



Reference equations for pulmonary diffusing capacity using segmented regression show similar predictive accuracy as GAMLSS models

Gerald Stanley Zavorsky ,¹ Jiguo Cao²

To cite: Zavorsky GS, Cao J. Reference equations for pulmonary diffusing capacity using segmented regression show similar predictive accuracy as GAMLSS models. *BMJ Open Res* 2022;**9**:e001087. doi:10.1136/bmjresp-2021-001087

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjresp-2021-001087>).

Received 23 August 2021
Accepted 24 January 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Pulmonary Services, Department of Medicine, UC Davis, Sacramento, California, USA

²Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia, Canada

Correspondence to

Dr Gerald Stanley Zavorsky; gerryzavorsky@gmail.com

ABSTRACT

Purpose To determine whether generalised additive models of location, scale and shape (GAMLSS) developed for pulmonary diffusing capacity are superior to segmented (piecewise) regression models, and to update reference equations for pulmonary diffusing capacity for carbon monoxide (DLCO) and nitric oxide (DLNO), which may be affected by the equipment used for its measurement.

Methods Data were pooled from five studies that developed reference equations for DLCO and DLNO (n=530F/546 M; 5–95 years old, body mass index 12.4–39.0 kg/m²). Reference equations were created for DLCO and DLNO using both GAMLSS and segmented linear regression. Cross-validation was applied to compare the prediction accuracy of the two models as follows: 80% of the pooled data were used to create the equations, and the remaining 20% was used to examine the fit. This was repeated 100 times. Then, the root-mean-square error was compared between both models.

Results In males, GAMLSS models were 7% worse to 3% better compared to segmented regression for DLCO and DLNO. In females, GAMLSS models were 2% worse to 5% better compared to segmented linear regression for DLCO and DLNO. The Hyp'Air Compact measured DLNO and alveolar volume (VA) that was approximately 16–20 mL/min/mm Hg and 0.2–0.4 L higher, respectively, compared to the Jaeger MasterScreen Pro. The measured DLCO was similar between devices after controlling for altitude.

Conclusions For the development of pulmonary function reference equations, we propose that segmented linear regression can be used instead of GAMLSS due to its simplicity, especially when the predictive accuracy is similar between the two models, overall.

INTRODUCTION

In 2000, a new approach for the development of reference equations for spirometry was described that allowed for a smooth transition between childhood and adulthood in a continuous fashion.¹ This modelling technique prevented discontinuities between paediatric and adult reference equations at the transition point, preventing misinterpretation. This methodology was based on a semiparametric regression approach of generalised

Key messages

What is already known on this topic

► Do segmented (piecewise) regression provide similar predictive accuracy compared to more complicated generalised additive models of location, scale and shape (GAMLSS) that the Global Lung Function Initiative Network uses for pulmonary function reference equations?

What this study adds

► Segmented linear regression for pulmonary function show similar predictive accuracy as GAMLSS models. Furthermore, the pooled data from five previously published studies (total pooled subjects=1076) demonstrate that the Hyp'Air Compact device measured pulmonary diffusing capacity for nitric oxide (DLNO) and alveolar volume (VA) that was approximately 18 mL/min/mm Hg and 0.3 L higher, respectively, compared with the Jaeger MasterScreen Pro (CareFusion, Germany; now Vyair Medical). However, it is not known which device measures DLNO and VA more correctly.

How this study might affect research, practice or policy

► The Hyp'Air Compact (Medisoft, Sorinnes, Belgium) measured DLNO and VA that was systematically higher than that of the Jaeger MasterScreen Pro. Thus, the two manufacturers should come together to resolve these between-machine differences. In the meantime, more comprehensive reference equations are updated here, accounting for the lung function testing device, where applicable.

additive models for location, scale and shape (GAMLSS) and was discussed again in the same journal in 2008.² GAMLSS allowed for age-related differences in between-subject variability, improving the definition of the lower limits of normal (LLN).²

Subsequently, in 2010, a European Respiratory Society (ERS) Task Force was created to create multiethnic, all-age reference equations for lung function for world use using GAMLSS models.³ This allowed for a single



reference source that would be able to monitor patients from childhood into old age.³ As there were over 400 published reference equations describing healthy lung function changes with age, sex and height, professionals were left with a decision of which equation to use.⁴ Thus, the Global Lung Function Initiative (GLI) Network was created to address discrepancies in lack of standardisation.⁴ These new 'Global' reference equations, using healthy subjects' data from around the world, were developed to model changes in lung size with age and height from childhood to adulthood.³ These complex growth patterns were modelled using GAMLSS that smoothed centile curves.⁵

Since 2012, the GLI Network produced three significant papers that were endorsed by the ERS and the American Thoracic Society and published in the European Respiratory Journal, which provided global reference equations for spirometry,⁶ pulmonary diffusing capacity for carbon monoxide (DLCO)^{7,8} and static lung volumes.⁹ Those articles presented reference equations using GAMLSS models. GAMLSS were introduced initially in 2005¹⁰ and updated in 2018,¹¹ allowing for a variety of smoothing functions. Besides pulmonary medicine, GAMLSS have been used in several fields such as exercise science,^{12,13} chemistry,¹⁴ hydrological science,¹⁵ genomics¹⁶ and psychology,¹⁷ to name a few; thus, GAMLSS models have pertinence across many disciplines.

However, GAMLSS are highly complex and challenging to implement (ie, see www.gamlss.com). One needs to understand distributions of a variable (and its properties), then decisions need to be made regarding the distribution of the response variable, the choice of explanatory variables, the link function (ie, monotonic functions of the distribution parameters) and the amount of smoothing and random effects.¹⁵ Thus, the application of GAMLSS models estimates time-varying quantiles, which are distribution dependent, so the selection of a suitable distribution is important.¹⁵ As such, there is a sophisticated understanding of physiology, statistics and computer programming that is involved in producing a proper model using GAMLSS.

However, segmented or 'broken-line' models are regression models that are simpler to use and should be the model of choice for the development of reference equations for lung function across the whole lifespan. Segmented regression is less complex, easier to comprehend and can be applied more readily applied as the formulas are easier to understand. Segmented regression allows for predictions to be made without experiencing discontinuities due to transitions from one prediction equation to the next. This is especially important in lung function prediction equations, in which one prediction equation is developed for children, and then another separate equation is developed for adults. Furthermore, once the equations are developed, a simple calculator can be used to obtain the predicted value without the use of splines.

Fitting piecewise or segmented terms in regression models for pulmonary function use age as the non-linear covariate with two-line segments connected at one breakpoint.^{18,19} From visual observation, this breakpoint occurs somewhere around 20 years of age forced vital capacity (FVC), forced expiratory volume in one second (FEV₁)^{18,20} and DLCO.²¹ Thus, it is the premise of this article to demonstrate segmented (piecewise) linear regression can be used more easily with similar prediction errors as GAMLSS models. We also believe that segmented regression models are more parsimonious compared with GAMLSS models, meaning that segmented regression could achieve goodness of fit using as few explanatory variables as possible. This reasoning comes from the idea of 'Occam's razor', which says that the simplest explanation is probably correct.

As such, the primary purpose of this study was to determine whether pulmonary diffusing capacity modelled using segmented linear regression with one breakpoint provides similar prediction accuracies as GAMLSS but without the use of complicated splines. It is our assumption that DLCO, pulmonary diffusing capacity for nitric oxide (DLNO) and alveolar volume (VA) could be modelled using separate segmented linear equations for each sex, which would be less complex compared to GAMLSS while providing similar prediction errors as GAMLSS. A secondary purpose was to update the pulmonary diffusing capacity prediction equations published by an ERS Task Force in 2017.²² Nearly 80% of the subjects used in the development of reference equations for the ERS task force in 2017 had pulmonary diffusing capacity measured by the Hyp'Air Compact device (Medisoft, Sorlines, Belgium). However, evidence suggests that the predicted DLNO is varied depending on the reference equation applied,^{23,24} which can be due to the different pulmonary function devices used between studies.²⁵ Thus, with a much larger pooled dataset to draw on, we also sought to evaluate between device discrepancies.

MATERIALS AND METHODS

Five previous studies that developed reference equations for DLNO in white individuals without cardiopulmonary disease were pooled and used in this study.²⁶⁻³⁰ Institutional Review Board approval was not needed as the deidentified data were obtained from previously published work. Data from three separate studies²⁶⁻²⁸ were obtained from a 2017 ERS task force on the technical standards of DLNO²²; another set was publicly available online,²⁹ and the fifth dataset set was created based on the anthropometric characteristics of another paper.³⁰ (Note: Munkholm *et al*³⁰ declined to provide us with their data after multiple repeated attempts. As such, we created simulated data that was statistically tested to be similar to their data using a statistical method called truncation. The procedures on how this fifth dataset was created are discussed in the online supplementary material).

Segmented (piecewise) linear regression models

Reference equations were created for DLCO, DLNO and VA using the 'R' language environment (<http://www.r-project.org>). The 'segmented' package was originally developed in 2008,³¹ based on previous work on piecewise fitting of at least on breakpoint³² (V.1.3–4, April 2021) generated the segmented models.³³ The covariate 'age squared' (Age^2) was used to estimate the single breakpoint for the entire age range of the data (5–95 years of age). Based on a visual plot between age^2 and either DLCO, DLNO or VA, an estimated starting value for the breakpoint is provided, and then an iterative procedure in R is used to estimate the breakpoint³² and the 95% CI of the breakpoint.³⁴

Other covariates used in the models were height (cm) or height², sex (1=male; 0=female), altitude (0–300 m), weight (kg) and the pulmonary function device. The brand of pulmonary function system was listed as a potential predictor of the model since there are discrepancies in DLNO depending on which equipment is used.³⁵ The devices used to measure pulmonary diffusing capacity were the Jaeger MasterScreen PFT Pro (CareFusion, Hochberg, Germany), Jaeger Masterlab Pro (Erich Jaeger, Würzburg, Germany) with NO chemiluminescence (77AM, Eco Physics, Switzerland) and the Hyp'Air Compact device (Medisoft).

Generalised additive models of location, scale and shape

The GAMLSS models developed here are implemented in a series of CRAN packages in the R language environment and are currently available for download at <http://www.r-project.org>.¹⁰ The Lamda-Mu-Sigma (LMS) method of Cole and Green was applied as an extension of the normal distribution that adjusts for skewness⁵ and is embedded in GAMLSS. The LMS method is equivalent to Box-Cox Cole and Green distribution (BCCG), BCCG (μ , σ , ν) and parameters μ , σ , ν are the approximate median, approximate coefficient of variation and approximate skewness parameters of the distribution of the response variable.¹¹ That is, μ controls the location, σ controls the scale and ν controls the skewness of the distribution as people grow and age.¹¹ The complex effects for the predictor variables on the dependent variable were modelled using splines, which allow the dependent variable to vary smoothly (non-linearly) as a function of a predictor. Thus, a continuous, smooth fit over the entire age range can be obtained using splines. The goodness of fit was assessed by Akaike's Information Criterion,³⁶ Bayesian Information Criterion,³⁷ Quantile-Quantile (Q-Q) plots³⁸ and worm plots.³⁹

The between-individual variability across age was assessed by obtaining the predicted SD divided by the predicted mean multiplied by 100. The predicted mean was determined by taking the median height at each age from the white US population⁴⁰ and applying a zero altitude for each model. The predicted SD was the residual

SD (RSD) obtained from the segmented linear regression models and the sigma value obtained from GAMLSS.

Prediction accuracy between models

To assess the prediction accuracy of the segmented linear regression and GAMLSS models, repeated random subsampling using the Holdout method was used that randomly sampled the complete dataset into two mutually exclusive subsets, a training set and a test set (also called a validation or Holdout set), repeated over several times.⁴¹ Eighty per cent of the pooled data was used to fit both models (training set), and then the fitted equation predicted the remaining 20% of the test subjects (validation set). This process was implemented for 100 replicates. The median, minimum, maximum and 95% CI of the root-mean-square error (ie, the square root of the average of the squared errors) from the 100 random samplings of the pooled data were compared between both models. The average correlation coefficients between each predicted value and the actual values obtained for 20% of the test data were also reported. The results of the repeated sampling would demonstrate whether GAMLSS or segmented linear regression models would be systematically favoured.

The LLN for both models was chosen as the fifth percentile. The LLN is the value below which there is only a 5% probability that the value from a population is normal. This was calculated by subtracting from the model the product of the one-sided area under the curve and the equation's RSD (−1.645 RSD).

Other analyses

Correlations were used to examine associations between variables. The GLI equations for DLCO⁷⁸ were also used to compare DLCO and VA against both segmented linear regression and GAMLSS models. A 2×3 repeated measures analysis of variance (RMANOVA) compared fitted z-scores between the three different types of prediction models (segmented linear regression, GAMLSS and GLI GAMLSS) for DLCO and VA and the pulmonary function device used. A 2×2 RMANOVA did the same for DLNO. A Passing-Bablok linear regression⁴² and Bland-Altman Plots⁴³ were used to examine the agreement of the LLN between models. To determine whether there was agreement in determining whether the measured value was below the LLN between models, a Kappa statistic was performed where 1 is less than the LLN and $0 \geq \text{LLN}$. The strength of for the Kappa statistic was: ≤ 0.20 = none; 0.21–0.39 = minimal, 0.40–0.59 = weak; 0.61–0.80 = moderate; ≥ 0.80 –0.90 = strong; ≥ 0.90 almost perfect.⁴⁴

Receiver-operating characteristic (ROC) analysis for evaluating performance DLCO, DLNO and VA between models was also examined.

To classify the impairment in DLNO, DLCO and VA based on z-scores, a linear regression analysis was performed between the average per cent predicted for DLNO, DLCO and VA that correspond to the average

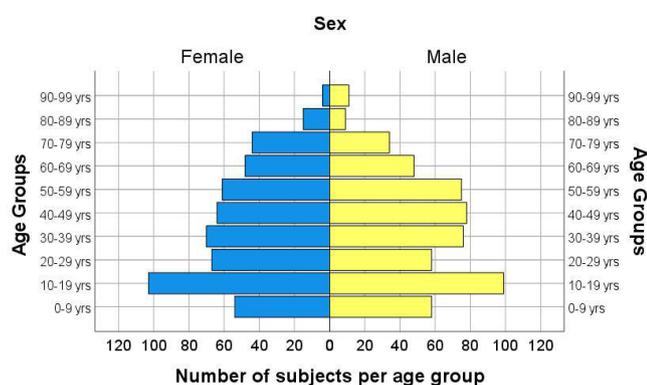


Figure 1 The pooled data used in the analysis display the number of subjects per age group. After removing outliers, 1076 subjects remained for analysis.

fitted z-scores for both models. This would allow an examination of the variability in per cent predicted values matched to z-score classifications.

Patient and public involvement

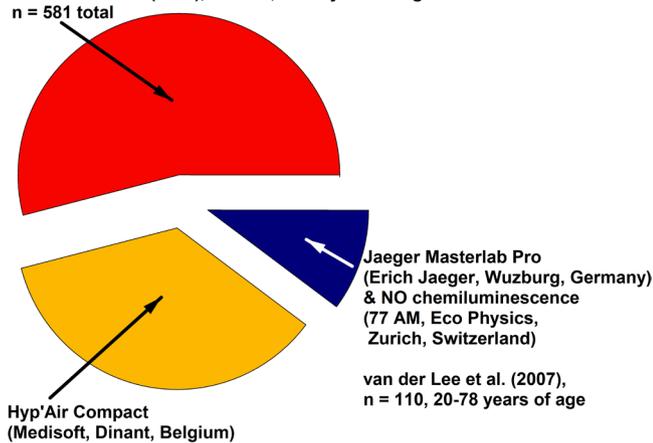
Neither patients nor members of the public were involved in the design, conduct, reporting or dissemination of this research study.

RESULTS

Pooled data from five studies were used to produce reference equations for DLCO, DLNO and VA.^{26–30} Age groups are displayed in figure 1 for a visual representation of the number of subjects in each age category. The five studies used three different pulmonary function machines. The numbers of subjects that were tested on each of these

Jaeger Masterscreen PFT Pro
(Carefusion, Hochberg, Germany)

Thomas et al. (2014), n = 309, 5–17 years of age
Munkholm et al. (2018), n = 272, 20–95 years of age
n = 581 total



Zavrosky et al. (2008), n = 114, 18–85 years of age
Aguilaniu et al. (2008), n = 271, 18–94 years of age
n = 385 total

Figure 2 A representative breakdown of the pooled data and the equipment used in the development of reference equations for pulmonary diffusing capacity.

pulmonary function machines are presented in figure 2. The two Jaeger pulmonary function systems were combined into one pulmonary function system since there were no meaningful differences between them.

Outliers were screened and removed from the analysis. About 7% of the complete dataset was eliminated during initial screening, in which multiple linear regression models were used to examine studentised residuals. Any raw data point that had a studentised residual ≥ 3.0 was eliminated. There were a similar number of males and females with wide age ranges and heights, totalling 1076 never-smokers. Fractional age was not available in the datasets. As DLNO is minimally affected by haemoglobin concentration,⁴⁵ DLNO was not adjusted for haemoglobin concentration. As well, DLCO was not adjusted for haemoglobin concentration since correcting for it does not improve the model fit for DLCO.⁷ There was a 2%–5% shared variance between breath-hold time and DLCO or DLNO (and no shared variance with VA). As such, breath-hold time was also not included as a covariate in the models. The subjects are presented in table 1.

Simulated raw data were created from the anthropometric characteristics of Munkholm *et al.*,³⁰ as that group was unwilling to provide us with the actual raw data. The simulated data represented 24% of the total data set and resembled the actual data (online supplemental tables S1, S2A,B, S3); thus, the simulated data were used in the overall analysis.

Measured DLCO and measured DLNO were highly correlated with each other. The Jaeger MasterScreen Pro produced a correlation of 0.922 between DLCO and DLNO ($R^2=0.85$), and the Hyp'Air Compact produced a correlation of 0.951 between DLCO and DLNO ($R^2=0.90$) (combined R^2 using both machines=0.87). For the Jaeger MasterScreen Pro, $DLNO=4.20 \cdot (DLCO)+8.42$, (adjusted $R^2=0.85$, $p<0.001$, with a residual $SE=14.1$ mL/min/mm Hg). The 95% CI of the slope 4.07 to 4.33. For the Hyp'Air Compact, $DLNO=4.69 \cdot (DLCO)+4.78$, (adjusted $R^2=0.90$, $p<0.001$, with a residual $SE=11.9$ mL/min/mm Hg). The 95% CI of the slope 4.54 to 4.85. Measured VA was correlated to measured DLCO ($r=0.88$, Jaeger MasterScreen Pro; $r=0.80$, Hyp'Air Compact).

The DLNO to DLCO ratio was relatively stable from 5 to 95 years of age (online supplemental figure S1). However, the Jaeger MasterScreen Pro yielded an approximate 0.29 units lower ratio compared with the Hyp'Air Compact due to its systematically larger DLNO values, with DLCO values being relatively unchanged between machine types. Prediction equations for the DLNO to DLCO ratio were not developed as the pulmonary function testing device (6.6% shared variance), altitude (2.2% shared variance), age (1.3% shared variance) and sex (0.6% shared variance) accounted for only 10% of the total shared variance.

Segmented linear reference equations and GAMLSS equations separated by sex are presented in tables 2 and 3. Segmented regression equations that include sex as a

Table 1 Pooled anthropometric data previously published studies from which reference equations were made^{26–30}

	Males (n=546)	Females (n=530)	Combined (n=1076)
Age (years)	38 (23) (5 to 95)	38 (23) (5 to 95)	38 (23) (5 to 95)
Weight (kg)	68.3 (19.5) (18.1 to 110.0)	57.1 (15.0) (14.8 to 101.2)	62.8 (18.3) (14.8 to 110.0)
Height (cm)	170 (17) (105 to 200)	159 (14) (109 to 182)	165 (16) (105 to 200)
Body mass index (kg/m ²)	23.0 (4.0) (14.0 to 35.5)	22.0 (3.9) (12.4 to 39.0)	22.5 (4.0) (12.4 to 39.0)
DLNO (mL/min/mm Hg)	138 (42) (36 to 235)	101 (27) (40 to 179)	120 (40) (36 to 235)
DLCO (mL/min/mm Hg)	29.3 (8.7) (8.5 to 49.9)	21.9 (5.7) (9.1 to 36.8)	25.7 (8.3) (8.5 to 49.9)
VA (L)	5.95 (1.71) (1.60 to 9.22)	4.52 (1.12) (1.70 to 7.50)	5.25 (1.62) (1.60 to 9.22)
KCO mL/min/mm Hg/L	5.0 (0.8) (2.1 to 7.2)	4.9 (0.8) (2.7 to 6.9)	5.0 (0.8) (2.1 to 7.2)
KNO mL/min/mm Hg/L	23.6 (4.0) (9.6 to 34.9)	22.6 (3.4) (10.8 to 31.5)	23.1 (3.7) (9.6 to 34.9)
DLNO/DLCO ratio	4.73 (0.56) (2.92 to 7.63)	4.63 (0.52) (2.64 to 6.98)	4.69 (0.54) (2.64 to 7.63)
Breath-hold time (s)	6.2 (1.4) (4.6 to 10.0)	6.2 (1.3) (4.8 to 10.0)	6.2 (1.3) (4.6 to 10.0)
Altitude of testing (m)	88 (114) (0 to 300)	86 (112) (0 to 300)	87 (113) (0 to 300)

*Mean (SD). Brackets represent ranges. The correlation (Spearman's rho) between height and weight was 0.66 for females and 0.72 for males.

