Longitudinal assessment of lung clearance index to monitor disease progression in children and adults with cystic fibrosis

Alex R Horsley, John Belcher, Katie Bayfield, Brooke Bianco, Steve Cunningham, Catherine Fullwood, Andrew Jones, Anna Shawcross, Jaclyn A Smith, Anirban Maitra, Francis J Gilchrist

ABSTRACT

Background Lung clearance index (LCI) is a valuable research tool in cystic fibrosis (CF) but clinical application has been limited by technical challenges and uncertainty about how to interpret longitudinal change. In order to help inform clinical practice, this study aimed to assess feasibility, repeatability and longitudinal LCI change in children and adults with CF with predominantly mild baseline disease.

Methods Prospective, 3-year, multicentre, observational study of repeated LCI measurement at time of clinical review in patients with CF <5 years, delivered using a rapid wash-in system.

Results 112 patients completed at least one LCI assessment and 98 (90%) were still under follow-up at study end. The median (IQR) age was 14.7 (8.6–22.2) years and the mean (SD) FEV₁ z-score was −1.2 (1.3). Of 81 subjects with normal FEV₁ (≥−2 z-scores), 63% had raised LCI (indicating worse lung function). For repeat stable measurements within 6 months, the mean (limits of agreement) change in LCI was 0.9% (−18.8% to 20.7%). A latent class growth model analysis identified four discrete clusters with high accuracy, differentiated by baseline LCI and FEV₁. Baseline LCI was the strongest factor associated with longitudinal change. The median total test time was under 19 min.

Conclusions Most patients with CF with well-preserved lung function show stable LCI over time. Cluster behaviours can be identified and baseline LCI is a risk factor for future progression. These results support the use of LCI in clinical practice in identifying patients at risk of lung function decline.

INTRODUCTION

Lung clearance index (LCI) derived from the multiple breath washout (MBW) test is an established research outcome for individuals with cystic fibrosis (CF). The test involves following an inert gas washed out from the lungs during tidal breathing. This makes it simple to perform from a patient’s perspective and correspondingly applicable to a wide range of ages and disease states. Key advantages of LCI over FEV₁ include increased sensitivity to early changes in airway obstruction, ability to be performed repeatedly even in very young children and a very stable upper limit of normal, even in growing lungs. LCI is recognised as a valuable endpoint in CF clinical trials, with guidelines for the performance and interpretation of the test and has played an important role in supporting registration of novel cystic fibrosis transmembrane conductance regulator (CFTR) modulator therapies.

Despite this increased acceptance and use of LCI in research, the technique has yet to become established in routine clinical care. From a technical perspective, the test takes more time than spirometry to perform (to which it would be additional) and the equipment is poorly mobile, making the test difficult to integrate into multidisciplinary CF clinic scheduling. Key clinical questions remain relating to natural variability in LCI progression over time and what constitutes a minimally important clinical change in LCI within patients.

To address these challenges and enable integration of LCI into clinical practice, we performed "LCI-SEARCH", a National Institute for Health Research-funded multicentre, all-age, prospective study of the value of LCI in clinical CF practice.
To reduce technical barriers, we employed a closed-circuit system which delivers more rapid wash-in, reducing overall test time, as well as allowing the system to be fully mobile. Since LCI appears to have most value in those with mild disease, this study specifically focused on routine follow-up in children and adults with well-preserved FEV₁ who were considered free of Pseudomonas. The objectives of this study were the following:

- To evaluate feasibility, acceptability and clinical value of repeated LCI measurements in clinical monitoring in CF outpatient clinics.
- To assess short-term reproducibility in patients with clinically stable CF with predominantly mild disease.
- To assess long-term trajectory of LCI in CF and identify risk factors for accelerated decline.

METHODS

This was a prospective, single-blind, observational study of children (≥5 years) and adults with CF, where routine LCI testing was integrated into clinical care in parallel with conventional clinic-based spirometry. Patients were recruited from three specialist CF centres in the UK: Wythenshawe Hospital, Manchester, Royal Manchester Children’s Hospital and University Hospital North Midlands. Patients had FEV₁ > 50% and were recruited from non-Pseudomonas clinics at each site (defined as currently free of chronic infection with Pseudomonas aeruginosa). Those who were subsequently reclassified as chronically infected or displayed new Pseudomonas infections during follow-up remained in the study. Patients and parents provided written informed consent and children assent.

Study visits

Patients were assessed at their usual clinic appointments, including both routine and emergency visits. Patients or parents completed a short questionnaire, comparing current symptoms with usual baseline and identifying any other symptoms of a pulmonary exacerbation. Study visits took place between November 2014 and February 2018. In order to establish repeatability and individual patient trajectories, LCI data were kept blind until the final 6 months of the study. Following this, they were revealed to clinical teams in real time at each clinic visit, along with a patient-specific report and graph showing all historical LCI data. During this period, clinicians were asked to rate the impact of having these LCI data on clinical decision-making using a 3-point scale (no impact, partial impact, strong impact). Adult patients were provided with a questionnaire to rate their experience of repeated LCI measurement. Details are provided in the online supplemental file.

Multiple breath washout

MBW was performed using a closed-circuit Innocor system (PulmoTrace ApS, Glamsbjerg, Denmark), as previously described and detailed in the online supplemental file. Detailed analysis and quality control were performed in a separate offline washout analysis package prepared in Igor Pro V6 (Wavemetrics, Lake Oswego, Oregon, USA), as previously described. Washout repeats were excluded if there was evidence of leak or a large difference between LCI or functional residual capacity (FRC) measurements (>25% from median). Final LCI and FRC values are the average of at least two reproducible repeats. The upper limit of normal for LCI was 6.9.

Statistical analysis

To assess between-test short-term repeatability of LCI, only paired measurements taken within 6 months of each other were included. Patients were required to be clinically stable at both assessments, defined as no additional oral or intravenous antibiotics within 14 days, deemed well by the reviewing physician and FEV₁ change <10% of previous measurement. Bland-Altman analysis and intra-class correlation (ICC) coefficient were used to describe change in LCI between visits.

Longitudinal analysis was only performed on those with at least four valid measurements while clinically stable. Latent class growth analysis (LCGA) was used to identify distinct trajectories of LCI data. This method is described in more detail in online supplemental file. LCGA was undertaken using the LCMM package in R.

Linear mixed modelling with multivariable adjustment was used to investigate clinical factors associated with change in LCI over time. A random effects model with an exchangeable correlation structure was used to estimate the ICC values and limits of agreement. Covariates were chosen a priori to include the following clinical data previously recognised to be associated with lung function decline: Pseudomonas status at study start (chronic and intermittent infection vs not infected), intravenous antibiotic courses, all antibiotic-treated exacerbations, age, body mass index and gender. Pancreatic status was added subsequently during data analysis. Baseline LCI and FEV₁ were included, but in order to reduce the number of factors only FEV₁ z-score was included of the spirometric indices.

Original sample size was planned to be ≥70 participants. This was based on an estimate of what was required for reasonably robust longitudinal modelling and not calculated against a specific outcome (since no longitudinal data sets existed at the time).

RESULTS

The study recruited 122 children and adults with CF, of whom 112 (92%) completed at least one successful LCI assessment and
98 (90%) were still under follow-up when the study finished. A CONSORT (Consolidated Standards of Reporting Trials) diagram of patient outcomes is shown in figure 1 and in the online supplemental file.

Summary demographic and clinical data for those with at least one LCI measurement are shown in table 1. The median (range) age at study entry was 14.7 (5.1–65.3) years. Forty-four subjects were adults (>18 years) assessed at the adult CF centre and 68 were children. Two subjects transitioned to adult care and had measurements performed at both paediatric and adult centres.

### Feasibility

Eight participants (7%) were unable to perform tests capable of producing LCI outputs: four patients after repeated attempts (including one adult) and four after attempts on a single visit. All were withdrawn or withdrew from the study. Two further subjects with equivocal sweat chloride and no CFTR gene mutations on full gene sequencing were reclassified as non-CF and excluded from all analyses. Of the 77 children consented, 70 (91%) were therefore able to perform LCI measurements, compared with 98% of adults. The remaining 112 patients completed a total of 913 LCI visits, with a median of 6 successful LCI measurements (IQR 4–9) per subject. In 67 visits, it was not possible to obtain the minimum requirement of two reproducible washout repeats, giving an overall visit success rate of 92.7%. Failed visits were more common in children (11.3% of all visits) than adults (3.3%) (p<0.001). The reasons for test failure are described in online supplemental file and figure E2). Overall 40 (36%) were pancreatic sufficient, 50 (45) were children (3.3%) (p<0.001). The reasons for test failure are described in online supplemental file and figure E2).

