Inter-test reproducibility of the lung clearance index measured by multiple breath washout

Esther Oude Engberink[1,2], Felix Ratjen[1,3], Stephanie D. Davis[4], George Retsch-Bogart[5], Reshma Amin[1,3] and Sanja Stanojevic[1,3]

Affiliations: [1]Division of Respiratory Medicine, Dept of Pediatrics, Hospital for Sick Children, Toronto, ON, Canada. [2]Division of Respiratory Medicine and Allergy, Dept of Pediatrics, Radboud University Medical Center Amalia Children's Hospital, Nijmegen, The Netherlands. [3]Dept of Medicine, University of Toronto, Toronto, ON, Canada. [4]Section of Pediatric Pulmonology, Allergy and Sleep Medicine, Dept of Pediatrics, Riley Hospital for Children, Indiana University School of Medicine, Indianapolis, IN, USA. [5]Division of Pediatric Pulmonology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

Correspondence: Sanja Stanojevic, Division of Respiratory Medicine, Sick Kids, 555 University Avenue, Toronto, ON, Canada, M5G 1X8. E-mail: Sanja.Stanojevic@SickKids.ca

ABSTRACT The lung clearance index (LCI) has strong intra-test repeatability; however, the inter-test reproducibility of the LCI is poorly defined.

The aim of the present study was to define a physiologically meaningful change in LCI in preschool children, which discriminates changes associated with disease progression from biological variability.

Repeated LCI measurements from a longitudinal cohort study of children with cystic fibrosis and age-matched controls were collected to define the inter-visit reproducibility of the LCI. Absolute change, the coefficient of variation, Bland–Altman limits of agreement, the coefficient of repeatability, intra-class correlation coefficient, and percentage changes were calculated.

LCI measurements (n=505) from 71 healthy and 77 cystic fibrosis participants (aged 2.6–6 years) were analysed. LCI variability was proportional to its magnitude, such that reproducibility defined by absolute changes is biased. A physiologically relevant change for quarterly LCI measurements in health was defined as exceeding ±15%. In clinically stable cystic fibrosis participants, the threshold was higher (±25%); however, for measurements made 24 h apart, the threshold was similar to that observed in health (±17%).

A percentage change in LCI greater than ±15% in preschool children can be considered physiologically relevant and greater than the biological variability of the test.

Introduction

The lung clearance index (LCI), measured by multiple breath washout (MBW), may be a useful physiological test to clinically monitor lung function in patients with cystic fibrosis. The LCI identifies early obstructive lung disease [1] and is feasible to perform in young children [2, 3]. In addition, the LCI has good intra-test repeatability [4–6]; however, there are limited longitudinal data describing inter-test reproducibility.

Investigators report differing findings for the inter-test reproducibility of the LCI due to variability in protocol designs, study population and MBW methodology [5–14]. Most notably, time intervals between repeated tests range from hours to years, and none report reproducibility for time intervals typically used to clinically track patients with cystic fibrosis (*e.g.* quarterly) [6, 11, 15]. Comparison between studies is further limited by the various methodologies that define reproducibility; coefficient of variation (%CV) [5, 7, 8], the intra-class correlation coefficient (ICC) [6], the limits of agreement from Bland–Altman plots [5, 9–11] or the coefficient of reproducibility [12, 13]. Reproducibility has most commonly been reported as Bland–Altman limits of agreement or coefficient of repeatability. Both of these methods assume that the measurement variability is independent of its magnitude, and analyses have only been tested in health or for a narrow range of LCI values in participants with cystic fibrosis [12, 15]. If this assumption is not valid across the spectrum of values observed in clinical practice, the use of the limits of agreement or coefficient of repeatability may introduce bias in the interpretation of clinically relevant between-visit changes in the LCI.

An additional challenge of defining outcome reproducibility is only using data from healthy individuals. This approach accounts for biological variability of the test in a healthy population, but may be too sensitive in disease as it assumes that any variability observed outside this range is driven by disease progression. In constrast, defining stability and assuring that observed changes between measurements in disease are not attributed to natural progression proves difficult; thereby, biological variability based on disease groups may be too conservative.

In this study, we aim to use longitudinal LCI data collected at multiple time intervals (*e.g.* 1 month, 3 months, 6 month, 9 months and 12 months) to define inter-visit reproducibility in health and stable cystic fibrosis. Furthermore, we aim to define reproducibility for repeated measurements made within 24 h to elucidate if variability in LCI is driven by the biological variability of the test, as changes observed within this time period are unlikely to be associated with changes in the disease state.