DLCO, pulmonary diffusing capacity for carbon monoxide; DLNO, pulmonary diffusing capacity for nitric oxide; VA, alveolar volume.

covariate are presented in online supplemental table S4. Weight was not a factor in any prediction equation since there was only a 1% shared variance between weight and DLCO or DLNO and 5% shared variance between weight and VA when controlling for height. The influence of the pulmonary function testing (PFT) device on DLCO was minor and therefore was not included in segmented reference equations. The Hyp'Air Compact PFT device produced an approximate 18 mL/min/mm Hg (15%) higher DLNO compared with the Jaeger MasterScreen Pro when all other variables were controlled for (online supplemental table S4). Controlling for all other variables, VA was found to be 0.76 L larger in men compared with females (online supplemental table S4). The Hyp'Air Compact PFT device was also found to produce a 0.28 L (5%) larger VA compared with the Jaeger MasterScreen PFT device. When standardising for the mean height (online supplemental table S10) and PFT device, both models show similar predicted values (figure 3A–C) and similar LLN (figure 4A–C).

Both segmented linear regression and GAMLSS models were fitted to the raw data (online supplemental table S5, online supplemental figures S2–S8). The mean (SD) z-scores of the fitted data in both segmented regression and GAMLSS models were alike. For DLCO and VA

(online supplemental figures S6, S8), the fitted z-scores made by the DLCO and VA GLI GAMLSS reference equations⁷ were affected using the Hyp'Air Compact device. For DLNO, the fitted scores were similar between models and pulmonary function devices used. There were no GLI reference equations made for DLNO. Q–Q plots demonstrate that the fitted z-scores for DLNO, DLCO and VA can be approximated by a normal distribution in both models (online supplemental figures S4 and S5); however, there were some outliers remaining when the per cent predicted values were fitted to the segmented regression models (online supplemental figure S4).

A correlational matrix of fitted z-scores between models shows strong associations in z-scores between models for DLNO and DLCO (online supplemental table S6). The predicted VA obtained from all models is highly associated with the measured VA (online supplemental table S7).

The coefficient of variation between subjects was larger in the segmented regression models at <10 years of age for DLCO, DLNO and VA (figure 5). Segmented regression also had a larger variability for DLNO at >60 years of age (figure 5). The variability was greater in those <10 and >70 years of age when using the segmented regression models (figure 5).

**Table 2** Reference equations using segmented regression

	Estimate	SE	95% CI	Adjusted R ²	RSE
DLCO, females (n=530) (mL/min/mm Hg) Breakpoint=24.3 (95% CI 22.7 to 25.8) years old					
Intercept ₁ (for 5.0–24.2 years old)	-11.82	1.87	-15.5 to 8.1	0.76	2.12 ₁
Intercept ₂ (for 24.3–95.0 years old)	-1.54				3.13 ₂
Age ² ₁ (for 5.0–24.2 years old)	0.01534	0.00183	0.012 to 0.019		
Age ² ₂ (for 24.3–95.0 years old)	-0.0018	0.000081	-0.002 to -0.002		
Height (cm)	0.183	0.014	0.156 to 0.210		
Altitude (m)	0.0041	0.0012	0.002 to 0.006		
DLCO, males (n=546) (mL/min/mm Hg) Breakpoint=22.7 (95% CI 21.2 to 24.0) years old					
Intercept ₁ (for 5.0–22.6 years old)	-15.22	2.3	-19.7 to 10.7	0.80	2.62 ₁
Intercept ₂ (for 22.7–95.0 years old)	2.5				4.35 ₂
Age ² ₁ (for 5.0–22.6 years)	0.0323	0.0038	0.025 to 0.039		
Age ² ₂ (for 22.7–95.0 years old)	-0.00246	0.00011	-0.009 to -0.008		
Height (cm)	0.206	0.017	0.173 to 0.239		
Altitude (m)	0.0041	0.0016	0.001 to 0.007		
DLNO, females (n=530) (mL/min/mm Hg) Breakpoint=22.6 (95% CI 20.6 to 24.5) years old					
Intercept ₁ (for 5.0–22.5 years old)	-66.43	8.4	-82.9 to 50.0	0.79	8.60 ₁
Intercept ₂ (for 22.6–95.0 years old)	-30.74				13.63 ₂
Age ² ₁ (for 5.0–22.5 years old)	0.0616	0.01	0.042 to 0.082		
Age ² ₂ (for 22.6–95.0 years old)	-0.00832	0.00034	-0.028 to 0.011		
Height (cm)	0.947	0.063	0.824 to 1.070		
PFT equipment	15.17	1.31	12.6 to 17.7		
DLNO, males (n=546) (mL/min/mm Hg) Breakpoint=22.2 (95% CI 20.7 to 23.5) years old					
Intercept ₁ (for 5.0–22.1 years old)	-87.15	10.5	-107.7 to 66.6	0.83	11.81 ₁
Intercept ₂ (for 22.2–95.0 years old)	-14.02				19.25 ₂
Age ² ₁ (for 5.0–22.1 years old)	0.1375	0.018	0.103 to 0.173		
Age ² ₂ (for 22.2–95.0 years old)	-0.012	0.00048	-0.013 to -0.011		
Height (cm)	1.086	0.08	0.93 to 1.24		
PFT equipment	18.00	1.83	14.4 to 21.6		
VA, females (n=530) (L) Breakpoint=30.3 (95% CI 28.0 to 32.4) years old					
Intercept ₁ (for 5.0–30.2 years old)	-4.16	0.32	-4.8 to -3.5	0.80	0.39 ₁
Intercept ₂ (for 30.3–95.0 years old)	-2.79				0.58 ₂
Age ² ₁ (for 5.0–30.2 years old)	0.00132	0.00017	0.001 to 0.002		
Age ² ₂ (for 30.3–95.0 years old)	-0.00018	0.00002	-0.0002 to -0.0001		
Height (cm)	0.05	0.0023	0.045 to 0.055		
PFT equipment	0.2545	0.054	0.15 to 0.36		
VA males (n=546) (L) Breakpoint=27.0 (95% CI 25.1 to 28.8) years old					
Intercept ₁ (for 5.0–26.9 years old)	-5.64	0.36	-6.4 to -4.9	0.86	0.46 ₁
Intercept ₂ (for 27.0–95.0 years old)	-3.61				0.73 ₂
Age ² ₁ (for 5.0–26.9 years old)	0.00265	0.0003	0.002 to 0.003		
Age ² ₂ (for 27.0–95.0 years old)	-0.00013	0.00002	-0.0003 to -0.0001		
Height (cm)	0.060	0.0026	0.055 to 0.065		
PFT equipment	0.241	0.07	0.11 to 0.37		

For the PFT equipment, 1=Hyp'Air Compact, and 0=Jaeger Masterscreen. For example, for a man who is 26.9 years old with the same height and equipment used, the predicted alveolar volume (VA) (L)=0.0027·(26.9²)+0.06·(180)+0.24-5.64=7.35 L with a lower limits of normal (LLN) of 7.35 - (0.46·1.645)=6.59 L. For a man 27 years old, 180 cm tall, and who had the measurement performed on the Hyp'Air, the predicted VA (L) = -0.00013·(27²)+0.06·(180)+0.24-3.61=7.34 L with the LLN=7.34 - (0.73·1.645)=6.14 L. DLCO, pulmonary diffusing capacity for carbon monoxide; DLNO, pulmonary diffusing capacity for nitric oxide; RSE, residual SE.

Table 3 Reference equations using generalised additive models of location, scale and shape models

	M=mu, median	(S)=sigma, coefficient of variation, which explains the variability around median	L, lamda, which is the index of skewness
Females (n=530)			
DLCO (mL/min/mm Hg)	$\exp(-4.481 + 1.406 \cdot \ln(\text{height}) + 0.194 \cdot \ln(\text{age}) + 0.0002 \cdot \text{altitude} + \text{Mspline})$	$\exp(0.642 \cdot \ln(\text{age}) - 1.018 \cdot \ln(\text{height}) + \text{Sspline})$	0.325
DLNO (mL/min/mm Hg)	$\exp(-3.777 + 0.144 \cdot \text{machine} + 1.510 \cdot \ln(\text{height}) + 0.3405 \cdot \ln(\text{age}) + \text{Mspline})$	0.1053 for Jaeger Masterscreen, 0.1401 for Hyp'Air	0.836
VA (L)	$\exp(-8.323 + 0.060 \cdot \text{machine} + 1.842 \cdot \ln(\text{height}) + 0.1705 \cdot \ln(\text{age}) + \text{Mspline})$	$\exp(-0.616 \cdot \ln(\text{height}) + 0.2485 \cdot \ln(\text{age}))$	0.577
Males (n=546)			
DLCO (mL/min/mm Hg)	$\exp(-5.163 + 1.500 \cdot \ln(\text{height}) + 0.3507 \cdot \ln(\text{age}) + 0.0002 \cdot \text{altitude} + \text{Mspline})$	$\exp(8.365 + 0.914 \cdot \ln(\text{age}) - 2.503 \cdot \ln(\text{height}) + \text{Sspline})$	0.632
DLNO (mL/min/mm Hg)	$\exp(-4.339 + 0.138 \cdot \text{machine} + 1.617 \cdot \ln(\text{height}) + 0.410 \cdot \ln(\text{age}) + \text{Mspline})$	$\text{Exp}(0.230 \cdot \text{machine} - 2.191)$	1.113
VA (L)	$\exp(-9.443 + 0.0569 \cdot \text{machine} + 2.076 \cdot \ln(\text{height}) + 0.169 \cdot \ln(\text{age}) + \text{Mspline})$	0.1016	0.0635

Height is in cm, age in years; Machine=1 for Hyp'Air Compact and 0 for the Jaeger Masterscreen; lower limits of normal (fifth percentile)= $\exp(\ln(M) + \ln(1 - 1.645 \cdot L \cdot S)/L)$; Per cent predicted = (measured/M)·100; Z-score = ((measured value/M)^L - 1)/(L·S); exp ()=natural exponential; ln()=natural logarithm; Mspline and Sspline correspond to the age-varying coefficients provided in the supplementary materials. Model is valid from ages 5–95 years of age and an altitude of 0–300 m. Note: If pulmonary diffusing capacity for carbon monoxide (DLCO) is measured at an altitude that is more than 300 m, we recommend converting the measured DLCO to sea level first, based on the data by Gray *et al.*⁸⁰ and then omitting the altitude covariate from the equation (as the converted DLCO will be at an altitude of 0 m). Adjusted DLCO to sea level (mL/min/mm Hg)=measured DLCO at altitude·(0.505+0.00065·barometric pressure in mm Hg at altitude). The formula to estimate barometric pressure at altitude in mm Hg is: $760 \cdot \exp(-0.284 \cdot \text{altitude in m} / (8.314 \cdot \text{Temperature in Kelvin}))$, where Kelvin = °C+273.15. (see: <https://planetcalc.com/938/>).

DLCO, pulmonary diffusing capacity for carbon monoxide; DLNO, pulmonary diffusing capacity for nitric oxide; VA, alveolar volume.

Both models had similar prediction accuracies (table 4). There was no clear model winner. Both models were comparable as the 95% CI of improvement overlapped zero for all cases. The average correlation coefficients of the predicted values associated with the actual values were similar between the two models (table 5).

There was a moderate agreement for DLCO and DLNO between both models (table 6). In the same vein, the Youden Index J (sensitivity+specificity–1) was determined from ROC analyses and described the overall diagnostic accuracy⁴⁶ (table 7). Diagnostic accuracy was the highest for DLCO, then DLNO and then VA when segmented regression was compared against GAMLSS, given that the estimated prevalence of abnormal results (values below the LLN) is 5%. Another ROC analysis was performed comparing DLNO to DLCO when the estimated prevalence of abnormal results is 5% in a population (online supplemental table S8). All characteristics of the ROC curve were similar between models for DLCO and DLNO (online supplemental table S8).

The derived LLN obtained from segmented linear regression models was compared with the derived LLN from GAMLSS models (online supplemental figures S9–S14). There were systematic and proportional differences between models.

The impairment in DLNO, DLCO and VA was classified based on z-scores (table 8). As the per cent predicted matched to the LLN (z-score = – 1.645) varies with age

(online supplemental table S9), and throughout a wide range of z-score values (online supplemental figure S15A,B), the classification of impairment is best defined via the z-scores. However, the per cent predicted value along with its variability is also provided in table 8 as it not only may be more intuitive than z-scores, but it can be an easier way for clinicians to assess the severity of a pulmonary function abnormality.

DISCUSSION

GAMLSS have been used by the GLI Network to develop reference equations for lung function for the world to use,^{6–9} but they are too complicated to implement (see online supplemental table S11 for a worked example). The first purpose of this study was to examine the accuracy of complicated GAMLSS models compared with simpler segmented (piecewise) linear regression models when developing reference equations for pulmonary diffusing capacity. We showed that segmented regression models are comparable to GAMLSS models in terms of prediction accuracy (tables 4 and 5). When identifying subjects below the LLN, there was a 61% and 66% true positive rate for DLCO and DLNO, respectively, when segmented regression was compared with GAMLSS, for which the estimated prevalence of abnormal results is 5% (table 7).

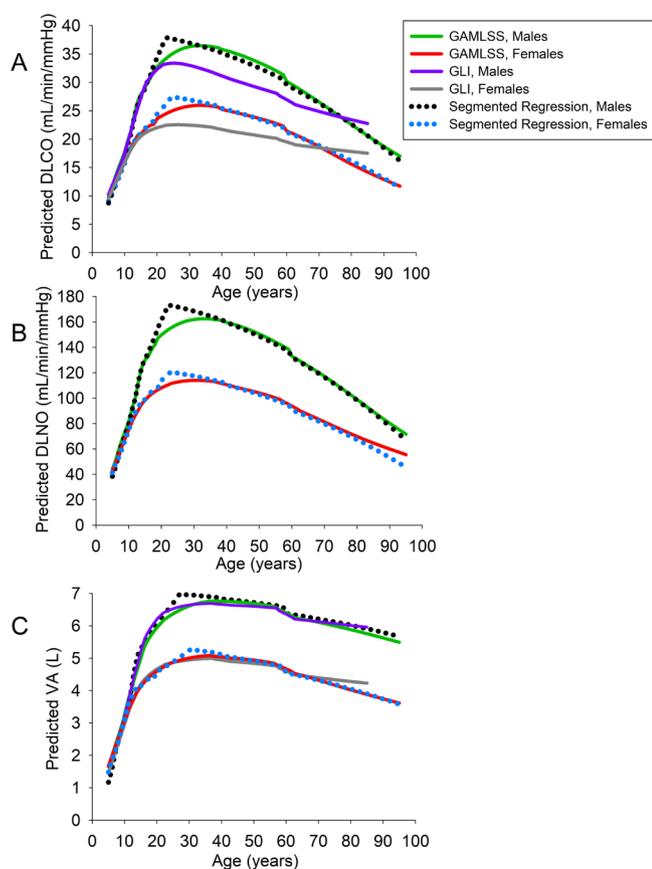


Figure 3 (A) predicted pulmonary diffusing capacity for carbon monoxide (DLCO) versus age, (B) predicted pulmonary diffusing capacity for nitric oxide (DLNO) versus age, (C) predicted alveolar volume (VA) versus age. The various fitted curves/lines are based on the median height for age and sex in the white US population,⁴⁰ an altitude of 0m and the Jaeger MasterScreen Pro equipment was used. Online supplemental table S10 in the supplement lists the heights with each age and sex. For DLCO and VA, the updated Global Lung Function Initiative (GLI) generalised additive models of location, scale and shape (GAMLSS) reference equations were included as a comparison.⁸ Notice that the (GLI) DLCO curves (*grey* in females, and *purple* in males) are lower compared with both GAMLSS and segmented regression models. The GLI GAMLSS prediction model is based on a 10s breath-hold, which allows for a more homogenous inspired gas penetration in the lung, and thus a lower DLCO compared with the 5–6s breath-hold manoeuvres. The GAMLSS and segmented linear regression curves/lines for DLCO, DLNO and VA are comparable.

When evaluating reference equations for lung function indices, there are limited studies comparing regression to GAMLSS. All the comparisons involve comparing FVC and FEV₁ between models, and none compared pulmonary diffusing capacity. Martinez-Briseño *et al*¹⁷ compared spirometric reference equations between similar models and determined that while GAMLSS displayed a slightly better fit over multiple linear regression, they were minimal. Brisman *et al*¹⁸ used a piecewise regression approach as discussed by Lubiński and Gólczewski,⁴⁹ and

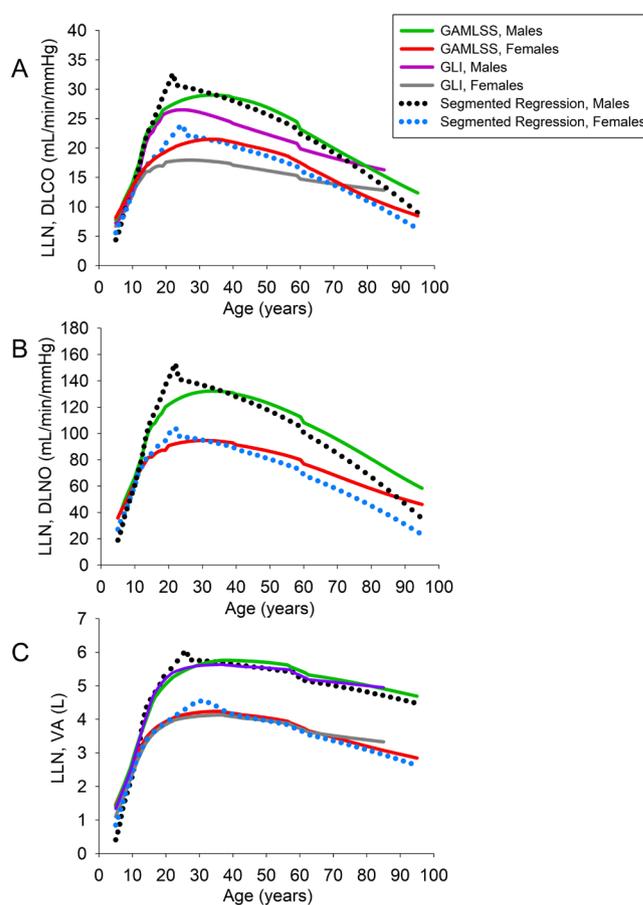


Figure 4 (A) pulmonary diffusing capacity for carbon monoxide (DLCO) versus age at the lower limits of normal (LLN), (B) pulmonary diffusing capacity for nitric oxide (DLNO) versus age at the LLN, (C) alveolar volume (VA) versus age at the LLN. The various fitted curves/lines are based on the median height for age and sex in the white US population,⁴⁰ an altitude of 0m, and the Jaeger MasterScreen Pro equipment was used. Online supplemental table S10 in the supplement lists the heights with each age and sex. For DLCO and VA, the updated Global Lung Function Initiative (GLI) generalised additive models of location, scale and shape (GAMLSS) reference equations were included as a comparison.⁸ Notice that the (GLI) DLCO curves (*grey* in females, and *purple* in males) are lower compared to both GAMLSS and segmented regression models. The GLI GAMLSS prediction model is based on a 10s breath-hold, which allows for a more homogenous inspired gas penetration in the lung, and thus a lower DLCO compared with the 5–6s breath-hold manoeuvres. The segmented linear regression lines for DLNO and DLNO tend to show a lower LLN compared with the GAMLSS models, especially after 60 years of age for DLNO and after 80 years of age for DLCO.

that the mean square errors of the models were similar to GAMLSS developed the GLI. In a follow-up study by Brisman *et al*,⁵⁰ they further determined that segmented linear regression should be used for the development of spirometric reference equations as the GLI GAMLSS equations identified too few subjects below the LLN.⁵⁰ Kubota *et al* also compared multiple linear regression

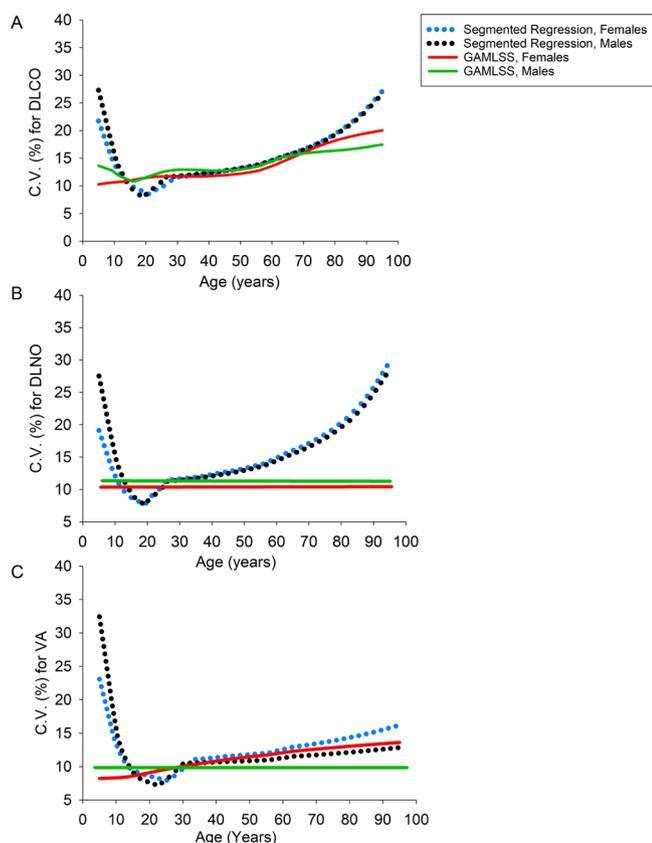


Figure 5 Between subject variability across age for (A) pulmonary diffusing capacity for carbon monoxide (DLCO) versus age, (B) pulmonary diffusing capacity for nitric oxide (DLNO) versus age, (C) alveolar volume (VA) versus age. The C.V. (%) = (predicted SD/predicted mean) \times 100. The predicted mean and SD for each age were calculated using the median height of the US population at each age.⁴⁰ The graph is also standardised for altitude (0 m). For DLCO and VA, the between subject variability unaffected by pulmonary function testing device. However, when measuring DLNO in females, those that were tested with the Hyp'Air Compact device showed a between subject variation of 14% compared with 10.5% in males throughout all ages (not shown). When using the Jaeger MasterScreen Pro, the between subject variation for DLNO was similar (panel B), 10.5%, females, 11.2% males). GAMLSS, generalised additive models of location, scale and shape.

against GAMLSS for FVC and FEV₁ in Japanese subjects.⁵¹ In that study, they claimed that their GAMLSS models more accurately reflected the transition in pulmonary function during young adulthood. However, they did not provide any information on prediction accuracy between models, nor did the study include children, and there was no real transition between adolescence and adulthood. Therefore, the results of this current study are particularly novel as we show similarity in prediction errors DLCO, DLNO and VA between GAMLSS and segmented linear regression.