### Patient population

In line with study objectives, patients generally had mild disease, with a mean FEV1 z-score of −1.2 (range −4.3 to 1.3). Nineteen (17%) had FEV1 >100% predicted (see also online supplemental file and figure E2). Overall 40 (36%) were pancreatic sufficient, but this proportion was significantly higher in adults (61% vs 19% in children, p=0.0001). Other indicators that adult patients were drawn from those with milder disease include a lower rate of P. aeruginosa (3%) patients with diabetes, 5 (4%) with CF liver disease and 19 (17%) with a coexistent diagnosis of asthma.
The median LCI at visit 1 was 7.7 (range 5.6–15.1) and there was no significant difference between children and adults (table 1). The relationship between FEV1 z-score and LCI at visit 1 is presented in figure 2. There was significant correlation between measurements (r=−0.43, 95% CI −0.58 to −0.25, p<0.0001). However, LCI was elevated in a large proportion of those with normal FEV1: of 81 subjects with normal FEV1 at visit 1, 51 (63%) had LCI of ≥6.9. The median total LCI test time was under 19 min, with shorter test times in children: median (IQR) 16.3 (13.5–19.3) min vs 21.2 (18.6–24.3) min in adults (p<0.0001).

Repeatability of LCI measurements
For assessment of LCI repeatability, 80 subjects contributed 313 valid data pairs of stable LCI measurements within 6 months of each other. Of the total 781 eligible LCI measurements, 315 (40%) were excluded from repeatability assessments due to the patient being unwell or on additional antibiotics. The median (IQR) interval between included measurements was 105 (70–154) days (approximately 3.5 months). The mean (SD) absolute difference in LCI was 0.01 (0.85), representing a mean (SD) change of 0.9% (10.1) of baseline LCI. The Bland-Altman limits of agreement were therefore −18.8% to 20.7% (see figure 3 and online supplemental figure E3). The ICC was 0.93 (95% CI 0.91 to 0.94) (see online supplemental file, table E1). As a sensitivity analysis, these analyses were repeated for all unique pairs of LCI (n=152). These produced highly similar results and are presented in the online supplemental file.

Longitudinal change in LCI
Latent class analysis
Latent class analysis identified four distinct clusters, based on the trajectory of LCI. These are shown in figure 4 and a descriptive summary is presented in table 2 with a list of clinical variables commonly used to describe CF lung disease. The four clusters were unevenly distributed, with the majority of subjects (58, 72%) being grouped together on the basis of LCI values which remained stable over the course of the study (cluster 1, 'stable near normal'). The other three clusters consisted of low LCI rising with time (cluster 2, 'near normal, increasing LCI', n=8); elevated LCI falling over time (cluster 3, 'abnormal, stable/improving', n=7); and high LCI increasing over time (cluster 4, 'abnormal, increasing', n=8). There was no difference in baseline clinical variables between clusters, with the exception of baseline LCI and the spirometric indices FEV1, FVC and forced expiratory flow z-scores. The spirometric indices are commonly highly associated with each other, and lower spirometry values and higher LCI were found in the clusters which experienced change in LCI over time.

LCGA was repeated for children and adults separately. Analysis is limited by sample size, but these exploratory analyses, presented in the online supplemental file, showed similar patterns of clustering into three groups (online supplemental figures E6 and E7). Also included in the online supplemental file is a sub-analysis of patients with normal-range FEV1 (z-score >−2) at visit 1, comparing those with normal LCI with those with high LCI (>6.9). Overall there was a greater mean change in LCI over the course of the study in those with normal FEV1.
and normal LCI at visit 1: mean (SE) 0.60 (0.23) units vs −0.07 (0.17) units (p = 0.02). However, those with high LCI were much more likely to be in one of the clusters showing change in LCI over time (29% vs 9%).

Linear mixed model analysis

The strongest factor associated with LCI change over time was baseline LCI (p < 0.001). The model also identified age, baseline LCI at visit 1: mean (SE) 0.60 (0.23) units vs −0.07 (0.17) units (p = 0.02). However, those with high LCI were much more likely to be in one of the clusters showing change in LCI over time (29% vs 9%).

**DISCUSSION**

In this multicentre, prospective study we have successfully introduced routine LCI measurements into clinical practice in children.
and adults. This has enabled the measurement of repeatability and change over time and identified different patient clusters based on LCI and trajectory of LCI. *Pseudomonas* infection is known to be associated with higher LCI values, so this study specifically targeted those free of chronic infection and with generally mild impairment in FEV₁. They may not therefore be representative of the entire CF population, which may explain why less change in LCI over time was seen than in studies with a greater proportion of *Pseudomonas*-infected patients. With well-preserved lung function increasingly common in older children and adults, and likely to be more so with CFTR modulator therapies, there is however a greater unmet need for clinically scalable sensitive lung function monitoring in this group of subjects and for knowledge of how this evolves over time.

A challenge of assessing stability of LCI in CF is that ventilation heterogeneity is an inherently unstable property due to shifting patterns of mucus that can change with physiotherapy and treatments as well as underlying disease state. Thus although within-session repeatability of LCI was good and similar to that of healthy controls, the between-session repeatability over 6 months was ±20%. This is narrower than that previously reported in preschool children and similar to that of school-aged children with CF. Tighter reproducibility may be seen over shorter time spans or in clinical trials, but for clinical practice this figure therefore seems relatively robust across different age groups and devices.

More relevant to clinical practice than simple paired measurements, we have also described longitudinal trajectories of LCI to determine risk factors for progressive disease. It was reassuring that the majority of these patients with predominantly mild CF fell into a cohort with stable LCI throughout the course of the study. Similar findings were reported recently in school-aged children. This has inevitably also made it harder to detect significant differences between the remaining cohorts. Nonetheless, 10% of participants showed evidence of LCI progression from a low (ie, well-preserved) baseline, and a similar proportion showed LCI rising from a higher baseline. Using linear mixed model analysis we identified age, baseline LCI, baseline FEV₁, *Pseudomonas* status and rate of exacerbations requiring intravenous antibiotics as differentiating factors. This matches observations in patients with more severe disease that those with lower lung function, older age and with chronic *Pseudomonas* are more likely to show longitudinal decline in lung function. Using these longitudinal data, we have also assessed the point-of-care impact of contemporaneous LCI results on clinical decision-making. This was challenging to deliver, requiring both pre-preparation of reports and rapid analysis and integration of real-time data. For these reasons numbers are limited, but in over half of cases clinicians identified that the measurement provided additional information about clinical status, above that from clinical review and spirometry.

There have been a small number of other longitudinal LCI studies, and a common picture of LCI in monitoring CF is emerging. For very young children, elevated preschool LCI seems to predict higher LCI at early school age and at adolescence. LCI seems to remain relatively stable during the early school years, but increases in adolescence, a time when predicted FEV₁ may be less useful due to rapid changes in lung size. Steeper changes in LCI over time are also seen with concurrent *Pseudomonas* infection. It has been recognised for some time that better tools to monitor lung function decline in CF are required and on the face of it LCI fits this bill well. There remain however significant challenges relating to the delivery of this measurement in routine practice, and this has so far only been successfully delivered in a handful of institutes. Barriers to routine use include technical, training and clinical factors. From a technical perspective, a range of devices are now available, requiring differing techniques and demonstrating poor correlation across systems, which has probably hindered clinical integration.

We addressed the practical challenges by using a system with closed-circuit wash-in, which allowed rapid, portable measurements and reduced total test time to a median of around 20 min. MBW assessment was well tolerated by patients, with no major or consistent issues identified with the procedure itself, although the biggest single user complaint remained one of time taken to complete testing. Delivering quality-controlled measurements to clinicians required real-time review and analysis by an experienced operator and may not be feasible in all clinics. What we have shown is that for the majority of patients with mild CF LCI remained stable. For most CF centres struggling with the practicalities of delivering real-time MBW, a more realistic and practical ambition would be to perform these less frequently, for example at annual review or on request (eg, where spirometry is equivocal or unreliable). Based on our observations, this would help to identify those with low and stable LCI as well as those with raised or progressively deteriorating LCI who would merit further assessment or treatment. This addresses one of the CF monitoring challenges recently posed by the UK National Institute for Health and Care Excellence. A combination of borderline-raised LCI, with or without impaired FEV₁, and increased need for rescue antibiotics may represent a group to target in future trials to investigate whether LCI trends can be reversed.

Delivery of LCI measurements in real-world clinical setting is both a strength and a limitation of this study. Intervals between visits were not standardised, with adult patients in particular being assessed more frequently than every 3 months. This has led to varying numbers of assessments for different subjects. On the other hand, this study was focused on assessing the value of LCI in a clinically relevant population (those with mild disease) and in a clinically realistic setting, where assessments may not be rigidly scheduled. Adult patients in this study appeared to be phenotypically somewhat different from the children, with a lower proportion with ‘classical’ CFTR mutations and pancreatic insufficiency. In order to resolve this, analyses were repeated for the adult and paediatric populations separately. Small numbers in some cohorts mean that these additional analyses should be considered as exploratory only.

In summary, we have shown that, with the appropriate resources, LCI can be routinely delivered in a clinical setting and is generally acceptable to patients. Most patients with CF with well-preserved lung function show stable LCI over time; however, cluster behaviours can be identified that could serve as interventional groups in future studies. We have reported on acceptable repeatability of clinical measurements and shown that baseline LCI is a risk factor for future progression of LCI. These results support the use of LCI in clinical practice in identifying patients at risk of lung function decline, but the measurement is challenging to deliver in routine practice and in many cases may be better suited to annual assessments.

