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; peer review: 2 approved, 1 approved with reservations]

Zhe Hui Hoo¹², Muhaned S.A. El-Gheryani¹², Rachael Curley¹², Martin J. Wildman¹²

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 \%FEV\textsubscript{1} in the GLI equation. A statistically significant p-value (0.016) was only obtained when using \%FEV\textsubscript{1} in Knudson equation extracted directly from spirometer machines.

**Conclusions:** Although the trend of reduced year-to-year fall in \%FEV\textsubscript{1} was robust, different data processing methods yielded varying results when year-to-year variation in \%FEV\textsubscript{1} was compared using a standard related group non-parametric statistical test. Observational studies with year-to-year variation in \%FEV\textsubscript{1} 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
**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 is scant attention paid to the methods of processing FEV₁ data. We previously reported a statistically significant reduction in year-to-year %ΔFEV₁ 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 %ΔFEV₁, thus may affect the interpretation of year-to-year variation in %ΔFEV₁.

Demographic data (age, gender, genotype, pancreatic status, CF related diabetes, Pseudomonas aeruginosa status), body mass index (BMI) and FEV₁ data were collected by two investigators (HZH and RC) 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, %ΔFEV₁ 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 %ΔFEV₁ (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 %ΔFEV₁ with GLI equation using an Excel Macro (Microsoft Excel 2013).

Best %ΔFEV₁, i.e. the highest %ΔFEV₁ reading in a calendar year for each study subject was used for analysis since it is most reflective of the true baseline %ΔFEV₁. Year-to-year %ΔFEV₁ change was calculated by subtracting best %ΔFEV₁ at Year 1 from Year 2 (i.e. negative values indicate fall in %ΔFEV₁ and positive values indicate increase in %ΔFEV₁). In addition to calculating year-to-year %ΔFEV₁ change using three different FEV₁ data processing methods, %ΔFEV₁ change calculated with GLI equation was also adjusted for baseline %ΔFEV₁ using reference values from Epidemiologic Study of CF (ESCF). The ESCF study found median %ΔFEV₁ change of −3%/year, −2%/year and −0.5%/year for baseline %ΔFEV₁ ≥100%, 40–99.9% and <40% respectively. Adjusted %ΔFEV₁ change was calculated by subtracting median ESCF %ΔFEV₁ change from actual %ΔFEV₁ change. Thus, an adjusted %ΔFEV₁ change >0 meant the subject’s year-to-year change in %ΔFEV₁ was less than expected (indicating better health outcome) whilst an adjusted %ΔFEV₁ change <0 meant the subject’s year-to-year change in %ΔFEV₁ was more than expected (indicating worse health outcome).

%ΔFEV₁ 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 compare 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 %ΔFEV₁ increased from 2014 onwards (see Table 1).

The %ΔFEV₁ increase was in part due to younger adults with higher %ΔFEV₁ transitioning from paediatric care because %ΔFEV₁ tended to decline from year to year (see Table 2). However, different year-to-year change in %ΔFEV₁ results were obtained with different FEV₁ data processing methods. There was statistically significant reduction in year-to-year fall in %ΔFEV₁ using %ΔFEV₁ readings as recorded in spirometer machines (p=0.016). Cleaning of height data and standardisation of %ΔFEV₁ calculation with Knudson equation did not alter the magnitude of year-to-year variation in %ΔFEV₁, but the p-value was no longer statistically significant (p=0.062). The use of
Table 1. Characteristics of study subjects from 2013 to 2016.

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

† Genotype status as defined by international consensus. Homozygous class I-III mutations indicate ‘severe genotype’.
‡ 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.
§ 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.
¶ 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.

<table>
<thead>
<tr>
<th>Methods of processing FEV1 data:</th>
<th>Change in %FEV1, median (IQR)</th>
<th>Friedman test p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) %FEV1, calculated with Knudson equation from spirometer machines used for analysis §</td>
<td>-2.0 (–6.0 to 1.0)</td>
<td>-1.0 (–3.3 to 2.0)</td>
</tr>
<tr>
<td>(2) FEV1, volume (in L) extracted and height data were cleaned, then %FEV1, calculated using Knudson equation §</td>
<td>-2.0 (–5.0 to 1.0)</td>
<td>-1.0 (–4.0 to 1.0)</td>
</tr>
<tr>
<td>(3) FEV1, volume (in L) extracted and height data were cleaned, then %FEV1, calculated using GLI equation ¶</td>
<td>-1.1 (–4.6 to 1.5)</td>
<td>-0.9 (–3.2 to 1.5)</td>
</tr>
<tr>
<td>(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 †</td>
<td>0.7 (–2.4 to 3.6)</td>
<td>1.1 (–1.4 to 3.5)</td>
</tr>
</tbody>
</table>

ESCF - Epidemiologic Study of cystic fibrosis

† 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.$
¢ FEV1, volume 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, 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.

¶ 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.