Nevertheless, the Q–Q plots for per cent predicted generated by GAMLSS demonstrate a better fit to the normal distribution compared with segmented regression

at the extreme ends of the plot. The Q–Q plot for VA, for example, shows that when the observed values are $\geq 140\%$ predicted, the expected normal value is much different; hence about 12 values deviate off the linear line (online supplemental figure S4). Similarly, there are 1–2 subjects for DLCO and DLNO in which the expected normal value was much different compared with the observed per cent predicted values. In comparison, there were no subjects that strayed off the per cent predicted Q–Q plot line when GAMLSS were used for DLNO, DLCO or VA, even at the extreme ranges (online supplemental figure S5). However, these instances are rare ($\leq 1\%$ of the subject pool), and when comparing models (table 4), the overall prediction accuracies were similar.

As the validity of different reference sets for DLNO has been questioned,^{23 24} the second purpose of this study was to update predictions equations from the ERS 2017 Technical standards document²² based on more available data so that between-machine comparisons could be verified. We confirmed that the Hyp'Air Compact measured DLNO values that were larger than the Jaeger MasterScreen Pro by 16–20 mL/min/mm Hg (13%–16%) (online supplemental table S4). These data agree with another study that demonstrated similar findings, although the differences between devices were slightly larger, at 22–26 mL/min/mm Hg (17%).²⁵ The slightly lower difference between devices observed in the current study is because our models include children, and their study did not. This study pooled all the available reference sets for DLNO that were published in the literature for white subjects from Europe^{27–30} and North America²⁶ and confirmed a systematic increase in DLNO when the Hyp'Air Compact PFT device was used. The pooled data also demonstrate a 0.2–0.4 L (6%–8% larger VA when the Hyp'Air Compact was used, which is slightly smaller than the between machine differences from Radtke *et al.*³⁵ The discrepancy estimating VA and the rates of alveolar uptake for nitric oxide per unit time and pressure (KNO) between the two systems could explain the discrepancy in DLNO and VA between devices.³⁵ Furthermore, as the Jaeger MasterScreen Pro uses a demand valve, whereas the Hyp'Air Compact uses a reservoir bag from which the mixture of gases is inspired, this would alter the expired to inspired nitric oxide ratio.

The results presented here are concerning since the lung function testing device is now an important covariate to consider when measuring DLNO and VA. A 2017 ERS Task Force Report on the standardisation of DLNO²² presented reference equations based on pooled data of three studies. However, about 75% of the pooled data from those three studies were based on DLNO data determined by the Hyp'Air Compact PFT system^{26–28}; yet 36% of the current pooled data was determined by the Hyp'Air Compact device. Thus, the results present a more balanced view of the between device findings, and we have updated the prediction equations here.

This study did not determine which pulmonary function testing device was more accurate, only that the two

**Table 4** Prediction accuracy between both models

	GAMLSS models				Segmented linear regression				Per cent improvement (95% CI)	
	AIC	BIC	Median	Range	AIC	BIC	Median	Range		
Males										
DLNO	4602	4649	17.7	15.3–20.5	4667	4697	17.4	15.2–19.9	–2% (–7% to 3%)	
DLCO	2984	3053	4.0	3.4–4.8	3040	3070	3.9	3.3–4.7	–2% (–7% to 3%)	
VA	959	1002	0.64	0.53–0.72	1082	1112	0.65	0.55–0.75	–2% (–4% to 8%)	
Females										
DLNO	4104	4151	12.1	10.0–13.8	4161	4191	12.3	10.1–14.0	2% (–1% to 5%)	
DLCO	2538	2602	2.8	2.4–3.3	2608	2638	2.8	2.4–3.4	1% (–2% to 5%)	
VA	670	717	0.50	0.39–0.59	788	818	0.50	0.41–0.61	4% (–1% to 9%)	

Note: a better model fit is usually indicated by a lower Akaike information criterion (AIC) or Bayesian information criterion (BIC). Thus, it may seem that generalised additive models of location, scale and shape (GAMLSS) are a better fit to the data. However, notice that this may not be correct. Under the per cent of improvement column, a positive percentage suggests that GAMLSS is the better model, a negative percentage value suggests segmented linear regression is the better model. One can see that both models are comparable because the 95% CI of the per cent of improvement overlaps zero. The 95% CI was developed after 100 random samplings of 80% of the pooled data. Under the Median and Range columns, the square root of the average of the squared errors is presented after 100 samplings of 80% of the pooled data.

DLCO, pulmonary diffusing capacity for carbon monoxide; DLNO, pulmonary diffusing capacity for nitric oxide; VA, alveolar volume.

devices were different. For us to determine which is a more accurate device, a comparison would have to be made against a gold standard device. Chemiluminescence NO analysers are considered the gold standard of NO analysers, but it is highly costly. Even so, van der Lee *et al* used a nitric oxide chemiluminescence analyser (along with the Jaeger Masterlab Pro system) in its development of reference equations for DLNO.²⁸ Our analysis showed no meaningful differences between DLNO measured by van der Lee *et al*²⁸ versus the studies that used the Jaeger MasterScreen PFT Pro with the NO electrochemical cell.³⁰ However, both the Jaeger Masterlab Pro system (with NO chemiluminescence) and Jaeger MasterScreen PFT Pro displayed lower DLNO values than the Hyp'Air Compact system.^{26 27} This would suggest that (1) either the Jaeger MasterScreen PFT Pro provides more accurate diffusing capacity values or

(2) the software calculations provided by Jaeger were different compared with the calculations of the Hyp'Air Compact device.

We also examined agreement between models using a kappa statistic and a ROC analysis. The kappa statistic showed moderate agreement between models for DLCO and DLNO and a weak agreement for VA (table 6). When comparing against GAMLSS, segmented regression demonstrated $\geq 97\%$ specificity (true negative rate) when the prevalence of an abnormal result in a population is 5% (ie, when 5% of the population is below the LLN). Moreover, when compared against GAMLSS, segmented regression was able to identify 75% of abnormal results for DLCO, 64% of abnormal results for DLNO, and 52% of abnormal results for VA, considering the prevalence of abnormal results in a population is 5%. This is termed the true positive rate.

Table 5 The correlation coefficients from the 100 samplings are compared between both models

	GAMLSS		Segmented regression	
	Average	95% CI	Average	95% CI
Males				
DLNO	0.91	0.88 to 0.93	0.91	0.88 to 0.94
DLCO	0.89	0.86 to 0.92	0.90	0.86 to 0.92
VA	0.93	0.91 to 0.95	0.92	0.90 to 0.94
Females				
DLNO	0.89	0.85 to 0.92	0.89	0.85 to 0.92
DLCO	0.87	0.83 to 0.90	0.86	0.82 to 0.90
VA	0.89	0.85 to 0.93	0.88	0.84 to 0.92

Eighty per cent of the pooled data was sampled 100 times, and the remaining 20% was used to test the fit of each model 100 times. DLCO, pulmonary diffusing capacity for carbon monoxide; DLNO, pulmonary diffusing capacity for nitric oxide; GAMLSS, generalised additive models of location, scale and shape; VA, alveolar volume.

Table 6 A breakdown of the percentage of subjects below the lower limits of normal (LLN), including the agreement between the two models for each variable

	DLCO	DLNO	VA
Number and percentage of the fitted data below the LLN (z score < -1.645)			
GAMLSS	60 (5.7%)	81 (7.5%)	54 (5.0%)
Segmented linear regression	71 (6.6%)	57 (5.3%)	40 (3.7%)
Percentage below the LLN by age group			
GAMLSS (5–49 years of age) (n=727)	42 (5.8%)	28 (3.9%)	37 (5.1%)
GAMLSS (50–95 years of age) (n=349)	18 (5.2%)	33 (9.5%)	17 (4.9%)
Segmented linear regression (5–49 years of age) (n=727)	53 (7.2%)	39 (5.4%)	31 (4.3%)
Segmented linear regression (50–95 years of age) (n=349)	18 (5.2%)	18 (5.2%)	9 (2.6%)
Agreement between the two models (Kappa statistic)	0.67 [0.57 to 0.76]	0.64 [0.54 to 0.74]	0.58 [0.46 to 0.70]

Agreement between models for each variable was determined by the Kappa statistic where 1 is less than the LLN and $0 \geq \text{LLN}$. Strength of agreement: ≤ 0.20 =none; 0.21–0.39=minimal, 0.40–0.59=weak; 0.61–0.80=moderate; ≥ 0.80 –0.90=strong; ≥ 0.90 almost perfect.⁴⁴ Brackets represent the 95% CI of the Kappa statistic.

DLCO, pulmonary diffusing capacity for carbon monoxide; DLNO, pulmonary diffusing capacity for nitric oxide; GAMLSS, generalised additive models of location, scale and shape; VA, alveolar volume.

The precision between both models was between 61% and 70%. That is, the probability that an actual abnormal result (ie, <LLN) identified by GAMLSS will also show an abnormal result using segment regression (aka precision) varies between 61% and 70% when the

prevalence of abnormal results in a population is 5% (table 7). Is this acceptable? Well, we must consider the week-to-week variability in pulmonary diffusing capacity. For example, the week-to-week variability (reproducibility) of DLCO is at least 3.8 mL/min/mm Hg in those

Table 7 Receiver-operating characteristic (ROC) analysis for evaluating the performance of both statistical models for pulmonary diffusing capacity for carbon monoxide (DLCO), pulmonary diffusing capacity for nitric oxide (DLNO) and alveolar volume (VA) when the estimated prevalence of an abnormal result in a population is 5% (ie, when 5% of the population is below the lower limits of normal (LLN))

	DLCO	DLNO	VA
Area under the ROC curve (AUC)	0.86 (0.84, 0.88)	0.81 (0.79, 0.83)	0.75 (0.73, 0.78)
Youden's J statistic	0.72 (0.59, 0.82)	0.62 (0.50, 0.72)	0.51 (0.38, 0.64)
Sensitivity	0.75 (0.62, 0.85)	0.64 (0.51, 0.76)	0.52 (0.38, 0.66)
Specificity	0.97 (0.96, 0.98)	0.98 (0.97, 0.99)	0.99 (0.98, 0.99)
Positive predictive value	61% (51%, 70%)	66% (54%, 76%)	70% (56%, 81%)
Negative predictive value	99% (98%, 99%)	98% (97%, 99%)	98% (97%, 98%)
Positive likelihood ratio	29.3 (19.5, 44.0)	34.5 (22.0, 59.2)	44.2 (23.8, 82.0)
Negative likelihood ratio	0.26 (0.17, 0.40)	0.37 (0.26, 0.51)	0.49 (0.37, 0.64)

Youden's J statistic (sensitivity+specificity–1): measures the effectiveness of using the segmented regression models as a diagnostic test compared with generalised additive models of location, scale and shape (GAMLSS).

Sensitivity (true positive rate): probability of an abnormal result (ie, DLCO, DLNO or VA<LLN) as identified by segmented regression models when GAMLSS also show an abnormal result.

Specificity (true negative rate): probability of a normal test result (ie, DLCO, DLNO or VA \geq LLN) as identified by segmented regression when GAMLSS also show a normal test result (\geq LLN) in that same variable.

Positive predictive value (precision): probability of an abnormal result (<LLN) in one variable as identified by GAMLSS when that same variable also shows an abnormal result as identified by segmented regression.

Negative predictive value: probability of a normal result (\geq LLN) in one variable as identified by GAMLSS when segmented regression also show a normal result in that same variable.

Positive likelihood ratio (true positive rate \div false positive rate): the ratio between the probability of an abnormal result (<LLN) identified by segmented regression given an abnormal test result (<LLN) as identified by GAMLSS and the probability of a normal test result as identified by segmented regression given a normal test result as identified by GAMLSS.

Negative likelihood ratio (false-negative rate \div true negative rate): the ratio between the probability of a normal test result identified by segmented regression (\geq LLN) when there is an abnormal test result as identified by GAMLSS (<LLN) and the probability of a normal test result as identified by segmented regression given a normal result as identified by GAMLSS.

AUC, The per cent chance that when GAMLSS is used detect abnormal results (values<LLN), segmented regression can also distinguish abnormal results in the same patient. Parentheses represent the 95% bootstrapped CI.

**Table 8** Classification of impairment in pulmonary diffusing capacity for nitric oxide, pulmonary diffusing capacity for carbon monoxide and alveolar volume, using the modified one-step NO–CO technique (4–6 s breath-hold manoeuvres)

	Severe decrease	Moderate decrease	Mild decrease	Normal	Increased
Suspected or prior evidence of lung disease					
z-score	–5.01 and below	–5.00 to –3.51	–3.50 to –1.65	–1.645 to +1.645	>+1.645
% Predicted	≤41 (3)%	42 (3) to 59 (3)%	60 (3) to 80 (3)%	81 (3) to 119 (3)%	>119 (3)%
Screening and case finding purposes only					
z-score	–5.01 and below	–5.00 to –3.51	–3.50 to –1.961	–1.96 to +1.96	>+1.96
% Predicted	≤41 (3)%	42 (3) to 59 (3)%	60 (3) to 77 (3)%	78 (3) to 123 (3)%	>123 (3)%

The z-scores should ultimately be used for the classification of diffusion impairment or low alveolar volume. The advantage of using z-scores to define the lower limits of normal (LLN) (as opposed to per cent predicted) is that the z-scores apply to all populations. However, the per cent (%) predicted may be more intuitive than z-scores, and the % predicted may be an easier way for clinicians to assess the severity of a pulmonary function abnormality. Nevertheless, there is a large SD of 2.8% (rounded to 3%) for each per cent predicted value that is matched to each z-score category, and the SD is depicted within the parentheses. The LLN is normally at the fifth percentile ($z = -1.645$). However, if interpreting multiple related lung function tests, there is an increase in false-positive rates when using the fifth percentile.⁸¹ As such, the LLN at a z-score of -1.96 is recommended for screening and case-finding purposes.⁶

with a cardiopulmonary disease^{52–54} and at least 3.6 mL/min/mm Hg in healthy individuals,^{53 55–57} but the differences in the LLN for DLCO between both models that equal to or exceeds 3.6 mL/min/mm Hg occurred in only 1.7% of the pooled data (18/1076). Furthermore, the 95% CI of the SD of the residuals multiplied by two between both models is less than its reproducibility (ie, <3.6 mL/min/mm Hg) (online supplemental figures S9A,B, S10A,B). Thus, the ability to classify subjects below the LLN for DLCO using either model is similar when considering the intersession variability in DLCO. What about the ability of the two models to identify the LLN for DLNO? Since the week-to-week variability of DLNO is approximately 13 mL/min/mm Hg in those with cardiopulmonary disease,^{53 54} and around 20 mL/min/mm Hg in healthy individuals,^{55 56} the differences in the LLN between for DLNO models that were equal to or exceeded 13 mL/min/mm Hg occurred in 10% of the pooled data (111/1076) (online supplemental figures S11A,B, S12A,B). Additionally, the differences in the LLN between DLNO models that equal to or exceed 20 mL/min/mm Hg occurred in only 3% of the pooled data (34/1076). Thus, the ability to classify the LLN for DLNO using either model is similar when considering the intersession variability in DLNO.

Segmented (piecewise) regression makes a series of assumptions: linearity (the relationship between X and the mean of Y is linear), homoscedasticity (the variance of the residual is the same for any value of X), independence (observations are independent of each other) and normality (for any fixed value of X, Y is normally distributed). Overall, there was linearity (table 2, (online supplemental table S4), homoscedasticity (online supplemental figure S2), independence (each subject is tested only once) and normality (online supplemental figure S4). Still, the Q–Q plot for the per cent predicted VA from segmented regression is not perfect; it has about 10 outliers (online supplemental figure S4).

Establishing categories on diffusion impairment based solely on per cent predicted values, as reported back in 2005,⁵⁸ is not appropriate. The LLN as expressed as a percentage of the predicted value changes with age for several different lung function indices, such as FEV₁,⁶ FVC⁶ and DLCO,⁷ and we have demonstrated this to be true for DLCO, DLNO and VA (online supplemental table S9). As such, the z-scores should be used to define the severity of diffusion impairment (table 8). Nevertheless, the per cent predicted value along with its variability is also provided in table 8 as it may be more intuitive than z-scores, and it could be an easier way for clinicians to assess the severity of a pulmonary function abnormality.

In the past, an upper limit of normal (ULN) was not formally established for spirometry⁶ because high values are not clinically meaningful. A ULN for DLCO was also not established when the 2017 GLI DLCO reference equations were published.⁷ Nonetheless, abnormally high DLCO values may be pathologic, even though they are rare. In those rare cases where high values are seen (ie, pulmonary haemorrhage, polycythaemia, obesity, asthma),^{59–64} the DLCO test is not the standard for diagnosis.⁷ Even so, we believe that there is a role for the ULN moving forward, and we incorporate it here. An increase in DLNO and DLCO above the ULN may not be pathologic; an increase could be caused by a negative intrathoracic pressure during breath-hold (Müller manoeuvre),^{65 66} or lung size may be very large. In fact, most nationally ranked swimmers have DLNO and DLCO above the ULN,⁶⁷ which are strongly associated with their large lung volumes. Regular swimming training throughout childhood and adolescence may have aided in the development of larger lungs,^{68–70} although this remains controversial.⁷¹ There also may be a selection bias such that those with larger lungs become good swimmers. Furthermore, given an apparent association of DLNO and DLCO with cardiorespiratory fitness,⁷² setting a ULN could identify individuals with supra-normal diffusing

capacities. As such, we agree with Quanjer *et al.*⁶ in that for those individuals suspected of lung disease, an LLN of the fifth percentile ($z = -1.645$) should be used; and if lung function testing is for screening and fact-finding purposes only, a value of the 2.5th and 97.5th percentile should be used (z -scores of ± 1.96). Nonetheless, this classification of diffusion impairment in [table 8](#) does not necessarily correlate with symptomatology, mortality and/or morbidity.