**Author affiliations**

1Division of Infection, Immunity and Respiratory Medicine, The University of Manchester Faculty of Biology, Medicine and Health, Manchester, UK
2Manchester Adult Cystic Fibrosis Centre, Manchester University NHS Foundation Trust, Manchester, UK
3School of Medicine, Keele University, Keele, UK
4Respiratory Medicine, Children’s Hospital at Westmead, Westmead, New South
Cystic fibrosis

Wales, Australia
5MRC Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK
6Statistics, Research and Innovation, Manchester University NHS Foundation Trust, Manchester, UK
7MHSC Centre for Biostatistics, University of Manchester, Manchester, UK
8Royal Manchester Children’s Hospital, Manchester University NHS Foundation Trust, Manchester, UK
9Academic Department of Child Health, University Hospitals of North Midlands NHS Trust, Stoke-on-Trent, UK
10Institute of Applied Clinical Sciences, Keele University, Keele, UK

Twitter Alex R Horsley @alexhorsley

Acknowledgements
The authors would like to thank all the patients and their families who took part in this study, and the clinical teams at the participating centres. This work was carried out at the NIHR Manchester Clinical Research Facility.

Contributors
ARH, JAS, AI conceived the study. ARH led the study. ARH, KB, BB, AS, AM and FJG were responsible for data acquisition. JB, CF and ARH carried out the data analyses, and together with SC, AI and FJG were responsible for data interpretation and drafting of the work. All authors have approved the final manuscript.

Funding
This research was funded by the National Institute for Health Research (NIHR), under grant code NIHRCS012-13. Additional support was provided by the NIHR Manchester Clinical Research Facility, the NIHR Manchester Biomedical Research Centre and the NIHR Clinical Research Network. ARH and JAS are supported by the NIHR Manchester Biomedical Research Centre. JAS is funded by a Wellcome Investigator Award (207504/B/17/Z) and is an NIHR Senior Investigator.

Disclaimer
Wellcome Investigator Award (207504/B/17/Z) and is an NIHR Senior Investigator.

Open access
This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iDs
Alex R Horsley http://orcid.org/0000-0003-1828-0058
Steve Cunningham http://orcid.org/0000-0001-7342-251X
Jaclyn A Smith http://orcid.org/0000-0001-8837-4928
Francis J Gilchrist http://orcid.org/0000-0002-1947-7621

REFERENCES
15. NICE. Cystic fibrosis: diagnosis and management NICE guideline. NG78, 2017.
Longitudinal assessment of lung clearance index to monitor disease progression in children and adults with cystic fibrosis

Alex R. Horsley, John Belcher, Katie J. Bayfield, Brooke G. Bianco, Steve Cunningham, Catherine L. Fullwood, Andrew M. Jones, Anna Shawcross, Jaclyn A. Smith, Anirban Maitra, Francis J Gilchrist

Online Data Supplement to https://doi.org/thoraxjnl-2021-216928

Additional Methods

1. **Clinical Data**

Use of additional antibiotics were recorded at each patient visit. A physician-diagnosed exacerbation was defined as additional antibiotics prescribed for a change in respiratory symptoms. Courses of consecutive antibiotics, without an antibiotic-free interval, were scored as a single exacerbation. Mixed use of oral and intravenous (IV) antibiotics (either simultaneously or sequentially) were scored as an exacerbation requiring IV antibiotics, regardless of IV course length. Exacerbations were expressed as events per year, based on the patient-specific length of follow-up.

This study took place before the widespread introduction of CFTR modulator therapies. One adult and three paediatric patients with class III gating mutations were on Ivcacftor. Three patients took part in blinded trials: one withdrew to do this; the other two adult patients remained in the study but their data during the blinded phase have not been used.

2. **Multiple Breath Washout (MBW)**

MBW was performed using a closed circuit Innocor™ system (PulmoTrace ApS, Glamsberg, Denmark), as previously described (1, 2). Wash-in was performed from a sealed bag filled with a mixture of room air and test gas (94% O₂, 1% SF₆ and 5% N₂O) up to a total bag volume of 4L, adjusted according to patient size. Expired air was depleted of CO₂ prior to re-inspiration. At the start of wash-in, participants took 5-6 slow deep inhalations before returning to tidal breathing. Final washin concentration of expired SF₆ was between approximately 0.1 and 0.4%, depending on the starting concentration in the bag and the ratio of bag volume to FRC, as previously described (2). At the end of wash-in, the
participants were switched to room air using fast-responding pneumatic valves within the breathing unit, and instructed to maintain tidal breathing. Washout was continued until expired end-tidal SF₆ concentration reached <2.5% of the starting concentration. There was no requirement for a delay between end of washout and start of next wash-in, and subjects started the next test as soon as they were able. Distraction was provided by showing age-appropriate movies or TV shows. In the case of adults, visual feedback of inspiratory volumes was also available to aid reproducibility of breathing patterns, and typically set at 10-15ml/kg. Washouts were performed in the outpatient clinic rooms or on the ward using a portable system as previously described(2).

Both children and adults used identical patient interfaces and mouthpieces, with the only difference being that a smaller filter was used in children (subjects <18yrs). Due to refinements in the patient interface over the 3 years of the study, total deadspace varied from 50 to 58mls for the paediatric setup and from 55-65mls for the adult setup. Children transitioning to adult care also transitioned from paediatric to adult setup. It has been assumed that these small changes in deadspace volume have not affected measurements.

Subjects completed three washouts. If one or more tests were obviously compromised (e.g. evidence of leak), then additional tests were added. Detailed analysis and quality control were performed in a separate offline custom-built washout analysis package prepared in Igor Pro v6 (Wavemetrics Inc., Lake Oswego, OR, USA), as previously described (2-5). Washout repeats were excluded if there was evidence of leak, or in case of large differences seen in LCI or FRC measurements (>25% from median) (6). Final LCI and FRC measurements quoted are the average of at least two reproducible repeats.

Operator training and quality control was led by the study lead (AH). Completed test files were sent electronically for centralised review by AH, who also analysed all washouts.

Washout test time was taken from the length of the washout file. This is the total time to complete all wash-in and washout tests, including additional tests required, any interval between tests, and analyser warm-up time (60 seconds). It does not include time taken to explain the test to the participants, or time taken to clean the apparatus between volunteers.

3. Patient experience questionnaire

Adult study participants were asked to complete a participant experience form immediately after testing, on a single occasion in the final 12 months of the study. The form provided opportunity for free-text feedback about the MBW test, asked subjects to identify the worst part, and provided visual analogue scales (VAS) out of 100 on

- “How easy was the test was to perform?” (“Not at all easy” to “Very easy”)
- “Rate the time taken to complete” (“Far too long” to “Just right”).
4. **Assessment of clinical impact**

For the final 6 months of the study period, clinicians were asked to rate the impact of LCI on clinical decision making. Patient data were loaded into the study database which was used to generate graphic reports showing all stored LCI and spirometry to date, along with times of exacerbation marked on the graph. Clinicians were provided with training to understand LCI and data on LCI variability generated in the first half of the study period. Assessment of impact depended upon having completed the LCI measurement before clinical review (in some cases not possible due to logistic issues) and having the data analysed and QC-checked before clinical review, which required the presence of an experienced operator (AH). Providing these conditions were met, the physician reviewing the patient recorded the clinical outcome immediately after reviewing the patient and rated the impact of the LCI measurement on that decision process as below:

1 – **None.** LCI data not relevant to clinical decision/outcome.

2 – **Partial.** LCI data played some role in clinical decision/outcome.

3 – **Strong.** LCI data were major factor in clinical decision/outcome.

Reasons for no impact could include all data in concordance (e.g. patient clinically stable with no change in LCI or other lung function measures) or patient clearly unwell and likely to receive treatment irrespective of LCI.

5. **Statistical analyses**

Data were analysed using Prism version 8 (GraphPad, San Diego, USA), R version 3.6.0 (Vienna, Austria) and Stata version 15.1 (IBM, New York, USA). Parametric data were expressed as mean (standard deviation) and nonparametric data expressed as median (interquartile range). Comparisons were performed using unpaired t test for normally distributed data, Mann-Whitney U test, or 2-tailed Fisher’s exact test for proportions. No adjustment was made for multiplicity and p<0.05 was considered statistically significant.

Target population size was 70 patients with regular follow-up, estimated to provide sufficient numbers for robust longitudinal monitoring. There was no formal power calculation and over-recruitment was permitted.

6. **Latent class growth analysis**

Latent class growth analysis (LCGA) is a person-centred method which can be considered as a special type of latent variable modelling(7-9). LCGA models allocate individuals into different groups or classes based on the shape of their latent growth curve trajectory(9). Thus each class is summarised...
by a latent growth curve with an estimated mean intercept and mean slope. The class’s intercept and slope are referred to as “latent parameters” since these parameters were unobserved prior to undertaking the analysis(10). In LCGA, the variance and covariance within each latent class is eliminated by fixing the variance of the intercept and slope to zero(7, 9, 10). Due to this lack of within-class variance, LCGA models therefore assume that all individual growth trajectories within a specific class are homogenous(7, 8, 10). LCGA can thus be thought of as a fixed effect model(11). This means that all class members have the same intercept, linear slope and quadratic slope(7, 10). Individuals are probabilistically assigned to the latent class which best reflects their latent trajectory; with individuals assigned to the class for which they have the highest posterior probability(12). The latent class growth analysis process sequentially increases the number of latent classes, until the optimal number of classes is determined (10).