Methods
Study population
Data were collected as part of a prospective multi-centre observational study of preschool children (aged 2.5–6 years) with cystic fibrosis and age-matched healthy controls from three North American cystic fibrosis centres between January 2013 and June 2015 [3]. MBW was measured at enrolment and 1, 3, 6, 9 and 12 months. All participants, both healthy and with cystic fibrosis, were free of respiratory infection for at least 4 weeks prior to the enrolment. Symptoms (cough) and treatments (oral or intravenous antibiotics) were recorded for all subsequent tests. Any measurements for which participants had either cough or pulmonary exacerbation (defined as respiratory symptoms and treatment with antibiotics) were excluded from these analyses. Thus, these data should reflect clinically stable visits. The study was approved by the Research Ethics Board at the Hospital for Sick Children (REB # 1000036303), Riley Children's Hospital (1401277863), and the Office of Human Research Ethics at University of North Carolina at Chapel Hill (13-1258). Informed written consent was obtained from the parent/guardian for all participants.

Reproducibility was also defined from the placebo arm of a randomised cross-over interventional trial (NCT 02276898) that included participants with cystic fibrosis aged ⩾7 years randomised to a single dose of either hypertonic saline or isotonic saline [14]. MBW was performed at baseline and 24 h later with at least 1 week washout between treatment arms [14]. The study was approved by the Research Ethics Board at the Hospital for Sick Children (REB # 1000024909) and informed written consent was obtained from the parent/guardian for all participants.

MBW outcomes
MBW tests were performed with the Exhalyzer D (EcoMedics AG, Duernten, Switzerland). Adaptions were made for testing preschool children [3]. Participants were enrolled if they could complete a single MBW trial, whereas a MBW test was considered successful if there were at least two technically acceptable trials. The LCI was calculated as the cumulative expired volume (CEV) divided by the functional residual capacity (FRC) at 1/40 of the starting gas concentration. LCI5 was calculated at 1/20 of the starting gas concentration. All MBW traces were reviewed for technical quality and appropriate breathing pattern [16]. Full details are previously published [3].

Spirometry

Spirometry was performed after MBW. Spirometry was performed according to the American Thoracic Society/European Respiratory Society standards [17]; different devices were used at each of the three sites. Absolute values of forced expiratory volume in 1 second (FEV_1) were standardised for height, age, sex and ethnicity using the Global Lung Function Initiative reference equations [18]. Mid-expiratory forced expiratory flow ($FEF_{25-75\%}$) was reported from the effort with the highest sum of FEV_1 and forced vital capacity.

Statistical analysis

Table 1 summarises the methods applied to calculate the inter-visit reproducibility for each pair-wise comparison of the six LCI measurements. An average reproducibility was calculated with the pooled pair-wise results of all time points using the mixed-effects model; the residual standard deviation from the random effects was used with the mean difference to calculate 95% confidence bands (mean differences ±1.96×standard deviation). To account for repeated observations in the same individuals, the mixed-effects model was specified with subject as the random factor. An exchangeable correlation structure was used. LCI values were right skewed, therefore non-parametric summaries are presented, whereas percentage change in LCI followed a normal distribution, and parametric summaries were used. The same analyses were conducted for FEV_1 and $FEF_{25-75\%}$ z-scores and % predicted. The sample size was a convenience sample size based on available data from the primary research study [3]. All statistical analyses were performed using Stata version 14 (StataCorp, College Station, TX, USA).

Results

77 participants with cystic fibrosis and 71 healthy preschool children had at least two MBW measurements performed during the study period. Of the 324 technically acceptable MBW measurements in healthy participants, 295 (91%) were asymptomatic; whereas, of the 343 technically acceptable MBW tests in cystic fibrosis participants, 210 (61%) were considered asymptomatic and clinically stable. The LCI was significantly higher in the cystic fibrosis group (median 8.9 (range 6.4–16.2)) at the enrolment visit compared with the healthy children (median 7.1 (range 6.1–8.1)). The reproducibility, defined as the absolute difference of LCI between two repeated measurements was close to zero in both health (mean difference: −0.03) and cystic fibrosis (mean difference: −0.05), regardless of the time interval between measurements; whereas the within-subject between-test %CV of the LCI was approximately double that in cystic fibrosis (7.7%) compared with healthy children (4.3%) (table 2). The ICC, representing the correlation between two measurements in the same individual, was much lower for healthy children (average 0.4), compared with cystic fibrosis participants (average 0.7).

