Author’s response to reviews


Authors:
Tafadzwa Madanhire (tafmadanhire@gmail.com)
Rashida Ferrand (ashida.Ferrand@lshtm.ac.uk)
Engi Attia (eattia@uw.edu)
Elopy Sibanda (ensibanda@gmail.com)
Simbarashe Rusakaniko (srusakaniko@gmail.com)
Andrea Rehman (Andrea.Rehman@lshtm.ac.uk)

Version: 1 Date: 15 Nov 2019

Author’s response to reviews:

University of Zimbabwe
College of Health Sciences
Parirenyatwa Group of Hospitals
Department of Community Medicine
New Health Science Building
3rd Floor
P.O.Box A178
Avondale, Harare

Dear Dr Alessandro Marcon
Editor, BMC Pulmonary Medicine
November 11, 2019

Dear Dr Marcon


Thank you for considering our manuscript for publication in the BMC Pulmonary Medicine. We extend our gratitude to the reviewers for their detailed and thoughtful comments to improve the manuscript. Please find below our responses to each of the comments.

Kind Regards

Tafadzwa Madanhire
Editor’s comments

Results, 2nd page. Omit "lower" from "lower 5th (per) centile": the "upper" one is called 95th percentile.
We have omitted the term lower to describe the Lower Limit to Normal for FEV1 and FVC. (Line 72,154)

Discussion, 1st page. Please revise "These are standardised values that reduce bias". It is not clear.
We have clarified the statement by referencing it to the ERS/GLI2012 predicted values. (Line 242-244)
“The GLI2012 SRE predict standardised z-score values that reduce bias, accounting for differences between ethnicity, anthropometric variables and lung function parameters.”

Could a selection of participants explain the statistically significant difference in age distribution between boys and girls? (Table 1) Please report descriptive statistics comparing children included vs children excluded.

Even though the difference was “statistically significant” this has arisen because of the large sample size (712 participants) which can detect a difference in mean age between boys (10.7 years; SD: 1.86) and girls (10.4 years; SD: 1.74) of 4.8 months with power of 0.839. The difference observed was only four months, which is not large on an absolute scale given the range of ages in the sample.

A description of the participants who were excluded from the study has been included and a comparative table inserted in the supplementary section as supplementary file 3. (Lines 181-184)

“On average, participants who were excluded from the study were older (11.6 years, SD: 1.45), than those considered for analysis. The ratios of girls to boys in the included and excluded study groups were different, with more girls [37(64.9%)] being excluded from the study. The mean BMI z-scores for excluded and included participants were -0.28(1.81) and 0.07(0.9) respectively.” (Supplementary file 1)

Table 3. Pearson's correlation coefficient for sex. Pearson's r cannot be used with a binary variable.
We thank the editor for the observation and we have revised the table and removed the variable sex in Table 3.

Figure 2 and 3, and supplementary Figures. Within a Figure, y- and x-axis should show the same range across different panels. Otherwise a visual comparison across Figure panels would be deceptive.
We have reviewed Figure 2,3 and supplementary figures to include the same range of values in the x- and y- axis. We would like to thank the editor for identifying this, since it will aid in comparing graphs.

Reviewer reports:
Sanja Stanojevic (Reviewer 1):

1. The GLI are not ERS equations per se, it would be more appropriate to just use GLI2012 when describing
We express our gratitude for the suggestion and have revised the whole document to specify the equations as GLI2012 as suggested by the reviewer.
2. In some cases the text says urban in other peri-urban please specify and use consistently.
We have revised the document to highlight consistently the study site as urban and peri-urban Zimbabwe.

3. In the abstract - please specify that z-scores using the African-American GLI equations are summarized in the results.
We thank the reviewer for the suggestion and we have included the comment to clearly specify how the spirometry z-scores were generated which reads:
“Spirometry z-scores were generated from African-American GLI2012 SRE which adjust for age, sex, ethnicity and height, and anthropometry z-scores were generated using the British (1990) reference equations which adjust for age and sex.”

4. Table 1 - it would be better to have the total column first. Which of the GLI equations were used to calculate the z-scores. Please specify Table 2 - instead of 'standard normal GLI' it would be more appropriate to say 'Caucasian GLI'.
We have restructured Table 1 to have the total column first and highlighted that the African American GLI2012 equations were used to generate spirometry z-scores.
We also renamed the Caucasian GLI as 'standard normal GLI' as advised by the reviewer.

5. Table 3 - I am not sure this table is very useful without accompanying confidence intervals (adding these correlation coefficients to the figures might be better).
We have included the 95% confidence intervals to the Pearson’s correlation coefficients between the spirometry z-scores to show the significance of these estimates. We have also added the correlation coefficients to Figure 2 and 3 as suggested by the reviewer. (Table 3, Figure 2 and Figure 3)

6. Figure 2/3 - adding the +1.645 line would be helpful to understand the upper and lower bounds.
We thank the reviewer for this suggestion and have included the +1.645 line in all plots from Figure 2 to help understand the distribution of the spirometry z-scores.

7. In the methods, the statement that z-scores are calculated as observed-expected/standard deviation is not incorrect, but with the GLI equation the calculation of z-scores is a bit more complicated. It might be simpler to state that GLI z-scores and LLN were calculated using the available calculators, which provide an age, height, sex and ethnic specific value.
We have revised the statement as the reviewer suggested to explain how the GLI z-scores and LLN were calculated using excel macro-calculators and now reads :(Line 145-147)
“The z-score and LLN values were calculated using the available Microsoft-Excel Macro calculators, which provide an age, height, sex and ethnic specific value.”