Latent class growth analysis (LCGA) was used to identify distinct trajectories of LCI data, enabling classes of people with similar trajectories to be identified. LCGA was undertaken using the LCMM package in R using full information maximum likelihood (FIML). After fitting a one-class quadratic LCGA model, the number of trajectory classes was increased sequentially. The statistical fit of each model was assessed by comparing Bayesian Information Criteria (BIC), Akaike Information Criteria (AIC), Size Adjusted BIC (SABIC), Entropy and the number of members per class (%) based on the most likely class membership. Lower values for BIC, AIC and SABIC indicate a better fitting model. Cluster sizes smaller than 1% of the total cohort are considered to be insufficient while entropy cut-offs of 1.0 (perfect), 0.8 (high), 0.6 (medium), and 0.4 (low) have been proposed(13).

After determining the optimal number of trajectory classes, the baseline LCI-specific characteristics of each trajectory class were compared descriptively using Stata version 15.1 and modelled by weighted multinomial logistic regression, the weights reflecting the uncertainty in estimating the latent cluster membership.
Supplementary Results

**Figure E1:** Consort diagram showing outcomes of adults and children with CF recruited to a longitudinal study of LCI measurements. Patients are shown in groups according to the site at which they were recruited. RMCH: Royal Manchester Children’s Hospital; UHN: University Hospital North Midlands.
**Feasibility of LCI**

Excluding those visits where LCI was not attempted, and excluding also the patients described in the main manuscript who were unable to perform LCI, there were 846 LCI measurements made on 112 subjects. In addition to this, there were another 6 visits where only a single usable washout repeat was obtained, and 61 visits where no usable repeat washouts were completed. These 67 failed assessments represented 7.3% of all visits where washout was attempted, giving a success rate of 92.7%.

Failed assessments were more common in children (52/462 visits, 11.3%) than adults (15/451 visits, 3.3%), p<0.001. Reasons for failure to obtain quality controlled washouts included patient-related issues such as inability to concentrate and complete a washout test (n=3, 4.5% all test failures). At another 22 visits (32.8%), test failure was due to washout technique issues (such as incomplete washin or washout, excessive breath volumes) that had not been successfully corrected at the time of testing. In 2 cases (3.0%), washout repeats were not reproducible enough to combine. The most common causes of washout failure however, accounting for 40 visits (59.7% of all failed visits), were technical issues relating to the washout system. Some of these were easily corrected, whilst one LCI machine in particular had a leaking valve which took longer to correct and resulted in the loss of several washout datasets.

Of 846 successful LCI visits, a full set of triplicate LCI repeats was available for 683 assessments (81%). 163 repeats were excluded, making up 6.4% of the total. The usual reasons for excluding a repeat were due to poor reproducibility or due to a washout not meeting quality control (eg inadequate washin, air leak). These data, and data on how many visits required a fourth washout to obtain a triplicate dataset, were not captured separately. Operators were encouraged to include a fourth washout if they suspected quality control was poor. Total test time includes all attempted washout repeats, whether included or not.

**Baseline FEV₁**

In the original protocol, mild disease was defined as those with best FEV₁ in last 6 months as >60% predicted. This lower limit was subsequently reduced in order to capture those with recent dips in FEV₁, an issue identified in some adults. At visit 1, some patients were additionally unable to achieve their recent best spirometry. Overall, 6 adults (14%) and 2 children (3%) had an FEV₁ below 60% predicted at visit 1, whilst 25 adults (57%) and 46 children (70%) had FEV₁ above 80% predicted. Distribution of FEV₁ at visit 1 is shown in Figure E2.
Figure E2: Distribution of FEV₁ as percent predicted, measured at visit 1. Overall, 57% of adults and 70% of children had FEV₁ within the normal range. Horizontal dotted lines show FEV₁ 80% predicted (above which FEV₁ may be considered within normal range) and FEV₁ 60% predicted (above which is considered mild impairment in FEV₁).
**Repeatability**

![Figure E3: Bland-Altman plot of absolute change in lung clearance index (LCl), defined as change at visit 2 compared to visit 1, against average. Central dotted line represents mean change, outer dotted lines represent upper and lower limits of agreement.](image-url)
Table E1: Repeatability of lung clearance index (LCI). Four methods of assessing repeatability of LCI are presented: Bland-Altman limits of agreement, coefficient of variation (%), Intraclass Correlation Coefficient (ICC), and coefficient of repeatability. Data are presented for the combined dataset and for adults and children separately. Change in LCI is expressed both as absolute change in LCI between visits (V1-V2) and as percent change ((V1-V2)/V1). ICC was only calculated for absolute change in LCI.

<table>
<thead>
<tr>
<th>Category</th>
<th>Summary Values Mean (sd)</th>
<th>Bland-Altman Limits of Agreement</th>
<th>Coefficient Variation (%)</th>
<th>ICC (95% CI)</th>
<th>Coefficient of Repeatability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>v1-v2</td>
<td>0.02 (0.82)</td>
<td>-1.58, 1.63</td>
<td>2.69</td>
<td>0.93 (0.91, 0.94)</td>
</tr>
<tr>
<td></td>
<td>(v1-v2)/v1 (%)</td>
<td>1.05 (9.76)</td>
<td>-18.08, 20.18</td>
<td>10.26</td>
<td></td>
</tr>
<tr>
<td>Children</td>
<td>v1-v2</td>
<td>-0.02 (0.85)</td>
<td>-1.68, 1.64</td>
<td>2.72</td>
<td>0.95 (0.93, 0.96)</td>
</tr>
<tr>
<td></td>
<td>(v1-v2)/v1 (%)</td>
<td>0.56 (9.98)</td>
<td>-18.99, 20.12</td>
<td>5.63</td>
<td></td>
</tr>
<tr>
<td>Adults</td>
<td>v1-v2</td>
<td>0.10 (0.73)</td>
<td>-1.33, 1.54</td>
<td>13.95</td>
<td>0.81 (0.71, 0.88)</td>
</tr>
<tr>
<td></td>
<td>(v1-v2)/v1 (%)</td>
<td>1.99 (9.12)</td>
<td>15.89, 19.87</td>
<td>21.78</td>
<td></td>
</tr>
</tbody>
</table>

Sensitivity analyses

Repeatability analyses were also conducted using only unique pairs of LCI, ie where both data points were only used in the first pairing in which they occurred. This resulted in 152 valid data pairs, with a median (IQR) interval between measurements of 91 (61-126) days. Mean (SD) absolute difference in LCI was 0.06 (0.84). Mean (SD) percent change in LCI was 1.6 (9.8), making the Bland-Altman limits of agreement -17.6 to 20.7%. These data are shown in Figures E4 and E5.

Analyses were also conducted for adults and children separately. The adults contributed 208 of the data pairs (66%). Median % difference in LCI at visit 2 was -0.5%, whilst for children the median % difference in LCI at visit 2 was 1.9%, p=0.015.
**Figure E4:** Bland-Altman plot of percent change in lung clearance index (LCI), defined as change at visit 2 compared to visit 1, against average, for unique data pairs only. Central dotted line represents mean change, outer dotted lines represent upper and lower limits of agreement.

**Figure E5:** Bland-Altman plot of absolute change in lung clearance index (LCI), defined as change at visit 2 compared to visit 1, against average for unique pairs of data only. Central dotted line represents mean change, outer dotted lines represent upper and lower limits of agreement.
**Latent Class Growth Analysis**

The 4-cluster solution was considered the best fit since it returned the lowest BIC, SABIC, AIC and yielded high entropy (Table E2). All class sizes were >1% of the total number of participants. The degree to which the trajectory classes captured distinct and important patterns in the data was assessed by estimating the average posterior probability for each cluster. These values are presented for the 4-cluster solution in Table E3 and follow the GRoLTS-Checklist (14). Individual and mean LCI trajectories for the 4-cluster solution for the combined dataset are shown in Figure E6, and described in the main manuscript text.

These analyses were repeated for the adults (n=29) and children (n=52) separately and are presented in Figures E6 and E7. Due to the smaller numbers of subjects in each of these cohorts, these analyses should be considered as exploratory only. In both cases the cohorts clustered into 3 groups; there were insufficient data to form the same four groups seen with the full dataset. Also in both cases the largest group was those with stable LCI. Posterior probabilities were high for almost all clusters (>0.9). In both adults and children, univariate modelling identified differences in baseline LCI across cohorts (P<0.0001). Differences in FEV\textsubscript{1} across the cohorts were only seen with the adult data (p=0.008). No other factors were significantly associated with clusters.

