-off. The rest of the analysis was performed as in the standard analysis. In order to prevent false LCI calculation based on small superficial breaths prior to washout completion, we performed a visual and numerical breath quality control based on inspiratory and expiratory volumes.

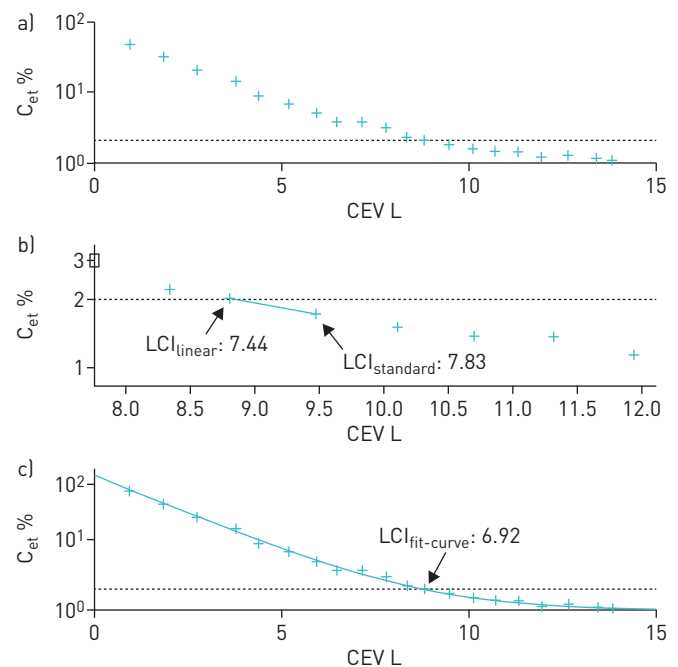


FIGURE 1 a) Overview of a nitrogen multiple-breath washout measurement in LungSim (NM GmbH); the crosses represent the end-tidal nitrogen concentration per breath, the line indicates 2.5% of the initial nitrogen concentration. The y-axis represents the nitrogen concentration (%) on a logarithmic scale (C_{et}), the x-axis represents the cumulative expiratory volume (CEV). b) Enlargement of the end of the washout curve: lung clearance index (LCI) is calculated at the breath after 2.5% (LCI_{standard}), or directly at 2.5% using the linear interpolation method (LCI_{linear}). c) LCI is calculated directly at $C_{\text{et}}=2.5\%$ using the fitting-curve method ($LCI_{\text{fit-curve}}$).

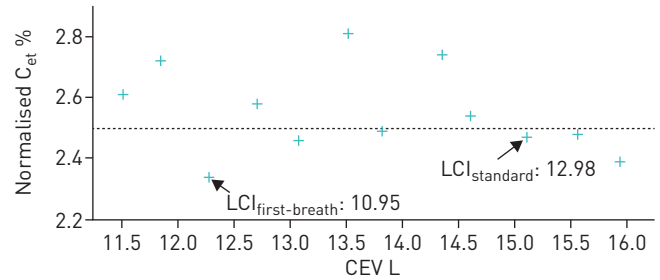


FIGURE 2 Nitrogen concentration curve at the end of the washout from a nitrogen multiple-breath washout measurement in a patient with primary ciliary dyskinesia with fluctuation around 2.5% analysed in LungSim (NM GmbH). The crosses represent the end-tidal nitrogen concentration per breath, the line indicates 2.5% of the initial nitrogen concentration. The y-axis represents the normalised nitrogen concentration [C_{et}] and the x-axis the cumulative expiratory volume (CEV). Lung clearance index (LCI) from the first of three consecutive breaths below the 2.5% cut-off [$LCI_{standard}$] and from the first breath below 2.5% [$LCI_{first-breath}$] are highlighted. $LCI_{first-breath}$ is calculated seven breaths prior to $LCI_{standard}$.

The system did not allow LCI calculations in breaths with inspiratory and/or expiratory volume less than half of the mean tidal volume.