8. In the methods it states that the ECCS/Polgar equations were used, but the ECCS are only available for those &gt;20 years, so I assume that only the Polgar equations are compared. Please include a citation for these equations. The statement in the discussion regarding the Polgar reference is not correct. The ECCS were developed in Coal workers, but the paediatric equations by Polgar were not, there equations combined multiple published equations and have their own set of limitations.
Results:

9. The gender differences in z-scores require further investigation. The calculation of z-scores should adjust for the sex-related differences in lung development, therefore the observed differences after adjustment suggest that the two groups are different due to other factors. We thank the reviewer for the comment. We revisited the comparisons of z-scores between girls and boys and found no differences by sex except for MMEF. Table 1.

10. Since z-scores were calculated for all of the GLI equations, it is important to highlight which results are described in the text. I assume after the comparisons are done in Table 2, that the rest of the results are for African-American equations but this is not clear. We like to thank the reviewer for the contribution and we have captured the suggestion to address the z-scores as generated from the African-American module. We have therefore renamed the z-scores equations as “African-American GLI2012 spirometric reference equations in the manuscript.

11. Reporting the % of subjects with values below the LLN would be more intuitive, as we would expect 5% to be below this value. It does look like for FEV1, a greater proportion than 5% have values below the LLN which will influence interpretation.

We have revised the explanations in relation to the percentage of participants below the LLN and also included the explanation on the implication of the observed percentages as below:
“The distribution of spirometry z-scores in relation to the 5th percentile (LLN) identified that for FEV1, 8.7% (7.9% of boys, 9.6% of girls) and for FVC, 5.8% (4.1% among boys, 7.6% among girls) had values below the LLN. However, the FEV1/FVC z-scores showed a different pattern with 18.4% (18.2% of boys, 18.6% of girls) of participants having values below the LLN indicating a deviation from the GLI2012 distribution.”

12. The comparison of GLI to Polgar should include a Bland-Altman plot.
We thank the reviewer for the suggestion and we have included the Bland-Altman plot as supplementary file 4. The Bland-Altman plots compares the performance of the GLI2012 and Polgar SRE in this sample for FVC, FEV1, FEV1/FVC and MMEF.
We have also included the description of the plots in the results section as:
“A Bland-Altman plot for the spirometry indices showed mean differences between the GLI2012 and Polgar SRE and evidence of proportional bias as the difference of GLI2012 and Polgar predicted values increased with the mean values of the two SREs. (Supplementary file 4: Regression coefficients).”
13. A multivariable regression to adjust for ses, school etc. might be more informative than the supplemental material. In the current version it is not clear whether univariate or multivariable regression results are presented. Is there evidence that some of the offset observed is explained by SES? A multivariable model might help explain this.
We would want to thank the reviewer for the suggestion to include regression analysis to help identify if the socio-economic groups are different from each other. We have included a multivariate regression analysis for FEV1, FVC and FEV1/FVC z-scores in supplementary 3. We reported a Wald p-value for the homogeneity of the coefficients of SES in the multivariate model.

14. In supplemental file 2 there seems to be a bias with z-scores and age in the high-income group. Additional data, including the average z-score (SD) and % below the LLN in each of these sub-groups (in a Table) would be informative. For direct comparison, the scales (y-axis in particular) should be the same for each figure. In the high income group, there is also a bias with zheight and FEVFVC

We have restructured the figures in supplementary file 3 to so that the scales are the same. We also included average z-score (SD) and % below the LLN for all the spirometric indices in the three categories as advised by the reviewer.

Discussion:

15. The dysnaptic growth differences during puberty are unlikely to have occurred in children between the ages of 7 and 13.. possible if puberty was shifted to younger ages, but then there would have been a clear age bias observed (which does not seem to be the case - Figure 2).
We note the comment by the reviewer and have adjusted the interpretation on the increased variation for FEV1/FVC in this population. We removed the suggestion that the variation of the ratio was partly due to changes that happen in puberty. The statement reads:
“However, the z-score standard deviations for the FEV1/FVC ratio were ≥ 1, indicating more variability than the reference population hence affecting the performance of the African-American GLI2012 LLN in this population. (Line 246-248)”

16. Avoid terms like diagnosis, since a diagnosis requires more than PFT results. Misclassification would be more appropriate.
We would like to thank the reviewer for the suggestion which improves the manuscript and have revised the term and corrected the document as suggested to read:
“By definition the LLN allows 5% of healthy people to be misclassified and higher variability in FEV1/FVC may increase misclassification of airway obstruction.”

17. Discussion, spirometry on its own cannot provide evidence of restrictive lung patterns, please consider re-wording.
We have revised the discussion section and reworded as suggested by the reviewer. We thank the reviewer for the suggestion and now reads:
“Our results suggest that the use of African-American GLI2012 SRE in Zimbabwean children can improve identification of a tendency towards a restrictive and obstructive lung function pattern.”
18. Discussion of the lower on average z-scores and what this means for use in clinical practice is important and currently not addressed in the manuscript. While overall the differences observed may simply be due to sampling bias, there is a chance that the LLN may over-classify children with low lung function. A comment about particular attention to interpretation of results around the LLN is needed to avoid over-interpretation of results in practice.