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Loglik</th>
<th>AIC</th>
<th>BIC</th>
<th>SABIC</th>
<th>entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-862.958</td>
<td>1735.916</td>
<td>1747.888</td>
<td>1732.12</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>-852.061</td>
<td>1720.122</td>
<td>1739.278</td>
<td>1714.048</td>
<td>0.8685657</td>
</tr>
<tr>
<td>3</td>
<td>-845.279</td>
<td>1712.559</td>
<td>1738.898</td>
<td>1704.207</td>
<td>0.8560668</td>
</tr>
<tr>
<td>4</td>
<td>-837.169</td>
<td>1702.339</td>
<td>1735.862</td>
<td>1691.71</td>
<td>0.877377</td>
</tr>
<tr>
<td>5</td>
<td>-837.169</td>
<td>1708.339</td>
<td>1749.045</td>
<td>1695.433</td>
<td>0.6253255</td>
</tr>
</tbody>
</table>

Table E2: Model fit results for latent class growth analysis of lung clearance index trajectories.

Optimal fitting and entropy are shown for the model using 4 clusters. Loglik: log likelihood. BIC: Bayesian Information Criteria. AIC: Akaike Information Criteria. SABIC: Size Adjusted BIC.
Figure E6: Clustering of longitudinal lung clearance index (LCI) data for the whole dataset performed using latent class growth analysis. Each individual patient’s LCI data are represented by a single line, with time from first measurement shown in days on the x-axis. Data are clustered around four discrete trajectories, with individual profiles shown in the colour of the class they are clustered with. Clusters are described in the text of the main manuscript and in Table E3, the posterior classification table (below).

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
<th>N (%)</th>
<th>Probability 1</th>
<th>Probability 2</th>
<th>Probability 3</th>
<th>Probability 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>Stable, near normal LCI</td>
<td>58 (72%)</td>
<td>0.9484</td>
<td>0.0198</td>
<td>0.0318</td>
<td>0.0000</td>
</tr>
<tr>
<td>Class 2</td>
<td>Near normal LCI, increasing</td>
<td>8 (10%)</td>
<td>0.0388</td>
<td>0.9051</td>
<td>0.0541</td>
<td>0.0000</td>
</tr>
<tr>
<td>Class 3</td>
<td>Abnormal LCI, stable/improving</td>
<td>7 (9%)</td>
<td>0.0744</td>
<td>0.0136</td>
<td>0.9006</td>
<td>0.0114</td>
</tr>
<tr>
<td>Class 4</td>
<td>Abnormal LCI, increasing</td>
<td>8 (10%)</td>
<td>0.0000</td>
<td>0.0446</td>
<td>0.0218</td>
<td>0.9336</td>
</tr>
</tbody>
</table>

Table E3: Posterior Classification table for the 4-cluster LCGA model: average posterior probability for each trajectory class, representing the mean probability of an individual having that cluster assignment given their observed data.
Figure E7: Clustering of longitudinal LCI data for the paediatric cohort only, performed using latent class growth analysis. Each individual patient’s LCI data are represented by a single line, with time from first measurement shown in days on the x axis. Posterior classification table is presented below, representing the mean probability of an individual having that cluster assignment given their observed data.

<table>
<thead>
<tr>
<th>Description</th>
<th>N (%)</th>
<th>Probability 1</th>
<th>Probability 2</th>
<th>Probability 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>Stable LCI</td>
<td>36 (69%)</td>
<td>0.9977</td>
<td>0.0019</td>
</tr>
<tr>
<td>Class 2</td>
<td>Rising LCI</td>
<td>8 (15%)</td>
<td>0.0076</td>
<td>0.9409</td>
</tr>
<tr>
<td>Class 3</td>
<td>Falling LCI</td>
<td>8 (15%)</td>
<td>0.0046</td>
<td>0.2286</td>
</tr>
</tbody>
</table>

Table E4: Posterior Classification table for the 3-cluster LCGA model for paediatric lung clearance index data: average posterior probability for each trajectory class.
**Figure E8:** Clustering of longitudinal LCI data for the adult cohort only, performed using latent class growth analysis. Each individual patient’s LCI data are represented by a single line, with time from first measurement shown in days on the x-axis. Posterior classification table is presented below, representing the mean probability of an individual having that cluster assignment given their observed data.

<table>
<thead>
<tr>
<th>Description</th>
<th>N (%)</th>
<th>Probability 1</th>
<th>Probability 2</th>
<th>Probability 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class 1</strong></td>
<td>Stable LCI</td>
<td>19 (65%)</td>
<td>0.9933</td>
<td>0.0065</td>
</tr>
<tr>
<td><strong>Class 2</strong></td>
<td>High stable LCI</td>
<td>5 (17%)</td>
<td>0.0576</td>
<td>0.9400</td>
</tr>
<tr>
<td><strong>Class 3</strong></td>
<td>Rising LCI</td>
<td>5 (17%)</td>
<td>0.0001</td>
<td>0.0717</td>
</tr>
</tbody>
</table>

Table E5: Posterior Classification table for the 3-cluster LCGA model for adult lung clearance index data: average posterior probability for each trajectory class
**Trajectory based on initial FEV₁ and LCI**

The group with normal-range FEV₁ (defined as z score >-2) but raised LCI were investigated further, and compared to those with normal FEV₁ and normal LCI. Such patients represent a group collectively identified as “normal” by spirometry but divided here by LCI to explore whether this measurement, at a single visit, could provide insight into future outcomes. This analysis was performed post-hoc, and was not a part of the original analysis plan.

This analysis was only conducted on those included in the longitudinal dataset. Of these, 63 patients had normal FEV₁ at their first visit, of whom 41 (65%) had elevated LCI (>6.9). Visit 1, LCI trajectory, and cluster distribution are shown below for these two groups.

<table>
<thead>
<tr>
<th>Group description</th>
<th>Normal FEV₁, normal LCI</th>
<th>Normal FEV₁, raised LCI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>22</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) FEV₁ z score</td>
<td>-0.70 (0.77)</td>
<td>-0.63 (0.89)</td>
<td>0.7</td>
</tr>
<tr>
<td>Median (IQR) LCI</td>
<td>6.49 (5.99-6.77)</td>
<td>7.86 (7.42 - 8.91)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mean (SE) group change in absolute LCI</td>
<td>0.60 (0.23)</td>
<td>-0.07 (0.17)</td>
<td>0.02</td>
</tr>
<tr>
<td>Mean rate of change per year in LCI (SE)</td>
<td>0.082 (0.076)</td>
<td>0.032 (0.091)</td>
<td>0.70</td>
</tr>
<tr>
<td>Cluster 1 (%)</td>
<td>20 (91)</td>
<td>29 (71)</td>
<td></td>
</tr>
<tr>
<td>Cluster 2 (%)</td>
<td>2 (9)</td>
<td>4 (10)</td>
<td></td>
</tr>
<tr>
<td>Cluster 3 (%)</td>
<td>0</td>
<td>6 (14.6)</td>
<td></td>
</tr>
<tr>
<td>Cluster 4 (%)</td>
<td>0</td>
<td>2 (4.9)</td>
<td></td>
</tr>
</tbody>
</table>

**Table E6:** Comparison of those in the longitudinal analysis who had normal-range FEV₁ at visit 1, divided by their LCI at visit 1. Clusters refer to the clusters identified in latent class growth analysis and described above (Table E3).

There was no difference in baseline FEV₁ between the groups. Median LCI was, inevitably, significantly higher in the group with raised LCI. Mean absolute change in LCI over the course of the study in the group with high LCI was close to zero, and the mean slope was not significantly non-zero. In the group with normal LCI, there was a significantly higher mean change in absolute LCI over the study of 0.6 units, but mean slope was not significantly non-zero. However there were also differences in cluster distribution, with almost all (91%) of those with normal LCI being in cluster 1, indicating stable LCI over time. Patients with elevated LCI and normal FEV₁ were more likely to be in one of the other clusters.
(29%) indicating change in LCI outcomes over time. Positive and negative changes in LCI (clusters 2 and 4, vs cluster 3) were equally likely, leading to an overall minimal change in mean LCI.

Participant experience questionnaires

Questionnaires were given to participants at the time of LCI review and handed in separately to the clinic nurse. Five previously-recruited subjects were not seen during the questionnaire period, making the eligible population 41 adult subjects (including those transitioned to adult care from paediatrics). Eighteen completed questionnaires were received (44% eligible population). Responses to free text were grouped into categories. Visual analogue scores (VAS) were taken from measurements of the point where the mark made by the patient crossed the score line, and are shown graphically in bins of 10mm (Figure E10).

Question 1: “How did you find the washout testing?”

32% identified that they had experienced no problems and a further 32% answered “ok”. Some added additional comments to say the test required little effort or time (n=3), one indicated that the first test had been the hardest, and one that it was harder when unwell. One respondent answered that the test was “long” and another that it was “boring”.

Question 2: What was the worst part of the test?

Responses to this question are shown in below in Figure E9.

Question 3: How could the test be improved?

7/18 respondents (39%) were unable to identify any ways to improve the test. Five (28%) felt the test time was too long, one that the apparatus needed to provide more leg room, two recommended more practice before starting testing. Three subjects felt there should be a better selection of films provided for distraction during tidal breathing.
Figure E9: Responses by participants to the question “What was the worst part of the [MBW] test?” Responses were free text and have been grouped into themes. Participants could identify more than one issue. Total number of separate items = 20 from 18 respondents.

