ORIGINAL RESEARCH

Paediatric reproducibility limits for the forced expiratory volume in 1 s

Sanja Stanojevic,[1,2] Nicole Filipow,[1] Felix Ratjen[1,3]

ABSTRACT

Background Current reproducibility standards for spirometry were derived using a small adult dataset and may not be optimal for interpretation of repeated measurements of lung function in children.

Objective To define reproducibility limits for forced expiratory volume in 1 s (FEV_1) change that represent the normal within-subject between-visit variability in healthy children and evaluate these limits as a tool to monitor children with cystic fibrosis (CF).

Methods Repeated FEV_1 measurements (3 months to 5 years apart) from healthy children from the Global Lung Function Initiative data repository were used to derive a conditional change score. Spirometry and clinical data from a CF clinical database was used to verify utility in clinical practice.

Results A reproducibility change score was derived from 47 938 FEV_1 measures from 7885 healthy children 6–18 years of age. The simple algorithm, which is conditional on the initial measurement, also accounts for age and time interval between measurements. The change score limits of reproducibility were much narrower than currently used cut-offs. Specifically, changes, considered as improvements using either a 12% or 10% relative change from baseline, are too wide for children. In CF, there was overall agreement between different approaches, with the distinct advantage that the change score was not biased by regression to the mean.

Conclusions Compared with current approaches to interpretation of repeated lung function measurements, the proposed change score was less biased and provides a simple alternative to reduce misinterpretation.

INTRODUCTION

Forced expiratory volume in 1 s (FEV_1), measured by spirometry, is an important predictor of respiratory morbidity and mortality.[1–3] Consequently, FEV_1 is the primary biological measure used to assess the lung function status of individuals with respiratory diseases.[4] An improvement in FEV_1 is also frequently used as an endpoint to assess the efficacy of interventional treatments.[5–7]

Monitoring changes in lung function over time is important to evaluate disease progression and to monitor response to interventions. Current guidelines suggest a relative change in FEV_1 of greater than 12% from baseline and a 200 mL absolute change in FEV_1 as a clinically meaningful change.[4] This recommendation traced back to a 1981 review describing a significant deterioration from normal as a reduction greater than 1.65 times the coefficient

Key messages

What is the key question?
► What is a meaningful change in forced expiratory volume in 1 s (FEV_1) in children?

What is the bottom line?
► Current reproducibility limits for FEV_1 were derived from adult data and can lead to misinterpretation of results in children. We derive a change score for FEV_1 based on a large dataset of healthy children that can be used to more accurately interpret serial FEV_1 measurements.

Why read on?
► Measurement of lung function is integral to the management of children with respiratory conditions. A less biased and more accurate approach to defining a meaningful change can help to guide treatment and management decisions.

of variation (CV) of repeated FEV_1 measurements in a healthy population.[8] This commonly used cut-off corresponds to 90% limits of reproducibility and a 10% false positive rate. In the 1976 study cited by the review, the CV was derived from weekly measurements of FEV_1 in 20 adult subjects (10 non-smokers, 10 smokers)[9]; no children were included.

More recently, Kirkby *et al* define FEV_1 annual reproducibility in nearly 1000 healthy 9-year-old children as an absolute change within ±1.3 z-scores.[10] It is not clear whether these limits apply across the paediatric age range or for intervals shorter or longer than 1 year. Previous studies in children have found the correlation between repeated measurements to decrease as the interval between measurements increase.[11 12] Further these limits have yet to be implemented in clinical use.

In clinical care of patients with cystic fibrosis (CF), a change greater than 10% from baseline is often used as a threshold for a meaningful change and worsening of FEV_1 of 10% or more is commonly used as part of definitions of a pulmonary exacerbation (PEx).[13] Whether a patient returns to 90% of baseline FEV_1 after treatment with antibiotic therapy is also typically defined as the threshold for treatment success. As lung function measurements are integral to monitor and manage respiratory conditions in children, there is a need to develop reproducibility standards based on

large and representative samples of healthy children to improve the interpretation of visit to visit changes.