In theory, the C_{et} for every washout breath is progressively lower. In practice, we often see that the C_{et} in the last washout breaths oscillates with values higher or lower than the cut-off of 2.5% (figure 2). We named this phenomenon a “tracer-gas fluctuation”. Based on the change in the end of the washout between this method and the standard method, we were able to define numerically tracer-gas fluctuations around the cut-off of 2.5%. Thus, we compared the washout breath number where LCI was measured (BrNr) between the standard analysis (BrNr_{standard}) and the analysis using the first breath below the cut-off (BrNr_{first-breath}) ($\Delta BrNr = BrNr_{standard} - BrNr_{first-breath}$). $\Delta BrNr \geq 2$ was a sign of fluctuation, and this was verified by visual control of the nitrogen concentration curve (figure 2).

Analysis with alternative methods to detect expiratory tracer gas concentration

Mean expiratory nitrogen concentration

For each breath, instead of C_{et} , the mean nitrogen concentration (C_{mean}) across 65–95% of the expired volume was used, while the rest of the analysis was performed as in the standard analysis. This algorithm is currently in use for certain setups (figure E3) [10, 11].

Median expiratory nitrogen concentration

For each breath, the median nitrogen concentration (C_{median}) of 65–95% of the expired volume was used (figure E3), while the rest of the analysis was performed as in the standard analysis. This algorithm can be optionally used in a commercially available setup [10].

Statistics

Data were analysed using GraphPad Prism version 5 for Windows (GraphPad Software, San Diego, CA, USA) and Stata (Stata Statistical Software: Release 13; StataCorp LP, College Station, TX, USA). Sample size was estimated based on previous controlled trials using LCI as the primary outcome, and considering the difference of one turnover as clinically significant [16, 17]. We used the paired t-test for comparisons of MBW outcomes of the same measurements analysed with different algorithms, and the unpaired t-test and one-way ANOVA with Tukey’s multiple comparison for the comparison of MBW outcomes between subjects. Relative changes from the standard analysis were visualised using the Bland–Altman method [18]. The intra-subject LCI variability was defined with the coefficient of variation (CV (%))= $SD/mean$). We estimated as clinically significant the difference in variability that is twice the SD of variability of the standard analysis per group. Receiver operating characteristics (ROC) analysis was used to estimate the ability of LCI calculated with the different methods to discriminate between health and lung disease (children with CF or PCD). Areas under the ROC curve were compared using the Chi-squared test. A linear regression analysis was used for associations between $LCI_{standard}$ values and 1) the end of the washout, as potentially dependent on C_{mean} and C_{median} ; and 2) fluctuations of nitrogen concentration at the end of the washout. A p-value<0.05 was accepted to indicate statistical significance.

Results

N_2 MBW measurements were performed in 20 healthy children, mean age 13.3 years (range: 7.6–15.9 years), 20 children with CF, mean age 9.9 years (4.6–16.6 years), and 17 children with PCD, mean

TABLE 1 Demographics of the study subjects

	Healthy	CF	PCD
Subjects n	20	20	17
Males	9	8	9
Age years	13.3±1.89	9.9±3.61	11.8±3.76
Weight kg	48.1±12.51	31.7±12.89	45.5±15.1
Weight z-score[#]	0.05±0.90	-0.33±0.79	0.62±1.22
Height cm	157.9±11.76	134.4±19.35	147.2±19.37
Height z-score[#]	0.19±1.06	-0.33±0.95	0.12±1.15
LCI_{standard}	7.10±0.56	10.13±2.00 ^{¶,+}	10.12±2.21 ^{¶,+}

Data are presented as mean±SD, unless otherwise stated. CF: cystic fibrosis; PCD: primary ciliary dyskinesia; LCI_{standard}: lung clearance index calculated according to the current recommendations [9].
[#]: z-scores calculated according to Centers for Disease Control and Prevention growth charts [30];
[¶]: p<0.001, unpaired t-test, compared with healthy subjects; ⁺: p=0.3, unpaired t-test between CF and PCD groups.

age 11.8 years (5.1–18.1 years) (table 1). As expected, LCI in children with CF and PCD was significantly higher compared with healthy children ($p<0.001$). LCI between children with CF and children with PCD did not differ significantly.

