Using different methods to process forced expiratory volume in one second (FEV₁) data can impact on the interpretation of FEV₁ as an outcome measure to understand the performance of an adult cystic fibrosis centre: A retrospective chart review [version 2; referees: 2 approved, 1 approved with reservations]

Zhe Hui Hoo, Muhaned S.A. El-Gheryani, Rachael Curley, Martin J. Wildman

1School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, S1 4DP, UK
2Sheffield Adult Cystic Fibrosis Centre, Northern General Hospital NHS Trust, Sheffield, S5 7AU, UK

Abstract

Background: Forced expiratory volume in one second (FEV₁) is an important cystic fibrosis (CF) prognostic marker and an established endpoint for CF clinical trials. FEV₁ is also used in observation studies, e.g. to compare different centre’s outcomes. We wished to evaluate whether different methods of processing FEV₁ data can impact on centre outcome.

Methods: This is a single-centre retrospective analysis of routinely collected data from 2013-2016 among 208 adults. Year-to-year %FEV₁ change was calculated by subtracting best %FEV₁ at Year 1 from Year 2 (i.e. negative values indicate fall in %FEV₁), and compared using Friedman test. Three methods were used to process %FEV₁ data. First, %FEV₁ calculated with Knudson equation was extracted directly from spirometer machines. Second, FEV₁ volume were extracted then converted to %FEV₁ using clean height data and Knudson equation. Third, FEV₁ volume were extracted then converted to %FEV₁ using clean height data and GLI equation. In addition, year-to-year variation in %FEV₁ calculated using GLI equation was adjusted for baseline %FEV₁ to understand the impact of case-mix adjustment.

Results: Year-to-year fall in %FEV₁ reduced with all three data processing methods but the magnitude of this change differed. Median change in %FEV₁ for 2013-2014, 2014-2015 and 2015-2016 was −2.0, −1.0 and 0.0 respectively using %FEV₁ in Knudson equation whereas the median change was −1.1, −0.9 and −0.3 respectively using %FEV₁ in the GLI equation. A statistically significant p-value (0.016) was only obtained when using %FEV₁ in Knudson equation extracted directly from spirometer machines.
Conclusions: Although the trend of reduced year-to-year fall in %FEV₁ was robust, different data processing methods yielded varying results when year-to-year variation in %FEV₁ was compared using a standard related group non-parametric statistical test. Observational studies with year-to-year variation in %FEV₁ as an outcome measure should carefully consider and clearly specify the data processing methods used.

Keywords
Cystic fibrosis, epidemiology, patient outcome assessment, forced expiratory volume

Corresponding author: Martin J. Wildman (martin.wildman@sth.nhs.uk)

Author roles: Hoo ZH: Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Supervision, Validation, Visualization, Writing – Original Draft Preparation; El-Gheryani MSA: Data Curation, Investigation, Writing – Review & Editing; Curley R: Data Curation, Formal Analysis, Investigation, Writing – Review & Editing; Wildman MJ: Conceptualization, Methodology, Project Administration, Supervision, Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2018 Hoo ZH et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

How to cite this article: Hoo ZH, El-Gheryani MSA, Curley R and Wildman MJ. Using different methods to process forced expiratory volume in one second (FEV₁) data can impact on the interpretation of FEV₁ as an outcome measure to understand the performance of an adult cystic fibrosis centre: A retrospective chart review [version 2; referees: 2 approved, 1 approved with reservations]
F1000Research 2018, 7:691 (https://doi.org/10.12688/f1000research.14981.2)

First published: 01 Jun 2018, 7:691 (https://doi.org/10.12688/f1000research.14981.1)
Introduction
Cystic fibrosis (CF) is a multi-system genetic condition but the
two main affected organs are lungs (resulting in recurrent infections
and respiratory failure) and gastrointestinal tract (resulting in fat
malabsorption and poor growth). Median survival has improved
to 45 years, in part because of improvement in care quality. An
important quality improvement initiative is benchmarking, which
involves identifying high-performing centres and the practices
associated with outstanding performance. Since forced
expiratory volume in one second (FEV₁) is an important CF
prognostic marker, it is often used as an outcome measure for
benchmarking.