A meaningful change in LCI can be defined as 0.9 units (table 2) based on the Bland–Altman limits of agreement (figure 1) and coefficient of repeatability calculated from healthy data. Thus, the probability that a change of more than 0.9 LCI units is due to chance is less than 5%, and can be considered clinically meaningful. The threshold is more than double (2 LCI units) when derived from measurements in

TABLE 1 Summary of statistical methods used to calculate the inter-visit reproducibility

Method	Calculation	Assumptions
Absolute differences	$LCI_{t2}-LCI_{t1}$	
Percentage change	$(LCI_{t2}-LCI_{t1})\times100/LCI_{t1}$	
Within-subject between-test standard deviation	Difference of each LCI value from the mean LCI for each subject	
Coefficient of variation [32]	Within-subject standard deviation×100/ within-subject mean	
Bland and Altman limits of agreement [36]	Mean difference $(LCI_{t2}-LCI_{t1})\pm1.96\times SD$ of the differences	The variance is proportional to the mean
Coefficient of reproducibility [37]	Within-subject standard deviation×2.77	The variance is proportional to the mean
Intra-class coefficient [35]	Proportion of observed variance between individuals calculated using a random effects model	

LCI: lung clearance index; t: time or test occasion.

TABLE 2 Reproducibility of lung clearance index between each pair of repeated measurements in preschool children

	Subjects n	Absolute Difference	%CV	Limits of agreement	CR	ICC	Percentage change (95% limits)
Healthy controls							
Enrolment–1 month	46	−0.11	4.39	−1.14–0.91	0.85	0.40	−1.4 (−16–13)
Enrolment–3 months	44	−0.18	4.26	−1.18–0.81	0.84	0.48	−2.4 (−16–11)
3–6 months	45	0.24	4.52	−0.87–1.34	0.89	0.52	3.7 (−12–19)
6–9 months	48	−0.19	4.41	−1.33–0.96	0.88	0.44	−2.3 (−17–13)
9–12 months	47	0.01	4.27	−1.10–1.13	0.85	0.59	0.4 (−15–16)
Enrolment–6 months	47	0.07	3.68	−0.88–1.02	0.74	0.49	1.1 (−12–14)
Enrolment–9 months	41	−0.13	4.58	−1.27–1.01	0.90	0.27	−1.7 (−17–14)
Enrolment–12 months	44	−0.07	4.43	−1.32–1.18	0.89	0.36	−0.9 (−18–16)
Average		−0.03	4.3%	−1.1–1.1	0.90	0.40	−0.14 (−15–15)
Cystic fibrosis							
Enrolment–1 months	21	0.36	8.30	−2.94–3.66	2.29	0.70	4.0 (−28–36)
Enrolment–3 months	38	−0.03	6.35	−2.03–1.98	1.54	0.88	0.6 (−22–23)
3–6 months	37	−0.07	7.96	−2.92–2.79	2.03	0.74	1.1 (−27–29)
6–9 months	33	−0.36	7.58	−3.13–2.40	1.98	0.68	−2.4 (−29–24)
9–12 months	35	0.12	5.53	−1.93–2.17	1.42	0.83	2.1 (−19–24)
Enrolment–6 months	35	−0.09	7.39	−2.61–2.43	1.90	0.54	0.6 (−26–27)
Enrolment–9 months	37	−0.08	9.01	−3.12–3.47	2.35	0.58	3.0 (−31–37)
Enrolment–12 months	35	0.10	9.23	−3.54–3.73	2.52	0.74	3.3 (−32–39)
Average		−0.05	7.7%	−2.9–2.8	2.00	0.70	1.27 (−25–27)

%CV: coeffiecient of variation; CR: coefficient of reproducibility; ICC: intra-class correlation coefficient. The average was calculated from the pooled pair-wise comparisons of all of the time points as the residual standard deviation from the random effects of a mixed-effects model.

clinically stable cystic fibrosis participants. However, both the limits of agreement and coefficient of repeatability assume that the within-subject standard deviation is proportional to the magnitude of the measurement, which we observed not to be the case (figure 2). The bias was not as apparent for healthy children in whom the range of LCI values is narrow, but this bias increases in cystic fibrosis, especially at higher LCI values (figure 2). Thus, using the limits derived in healthy children in cystic fibrosis would lead to over-estimation of clinically relevant changes in those with higher LCI values (*i.e.* greater disease severity).

The reproducibility of repeated LCI measurements for each of the time intervals, calculated as a percentage change, was, on average, very close to zero in both health and stable cystic fibrosis (table 2) and independent of the magnitude of LCI (figure 3). To assess the percentage change in LCI that would be considered a threshold beyond the biological variability of the test, we used the average of the pair-wise repeated measurements; 95% of the percentage change observed in healthy preschool children was between ±15%. In cystic fibrosis participants, the limits were higher (±25%). The percentage change was similar for
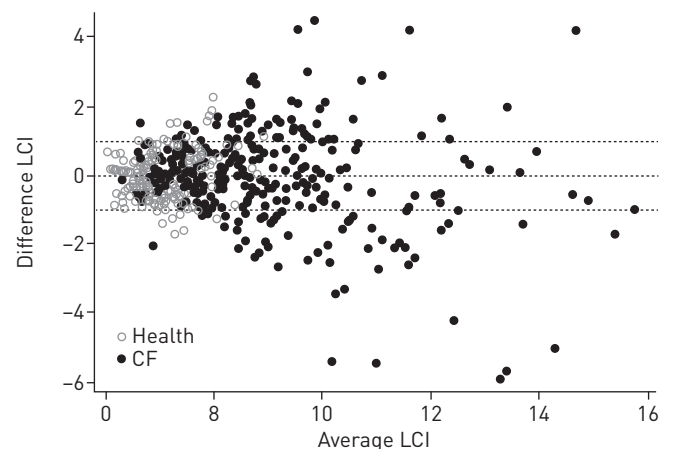