The objectives of this study were to (1) define individual-reproducibility limit for FEV_1 change that represents the normal within-subject variability of measurements in health and (2) evaluate whether these reproducibility limits can be used to effectively monitor lung function changes in children with CF.

METHODS

Data sources

Anonymised longitudinal spirometry measurements from healthy Caucasian children and adolescents (≤18) were obtained from the Global Lung Function Initiative (GLI) data repository (www.lungfunction.org). Spirometry in each dataset was performed according to contemporaneous standards. Subjects with at least two spirometry measurements, made at least 1 day apart were included. Measurements with missing anthropometric data (eg, height, sex) and implausible height measurements (eg, serial height measurements that decreased by more than 5 cm) were excluded.

Spirometry and clinical data from all paediatric patients (age range 3.8–17.9 years) captured within the Toronto CF database between 2009 and 2017 were obtained (n=282). Exclusions were the same as described in the healthy data and included the removal of visits less than 180 days prior to lung transplant. Treatment with intravenous antibiotics for respiratory indications was considered a physician defined PEx. Stable visits were defined as any clinical encounter that did not overlap with active treatment with either oral or intravenous antibiotics. A physician defined PEx was called an oPEx if treated with oral antibiotics or (iPEx) if treated with intravenous antibiotics, with the following restrictions: (1) minimum of 7 days of treatment, (2) maximum of 4 months of treatment for iPEx or 28 days for oPEx, (3) Minimum of 21 days between antibiotic courses (iPEx) to be considered independent events and (4) minimum of 1 day between oral and IV antibiotics to be considered separate events, otherwise the event was denoted an iPEx. The effects of antibiotic treatment were summarised as the change in FEV_1 from the start of antibiotic treatment (−7 days to 4 days) to the end of treatment (−7 to +7 days). PEx events treated with oral antibiotics were not considered as too few of these events had complete spirometry measurements at the start and end of treatment.

Participants (or parents/guardians) in each of the studies that contributed data provided informed consent for their data to be collected for the primary study. Permission from the study investigators was obtained to use these anonymised data sources for secondary analyses.

Interpretation of spirometry

ATS/ERS reproducibility: The average CV of repeated FEV_1 measurements was calculated as the SD of all FEV_1 measurements (% predicted) in an individual divided by the mean of all measurements in the individual to replicate the approach used by McCarthy *et al*.[9] Per cent predicted was used instead of litres since FEV1 increase as children grow, and therefore, the calculations are not equivalent in children. z-scores: FEV_1 z-scores ($zFEV_1$) for each visit were calculated using the GLI reference equations.[14] A $zFEV_1$ indicates how many SD a measurement is from the predicted value. The absolute difference in $zFEV_1$ was calculated as ($zFEV_1$ current − $zFEV_1$ initial).

Relative change in FEV_1: A relative change in FEV_1 was defined as (initial FEV_1 − current FEV_1)/initial FEV_1. Since FEV_1 (measured in litres) increases as children grow, relative change was calculated from the per cent predicted values which standardise measurements for height, age and sex. Per cent predicted was derived using the GLI reference equations.[14]

Change score

We developed a change score for lung function measurements based on the method originally proposed for the interpretation of serial anthropometric measurements in children.[14]

$$Change\ Score = \frac{zFEV_{1tf} - (r \times zFEV_{1ti})}{\sqrt{1 - r^2}}$$

The change score summarises the range of values that can be expected between two repeated measurements in healthy individuals, given the initial measurement. FEV_1 values measured in litres were converted to z-scores to adjust for the expected changes in lung function with normal growth and development. Once standardised we can assume that the change between repeated lung function measurements is constant, or that the difference between two measurements should be normally distributed around zero. Furthermore, repeated measurements in the same person will be correlated (r), such that the first measurement can inform the second; however, the correlation is not perfect ($r \neq 1$). In the case of lung function measurements, previous studies have shown that correlation between two measurements in the same individual is lower the longer the time interval between measurements.[11 12] The variability of measurements between individuals also depends on age,[15] therefore, the variability of measurements within an individual may also vary with age.

