Author’s response to reviews

Title: Considerations in the use of different spirometers in epidemiological studies

Authors:

Edith Milanzi (e.b.milanzi@uu.nl; milanziedith@yahoo.com)
Gerard Koppelman (g.h.koppelman@umcg.nl)
Marieke Oldenwening (M.Oldenwening@uu.nl)
Sonja Augustijn (s.w.s.augustijn@umcg.nl)
Bernadette Aalders-de Ruijter (B.M.Aalders@uu.nl)
Martijn Farenhorst (m.p.farenhorst@umcg.nl)
Judith Vonk (j.m.vonk@umcg.nl)
Marjan Tewis (m.tewis@uu.nl)
Bert Brunekreef (B.Brunekreef@uu.nl)
Ulrike Gehring (U.Gehring@uu.nl)

Version: 1 Date: 06 Apr 2019

Author’s response to reviews:

Dear Editor,

Enclosed please find a marked up copy of the revised version of our manuscript entitled ‘Considerations in the use of different spirometers in epidemiological studies’ (Ref: ENHE-D-18-00397.”, taking into account the reviewer’s comments. Point-by-point responses to the comments of the reviewer are provided below.

Substantial revisions were made to the text while addressing each of the concerns raised. We now extensively address the justification of the study design as well as limitations of the study. Overall we believe that these have significantly improved our manuscript over the previous version.
We appreciate the opportunity to submit a revised version of our manuscript. We thank you for your consideration of our manuscript and look forward to your decision on the revised version.

Sincerely,

Ulrike Gehring, PhD (on behalf of all co-authors)

Reviewer #1: GENERAL COMMENT

The manuscript aims at assessing comparability between measurements obtained by different spirometers in two series of 49 children. In general, the manuscript is well written and is of interest, but there are some issues that should be dealt with by the Authors.

MAJOR COMMENTS

Methods

Comment 1

Page 4, lines 56-58: It seems that anthropometric data were not objectively measured but only asked to the studied subjects. If this is the case, advantages (if any) and limitations of such a method should be discussed.

Response 1:

Indeed, weight and height of the participants were self-reported and not objectively measured. Since we compared spirometers within persons, i.e. within the same group of subjects, under the same conditions, height and weight of the participants did not differ between the two spirometers and consequently do not affect the observed relationship between the two spirometers. However, in reference to the GLI equations used to estimate percent predicted lung function for the two spirometers (presented in Table 1), the self-reported height could bias estimation of percent predicted lung function per spirometer (self-reported weight as it is not a predictor in the GLI equations).

We have included this limitation on page 14-15, lines 310-320 of the revised marked manuscript as follows; “We used self-reported instead of measured height and weight for the in total 98 volunteers that participated in the comparisons of the spirometers. Since spirometers were compared within persons, and consequently height and weight did not differ between the spirometers that were compared within a series, this does not affect the observed differences between spirometers. Self-reported height might be a source of bias in the GLI equations as
height values may be over-/underreported. Weight is not used in the GLI equations to estimate percent predicted lung function and therefore poses no risk of bias. Studies of the agreement between self-reported and measured weight and height provided inconsistent results, some suggested good agreement (26, 27), while others reported significant discrepancies mainly in overweight/obese individuals (28, 29).


Comment 2

Page 6, lines 7 - 15: this procedure is based on some assumptions which might turn to be undemonstrated. First, the Authors report that their 49 volunteers were "healthy", possibly meaning that they were disease- and symptom-free. However, no information is provided on individual characteristics, including environmental exposures. Other assumptions are cited in the comments to Discussion.

Response 2:

In relation to the response above, for the purposes of the comparison of lung function measurements from the different spirometers, other characteristics such as environmental exposures were not assessed as the comparisons have been performed within persons and in
random order and consequently individual characteristics/ environmental exposures did not influence the observed difference between the spirometers. This has now been mentioned in the methods section on page 6 lines 115-117 of the revised marked manuscript as follows; “In both series, all volunteers performed tests on both spirometers in random order but in immediate succession to eliminate confounding by individual characteristics”

Discussion

Comment 3

Page 8- lines 15-25: It would be worthwhile to quote also other approaches to assessment of quality of spirometry test performance (e.g. Enright PL et al, Am Rev Respir Dis. 1991; Enright PL et al, Chest. 2000) and to inter-laboratory comparison of spirometry measurements in international epidemiological studies (e.g. Viegi G et al, Respir Med 2000)

Response 3:

Thank you for the suggestion, the suggested references have been added on page 13, lines 272-275 of the marked revised manuscript as follows;

“Several studies have conducted similar experiments comparing different types of spirometers, handheld/office and standard laboratory spirometers both in clinical and research settings (2-4, 19-22), with the comparisons also used as quality control procedure in international multicentre epidemiological studies (23, 24).”