FIGURE 1 Bland–Altman plot of the difference in repeated lung clearance index (LCI) measurements. The difference between LCI measurements is greater at higher LCI values. CF: cystic fibrosis.

FIGURE 2 The within–subject between-test standard deviation of the lung clearance index (LCI) was found to be proportional to the magnitude, especially in cystic fibrosis (CF) subjects.
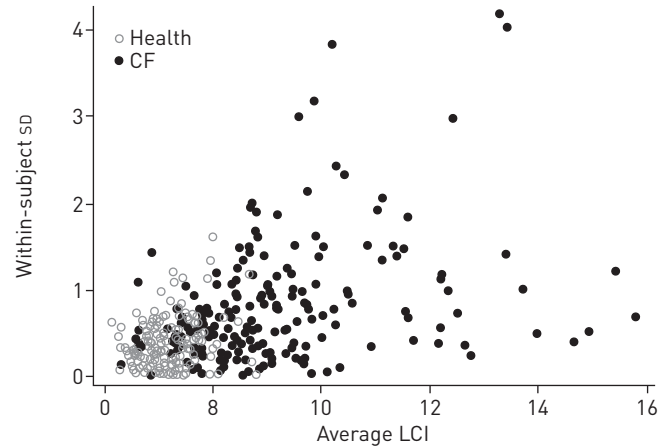
measurements made 1 month apart compared with those made 3 months apart, in both health and cystic fibrosis (table 2). In addition, the percentage change in LCI was similar for each of the pair-wise comparisons across the 12-month period in healthy children and cystic fibrosis participants, albeit the range was much wider for the cystic fibrosis group (figure 4). LCI_5, at 1/20 of the starting concentration, was more variable than the standard LCI (at 1/40 of the starting gas concentration), with 95% confidence bands in health ranging from −35% to 34%. The observed percentage change for FEV_1 were similar to those observed for LCI (95% confidence bands in health (−19–16%) and cystic fibrosis (−29–26%)); however, the reproducibility of $FEF_{25-75\%}$ were more than double that observed for FEV_1 and LCI in both health and cystic fibrosis (table 3). Since interpretation of % predicted can be biased, we include the absolute changes in z-scores which were also more variable in cystic fibrosis than in health, and were more variable for z $FEF_{25-75\%}$ than for $zFEV_1$.

LCI reproducibility was also defined for measurements in cystic fibrosis participants made 24 h apart. The average percentage change was −1.3% (95% confidence bands ±17%). When the limits defined in healthy preschool children were applied, 16 (94%) out of 17 measurements made 24 h apart in cystic fibrosis participants had a percentage change in LCI within ±15% of the baseline measurement and 100% within ±25%. In contrast, using an absolute change, only 9 (53%) out of 17 of the 24 h measurements in cystic fibrosis participants were within 0.9 units in LCI (the coefficient of repeatability derived from the healthy preschool children).

Discussion

We comprehensively assessed the biological variability of repeated LCI measurements at time points relevant for clinical care of cystic fibrosis patients. These data highlight that the interpretation of LCI in terms of absolute change (*i.e.* 1 unit) is prone to bias, especially in cystic fibrosis. A percentage change in LCI of ±15% could be considered greater than the biological variability of the test in healthy children, and is comparable to measurements made 24 h apart in cystic fibrosis. In clinically stable cystic fibrosis patients measured at longer time intervals (1–3 months), the biological variability was within ±25%. The greater variability likely represents asymptomatic and spontaneous variations in disease not captured by

