



ARTICLE

# Development and validation of models to predict respiratory function in persons with long-term spinal cord injury

Anja M. Raab<sup>1</sup> · Sonja de Groot<sup>2,3</sup> · David J. Berlowitz<sup>4</sup> · Marcel W. M. Post<sup>5,6</sup> · Jacinthe Adriaansen<sup>5</sup> · Maria Hopman<sup>7</sup> · Gabi Mueller<sup>1</sup>

Received: 4 March 2019 / Revised: 31 May 2019 / Accepted: 31 May 2019 / Published online: 19 June 2019  
© The Author(s), under exclusive licence to International Spinal Cord Society 2019

## Abstract

**Study Design** Multicenter, cross-sectional study.

**Objectives** To validate previously developed respiratory function prediction models for persons with long-term spinal cord injury (SCI) and if necessary develop and validate new models.

**Setting** Ten SCI rehabilitation centers.

**Methods** Five respiratory function parameters were measured in adults with chronic, traumatic, motor complete SCI (C4–T12). First, the models published in 2012 were validated using Bland–Altman plots. Then, new models were calculated using 80% of the dataset by multiple regression analysis with the candidate predictors gender, age, height, weight, time post injury (TPI), lesion level, and smoking. In a third step, the new models were validated using the other 20% of the dataset by Bland–Altman plots.

**Results** In total 613 participants were included. For persons with long-term SCI, the 2012 models were poorly predictive, especially for respiratory muscle strength ( $R^2 = 0.4$ ). Significant predictors for all respiratory function parameters in the new models ( $R^2 = 0.7–0.8$ ) were lesion level, gender and weight. Small effects on single outcome parameters were observed for TPI and age whereas smoking had no effect. For the new models the mean differences between measured and predicted values for respiratory muscle strength were  $4.0 \pm 36.0$  cm H<sub>2</sub>O and for lung function parameters  $-0.5 \pm 1.2$  L (FVC),  $-0.3 \pm 0.9$  L (FEV1) and  $-0.5 \pm 2.0$  L/s (PEF).

**Conclusion** We did not find better models for lung function in long-term SCI but those for respiratory muscle strength showed better accuracy.

**Sponsorship** The content of this publication was developed under grant from Wings for Life, grant number WFL-CH-017/14.

## Introduction

Prediction of respiratory function is an essential component of understanding respiratory status in daily clinical practice,

but the prediction of respiratory function based on able-bodied values has limitations in persons with spinal cord injury (SCI). The respiratory function in SCI is reduced due to the partly or complete impairment of the respiratory muscles [1, 2]. This impairment is dependent on the completeness and level of the lesion [3, 4]. Immediately after the onset of the injury the respiratory function is reduced but during inpatient rehabilitation and the first year thereafter the

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41393-019-0313-1>) contains supplementary material, which is available to authorized users.

✉ Anja M. Raab  
anja.raab@paraplegie.ch

- 1 Clinical Trial Unit, Swiss Paraplegic Centre, Nottwil, Switzerland
- 2 Amsterdam Rehabilitation Research Center, Reade, Amsterdam, The Netherlands
- 3 University of Groningen, University Medical Center Groningen, Center for Human Movement Sciences, Groningen, The Netherlands

- 4 Institute for Breathing and Sleep, Austin Health, Heidelberg, Australia
- 5 Center of Excellence for Rehabilitation Medicine, Brain Center Rudolf Magnus, University, Medical Center Utrecht and De Hoogstraat Rehabilitation, Utrecht, The Netherlands
- 6 University of Groningen, University Medical Center Groningen, Center for Rehabilitation, Groningen, The Netherlands
- 7 Department of Physiology, Radboud University Nijmegen, Nijmegen, The Netherlands

respiratory function can partly recover [1, 5–9]. At some point this initial improvement turns into a decline [5] which exceeds the normal age-related decline [10]. This age-related decline in able-bodied persons is mainly related to gender and height [11–13]. In persons with SCI the decline in respiratory function in the first years after rehabilitation also appears associated with higher body mass index, lower inspiratory muscle strength and declined physical fitness [6, 10, 14]. In general, large interpersonal differences in changes of respiratory function might be possible [10]. The intention of this study is to promote regular measurements of the respiratory function by spirometry but also give the clinicians an additional tool for better individualized interpretation of the measured values. SCI-specific prediction models can help identify patients whose respiratory function is below the norm (i.e. predicted values) and thus, clinicians can deliver targeted interventions to this group of patients. Individual pulmonary diagnostics and therapy are therefore central treatment targets to maintain health and quality of life [15].

In 2012 a reference value calculator was developed to obtain an overview of respiratory function status while taking into account the spinal cord lesion level [15]. To develop these initial models, 440 persons were tested of whom 150 were between six months and two years post injury; a period when lung function is known to improve after a SCI. As such, it is unclear whether these models are accurate for persons several years post injury.

Therefore, the aim of this study was to test the accuracy of the 2012 models in persons with long-term SCI and if necessary develop and validate new prediction models of respiratory function. We hypothesize that the lesion-specific reference models published in 2012 [15] are not sufficiently accurate for long-term injured persons and that new models are required to improve decision making about treatment [16].