We thank the reviewer for this contribution and we have tried to address the implications of the negative spirometry z-scores in the discussion section as highlighted below:

“The negative mean spirometry z-scores for all the indices implies the LLN should be cautiously interpreted by practitioners, to avoid over-classifying children with low lung function.”

19. The conclusions generalizes the results to all children in Zimbabwe. The authors may want to add a disclaimer that this assumes that the schools in this study are representative of all children.

We have included a disclaimer in the conclusions stated as:

“In conclusion, the African-American GLI2012 SRE are appropriate for predicting lung function in Zimbabwean school-going urban and peri-urban children aged 7-13 years”

Helmi Ben Saad (Reviewer 2):

1. General remarks

*Avoid, over all the paper, the active voice and the use of the first person. Apply for the passive voice.

We have revised the manuscript and used the passive voice as the reviewer suggested.

*Be careful for abbreviations. Replace FEF25-75 by MMEF (maximal-mid expiratory flow).

We thank the reviewer for the contribution and we have replaced the FEF25-75 with MMEF in the main text, Figures and, Supplementary files.

*In the abstract and in the main manuscript: Replace reference values, reference equations by SRE (for spirometric reference equation(s)). In other words, homogenize your terms and avoid elegant variation of terms.

We have replaced the term reference values and reference equations with spirometric reference equations (SRE) in the manuscript as the reviewer suggested.

*Over all the paper, replace "restriction" by "tendency through restriction". In fact, the diagnosis of a restrictive ventilatory defect is based on a TLC < LLN (ATS/ERS-2005). The combination of a FEV1/FVC < LLN and a FVC < LLN oriented towards a tendency through a restrictive defect.

We have revised the manuscript and replaced restriction with “tendency through restriction” as suggested by the reviewer to read:

“The use of the African-American GLI2012 SRE to improve identification of a tendency towards restrictive and obstructive lung function patterns in this population is recommended.”

*Avoid redundancy between text and tables (results section).

We appreciate the comment from the reviewer and have revised the sections that follows manuscript tables to reduce redundancy.
2. Title

In order to shorten the abstract, authors are asked to introduce in the title three abbreviations (ERS, GLI2012 and SRE); The new title will be:


We would like to thank the reviewer for the contribution. However, we feel abbreviations should not be included in the title but rather starting from the abstract if used more than once. We have only corrected the title by removing “European Respiratory Society”, and now reads:


3. Abstract

*Background. Change as: The ERS/GLI2012 provide multi-ethnic SRE for the 3-95-year-old age range, but Sub-Saharan African populations are not represented. This study aimed to evaluate the fit of the ERS/GLI2012 ERS to a population of healthy urban Zimbabwean school children (7-13 years).

We have changed the sentences in abstract background as the reviewer has suggested. We however excluded the abbreviation ERS as shown below: (Line 22-25)

“The 2012 Global Lung Function Initiative GLI2012 provide multi-ethnic SRE for the 3-95 year-old age range, but Sub-Saharan African populations are not represented. This study aimed to evaluate the fit of the African-American GLI2012 SRE to a population of healthy urban Zimbabwean school children (7-13 years).”

*Methods: Authors are asked to write a clear hypothesis in the method section. See the abstract of your reference N°11 and the abstract of the following African study: Ketfi A, et al. The multi-ethnic global lung initiative 2012 (GLI-2012) norms reflect contemporary adult's Algerian spirometry. PLoS One. 2018 Sep 4;13(9):e0203023. For example, state that: If the average Z-score deviated by "&lt; ± 0.5" from the overall mean, the ERS/GLI2012 norms would be considered as reflective of contemporary Zimbabwean children spirometry.

We would like to thank the reviewer for this contribution and we have included a hypothesis statement in the abstract as suggested by the reviewer. (Line 32-35)

“The African-American GLI2012 z-score distribution for the four spirometry measurements (FVC, FEV1, FEV1/FVC and MMEF) were evaluated against a standard normal distribution (with z-scores deviations within ± 0.5 reflecting Zimbabwean children) across age, height, BMI and school (for socioeconomic status) to assess for bias”

*Methods. The following sentence (L37-42) is a source of confusion: "We also compared the percent predicted values between the ERS/GLI2012 and Polgar reference values (which are in current clinical use in Zimbabwe)." Why? Because in your main manuscript (P14L14-17) you stated that "Most physicians in Zimbabwe use the Polgar reference values for diagnosis of lung disease, which were developed from Caucasian male workers in coal mines and steel industries in the 1970s." You insist on the term "male"? What about female, which SRE are applied in Zimbabwe? Please clarify or correct

We like to thank the reviewer for the comment and have revised on how the Polgar reference values were originally developed, that is, they were developed from North America, Europe and Japan 6-18 age populations. (Line 271-273),
“Therefore, we feel it would be informative to compare the performance of the Polgar reference equations (currently used in Zimbabwe) against the GLI2012 reference equations being proposed.”

*Results. L45 to 48, change as "The validation dataset contained acceptable spirometry data from 712 participants (344 girls, mean age:10.5 years (SD 1.81)).
We like to thank the reviewer for the contribution and have made the change as suggested and now reads: (Line 38, 39)
“The validation dataset contained acceptable spirometry data from 712 participants (344 girls, mean age: 10.5 years (SD 1.81)).”