FIGURE 3 The percentage change in lung clearance index (LCI) between two test occasions was found to be independent of the magnitude of the LCI.
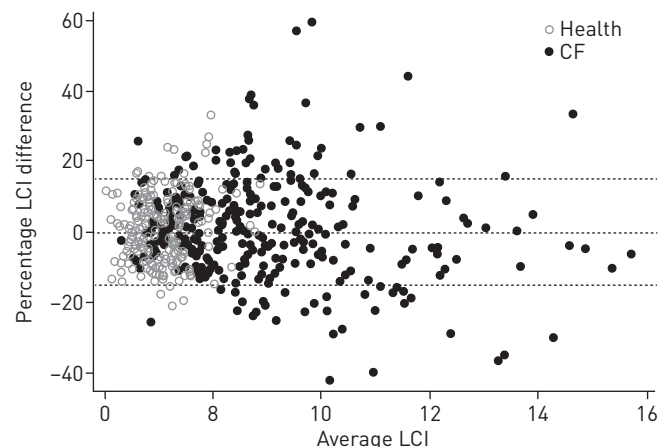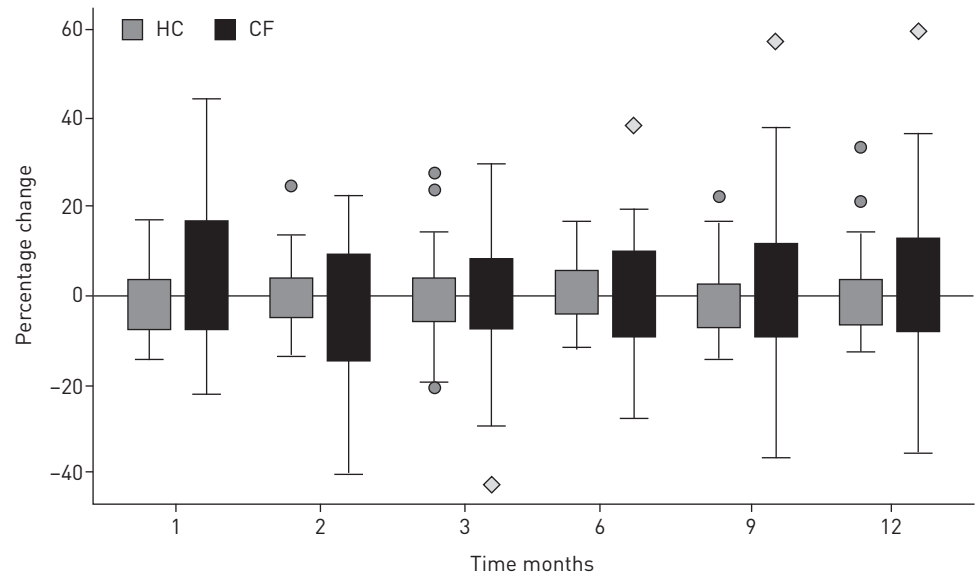
FIGURE 4 Percentage changes between LCI measurements were independent of the time interval between measurements in both health and in stable cystic fibrosis disease. Boxplots indicate the median value (centre line); inter-quartile range (box) and minimum and maximum values, excluding outliers greater than three-times the lower quartile (error bars).

clinical symptoms; however, differentiating biological variability from changes in disease state in this population is challenging and requires further investigation. Our previous work highlighted the population-level changes in LCI in preschool children over time and identified changes in LCI are associated with lower respiratory symptoms [3]. The in-depth analyses of the individual-level biological variability of the LCI presented here provide further evidence and tools to interpret clinical changes in LCI in young children with cystic fibrosis.

Changes in LCI >15% can be considered physiologically relevant and greater than the biological variability of the test in health. At the healthy spectrum of LCI (*i.e.* LCI=7), a 15% change in LCI reflects a change of approximately 1 unit, which coincides with the Bland–Altman limits of agreement and coefficient of repeatability defined in this study. The inter-visit %CV and absolute change in LCI are consistent with previous literature; however the range of observed LCI values in previous studies was narrower, thus the relationship between the variability and magnitude may not have been apparent [6, 12, 15]. For instance, SINGER *et al.* [12] reported a coefficient of repeatability of 0.96 units for measurements in cystic fibrosis made 24 h apart, which was similar to what we observed in healthy participants. The LCI values observed by SINGER *et al.* [12] ranged from 7.3 to 11.5 units in cystic fibrosis; thereby, are not representative of the spectrum of lung disease observed in patients with cystic fibrosis. In our data, at the severe spectrum of cystic fibrosis lung disease (*i.e.* LCI=16), the LCI would need to change by more than 2.5 units to be considered physiologically relevant, highlighting the limitations of using absolute LCI as a threshold for a relevant change. We also observed greater variability for the LCI_5 between visits than for LCI (at 1/40 the starting N_2 concentration), which is opposite to the within-visit variability [3] but corresponds to the effects observed for interventional studies [19].