There are reasons why the shared variance between DLNO and DLCO is not 100%. Approximately 70%–80% of the barrier to carbon monoxide uptake resides within the red cell (ie, red cell resistance), while the remaining 25% or so is in the alveolar membrane (see [figure 1](#) elsewhere²²). In contrast, the main barrier to NO uptake resides between the alveolar and red cell membranes (about 60%) (ie, membrane resistance).⁷³ Thus, DLNO is better represented by gas transfer through the alveolar-capillary membrane compared with DLCO, and DLNO is more affected by changes in lung volume.⁷³ Thus, DLNO provides a more sensitive evaluation of fibrotic changes in the lung compared with DLCO, and DLCO provides a more sensitive evaluation of pulmonary vascular disorders than DLNO. Unlike DLCO, DLNO is relatively unaffected by changes in haemoglobin concentration⁴⁵ or carboxyhaemoglobin concentration.⁷⁴

From the pooled data in this study of non-diseased subjects, 88% of the variance in DLNO is shared by DLCO yet DLNO z -scores share about 39%–47% of the variance in DLCO z -scores (online supplemental table S6). Indeed, it seems logical that measuring DLNO and DLCO together would provide a better assessment of a patient's pulmonary condition than measuring either one of them on its own since approximately 53%–61% of the total variance between the fitted DLCO z -scores and fitted DLNO z -scores are not shared. Moreover, the fact that there is a low true positive rate and a low positive predictive value between DLNO and DLCO when the prevalence of an abnormal result is 5% further demonstrates that DLNO and DLCO measure different things, even though there is considerable overlap (online supplemental table S8). Regardless of whether segmented linear regression or GAMLSS models are used in predicting DLNO and DLCO, there is only a 38%–42% probability that when DLNO is abnormal ($<LLN$), DLCO is also abnormal (online supplemental table S8). Thus, it behoves us to measure both DLNO and DLCO together to better understand a patient's potential lung pathology.

The current GLI DLCO reference equations⁷⁸ and the reference equations updated here for DLNO, DLCO and VA are for white subjects only. As there are slight but essential differences in DLNO,^{75 76} DLCO^{77 78} and VA between various ethnic groups,^{75 76} it is crucial to develop multi-ethnic reference equations⁷⁹. For example, lung disease could be overdiagnosed by about 8% in the black population if reference equations for white subjects were used.⁷⁵ This false-positive misdiagnosis could increase patient

stress, and healthcare resources would be extended, resulting in a higher cost for a non-illness.⁷⁵

In conclusion, when developing pulmonary function reference equations, we propose that segmented (piece-wise) linear regression can be used instead of GAMLSS due to its simplicity, especially when overall prediction errors are similar between the two types of models. Still, the Q–Q plots of observed versus expected per cent predicted reveals a better fit to the normal distribution when GAMLSS models are used, but only at the upper end of per cent predicted (ie, $\geq 140\%$ predicted), and these were rare occurrences. These reference equations for DLNO, DLCO and VA developed here are robust and should be used moving forward for any clinical assessment that uses the NO–CO double diffusion technique and breath-hold time of about 6 seconds. Since the Hyp'Air Compact device measures DLNO and VA that is systematically higher than that of the Jaeger MasterScreen Pro, we urge the two manufacturers to come together to resolve these differences.

Acknowledgements GSZ and JC thank the authors that provided the data for the development of these updated reference equations.

Contributors Conception and design (GSZ); statistical analyses (JC primary, GSZ secondary); interpretation (all authors); guarantor (GSZ). All authors edited and approved the final version of the manuscript.

Funding JC received monetary compensation for developing the GAMLSS and segmented regression models.

Competing interests GSZ is a Global Lung Function Initiative Network member, which published reference equations for pulmonary diffusing capacity and static lung volumes using GAMLSS. GSZ is the current co-chair of the European Respiratory Society Task Force on the interpretation of pulmonary diffusing capacity for nitric oxide.

Patient consent for publication Not applicable.

Ethics approval This study does not involve human participants.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The pooled data datasets used in this current study are available from the corresponding author [GSZ] on reasonable request. It is required that should the complete dataset be shared, then any abstract, conference proceedings, or article that will be published related to this dataset will have GSZ as one of its co-authors.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Gerald Stanley Zavorsky <http://orcid.org/0000-0002-4473-1601>

REFERENCES

- 1 Pistelli F, Bottai M, Viegi G, *et al.* Smooth reference equations for slow vital capacity and flow-volume curve indexes. *Am J Respir Crit Care Med* 2000;161:899–905.

- 2 Stanojevic S, Wade A, Stocks J, *et al.* Reference ranges for spirometry across all ages: a new approach. *Am J Respir Crit Care Med* 2008;177:253–60.
- 3 Cooper BG, Stocks J, Hall GL, *et al.* The global lung function initiative (Gli) network: bringing the world's respiratory reference values together. *Breathe* 2017;13:e56–64.
- 4 Hall GL, Stanojevic S, *et al.* GLI Network Executive. The global lung function initiative (Gli) network ERS clinical research collaboration: how international collaboration can shape clinical practice. *Eur Respir J* 2019;53:1802277.
- 5 Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992;11:1305–19.
- 6 Quanjer PH, Stanojevic S, Cole TJ, *et al.* Multi-Ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012;40:1324–43.
- 7 Stanojevic S, Graham BL, Cooper BG, *et al.* Official ERS technical standards: global lung function initiative reference values for the carbon monoxide transfer factor for Caucasians. *Eur Respir J* 2017;50:1700010.
- 8 "Official ERS technical standards: Global Lung Function Initiative reference values for the carbon monoxide transfer factor for Caucasians." Sanja Stanojevic, Brian L. Graham, Brendan G. Cooper, Bruce R. Thompson, Kim W. Carter, Richard W. Francis and Graham L. Hall on behalf of the Global Lung Function Initiative T_{lco} working group. *Eur Respir J* 2017; 50: 1700010. *Eur Respir J* 2020;56:1750010.
- 9 Hall GL, Filipow N, Ruppel G, *et al.* Official ERS technical standard: global lung function initiative reference values for static lung volumes in individuals of European ancestry. *Eur Respir J* 2021;57:2000289.
- 10 Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape (with discussion). *J Royal Statistical Soc C* 2005;54:507–54.
- 11 Stasinopoulos MD, Rigby RA, Bastiani FD. GAMLSS: a distributional regression approach. *Stat Modelling* 2018;18:248–73.
- 12 Vaccari F, Fiori F, Bravo G, *et al.* Physical fitness reference standards in Italian children. *Eur J Pediatr* 2021;180:1789–98.
- 13 Albrecht BM, Stalling I, Bammann K. Sex- and age-specific normative values for handgrip strength and components of the Senior Fitness Test in community-dwelling older adults aged 65–75 years in Germany: results from the OUTDOOR ACTIVE study. *BMC Geriatr* 2021;21:273.
- 14 Li K, Hu L, Peng Y, *et al.* Comparison of four algorithms on establishing continuous reference intervals for pediatric analytes with age-dependent trend. *BMC Med Res Methodol* 2020;20:136.
- 15 Debele SE, Bogdanowicz E, Strupczewski WG. Around and about an application of the GAMLSS package to non-stationary flood frequency analysis. *Acta Geophys.* 2017;65:885–92.
- 16 Khondoker MR, Glasbey CA, Worton BJ. A comparison of parametric and nonparametric methods for normalising cDNA microarray data. *Biom J* 2007;49:815–23.
- 17 Timmerman ME, Voncken L, Albers CJ. A tutorial on regression-based norming of psychological tests with GAMLSS. *Psychol Methods* 2021;26:357–73.
- 18 Sherrill DL, Lebowitz MD, Knudson RJ, *et al.* Methodology for generating continuous prediction equations for pulmonary function measures. *Comput Biomed Res* 1991;24:249–60.
- 19 Sherrill DL, Lebowitz MD, Knudson RJ, *et al.* Continuous longitudinal regression equations for pulmonary function measures. *Eur Respir J* 1992;5:452–62.
- 20 Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;159:179–87.
- 21 Paoletti P, Viegi G, Pistelli G, *et al.* Reference equations for the single-breath diffusing capacity. A cross-sectional analysis and effect of body size and age. *Am Rev Respir Dis* 1985;132:806–13.
- 22 Zavorsky GS, Hsia CCW, Hughes JMB, *et al.* Standardisation and application of the single-breath determination of nitric oxide uptake in the lung. *Eur Respir J* 2017;49:1600962.
- 23 Radtke T, Hua-Huy T, Dressel H, *et al.* Of the need to reconcile discrepancies between two different reference equations for combined single-breath $D_{LNO}-D_{LCO}$ in systemic sclerosis. *Eur Respir J* 2019;53:1802109.
- 24 Austin M, Smith L, Eaden J. Validation of nitric oxide transfer factor predicted equations. *Eur Respir J* 2019;54:PA3907.
- 25 Radtke T, de Groot Q, Haile SR, *et al.* Lung diffusing capacity for nitric oxide measured by two commercial devices: a randomised crossover comparison in healthy adults. *ERJ Open Res* 2021;7:00193-2021.
- 26 Zavorsky GS, Cao J, Murias JM. Reference values of pulmonary diffusing capacity for nitric oxide in an adult population. *Nitric Oxide* 2008;18:70–9.
- 27 Aguilaniu B, Maitre J, Glénet S, *et al.* European reference equations for CO and NO lung transfer. *Eur Respir J* 2008;31:1091–7.
- 28 van der Lee I, Zanen P, Stigter N, *et al.* Diffusing capacity for nitric oxide: reference values and dependence on alveolar volume. *Respir Med* 2007;101:1579–84.
- 29 Thomas A, Hanel B, Marott JL, *et al.* The single-breath diffusing capacity of CO and NO in healthy children of European descent. *PLoS One* 2014;9:e113177.
- 30 Munkholm M, Marott JL, Bjerre-Kristensen L, *et al.* Reference equations for pulmonary diffusing capacity of carbon monoxide and nitric oxide in adult Caucasians. *Eur Respir J* 2018;52:1500677.
- 31 Muggeo VMR. Segmented: an R package to fit regression models with broken-line relationships. *R News* 2008;8/1:20–5.
- 32 Muggeo VMR. Estimating regression models with unknown breakpoints. *Stat Med* 2003;22:3055–71.
- 33 Segmented-package: Segmented relationships in regression models with breakpoints / changepoints estimation [program]. Version 1.3-4. 2021.
- 34 Muggeo VMR. Interval estimation for the breakpoint in segmented regression: a smoothed score-based approach. *Aust N Z J Stat* 2017;59:311–22.
- 35 Radtke T, de Groot Q, Haile SR, *et al.* Lung diffusing capacity for nitric oxide measured by two commercial devices: a randomised crossover comparison in healthy adults. *ERJ Open Res* 2021;7:00193-2021.
- 36 Akaike H. New look at Statistical-Model identification. *Ieee T Automat Contr* 1974;Ac19:716–23.
- 37 Schwarz G. Estimating the dimension of a model. *Ann Stat* 1978;6:461–4.
- 38 Wilk MB, Gnanadesikan R. Probability plotting methods for the analysis of data. *Biometrika* 1968;55:1–17.
- 39 van Buuren S, Fredriks M. Worm plot: a simple diagnostic device for modelling growth reference curves. *Stat Med* 2001;20:1259–77.
- 40 Fryar CD, Carroll MD, Gu Q. *Anthropometric reference data for children and adults: United States, 2015–2018. analytical and epidemiological studies.* Washington, D.C.: Vital Health Statistics: National Center for Helath Statistics, 2021: 36 pages.
- 41 . A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJ); 1995 Montreal;* August 20–25, Quebec, Canada, 1995.
- 42 Passing H. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, part I. *J Clin Chem Clin Biochem* 1983;21:709–20.
- 43 Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician* 1983;32:307–17.
- 44 McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* 2012;22:276–82.
- 45 van der Lee I, Zanen P, Biesma DH, *et al.* The effect of red cell transfusion on nitric oxide diffusing capacity. *Respiration* 2005;72:512–6.
- 46 Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.
- 47 Martínez-Briseño D, Gochicoa-Rangel L, Torre-Bouscoulet L, *et al.* Comparing Spirometric reference values from childhood to old age estimated by LMS and linear regression models. *Arch Bronconeumol* 2021;57:172–8.
- 48 Brisman J, Kim J-L, Olin A-C, *et al.* A physiologically based model for Spirometric reference equations in adults. *Clin Physiol Funct Imaging* 2016;36:77–84.
- 49 Lubiński W, Gólczewski T. Physiologically interpretable prediction equations for Spirometric indexes. *J Appl Physiol* 2010;108:1440–6.
- 50 Brisman J, Kim J-L, Olin A-C, *et al.* Spirometric reference equations for Swedish adults. *Clin Physiol Funct Imaging* 2017;37:640–5.
- 51 Kubota M, Kobayashi H, Quanjer PH, *et al.* Reference values for spirometry, including vital capacity, in Japanese adults calculated with the LMS method and compared with previous values. *Respir Investig* 2014;52:242–50.
- 52 Robson AG, Innes JA. Short term variability of single breath carbon monoxide transfer factor. *Thorax* 2001;56:358–61.
- 53 Radtke T, Benden C, Maggi-Beba M, *et al.* Intra-session and inter-session variability of nitric oxide pulmonary diffusing capacity in adults with cystic fibrosis. *Respir Physiol Neurobiol* 2017;246:33–8.
- 54 Magini A, Contini M, Spadafora E, *et al.* Variability in pulmonary diffusing capacity in heart failure. *Respir Physiol Neurobiol* 2020;280:103473.
- 55 Desjardin A, Creveuil C, Bergot E, *et al.* Assessment of concordance between diffusion of carbon monoxide through the lung using the 10 S breath-hold method, and the simultaneous NO/CO technique, in healthy participants. *Respir Physiol Neurobiol* 2020;273:103319.

- 56 Murias JM, Zavorsky GS. Short-Term variability of nitric oxide diffusing capacity and its components. *Respir Physiol Neurobiol* 2007;157:316–25.
- 57 Hegewald MJ, Jensen RL, Teeter JG, *et al.* Long-Term intersession variability for single-breath diffusing capacity. *Respiration* 2012;84:377–84.
- 58 Pellegrino R, Viegi G, Brusasco V, *et al.* Interpretative strategies for lung function tests. *Eur Respir J* 2005;26:948–68.
- 59 DeCato TW, Hegewald MJ. Breathing red: physiology of an elevated single-breath diffusing capacity of carbon monoxide. *Ann Am Thorac Soc* 2016;13:2087–92.
- 60 Ewan PW, Jones HA, Rhodes CG, *et al.* Detection of intrapulmonary hemorrhage with carbon monoxide uptake. application in Goodpasture's syndrome. *N Engl J Med* 1976;295:1391–6.
- 61 Greening AP, Patel K, Goolden AW, *et al.* Carbon monoxide diffusing capacity in polycythaemia rubra vera. *Thorax* 1982;37:528–31.
- 62 Saydain G, Beck KC, Decker PA, *et al.* Clinical significance of elevated diffusing capacity. *Chest* 2004;125:446–52.
- 63 Greening AP, Hughes JM. Serial estimations of carbon monoxide diffusing capacity in intrapulmonary haemorrhage. *Clin Sci* 1981;60:507–12.
- 64 Collard P, Njinou B, Nejadnik B, *et al.* Single breath diffusing capacity for carbon monoxide in stable asthma. *Chest* 1994;105:1426–9.
- 65 Jorgenson CC, Chase SC, Olson LJ, *et al.* Assessment of thoracic blood volume by computerized tomography in patients with heart failure and periodic breathing. *J Card Fail* 2018;24:479–83.
- 66 Smith TC, Rankin J. Pulmonary diffusing capacity and the capillary bed during Valsalva and Müller maneuvers. *J Appl Physiol* 1969;27:826–33.
- 67 Burtch AR, Ogle BT, Sims PA, *et al.* Controlled frequency breathing reduces inspiratory muscle fatigue. *J Strength Cond Res* 2017;31:1273–81.
- 68 Armour J, Donnelly PM, Bye PT. The large lungs of elite swimmers: an increased alveolar number? *Eur Respir J* 1993;6:237–47.
- 69 Yost LJ, Zauner CW, Jaeger MJ. Pulmonary diffusing capacity and physical working capacity in swimmers and non-swimmers during growth. *Respiration* 1981;42:8–14.
- 70 Andrew GM, Becklake MR, Guleria JS, *et al.* Heart and lung functions in swimmers and nonathletes during growth. *J Appl Physiol* 1972;32:245–51.
- 71 Bovard JM, Welch JF, Houghton KM, *et al.* Does competitive swimming affect lung growth? *Physiol Rep* 2018;6:e13816.
- 72 Zavorsky GS, Smoliga JM. The association between cardiorespiratory fitness and pulmonary diffusing capacity. *Respir Physiol Neurobiol* 2017;241:28–35.
- 73 Borland CDR, Dunningham H, Bottrill F, *et al.* Significant blood resistance to nitric oxide transfer in the lung. *J Appl Physiol* 2010;108:1052–60.
- 74 Zavorsky GS. The rise in carboxyhemoglobin from repeated pulmonary diffusing capacity tests. *Respir Physiol Neurobiol* 2013;186:103–8.
- 75 Zavorsky GS, Almamary AS, Alqahtani MK, *et al.* The need for race-specific reference equations for pulmonary diffusing capacity for nitric oxide. *BMC Pulm Med* 2021;21:232.
- 76 Simaga B, Forton K, Motoji Y, *et al.* Lung diffusing capacity in sub-Saharan Africans versus European Caucasians. *Respir Physiol Neurobiol* 2017;241:23–7.
- 77 Pesola GR, Sunmonu Y, Huggins G, *et al.* Measured diffusion capacity versus prediction equation estimates in blacks without lung disease. *Respiration* 2004;71:484–92.
- 78 Neas LM, Schwartz J. The determinants of pulmonary diffusing capacity in a national sample of U.S. adults. *Am J Respir Crit Care Med* 1996;153:656–64.
- 79 Weiner DJ, Graham B, Stanojevic S. Ethnically diverse normative data for diffusing capacity and lung volumes: another research priority. *Ann Am Thorac Soc* 2020;17:128.
- 80 Gray G, Zamel N, Crapo RO. Effect of a simulated 3,048 meter altitude on the single-breath transfer factor. *Bull Eur Physiopathol Respir* 1986;22:429–31.
- 81 Vedal S, Crapo RO. False positive rates of multiple pulmonary function tests in healthy subjects. *Bull Eur Physiopathol Respir* 1983;19:263–6.

ONLINE SUPPLEMENTARY MATERIAL**Reference equations for pulmonary diffusing capacity using segmented regression show similar predictive accuracy as GAMLSS models**

¹Gerald S. Zavorsky Ph.D., ²Jiguo Cao Ph.D.

¹Pulmonary Services, University of California, Davis, Medical Center, Sacramento, California, United States

²Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia, Canada

This supplement has been peer-reviewed.

Gerald Zavorsky: ORCID ID: <https://orcid.org/0000-0002-4473-1601>

How simulated raw data was created from the paper by Munkholm and colleagues

After repeated unsuccessful attempts to obtain the raw data from Munkholm and colleagues¹, it was decided to simulate their data so that the reference equations for pulmonary diffusing capacity, originally published in 2017 in the European Respiratory Journal², could be updated. Only adults were used in their study¹, so this newly created or "simulated" raw data had the same number of adult women and adult men as their paper (142 females, 140 males). The age data (range = 18 to 97 years of age) was generated from the normal distribution with their mean and standard deviation (53 ± 23 years old for women, 54 ± 22 years old for men) using a statistical technique called truncation³. A truncated sample can be thought of as being equivalent to an underlying sample with all values outside the bounds entirely omitted. The "*rtruncnorm*" library (version 1.0-8) in R language (<https://www.r-project.org/>) was used for random generation of truncated normally distributed values. There is an approximate correlation of 0.6 between height and weight when both sexes are combined in previous studies that generated reference equations for DLNO in adult subjects⁴⁻⁶(pooled data), the mean and SD for height (165 ± 7 cm for women, 179 ± 8 cm for men) and weight (64.6 ± 9.0 kg for women, 78.5 ± 11.0 cm for men) were generated from a bivariate normal distribution. The mean and standard deviation for males and females from Munkholm *et al.*¹ and the correlation coefficient ~ 0.6 was truncated by the range of height (range = 149-184 cm in women, 156-198 cm in men) and weight (45-97 kg in women, and 52 to 109 kg in men). The data for DLCO, DLNO, and VA are generated from the linear model from their paper¹, and the residuals are generated from the normal distribution with mean 0 and the standard deviation provided for its residual standard errors¹.