*Conclusion. Avoid the active voice, avoid the term 'restricted"…
We have revised the active voice in the abstract conclusion and also replaced the term restriction by a tendency towards restriction to read as (Line 48-50)
“The use of the African-American GLI2012 SRE to improve identification of a tendency towards restrictive and obstructive lung function patterns in this population is recommended.”

4. Keywords
Use key words, not cited neither in the title nor in the abstract…
We thank the reviewer for the contribution and we have subsequently revised the terms cited under Key words section in the manuscript.

5. Introduction
*P6L10-12: add references after the following sentence: "Studies validating the ERS/GLI2012 reference equations have made varying conclusions, with some indicating a poor fit for local populations."
We have added references to the sentence as advised by the reviewer.

6. Methods
*P7L4-7: add an adequate reference for the following sentence "living at least three days per week with people smoking cigarettes)"
We thank the reviewer for the contribution and we have included citations on the effect of passive smoking on lung health and lung function indices. The study excluded participants with regular exposure to passive smoke because there is evidence of reduction in lung function indices among these participants.

*P7L35-40: the reference (ref 26) to argue the section related to spirometry is WRONG. Authors stated "ATS/ERS guidelines" and argues by a reference 26 related to how to write a spirometry report. The good reference is your reference number 28 (Miller. ERJ 2005). The same remark is valid for the sentence P7L53-55 (Participants ….analysis"…)
We thank the reviewer for the comment and have corrected the citation on the spirometry guidelines to Miller.ERJ 2005 as suggested.

*P7L52-57: the FVC and FEV1 reproducibility criteria should be reviewed. According to the ERS/ATS (ref 27): the between-manuever evaluation criteria are: an adequate test requires a minimum of three acceptable FVC maneuvers. Acceptable repeatability is achieved when the difference between the largest and the next largest FVC is ≤0.150 L and the difference between the largest and next largest
FEV1 is \( \leq 0.150 \) L. For those with an FVC of \( \leq 1.0 \) L, both these values are \( 0.100 \) L." Moreover, the ATS/ERS stated that "The largest FVC and the largest FEV1 (BTPS) should be recorded after examining the data from all of the usable curves, even if they do not come from the same curve." Please correct/clarify

We like to thank the reviewer for the comment and clarity of the FVC and FEV1 reproducibility criteria. We have therefore taken heed and clarified the steps taken to explain the methods used in ensuring that the participants’ efforts meet acceptable repeatability. (Line 128-129)

“The best effort of manoeuvres was defined as the largest sum of FVC and FEV1 within 0.15 litres (FVC \( \gt 1.0 \) litres) and 0.1 litres (FVC \( \leq 1.0 \) litres) of each other after considering the time of exhalation.”

*P8L4-7. Can authors add an adequate reference to argue their sentence "All volume-time curves were first checked by the diagnostic software, assessing the longevity of the exhalation phase (\( \geq 6 \) seconds in \( \geq 9 \) year-olds and \( \geq 3 \) seconds in \( \lt 9 \) year-olds)." I recommend the lecture of the following recent Letter to Editor: "Saad HB. Encouraging the publication of spirometric norms in healthy children from Africa. Int J Tuberc Lung Dis. 2019 Jun 1;23(6):764-765." The issue related the expiratory time was discussed.

We thank the reviewer for suggesting the paper to reference the time for exhalation phase. We therefore added the reference as suggested to:

“All volume-time curves were first checked by the diagnostic software, assessing the longevity of the exhalation phase (\( \geq 6 \) seconds in \( \geq 10 \) year-olds and \( \geq 3 \) seconds in \( \lt 10 \) year-olds).”

*P8L42: replace FEF25-75 by MMEF and define it.

We agree with the reviewer’s comments and have replaced the FEF25-75 with Maximal-Mid Expiratory Flow (MMEF) as suggested.

*P8L47: delete the following sentence "in the form \( \log(Y) = a + b*\log(\text{height}) + c*\log(\text{age}) + d*\text{Ethn} + \text{spline factor} + \text{error}\)."

We thank the reviewer for the suggestion and have deleted the GLI2012 formula.

*P9L26: SES was previously used P6L19. So P6L19 add SES after socioeconomic status.

We agree with the reviewer and we have therefore included the abbreviation for socioeconomic status (SES) as suggested. (Line 159-160)

7. Results

*P10L21: delete the following sentence "Mean height (centimetres) for girls was 139.6 (115.1-168.4) and for boys was 140.2 (116.4-166.4) and mean weight (kilograms) was 33.9 (19.5-56.1) for girls and 34.8 (18.7-59.7) for boys"

*P15L21-23: replace by "Boys had higher mean height and BMI for-age z-scores"

We thank the reviewer for the contribution and we had deleted and imputed with the reviewer’s suggestion to reduce redundancy in tables and text and now reads: (Line 177-180)

“However, boys had a higher mean age (10.7 years) than girls (10.3 years), height, mean weight, BMI-for-age, FEV1 and FVC z-scores, congruent with other studies that have also highlighted sex-related differences in anthropometry and lung function indices in children of the same age.”
*P10L32: replace as "After performing a Shapiro Wilk test, the FEV1/FVC (for both sexes) and MMEF (for boys) z-scores……."
We agree with the reviewer and therefore had revised the sentence as suggested. (Line 187-188)
“The Shapiro Wilk test, highlighted that the FEV1/FVC (for both sexes) and MMEF (for boys) z-scores generated from our sample were not a perfect standard normal distribution (mean≠0, SD≠1).”