†§ %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 \%FEV\textsubscript{i} although the trend of reduced year-to-year fall in \%FEV\textsubscript{i} persisted (p=0.135). Adjustment for baseline \%FEV\textsubscript{i} 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 \%FEV\textsubscript{i} calculated from clean FEV\textsubscript{i} data using Knudson equation\textsuperscript{17} vs year-to-year variation in \%FEV\textsubscript{i} calculated from clean FEV\textsubscript{i} data using GLI equation\textsuperscript{14} indicate the tendency for Knudson equation\textsuperscript{7} to over-estimate the magnitude of year-to-year fall in \%FEV\textsubscript{i} by a mean difference of 0.1–0.4% (see Figure 1).

**Dataset 1. Sheffield forced expiratory volume in one second (FEV\textsubscript{1}) data**

http://dx.doi.org/10.5256/f1000research.14981.d205603

**Discussion**

We demonstrated that different centre-level year-to-year variation in \%FEV\textsubscript{i} results were obtained using different FEV\textsubscript{i} data processing methods. In particular, year-to-year fall in \%FEV\textsubscript{i} was smaller in magnitude when \%FEV\textsubscript{i} was calculated using GLI equation\textsuperscript{14} instead of Knudson equation\textsuperscript{7}. 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 \%FEV\textsubscript{i} decline from childhood to adulthood with GLI equation, whereas there was accelerated \%FEV\textsubscript{i} decline during adolescence and young adulthood when \%FEV\textsubscript{i} was calculated with Knudson equation\textsuperscript{7}. One advantage of using the GLI equation, which is seamless across all ages, is that it improves the interpretation of \%FEV\textsubscript{i} decline\textsuperscript{14,20}. Another advantage is that \%FEV\textsubscript{i} decline can be adjusted for baseline \%FEV\textsubscript{i} using ESCF reference values (since the ESCF values for \%FEV\textsubscript{i} decline were calculated using the GLI equation\textsuperscript{14}).

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 \%FEV\textsubscript{i} were derived using a cohort from around 15 years ago\textsuperscript{20}, and may not represent the current population. Our results nonetheless highlighted that year-to-year variation in \%FEV\textsubscript{i} can be extremely sensitive to the FEV\textsubscript{i} data processing methods. This is one of the challenges of using year-to-year variation in \%FEV\textsubscript{i} to infer quality of care. Another challenge is that \%FEV\textsubscript{i} 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% FEV\textsubscript{i} difference at the 95% significance level\textsuperscript{19}. 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 indicator\textsuperscript{27}.

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

**Table 3. Discrepancies in year-to-year \%FEV\textsubscript{i} variation with different methods of processing forced expiratory volume in one second (FEV\textsubscript{i}) data among adults aged ≥18 years.**

<table>
<thead>
<tr>
<th>Methods of processing FEV\textsubscript{i} data:</th>
<th>Change in %FEV\textsubscript{i}, median (IQR)</th>
<th>Friedman test p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) %FEV\textsubscript{i} (calculated with Knudson equation) extracted from spirometer machines used for analysis</td>
<td>-2.0 (-6.0 to 1.0)</td>
<td>-1.0 (-3.0 to 2.0)</td>
</tr>
<tr>
<td>(2) FEV\textsubscript{i} volume (in L) extracted and height data were cleaned, then %FEV\textsubscript{i} calculated using Knudson equation</td>
<td>-2.0 (-5.0 to 1.0)</td>
<td>-1.0 (-4.0 to 1.0)</td>
</tr>
<tr>
<td>(3) FEV\textsubscript{i} volume (in L) extracted and height data were cleaned, then %FEV\textsubscript{i} calculated using GLI equation</td>
<td>-1.3 (-4.6 to 1.3)</td>
<td>-1.0 (-3.2 to 1.4)</td>
</tr>
<tr>
<td>(4) FEV\textsubscript{i} volume (in L) extracted and height data were cleaned, then %FEV\textsubscript{i} calculated using GLI equation, then change %FEV\textsubscript{i} adjusted for baseline %FEV\textsubscript{i} using ESCF reference values</td>
<td>0.5 (-2.4 to 3.3)</td>
<td>1.0 (-1.4 to 3.4)</td>
</tr>
</tbody>
</table>
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


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text | Free Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract


Published Abstract | Publisher Full Text | Free Full Text


Published Abstract


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text | Free Full Text


Published Abstract | Publisher Full Text | Free Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text


Published Abstract | Publisher Full Text | Free Full Text


Published Abstract | Publisher Full Text | Free Full Text


http://www.doi.org/10.5256/f1000research.14981.d20560
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 GLI in deriving the predicted FEV1 that is used in the calculation of FEV1%.
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.

*Reviewer Expertise:* Stats

I confirm that I have read this submission and 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.

Reviewer Report 31 October 2018

https://doi.org/10.5256/f1000research.17436.r37306
Pierre-Régis Burgel
Pulmonary Department and Adult CF Centre
Groupe Hospitalier Cochin-Hotel Dieu, Paris Descartes University, Paris, France

No further comment

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

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?
Partly

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

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

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

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

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
Pierre-Régis Burgel
Pulmonary Department and Adult CF Centre
Groupe 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 and over? This would minimize 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.

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

I confirm that I have read this submission and 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, 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

Reviewer Report 06 July 2018

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

© 2018 McKone E. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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\textsuperscript{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**


**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 confirm that I have read this submission and 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, 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 (Knudsen 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 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
The benefits of publishing with F1000Research:

- 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