## Methods

### Design and setting

A multicenter, cross-sectional study with 10 SCI rehabilitation centers was performed (The Netherlands ( $n = 8$ ), Australia ( $n = 1$ ), Switzerland ( $n = 1$ )).

In the Netherlands, data from the research program ‘Active Lifestyle Rehabilitation Interventions in aging Spinal Cord Injury’ (ALLRISC) have been used [17]. Reporting follows the TRIPOD checklist [18].

### Study population

Inclusion criteria for this study were: long-term SCI (>2 years after injury, median 20 years) due to trauma, at least

18 years of age, motor complete injury (American Spinal Injury Association Impairment Scale (AIS) A and B), right and left motor level between C4–T12. Only data from motor complete injured persons were analyzed to increase the homogeneity of the sample.

Persons with acute or chronic respiratory diseases were excluded, as well as persons with severe scoliosis, progressive neurological diseases, bronchodilators or any other medication that could have adversely affected respiratory function at the time of assessment. Persons with respiratory failure, ventilator or tracheostomy dependency, traumatic brain injury or mental disorders were excluded. Persons with obstructive sleep apnea were not excluded.

### Procedure

The outcome measures were forced vital capacity (FVC), forced expiratory volume in one second ( $FEV_1$ ), peak expiratory flow (PEF), maximal inspiratory ( $PI_{max}$ ) and maximal expiratory pressure ( $PE_{max}$ ). Gender, age, height (measured or asked and not arm span as a surrogate), weight and time post injury (TPI) at the time of the respiratory function measurement were recorded. Information on smoking history was also collected. Each center entered their anonymized data in a centralized study database. In the Swiss center the data were collected from November 2002 to June 2018, in the Dutch centers from November 2011 to February 2014 and in the Australian center from April 1996 to May 2014. The lung function measurements in the Netherlands and in Australia were performed for research purposes [19, 20]. The lung function measurements in Switzerland were performed during clinical practice. Measurements of respiratory function were performed only once per participant in all centers by appropriately trained personnel according to the ATS/ERS guidelines [21]. Participants were sitting upright in their own wheelchair and breathed through a mouthpiece while wearing a nose clip. Each measurement was repeated until three reproducible measurements were registered and the highest value was quoted for further analysis. The spirometers for measurement of FVC,  $FEV_1$ , and PEF were calibrated daily. To measure FVC and  $FEV_1$ , the participants were instructed to exhale fully from total lung capacity. The maximum airflow during this forced expiration was measured to assess PEF.  $PI_{max}$  and  $PE_{max}$  were measured using a respiratory pressure meter (Micro RPM, Micro Medical, Hoechberg, Germany). For the measurements the participants carried out their maximal in- and expiratory maneuvers from residual volume and total lung capacity, respectively [22]. Abdominal binders were removed for all measurements.

## Statistical analyses

Demographic data are presented as frequency or as median (25% and 75% quartile). Depending on the distribution of the data, an independent *t*-test or a Mann–Whitney U test was used to investigate differences between the model- and validation sample for all numerical data. Differences between the model- and the validation sample of categorical data were compared using chi-square tests. Lesion level was treated as a categorical variable with four different lesion groups, two for tetraplegia grouped in lesion level C4–C5 and C6–C8 and two for paraplegia grouped in lesion level T1–T6 and T6–T12. These lesion level groups were built based on the level of innervation of the main respiratory muscles. Lesion group T1–T6 was the reference group.

All analysis except the multilevel regression analysis were performed using SPSS (Version 18.0.3, IBM, Somers, NY, USA). Statistical significance was set at  $\alpha \leq 0.05$ . The multilevel regression analysis were done with the multi-level modelling program MLwin (MLwin, version 1.1; Center for Multilevel Modelling, Institute for Education, London, UK) [23, 24].

The project was divided into three phases as illustrated in Fig. 1.

Phase 1– Accuracy of the statistical models published in 2012

Bland–Altman plots were used [25] to quantify the agreement and to evaluate a bias between the predicted values (calculated with the lesion-specific reference values published in 2012 [15]) and measured respiratory function values. The mean differences between measured and predicted values were assessed against their mean and the limits of agreement (mean  $\pm$  1.96 standard deviations (SD) of the difference). A visual examination of the plot allowed us to evaluate the global agreement between the two different methods [25]. Adding a regression line of the

difference and confidence interval limits into the Bland–Altman plots assisted with describing any proportional difference [25]. To test how much the measured test scores are spread around the ‘true’ score the standard error of measurement (SEM) was calculated. T-tests were performed to test for proportional bias in the old models (measured values versus predicted values). The ICCs (two-way random, absolute agreement) were calculated to determine the correlation between the predicted and measured respiratory function value. Values for ICCs indicate the following reliability: less than 0.5 poor, between 0.5 and 0.75 moderate, between 0.75 and 0.9 good, greater than 0.9 excellent [26]. The ICCs and Bland–Altman tests are appropriate for reliability analysis and it is recommended that both are used [27].