Only 25% of the total pooled data (272 out of 1076 subjects) included the simulated Munkholm data. Of all the subjects that had their lung function tested using the Jaeger MasterScreen Pro or MasterLab Pro lung function device, 40% of these subjects are from the simulated Munkholm data (see **Figure 2** in their article). The results of the accuracy of the simulated data compared to their prediction equations are presented in the following four pages (**Tables S1, S2A, S2B, S3**).

Table S1. A comparison between anthropometric characteristics generated from simulated data created from Munkholm and colleagues¹, and the actual data.

Female (N=142)	Munkholm <i>et al.</i> (2018)	Simulated raw data created from Munkholm's population characteristics
Age (yrs)	53 (23) [18-97]	55 (18) [20-95]
Height (cm)	165 (7) [149-184]	166 (7) [152-182]
Weight (kg)	64.6 (9.0) [45.1-97.0]	65.0 (9.0) [45.1-92.2]
BMI (kg/m ²)	23.6 (2.7) [18.1-29.8]	23.5 (3.0) [15.9 to 32.1]
Male (N=140)		
Age (yrs)	54 (22) [18-97]	55 (19) [19-95]
Height (cm)	179 (8) [156-198]	179 (8) [160-197]
Weight (kg)	78.5 (11.0) [52.0-108.8]	78.4 (11.0) [54.7-100.7]
BMI (kg/m ²)	24.4 (2.6) [18.0-30.0]	24.4 (2.8) [18.0-32.5]

Mean and SD. The SD are in parentheses; the range is in brackets.

Table S2A. The simulated raw data created from Munkholm and colleagues¹ female study population characteristics is nearly equivalent to their actual raw data.

Female	Age	Age ²	height	Constant	R ²	SEE
DLNO Munkholm <i>et al.</i> (2018) (mL/min/mmHg)		-0.00753	0.766	-2.36	0.80	11.4
DLNO based on simulated raw data from Munkholm's (2018) population characteristics (mL/min/mmHg)		-0.00769	0.799	-7.74	0.66	12.0
DLCO Munkholm <i>et al.</i> (2018) (mL/min/mmHg)		-0.00166	0.192	-3.58	0.77	2.8
DLCO based on simulated raw data from Munkholm's (2018) population characteristics (mL/min/mmHg)		-0.00170	0.251	-12.84	0.67	2.8
VA Munkholm <i>et al.</i> (2018) (L)	0.039	-0.000426	0.047	-3.55	0.53	0.49
VA based on simulated raw data from Munkholm's (2018) population characteristics (L)		-0.000126	0.035	-0.48	0.30	0.54

DLNO = pulmonary diffusing capacity for nitric oxide; DLCO = pulmonary diffusing capacity for carbon monoxide; VA = alveolar volume.

Table S2B. The simulated raw data created from Munkholm and colleagues¹ male study population characteristics is nearly equivalent to their actual raw data.

Male	Age	Age ²	height	Constant	R ²	SEE
DLNO Munkholm <i>et al.</i> (2018) (mL/min/mmHg)		-0.0125	0.97	5.72	0.82	16.6
DLNO based on simulated raw data from Munkholm's (2018) population characteristics (mL/min/mmHg)		-0.0119	1.151	-7.74	0.75	15.3
DLCO Munkholm <i>et al.</i> (2018) (mL/min/mmHg)		-0.00258	0.252	-5.01	0.81	3.7
DLCO based on simulated raw data from Munkholm's (2018) population characteristics (mL/min/mmHg)		-0.00248	0.22	0.14	0.68	3.7
VA Munkholm <i>et al.</i> (2018) (L)	0.0387	-0.000442	0.077	-7.90	0.59	0.69
VA based on simulated raw data from Munkholm's (2018) population characteristics (L)		-0.000090	0.073	-6.16	0.48	0.64

DLNO = pulmonary diffusing capacity for nitric oxide; DLCO = pulmonary diffusing capacity for carbon monoxide; VA = alveolar volume.

Table S3. Reference equations from Munkholm *et al.*¹ are compared to estimated equations based on simulated raw data created from Munkholm's study population characteristics.

	DLNO	DLCO	VA
Correlation coefficient	0.99	0.99	0.99
Coefficient of Variation (%)	1.4%	1.2%	2.1%
Bland-Altman Differences 95% CI (mL/min/mmHg)	-4 to +5	-0.9 to +0.9	-0.34 to +0.34
Paired <i>t</i> -test differences (mL/min/mmHg)	+0.5 (2.4)	0 (0.5)	0 (0.2)

For the paired *t*-test differences, the numbers in parentheses indicate the SD. DLNO = pulmonary diffusing capacity for nitric oxide; DLCO = pulmonary diffusing capacity for carbon monoxide; VA = alveolar volume.

Table S4. Reference equations using segmented regression with sex and equipment devices included in the equation.

DLCO (n = 1076) (mL/min/mmHg) Breakpoint = 24.3 yrs old [95% CI = 23.1 to 25.5 yrs old] AIC = 5801; BIC = 5841	Estimate	SE	95% CI	Adjusted R ²	RSE
Intercept ₁ (for 5.0 to 24.2 yrs old)	-19.49	1.46	-12.4, -16.6	0.81	2.86 ₁
Intercept ₂ (for 24.3 to 95.0 yrs old)	-6.74				3.90 ₂
Age ² ₁ (for 5.0 to 24.2 yrs old)	0.0194	0.0017	0.016, 0.023		
Age ² ₂ (for 24.3 to 95.0 yrs old)	-0.0021	0.00007	-0.002, -0.002		
Height (cm)	0.223	0.011	0.20, 0.24		
Altitude (m)	0.005	0.00106	0.003, 0.007		
Sex	4.86	0.25	4.38, 5.34		
DLNO (n = 1076) (mL/min/mmHg) Breakpoint = 23.5 yrs old [95% CI = 21.9 to 24.9 yrs old] AIC = 9044; BIC = 9084	Estimate	SE	95% CI	Adjusted R ²	RSE
Intercept ₁ (for 5.0 to 23.4 yrs old)	-108.1	6.80	-121.4, -94.8	0.84	13.31 ₁
Intercept ₂ (for 23.5 to 95.0 yrs old)	-61.0				17.36 ₂
Age ² ₁ (for 5.0 to 23.4 yrs)	0.0755	0.0092	0.058, 0.094		
Age ² ₂ (for 23.5 to 95.0 yrs old)	-0.010	0.003	-0.016, 0.004		
Height (cm)	1.17	0.052	1.07, 1.27		
PFT equipment	18.0	1.21	15.6, 20.4		
Sex	23.9	1.12	21.7, 26.1		
VA (n = 1076) (L) Breakpoint = 29.5 yrs old [95% CI = 27.7 to 31.2 yrs old] AIC = 2018; BIC = 2058	Estimate	SE	95% CI	Adjusted R ²	RSE
Intercept ₁ (for 5.0 to 29.4 yrs old)	-5.93	0.243	-6.41, -5.45	0.86	0.52 ₁
Intercept ₂ (for 29.5 to 95.0 yrs old)	-4.39				0.67 ₂
Age ² ₁ (for 5.0 to 29.4 yrs old)	0.0016	0.0002	0.001, 0.002		
Age ² ₂ (for 29.5 to 95.0 yrs old)	-0.00015	0.00001	0.000, 0.000		
Height (cm)	0.060	0.002	0.056, 0.064		
PFT equipment	0.280	0.045	0.192, 0.368		
Sex	0.76	0.042	0.68, 0.85		

SE = standard error, RSE = residual standard error. AIC = Akaike information criterion; BIC = Bayesian information criterion; For the PFT equipment, 1 = Hyp'Air Compact, and 0 = Jaeger MasterScreen Pro. Sex, 1 = male, 0 = female. Only one age² and one intercept is used, depending on the age of the subject. Age²₁ and intercept₁ is used before the breakpoint. Age²₂ and intercept₂ is used at the breakpoint and beyond. For example, for a male who is 23.0 years old, 180 cm tall, and had his measurements performed on the Hyp'Air device, the predicted DLNO (mL/min/mmHg) = 0.0755·(23²) + 1.17·(180) + 18·(1) + 23.9·(1) – 108.1 = 184.3 mL/min/mmHg with an LLN of 184.3 – (13.31·1.645) = 162.4 mL/min/mmHg. For a male 24 years old, with the same height and equipment used, the predicted DLNO (mL/min/mmHg) = -0.01·(24²) + 1.17·(180) + 18·(1) + 23.9·(1) – 61.0 = 185.7 mL/min/mmHg with an LLN of 185.7 – (17.36·1.645) = 157.1 mL/min/mmHg.

NOTE: If the DLCO is measured at an altitude that is more than 300 meters, we recommend converting the measured DLCO to sea level first, based on the data by Gray and colleagues⁷, and then omitting the altitude covariate from the equation (as the converted DLCO will be adjusted to an altitude of 0 meters). Adjusting the DLCO to sea level (mL/min/mmHg) = measured DLCO at altitude · (0.505 + 0.00065·barometric pressure in mmHg at altitude). The formula to estimate barometric pressure at altitude in mm Hg is: 760·exp[- 0.284·altitude in meters / (8.314·Temperature in Kelvin)], where Kelvin = °C + 273.15. (see: <https://planetcalc.com/938/>)

Table S5. Mean z-scores, skewness, and mean percent predicted values fitted to the reference equations.

Fitted z-scores [mean (SD)]	GAMLSS	Segmented linear regression	GLI
DLCO _{MasterScreen}	-0.03 (0.95) [-0.02]	-0.03 (0.90) [-0.03]	0.28 (0.90) [0.05]
DLCO _{Hyp'Air}	0.11 (1.08) [-0.03]	0.04 (1.38) [0.07]	0.99 (1.11) [-0.27]
DLNO _{MasterScreen}	-0.02 (1.02) [0.03]	-0.01 (0.88) [-0.12]	N/A
DLNO _{Hyp'Air}	-0.01 (1.00) [-0.06]	0.01 (1.17) [0.03]	N/A
VA _{MasterScreen}	-0.01 (0.98) [-0.02]	0.04 (0.93) [0.23]	0.02 (0.91) [0.13]
VA _{Hyp'Air}	-0.03 (1.04) [0.01]	0.05 (1.09) [0.18]	0.63 (1.09) [0.12]
Fitted % predicted [Mean (SD)]	GAMLSS	Segmented linear regression	GLI
DLCO _{MasterScreen}	100 (12)	100 (12)	105 (14)
DLCO _{Hyp'Air}	102 (15)	100 (15)	116 (18)
DLNO _{MasterScreen}	100 (11)	100 (12)	N/A
DLNO _{Hyp'Air}	100 (14)	100 (14)	N/A
VA _{MasterScreen}	100 (10)	101 (14)	101 (10)
VA _{Hyp'Air}	100 (11)	100 (11)	107 (12)

n = 691 for the Jaeger MasterScreen Pro; n = 385 for the Hyp'Air Compact. DLNO = pulmonary diffusing capacity for nitric oxide; DLCO = pulmonary diffusing capacity for carbon monoxide; VA = alveolar volume; N/A = not applicable. Brackets represent the skewness, and parentheses represent the SD.

Table S6. Pearson's r Correlational matrix of fitted z-scores between models.

	DLCO z-scores GAMLSS	DLCO z-scores segmented linear regression	DLCO z-scores GLI	DLNO z-scores GAMLSS	DLNO z-scores segmented linear regression	VA z- scores GAMLSS	VA z- scores segmented linear regression	VA z- scores GLI
DLCO z-scores GAMLSS		0.94	0.82	0.62	0.65	0.44	0.43	0.41
DLCO z-scores segmented linear regression	0.94		0.80	0.59	0.69	0.42	0.45	0.40
DLCO z-scores GLI	0.82	0.80		0.51	0.54	0.35	0.33	0.48
DLNO z-scores GAMLSS	0.62	0.59	0.52		0.93	0.46	0.43	0.42
DLNO z-scores segmented linear regression	0.65	0.69	0.54	0.93		0.47	0.49	0.44
VA z-scores GAMLSS	0.44	0.42	0.35	0.46	0.47		0.95	0.91
VA z-scores Segmented linear regression	0.43	0.45	0.33	0.43	0.49	0.95		0.88
VA z-scores GLI	0.41	0.40	0.48	0.42	0.44	0.91	0.88	

$p < 0.001$ for all. The correlations between the Global Lung Function Initiative (GLI) fitted z-scores (<http://gli-calculator.ersnet.org/index.html>), and all others include 1065 subjects as GLI fitted z-scores do not go above 85 years of age. Otherwise, 1076 subjects. DLNO = pulmonary diffusing capacity for nitric oxide; DLCO = pulmonary diffusing capacity for carbon monoxide; VA = alveolar volume. Note: these correlations are based on non-diseased subjects, so the range of z-scores are narrow (about -3.0 to +3.0). If diseased subjects were included then the range of z-scores would be wider and the correlations would likely be higher than what is shown, here.

Table S7. Pearson's r correlations between measured variables and the fitted predicted values from various models.

	Measured DLCO	Measured DLNO	Measured VA
Predicted DLCO segmented linear regression	0.91	0.92	0.88
Predicted DLCO GAMLSS	0.91	0.91	0.88
Predicted DLNO segmented linear regression	0.91	0.93	0.86
Predicted DLNO GAMLSS	0.90	0.93	0.87
Predicted VA segmented linear regression	0.86	0.85	0.93
Predicted VA GAMLSS	0.86	0.85	0.94
Predicted DLCO GLI	0.88	0.88	0.88
Predicted VA GLI	0.84	0.83	0.93

$p < 0.001$ for all. Both equipment devices were combined. DLNO = pulmonary diffusing capacity for nitric oxide; DLCO = pulmonary diffusing capacity for carbon monoxide; VA = alveolar volume. $n = 1076$ for GAMLSS and segmented linear regression; $n = 1065$ for GLI. The correlations between the predicted values calculated by the Global Lung Function Initiative (GLI) (<http://gli-calculator.ersnet.org/index.html>) and the current GAMLSS and segmented regression models were based on 1052 subjects as GLI predictions do not go above 85 years of age.

Table S8. Receiver-Operating Characteristic (ROC) analysis for evaluating the performance of DLNO compared to DLCO when the estimated prevalence of an abnormal result in a population is five percent (i.e., when 5% of the population is below the LLN based on DLCO).

	Segmented regression	GAMLSS
Area under the ROC Curve (AUC)	0.68 [0.65, 0.70]	0.69 [0.66, 0.72]
Youden's J statistic	0.35 [0.24, 0.47]	0.38 [0.25, 0.50]
Sensitivity	0.38 [0.26, 0.50]	0.42 [0.29, 0.55]
Specificity	0.97 [0.96, 0.98]	0.97 [0.95, 0.98]
Positive predictive value	40% [30%, 52%]	38% [29%, 49%]
Negative predictive value	97% [96%, 97%]	97% [96%, 98%]
Positive likelihood ratio	12.7 [8.0, 20.2]	11.8 [7.6, 18.2]
Negative likelihood ratio	0.64 [0.53, 0.77]	0.60 [0.49, 0.75]

AUC = The percent chance that when using DLCO to detect an abnormal result, the DLNO can also distinguish an abnormal result in the same patient. Brackets represent the 95% bootstrapped confidence interval.

Youden's J statistic (sensitivity + specificity – 1): Measures the effectiveness of using DLNO as a diagnostic test compared to DLCO.

Sensitivity (true positive rate): Probability that DLNO is abnormal (< LLN) when DLCO is also abnormal (< LLN).

Specificity (true negative rate): Probability that DLNO is normal (\geq LLN) when DLCO is also normal (\geq LLN).

Positive predictive value (precision): Probability DLCO is abnormal (< LLN) when DLNO is also abnormal (< LLN).

Negative predictive value: Probability that DLCO is normal (\geq LLN) when DLNO is also normal (\geq LLN).

Positive likelihood ratio (True positive rate \div false positive rate): The ratio between the probability that DLNO is abnormal (< LLN) given that DLCO is abnormal (< LLN) and the probability that DLNO is abnormal (< LLN) given that DLCO is normal (\geq LLN).

Negative likelihood ratio (false-negative rate \div true negative rate): The ratio between the probability that DLNO is normal (\geq LLN) given that the DLCO is abnormal (< LLN) and the probability that the DLNO is normal (\geq LLN) given that the DLCO is normal (DLCO is \geq LLN).

Table S9. The approximate percent predicted at the lower limit of normal and its interquartile range varies across age groups between the two different models.

Segmented Regression	5-7 years (n=51)	8-10 years (n=93)	11-50 years (n=594)	51-70 years (234)	71-95 years (n=104)
DLNO	68 [61-73]	78 [74-80]	82 [81-84]	78 [76-79]	71 [67-74]
DLCO	66 [60-70]	75 [72-77]	82 [80-86]	77 [76-79]	71 [67-79]
VA	64 [55-70]	76 [73-79]	83 [82-86]	81 [80-82]	79 [77-81]
GAMLSS	5-7 years (n=51)	8-10 years (n=93)	11-50 years (n=594)	51-70 years (234)	71-95 years (n=104)
DLNO	83 [83-83]	83 [83-83]	83 [77-83]	82 [77-83]	82 [77-83]
DLCO	84 [80-84]	83 [81-84]	82 [80-83]	79 [77-80]	74 [72-75]
VA	87 [85-87]	86 [85-87]	85 [84-85]	85 [81-85]	81 [80-85]

The 25th to 75th percentile of the percent predicted at the LLN is a weighted average in brackets with its median outside. These values correspond to a z-score of -1.645 . Notice that in the segmented regression models, the youngest and oldest age groups have a much lower median percent predicted compared to the middle age groups for DLNO and DLCO. For VA (segmented regression), the youngest age group is much different compared to all other age groups. With the GAMLSS models, the percent predicted at the LLN is very consistent for all age groups except for DLCO.

Table S10. The US Department of Health median values for height at each age for white subjects⁸.

Age (yrs)	Height female (cm)	Age (yrs)	Height male (cm)
5	112.1	5	112.4
6	119.3	6	118.0
7	123.7	7	126.1
8	129.8	8	131.8
9	136.5	9	136.4
10	142.3	10	141.1
11	150.8	11	148.3
12	154.3	12	153.9
13	157.7	13	163.6
14	161.2	14	170.0
15	160.0	15	172.7
16	161.7	16	172.6
17	162.4	17	174.9
18	162.3	18	175.5
19	161.2	19-39	178.0
20-39	164.4	40-59	177.5
40-59	163.1	60-95	174.3
60-95	159.8		

Table S11. A worked example in traditional units for calculating the predicted, percent predicted, and DLCO z-score in a hypothetical individual using the GAMLSS model.

A white female 65 yrs old, 167 cm tall) has a measured DLCO of 17.0 mL/min/mmHg at 300-m elevation. To predict the DLCO value for her age, sex, height, ethnicity, and elevation, see the reference equation in Table 2B of the main article and see the supplementary excel spreadsheet for the spline tables.

From the excel tables, the Mspline for DLCO, female = -0.458, and the Sspline for DLCO, female = 0.5711

M (predicted value) = $\exp[-4.481 + 1.406 \cdot \ln(\text{height}) + 0.194 \cdot \ln(\text{age}) + 0.0002 \cdot (\text{altitude}) + \text{Mspline}]$

$M = \exp[-4.481 + 1.406 \cdot \ln(167) + 0.194 \cdot \ln(65) + 0.0002 \cdot (300) - 0.458] = \exp(3.1267)$

M (predicted value) = **22.8 mL/min/mmHg**

S (variability around the median) = $\exp[0.642 \cdot \ln(\text{age}) - 1.018 \cdot \ln(\text{height}) + \text{Sspline}]$

$S = \exp[0.642 \cdot \ln(65) - 1.018 \cdot \ln(167) + 0.5711] = \exp(-1.959)$

S = **0.141**

L (index of skewness) = **0.325**

% predicted = (measured/M) · 100

% predicted = $(17.0/22.8) \cdot 100 = 74.6\%$

Lower limit of Normal (LLN) (5th percentile) = $\exp[\ln(M) + \ln(1 - 1.645 \cdot L \cdot S)/L]$

Lower limit of Normal (LLN) (5th percentile) = $\exp[\ln(22.8) + (\ln(1 - 1.645 \cdot 0.325 \cdot 0.141)/0.325)]$

Lower limit of Normal (LLN) (5th percentile) = $\exp[3.1268 + (\ln(0.9246)/0.325)]$

Lower limit of Normal (LLN) (5th percentile) = $\exp[3.1268 + (-0.0784/0.325)]$

Lower limit of Normal (LLN) (5th percentile) = $\exp[3.1268 - 0.2412] = \exp(2.8856)$

Lower limit of Normal (LLN) (5th percentile) = **17.9 mL/min/mmHg**

Z-score = $[(\text{measured}/M)^L - 1]/(L \cdot S)$

Z-score = $[(17.0/22.8)^{0.325} - 1]/(0.325 \cdot 0.141)$

Z-score = $[(0.7456)^{0.325} - 1]/0.0458$

Z-score = $[0.909 - 1]/0.0458 = -1.987$.