*P10L34: it is unusual to write reference in the Results section.
We thank the reviewer for the contribution and we had revised the Results section accordingly as advised by the reviewer removing the term ‘reference’.

*P11L14-19: replace as "among" by "of" as previously done.
We agree with the reviewer and we therefore corrected as suggested. (Line 201-205)
“The distribution of spirometry z-scores in relation to the 5th percentile (LLN) identified that for FEV1, 8.7% (7.9% of boys, 9.6% of girls) and for FVC, 5.8% (4.1% of boys, 7.6% of girls) had values below the LLN. However, the FEV1/FVC z-scores showed a different pattern with 18.4% (18.2% of boys, 18.6% of girls) of participants having values below the LLN indicating a deviation from the GLI2012 distribution.”

*P11L40: (Figures 3 a, c, e),
We are not quite sure what the reviewer meant by this comment. However, we had formatted Figure 3 as indicated by the first reviewer and included correlation estimates on the individual scatter plots. (Figure 3)

8. Figure 1
*Define all applied abbreviations (BMI, ATS/ERS…)
*14 participants had outlying values? What is an outlying value? Define it in the method section? Use "pupils" rather than "students"?

We included the definitions for the abbreviations, used to develop Figure 1. We also thank the reviewer for the comment on outlying values and therefore included the definition in the method section and Figure 1. (Figure 1, Line 104-105)

9. Figure 2
*Define all applied abbreviations (BMI, ATS/ERS…)
*Replace FEF25-75 by MMEF
We thank the reviewer for the comment and we have included a key section under Figure 2 to define the abbreviations used (Body Mass Index (BMI), American Thoracic Society (ATS), European Respiratory Society (ERS) and also replaced FEF25-75 with MMEF in Figure 2.

10. Figure 3
*Define all applied abbreviations (BMI, ATS/ERS…)
*Replace FEF25-75 by MMEF
We have included a key section under Figure 2 to define the abbreviations used and replaced FEF25-75 with MMEF in Figure 3.

11. Figure 4
*Define all applied abbreviations (BMI, ATS/ERS…)
*Replace FEF25-75 by MMEF
We agree with the reviewer and have included a key section under Figure 2 to define the abbreviations used and replaced FEF25-75 with MMEF in Figure 4.

12. Table 1
*Define all applied abbreviations
*Use boys/girls rather than males/females.
We have included definitions for abbreviations included in table 1 and revised the terms to boys and girls as corrected by the reviewer.

13. Table 2
*Define all applied abbreviations
We thank the reviewer for the comments and have included definitions for abbreviations included in table 2.

14. Table 3
*Define all applied abbreviations
We apologise for the lack of abbreviations in this section of the manuscript and therefore had corrected and included abbreviations as suggested for all the terms in table 3.

15. Discussion
*P12L58: SDs
We have corrected the term SDs in the discussion as suggested by the reviewer

*P14L22: replace "more than 74,000" by "74,117"..
We thank the reviewer for the comment and therefore corrected to include the actual number of participants that were included in the GLI 2012 study and is now stated as: (Line 274-275)
“In contrast, the GLI2012 SRE produced standardised reference values from 74,117 healthy participants worldwide.”

*P14L27 and P15L34: add a comma before "respectively"
We agree with what the reviewer suggested and therefore have made corrections to the two sentences.
“Mean comparisons of percent predicted GLI2012 reference-derived values against the Polgar values in this population showed substantially higher lung function prediction (5.6%, 9.1% and 3.6% in FVC, FEV1 and FEV1/FVC, respectively).”

*Study limitation: what about FEF25, FEF50 and FEF75?
The study set to specify the performance of the African-American GLI 2012 spirometric reference equations for FVC, FEV1, FEV1/FVC and MMEF because they are the lung function indices mostly used in the Zimbabwean setting to help identify patients with a tendency towards a restrictive of obstructive lung pattern. Therefore, evaluating the performance of FEF25, FEF50 and FEF75 was not part of the scope of this study.

16. References
Please homogenize all references
We thank the reviewer for the contribution and we had formatted and homogenized the references.

Ref 4. Journal name?
We have included the journal name for reference 4 as highlighted by the reviewer.

Ref 5. Verify the journal name.
We have verified and corrected the journal name for reference 5

Ref 10. Journal name? website?
We have corrected reference 10, we have included the website from which the citation has been adopted from.

Ref 12. Respir Care
We thank the reviewer for pointing on the mentioned reference and we have clarified as Respiratory Care journal.

Ref 32. Journal name?
We have corrected the reference highlighted in by the reviewer and included the journal name.

Ref 36: delete "1,2" at the end of the title.
We thank the reviewer for the comment and therefore had corrected accordingly.

Ref 43. Plos One
This has been corrected.

Ref 5, 7, 18, 19, 26, 34, 36, 37, 43, 46, 47. Avoid capital letters in the title
We have corrected the references listed by the reviewer.