Different statistical methods of analysing FEV₁ data can yield
different results, but there are scant attention paid to the methods
of processing FEV₁ data. We previously reported a statistically
significant reduction in year-to-year %FEV₁ fall for our CF
centre from 2013–2016. We now set out to understand the impact of using different FEV₁ data processing methods on our
CF centre’s outcome.

Methods
This is a single-centre retrospective analysis of routinely
collected clinical data from 2013–2016. Regulatory approval for
the analysis was obtained from NHS Health Research Authority
(IRAS number 210313). All adults with CF diagnosed according
to the UK CF Trust criteria aged ≥16 years were included,
except those with lung transplantation or on ivacaftor. These
treatments have transformative effects on %FEV₁, thus may affect the interpretation of year-to-year variation in %FEV₁.

Demographic data (age, gender, genotype, pancreatic status, CF
related diabetes, Pseudomonas aeruginosa status), body mass
index (BMI) and FEV₁ data were collected by two investiga-
tors (HZH and RC / HZH and MEG) independently reviewing
paper notes and electronic records. Where data from the two
investigators differ, the original data from paper notes or electronic
records were reviewed to by both investigators to ensure the
accuracy of abstracted data. This process ensures the accuracy
of abstracted data and helps avoid potential bias from inaccurate
or inconsistent data collection. FEV₁ data were processed with
three different methods prior to analysis. First, %FEV₁ readings
(calculated with Knudson equation and available in whole
numbers) were directly extracted from spirometer machines.
Second, FEV₁ volumes (in litres, to two decimal places) were
extracted and clean height data were used to calculate %FEV₁ (as
whole numbers) with Knudson equation. Third, FEV₁ volumes
(in litres, to two decimal places) were extracted and clean
height data were used to calculate %FEV₁ with GLI equation using
an Excel Macro (Microsoft Excel 2013).

Best %FEV₁, i.e. the highest %FEV₁ reading in a calendar year
for each study subject was used for analysis since it is most reflec-
tive of the true baseline %FEV₁. Year-to-year %FEV₁ change
was calculated by subtracting best %FEV₁ at Year 1 from Year 2
(i.e. negative values indicate fall in %FEV₁ and positive values
indicate increase in %FEV₁). In addition to calculating year-to-
year %FEV₁ change using three different FEV₁ data processing
methods, %FEV₁ change calculated with GLI equation was
also adjusted for baseline %FEV₁ using reference values from
Epidemiologic Study of CF (ESCF). The ESCF study found
median %FEV₁ change of 3%/year, 2%/year and 0.5%/year
for baseline %FEV₁ ≥100%, 40–99.9% and <40% respectively.
Adjusted %FEV₁ change was calculated by subtracting median
ESCF %FEV₁ change from actual %FEV₁ change. Thus, an
adjusted %FEV₁ change >0 meant the subject’s year-to-year
change in %FEV₁ was less than expected (indicating better health
outcome) whilst an adjusted %FEV₁ change <0 meant the
subject’s year-to-year change in %FEV₁ was more than expected
(indicating worse health outcome).

%FEV₁ change from 2013–2014 to 2015–2016 calculated using
different FEV₁ data processing methods were compared using
Friedman test. Bland-Altman analyses were also used to com-
pare year-to-year variation in FEV₁ as calculated with Knudson
equation against year-to-year variation in FEV₁ as calculated
with GLI equation, to understand the impact of using different
reference equations. Analyses were performed using SPSS v24
(IBM Corp) and Prism v7 (GraphPad Software). P-value <0.05
was considered statistically significant.

Results
This analysis included 208 adults, with 147 adults providing data
for all four years. Overall, the cohort was ageing but
baseline %FEV₁ increased from 2014 onwards (see Table 1).

The %FEV₁ increase was in part due to younger adults with
higher %FEV₁ transitioning from paediatric care because
%FEV₁ tended to decline from year to year (see Table 2).
However, different year-to-year change in %FEV₁ results
were obtained with different FEV₁ data processing methods. There
was statistically significant reduction in year-to-year fall in %FEV₁
using %FEV₁ readings as recorded in spirometer machines
(p=0.016). Cleaning of height data and standardisation of
%FEV₁ calculation with Knudson equation did not alter the
magnitude of year-to-year variation in %FEV₁, but the p-value
was no longer statistically significant (p=0.062). The use of

Amendments from Version 1
As recommended by Prof McKone, we have used Bland-Altman
analyses to compare different reference equations (Knudson vs
GLI).