Her z-score is below the LLN for the 5th percentile and 2.5th percentile. Her z-score of -1.987 is more negative than a z-score of -1.645 AND -1.96.

NOTE: If the DLCO is measured at an altitude that is more than 300 meters, we recommend converting the measured DLCO to sea level first based on the data by Gray and colleagues⁷, and then omitting the altitude covariate from the equation (as the converted DLCO will be at an altitude of 0 meters). Adjusted DLCO to sea level (mL/min/mmHg) = measured DLCO at altitude · (0.505 + 0.00065 · barometric pressure in mmHg at altitude). The formula to estimate barometric pressure at altitude in mm Hg is: $760 \cdot \exp[-0.284 \cdot \text{altitude in meters} / (8.314 \cdot \text{Temperature in Kelvin})]$, where Kelvin = °C + 273.15. (see: <https://planetcalc.com/938/>).

Figure S1. The DLNO to DLCO ratio across age groups and lung function devices. The ratio is relatively independent of age. Those that had the DLNO/DLCO ratio measured on the Hyp'Air Compact device (n=385, solid green bars) had a higher ratio compared to those that had the ratio using the Jaeger MasterScreen Pro (n=691, solid blue bars) (4.87 ± 0.44 vs. 4.58 ± 0.57). The Hyp'Air device provided a ratio that was 0.23 to 0.35 units higher than the Jaeger MasterScreen Pro (95% CI, $p < 0.001$). This is because the Hyp'Air Compact measures DLNO that is about 16 to 20 mL/min/mmHg more compared to the Jaeger MasterScreen Pro, yet both devices have similar DLCO values. Open circles are outliers [3^{rd} quartile + $1.5 \cdot$ interquartile range or 1^{st} quartile - $1.5 \cdot$ interquartile range]. Extreme outliers are presented by an asterisk [more than the 3^{rd} quartile + $3 \cdot$ interquartile range or 1^{st} quartile - $3 \cdot$ interquartile range].

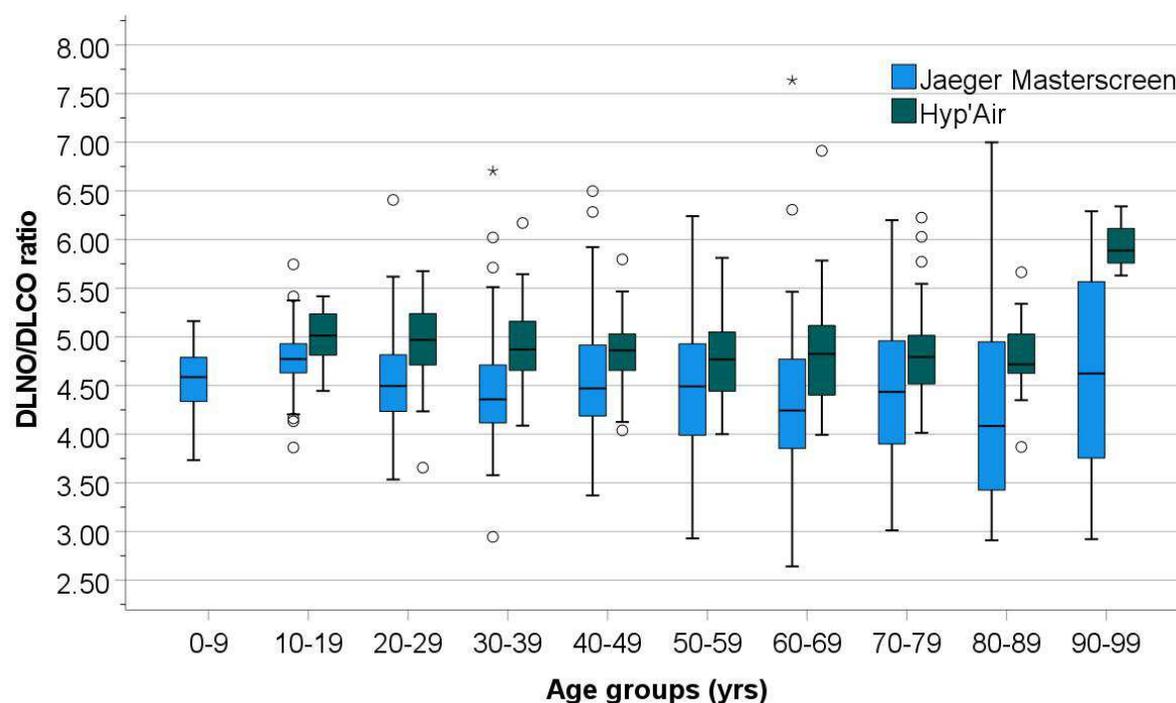


Figure S2. Fitted z-scores and % predicted DLNO, DLCO, and VA using segmented linear regression models. Both devices are combined (n=1076).

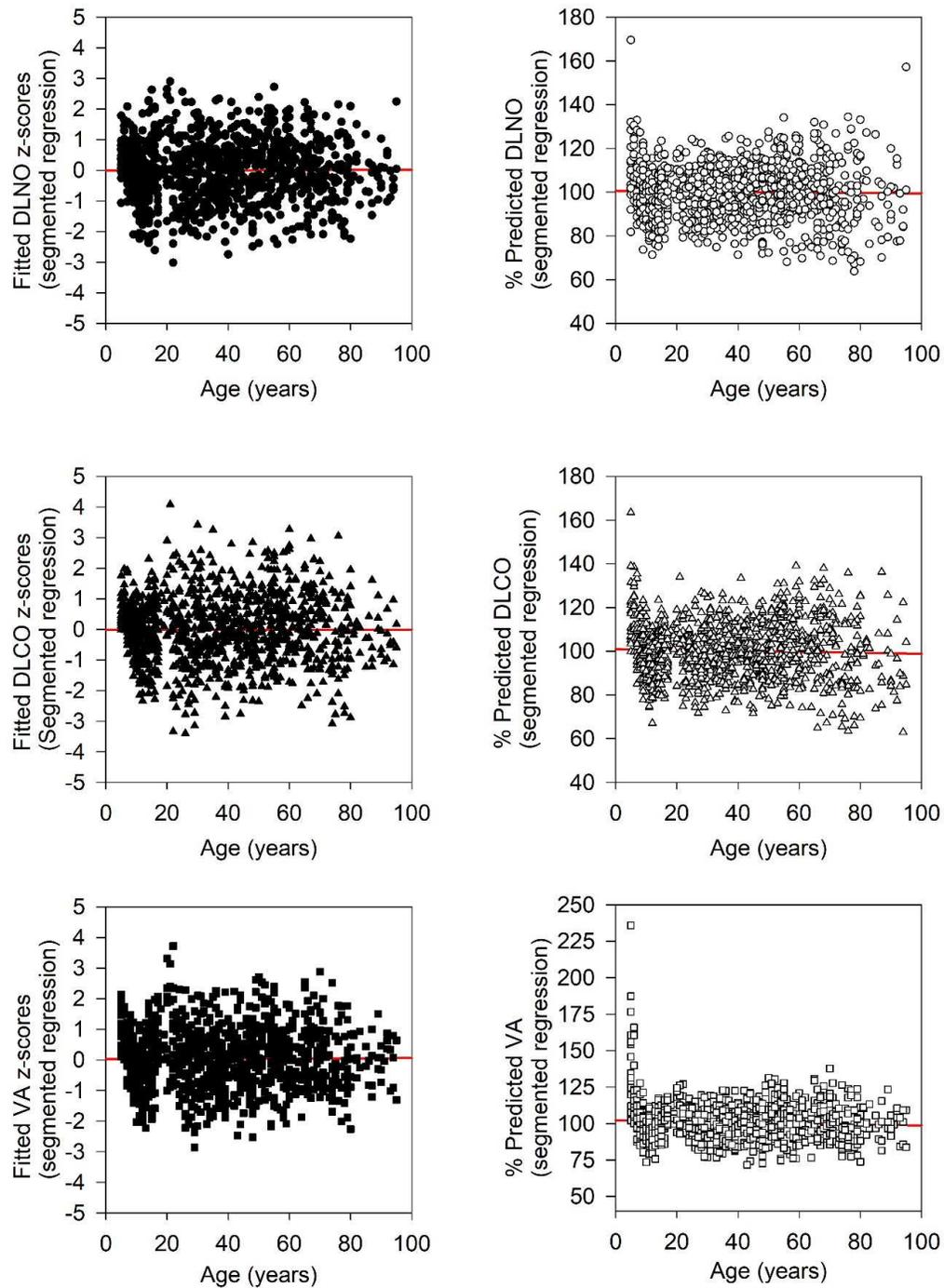


Figure S3. Fitted z-scores and % predicted DLNO, DLCO, and VA using GAMLSS models. Both devices are combined (n = 1076).

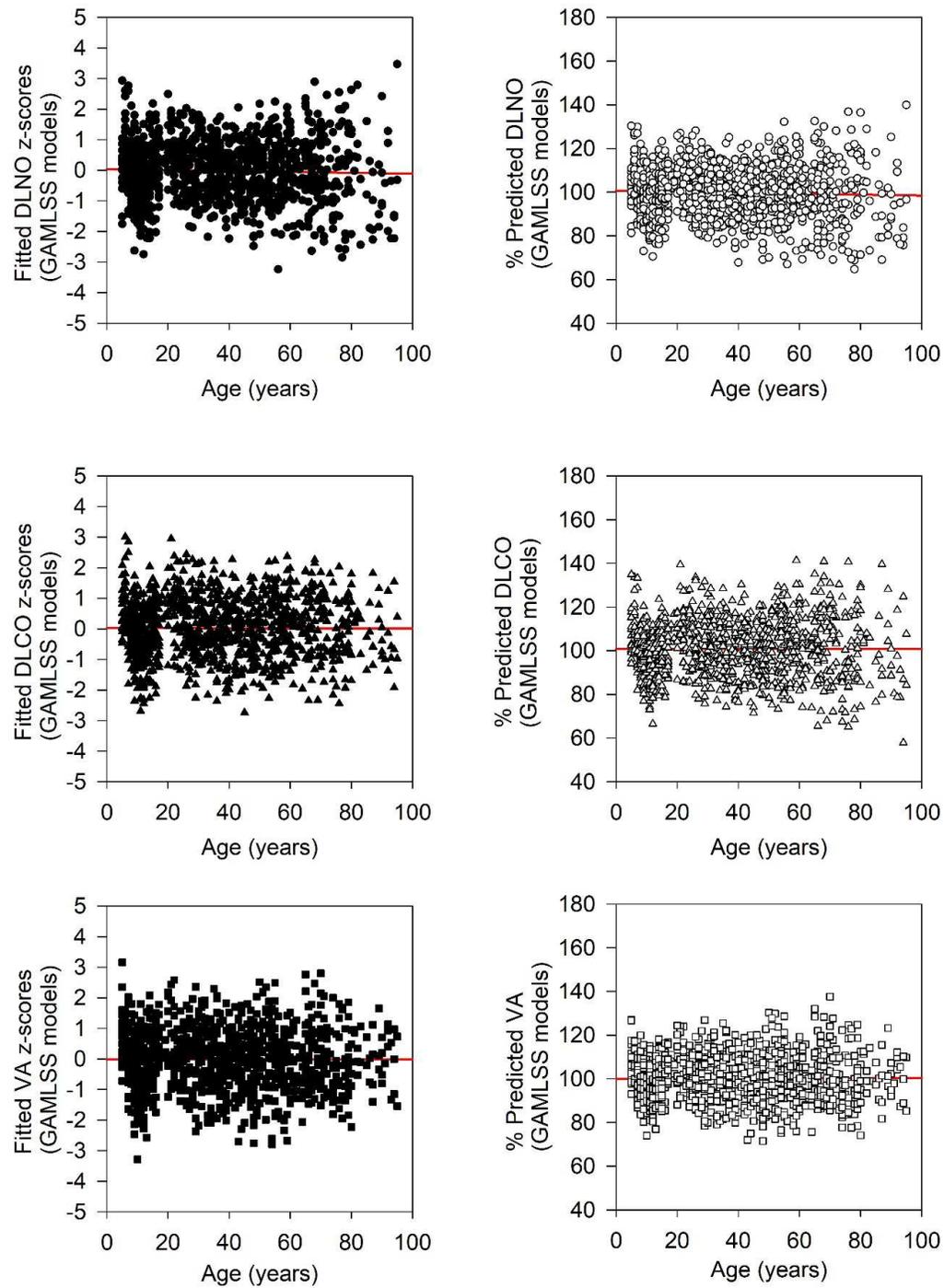


Figure S4. Quantile-Quantile (Q-Q) plots for DLNO, DLCO, and VA as generated by segmented regression models. These graphs depict the observed quantiles of a variable's distribution (red circles) against the quantiles that we would expect to see if the data is approximately normally distributed (solid black line). A normal distribution can approximate the z-scores data since most data points fall on the straight line. As for the percent predicted data, DLNO, DLCO, and VA shows two, one, and 12 outliers, respectively.

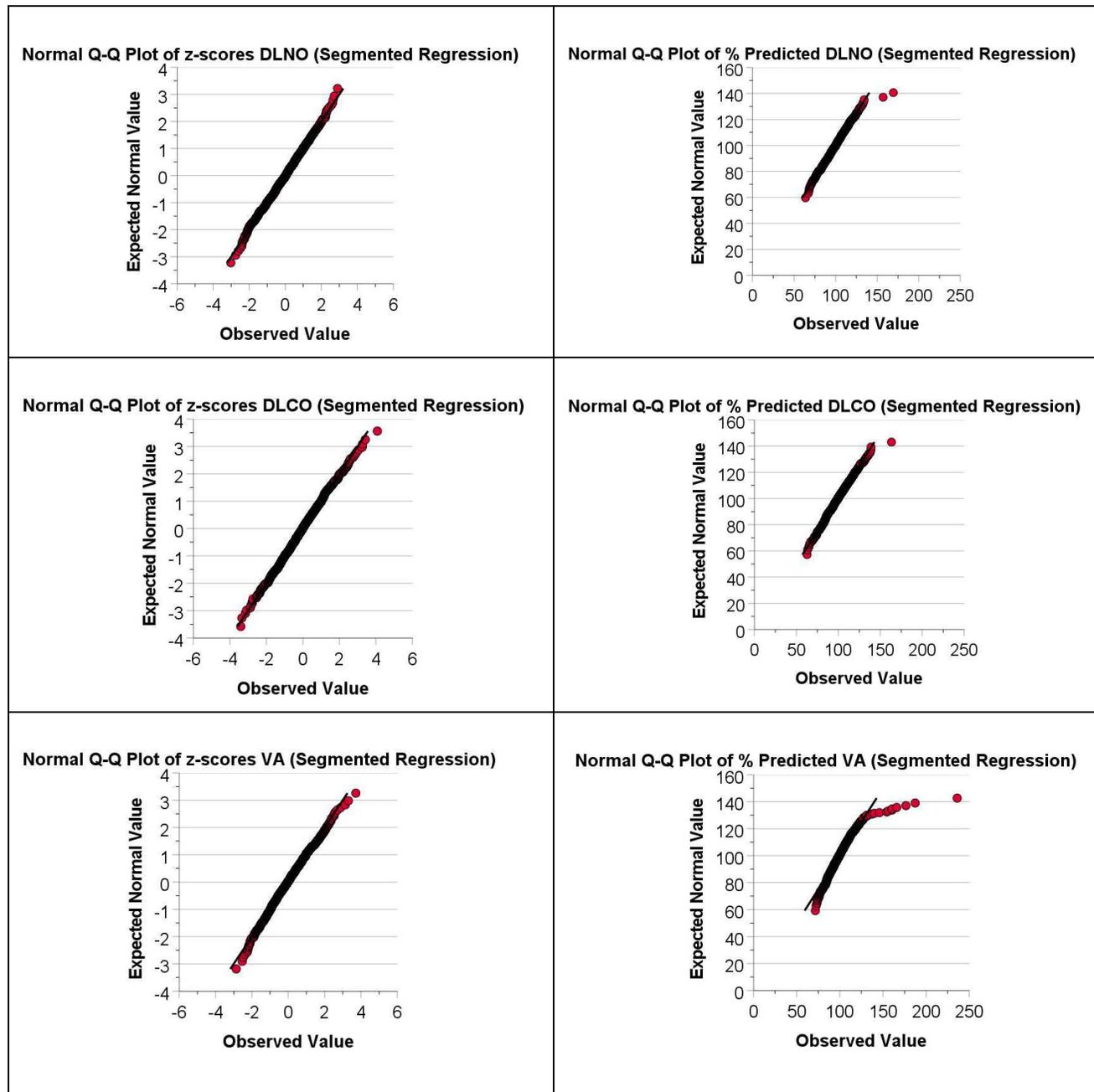


Figure S5. Quantile-Quantile (Q-Q) plots for DLNO, DLCO, and VA as generated by GAMLSS models. These graphs depict the observed quantiles of a variable's distribution (red circles) against the quantiles that we would expect to see if the data is approximately normally distributed (solid black line). The z-scores and % predicted values for all variables could be approximated by a normal distribution since almost all data points are straight.

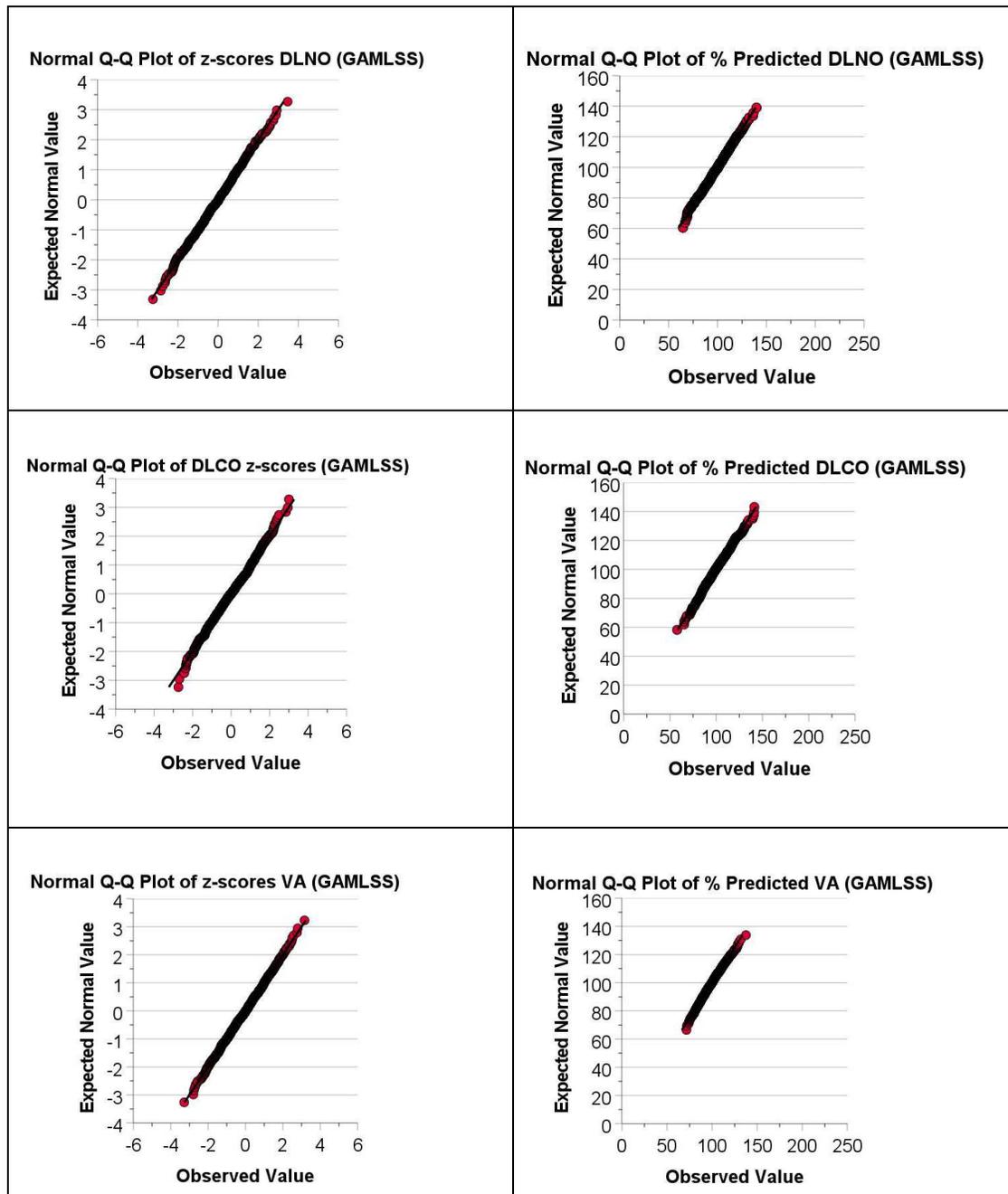


Figure S6. The mean and SD of the fitted scores for DLCO per model are displayed against the pulmonary function device used. There is a statistical difference between the fitted z-scores per model ($p < 0.001$) and pulmonary function device used ($p < 0.001$), and the interaction between model and device ($p < 0.001$). When the data were fitted to the Global Lung Function (GLI) Initiative GAMLSS reference equations, the DLCO z-scores were about 0.3 SD units higher than the z-scores obtained from the current study's segmented regression and GAMLSS two models when the data was obtained from the MasterScreen Pro device. However, when the data originated from the Hyp'Air Compact device, the z-scores fitted to the GLI reference equations were about 1.0 SD units higher than the z-scores obtained from the other two models ($p < 0.001$). The GLI GAMLSS reference equations are based on a 10 s breath-hold time, allowing for a better homogenous gas penetration in the lung. Indeed, a 5-6 s breath-hold time will result in poorer gas penetration in the lung. The Hyp'Air Compact device is represented by open white circles, and the Jaeger MasterScreen Pro is represented by solid, black-filled circles.