Helena Backman (Reviewer 3):

Main comments:
The aim of the study, as stated in the end of the introduction, seems to be to "evaluate the performance of the ERS/GLI2012 equations among urban and peri-urban Zimbabwean children aged 7-13 years against a standard normal distribution". I don't think that this is what the authors have done in the current study. The first part of this sentence is correct but the ending "against a standard normal distribution" seems questionable to me and should either be justified or omitted. I also hesitate to agree with the authors on their conclusion that evaluation of restrictive and obstructive spirometry should be based on the GLI reference values. I agree that the overall fit is evaluated in this paper, but we have no clinical data suggesting that these reference values identify obstructive or restrictive spirometry. This is a separate study.

We thank the reviewer for the comment and would like to highlight that, the aim of the study is to evaluate the performance of the GLI2012 spirometric equations among urban and peri-urban Zimbabwean children aged 7-13 years. This was done by evaluating the spirometry indices against the standard normal distribution (mean=0, SD=1, which is the distribution of the reference equations) and identify their distribution. Therefore the last part of the aim highlighted by the reviewer "against a standard normal distribution" is focused on the methods and analysis done to achieve this.

We agree with the reviewer that identification of a restrictive and obstructive spirometry pattern requires clinical data. Therefore we had revised after taking the previous reviewer’s suggestion and therefore highlight that spirometry help improve identification of a tendency towards restrictive and obstructive lung patterns. (Line 49, 280, 287)
In the Introduction (second page), line 12: I would like the authors to justify, as a response to this comment and perhaps not necessarily so in the manuscript, their claim that "However, the FEV1/FVC ratio has consistently demonstrated a better fit across populations than the other lung function measurements". I am not convinced that this is true.

The notion that FEV1/FVC ratio has better fit across the populations was considered after looking at studies that validated the GLI reference equations in their respective populations. Regardless of the conclusion of these studies, the FEV1/FVC ratio had the smallest absolute deviation from the standard normal distribution which shows a perfect fit. Some of the studies include in North Africa; Saad B et al (2013), Sweden; Backman et al (2015), North East Asians; Zhang et al (2018), Central Africa; Arigliani et al (2017), Australia and New Zealand; Brazzale et al (2016)

Furthermore, Lu metal;(2012) also noted the performance of FEV1/FVC ratio patterns were similar regardless of race in a study validating the GLI-2012 multi-ethnic spirometry reference equations in London school children (5-11 years)

This supports and hence resulted in our comment on the performance of FEV1/FVC ratio in this paper.

And in the following sentence the authors list possible reasons for poor fit of spirometry reference equations such as air pollution, malnutrition and low socioeconomic status. This statement is in one way true but on the other hand it is sort of incomplete and "one-eyed". The reasons for poor fit can also be due to the reference equations per se (and not only because of the harmful exposures and other factors in the population they are applied on) such as inappropriate modelling of the reference equation per se, inappropriate sample that the reference equation is based on, and so on. The population that the reference equations are applied in may also be healthier, less exposed and of higher socioeconomic status than for instance the GLI population, possibly resulting in "poor fit" or "erroneous estimations" as stated on line 22. I think this aspect should be considered better not only in the introduction, but throughout the manuscript.

We like to thank the reviewer for the comment and we agree that the reasons leading to misfit are more than those highlighted in this paper. We have included poor sampling as part of the reasons leading to misfit of the reference equations as: (Line 84-85)

“Possible reasons for poor fit of SRE include poor sampling and environmental factors such as exposure to indoor and/or ambient air pollution, malnutrition, and low socioeconomic status (SES), which may result in lower lung volumes on a population level, leading to erroneous estimations”

Thus, to a larger extent, the paper focused on environmental factors that can lead to an over-estimation (lower values at population level) by spirometric reference values.

I am not sure I agree that this is a random sample as stated in the last part of the discussion. It does not necessarily mirror the distribution in the Zimbabwean society. Instead, in the last part of the discussion where strengths and weaknesses are described, please clarify that these children are from three schools randomly selected from three areas with different socioeconomic standard. And further on this topic I would prefer that the conclusion is adjusted with the clarification that these children all come from the region of Harare, and are not randomly selected form the general population of Zimbabwe, as the current wording could implicate.
We thank the reviewer for the comment and state that the participants were randomly selected from schools. Also, the 3 schools were randomly selected from the 3 socio-economic zones. Therefore, we have clarified that, the participants in the study mirror the distribution of school-going children in urban and peri-urban Zimbabwe.

We also corrected the conclusion as the reviewer suggested to focus on children in urban and peri-urban Zimbabwe. (Line 310-311)

I find the methods regarding data collection and spirometry sufficiently described. However, it should be clarified if all included children were indeed Zimbabwean and if other "ethnicity/race" was an exclusion criterion.

We thank the reviewer for the comment and thus have added the part in the methodology to highlight the ethnicity of the participants selected in the study indicated as: (Line 27, 97)

“Between June and October 2018, black-Zimbabwean children aged 7 to 13 years were recruited from three schools in Harare randomly selected from three economic zones classified as high, medium and low-income status by the Ministry of Education.”

Minor comments:

The abbreviation ECSC stands for "European Coal and Steel Community" and not "European Coal and Steel Company".

We agree with the reviewer and therefore had revised the abbreviation as suggested. (Line 331)

In the Introduction (first page), line 15: Please consider replacing the word "are" with "can be".

In the Introduction (first page), line 45: Please consider replacing the word "are" with "can be". The GLI task force recommends the use of the 2.5th percentile as LLN for epidemiological studies, and thus the 5th percentile is not always and per definition the golden standard.