As recommended by Prof Burgel, we have:
1. Performed a sensitivity analysis for the results in Table 2
   using only adults aged 18 years and above - we have also
done the same for the Bland-Altman analyses that were
added following suggestion from Prof McKone
2. Replaced the term “FEV1 decline” with “year-to-year FEV1
   variation”

See referee reports
Table 1. Characteristics of study subjects from 2013 to 2016.

|               | 2013                  | 2014                  | 2015                  | 2016                  |
|---------------|------------------------|------------------------|------------------------|------------------------|
| Excluded      | Lung transplantation, n | 6                      | 7                      | 9                      | 7                      |
|               | On ivacaftor, n        | 7                      | 7                      | 9                      | 9                      |
| Included, n   |                        | 166                    | 170                    | 185                    | 186                    |
| Age in years, median (IQR) | 25 (19 – 31)          | 26 (20 – 32)          | 27 (20 – 34)          | 27 (21 – 34)          |
| Female, n (%) |                        | 76 (45.8)              | 80 (47.1)              | 87 (47.0)              | 90 (48.4)              |
| Genotype status: |                        |                        |                        |                        |
| ≥1 known mutation(s), n (%) | 11 (6.6)              | 13 (7.6)              | 16 (8.6)              | 15 (8.1)              |
| ≥1 class IV-V mutation(s), n (%) | 26 (15.7)            | 29 (17.1)            | 36 (19.5)            | 34 (18.3)            |
| Homozygous class I-III, n (%) | 129 (77.7)           | 128 (75.3)           | 133 (71.9)           | 137 (73.7)           |
| Pancreatic insufficient, n (%) | 137 (82.5)           | 135 (79.4)           | 142 (76.8)           | 145 (78.0)           |
| CF related diabetes, n (%) | 39 (23.5)             | 42 (24.7)             | 42 (22.7)             | 54 (29.0)             |
| P. aeruginosa status: |                        |                        |                        |                        |
| No P. aeruginosa, n (%) | 60 (36.1)             | 57 (33.5)             | 74 (40.0)             | 78 (41.9)             |
| Intermittent P. aeruginosa, n (%) | 37 (22.3)            | 36 (21.2)            | 31 (16.8)            | 29 (15.6)            |
| Chronic P. aeruginosa, n (%) | 69 (41.6)             | 77 (45.3)             | 80 (43.2)             | 79 (42.5)             |
| BMI, median (IQR) | 22.3 (19.7 – 24.6)    | 22.7 (20.0 – 25.0)    | 23.0 (20.3 – 26.0)    | 23.2 (20.4 – 26.0)    |
| Best %FEV₁, median (IQR) | 78.7 (54.1 – 92.5)    | 76.6 (54.4 – 89.7)    | 77.8 (60.4 – 89.0)    | 78.5 (58.5 – 89.6)    |

1 Genotype status as defined by international consensus. Homozygous class I-III mutations indicate ‘severe genotype’.
2 Pancreatic insufficiency was diagnosed by the clinical team on the basis of ≥2 faecal pancreatic elastase levels <200µg/g stool and symptoms consistent with maldigestion and malabsorption, in accordance to the UK Cystic Fibrosis (CF) Trust guideline.
3 CF related diabetes was diagnosed by the clinical team on the basis of oral glucose tolerance test and continuous subcutaneous glucose monitoring results, in accordance to the UK CF Trust guideline.
4 “Pseudomonas aeruginosa status was determined according to the Leeds criteria.

Table 2. Discrepancies in year-to-year %FEV₁ variation with different methods of processing forced expiratory volume in one second (FEV₁) data.

| Methods of processing FEV₁, data:                                      | Change in %FEV₁, median (IQR) | Friedman test p-values |
|-----------------------------------------------------------------------|--------------------------------|------------------------|
| (1) %FEV₁, (calculated with Knudson equation) extracted from spirometer machines used for analysis | -2.0 (-6.0 to 1.0)               | 0.016                  |
| (2) FEV₁, volume (in L) extracted and height data were cleaned, then %FEV₁, calculated using Knudson equation | -2.0 (-5.0 to 1.0)               | 0.062                  |
| (3) FEV₁, volume (in L) extracted and height data were cleaned, then %FEV₁, calculated using GLI equation | -1.1 (-4.6 to 1.5)               | 0.135                  |
| (4) FEV₁, volume (in L) extracted and height data were cleaned, then %FEV₁, calculated using GLI equation, then change %FEV₁, adjusted for baseline %FEV₁, using ESCF reference values | 0.7 (-2.4 to 3.6)               | 0.016                  |