Linear interpolation analysis

LCI_{linear} values were systematically lower than LCI_{standard} in all groups (figure 3a, table 2). Relative mean±SD difference from LCI_{standard} was $-1.7\pm 1.3\%$, with no statistically significant differences between groups (table 2). Intra-subject variability (CV %) in LCI was comparable between both methods (table 3). The effect on LCI was mainly due to decreased CEV_{linear} values in all groups, while FRC was minimally affected (maximum difference, $FRC_{standard}-FRC_{linear}=0.03$ L) (table E1). The ability of this method to discriminate between health and disease did not differ significantly from the standard analysis (figure 4a).

Fitting-curve analysis

Using the nonlinear curve-fitting method, the washout curve did not reach the standard 2.5% cut-off in six (10%) out of 60 measurements of healthy children, and 11 (22%) out of 51 measurements of children with PCD. LCI_{fit-curve} values were in a nonsystematic way lower than LCI_{standard} in all groups (figure 3b and table 2). Relative mean±SD difference from LCI_{standard} was $-11.1\pm 15.7\%$ and varied significantly between groups (table 2). The intra-subject variability in LCI remained unchanged in healthy and CF, but was higher in PCD compared with the standard analysis (table 3). Using this method, the discrimination between health and disease was poorer in comparison with the standard method (figure 4a).

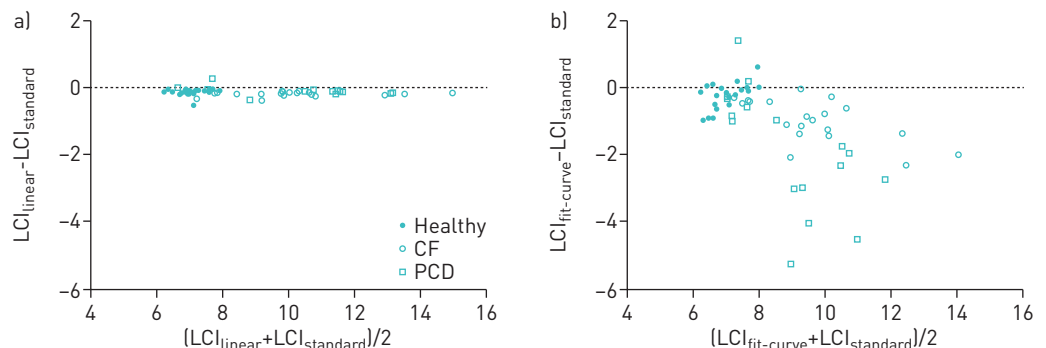


FIGURE 3 The linear interpolation analysis and the fitting-curve analysis give lower values for lung clearance index (LCI). a) Bland-Altman plots of the absolute difference between LCI_{linear} and LCI_{standard} plotted versus the mean of LCI_{linear} and LCI_{standard} for 20 healthy children, 20 children with cystic fibrosis (CF) and 17 children with primary ciliary dyskinesia (PCD) (one-way ANOVA, $p=0.07$). b) Bland-Altman plots of relative difference between LCI_{fit-curve} and LCI_{standard} plotted versus mean of LCI_{fit-curve} and LCI_{standard} for 20 healthy children, 20 children with CF and 17 children with PCD (one-way ANOVA, $p<0.001$).

TABLE 2 Lung clearance index values at 2.5% cut-off ($LCI_{2.5}$) in healthy controls, children with CF and children with PCD, calculated with the different analysis methods in LungSim, and relative difference from the LCI calculated with the standard analysis (ΔLCI)