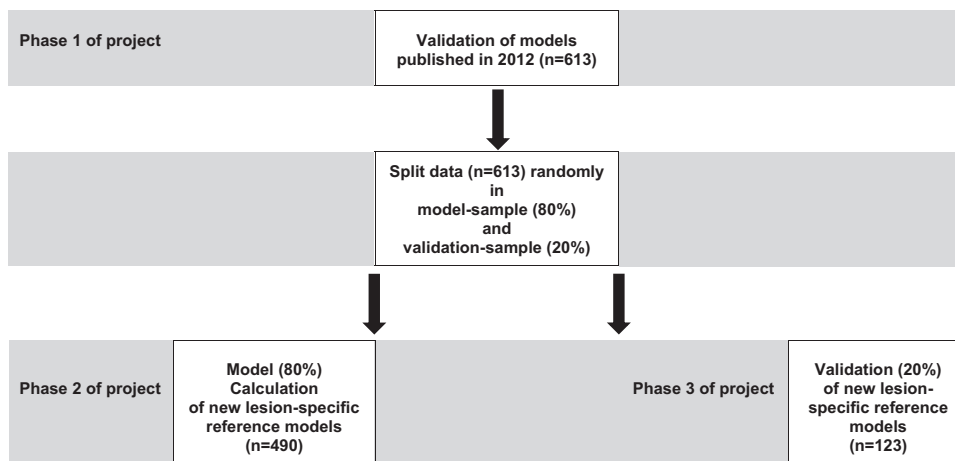
Data splitting procedure (preparation for phase 2 and 3)

The data were split into two samples: (1) a sample to develop the new statistical models, the ‘model sample’ (80%) (phase 2) and (2) a sample to cross-validate the reliability of the models, the ‘validation sample’ (20%) (phase 3). Random numbers were generated and the data split on basis of the order of these numbers (model 80% and validation 20%) [28, 29].

Phase 2 - Development of lesion-specific reference models

For the development of the new lesion-specific reference models we principally followed the 2012 statistical procedure but now added the parameter country as level. For each of the respiratory function parameters FVC, FEV<sub>1</sub>, PEF, PI<sub>max</sub>, and PE<sub>max</sub> one model was developed to determine the relationship of personal and lesion characteristics with respiratory function. The hierarchy in the data was as follows: individual participants (level 1) who were grouped in the participating centers (level 2) and the participating countries (level 3). In order to calculate the influence of the lesion level, three dummy variables were used and the lesion group T1–T6 with the most participants was defined as reference group. Further factors

Fig. 1 Flow-chart of the project



potentially influencing respiratory function, such as gender (male = 1, female = 0), age (years), height (cm), weight (kg), TPI (years) were added to a basic univariate multilevel regression equation. Information about smoking history (pack-years (years), ever (0 = never smoker, 1 = ever smoker), former (0 = no former smoker, 1 = former smoker) or current smoker (0 = no current smoker, 1 = current smoker)) were added to the basic univariate multilevel regression equation separately. Independent variables with  $p$ -values  $\leq 0.1$  were included in a subsequent multivariable equation. Model fit was assessed with the  $-2$  Log likelihood for the equations. A backward selection procedure was then carried out, excluding non-significant determinants ( $p \geq 0.05$ ) in order to create a final multivariable equation.

The predictive ability of each of the five models was judged based on the adjusted  $R^2$  (explained variance) as a statistical measure of accuracy.

### Phase 3 - Accuracy of the lesion-specific reference models

The validation sample was used to test the predictive value of the developed models. As in phase 1 of the project the predicted respiratory function values (calculated with the new models) were compared with the measured respiratory function values. Adding a regression line of the difference and confidence interval limits into the Bland–Altman plots assisted with describing any proportional difference [25].  $T$ -tests were performed to test for proportional bias in the new models (measured values versus predicted values). For the graphical illustration of the validation also residual plots were used. The SEM and ICCs were calculated.

## Results

A total of 613 participants were analysed; 346 participants (56%) with tetraplegia and 267 (44%) with paraplegia. The Swiss rehabilitation center provided data from 304 persons, the Dutch 215, and the Australian center 94 persons with long-term SCI. Demographic data are presented in Table 1. No differences in personal and lesion characteristics were found between the model- and the validation sample (Table 1). No measurements of respiratory muscle pressures from the Dutch centers and no measurements of PEF from the Australian center were available.

### Phase 1 - Accuracy of the statistical models published in 2012

The predictions of the models published in 2012 did not match the measured data of persons with long-term SCI well. The  $t$ -tests to test for proportional bias (measured versus predicted values) are significant for FEV<sub>1</sub> ( $p = 0.033$ ), PEF ( $p < 0.000$ ), PI<sub>max</sub> ( $p < 0.000$ ), PE<sub>max</sub> ( $p < 0.000$ ) and almost for FVC ( $p = 0.053$ ), that means that the measured values are significantly lower or higher than the predicted values. Thus, there is a systematic difference in the old models for all five respiratory function parameters and the models give an over- or underestimation. Figure 2 demonstrates this proportional bias for all of the variables examined. The 2012 equations underestimated the actual values in the lower range and overestimated in the upper range for all five parameters. The ICCs for the lung volumes (0.62–0.64) were moderate but the ICCs for the respiratory muscle strength values were only