ESCF - Epidemiologic Study of cystic fibrosis

1 The vast majority of the %FEV₁ data were from spirometer machines at the Sheffield Adult cystic fibrosis (CF) centre, which were calculated with Knudson equation in whole numbers. Some %FEV₁ data were from spirometer machines at the Pulmonary Function Unit which operationalised the Knudson equation differently, by calculating age to one decimal place to determine the predicted FEV₁. These spirometer machines also provided %FEV₁ to two decimal places, but this was rounded to whole numbers for the purpose of analysis. These results were presented at the 2017 North American CF Conference and were published as an abstract in Pediatric Pulmonology.

2 FEV₁ volumes were available in litres to two decimal places from spirometer machines. Height data were also extracted to allow the calculation of predicted FEV₁. This led us to uncover the inconsistency recording of height, which affected 30–40% of the study subjects and would have introduced erroneous variability to the %FEV₁ because all equations for predicted %FEV₁ are dependent on height. Height data were cleaned to weed out error. Where there was uncertainty regarding the height, the higher value was used to obtain a conservative estimate of %FEV₁. To replicate calculation process of the spirometer machines at the Sheffield Adult CF centre, age was rounded down to a whole number and predicted FEV₁ in volume were calculated to two decimal places using Knudson equation. This was used to derive the %FEV₁, which was then rounded to whole numbers for the purpose of analysis.

3 %FEV₁, height data were extracted as above. %FEV₁ was calculated using the GLI equation using an Excel Macro available at the European Respiratory Society website.

4 %FEV₁ calculated using the GLI equation as described above, then adjusted for baseline %FEV₁ as described in the ‘Methods’ section. An adjusted %FEV₁, change of >0 meant the subject’s year-to-year fall in %FEV₁ was less than expected for his / her baseline %FEV₁, indicating better health outcomes.
GLI equation altered the magnitude of year-to-year variation in %FEV₁ although the trend of reduced year-to-year fall in %FEV₁ persisted (p=0.135). Adjustment for baseline %FEV₁ further increased the p-value (p=0.210).

Similar results were obtained when restricting the analyses to those aged ≥18 years (see Table 3). Bland-Altman analyses comparing year-to-year variation in %FEV₁ calculated from clean FEV₁ data using Knudson equation¹⁷ vs year-to-year variation in %FEV₁ calculated from clean FEV₁ data using GLI equation¹⁴ indicate the tendency for Knudson equation¹⁷ to over-estimate the magnitude of year-to-year fall in %FEV₁ by a mean difference of 0.1–0.4% (see Figure 1).

Discussion
We demonstrated that different centre-level year-to-year variation in %FEV₁ results were obtained using different FEV₁ data processing methods. In particular, year-to-year fall in %FEV₁ was smaller in magnitude when %FEV₁ was calculated using GLI equation¹⁴ instead of Knudson equation¹⁷. This is in part due to the demographic of our centre which has a relatively young adult population. A previous study found a near-linear %FEV₁ decline from childhood to adulthood with GLI equation, whereas there was accelerated %FEV₁ decline from childhood to adulthood with Knudson equation¹⁷. One advantage of using the GLI equation, which is seamless across all ages, is that it improves the interpretation of %FEV₁ decline¹⁴. Another advantage is that %FEV₁ decline can be adjusted for baseline %FEV₁ using ESCF reference values (since the ESCF values for %FEV₁ decline were calculated using the GLI equation¹⁷).