Comment 4
Page 8 - lines 27-40: However, there are some papers in the literature showing that the GLI equations may not be suitable for all the populations (e.g. in North Africa, Ben Saad H et al, Respir Med 2013; in Sweden, Backman H et al, BMC Pulm Med. 2015; in Italy, Fasola S et al, Respir Med 2017). It would be useful to check the applicability of GLI equations in the Netherlands, before assuming that they are a gold standard.

Response 4:
Indeed, it has been reported that the GLI reference equations may not be suitable for all populations. Studies in the Netherlands have shown that measurements in the Dutch population
generally agree with the Global Lung Initiative references values in adults (van Oostrom, 2018). While we cannot claim the equations to be the gold standard for the Dutch population, we believe these equations are most likely suitable for our current study population as the Masterscreen-EasyOne comparison population, was 100% Dutch. The ethnicity frequencies for the current study population have now been reported in Table 1. A few sentences have also been included in the discussion to highlight this on page 12, lines 263-268 of the revised marked manuscript as follows:

“It is important to note that the GLI reference equations are not universally applicable. However, these equations are based on an extensive database and studies in the Netherlands have shown that measurements in the Dutch population generally agree with the GLI references values in adults (17). We therefore believe these equations are most likely suitable for our current study population as the Masterscreen-EasyOne comparison population was 100% Dutch.”


Comment 5

Page 9, lines 34-44: two different technicians performed the first test series, one for the Masterscreen and one for the EasyOne. Even if they were both well-trained, this might represent an important source of bias, since it is unclear if the observed differences have to be ascribed to the spirometer or the technician. Moreover, it is not possible to adjust for the technician effect, since the two factors "spirometer" and "technician" are perfectly associated. The fact that two different technicians performed the Masterscreen and the EasyOne measurements need to be explicated in the Methods, and the consequent risk of bias clearly highlighted in the Discussion as a study limitation.

Response 5:

Thank you for the important observation. Indeed, the potential technician bias is important. It should be noted however that this was done by design to reflect how in a multicentre research setting, one technician in one center can operate one spirometer and another technician in another center can operate a different spirometer and how this can introduce bias in the measurements. This has been mentioned in the methods section on page 6 lines 108-112 of the marked revised manuscript as follows;

“Two highly experienced and trained technicians conducted spirometry measurements in the first test series (one for the Masterscreen pneumotachograph and one for the EasyOne1). We let each technician use a different spirometer by design to reflect a real-life multicentre research setting
where different spirometers are used in different centers by different technicians”. In the discussion, on pages 14 lines 299-307 of the marked revised manuscript, we have highlighted the implications this might have on the measurements obtained and consequently epidemiological associations and how the provided correction can help to correct for both the device and the technician effect as follows;

“We designed the comparison of the Masterscreen pneumotachograph and EasyOne spirometers such that different technicians operated the different spirometers to imitate a real multicentre study. While the technicians were highly trained and experienced, due to the study design it was impossible to disentangle differences between spirometers from differences between technicians. Consequently, part of the observed difference between spirometers may be attributable to differences between technicians. The provided correction equation thus simultaneously corrects for the technician and device effect and may not be generalizable to other studies where different technicians are involved. However, it is expected that the calibration method can be applied accordingly.”

SPECIFIC COMMENTS

Comment 6
- Abstract, line 15; page 4, line 5; page 6, line 18: technically speaking, the Authors have not "computed a correction factor", but they have rather provided a "correction equation".

Response 6:
The phrase has now been corrected replacing ‘computed” with ‘provided” and “factor” with “equation” in the mentioned places on page 3; lines 47-48, page 6; line 99, page 9; line 181-182 of the marked revised manuscript.

Comment 7
- Page 4, line 50: there is a missing open parenthesis before "… n=4 for each of the two series)".

Response 7:
The missing parenthesis has been added on page 7 line 123 of the revised marked script

Comment 8
- Page 5, line 13: Yorba Linda, not Yoba Linda
Response 8:
The typo has been corrected on page 8 line 148 of the revised marked script

Comment 9
Page 6, line 53: the percentage differences should be better explained:

6.3% appears to derive from 100(1-3.54/3.78) (% decrease switching from Masterscreen to EasyOne);

8.4% appears to derive from 100(4.78/4.41-1) (% increase switching from EasyOne to Masterscreen)

Similarly, 1.15% and 0.9% are % increase switching from EasyOne2 to EasyOne1.