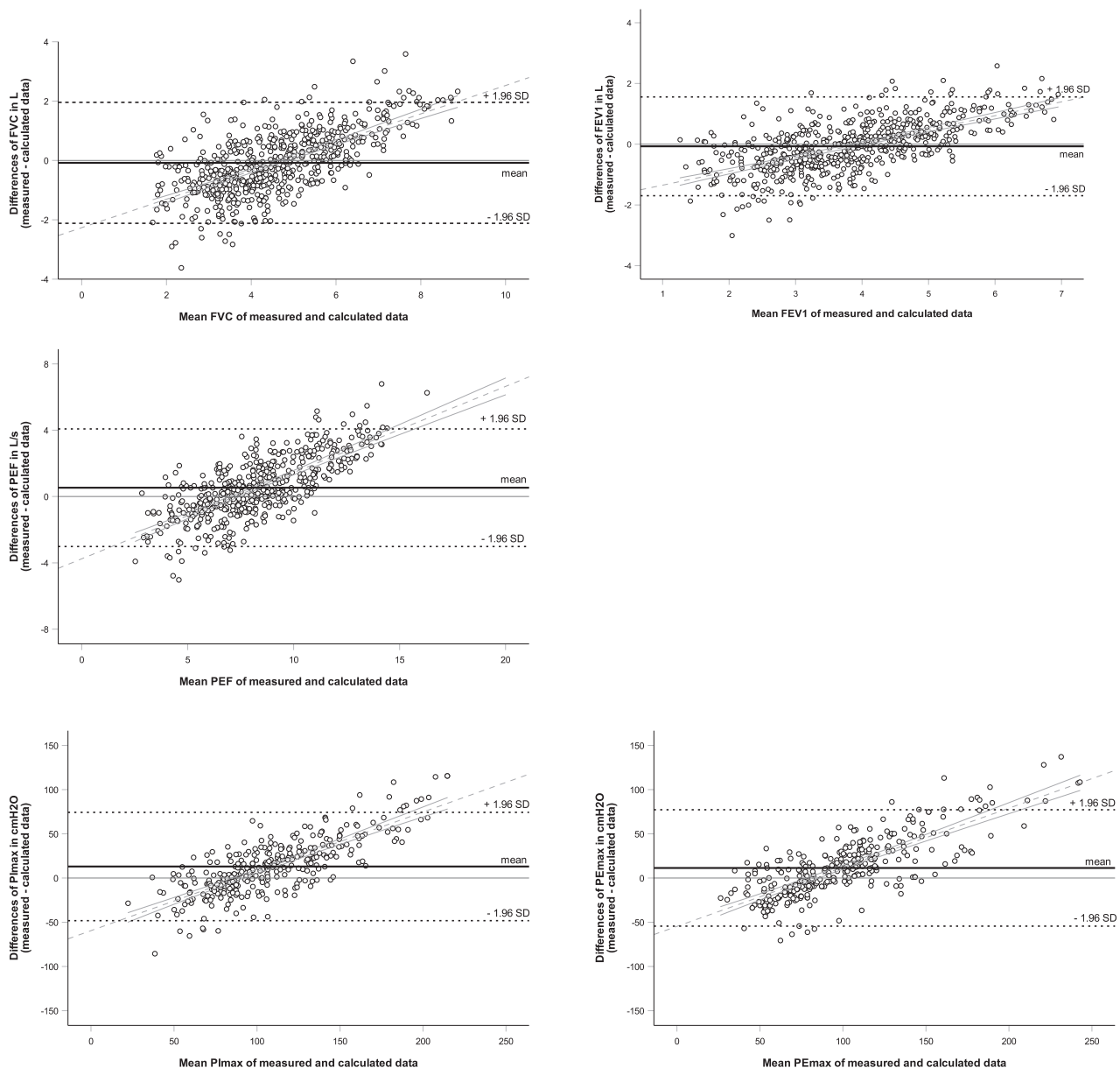
**Table 1** Characteristics of participants of the whole sample ( $n = 613$ ), divided into a model- and a validation-sample

Characteristics	Model-sample (80%) $n = 490$	Validation-sample (20%) $n = 123$	$p$ -value (Model-sample vs. Validation-sample)
Country	CH = 247 (50%) AUS = 77 (16%) NL = 166 (34%)	CH = 57 (46%) AUS = 17 (14%) NL = 49 (40%)	0.46 ‡
Male/Female [n]	400/90 (82%/18%)	95/28 (77% / 23%)	0.27 ‡
Lesiongroup	C4–C5 = 155 (32%) C6–C8 = 125 (26%) T1–T6 = 140 (29%) T7–T12 = 70 (14%)	C4–C5 = 29 (24%) C6–C8 = 37 (30%) T1–T6 = 42 (34%) T7–T12 = 15 (12%)	0.24 ‡
Age (years)	48 (40 to 58)	49 (41 to 58)	0.78 †
Time post injury (years)	20 (11 to 29)	22 (14 to 30)	0.09 †
Height (cm)	178 (172 to 184)	178 (170 to 184)	0.93 †
Weight (kg) ( $n = 489/123$ )	77 (67 to 89)	76 (63 to 86)	0.26 †
Ever smoker [n]/%	216 out of 431 (50%)	47 out of 103 (46%)	
Pack years ( $n = 430/110$ )	0.00 (0.00 to 8.5)	0.00 (0.00 to 15)	0.87 †