The limitation for all single-centre analysis is the potential lack of generalisability. Another limitation of our analysis is that the ESCF reference values used to adjust year-to-year variation in %FEV₁ were derived using a cohort from around 15 years ago¹⁸, and may not represent the current population. Our results nonetheless highlighted that year-to-year variation in %FEV₁ can be extremely sensitive to the FEV₁ data processing methods. This is one of the challenges of using year-to-year variation in %FEV₁ to infer quality of care. Another challenge is that %FEV₁ lacks sensitivity as an outcome measure. A recent sample size estimation using the UK CF registry data suggests that 273 adults per centre are needed to detect a 5% FEV₁ difference at the 95% significance level¹⁹. The sensitivity of measures used to detect variations in care quality is particularly pertinent to CF because a relatively small population is spread across many centres. Indeed, only 6/28 (21.4%) of all UK adult CF centres have ≥273 adults. That means process measures, e.g. medication adherence, is important to detect variations in quality of CF care. Mant & Hicks previous demonstrated that measuring processes of care proven in randomised controlled trials to reduce death allows detection of meaningful differences in care quality for myocardial infarction with just 75 cases, whereas 8179 cases would be needed if mortality was used as the quality indicator²⁰.

Given the limitations of FEV₁ as an outcome measure in CF, results of centre comparisons based on FEV₁ data should be carefully interpreted. Observational studies with year-to-year variation in %FEV₁ as an outcome measure should carefully consider and clearly specify the data processing methods used.

Table 3. Discrepancies in year-to-year %FEV₁ variation with different methods of processing forced expiratory volume in one second (FEV₁) data among adults aged ≥18 years.

| Methods of processing FEV₁ data: | Change in %FEV₁, median (IQR) | Friedman test p-values |
|---------------------------------|-------------------------------|------------------------|
| (1) %FEV₁ (calculated with Knudson equation) extracted from spirometer machines used for analysis | -2.0 (-6.0 to 1.0) | 0.016 |
| (2) FEV₁ volume (in L) extracted and height data were cleaned, then %FEV₁ calculated using Knudson equation | -2.0 (-5.0 to 1.0) | 0.029 |
| (3) FEV₁ volume (in L) extracted and height data were cleaned, then %FEV₁ calculated using GLI equation | -1.3 (-4.6 to 1.3) | 0.090 |
| (4) FEV₁ volume (in L) extracted and height data were cleaned, then %FEV₁ calculated using GLI equation, then change %FEV₁ adjusted for baseline %FEV₁ using ESCF reference values | 0.5 (-2.4 to 3.3) | 0.149 |
Ethical considerations
Regulatory approval for the analysis was obtained from NHS Health Research Authority (IRAS number 210313).

Data availability
Dataset 1: Sheffield forced expiratory volume in one second (FEV₁) data 10.5256/f1000research.14981.d20560

Competing interests
No competing interests were disclosed.

Grant information
The author(s) declared that no grants were involved in supporting this piece of work.

References
1. Elborn JS: Cystic fibrosis. Lancet. 2016; 388(10059): 2519–2531. PubMed Abstract | Publisher Full Text
2. Stevens DP, Marshall BC: A decade of healthcare improvement in cystic fibrosis: lessons for other chronic diseases. BMJ Qual Saf. 2014; 23 Suppl 1:
3. Boyle MP, Sabadosa KA, Quinton HB, et al.: Key findings of the US Cystic Fibrosis Foundation’s clinical practice benchmarking project. BMJ Qual Saf.
Benchmarking to improve the quality of cystic fibrosis care. Curr Opin Pulm Med. 2012; 18(6): 596–601.

Schechter MS: Benchmarking to improve the quality of cystic fibrosis care. Curr Opin Pulm Med. 2012; 18(6): 596–601. PubMed Abstract | Publisher Full Text

5. Stern M, Niemann N, Wiedernann B, et al.: Benchmarking improves quality in cystic fibrosis care: a pilot project involving 12 centres. Int J Qual Health Care. 2011; 23(3): 349–356. PubMed Abstract | Publisher Full Text

6. Liu TG, Adler FR, Fitzsimmons SC, et al.: Predictive 5-year survivorship model of cystic fibrosis. Am J Epidemiol. 2001; 153(4): 345–352. PubMed Abstract | Publisher Full Text | Free Full Text

7. Corey M, Edwards L, Levison H, et al.: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1(8476): 307–310. PubMed Abstract | Publisher Full Text