Analysis	Healthy (n=20)		CF (n=20)		PCD (n=17)	
	Mean \pm SD $LCI_{2.5}$	ΔLCI (95% CI) %	Mean \pm SD $LCI_{2.5}$	ΔLCI (95% CI) %	Mean \pm SD $LCI_{2.5}$	ΔLCI (95% CI) %
Standard	7.11 \pm 0.45		10.13 \pm 1.98		10.12 \pm 2.21	
Linear	6.98 \pm 0.46***	-2.0 [-2.7- -1.3]	9.95 \pm 1.98***	-2.0 [-2.4- -1.5]	9.79 \pm 2.14***	-1.1 [-1.9- -0.3]
Fit-curve	6.88 \pm 0.68***	-3.6 [-6.4- 0.8]	9.16 \pm 1.56***	-9.7 [-12.4- -7.1]	8.04 \pm 1.22***	-19.7 [-29.8- -9.6]
1st breath	7.06 \pm 0.57	-0.6 [-1.2- 0.0]	9.99 \pm 1.90***	-1.4 [-2.3-0.4]	9.55 \pm 1.92***	-5.1 [-6.0- -2.0]
C_{mean}	6.95 \pm 0.47***	-2.5 [-3.6- -1.3]	9.56 \pm 1.60***	-5.4 [-7.2- -3.5]	9.39 \pm 1.82***	-5.1 [-7.4- -2.8]
C_{median}	6.94 \pm 0.47***	-2.5 [-3.6- -1.5]	9.61 \pm 1.60***	-4.9 [-6.7- -3.1]	9.49 \pm 2.68***	-5.6 [-7.5- -2.7]

CF: cystic fibrosis; PCD: primary ciliary dyskinesia; linear: linear interpolation method; fit-curve: fitting-curve method; 1st breath: LCI calculated at the first breath below a cut-off; C_{mean}: mean expiratory nitrogen concentration; C_{median}: median expiratory nitrogen concentration. ***: p<0.001, paired t-test compared with standard analysis.

CEV_{fit-curve}, but also FRC_{fit-curve}, values were lower than standard in all groups (table E1). In the CF and PCD groups, mean $LCI_{fit-curve}$ was also significantly lower than mean LCI_{linear} (p<0.001 for CF, p<0.0001 for PCD).

LCI at the first breath below the 2.5% cut-off

LCI from the first breath below 2.5% was lower in all groups compared with $LCI_{standard}$ (table 2 and figure E4). Relative mean \pm SD difference from $LCI_{standard}$ was $-1.8\pm 4.3\%$ and varied significantly between groups (table 2). The intra-subject variability in LCI (table 3) and the ability to discriminate between health and disease (figure 4a) were similar to the standard analysis.

At least one N₂MBW measurement with fluctuations around the cut-off of 2.5% was observed in seven (35%) out of 20 healthy children, nine (45%) out of 20 children with CF and 12 (71%) out of 17 children with PCD (figure E5). A lower number of measurements with fluctuations were observed around the cut-off of 5% (figure E6). The presence of fluctuations was further associated with higher LCI values in both CF (coefficient: 0.47, R²: 0.14, p=0.003, CI: 0.16–0.77) and PCD (coefficient: 0.29, R²: 0.31, p<0.0001, CI: 0.16–0.42) patients, but not in healthy children (coefficient: 0.15, R²: 0.05, p=0.1, CI: -0.03–0.33) (figure E5).

Effect of mean/median nitrogen expiratory concentration on LCI values

LCI_{meanN_2} as well as $LCI_{medianN_2}$ values were significantly lower compared with standard values in all groups, but the LCI intra-subject variability was not changed (figure 5, tables 2 and 3). The relative mean \pm SD difference from $LCI_{standard}$ was $-4.2\pm 5.3\%$ for LCI_{meanN_2} and $-4.1\pm 5.2\%$ for $LCI_{medianN_2}$, and was significantly lower in healthy subjects compared with disease groups in both analyses (table 2).

TABLE 3 Coefficient of intra-subject variation in lung clearance index at 2.5% cut-off (CV_{LCI}) in healthy children, children with CF and children with PCD, analysed with different methods

Analysis	CV_{LCI} %		
	Healthy (n=20)	CF (n=20)	PCD (n=17)
Standard	4.7 \pm 3.5	5.7 \pm 2.5	7.1 \pm 3.8
Linear	4.3 \pm 3.3	5.4 \pm 2.8	7.2 \pm 4.0
Fit-curve	5.1 \pm 3.2	6.0 \pm 5.3	17.6 \pm 11.0***
1st breath	4.6 \pm 3.1	5.1 \pm 2.5	5.3 \pm 3.2
C_{mean}	5.4 \pm 3.7	5.3 \pm 2.9	5.9 \pm 3.7
C_{median}	5.2 \pm 3.5	5.3 \pm 3.0	5.9 \pm 3.6

Data are presented as mean \pm SD. CF: cystic fibrosis; PCD: primary ciliary dyskinesia; linear: linear interpolation method; fit-curve: fitting-curve method; 1st breath: LCI calculated at the first breath below a cut-off; C_{mean}: mean expiratory nitrogen concentration; C_{median}: median expiratory nitrogen concentration. ***: p<0.001, paired t-test compared with standard analysis.