To estimate the correlation between repeated measurements (r), which is necessary to calculate a change score, we developed a linear equation using the Pearson correlation coefficient estimated from all pairwise measurements in the same individual. Estimates were obtained using bootstrapping to account for the extra uncertainty. Sensitivity analyses included removal of obvious outliers, removal of the first test occasion, limiting the observation to the first two test occasions or using integer ages.

The change score can be interpreted as a deterioration (negative change score) or improvement (positive change score). The further away the change score is from zero the less likely it is within the normal reproducibility limits. A change score within ±1.96 is within the limits of normal and has an inherent 5% false positive rate. A change score within ±1.65 would have a 10% false positive rate.

RESULTS

A total of 50 265 FEV_1 measurements from 9307 healthy individuals (5–18 years of age; 61% were less than 12 years) were obtained (figure 1). Less than 5% (n=2327) of measurements were excluded, the majority were (n=1404) were excluded because there was only a single lung function measurement available. The remaining 7885 subjects (table 1) had 47 938 FEV_1 measurements. On average each subject had six measurements (range: 2–13), with an average time interval between repeated measurements of 0.9 years (range: 1.2 months to 11 years). In total there were 156 723 paired measurements (eg, a subject with 4 FEV_1 measurements would contribute six visit pairs with varying age at first visit and time intervals between visits).

ATS/ERS reproducibility: The average CV between repeated FEV_1 measurements in the same individual was 5.2%, which results in a reproducibility range (1.65*CV) of ±8.6% and therefore lower than the currently recommended 12% relative change. z-scores: The average $zFEV_1$ (ie, how far each observation was
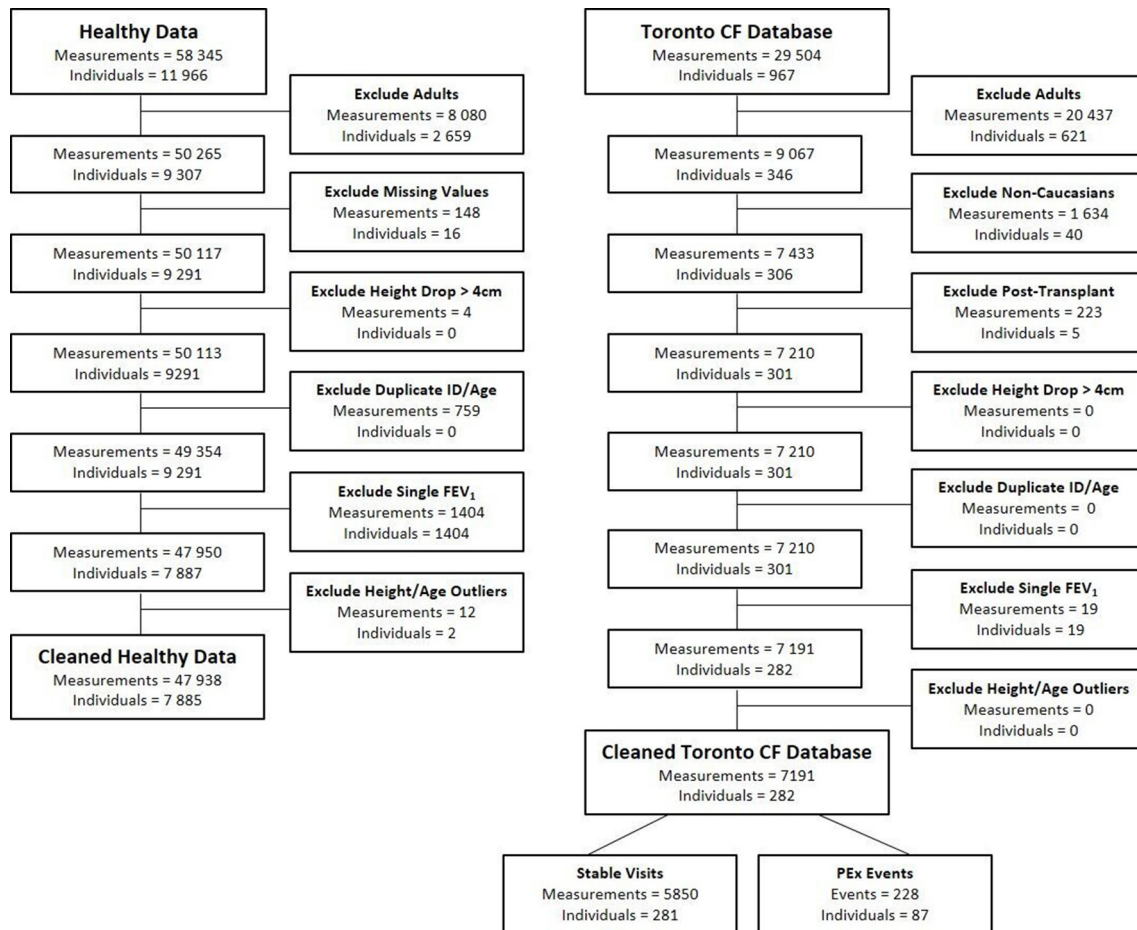
Figure 1 Flow diagram of study participants. CF, cystic fibrosis; FEV$_1$, forced expiratory volume in 1 s; PEx, pulmonary exacerbation.