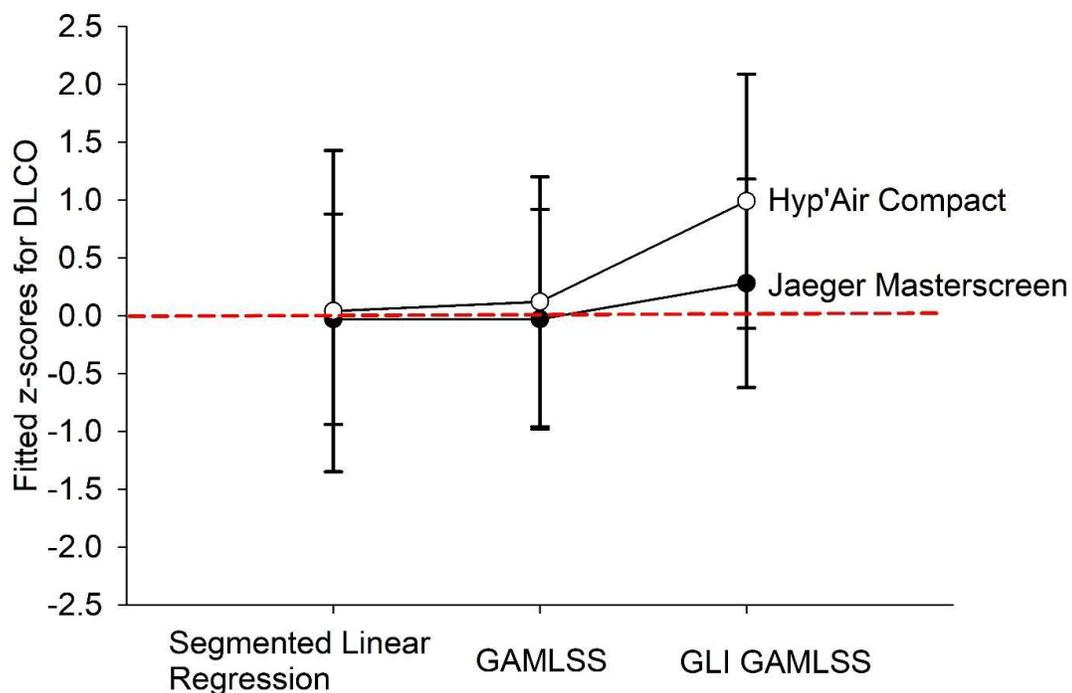


Figure S7. The mean and SD of the fitted z-scores for DLNO per model are displayed against the pulmonary function device used. There is no statistical difference between the fitted z-scores per model ($p = 0.11$), pulmonary function device used ($p = 0.86$), or the interaction between model and device ($p = 0.67$). Open white circles represent the Hyp'Air Compact device, and the Jaeger MasterScreen Pro is represented by solid, black-filled circles.

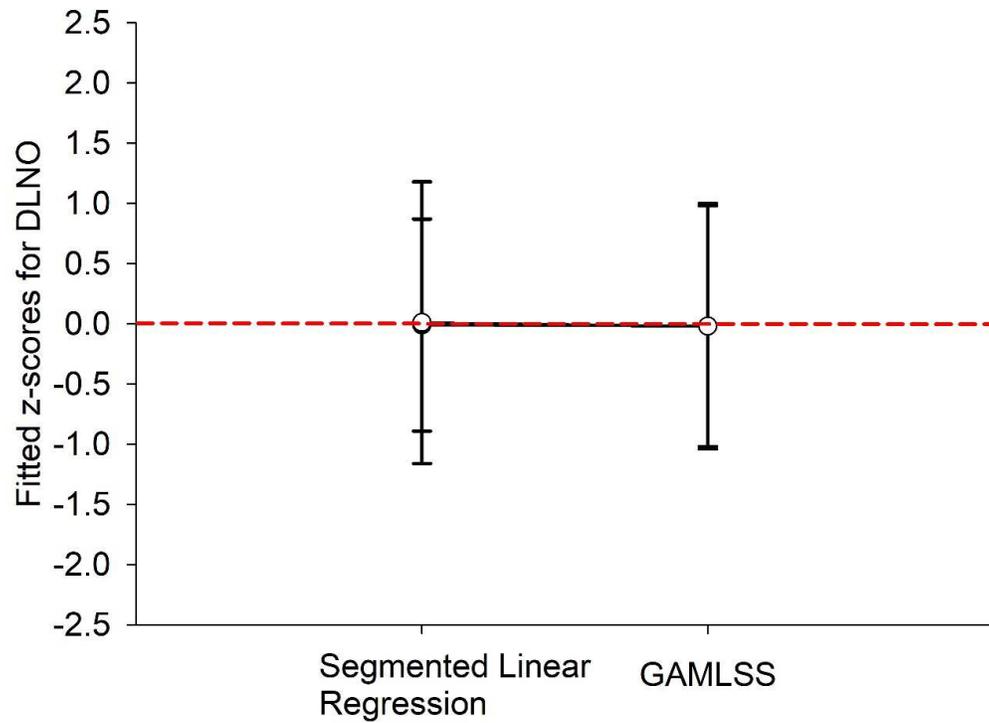


Figure S8. The mean and SD of the fitted z-scores for alveolar volume (VA) per model are displayed against the pulmonary function device used. There is a statistical difference between the fitted z-scores per model ($p < 0.001$) and pulmonary function device used ($p = 0.001$), as well as the interaction between model and device ($p < 0.001$). When the data originated from the Hyp'Air Compact device, the z-scores fitted to the GLI reference equations were 0.6 SD units higher than the z-scores obtained from the other two models ($p < 0.001$). The GLI GAMLSS reference equations are based on a 10 s breath-hold time, allowing for a better homogenous gas penetration in the lung. Indeed, a 5-6 s breath-hold time will result in poorer gas penetration in the lung. The Hyp'Air Compact device is represented by open white circles, and the Jaeger MasterScreen Pro is represented by solid, black-filled circles.

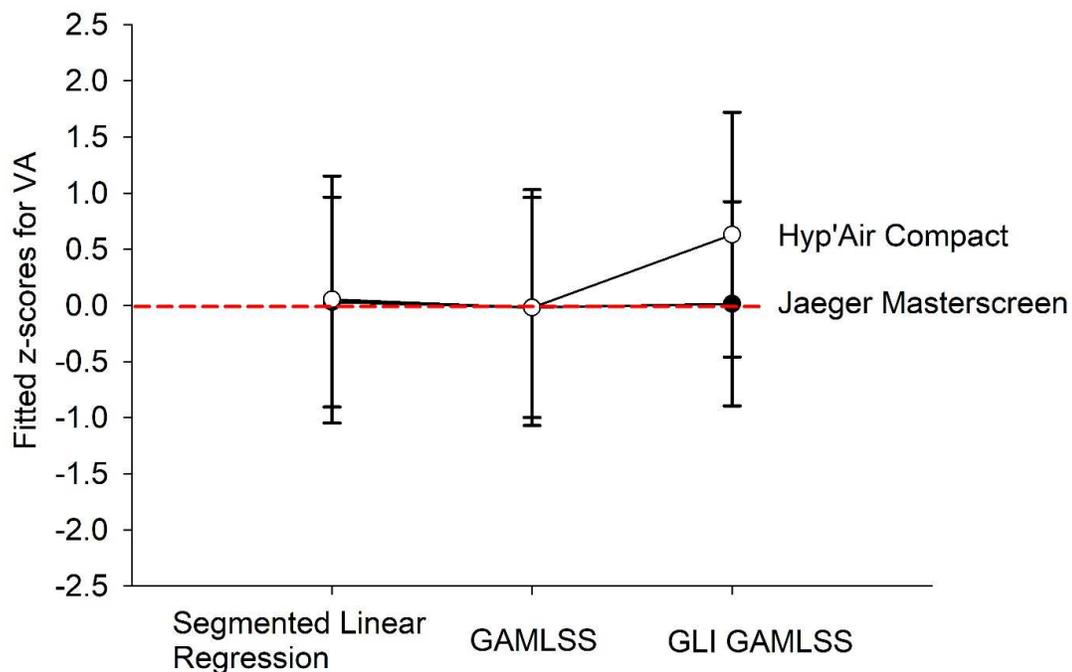


Figure S9A. Comparison between LLN for DLCO obtained from segmented linear regression models and the LLN for DLCO obtained from this study's GAMLSS equations. Both devices are combined ($n = 1076$). Units are mL/min/mmHg. The dotted black line is the line of identity; the solid red line is the regression line. The mean absolute difference between the LLN from both models was 1.1 (SD 1.1) mL/min/mmHg [range = 0 to 7.4 mL/min/mmHg]. Systematic differences (y-intercept) = 1.98 [95% CI = 1.77 to 2.20]; Proportional differences (slope) = 0.91 [0.90 to 0.92], Random differences (residual standard deviation) = 0.96 [95% CI = -1.88 to 1.88]; Linear model validity (Cusum test for linearity) = $p < 0.01$; Spearman rank correlation coefficient = 0.98.

LLN for DLCO
GAMLSS = $0.91 \cdot (\text{LLN for DLCO Segmented regression}) + 1.98$

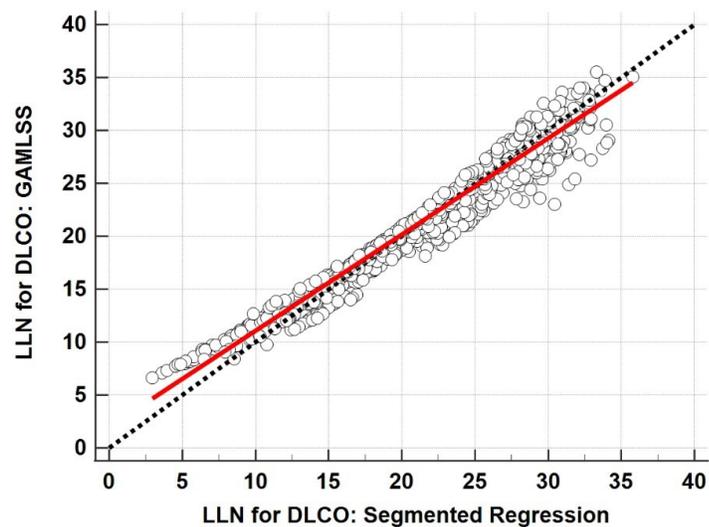


Figure S9B. Residuals plot by rank number. The residuals are the differences between the predicted values and the observed values. Red boxes indicate outliers (4 SD).

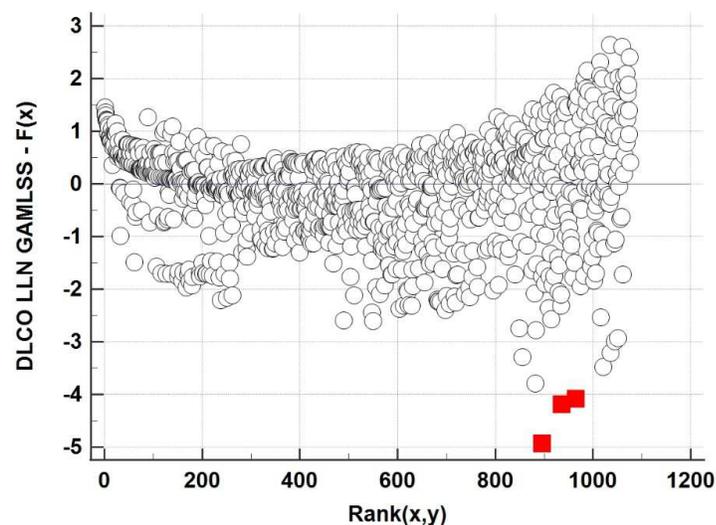


Figure S10A (Top panel). This plot demonstrates that when the LLN for DLCO tends to be on the low end (i.e., < 12.5 mL/min/mmHg), the GAMLSS models usually have a higher LLN. When the LLN for DLCO > 25 mL/min/mmHg, the segmented linear regression models can sometimes have a much higher LLN compared to GAMLSS models. Fifty-one subjects (4.7%) fall outside of ± 3.0 mL/min/mmHg difference. Both lung function devices are combined ($n = 1076$). Units are mL/min/mmHg. **Figure S10B (Lower panel).** The absolute difference in the LLN between both models versus the mean LLN from both models (mL/min/mmHg). Kendall's Tau correlation coefficient = -0.05 [95% CI = -0.10 to -0.001] ($p = 0.01$).

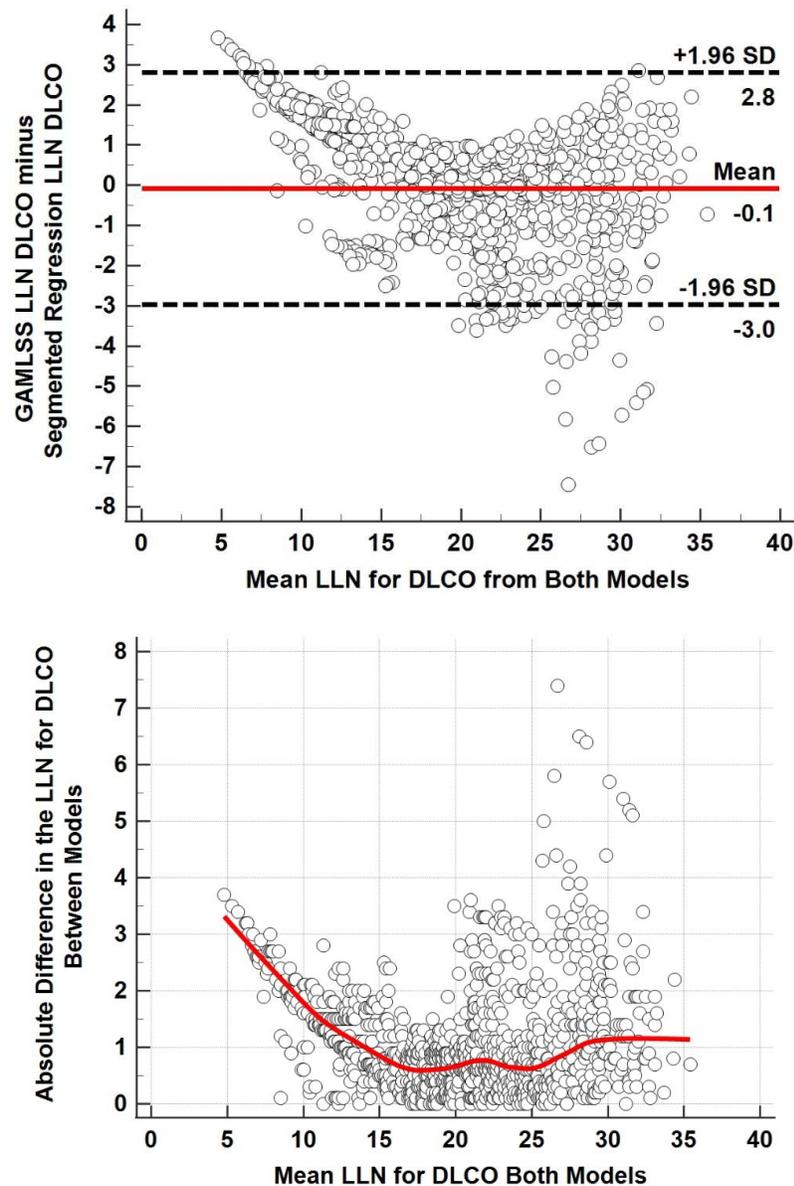


Figure S11A. Comparison between LLN for DLNO obtained from segmented linear regression models and the LLN for DLNO obtained from this study's GAMLSS equations. Both devices are combined ($n = 1076$). Units are mL/min/mmHg. The dotted black line is the line of identity; the solid red line is the regression line. The mean absolute difference between the LLN from both models was 6.2 (SD 5.5) mL/min/mmHg [range = 0 to 37 mL/min/mmHg]. Systematic differences (y-intercept) = 14.0 [95% CI = 12.9 to 15.0]; Proportional differences (slope) = 0.86 [0.85 to 0.87], Random differences (residual standard deviation) = 4.7 [95% CI = -9.2 to 9.2]; Linear model validity (Cusum test for linearity) = $p < 0.01$; Spearman's rank correlation coefficient = 0.98.

LLN for DLNO
GAMLSS = 0.86 · (LLN
for DLNO Segmented
Regression) + 14.0

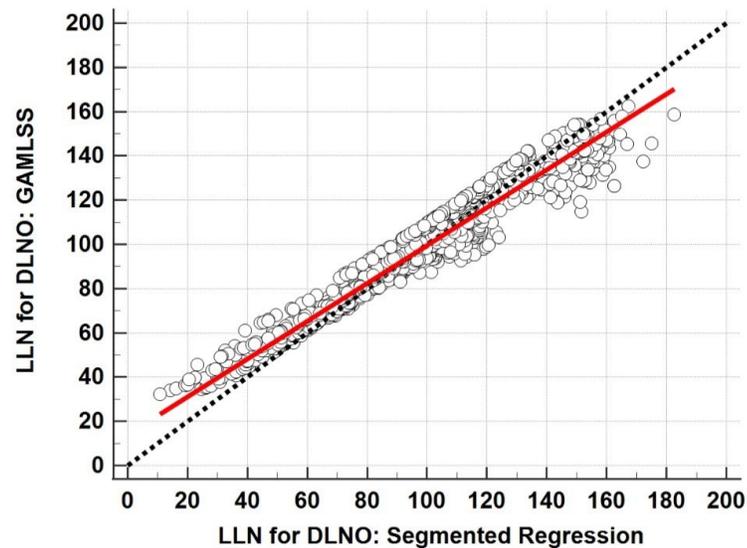


Figure S11B. Residuals plot by rank number. The residuals are the differences between the predicted values and the observed values. Red boxes indicate outliers (4 SD).

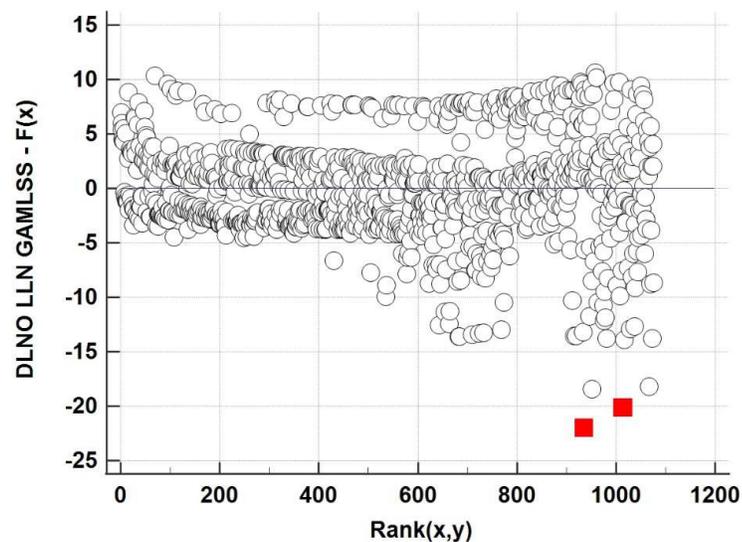


Figure S12A (Top panel). This plot demonstrates that when the DLNO is on the low end (< 60 mL/min/mmHg), the GAMLSS models usually have a higher LLN, and when the LLN for DLNO > 130 mL/min/mmHg, the segmented linear regression models can sometimes have a much higher LLN compared to GAMLSS models. Sixty-nine subjects (6.4%) fall outside of ± 16.0 mL/min/mmHg. Both lung function devices are combined ($n = 1076$). Units are mL/min/mmHg.