All data are reported as median (with 25% & 75% quartiles) unless otherwise indicated

CH Switzerland; AUS Australia; NL The Netherlands

‡Chi-square test; †Mann–Whitney U test



**Fig. 2** (phase 1). Bland-Altman plots of differences between the measured respiratory function values (FVC, FEV<sub>1</sub>, PEF, PI<sub>max</sub>, and PE<sub>max</sub>) of each participant and the calculated respiratory function values using the lesion-specific reference models published in 2012. The bold line represents the proportional bias and dashed lines show the limits of agreement. Dotted, diagonal line represents the regression line and confidence interval limits are presented as continuous,

diagonal lines. FVC regression line (95% CI):  $y = -2.27 (4.545 \text{ to } 4.717) + 0.48 (0.832 \text{ to } 0.997) *x$ . FEV<sub>1</sub> regression line (95% CI):  $y = -1.81 (3.786 \text{ to } 3.928) + 0.46 (0.810 \text{ to } 0.981) *x$ . PEF regression line (95% CI):  $y = -3.76 (7.537 \text{ to } 7.856) + 0.52 (0.959 \text{ to } 1.128) *x$ . PI<sub>max</sub> regression line (95% CI):  $y = -59.5 (93.529 \text{ to } 98.852) + 0.67 (0.838 \text{ to } 0.995) *x$ . PE<sub>max</sub> regression line (95% CI):  $y = -54.66 (85.585 \text{ to } 91.181) + 0.67 (0.839 \text{ to } 0.997) *x$

poor to moderate with 0.41 for PI<sub>max</sub> and 0.40 for PE<sub>max</sub>, respectively (Table 2) [26]. The residual- and scatterplots can be found in online supplement [25, 30, 31].

## Phase 2 - Development of lesion-specific reference models

Due to the poor accuracy of the models published in 2012 for long-term SCI, new statistical models to predict

respiratory reference values were developed with the model sample (80% of the total sample) (Table 3). The remaining candidate predictors after the backward regression models for all five respiratory function parameters were lesion level and gender. Age, height, weight, and TPI had an additional significant influence on a sub-set of parameters only. We explored different forms of TPI in advance (observed, linear, logarithmic, exponential fittings) and did not find substantial differences between the four fittings. Based on those

**Table 2** Phase 1: Accuracy for prediction of lung function and respiratory muscle strength for the whole sample of persons with long-term SCI with the models published in 2012

	FVC (L)	FEV <sub>1</sub> (L)	PEF (L/s)	PI <sub>max</sub> (cmH <sub>2</sub> O)	PE <sub>max</sub> (cmH <sub>2</sub> O)
<i>n</i>	610	613	500	333	333
Measured mean (SD)	3.0 (1.2)	2.5 (0.97)	5.7 (2.2)	79 (33)	69 (35)
Predicted mean (SD)	3.1 (0.72)	2.6 (0.59)	5.2 (1.2)	66 (16)	58 (16)
Average of measured and predicted values (SD)	3.1 (0.85)	2.5 (0.69)	5.4 (1.5)	72 (21)	64 (22)
ICC (95% CI)	0.63 (0.56 to 0.68)	0.64 (0.58 to 0.69)	0.62 (0.52 to 0.69)	0.41 (0.24 to 0.54)	0.40 (0.24 to 0.52)
Bland–Altman					
Difference (mean (SD))	−0.08 (1.0)	−0.07 (0.82)	0.53 (1.8)	13 (31)	11.4 (33)
Standard error of difference	0.04	0.03	0.08	1.7	1.8
95% CI for difference	−0.16 to 0.00	−0.14 to −0.01	0.37 to 0.69	9.7 to 16	7.8 to 14
95% limits of agreement	−2.1 to 2.0	−1.7 to 1.6	−3.0 to 4.1	−48 to 74	−54 to 77
Standard error of measurement (95% CI)	0.44 (−0.44 to 1.3)	0.35 (−0.34 to 1.1)	0.72 (−0.69 to 2.1)	13.6 (−13 to 40)	12 (−12 to 37)

CI confidence interval, ICC intraclass correlation coefficient, SCI spinal cord injury, SD standard deviation FVC forced vital capacity (L), FEV<sub>1</sub> forced expiratory volume in 1 s (L), PEF peak expiratory flow (L/s), PI<sub>max</sub> peak inspiratory pressure (cmH<sub>2</sub>O), PE<sub>max</sub> peak expiratory pressure (cmH<sub>2</sub>O)

**Table 3** Phase 2: Regression coefficients ( $\beta$ ) and 95% CI of the new models from the multilevel regression analysis of respiratory function parameters