from the distribution of healthy individuals in the GLI reference population)[15] of the first measurement was 0.12 (SD=1.0), indicating good overall fit with the GLI population. The change in zFEV$_1$ between measurements (zFEV$_{1\ current}$ − zFEV$_{1\ initial}$), was −0.02, 95% limits of agreement −0.8; 1.2 (range −4.7 to 6.5).

There was also a correlation observed between the first (zFEV$_{1\ initial}$) and difference between two measurements (figure 2). A healthy individual with a higher initial value is more likely to have a lower measurement at the second visit, whereas someone with a lower initial value is more likely to have a higher value.

Relative change in FEV$_1$: A total of 89% of the healthy observations had a relative change from initial visit within ±10%; however, children with lower initial values were more likely to have a change greater than 10% than children with values in the normal range, and children with higher FEV$_1$ % predicted values were more likely to have a drop greater than 10% (figure 3).

Change score: The correlation between repeated FEV$_1$ measures in the same subject decreased with time (figure 4A). Measurements made closer together in time were more correlated than those made further apart in time. Repeated measurements

Table 1 Demographic characteristics of the study population

	Healthy (n=7885)		CF (n=282)	
	Mean (SD)	**Range**	**Mean (SD)**	**Range**
Age at first measurement (years)	8.5 (1.9)	5.0–17.1	10.6 (3.1)	3.8–17.7
Height at first measurement (cm)	131.3 (12.4)	102.0–188.0	137.0 (23.3)	99.0–184.0
Height-for-age centile at first measurement (%)	54.0 (28.4)	0.02–100.0	32.9 (27.5)	0.05–96.3
No of FEV$_1$ measurements	47938		7191	
No of FEV$_1$ measurements/subject	6.1 (3.0)	2.0–13.0	25.5 (20.9)	2.0–106.0
Time between successive measurements (years)	0.9 (0.7)	0.1–11.0	0.18 (0.17)	0.003–5.7
Time between first and last measurement (years)	4.6 (3.2)	0.2–12.0	4.5 (2.9)	0.04–8.9
	Female	**Male (%)**	**Female (%)**	**Male (%)**
Sex (individuals) n (%)	3746 (47.5)	4139 (52.5)	147 (52.1)	135 (47.9)

CF, cystic fibrosis; FEV$_1$, forced expiratory volume in 1 s.
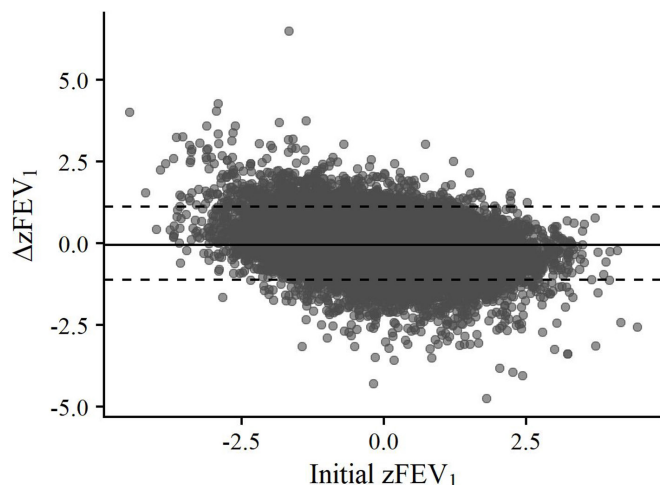
Figure 2 The change in FEV_1 between repeated measurements was strongly correlated with the initial measurement. There is a regression to the mean observed between repeated measurements. FEV_1, forced expiratory volume in 1 s; $zFEV_1$, FEV_1 z-score.