Figure S12B (Lower panel). The absolute difference in the LLN between both models versus the mean LLN from both models (mL/min/mmHg). Kendall's Tau correlation coefficient = 0.06 [95% CI = 0.01 to 0.10] ($p = 0.006$).

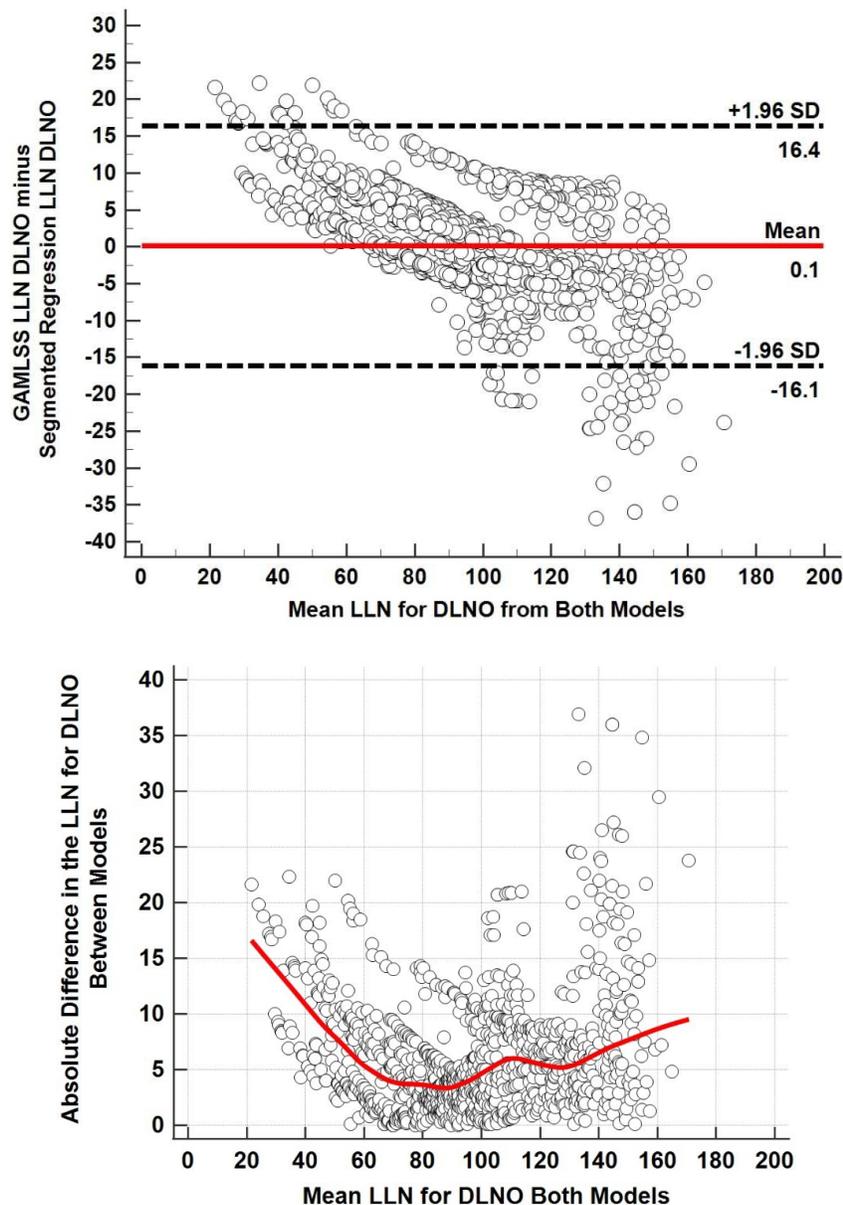


Figure S13A. Comparison between LLN for alveolar volume (VA) obtained from segmented linear regression models and the LLN for VA obtained from this study's GAMLSS equations. Both devices are combined ($n = 1076$). Units are in L. Dotted black line is the line of identity; the solid red line is the regression line. The mean absolute difference between the LLN from both models was 0.18 (SD 0.16) L [range = 0.0 to 1.38 L]. Systematic differences (y-intercept) = 0.22 [0.18 to 0.26]; Proportional differences (slope) = 0.97 [0.96 to 0.98], Random differences (residual standard deviation) = 0.14 [-0.28 to 0.28]; Linear model validity (Cusum test for linearity) = $p < 0.01$; Spearman rank correlation coefficient = 0.99.

$$\text{LLN VA GAMLSS} = 0.968 \cdot (\text{LLN for VA Segmented Regression}) + 0.22$$

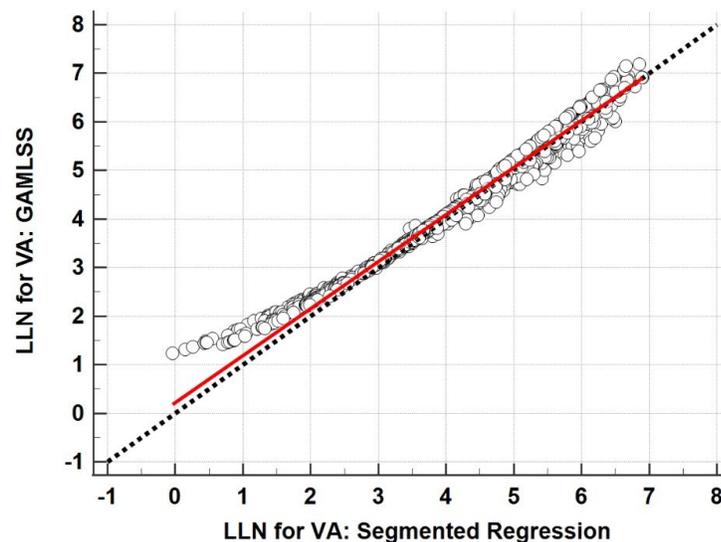


Figure S13B. Residuals plot by rank number. The residuals are the differences between the predicted values and the observed values. Red boxes indicate outliers (4 SD).

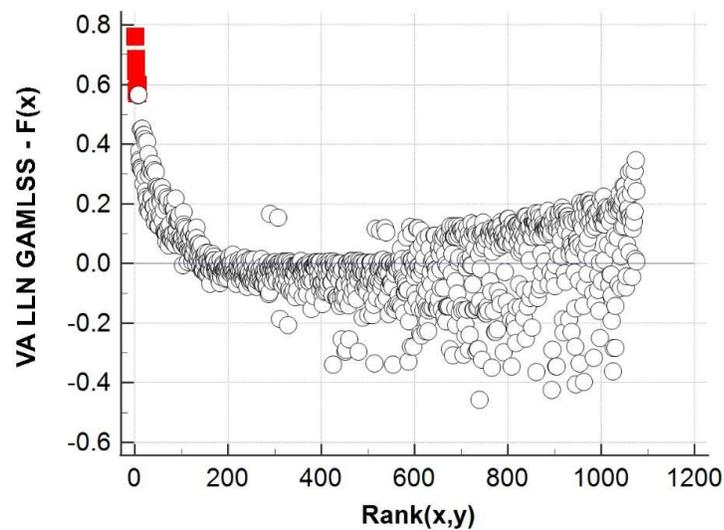


Figure S14A (Top panel). This plot demonstrates that when the VA is on the low end (< 2.5 L), the GAMLSS models developed in this study tend to have a higher LLN. When VA is > 4 L, the segmented linear regression models can sometimes have a much higher LLN compared to GAMLSS models. Seventy-nine subjects (7.3%) fall outside of ± 0.40 L. Both lung function devices are combined ($n = 1076$). Units are L. **Figure S14B (Lower panel).** The absolute difference in the LLN between both models versus the mean LLN from both models (L). Kendall's Tau correlation coefficient = 0.07 [95% CI = 0.02 to 0.13] ($p = 0.0004$).

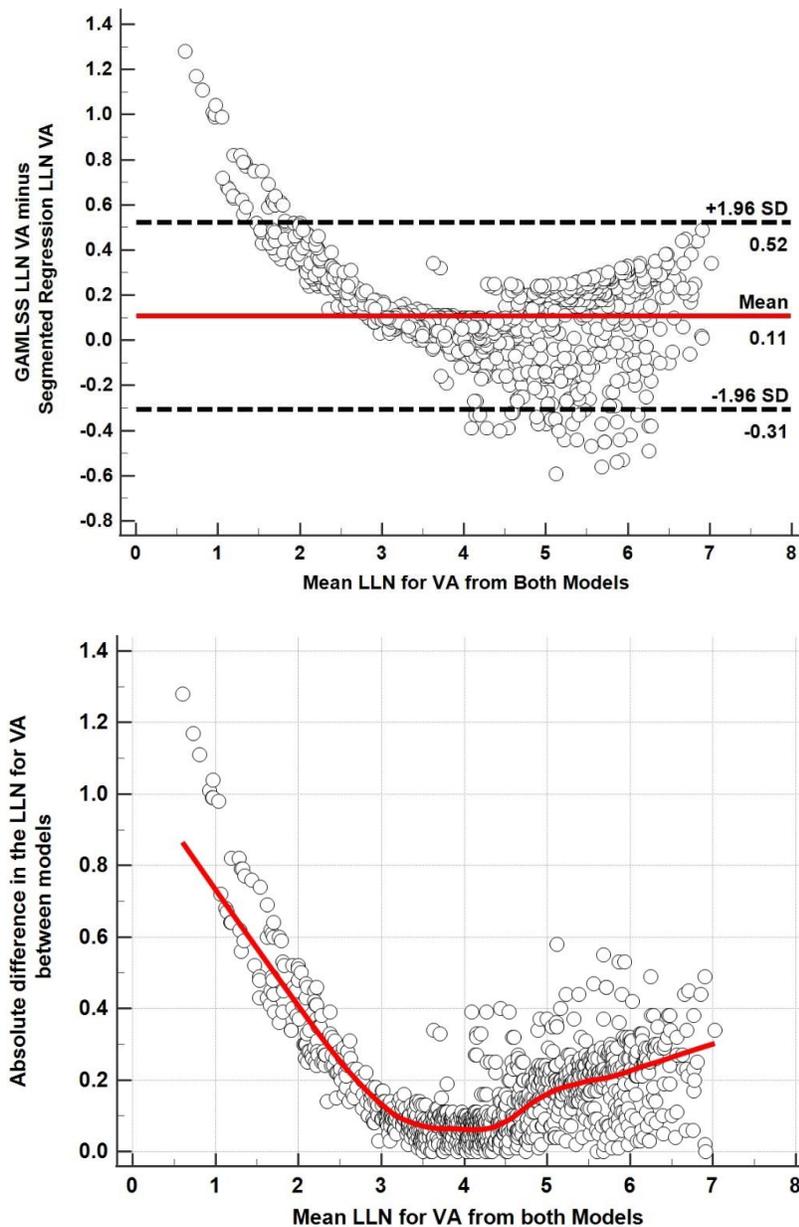
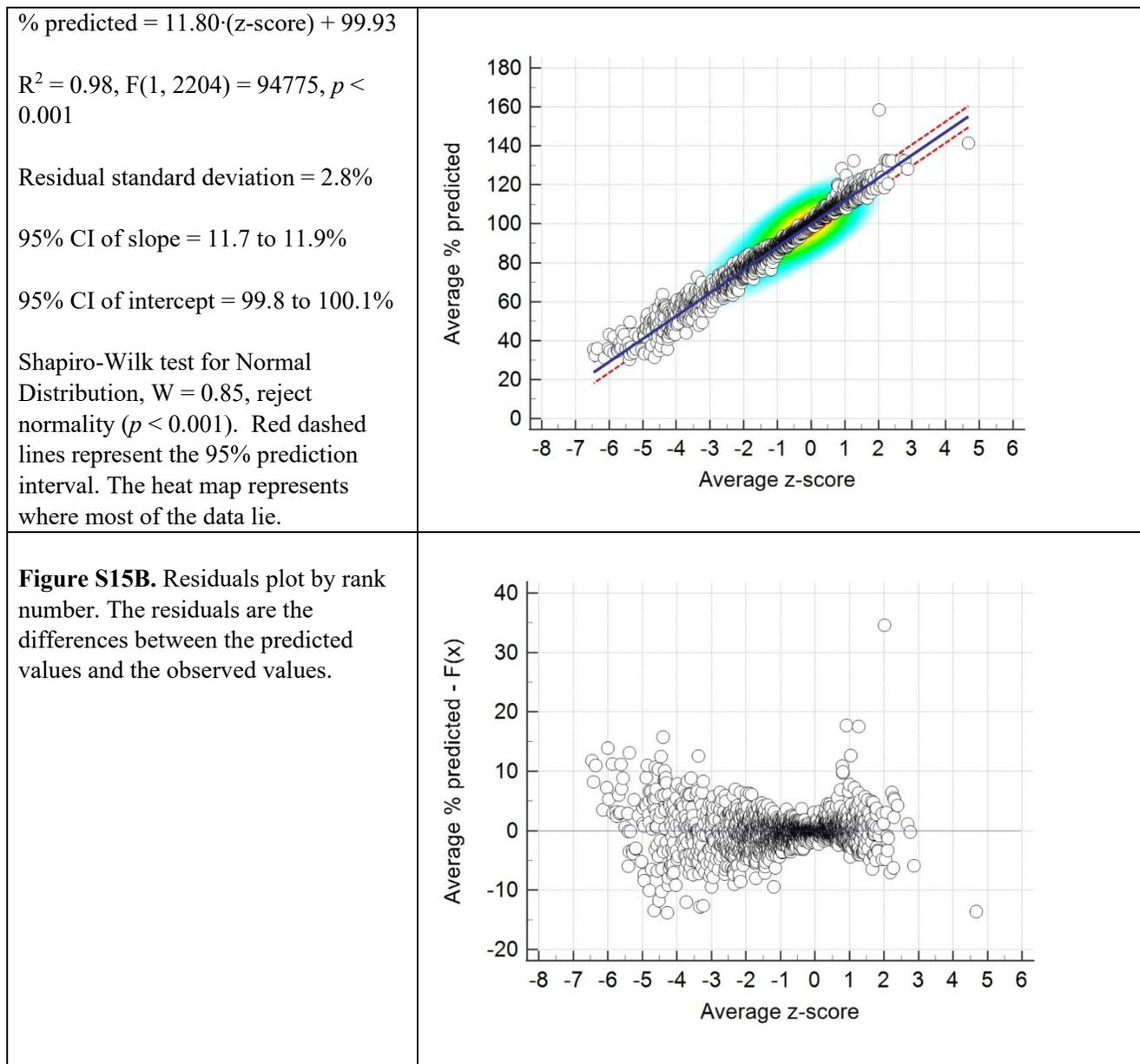


Figure S15A. Regression analysis demonstrates the linear association between percent predicted values and z-scores in both healthy and diseased subjects combined (n=2206). Each circle represents an average z-score combining DLNO, DLCO, and VA from both GAMLSS and segmented regression models. The data consisted of pooled data on healthy subjects ^{1 4-6 9} [n = 1076 subjects 5 to 95 years old, 546 males, 530 females] as well as diseased patients [n = 683 patients, 8 to 87 years of age, 289 males, 394 females]¹⁰⁻²⁰. Also, 447 control subjects from studies with cardiopulmonary disease were used^{13 14 17 18}. Three-hundred twenty-nine out of 2206 subjects (15%) include unpublished data from Cochin Hospital in Paris, France (Courtesy of Dr. Anh-Tuan Dinh Xuan, with permission). The mean breath-hold times were 6.1 s (SD = 1.9 s) for all data.



References

1. Munkholm M, Marott JL, Bjerre-Kristensen L, et al. Reference equations for pulmonary diffusing capacity of carbon monoxide and nitric oxide in adult Caucasians. *Eur Respir J* 2018;52(1):1500677. doi: 10.1183/13993003.00677-2015
2. Zavorsky GS, Hsia CC, Hughes JM, et al. Standardisation and application of the single-breath determination of nitric oxide uptake in the lung. *Eur Respir J* 2017;49(2):1600962. doi: 10.1183/13993003.00962-2016
3. Robert CP. Simulation of Truncated Normal Variables. *Stat Comput* 1995;5(2):121-25. doi: Doi 10.1007/Bf00143942
4. Aguilaniu B, Maitre J, Glenet S, et al. European reference equations for CO and NO lung transfer. *Eur Respir J* 2008;31(5):1091-7. doi: 10.1183/09031936.00063207
5. van der Lee I, Zanen P, Stigter N, et al. Diffusing capacity for nitric oxide: reference values and dependence on alveolar volume. *Respir Med* 2007;101(7):1579-84. doi: 10.1016/j.rmed.2006.12.001
6. Zavorsky GS, Cao J, Murias JM. Reference values of pulmonary diffusing capacity for nitric oxide in an adult population. *Nitric Oxide* 2008;18(1):70-9. doi: 10.1016/j.niox.2007.10.002
7. Gray G, Zamel N, Crapo RO. Effect of a simulated 3,048 meter altitude on the single-breath transfer factor. *Bull Eur Physiopathol Respir* 1986;22(5):429-31. [published Online First: 1986/09/01]
8. Fryar CD, Carroll MD, Gu Q, et al. Anthropometric Reference Data for Children and Adults: United States, 2015-2018. Analytical and Epidemiological Studies. National Center for Health Statistics. Washington, DC: Vital Health Statistics, 2021:36 pages.
9. Thomas A, Hanel B, Marott JL, et al. The single-breath diffusing capacity of CO and NO in healthy children of European descent. *PLoS One* 2014;9(12):e113177. doi: 10.1371/journal.pone.0113177
10. Radtke T, Benden C, Maggi-Beba M, et al. Intra-session and inter-session variability of nitric oxide pulmonary diffusing capacity in adults with cystic fibrosis. *Respir Physiol Neurobiol* 2017;246:33-38. doi: 10.1016/j.resp.2017.08.002

11. van der Lee I, Zanen P, Biesma DH, et al. The effect of red cell transfusion on nitric oxide diffusing capacity. *Respiration* 2005;72(5):512-6. doi: 10.1159/000087676
12. Moinard J, Guénard H. Determination of lung capillary blood volume and membrane diffusing capacity in patients with COLD using the NO-CO method. *Eur Respir J* 1990;3(3):318-22.
13. van der Lee I, Gietema HA, Zanen P, et al. Nitric oxide diffusing capacity versus spirometry in the early diagnosis of emphysema in smokers. *Respir Med* 2009;103(12):1892-7. doi: 10.1016/j.rmed.2009.06.005
14. Magini A, Apostolo A, Salvioni E, et al. Alveolar-capillary membrane diffusion measurement by nitric oxide inhalation in heart failure. *Eur J Prev Cardiol* 2015;22(2):206-12. doi: 10.1177/2047487313510397
15. Apostolo A, Paolillo S, Contini M, et al. Comprehensive effects of left ventricular assist device speed changes on alveolar gas exchange, sleep ventilatory pattern, and exercise performance. *J Heart Lung Transplant* 2018;37(11):1361-71. doi: 10.1016/j.healun.2018.07.005
16. Magini A, Contini M, Spadafora E, et al. Variability in pulmonary diffusing capacity in heart failure. *Respir Physiol Neurobiol* 2020;280:103473. doi: 10.1016/j.resp.2020.103473
17. Barisione G, Brusasco C, Garlaschi A, et al. Lung diffusing capacity for nitric oxide as a marker of fibrotic changes in idiopathic interstitial pneumonias. *J Appl Physiol* 2016;120(9):1029-38. doi: 10.1152/jappphysiol.00964.2015
18. Sorensen JK, Buchvald F, Berg AK, et al. Ventilation inhomogeneity and NO and CO diffusing capacity in ex-premature school children. *Respir Med* 2018;140:94-100. doi: 10.1016/j.rmed.2018.06.006
19. Dressel H, Filser L, Fischer R, et al. Lung diffusing capacity for nitric oxide and carbon monoxide: dependence on breath-hold time. *Chest* 2008;133(5):1149-54. doi: 10.1378/chest.07-2388
20. Degano B, Soumagne T, Delaye T, et al. Combined measurement of carbon monoxide and nitric oxide lung transfer does not improve the identification of pulmonary hypertension in systemic sclerosis. *Eur Respir J* 2017;50(4) doi: 10.1183/13993003.01008-2017