Question 4 (VAS): How easy did you find it to complete test?
Participants were asked to mark their experience on a 100mm VAS from “Not at all easy” to “Very easy”. Responses are shown in Figure E10. 78% scored over >80mm for ease of testing.

Question 5 (VAS): 5. How do you rate the time taken to complete one set of tests?
Participants were asked to mark their experience on a 100mm VAS from “Far too long” to “Just right”. Responses are shown in Figure E10. 78% scored over >60mm for “Just right”. 
**Figure E10:** Summary of individual participant responses to two visual analogue score questions on their experience of multiple breath washout testing. Scores were measured and placed into 10mm bins. Ease of testing (shown in green) was scored from 0mm – “Not at all easy” to 100mm - “Very easy”. Time taken (shown in blue) was scored from 0mm – “Far too long” to 100mm - “Just right”.

**Impact**

**Figure E11:** Comparison of visit LCI and the clinician-rated impact of the measurement on the clinical decision making. There were no statistically significant differences between the groups (p=0.9)
References


Longitudinal assessment of lung clearance index to monitor disease progression in children and adults with cystic fibrosis

Alex R. Horsley, John Belcher, Katie J. Bayfield, Brooke G. Bianco, Steve Cunningham, Catherine L. Fullwood, Andrew M. Jones, Anna Shawcross, Jaclyn A. Smith, Anirban Maitra, Francis J Gilchrist

Online Data Supplement to https://doi.org/thoraxjnl-2021-216928

Additional Methods

1. Clinical Data
Use of additional antibiotics were recorded at each patient visit. A physician-diagnosed exacerbation was defined as additional antibiotics prescribed for a change in respiratory symptoms. Courses of consecutive antibiotics, without an antibiotic-free interval, were scored as a single exacerbation. Mixed use of oral and intravenous (IV) antibiotics (either simultaneously or sequentially) were scored as an exacerbation requiring IV antibiotics, regardless of IV course length. Exacerbations were expressed as events per year, based on the patient-specific length of follow-up. This study took place before the widespread introduction of CFTR modulator therapies. One adult and three paediatric patients with class III gating mutations were on Ivacaftor. Three patients took part in blinded trials: one withdrew to do this; the other two adult patients remained in the study but their data during the blinded phase have not been used.

2. Multiple Breath Washout (MBW)
MBW was performed using a closed circuit Innocor™ system (PulmoTrace ApS, Glamsberg, Denmark), as previously described (1, 2). Wash-in was performed from a sealed bag filled with a mixture of room air and test gas (94% O₂, 1% SF₆ and 5% N₂O) up to a total bag volume of 4L, adjusted according to patient size. Expired air was depleted of CO₂ prior to re-inspiration. At the start of wash-in, participants took 5-6 slow deep inhalations before returning to tidal breathing. Final washin concentration of expired SF₆ was between approximately 0.1 and 0.4%, depending on the starting concentration in the bag and the ratio of bag volume to FRC, as previously described (2). At the end of wash-in, the
participants were switched to room air using fast-responding pneumatic valves within the breathing unit, and instructed to maintain tidal breathing. Washout was continued until expired end-tidal SF₆ concentration reached <2.5% of the starting concentration. There was no requirement for a delay between end of washout and start of next wash-in, and subjects started the next test as soon as they were able. Distraction was provided by showing age-appropriate movies or TV shows. In the case of adults, visual feedback of inspiratory volumes was also available to aid reproducibility of breathing patterns, and typically set at 10-15ml/kg. Washouts were performed in the outpatient clinic rooms or on the ward using a portable system as previously described.

Both children and adults used identical patient interfaces and mouthpieces, with the only difference being that a smaller filter was used in children (subjects <18yrs). Due to refinements in the patient interface over the 3 years of the study, total deadspace varied from 50 to 58mls for the paediatric setup and from 55-65mls for the adult setup. Children transitioning to adult care also transitioned from paediatric to adult setup. It has been assumed that these small changes in deadspace volume have not affected measurements.

Subjects completed three washouts. If one or more tests were obviously compromised (e.g. evidence of leak), then additional tests were added. Detailed analysis and quality control were performed in a separate offline custom-built washout analysis package prepared in Igor Pro v6 (Wavemetrics Inc., Lake Oswego, OR, USA), as previously described (2-5). Washout repeats were excluded if there was evidence of leak, or in case of large differences seen in LCI or FRC measurements (>25% from median) (6). Final LCI and FRC measurements quoted are the average of at least two reproducible repeats. Operator training and quality control was led by the study lead (AH). Completed test files were sent electronically for centralised review by AH, who also analysed all washouts.

Washout test time was taken from the length of the washout file. This is the total time to complete all wash-in and washout tests, including additional tests required, any interval between tests, and analyser warm-up time (60 seconds). It does not include time taken to explain the test to the participants, or time taken to clean the apparatus between volunteers.

3. Patient experience questionnaire

Adult study participants were asked to complete a participant experience form immediately after testing, on a single occasion in the final 12 months of the study. The form provided opportunity for free-text feedback about the MBW test, asked subjects to identify the worst part, and provided visual analogue scales (VAS) out of 100 on

- “How easy was the test was to perform?” (“Not at all easy” to “Very easy”)
- “Rate the time taken to complete” (“Far too long” to “Just right”).

2
4. **Assessment of clinical impact**

For the final 6 months of the study period, clinicians were asked to rate the impact of LCI on clinical decision making. Patient data were loaded into the study database which was used to generate graphic reports showing all stored LCI and spirometry to date, along with times of exacerbation marked on the graph. Clinicians were provided with training to understand LCI and data on LCI variability generated in the first half of the study period. Assessment of impact depended upon having completed the LCI measurement before clinical review (in some cases not possible due to logistic issues) and having the data analysed and QC-checked before clinical review, which required the presence of an experienced operator (AH). Providing these conditions were met, the physician reviewing the patient recorded the clinical outcome immediately after reviewing the patient and rated the impact of the LCI measurement on that decision process as below:

1 - **None.** LCI data not relevant to clinical decision/outcome.

2 - **Partial.** LCI data played some role in clinical decision/outcome.

3 - **Strong.** LCI data were major factor in clinical decision/outcome.

Reasons for no impact could include all data in concordance (e.g. patient clinically stable with no change in LCI or other lung function measures) or patient clearly unwell and likely to receive treatment irrespective of LCI.

5. **Statistical analyses**

Data were analysed using Prism version 8 (GraphPad, San Diego, USA), R version 3.6.0 (Vienna, Austria) and Stata version 15.1 (IBM, New York, USA). Parametric data were expressed as mean (standard deviation) and nonparametric data expressed as median (interquartile range). Comparisons were performed using unpaired t test for normally distributed data, Mann-Whitney U test, or 2-tailed Fisher’s exact test for proportions. No adjustment was made for multiplicity and p<0.05 was considered statistically significant.

Target population size was 70 patients with regular follow-up, estimated to provide sufficient numbers for robust longitudinal monitoring. There was no formal power calculation and over-recruitment was permitted.

6. **Latent class growth analysis**

Latent class growth analysis (LCGA) is a person-centred method which can be considered as a special type of latent variable modelling(7-9). LCGA models allocate individuals into different groups or classes based on the shape of their latent growth curve trajectory(9). Thus each class is summarised...
by a latent growth curve with an estimated mean intercept and mean slope. The class’s intercept and slope are referred to as “latent parameters” since these parameters were unobserved prior to undertaking the analysis. In LCGA, the variance and covariance within each latent class is eliminated by fixing the variance of the intercept and slope to zero. Due to this lack of within-class variance, LCGA models therefore assume that all individual growth trajectories within a specific class are homogenous. LCGA can thus be thought of as a fixed effect model. This means that all class members have the same intercept, linear slope and quadratic slope. Individuals are probabilistically assigned to the latent class which best reflects their latent trajectory; with individuals assigned to the class for which they have the highest posterior probability. The latent class growth analysis process sequentially increases the number of latent classes, until the optimal number of classes is determined.

Latent class growth analysis (LCGA) was used to identify distinct trajectories of LCI data, enabling classes of people with similar trajectories to be identified. LCGA was undertaken using the LCMM package in R using full information maximum likelihood (FIML). After fitting a one-class quadratic LCGA model, the number of trajectory classes was increased sequentially. The statistical fit of each model was assessed by comparing Bayesian Information Criteria (BIC), Akaike Information Criteria (AIC), Size Adjusted BIC (SABIC), Entropy and the number of members per class (%) based on the most likely class membership. Lower values for BIC, AIC and SABIC indicate a better fitting model. Cluster sizes smaller than 1% of the total cohort are considered to be insufficient while entropy cut-offs of 1.0 (perfect), 0.8 (high), 0.6 (medium), and 0.4 (low) have been proposed.