	FVC ( <i>n</i> = 486) $\beta$ [95% CI]	FEV <sub>1</sub> ( <i>n</i> = 488) $\beta$ [95% CI]	PEF ( <i>n</i> = 393) $\beta$ [95% CI]	PI <sub>max</sub> ( <i>n</i> = 276) $\beta$ [95% CI]	PE <sub>max</sub> ( <i>n</i> = 276) $\beta$ [95% CI]
R <sup>2</sup>	0.78	0.74	0.69	0.69	0.69
Constant	0.52	1.2	4.7	62	71
Lesion groups					
Δ Lesion group C4–C5	−0.76 [−0.97 to −0.55]	−0.69 [−0.87 to −0.51]	−2.06 [−2.53 to −1.59]	−25 [−35 to 35]	−27 [−37 to −17]
Δ Lesion group C6–C8	−0.35 [−0.57 to −0.12]	−0.28 [−0.46 to −0.09]	−1.07 [−1.11 to −1.02]	−16 [−27 to −6.8]	−17 [−28 to −6.4]
Δ Lesion group T7–T12	0.10 [−0.16 to 0.369]	0.08 [−0.14 to 0.30]	0.31 [−0.20 to 0.83]	−6.2 [−18 to 6.3]	10 [−2.6 to 23]
Gender (male)	0.56 [0.31 to 0.82]	0.36 [0.15 to 0.57]	1.14 [0.64 to 1.64]	18 [8.9 to 28]	18 [8.1 to 28]
Age (years)	n.s.	−0.02 [−0.02 to −0.01]	n.s.	n.s.	n.s.
Height (cm)	0.02 [0.01 to 0.03]	0.01 [0.00 to 0.02]	n.s.	n.s.	n.s.
Weight (kg)	n.s.	0.01 [0.00 to 0.01]	0.02 [0.00 to 0.03]	0.24 [−0.00 to 0.47]	n.s.
Time post injury (years)	−0.01 [−0.02 to −0.00]	n.s.	n.s.	n.s.	n.s.
Smoking history	n.s.	n.s.	n.s.	n.s.	n.s.

\*Significant influencing factor ( $p \leq 0.05$ ); n.s.: not significant and removed from the final model

FVC forced vital capacity (L), FEV<sub>1</sub> forced expiratory volume in 1 s (L), PEF peak expiratory flow (L/s), PI<sub>max</sub> peak inspiratory pressure (cmH<sub>2</sub>O), PE<sub>max</sub> peak expiratory pressure (cmH<sub>2</sub>O)

$\beta$ : regression coefficient for each independent variable

CI: confidence interval

Δ: lesion group dummies with lesion group T1–T6 as reference

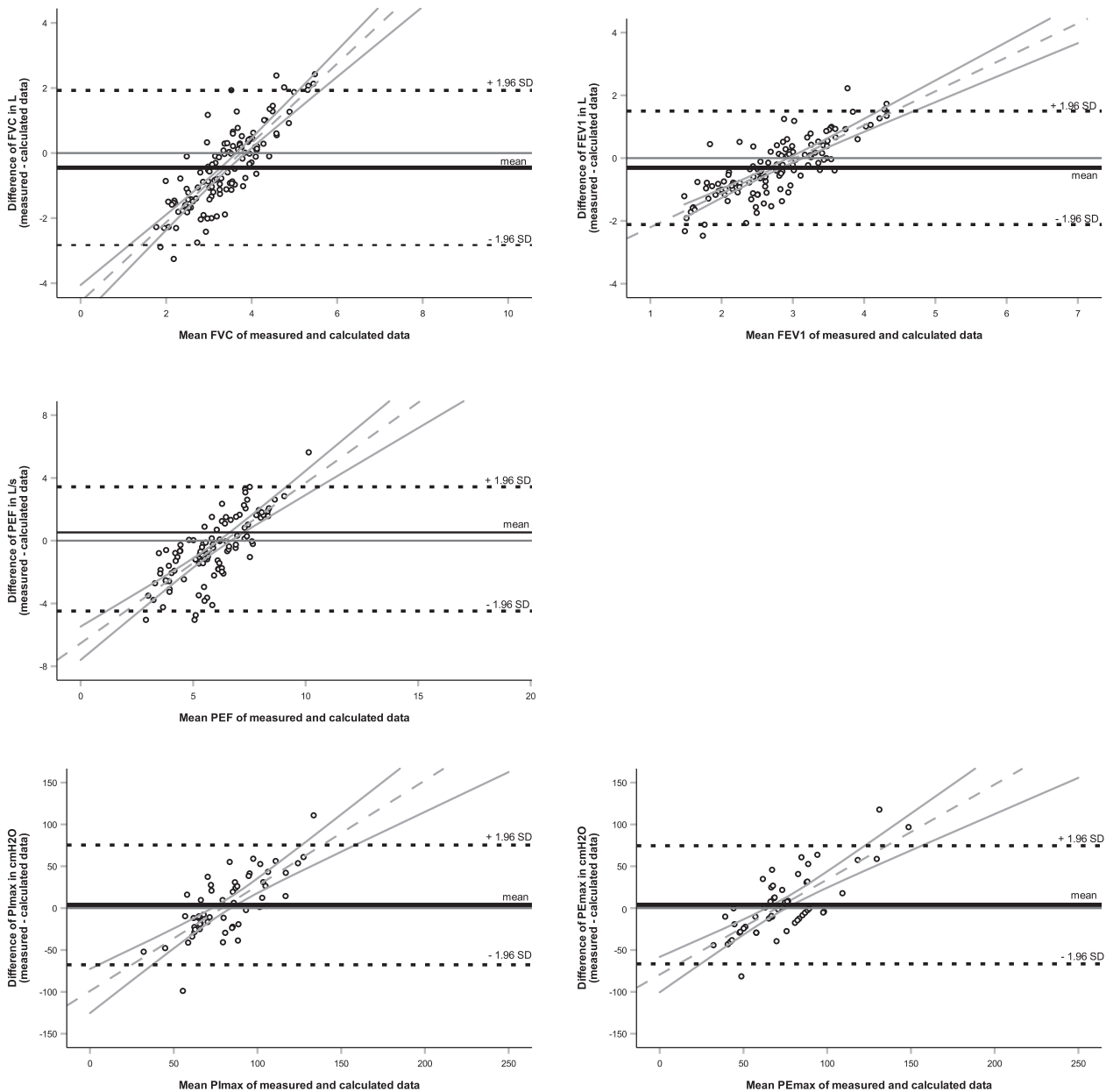
Smoking history includes the variables pack-years, ever smoker, former smoker, current smoker