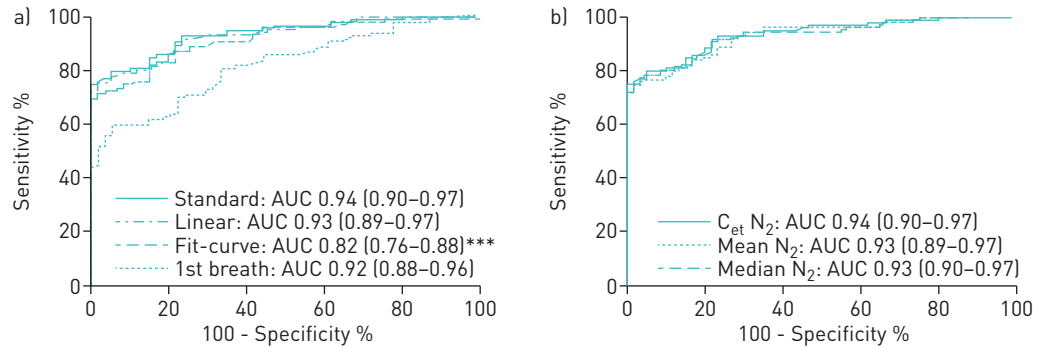


FIGURE 4 Discriminatory performance (receiver operating characteristics (ROC) analysis) of lung clearance index (LCI) analysed with the different methods. a) ROC curves of LCI calculated using the standard analysis (standard), the linear interpolation (linear), the fit-curve method (fit-curve) and the method that uses the first breath below the 2.5% cut-off (1st breath). b) ROC curves of LCI calculated using the end-tidal nitrogen concentration (standard analysis, C_{et} N₂), the mean nitrogen concentration (mean N₂) and the median nitrogen concentration (median N₂) per breath. AUC: area under the curve. ***: p<0.001, chi-squared test, compared with the standard analysis.

Both LCI_{meanN2} and LCI_{medianN2} values were calculated earlier in the washout curve, compared with LCI_{standard} (table E2). Interestingly, the higher the degree of ventilation inhomogeneity (LCI values), the earlier the breath number in the washout curve where the LCI_{meanN2} is calculated, compared with LCI_{standard} (coefficient: 0.66, R²: 0.47, p<0.001, CI: 0.57–0.74). We observed similar findings in LCI_{medianN2}. Minimal but statistically significant differences were found between LCI_{meanN2} and LCI_{medianN2} in the healthy (p=0.047) and CF (p=0.038) groups, but not in the PCD group (p=0.6), although the breath number for the LCI calculation remained the same (table E2).

Discussion

In this proof-of-principle study we reported alternative methods for MBW analysis, and their influence on MBW outcomes. We used different methods aiming to define the end of the washout with high precision, which gave LCI values lower than the standard analysis, without improving the intra-subject variability in LCI. Alternative ways to detect expiratory tracer gas concentration had a great influence on LCI values. The ability to discriminate between health and disease was excellent for all methods and only declined using the nonlinear curve-fitting method.

To our knowledge, this is the first study that reports mathematical ways to analyse LCI directly at the cut-off. Both fitting methods influence mainly the end of the washout curve. However, they have a principal difference. The fit-curve method is based on breath-by-breath nitrogen concentration of the whole washout, and thus depends on the breathing pattern across the whole test. The linear interpolation method uses only two values around the cut-off at the end of the test. This explains why LCI_{linear} values