After determining the optimal number of trajectory classes, the baseline LCI-specific characteristics of each trajectory class were compared descriptively using Stata version 15.1 and modelled by weighted multinomial logistic regression, the weights reflecting the uncertainty in estimating the latent cluster membership.
Supplementary Results

**Figure E1:** Consort diagram showing outcomes of adults and children with CF recruited to a longitudinal study of LCI measurements. Patients are shown in groups according to the site at which they were recruited. RMCH: Royal Manchester Children’s Hospital; UHN: University Hospital North Midlands.
Feasibility of LCI

Excluding those visits where LCI was not attempted, and excluding also the patients described in the main manuscript who were unable to perform LCI, there were 846 LCI measurements made on 112 subjects. In addition to this, there were another 6 visits where only a single usable washout repeat was obtained, and 61 visits where no usable repeat washouts were completed. These 67 failed assessments represented 7.3% of all visits where washout was attempted, giving a success rate of 92.7%.

Failed assessments were more common in children (52/462 visits, 11.3%) than adults (15/451 visits, 3.3%), p<0.001. Reasons for failure to obtain quality controlled washouts included patient-related issues such as inability to concentrate and complete a washout test (n=3, 4.5% all test failures). At another 22 visits (32.8%), test failure was due to washout technique issues (such as incomplete washin or washout, excessive breath volumes) that had not been successfully corrected at the time of testing. In 2 cases (3.0%), washout repeats were not reproducible enough to combine. The most common causes of washout failure however, accounting for 40 visits (59.7% of all failed visits), were technical issues relating to the washout system. Some of these were easily corrected, whilst one LCI machine in particular had a leaking valve which took longer to correct and resulted in the loss of several washout datasets.

Of 846 successful LCI visits, a full set of triplicate LCI repeats was available for 683 assessments (81%). 163 repeats were excluded, making up 6.4% of the total. The usual reasons for excluding a repeat were due to poor reproducibility or due to a washout not meeting quality control (eg inadequate washin, air leak). These data, and data on how many visits required a fourth washout to obtain a triplicate dataset, were not captured separately. Operators were encouraged to include a fourth washout if they suspected quality control was poor. Total test time includes all attempted washout repeats, whether included or not.

Baseline FEV\textsubscript{1}

In the original protocol, mild disease was defined as those with best FEV\textsubscript{1} in last 6 months as >60% predicted. This lower limit was subsequently reduced in order to capture those with recent dips in FEV\textsubscript{1}, an issue identified in some adults. At visit 1, some patients were additionally unable to achieve their recent best spirometry. Overall, 6 adults (14%) and 2 children (3%) had an FEV\textsubscript{1} below 60% predicted at visit 1, whilst 25 adults (57%) and 46 children (70%) had FEV\textsubscript{1} above 80% predicted. Distribution of FEV\textsubscript{1} at visit 1 is shown in Figure E2.
**Figure E2:** Distribution of FEV₁ as percent predicted, measured at visit 1. Overall, 57% of adults and 70% of children had FEV₁ within the normal range. Horizontal dotted lines show FEV₁ 80% predicted (above which FEV₁ may be considered within normal range) and FEV₁ 60% predicted (above which is considered mild impairment in FEV₁).
Repeatability

**Figure E3**: Bland-Altman plot of absolute change in lung clearance index (LCl), defined as change at visit 2 compared to visit 1, against average. Central dotted line represents mean change, outer dotted lines represent upper and lower limits of agreement.
Table E1: Repeatability of lung clearance index (LCI). Four methods of assessing repeatability of LCI are presented: Bland-Altman limits of agreement, coefficient of variation (%), Intraclass Correlation Coefficient (ICC), and coefficient of repeatability. Data are presented for the combined dataset and for adults and children separately. Change in LCI is expressed both as absolute change in LCI between visits (V1-V2) and as percent change ((V1-V2)/V1). ICC was only calculated for absolute change in LCI.

<table>
<thead>
<tr>
<th>Category</th>
<th>Summary Values Mean (sd)</th>
<th>Bland-Altman Limits of Agreement</th>
<th>Coefficient Variation (%)</th>
<th>ICC (95% CI)</th>
<th>Coefficient of Repeatability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>v1-v2</td>
<td>0.02 (0.82)</td>
<td>-1.58, 1.63</td>
<td>2.69</td>
<td>0.93 (0.91, 0.94)</td>
</tr>
<tr>
<td></td>
<td>(v1-v2)/v1 (%)</td>
<td>1.05 (9.76)</td>
<td>-18.08, 20.18</td>
<td>10.26</td>
<td></td>
</tr>
<tr>
<td>Children</td>
<td>v1-v2</td>
<td>-0.02 (0.85)</td>
<td>-1.68, 1.64</td>
<td>2.72</td>
<td>0.95 (0.93, 0.96)</td>
</tr>
<tr>
<td></td>
<td>(v1-v2)/v1 (%)</td>
<td>0.56 (9.98)</td>
<td>-18.99, 20.12</td>
<td>5.63</td>
<td></td>
</tr>
<tr>
<td>Adults</td>
<td>v1-v2</td>
<td>0.10 (0.73)</td>
<td>-1.33, 1.54</td>
<td>13.95</td>
<td>0.81 (0.71, 0.88)</td>
</tr>
<tr>
<td></td>
<td>(v1-v2)/v1 (%)</td>
<td>1.99 (9.12)</td>
<td>15.89, 19.87</td>
<td>21.78</td>
<td></td>
</tr>
</tbody>
</table>

Sensitivity analyses

Repeatability analyses were also conducted using only unique pairs of LCI, ie where both data points were only used in the first pairing in which they occurred. This resulted in 152 valid data pairs, with a median (IQR) interval between measurements of 91 (61-126) days. Mean (SD) absolute difference in LCI was 0.06 (0.84). Mean (SD) percent change in LCI was 1.6 (9.8), making the Bland-Altman limits of agreement -17.6 to 20.7%. These data are shown in Figures E4 and E5.

Analyses were also conducted for adults and children separately. The adults contributed 208 of the data pairs (66%). Median % difference in LCI at visit 2 was -0.5%, whilst for children the median % difference in LCI at visit 2 was 1.9%, p=0.015.
**Figure E4:** Bland-Altman plot of percent change in lung clearance index (LCI), defined as change at visit 2 compared to visit 1, against average, for unique data pairs only. Central dotted line represents mean change, outer dotted lines represent upper and lower limits of agreement.

**Figure E5:** Bland-Altman plot of absolute change in lung clearance index (LCI), defined as change at visit 2 compared to visit 1, against average for unique pairs of data only. Central dotted line represents mean change, outer dotted lines represent upper and lower limits of agreement.
**Latent Class Growth Analysis**

The 4-cluster solution was considered the best fit since it returned the lowest BIC, SABIC, AIC and yielded high entropy (Table E2). All class sizes were >1% of the total number of participants. The degree to which the trajectory classes captured distinct and important patterns in the data was assessed by estimating the average posterior probability for each cluster. These values are presented for the 4-cluster solution in Table E3 and follow the GRoLTS-Checklist (14). Individual and mean LCI trajectories for the 4-cluster solution for the combined dataset are shown in Figure E6, and described in the main manuscript text.

These analyses were repeated for the adults (n=29) and children (n=52) separately and are presented in Figures E6 and E7. Due to the smaller numbers of subjects in each of these cohorts, these analyses should be considered as exploratory only. In both cases the cohorts clustered into 3 groups; there were insufficient data to form the same four groups seen with the full dataset. Also in both cases the largest group was those with stable LCI. Posterior probabilities were high for almost all clusters (>0.9). In both adults and children, univariate modelling identified differences in baseline LCI across cohorts (P<0.0001). Differences in FEV₁ across the cohorts were only seen with the adult data (p=0.008). No other factors were significantly associated with clusters.

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Loglik</th>
<th>AIC</th>
<th>BIC</th>
<th>SABIC</th>
<th>entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-862.958</td>
<td>1735.916</td>
<td>1747.888</td>
<td>1732.12</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>-852.061</td>
<td>1720.122</td>
<td>1739.278</td>
<td>1714.048</td>
<td>0.8685657</td>
</tr>
<tr>
<td>3</td>
<td>-845.279</td>
<td>1712.559</td>
<td>1738.898</td>
<td>1704.207</td>
<td>0.8560668</td>
</tr>
<tr>
<td>4</td>
<td>-837.169</td>
<td>1702.339</td>
<td>1735.862</td>
<td>1691.71</td>
<td>0.877377</td>
</tr>
<tr>
<td>5</td>
<td>-837.169</td>
<td>1708.339</td>
<td>1749.045</td>
<td>1695.433</td>
<td>0.6253255</td>
</tr>
</tbody>
</table>

Table E2: Model fit results for latent class growth analysis of lung clearance index trajectories.

Optimal fitting and entropy are shown for the model using 4 clusters. Loglik: log likelihood. BIC: Bayesian Information Criteria. AIC: Akaike Information Criteria. SABIC: Size Adjusted BIC.
Figure E6: Clustering of longitudinal lung clearance index (LCI) data for the whole dataset performed using latent class growth analysis. Each individual patient’s LCI data are represented by a single line, with time from first measurement shown in days on the x axis. Data are clustered around four discrete trajectories, with individual profiles shown in the colour of the class they are clustered with. Clusters are described in the text of the main manuscript and in Table E3, the posterior classification table (below).