in young children were less correlated than repeated measurements in older children (figure 4B). In a multivariable regression analysis, both time and age were independently associated with the correlation between two measurements (r=0.642 (95% CI 0.62 to 0.66) $-0.04_{t\,years}$ (95% CI -0.041 to -0.033) $+0.020\,a$ years (95% CI 0.018; 0.021)); where t is the time interval between measurements in years, and a is the initial age in years. The observed age and time-dependent correlation was consistent across multiple sensitivity analyses (online supplementary table S1 and online supplement figure S2).

The change score in the healthy individuals was normally distributed with a mean of -0.12 (SD 0.97), 96% of the change scores fell within ±1.96 change scores (online supplementary figure S1). The change score was independent of the initial FEV_1 value (figure 5A).

Application in Cystic Fibrosis
A total of 282 subjects with CF in the Toronto CF database had 7191 FEV_1 measurements. On average each subject had
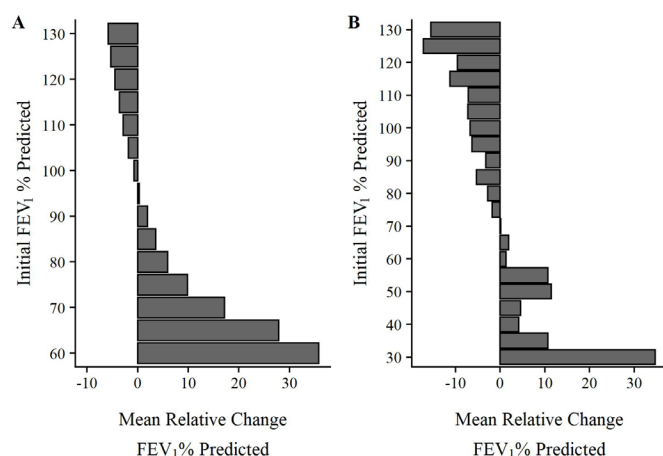
Figure 3 Waterfall plots of the relative change in FEV_1 % predicted in (A) healthy subjects and (B) children with CF averaged across groups of initial FEV_1. Between-visit change in FEV_1 was calculated from all within-patient pairwise measurements of individuals. FEV_1, forced expiratory volume in 1 s.
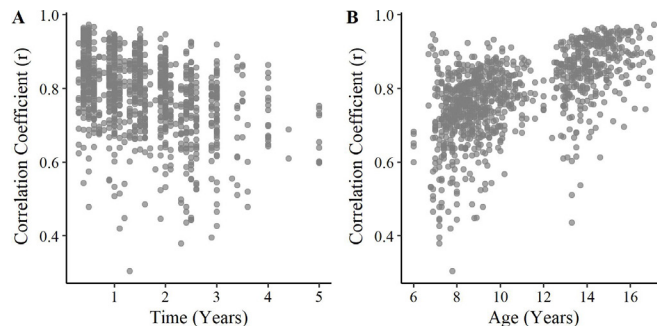

Figure 4 Correlations of repeated $zFEV_1$ measurements against A) time between measurements and (B) age at the first measurement for 1015 pairs of correlations. FEV_1, forced expiratory volume in 1 s z-scores.

25 measurements (range: 2–106), with an average time interval between repeated measurements of 2.2 months (range: 0.03 months to 5.8 years).