The observed within-subject between-test reproducibility of LCI was relatively high (±15%), but consistent with FEV_1 reproducibility defined using similar methods in the same study population. Absolute z-scores changes ranged from −1.3 to 1 z-scores, corresponding to approximately −19% to 16% percentage change when using % predicted. This z-score range is similar to a previous study in children reporting reproducibility for measurements made 12 months apart [20]. Furthermore, multiple studies demonstrate that the "signal" such as treatment differences, or changes related to worsening symptoms, from both of these lung function tests is much greater than the "noise" or biological variability and can be useful for tracking cystic fibrosis lung disease over time [1, 3, 21–24]. The similar within-subject variability observed for LCI and FEV_1 is in contrast to the lower between-subject variability observed for LCI compared to FEV_1 in interventional studies [7, 8, 25]. This paradox is not unexpected, since changes at a group level are often lower than changes observed within an individual [26, 27]. Compared with the reproducibility of the LCI and FEV_1, the reproducibility of the $FEF_{25-75\%}$ was rather poor (absolute z-score changes ranged from −2.5 to 1.8 z-scores, which corresponds to approximately −55% to 65% percentage changes when

TABLE 3 Reproducibility of FEV1 and FEF25-75% z-scores for absolute change (95% limits of agreement) and % predicted for percentage change (95% confidence bands)

	Subjects n	FEV1 z-scores absolute change	FEV1 % predicted percentage change	Subjects n	FEF25-75% z-scores absolute change	FEF25-75% % predicted percentage change
Healthy controls						
Enrolment–1 month	31	−0.1 (−1.4–1.1)	−1.6 (−20–17)	21	−0.4 (−1.9–1.2)	−7.8 (−47–31)
Enrolment–3 months	26	−0.3 (−1.5–0.9)	−3.5 (−22–15)	15	−0.1 (−1.2–1)	−1.3 (−36–34)
3–6 months	28	−0.1 (−1.3–1.2)	0.4 (−15–16)	17	−0.1 (−1.4–1.1)	−2.1 (−34–30)
6–9 months	31	0.03 (−1.3–1.2)	−0.6 (−17–16)	21	0.02 (−1.3–1.2)	1.2 (−32–34)
9–12 months	31	0.003 (−1.2–1.2)	0.7 (−18–19)	25	0.1 (−1.3–1.4)	3.7 (−39–46)
Enrolment–6 months	31	−0.2 (−1.6–1.3)	−2.0 (−26–22)	15	−0.2 (−1.5–1.0)	−4.7 (−38–28)
Enrolment–9 months	28	−0.2 (−1.5–1.0)	−2.9 (−25–19)	16	−0.1 (−1.5–1.3)	−1.8 (−39–35)
Enrolment–12 months	28	−0.4 (−1.6–0.9)	−4.4 (−22–14)	16	−0.5 (−2.6–1.5)	−9.8 (−58–38)
Average		−0.1 (−1.3–1.0)	−1.5 (−19–16)		−0.1 (−1.3–1.1)	−1.2 (−31–34)
Cystic fibrosis						
Enrolment–1 months	16	−0.1 (−1.6–1.4)	−1.5 (−26–23)	13	−0.5 (−2.5–1.5)	−10.4 (−55–34)
Enrolment–3 months	18	−0.2 (−1.8–1.4)	−2.3 (−28–23)	11	−0.4 (−3.3–2.5)	−7.0 (−75–61)
3–6 months	19	−0.3 (−2.0–1.3)	−3.8 (−27–20)	14	−0.1 (−1.9–1.7)	1.0 (−46–48)
6–9 months	19	0.2 (−1.8–2.2)	3.6 (−33–40)	15	−0.2 (−2.3–1.9)	−3.7 (−74–67)
9–12 months	26	0.1 (−1.6–1.8)	2.4 (−25–30)	21	0.2 (−1.5–2.0)	12.1 (−57–81)
Enrolment–6 months	16	−0.4 (−2.2–1.4)	−5.3 (−32–22)	13	−0.6 (−3.5–2.3)	−8.5 (−77–60)
Enrolment–9 months	19	−0.03 (−1.8–1.8)	0.3 (−28–28)	12	−0.6 (−3.2–1.9)	−11.0 (−74–52)
Enrolment–12 months	24	−0.2 (−2.1–1.8)	−1.2 (−31–29)	20	−0.6 (−2.7–1.5)	−13.2 (−63–36)
Average		−0.1 (−1.8–1.6)	−1.3 (−29–26)		−0.4 (−2.5–1.8)	−5.0 (−55–65)

FEV1: forced expiratory volume in 1 s; FEF25–75%: mid-expiratory forced expiratory flow. The average was calculated from the pooled pair-wise comparisons of all of the time points as the residual standard deviation from the random effects of a mixed-effects model.

using % predicted). This is consistent with previous literature which suggests both the within- [28] and between-subject variability [29] of the FEF25-75% is too high to track individuals over time, despite the sensitivity of FEF25-75% to detect differences at the population level [30].