We have made the revisions as suggested by the reviewer to help define the LLN as (Line 71-72)

“The GLI2012 spirometric equations provide age-, height-, sex- and race-specific reference values. These equations provide lower-limit-of-normal (LLN) values, which can be defined as the 5th percentile values (z-score < -1.64) of the healthy, non-smoking population.”

In the second page of Results, line 11: consider changing the word "centile" to percentile". On the same page on line 19, consider replacing the word "normality" with "the GLI distribution" or similar. And further on the same page, line 33-35, the authors state that "...the strengths of associations (β coefficients) from linear regression were within ±0.2 (Supplementary File 2).". This statement should be changed as the absolute value of β coefficients says nothing about the strengths of associations. The strength of association cannot be evaluated by the value of β alone, but rely heavily also on the values of the X-variable. For instance the value of β is affected substantially if e.g. height would be an X-variable included measured in centimetres or meters.

We thank the reviewer for the comments and we have replaced the word ‘centile’ and normality by percentile and GLI distribution respectively as:
“The distribution of spirometry z-scores in relation to the 5th percentile (LLN) identified that for FEV1, 8.7% (7.9% of boys, 9.6% of girls) and for FVC, 5.8% (4.1% of boys, 7.6% of girls) had values below the LLN.”

We have revised the statement on the β coefficients from the results section in the manuscript to read:
(Line 209-211)
“The linear associations between spirometry, anthropometric variables and school income level were not uniform, and the β coefficients from linear regression were within ±0.2”

Please consider adding the words "With perfect fit" before the sentence "The z-scores developed from the ERS/GLI…” at the second page of Discussion, lines 37-38.
We have revised the statement and included the reviewer’s suggestions as highlighted in the above comment and now reads:
“With a perfect fit, the z-scores developed from the GLI2012 spirometric reference equations should show a lack of association with race and anthropometric variables since they are independent variables for generating the LLN.”

The reference list is not appropriately formatted. Some references are duplicated and sometimes the journal names are missing. Please revise.
We have revised and formatted the reference list as advised by the two reviewers.

On the third Discussion page, please consider replacing the word "better" by the word "higher" on line 25. On the same page, please consider revising the sentence on lines 35-38 as previously commented on.
We thank the reviewer for the suggestion and has revised the sentence as suggested by the reviewer.
(Line 276)
“Mean comparisons of percent predicted GLI2012 reference-derived values against the Polgar values in this population showed substantially higher lung function prediction.”

Please clarify the statement "Using the LLN as compared to the Polgar fixed cut-offs of 70%.." on the third discussion page, lines 52-57. I don't think that the fixed cut-off of FEV1/FVC<0.7 is generally referred to as the "Polgar fixed cut-off" but more commonly it is stated that this cut-off is generally used on many clinics due to simplicity and that this cut-off is recommended by GOLD. This could be corrected.
We have revised the statement and removed the term Polgar and specified the cut-offs of 70%. (Line 287-289)
“Using the LLN as compared to the fixed cut-offs of 70% reduces errors in the diagnosis of airflow obstruction, and the Global Initiative for Chronic Obstructive Lung Disease (GOLD) group acknowledges its effect in reducing false-positive and false-negative results.”

Finally, I suggest adding mean and SD to the captions for Figures 3 and 4.
We thank the reviewer for the suggestion. However, Figure 3 aims at showing the relationships between spirometry z-scores and anthropometry z-scores. We included the correlation coefficients in Figure 3 captions to help show the scatterplots patterns.
Also, Figure 4 compares the mean percentage predicted values for the spirometry indices from the GLI and Polgar reference equations. We however believe that, the comparative bar graph shows the comparisons and how the two reference sets of equations are performing in this population.
Introduction

- Page 4 line 40: The authors appear to have listed regions/groups of data rather than (understandably) listing the 33 countries that contributed data. However, several contributors of data are missing from this sentence including Europe and Australia and in fact, data from Japan was not included in the GLI dataset due to publishing embargo. This sentence needs some clarification.

We thank the reviewer for the comment and we have revised the list of region contributors to the GLI 2012 data. We also excluded Japan as advised by the reviewer. (Line 67-68)

“The data used to generate the GLI2012 SRE were collected from Europe, Australia, Latin America, East Asia, India, North America and North Africa.8 The GLI2012 provide race-specific equations for Caucasians, African-Americans, South East Asians and North East Asians.”

- Following on from the point above, it would be beneficial for readers to know what the race-specific reference equations are (i.e. Caucasian, African-American etc).

We agree with the reviewer to this suggestion and therefore included the specific reference equations that are under the GLI2012 modules. (Line 68-70)

“The GLI2012 provide race-specific equations for Caucasians, African-Americans, South East Asians and North East Asians.”

Methods

- Could the authors provide some additional information into the differences between the three economic zones? Furthermore, what was the breakdown of participants from each of the schools? Was there an even distribution in numbers recruited, age range and gender ratios?

We thank the reviewer for the comment and we agree with the reviewer to add the distribution of participants in Supplementary file 3: Table 1.

Notably, the three schools were selected each from the 3 economic zones classified by the Ministry of Primary and Secondary Education in Zimbabwe.

“The schools were classified after taking into account, the economic status of the school and location of the school.”

- Page 7 line 7: The authors have written "exhalation phase (≥ 6 seconds in ≥9 year olds and ≥3 seconds in <9 year olds)” however current guidelines list 10 years of age.