plots we decided that other fittings than the linear fitting did not improve the models. All significant predictors can be seen in Table 3.

The predictors included in the new models explained 69–78% of the variance ( $R^2$ ) for lung volumes and 69% for both respiratory muscle strength parameters (Table 3).

### Phase 3 - Accuracy of the lesion-specific reference models

The reliability of the new models were tested with the 'validation sample' (20% of the total sample) (Fig. 1). The  $t$ -tests to test for proportional bias (measured versus



**Fig. 3** (phase 3). Bland–Altman plots of differences between the measured respiratory function values (FVC, FEV<sub>1</sub>, PEF, PI<sub>max</sub>, and PE<sub>max</sub>) of each participant and the calculated respiratory function values using the new developed models with the representation of the limits of agreement (dashed, horizontal lines). The bold line represents the proportional bias and dashed lines show the limits of agreement. Dotted, diagonal line represents the regression line and confidence interval limits are presented as continuous, diagonal lines. FVC

regression line (95% CI):  $y = -4.57 (-5.08 \text{ to } -4.06) + 1.22 (1.07 \text{ to } 1.37) *x$ . FEV<sub>1</sub> regression line (95% CI):  $y = -3.29 (-3.70 \text{ to } -2.87) + 1.08 (0.94 \text{ to } 1.23) *x$ . PEF regression line (95% CI):  $y = -6.54 (-7.60 \text{ to } -5.47) + 1.02 (0.85 \text{ to } 1.20) *x$ . PI<sub>max</sub> regression line (95% CI):  $y = -99.05 (-125.42 \text{ to } -72.67) + 1.26 (0.95 \text{ to } 1.57) *x$ . PE<sub>max</sub> regression line (95% CI):  $y = -79.40 (-100.56 \text{ to } -58.25) + 1.14 (0.86 \text{ to } 1.41) *x$

**Table 4** Phase 3: Accuracy for prediction of lung function and respiratory muscle strength for the validation sample with the developed statistical models

	FVC (L)	FEV <sub>1</sub> (L)	PEF (L/s)	PI <sub>max</sub> (cmH <sub>2</sub> O)	PE <sub>max</sub> (cmH <sub>2</sub> O)
<i>n</i>	122	122	104	56	56
Measured mean (SD)	3.2 (1.4)	2.6 (1.1)	5.6 (2.3)	83 (37)	75 (39)
Predicted mean (SD)	3.6 (0.47)	2.9 (0.42)	6.0 (1.1)	83 (13)	75 (14)
Average of measured and predicted values (SD)	3.4 (0.83)	2.8 (0.68)	5.9 (1.5)	81 (21)	73 (23)
ICC (95% CI)	0.43 (0.19 to 0.60)	0.51 (0.31 to 0.66)	0.53 (0.30 to 0.68)	0.28 (−0.23 to 0.58)	0.43 (0.02 to 0.66)
Bland–Altman					
Difference (mean (SD))	−0.45 (1.2)	−0.31 (0.92)	−0.53 (2.0)	3.9 (36)	4.0 (36)
Standard error of difference	0.11	0.08	0.20	4.9	4.8
95% CI for difference	−0.67 to −0.23	−0.47 to −0.14	−0.92 to −0.13	−5.9 to 13	−5.7 to 13
95% limits of agreement	−2.8 to 1.9	−2.1 to 1.5	−4.5 to 3.4	−67 to 75	−66 to 74
Standard error of measurement (95% CI)	0.35 (−0.34 to 1.1)	0.29 (−0.28 to 0.87)	0.73 (−0.70 to 2.2)	11 (−11 to 34)	11 (−10 to 32)