Response 9:
In response to this comment, the percent differences have been further explained as suggested, the text has been revised on page 10 lines 199-203 of the revised marked script as follows;

“The 0.24 L and 0.37 L mean differences, correspond to a 6.3% decrease in FEV1 switching from the Masterscreen pneumotachograph to the EasyOne1 spirometer and 8.4% decrease in FVC switching from the Masterscreen pneumotachograph to the EasyOne1 spirometer respectively…. “

And lines 205-208 of the same page “The mean differences correspond to a 1.1 % decrease in FEV1 switching from the EasyOne1 to the EasyOne2 spirometer and 0.9% decrease in FVC switching from the EasyOne1 to the EasyOne2 spirometer.”

Comment 10
- Page 8, line 13: "…. Measurements were comparable." However, with a systematic significant difference.

Response 10:
In response to this comment we revised the text on page 12 lines 244-246 of the revised marked manuscript as follows: “We also investigated the agreement between two EasyOne spirometers of the same generation and found that measurements were comparable, but with a small significant difference.”

Comment 11
- Page 8, line 40: are the Authors able to ascertain if the study population falls under GLI 2012 database?

Response 11:
See response 4. Validation of the GLI reference values using local controls generally requires larger sample sizes to take into account sampling error. As such we cannot strictly claim that the current study population falls under the GLI database with our sample size, but a study involving the Dutch population has shown good agreement with the GLI 2012 reference values (van Oostrom, PloSOne. 2018).


Comment 12
- Page 10, line 5: However, a formal computation of the sample size to test the hypothesis would be useful.

Response 12:
We have added a statement detailing sample size calculation in the statistical analysis section on page 9, lines 169-171 as follows; “Sample size was obtained based on a standard deviation (SD) for FEV1 of 0.5L. With a significance level of 0.05, 44 volunteers were required to detect a mean difference of 0.3L between the spirometers with 80% power.”

Comment 13
- Page 10, line 6: before the Conclusion, the Authors should also underline the study limitations, such as potential bias due to technician effect, lack of assessment of external validity of the proposed correction method.
Response 13:

Limitations of the study have been outlined in the discussion on page 14, lines 296-320 of the marked revised manuscript as below:

“However, the following limitations should be considered; two experienced technicians performed the first test series (one for the Masterscreen pneumotachograph and one for the EasyOne) and one of them performed all measurements of the second test series. We designed the comparison of the Masterscreen pneumotachograph and EasyOne spirometers such that different technicians operated the different spirometers to imitate a real multicentre study. While the technicians were highly trained and experienced, due to the study design it was impossible to disentangle differences between spirometers from differences between technicians. Consequently, part of the observed difference between spirometers may be attributable to differences between technicians. The provided correction equation thus simultaneously corrects for the technician and device effect and may not be generalizable to other studies where different technicians are involved. However, it is expected that the calibration method can be applied accordingly. We were not able to assess the external validity of the correction for spirometry measurements outside the PIAMA population, but it has been used before to correct spirometry measurements (6) and the method has been validated in other fields of epidemiology (25). We used self-reported instead of measured height and weight for the 98 (in total) volunteers that participated in the comparisons of the spirometers. Since spirometers were compared within persons, and consequently height and weight did not differ between the spirometers that were compared within a series, this does not affect the observed differences between spirometers. Self-reported height might be a source of bias in the GLI equations as height values may be over-/underreported. Weight is not used in the GLI equations to estimate percent predicted lung function and therefore poses no risk of bias. Studies of the agreement between self-reported and measured weight and height provided inconsistent results, some suggested good agreement (26, 27), while others reported significant discrepancies mainly in overweight/obese individuals (28, 29).”


Comment 14

- Page 10, lines 16-17: It should also be pointed out that a correction may be important also for estimating prevalence rates of airflow obstruction in a population when different spirometers are used in different occasions (see, for instance, Maio S et al, Respir Med 2016).

Response 14:

Thank you for the suggestion, we have incorporated this in the discussion on page 12 in lines 255-263 of the revised marked manuscript as follows:

“The lower percent predicted lung function for the EasyOne1 suggests that the EasyOne spirometer may be more likely to overestimate the percentage of subjects with a clinically low lung function in a setting where different spirometers are used. This has been previously demonstrated in a comparison involving the EasyOne spirometer and a water-sealed spirometer (Collins, Stead-Wells) where underestimated values of both FEV1 and FVC from the EasyOne spirometer and consequently higher prevalence rates of airway obstruction were observed (16).”