To illustrate how the change score could be used in practice we will consider a 6 year old patient with CF who has an improvement in FEV_1 from 79% (-1.64 $zFEV_1$) to 92% predicted (-0.66 $zFEV_1$) between two annual visits (r=0.686). This corresponds to 16% relative increase, and a conditional z-score for change of 0.64, which is well within the limits of normal variability. In contrast a 14-year-old male patient (170 cm) with a lung function drop from 90.6% (-0.78 $zFEV_1$) to 80.6% predicted (-1.60 $zFEV_1$) at 3 months (r=0.907), has a corresponding 11% relative drop and a conditional change score of -2.12, which is outside the limits of normal variability. The same drop in lung function at a 4-year interval (r=0.769) corresponds to a conditional change (Z_c=-1.56) within the limits of normal variability.

Of the 7191 clinic visits (282 subjects) in the Toronto CF Database, 5850 were identified as stable. Each patient had an average of 17 stable clinical encounters (range: 2 to 93). The average length of time between all repeated stable measurements was 2.4 years (range 0.9 days to 8.9 years). Overall 75% of the visit-to-visit changes in FEV_1 between two stable visits (not treated with either oral or IV antibiotics) were within ±1.96 change scores. The change score between stable visits was also constant across initial FEV_1 values (figure 6).

There were 228 PEx events (87 subjects) treated with IV antibiotics in the Toronto CF database with spirometry measurements at the start (-7 days to 4 days) and end of antibiotic therapy (-7 to $+7$ days). The average treatment duration was 16 days (range: 2.0–91.0 days). Of these events, 35% showed
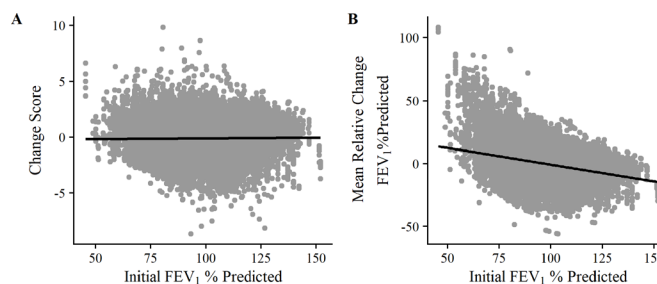
Figure 5 In healthy individuals, relationship between initial FEV_1 and (A) change score (linear mixed-effects regression 4.15–0.09*initial FEV_1) and (B) per cent predicted (linear mixed-effects regression 50.12–1.02*initial FEV_1). The change score is less influenced by initial value than relative change in FEV_1 per cent predicted. Regression coefficients centred at 50. FEV_1, forced expiratory volume in 1 s.

Stanojevic S, et al. Thorax 2020;**75**:891–896. doi:10.1136/thoraxjnl-2020-214817
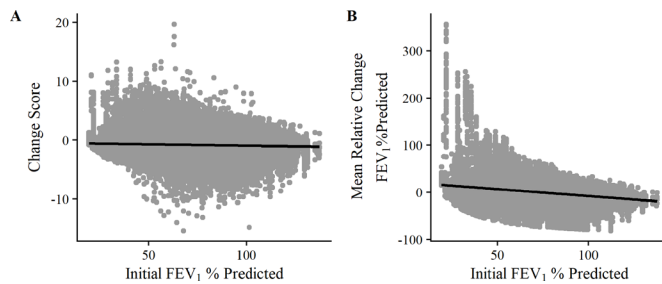
Figure 6 In stable CF, relationship between initial FEV$_1$ and (A) change score (linear mixed-effects regression 2.79–0.10*initial FEV$_1$) and (B) per cent predicted (linear mixed-effects regression 45.38–1.35*initial FEV$_1$). The change score is less influenced by initial value than relative change in FEV$_1$ per cent predicted. Regression coefficients centred at 50. CF, cystic fibrosis; FEV$_1$, forced expiratory volume in 1 s.

changes greater than a 10% relative change in % predicted FEV$_1$ and a change score greater than 1.96 units. An additional 37% did not show improvements with either a 10% relative change in FEV$_1$ % predicted nor the change score. Nearly one quarter (52/228) of visits had an improvement greater than 10% that were within 1.96 change scores.