8. Konstan MW, VanDevanter DR, Sawicki GS, et al.: Association of High-Dose Ibuprofen Use, Lung Function Decline, and Long-Term Survival in Children with Cystic Fibrosis. Ann Am Thorac Soc. 2018; 15(4): 485–493. PubMed Abstract | Publisher Full Text

9. Wagener JS, Elkin EP, Pasta DJ, et al.: Pulmonary function outcomes for assessing cystic fibrosis care. J Cyst Fibros. 2011; 10(2): 153–160. PubMed Abstract | Publisher Full Text | Free Full Text

10. Szczesniak R, Heltshe SL, Stanojevic S, et al.: Use of FEV1 in cystic fibrosis epidemiologic studies and clinical trials: A statistical perspective for the clinical researcher. J Cyst Fibros. 2017; 16(3): 318–326. PubMed Abstract | Publisher Full Text | Free Full Text

11. Hoo ZH, Curley R, Walters SJ, et al.: Real world evidence of sustained improvement in objective adherence to maintenance inhaled therapies in an adult cystic fibrosis centre [abstract]. Pediatr Pulmonol. 2017; 52(Suppl 47): S488. Reference Source

12. Inci I, Stanimirov O, Bendin C, et al.: Lung transplantation for cystic fibrosis: a single center experience of 100 consecutive cases. Eur J Cardiothorac Surg. 2012; 41(2): 435–440. PubMed Abstract | Publisher Full Text

13. Lynch JP 3rd, Sayah DM, Belpiero JA, et al.: Lung transplantation for cystic fibrosis: results, indications, complications, and controversies. Semin Respir Crit Care Med. 2015; 36(2): 299–320. PubMed Abstract | Publisher Full Text | Free Full Text

14. Ramsey BW, Davies J, McElroy NG, et al.: A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. N Engl J Med. 2011; 365(18): 1663–1672. PubMed Abstract | Publisher Full Text | Free Full Text

15. Gilbert EH, Lowerstein SR, Kodzi-McLain J, et al.: Chart reviews in emergency medicine research: Where are the methods? Ann Emerg Med. 1996; 27(3): 305–308. PubMed Abstract | Publisher Full Text

16. Knudson RJ, Lebowitz MD, Holberg CJ, et al.: Changes in the normal maximal expiratory flow-volume curve with growth and aging. Am Rev Respir Dis. 1983; 127(6): 725–734. PubMed Abstract

17. 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(6): 1324–1343. PubMed Abstract | Publisher Full Text | Free Full Text

18. Knudson RJ, Lebowitz MD, Holberg CJ, et al.: Changes in the normal maximal expiratory flow-volume curve with growth and aging. Am Rev Respir Dis. 1983; 127(6): 725–734. PubMed Abstract
Open Peer Review

Current Referee Status: ✅ ✅ ✗

Clive Osmond
MRC LifeCourse Epidemiology Unit, University of Southampton, Southampton, UK

Thank you for sending me this interesting note. A few thoughts on the analysis from a statistician

1. It’s an interesting, though sobering, fact that between 30 and 40 percent of the machine-entered heights are incorrect. Normally the tendency would be for such errors to obscure, rather than generate, associations. This now-known, high error rate makes it less interesting to explore this section of the results.

2. What does a Friedman test measure? It’s a non-parametric version of a repeated measures one-way analysis of variance. Two issues are worth considering. It requires a complete table, so only those subjects with all four years of data may be included. Secondly it produces a three degree of freedom test, which is not very well directed to address the most likely question of interest. We might be most interested in detecting a smooth, linear trend over time. However just as much weight is being given to detect non-linear patterns such as curvature {low, high, high, low} and saw-tooth {low, high, low, high}. I don’t know how centres are compared officially. Comparison of neighbouring years’ data would be unstable. Also, using these non-linear components could be very misleading. I hope that the linear trend is used.

3. Are there any alternative analyses that would address these two issues? Certainly. To begin with, let’s use the original data for FEV1% and not worry about their normality. You could fit a mixed model to these data once you stack them in long format (“varstocases” in SPSS). This would enable you to use all data, not just data for those with a complete set. It would also enable you to extract a one degree of freedom test for trend across the four years. This should be a more powerful approach. I now see that the other referees refer to this as well, though I don’t agree that you need to have at least three observations per subject.