We thank the reviewer for the comment and we have corrected the time of the exhalation times and age groups as the current guidelines explains. (Line 128-129)

“All volume-time curves were first checked by the diagnostic software, assessing the longevity of the exhalation phase (≥ 6 seconds in ≥ 10 year-olds and ≥ 3 seconds in < 10 year-olds)30”
The rationale for the comparison of the GLI equations against the ECSC references needs to be mentioned in the aims. It is not until the discussion that mention is made of the ECSC references being current clinical practice.
We thank the reviewer for this contribution and we have added a rationale for the comparison of the GLI and the ECSC equations in this population. (Line 93-94)

“The GLI2012 spirometric reference equations were also compared against the Polgar equations because they are currently used in clinical practice.”

Page 8 line 58: Please add a reference the last sentence of the methods section.
We have added the reference to the sentence as suggested by the reviewer. (Line 172)

The abstract mentions that participants with a BMI z-score < -2 were excluded however this is not mentioned in the manuscript.
We thank the reviewer for pointing out this important aspect and we have included this exclusion criteria item in the methods section. (Line 104-105)

“Participants with outlying FVC and FEV1 values (Median ± 1.5 [Interquartile Range]) and BMI z-score < -2 were also excluded from the analysis dataset”

Results

- Could the authors please clarify the number of participants who failed to perform technically acceptable spirometry. In the manuscript, (page 9 line 10) 33 participants were excluded however in figure 1, 19 are listed as failing ATS/ERS guidelines and 14 had outlying values. What was considered 'outlying values'? It would be inappropriate to remove these participants if they simply had too high or too low values.
A total of 33 participants were excluded from the study on technical terms, that is, because they failed to meet the ATS/ERS spirometry guidelines or their volumes were regarded as outliers during data cleaning. Since spirometry is effort based, all outlying values (Median± 1.5 (Interquartile range)) were excluded for technical grounds during data cleaning because we wanted to maintain a normal population based on the effort put during the spirometry manoeuvres. Thus 19 participants who were strictly excluded on ATS/ERS grounds and 14 with outlying values were all excluded from the study.

- I admit that I am unfamiliar with the shapiro-wilk test and Q-Q plots however I found this section difficult to read and understand as they appeared to say opposite things. Shapiro-wilk test = not normally distributed, Q-Q plots = normally distributed.
The study used the Shapiro Wilk test to evaluate the hypothesis that all the spirometry z-score are normally distributed at 95% confidence level. A p-value < 0.05 would suggest lack of normality for the spirometry z-scores. The results for the Shapiro Wilk test provide strict evaluation against the standard normal distribution. However, the Q-Q plots were mainly used to show the extent to which FEV1/FVC (for both sexes) and MMEF (for boys) z-scores deviated from normality. (Line 187-193)

- Page 9 line 49: The sentence "compared to other ERS/GLI2012 ethnic modules which were also generally within ±0.5 of zero" is not accurate. Based off table 2, only South-East Asian for FVC and Caucasian/North-East Asian/Other for FEV1/FVC meet this criterion. In it's current format this sentence suggests that other reference equations may also be appropriate for this population which is not the case. This sentence also appears in the results section of the abstract.
We thank the reviewer for the comment and we have revised the sentence to the intended meaning as:
“The African-American module gave the smallest absolute differences (closest to zero) as compared to other GLI2012 ethnic modules which were also generally out of the range of ±0.5.”

- Can the authors please justify why the scatterplots are not further separated according to gender? When looking at the numbers provided in the text, (page 10 lines 9-14) more male participants had z-scores above the LLN for both FEV1 and FVC compared to females. Was the distribution of females below the LLN consistent across the age range? Could there be bias in the female participants?

We thank the reviewer for the comment and we decided not to make z-score scatterplots comparisons by sex because there were no differences for most (FVC, FEV1 and FEV1/FVC) spirometry z-scores by sex. We expected no differences in spirometry z-scores by sex because is adjusted for, in the GLI2012 equations. The analysis of spirometry scatterplots were however performed on variables that differed by sex in this sample (age, height and BMI z-scores).

Discussion

- I feel that the opening paragraph needs to better highlight which of the GLI2012 reference equations is the most appropriate for this population. For example, the final sentence states "the ERS/GLI2012 reference values are applicable" … which ones? All of them?
We thank the reviewer for the suggestion and we have therefore clarified the GLI2012 module that applies in this population as African-American GLI2012 SRE module. (Line 231-235)

- Similarly to the above point, the sentence "mean z-scores for all the spirometry variables were within 0.5 z-scores from zero" … this was not true for all of the GLI2012 equations (Caucasians, North-East and South-East Asian). This clarity should be included in several sentences throughout the discussion as well as in the abstract.
We have clarified the specific GLI2012 reference equation module (African-American) that was compared against the local population and now addressed as African-American GLI2012 SRE (Line 234-235)

- Did the authors collect any information on other known influencing factors to lung development including low birthweight/preterm birth, adverse childhood respiratory events and number of hospitalisations? If not, this should be acknowledged as a limitation.
We thank the reviewer for the contribution. During data collection, the study collected data on the history of respiratory conditions but not on birthweight and preterm birth status. Therefore we have included the missing parts suggested as limitations to this study as suggested as: (Line 307-308)
“The study did not capture birthweight and preterm status which is associated with the general lung development in children.”