DISCUSSION

Limits of reproducibility for FEV$_1$ were derived from a large dataset of healthy children to provide a standardised approach to interpretation of repeated lung function measurements. Currently used reproducibility limits for FEV$_1$ (eg, 12% relative change from baseline) are too wide for children and may miss important changes in lung function. In addition, fixed cut-offs (eg, 12% relative change or 10% relative change) of reproducibility are biased by the magnitude of the initial measurements such that children with lung function at the extremes of the range will have changes in lung function misinterpreted. The proposed change score is a simple algorithm that only requires knowledge of the two measured values converted to z-scores and the age of the subject at each visit. This information is readily available in pulmonary function test reports and the algorithm can easily be incorporated into trend reports.

Given the clinical need to track lung function over time for patients with respiratory diseases, a robust and unbiased measure to determine what constitutes meaningful changes is imperative. The use a relative change in FEV$_1$ per cent predicted is the widespread standard, however, we observed that this measure is biased to extreme (high and low) FEV$_1$ values in both healthy children and children with CF. Consistent with previous literature, baseline lung function in both health and disease influences how much lung function declines with symptoms, and the magnitude of treatment response.[12 16] This is particularly important for FEV$_1$ measurements in young children with CF, where lung function is both higher (often within the normal range) and more variable. These data also suggest that an improvement in FEV$_1$ of more than 10% from the start of antibiotic therapy may overestimate treatment response, as the observed changes were often within the reproducibility limits of FEV$_1$ defined by the change score. Together with the observation that healthy children with lower lung function are more likely to have improvements in lung function at the subsequent visit, the improvements seen following treatment after a drop in lung function may in part be explained by regression to the mean, and thus overestimate treatment response.

The decreasing correlation between repeated measurements in the same individual also has implications for how baseline lung function is defined, and therefore, whether there is a significant deterioration. People with CF are typically seen in clinic every 3 months, or during acute episodes of worsening. Lung function measurements from a stable visit within 3 months will be more closely correlated with the lung function measurement during the acute episode than a baseline measurement from 6 or 12 months before the episode. In some cases, the interval between stable visits may be much longer and the use of fixed cut-offs may impact the interpretation of visit to visit changes. Another consideration is that the change score is a statistically derived cut-off that represents a false positive rate of 5%. The specific cut-point can be modified based on the patient group to identify the optimal trade-off between sensitivity and specificity for the specific context and pretest probability.

In children, the 12% ATS/ERS reproducibility limits are much wider than the limits calculated from this large dataset of repeated measurements in children.[4] Further, the ATS/ERS limits and the change score are not directly comparable since the ATS/ERS limit of a 12% where the false positive rate is 10% and therefore not directly comparable to the 5% proposed by using 1.96 change scores. Recalculate the limits to be equivalent resulting in limits of reproducibility that are 10.2%, and still overestimate the range and not resolve the regression to the mean observed. Our results are consistent with previously reported reproducibility limits in children defined for 9-year-old children measured at a 1-year interval (range 9–16 months).[10] The change score has the distinct advantage of spanning a much wider age range and time intervals and is conditional on the initial measurement.

There are several limitations to our approach that warrant discussion. First, the lack of repeated measurements in adults limits the use of the conditional change score to children only. We were unable to identify any longitudinal lung function data in adults with measurements made within intervals of less than 1 year. Furthermore, we were limited to data for FEV$_1$, which can be misleading in children if flow limitation is not achieved. The median FEV$_1$/FVC ratio in healthy children is above 0.85[15] highlighting that most healthy children can expire nearly their total lung volume in approximately 1 s. Longitudinal data for FEV$_{0.75}$ is needed to provide appropriate limits of reproducibility for children that cannot reach flow limitation.[17] Alternatively, we need to identify more sensitive outcomes to measure lung function in children. Although the algorithm was developed using a large collection of healthy data in children, these datasets are not contemporaneous, such that they require validation in external datasets to confirm generalisability to current population characteristics, equipment and quality control methods. It is challenging to demonstrate that the conditional change score is superior to established limits of reproducibility as there is no clear gold standard. Use of a physician decision, such as whether or not to treat pulmonary symptoms with antibiotics, as a benchmark may be biased by indication and therefore not an ideal comparison. It will be necessary to prospectively evaluated whether tracking of lung function using this approach will result in long term improvements in outcomes. We used the GLI reference equations to calculate z-scores in order to account for the changes associated with lung growth during childhood. The simple algorithm can be applied to z-scores calculated from other reference equations. Ultimately reference equations derived from longitudinal data, where individual trajectories can be identified would be ideal; however, the practical and logistical considerations are not trivial.