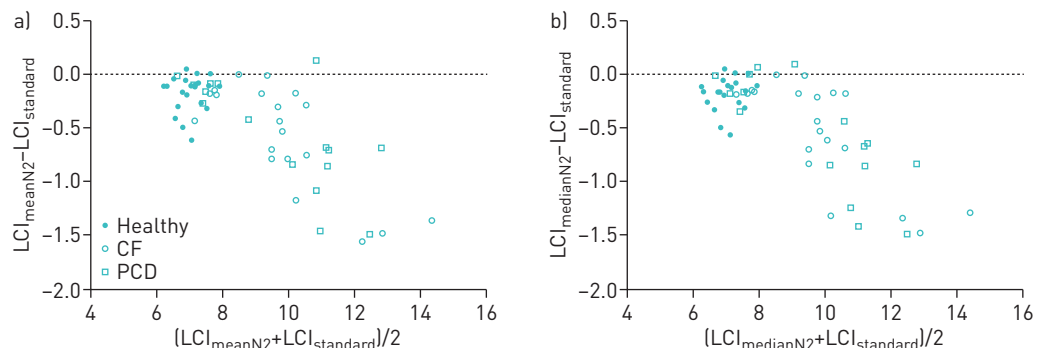


FIGURE 5 Lung clearance index (LCI) values based on the end-tidal nitrogen washout curve (LCI_{standard}) are higher compared with the mean or median expiratory nitrogen curve (LCI_{meanN2} and LCI_{medianN2}, respectively). Bland-Altman plots of relative difference a) between LCI_{meanN2} and LCI_{standard} plotted versus the mean of LCI_{meanN2} and LCI_{standard}; and b) between LCI_{medianN2} and LCI_{standard} plotted versus the mean of LCI_{medianN2} and LCI_{standard} for 20 healthy children, 20 children with cystic fibrosis (CF) and 17 children with primary ciliary dyskinesia (PCD) (one-way ANOVA, p<0.001).

were slightly, but systematically lower than LCI_{standard} , while $LCI_{\text{fit-curve}}$ values were more variable. The differences found in LCI values from the different analysis methods were mainly due to differences in CEV. FRC was less affected by the different algorithms.

Surprisingly none of the fitting methods resulted in lower intra-test variability of LCI. We speculate that breathing pattern and sensors, as well as the time interval between the washouts [19, 20], have a greater influence on intra-test variability. We only assessed indices based on “areas under the curve”, such as LCI. These indices are known to be inherently less variable compared with phase III slope indices.

Our findings support the present consensus recommendation for three consecutive breaths below the 2.5% cut-off. Intra-test variability and detection of lung disease were excellent. Consideration of three breaths at the end of the washout prevents false early test termination. The latter may occur due to nitrogen fluctuations around the cut-off even in healthy children. Fluctuations may relate to several factors, for example lower signal-to-noise ratio, breathing pattern, opening of slowly ventilated compartments and back-diffusion of tissue-nitrogen [8, 9, 21]. We assume that variable ventilation of lung compartments was important, as the fluctuations were greater in CF and PCD. Increased back-diffusion or breathing variability in diseased children appears counterintuitive. Serial opening of slowly ventilated lung compartments seems to be a distinct feature in patients with abnormal ventilation distribution efficiency [8, 9].

Differences in LCI between end-expiratory and mean or median expiratory nitrogen concentrations were small and in accordance with previous reports [11, 22]. As expected, the impact was disease-dependent. Because phase III slopes were steeper (from visual inspection) in measurements from CF and PCD patients, averaging expiratory nitrogen across the phase III underestimated later end-tidal nitrogen [10]. Yet intra-test variability and differentiation between health and disease were not impaired, suggesting that averaging nitrogen across larger portions of washout breaths is appropriate.

In our study, the influence of alternative algorithms was more prominent in LCI values from patients with CF or PCD. This is not surprising, as the uneven ventilation and the late opening of slowly ventilated areas mostly affect the end of the test. Moreover, the part of expiration that the tracer gas concentration is determined from is more critical in patients with CF or PCD, as mentioned earlier. Therefore, in accordance to similar studies [23, 24], our analysis suggests that it is essential to include measurements from subjects with lung disease in software/algorithm validation studies, as those measurements are more sensitive to small changes in the analysis process.

The alternative algorithms proposed here can be applied to any MBW setup, as they are independent of the tracer gas and hardware/software used. It will be interesting to investigate the influence of those algorithms using foreign tracer gases, considering the differences in washout behaviour between nitrogen and SF_6 [8, 21].