Numerous statistical approaches exist to calculate the biological variability of a test, each with its own advantages, disadvantages and statistical assumptions that may be specific to an outcome. Generally, less variable measurements are more precise, thus better at tracking individuals over time [31]. The Bland–Altman limits of agreement are widely used to define a threshold of measurement error and biological variability, but assume that the error is independent of the measurement itself [32]. Our findings clearly demonstrate that, for LCI, this assumption does not hold true. The effect was less obvious in healthy participants (Bland–Altman limits of agreement defined in health); therefore, likely to impact the interpretation of changes observed in cystic fibrosis, where the bias is more pronounced. The discordance with previous studies is not unexpected since these were either limited to healthy subjects, or a narrow range of LCI values [12, 15]. In some previous studies within-test repeatability criteria of the FRC and LCI have been applied, such that higher LCI measurements with greater variability were excluded [6]. MBW data in our study relied on a standardised quality control protocol that examined trials for technical quality and variability of the breathing pattern, thus no exclusions were made based on measurement variability.

Previous studies define inter-test reproducibility as the %CV [5, 7, 8] or the ICC [6]. Both of these are useful parameters to summarise consistency and agreement, but do not provide a meaningful limit of reproducibility [33]. Furthermore, both the %CV and ICC are better suited for situations where there are more than two repeated measurements in the same individual [34] which were not available for most of the previously published data sets. The %CV observed in our study was comparable to values previously published [5, 7, 8]; whereas the ICC was much lower. O'NEILL et al. [6] report that the ICC of LCI measurements made in cystic fibrosis participants approximately 8 months apart (range 67–614 days) was 0.96. The discordance in results likely reflects the inherent sensitivity of the ICC to the magnitude (or spread) of values in the dataset [33]. Indeed, the ICC calculated from the healthy children in our study was much lower than the 0.6 limit suggested as a threshold for a meaningful test [35], but the spread of values in health was smaller. A similar bias can be expected if results in cystic fibrosis participants represent only a small segment of the spectrum of disease.

Limitations
Our findings are limited by the narrow age range of preschool data, and may not be generalisable to older patients with more advanced lung disease. These observations were based on patients followed at three centres and there is a need to validate in a different study populations. Further longitudinal studies in older cystic fibrosis patients across a range of disease severity are necessary to better interpret changes seen in the clinical setting. The lack of a gold standard to define the variability of LCI (or FEV1) in cystic fibrosis limited our ability to distinguish variability due to asymptomatic disease from inherent variability of ventilation inhomogeneity in those with cystic fibrosis; the 24 h reproducibility data from a small number of older patients does suggest that biological variability of LCI is similar in health and cystic fibrosis, and the poorer reproducibility observed at longer time intervals is associated with progression of lung disease. Additional studies, potentially including images of both structural changes as well as images of ventilation (e.g. hyper-polarised gas magnetic resonance imaging) may help to better understand LCI variability observed in cystic fibrosis patients who clinically appear stable. Finally, the spirometry collected in preschool children was limited to a smaller number of observations, and thus the results may be more variable than those observed in older children and adults. These were included as a reference and comparison, and were not meant to define reproducibility of preschool spirometry.

Conclusion
Our findings suggest that the reproducibility of LCI between two consecutive measurements should be interpreted as percentage changes, where changes greater than ±15% represent changes greater than the biological variability of the test and are physiologically relevant.

R. Amin and S. Stanojevic. Drafting the work or revising it critically for important intellectual content: E. Oude Engberink, F. Ratjen, S.D. Davis, G. Retsch-Bogart, R. Amin and S. Stanojevic. Final approval of the version to be published: E. Oude Engberink, F. Ratjen, S.D. Davis, G. Retsch-Bogart, R. Amin and S. Stanojevic. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: E. Oude Engberink, F. Ratjen, S.D. Davis, G. Retsch-Bogart, R. Amin and S. Stanojevic. S. Stanojevic is the guarantor and accepts full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

References

1   Aurora P, Stanojevic S, Wade A, *et al.* Lung clearance index at 4 years predicts subsequent lung function in children with cystic fibrosis. *Am J Respir Crit Care Med* 2011; 183: 752–758.

2   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.

3   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.

4   Fuchs SI, Gappa M. Lung clearance index: clinical and research applications in children. *Paediatr Respir Rev* 2011; 12: 264–270.

5   Fuchs SI, Eder J, Ellemunter H, *et al.* Lung clearance index: normal values, repeatability, and reproducibility in healthy children and adolescents. *Pediatr Pulmonol* 2009; 44: 1180–1185.

6   O'Neill K, Tunney MM, Johnston E, *et al.* Lung clearance index in adults and children with cystic fibrosis. *Chest* 2016; 150: 1323–1332.

7   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.

8   Amin R, Subbarao P, Lou W, *et al.* The effect of dornase alfa on ventilation inhomogeneity in patients with cystic fibrosis. *Eur Respir J* 2011; 37: 806–812.

9   Belessis Y, Dixon B, Hawkins G, *et al.* Early cystic fibrosis lung disease detected by bronchoalveolar lavage and lung clearance index. *Am J Respir Crit Care Med* 2012; 185: 862–873.