CONCLUSIONS

The conditional change score for FEV_1 may help to facilitate interpretation of repeated lung function measurements in children to identify measurements that are outside the measurement variability observed in health. Once validated the relatively simple algorithm can be easily implemented into commercial software to facilitate monitoring of disease progression and to evaluate treatment response.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Research Ethics Board approval was obtained to reanalyse the existing data sources (Hospital for Sick Children REB #1000057344).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available on reasonable request. Healthy data are open access and available from the GLI Network data repository (www.lungfunction.org). Data from the Toronto CF database are available on reasonable request from the Hospital for Sick Children Research Ethics Board (ask.crs@sickkids.ca).

REFERENCES

1 Kerem E, Reisman J, Corey M, *et al*. Prediction of mortality in patients with cystic fibrosis. *N Engl J Med* 1992;326:1187–91.
2 Magnussen C, Ojeda FM, Rzayeva N, *et al*. FEV1 and FVC predict all-cause mortality independent of cardiac function - Results from the population-based Gutenberg Health Study. *Int J Cardiol* 2017;234:64–8.
3 Sabia S, Shipley M, Elbaz A, *et al*. Why does lung function predict mortality? Results from the Whitehall II Cohort Study. *Am J Epidemiol* 2010;172:1415–23.
4 Pellegrino R, Viegi G, Brusasco V, *et al*. Interpretative strategies for lung function tests. *Eur Respir J* 2005;26:948–68.
5 Stanojevic S, Ratjen F. Physiologic endpoints for clinical studies for cystic fibrosis. *J Cyst Fibros* 2016;15:416–23.
6 Szczesniak R, Heltshe SL, Stanojevic S, *et al*. Use of FEV_1 in cystic fibrosis epidemiologic studies and clinical trials: A statistical perspective for the clinical researcher. *J Cyst Fibros* 2017;16:318–26.
7 Cazzola M, MacNee W, Martinez FJ, *et al*. Outcomes for COPD pharmacological trials: from lung function to biomarkers. *Eur Respir J* 2008;31:416–69.
8 Pennock BE, Rogers RM, McCaffree DR. Changes in measured spirometric indices. What is significant? *Chest* 1981;80:97–9.
9 McCarthy DS, Craig DB, Cherniack RM. Effect of modification of the smoking habit on lung function. *Am Rev Respir Dis* 1976;114:103–13.
10 Kirkby J, Bountziouka V, Lum S, *et al*. Natural variability of lung function in young healthy school children. *Eur Respir J* 2016;48:411–9.
11 Schrader PC, Quanjer PH, Borsboom G, *et al*. Evaluating lung function and anthropometric growth data in a longitudinal study on adolescents. *Hum Biol* 1984;56:365–81.
12 Burrows B, Lebowitz MD, Camilli AE, *et al*. Longitudinal changes in forced expiratory volume in one second in adults. Methodologic considerations and findings in healthy nonsmokers. *Am Rev Respir Dis* 1986;133:974–80.
13 Fuchs HJ, Borowitz DS, Christiansen DH, *et al*. Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. The Pulmozyme Study Group. *N Engl J Med* 1994;331:637–42.
14 Cole TJ. Presenting information on growth distance and conditional velocity in one chart: practical issues of chart design. *Stat Med* 1998;17:2697–707.
15 Quanjer PH, Stanojevic S, Cole TJ, *et al*. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012;40:1324–43.
16 Hibbert ME, Lannigan A, Landau LI, *et al*. Lung function values from a longitudinal study of healthy children and adolescents. *Pediatr Pulmonol* 1989;7:101–9.
17 Joseph-Bowen J, de Klerk NH, Firth MJ, *et al*. Lung function, bronchial responsiveness, and asthma in a community cohort of 6-year-old children. *Am J Respir Crit Care Med* 2004;169:850–4.