The strengths of our study relate to a sufficiently large sample size which, for example, previously allowed the detection of treatment effects in CF [16] or physiological phenotypes [25]. MBW testing was carried out according to the current guidelines in the same device and recording software.

The intra-test variability in LCI values was chosen to measure the robustness of different algorithms, as previously described [24, 26]. Despite the high quality of the recordings, intra-test variability was widely scattered, which possibly facilitates comparison to clinical testing situations. However, according to a *post hoc* power analysis, the study was underpowered to detect significant changes in intra-subject variability, as this would require a much larger sample ($n=730$ with 90% power at the 0.05 level) that is non-realistic in a clinical setting. Moreover, all tests were performed at a single time-point, so we were not able to assess the intra-subject between-test repeatability [27].

The healthy subjects in our study were slightly older than our patients. Yet intra-test variability did not differ and ventilation inhomogeneity indices marginally relate to body size in the age range assessed [28]. Effects of the different analysis methods could therefore be easily assessed. We avoided mathematical extrapolation and considered both the full washout and its end. For comparison we used clinically relevant outcomes such as ROC curves to discriminate between health and disease. However, we did not assess the sensitivity to capture lung function response to interventions or lung function dynamics over time. We also did not assess the relationship with structural changes of the airways which have been described previously [29].

This methodological report has several clinical implications. We were able to show that MBW outcomes are prone to the analysis method used, and thus it is of high priority to stick to the analysis outlined in the current recommendations [9]. Alternative analysis methods should be clearly stated by the manufacturers and the users. The differences in the outcomes provided here do not allow any direct comparison of results analysed with different methods, and limit the use of normative values not only to the same setup but additionally to the same analysis algorithms.

Overall, the data support current recommendations to measure LCI in children. Standard LCI is characterised by low intra-test variability and good discrimination between children with CF or PCD and controls. However, we show that the use of different analysis algorithms may considerably influence LCI. Reference equations should be based on appropriate normative data obtained by the same hardware and software with appropriate settings. Transparent reporting of algorithms is necessary in both research and clinical applications.

Conflict of interest: None declared.