4. If you are worried about the normality (though the published quartiles are not that alarming) then two alternatives would be (1) to find a normalising transformation that would apply to the stacked column of FEV1% values, or (2) to use a rank-based transformation (“Fisher-Yates”) available in SPSS as “rank y/normal into z.”

5. What might be the mathematics underlying any difference in slope obtained by the Knudson and GLI methods? I have tried to abstract the formulae used by Knudson and by GLIn deriving the predicted FEV1 that is used in the calculation of FEV1%. For a specific example I have chosen males (slightly more common in this study) aged 25 to 28 years (somewhere near the median age) with height of 175cm (just below mean UK adult height).

The Knudson equation has a functional form
FEV1 predicted = 5.1228 – 0.0292.age.
FEV1% = FEV1/(FEV1 predicted) can then be differentiated to see how it varies with changes in age and FEV1
However the GLI equivalent is given as point estimates from a Cole-Green LMS fitting procedure. The penalised cubic splines are not given, so no functional form is available.

The table shows, just for this combination, how the predicted values compare. Those from Knudson are slightly lower and decrease slightly more rapidly with age. Such differences, and those from other combinations, will work together to determine how FEV1% might be expected to change with age and observed FEV.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Referee Expertise: Stats

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Referee Report 31 October 2018
https://doi.org/10.5256/f1000research.17436.r37306

Pierre-Régis Burgel
Pulmonary Department and Adult CF CentreGroupe Hospitalier Cochin-Hotel Dieu, Paris Descartes University, Paris, France

No further comment

Competing Interests: No competing interests were disclosed.
Referee Expertise: Adult pulmonologist with experience in the care of adults with cystic fibrosis. Researcher.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 17 July 2018
https://doi.org/10.5256/f1000research.16309.r34826

Pierre-Régis Burgel
Pulmonary Department and Adult CF CentreGroupe Hospitalier Cochin-Hotel Dieu, Paris Descartes University, Paris, France

The authors performed a retrospective analysis of FEV1% predicted data over 3 years in an adult CF center in the UK. They examined FEV1 decline from year to year by calculating variation in best FEV1 during two consecutive years and examined the impact of using data obtained using Knudson equation (directly extracted from the spirometer or recalculated with the appropriate height) vs. GLI equation. They also performed an adjustment using ESCF data.

The authors concluded that trends in FEV1 decline were robust among methods, although the results were somewhat different using different methods/equations.

The study has some interest in highlighting problems associated with these type of calculations, especially when used for benchmarking (as in the UK).

I have the following comments for improvement:

1. An important drawback of Knusdon equation is related to the change of FEV1 in the transition from pediatric to adults. This is why the GLI is nowadays often used in mixed pediatric/adult population. The authors used the UK definition of adults (over 16 years) and suggested that some of the difference in their results between Knudson and GLI data are due to the younger patients in this cohorts. I would be happier if the authors could perform a sensitivity analysis using only patients 18 years an over? This would miniminze the Knusdon/GLI age bias and would make these results more relevant to the adult centres outside of UK. Looking at Table 1, it seems that only a minority of patients were below 18 years.

2. I think the word FEV1 decline is inappropriate in this manuscript. A year to year variation (even over 3 years) is not a decline. For calculating a decline, you would need multiple data points (at the very least 3 data points) and perform more complicated analyses (e.g., mixed model analysis). I would suggest to remove the word decline from the manuscript as the main goal of the authors did not appear to be FEV1 decline but mostly year to year FEV1 variation which is used for benchmarking in the UK.

Is the work clearly and accurately presented and does it cite the current literature?
Yes
Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
No source data required

Are the conclusions drawn adequately supported by the results?
Partly

**Competing Interests:** No competing interests were disclosed.

**Referee Expertise:** Adult pulmonologist with experience in the care of adults with cystic fibrosis. Researcher.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

---

**Author Response 12 Aug 2018**

**Zhe Hui Hoo, University of Sheffield, UK**

We thank Prof Burgel for the review and we will iterate the manuscript taking into account the two very useful suggestions, i.e.
1. we will perform a sensitivity analysis for the results in Table 2 using only adults aged 18 years and above
2. we will replace the term "FEV1 decline" with "year-to-year FEV1 variation"