10   Horsley AR, Gustafsson PM, Macleod KA, *et al.* Lung clearance index is a sensitive, repeatable and practical measure of airways disease in adults with cystic fibrosis. *Thorax* 2008; 63: 135–140.

11   Singer F, Houltz B, Latzin P, *et al.* A realistic validation study of a new nitrogen multiple-breath washout system. *PloS One* 2012; 7: e36083.

12   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.

13   Sonnappa S, Bastardo CM, Wade A, *et al.* Repeatability and bronchodilator reversibility of lung function in young children. *Eur Respir J* 2013; 42: 116–124.

14   Amin R, Stanojevic S, Kane M, *et al.* A randomized controlled trial to evaluate the lung clearance index as an outcome measure for early phase studies in patients with cystic fibrosis. *Respir Med* 2016; 112: 59–64.

15   Fuchs SI, Toussaint S, Edlhaimb B, *et al.* Short-term effect of physiotherapy on variability of the lung clearance index in children with cystic fibrosis. *Pediatr Pulmonol* 2010; 45: 301–306.

16   Jensen R, Stanojevic S, Klingel M, *et al.* A systematic approach to multiple breath nitrogen washout test quality. *PloS One* 2016; 11: e0157523.

17   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: 1304–1345.

18   Quanjer PH, Stanojevic S, Cole TJ, *et al.* Multi-ethnic reference values for spirometry for the 3–95-year age range: the global lung function 2012 equations. *Eur Respir J* 2012; 40: 1324–1343.

19   Stanojevic S, Jensen R, Sundaralingam D, *et al.* Alternative outcomes for the multiple breath washout in children with CF. *J Cyst Fibros* 2015; 14: 490–496.

20   Kirkby J, Bountziouka V, Lum S, *et al.* Natural variability of lung function in young healthy school children. *Eur Respir J* 2016; 48: 411–419.

21   Kieninger E, Singer F, Fuchs O, *et al.* Long-term course of lung clearance index between infancy and school-age in cystic fibrosis subjects. *J Cyst Fibros* 2011; 10: 487–490.

22   Taylor-Robinson D, Whitehead M, Diderichsen F, *et al.* Understanding the natural progression in %FEV1 decline in patients with cystic fibrosis: a longitudinal study. *Thorax* 2012; 67: 860–866.

23   Corey M, Edwards L, Levison H, *et al.* Longitudinal analysis of pulmonary function decline in patients with cystic fibrosis. *J Pediatr* 1997; 131: 809–814.

24   Corey M, Farewell V. Determinants of mortality from cystic fibrosis in Canada, 1970–1989. *Am J Epidemiol* 1996; 143: 1007–1017.

25   Davies J, Sheridan H, Bell N, *et al.* Assessment of clinical response to ivacaftor with lung clearance index in cystic fibrosis patients with a G551D-CFTR mutation and preserved spirometry: a randomised controlled trial. *Lancet Respir Med* 2013; 1: 630–638.

26   Cella D, Bullinger M, Scott C, *et al.* Group *versus* individual approaches to understanding the clinical significance of differences or changes in quality of life. *Mayo Clin Proc* 2002; 77: 384–392.

27   US Department of Health and Human Services FDA Center for Drug Evaluation and Research, US Department of Health and Human Services FDA Center for Biologics Evaluation and Research and US Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes* 2006; 4: 79.

28   Bakker EM, Borsboom GJ, van der Wiel-Kooij EC, *et al.* Small airway involvement in cystic fibrosis lung disease: Routine spirometry as an early and sensitive marker. *Pediatr Pulmonol* 2013; 48: 1081–1088.

29   Quanjer PH, Weiner DJ, Pretto JJ, *et al.* Measurement of FEF25-75% and FEF75% does not contribute to clinical decision making. *Eur Respir J* 2014; 43: 1051–1058.

30   Lum S, Gustafsson P, Ljungberg H, *et al.* Early detection of cystic fibrosis lung disease: multiple-breath washout versus raised volume tests. *Thorax* 2007; 62: 341–347.

31 Vaz S, Falkmer T, Passmore AE, *et al.* The case for using the repeatability coefficient when calculating test-retest reliability. *PloS One* 2013; 8: e73990.
32 Bland JM, Altman DG. Measurement error proportional to the mean. *BMJ* 1996; 313: 106.
33 Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med* 1990; 20: 337–340.
34 Bruton A, Conway JH, Holgate ST. Reliability: What is it, and how it measured. *Physiotherapy* 2000; 86: 94–99.
35 Chinn S. Statistics in respiratory medicine. 2. Repeatability and method comparison. *Thorax* 1991; 46: 454–456.
36 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.
37 Bland JM, Altman DG. Measurement error. *BMJ* 1996; 313: 744.