References

- 1 Subbarao P, Milla C, Aurora P, *et al.* Multiple-breath washout as a lung function test in cystic fibrosis. A Cystic Fibrosis Foundation Workshop Report. *Ann Am Thorac Soc* 2015; 12: 932–939.
- 2 Horsley A. Lung clearance index in the assessment of airways disease. *Respir Med* 2009; 103: 793–799.
- 3 Aurora P, Bush A, Gustafsson P, *et al.* Multiple-breath washout as a marker of lung disease in preschool children with cystic fibrosis. *Am J Respir Crit Care Med* 2005; 171: 249–256.
- 4 Gustafsson PM, Aurora P, Lindblad A. Evaluation of ventilation maldistribution as an early indicator of lung disease in children with cystic fibrosis. *Eur Respir J* 2003; 22: 972–979.
- 5 Boon M, Vermeulen FL, Gysemans W, *et al.* Lung structure-function correlation in patients with primary ciliary dyskinesia. *Thorax* 2015; 70: 339–345.
- 6 Nyilas S, Schlegtendal A, Yammine S, *et al.* Further evidence for an association between LCI and FEV₁ in patients with PCD. *Thorax* 2015; 70: 896.
- 7 Poncin W, Singer F, Aubriot AS, *et al.* Agreement between multiple-breath nitrogen washout systems in children and adults. *J Cyst Fibros* 2017; 16: 258–266.
- 8 Jensen R, Stanojevic S, Gibney K, *et al.* Multiple breath nitrogen washout: a feasible alternative to mass spectrometry. *PLoS One* 2013; 8: e56868.
- 9 Robinson PD, Latzin P, Verbanck S, *et al.* Consensus statement for inert gas washout measurement using multiple- and single- breath tests. *Eur Respir J* 2013; 41: 507–522.
- 10 Anagnostopoulou P, Egger B, Lura M, *et al.* Multiple breath washout analysis in infants: quality assessment and recommendations for improvement. *Physiol Meas* 2016; 37: L1–L15.
- 11 Verbanck S, Paiva M, Schuermans D, *et al.* Relationships between the lung clearance index and conductive and acinar ventilation heterogeneity. *J Appl Physiol (1985)* 2012; 112: 782–790.
- 12 Yammine S, Nyilas S, Casaulta C, *et al.* Function and ventilation of large and small airways in children and adolescents with inflammatory bowel disease. *Inflamm Bowel Dis* 2016; 22: 1915–1922.
- 13 Nyilas S, Schlegtendal A, Singer F, *et al.* Alternative inert gas washout outcomes in patients with primary ciliary dyskinesia. *Eur Respir J* 2017; 49: 1600466.
- 14 Singer F, Kieninger E, Abbas C, *et al.* Practicability of nitrogen multiple-breath washout measurements in a pediatric cystic fibrosis outpatient setting. *Pediatr Pulmonol* 2013; 48: 739–746.
- 15 Yammine S, Singer F, Abbas C, *et al.* Multiple-breath washout measurements can be significantly shortened in children. *Thorax* 2013; 68: 586–587.
- 16 Amin R, Subbarao P, Jabar A, *et al.* Hypertonic saline improves the LCI in paediatric patients with CF with normal lung function. *Thorax* 2010; 65: 379–383.
- 17 Oude Engberink E, Ratjen F, Davis SD, *et al.* Inter-test reproducibility of the lung clearance index measured by multiple breath washout. *Eur Respir J* 2017; 50: 1700433.
- 18 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.
- 19 Salamon ER, Gain KR, Hall GL. Defining the appropriate waiting time between multiple-breath nitrogen washout measurements. *Eur Respir J* 2015; 45: 1489–1491.
- 20 Jensen R, Stanojevic S, Klingel M, *et al.* A systematic approach to multiple breath nitrogen washout test quality. *PLoS One* 2016; 11: e0157523.
- 21 Yammine S, Lenherr N, Nyilas S, *et al.* Using the same cut-off for sulfur hexafluoride and nitrogen multiple-breath washout may not be appropriate. *J Appl Physiol (1985)* 2015; 119: 1510–1512.
- 22 Luijendijk SC, van der Grinten CP. The ratio of the alveolar ventilations of SF₆ and He in patients with lung emphysema and in healthy subjects. *Respir Physiol Neurobiol* 2002; 130: 69–77.
- 23 Anagnostopoulou P, Yammine S, Schmidt A, *et al.* False normal lung clearance index in infants with cystic fibrosis due to software algorithms. *Pediatr Pulmonol* 2015; 50: 970–977.
- 24 Gronbaek J, Hallas HW, Arianto L, *et al.* New time-saving predictor algorithm for multiple breath washout in adolescents. *Pediatr Res* 2016; 80: 49–53.
- 25 Nyilas S, Singer F, Kumar N, *et al.* Physiological phenotyping of pediatric chronic obstructive airway diseases. *J Appl Physiol (1985)* 2016; 121: 324–332.
- 26 Latzin P, Sauteur L, Thamrin C, *et al.* Optimized temperature and deadspace correction improve analysis of multiple breath washout measurements by ultrasonic flowmeter in infants. *Pediatr Pulmonol* 2007; 42: 888–897.
- 27 Stanojevic S, Davis SD, Retsch-Bogart G, *et al.* Progression of lung disease in preschool patients with cystic fibrosis. *Am J Respir Crit Care Med* 2017; 195: 1216–1225.
- 28 Lum S, Stocks J, Stanojevic S, *et al.* Age and height dependence of lung clearance index and functional residual capacity. *Eur Respir J* 2013; 41: 1371–1377.
- 29 Ramsey KA, Rosenow T, Turkovic L, *et al.* Lung clearance index and structural lung disease on computed tomography in early cystic fibrosis. *Am J Respir Crit Care Med* 2016; 193: 60–67.
- 30 Centers for Disease Control and Prevention, National Center for Health Statistics. CDC growth charts: United States. www.cdc.gov/growthcharts/ Date last updated: September 9, 2010. Date last accessed: March 30, 2016.