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
<th>N (%)</th>
<th>Probability 1</th>
<th>Probability 2</th>
<th>Probability 3</th>
<th>Probability 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>Stable, near normal LCI</td>
<td>58 (72%)</td>
<td>0.9484</td>
<td>0.0198</td>
<td>0.0318</td>
<td>0.0000</td>
</tr>
<tr>
<td>Class 2</td>
<td>Near normal LCI, increasing</td>
<td>8 (10%)</td>
<td>0.0388</td>
<td>0.9051</td>
<td>0.0541</td>
<td>0.0020</td>
</tr>
<tr>
<td>Class 3</td>
<td>Abnormal LCI, stable/improving</td>
<td>7 (9%)</td>
<td>0.0744</td>
<td>0.0136</td>
<td>0.9006</td>
<td>0.0114</td>
</tr>
<tr>
<td>Class 4</td>
<td>Abnormal LCI, increasing</td>
<td>8 (10%)</td>
<td>0.0000</td>
<td>0.0446</td>
<td>0.0218</td>
<td>0.9336</td>
</tr>
</tbody>
</table>

Table E3: Posterior Classification table for the 4-cluster LCGA model: average posterior probability for each trajectory class, representing the mean probability of an individual having that cluster assignment given their observed data.
Figure E7: Clustering of longitudinal LCI data for the paediatric cohort only, performed using latent class growth analysis. Each individual patient’s LCI data are represented by a single line, with time from first measurement shown in days on the x axis. Posterior classification table is presented below, representing the mean probability of an individual having that cluster assignment given their observed data.

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
<th>N (%)</th>
<th>Probability 1</th>
<th>Probability 2</th>
<th>Probability 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>Stable LCI</td>
<td>36 (69%)</td>
<td>0.9977</td>
<td>0.0019</td>
<td>0.0004</td>
</tr>
<tr>
<td>Class 2</td>
<td>Rising LCI</td>
<td>8 (15%)</td>
<td>0.0076</td>
<td>0.9409</td>
<td>0.0515</td>
</tr>
<tr>
<td>Class 3</td>
<td>Falling LCI</td>
<td>8 (15%)</td>
<td>0.0046</td>
<td>0.2286</td>
<td>0.7668</td>
</tr>
</tbody>
</table>

Table E4: Posterior Classification table for the 3-cluster LCGA model for paediatric lung clearance index data: average posterior probability for each trajectory class
**Figure E8**: Clustering of longitudinal LCI data for the adult cohort only, performed using latent class growth analysis. Each individual patient’s LCI data are represented by a single line, with time from first measurement shown in days on the x-axis. Posterior classification table is presented below, representing the mean probability of an individual having that cluster assignment given their observed data.

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
<th>N (%)</th>
<th>Probability 1</th>
<th>Probability 2</th>
<th>Probability 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>Stable LCI</td>
<td>19 (65%)</td>
<td>0.9933</td>
<td>0.0065</td>
<td>0.0002</td>
</tr>
<tr>
<td>Class 2</td>
<td>High stable LCI</td>
<td>5 (17%)</td>
<td>0.0576</td>
<td>0.9400</td>
<td>0.0024</td>
</tr>
<tr>
<td>Class 3</td>
<td>Rising LCI</td>
<td>5 (17%)</td>
<td>0.0001</td>
<td>0.0717</td>
<td>0.9282</td>
</tr>
</tbody>
</table>

*Table E5*: Posterior Classification table for the 3-cluster LCGA model for adult lung clearance index data: average posterior probability for each trajectory class.
**Trajectory based on initial FEV$_1$ and LCI**

The group with normal-range FEV$_1$ (defined as z score $>-2$) but raised LCI were investigated further, and compared to those with normal FEV$_1$ and normal LCI. Such patients represent a group collectively identified as “normal” by spirometry but divided here by LCI to explore whether this measurement, at a single visit, could provide insight into future outcomes. This analysis was performed post-hoc, and was not a part of the original analysis plan.

This analysis was only conducted on those included in the longitudinal dataset. Of these, 63 patients had normal FEV$_1$ at their first visit, of whom 41 (65%) had elevated LCI ($>6.9$). Visit 1, LCI trajectory, and cluster distribution are shown below for these two groups.

<table>
<thead>
<tr>
<th>Group description</th>
<th>Normal FEV$_1$, normal LCI</th>
<th>Normal FEV$_1$, raised LCI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>22</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) FEV$_1$ z score</td>
<td>-0.70 (0.77)</td>
<td>-0.63 (0.89)</td>
<td>0.7</td>
</tr>
<tr>
<td>Median (IQR) LCI</td>
<td>6.49 (5.99-6.77)</td>
<td>7.86 (7.42 - 8.91)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mean (SE) group change in absolute LCI</td>
<td>0.60 (0.23)</td>
<td>-0.07 (0.17)</td>
<td>0.02</td>
</tr>
<tr>
<td>Mean rate of change per year in LCI (SE)</td>
<td>0.082 (0.076)</td>
<td>0.032 (0.091)</td>
<td>0.70</td>
</tr>
<tr>
<td>Cluster 1 (%)</td>
<td>20 (91)</td>
<td>29 (71)</td>
<td></td>
</tr>
<tr>
<td>Cluster 2 (%)</td>
<td>2 (9)</td>
<td>4 (10)</td>
<td></td>
</tr>
<tr>
<td>Cluster 3 (%)</td>
<td>0</td>
<td>6 (14.6)</td>
<td></td>
</tr>
<tr>
<td>Cluster 4 (%)</td>
<td>0</td>
<td>2 (4.9)</td>
<td></td>
</tr>
</tbody>
</table>

**Table E6**: Comparison of those in the longitudinal analysis who had normal-range FEV$_1$ at visit 1, divided by their LCI at visit 1. Clusters refer to the clusters identified in latent class growth analysis and described above (Table E3).

There was no difference in baseline FEV$_1$ between the groups. Median LCI was, inevitably, significantly higher in the group with raised LCI. Mean absolute change in LCI over the course of the study in the group with high LCI was close to zero, and the mean slope was not significantly non-zero. In the group with normal LCI, there was a significantly higher mean change in absolute LCI over the study of 0.6 units, but mean slope was not significantly non-zero. However there were also differences in cluster distribution, with almost all (91%) of those with normal LCI being in cluster 1, indicating stable LCI over time. Patients with elevated LCI and normal FEV$_1$ were more likely to be in one of the other clusters.
(29%) indicating change in LCI outcomes over time. Positive and negative changes in LCI (clusters 2 and 4, vs cluster 3) were equally likely, leading to an overall minimal change in mean LCI.

**Participant experience questionnaires**

Questionnaires were given to participants at the time of LCI review and handed in separately to the clinic nurse. Five previously-recruited subjects were not seen during the questionnaire period, making the eligible population 41 adult subjects (including those transitioned to adult care from paediatrics). Eighteen completed questionnaires were received (44% eligible population). Responses to free text were grouped into categories. Visual analogue scores (VAS) were taken from measurements of the point where the mark made by the patient crossed the score line, and are shown graphically in bins of 10mm (Figure E10).

**Question 1: “How did you find the washout testing?”**

32% identified that they had experienced no problems and a further 32% answered “ok”. Some added additional comments to say the test required little effort or time (n=3), one indicated that the first test had been the hardest, and one that it was harder when unwell. One respondent answered that the test was “long” and another that it was “boring”.

**Question 2: What was the worst part of the test?**

Responses to this question are shown in below in Figure E9.

**Question 3: How could the test be improved?**

7/18 respondents (39%) were unable to identify any ways to improve the test. Five (28%) felt the test time was too long, one that the apparatus needed to provide more leg room, two recommended more practice before starting testing. Three subjects felt there should be a better selection of films provided for distraction during tidal breathing.
Figure E9: Responses by participants to the question “What was the worst part of the [MBW] test?” Responses were free text and have been grouped into themes. Participants could identify more than one issue. Total number of separate items = 20 from 18 respondents.

Question 4 (VAS): How easy did you find it to complete test?

Participants were asked to mark their experience on a 100mm VAS from “Not at all easy” to “Very easy”. Responses are shown in Figure E10. 78% scored over >80mm for ease of testing.

Question 5 (VAS): 5. How do you rate the time taken to complete one set of tests?

Participants were asked to mark their experience on a 100mm VAS from “Far too long” to “Just right”. Responses are shown in Figure E10. 78% scored over >60mm for “Just right”.
**Figure E10:** Summary of individual participant responses to two visual analogue score questions on their experience of multiple breath washout testing. Scores were measured and placed into 10mm bins. Ease of testing (shown in green) was scored from 0mm – “Not at all easy” to 100mm - “Very easy”. Time taken (shown in blue) was scored from 0mm – “Far too long” to 100mm - “Just right”.

**Impact**

**Figure E11:** Comparison of visit LCI and the clinician-rated impact of the measurement on the clinical decision making. There were no statistically significant differences between the groups (p=0.9)
References