**Competing Interests:** No competing interests were disclosed

---

**Referee Report 06 July 2018**

https://doi.org/10.5256/f1000research.16309.r34828

**Edward McKone**

Department of Respiratory Medicine, St Vincent’s Hospital, Dublin, Ireland

FEV1 as a percent of predicted is widely used as an outcome measure in patients with cystic fibrosis and is one of the metrics used to compare centres or countries in benchmarking exercises. This manuscript presents data showing that differences in data processing and the use of different reference equations used to estimate FEV1 as a percent predicted can result varying estimates of lung disease changes and potentially impact comparisons of centres/countries.
The paper supports the standardization of FEV1 collection and reference equations which is currently in development by CF International Registries. It also highlights that different approaches to data collection can impact the interpretation of statistical analyses.

Comments:

Differences in FEV1 percent predicted using different equations is well known (Rosenfeld et al \(^1\) and more recently in the cited UK/US comparison study). For this reason, the GLI have been recently accepted as the standard for most CF registries.

Although year to year subtraction is a method of looking at longitudinal changes, regression methodology is preferable to analyse these changes, especially, as in this case, where you have 3 time points. This also allows to adjust for baseline factors such as lung disease severity.

The method of adjustment for baseline lung function is a bit crude. The medians subtracted are from a US population over 10 years ago and are likely to overestimate lung function decline in this population. In the Morgan et al, J Pediatr 2016 paper cited, the benefits of using this type of adjustment was shown using regression.

Did their statistical approach factor in that these were repeated measures in the same patients?

Bland & Altman plots comparing different reference equations could be considered.

The results suggest that height inaccuracy is impacting the results. As this is a single centre study, it is difficult to determine is this is a more universal problem.

References
1. Rosenfeld M, Pepe MS, Longton G, Emerson J, FitzSimmons S, Morgan W: Effect of choice of reference equation on analysis of pulmonary function in cystic fibrosis patients. *Pediatr Pulmonol*. 2001; 31 (3): 227-37 PubMed Abstract

---

Is the work clearly and accurately presented and does it cite the current literature?  
Yes

Is the study design appropriate and is the work technically sound?  
Partly

Are sufficient details of methods and analysis provided to allow replication by others?  
Yes

If applicable, is the statistical analysis and its interpretation appropriate?  
Partly

Are all the source data underlying the results available to ensure full reproducibility?  
Yes

Are the conclusions drawn adequately supported by the results?
Yes

**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

**Author Response 12 Aug 2018**  
**Zhe Hui Hoo, University of Sheffield, UK**

We thank Prof McKone for the review and we will iterate the manuscript taking into account the suggestion to compare the different reference equations (Knudson vs GLI) using Bland-Altman analysis.

We concur the GLI has been recently accepted as the standard for most CF registries.

We concur that regression analyses is preferable to determine FEV1 decline. As recommended by Prof Burgel, we will replace the term "FEV1 decline" with "year-to-year FEV1 variation" in the revised manuscript.

We concur that the method used to adjust year-to-year FEV1 variation for baseline FEV1 is crude. The displayed data from the ESCF paper is only presented according to the four FEV1 categories, hence our choice of adjustment method. Given the limited number of subjects within the Sheffield dataset, we felt is it is more appropriate to use reference values for suitably large datasets instead of simply calculating the predicted %FEV1 change using the Sheffield dataset. There are more recent reference values for FEV1 from the ECFSPR (Boëlle et al, 2012) and Canadian registry (Kim et al, 2018); however those papers do not provide reference values for year-to-year FEV1 variation.

Our statistical method account for repeated FEV1 measures since:
1. by using best FEV1, there is only x1 FEV1 reading per person per year
2. only x1 FEV1 reading per person was used to calculate the year-to-year FEV1 variation

As mentioned in the discussion section, we concur that a single-centre study may not be generalisable. However, inaccurate data recording within routine datasets (e.g. CF registries) is unlikely to be an isolated problem. For example, the letter by Hartley et al (2016) in JCF revealed that 6% of the adults with CF at the Manchester Adult CF Centre had incorrect genotype data recorded in the UK CF registry.

**Competing Interests:** No competing interests were disclosed
• Your article is published within days, with no editorial bias
• You can publish traditional articles, null/negative results, case reports, data notes and more
• The peer review process is transparent and collaborative
• Your article is indexed in PubMed after passing peer review
• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com