CI confidence interval, ICC intraclass correlation coefficient, SD standard deviation FVC forced vital capacity (L), FEV<sub>1</sub> forced expiratory volume in 1 s (L), PEF peak expiratory flow (L/s), PI<sub>max</sub> peak inspiratory pressure (cmH<sub>2</sub>O). PE<sub>max</sub> peak expiratory pressure (cmH<sub>2</sub>O)

predicted values) are significant for FVC ( $p < 0.000$ ), FEV<sub>1</sub> ( $p < 0.000$ ) and PEF ( $p < 0.000$ ). That means that the measured values are significantly lower or higher than the predicted values, but not significant for PI<sub>max</sub> ( $p = 0.43$ ) or PE<sub>max</sub> ( $p = 0.41$ ). Thus, there is a systematic difference with an over- or underestimation in the new models for all three lung function parameters but not for the respiratory muscle strength parameters. Figure 3 shows the predicted and measured respiratory function values in Bland–Altman plots. The ICCs between the measured and predicted lung function values for the new models ranged from 0.28 (PI<sub>max</sub>) to 0.55 (PEF), i.e., between poor and moderate reliability (Table 4) [26]. The residual- and scatterplots can be found in the online supplement [25, 30, 31].

## Discussion

The prediction models published in 2012 showed to be not accurate enough for persons with long-term SCI and new models needed to be developed. The reason is that the 2012 equations underestimated the actual values in the lower range and overestimated them in the upper range for all five respiratory function parameters. For clinical practice, comparison of measured values with population specific reference values is important in order to prescribe preventive treatment. One possible treatment could be the increase of inspiratory muscle strength by respiratory muscle training as shown in a previous publication of our research group where we also used reference values [32]. Using lesion-specific relative values of respiratory function is much more sensitive in the SCI-population than just absolute values or reference values from able-bodied persons [32]. For lung function we did not find better models for persons with long-term SCI but the models for respiratory muscle strength improved.

## Phase 1 - Accuracy of the statistical models published in 2012

There was a need to test the accuracy of the first published models for persons with more than two years TPI. The ICCs for the lung volumes were good and also the Bland–Altman plots only showed small differences between the measured and the predicted values (Table 2). According to the ATS/ERS documents the acceptable difference between the measured and the predicted value should be below 0.150 L for FVC, should not exceed 20% for FEV<sub>1</sub> and should be below 0.67 L/s for PEF [33]. In our results these differences between measured and predicted values are within these acceptable differences, however, the 95% limits of agreement for FVC are between 2.12 L and 1.96 L which represents a wide range (Table 2). For PI<sub>max</sub> and PE<sub>max</sub> the ICCs were poor, the differences and limits of agreement between the measured and the predicted values relatively wide (Table 2). Normal ranges for respiratory muscle strength are wide and the inter-individual differences between measurements in muscle strength is considerably greater than for lung function [34]. Due to these findings we judged the old models not accurate, especially for respiratory muscle strength for a long-term SCI population.

## Phase 2 - Development of lesion-specific reference models

New models to predict respiratory function values in long-term SCI were developed with lesion level, gender and weight as the main candidate predictors. There are some parallels between the ‘old’ and the newly developed models but also some fundamental differences. Lesion level, as a SCI-specific parameter is important for all models and similar to able-bodied persons, also gender had an influence



on all models (Table 3) [15]. Typically women have smaller vital capacities and maximal expiratory flow rates, reduced airway-diameters and smaller diffusion surfaces than age- and height-matched men [11]. Height is similar between both models [15] (Table 3), only in PEF there was no additional increase with increasing height (Table 3). In able-bodied persons a 1% increase in height corresponds to a 2.5% increase in FVC and FEV<sub>1</sub> [13].

In the ‘old’ models, increasing age had a negative effect on all five respiratory function parameters [15]. In the newly developed models age had only a negative effect on FEV<sub>1</sub> with a decrease of 15 ml per year (Table 3). This is about half the age-related decline shown in the old models and in able-bodied persons where FEV<sub>1</sub> declines by up to 30 ml per year [12, 35]. The aging lung is likely to have experienced exposures to environmental toxins and reductions in physiological capacity [12], and the chest-wall compliance is reduced due to stiffening of the rib cage [1]. In the “old” models TPI was positively associated with PE<sub>max</sub> [15]. The newly developed models with the long-term SCI population showed that the longer the TPI the lower the FVC (Table 3). For FVC the TPI was even a stronger predictor than age. Large interpersonal differences in change of FVC can happen in the first five years after rehabilitation [10]. If a clinician is interested to see what the influence is of 10 years aging or TPI instead of 1 year aging or TPI one can multiply the beta of age or TPI by 10 (Table 3). Smoking history conferred no predictive power even though it is well known that starting smoking is related to a rapid decline of lung function while quitting smoking has a beneficial effect on lung function [36]. Our findings are supported by literature where smoking history did also not show differences over time for FVC and FEV<sub>1</sub> in males with traumatic tetraplegia and in persons with either tetra- or paraplegia with AIS A-D [10, 37]. In our study all those participants with chronic lung diseases have been excluded and thus, an under-representation of the average population-based respiratory function status may occur. When estimating smoking history, recall bias may be an issue due to the retrospective design among ex-smokers [38]. Recall bias means when persons remember past events, they do not usually have a complete or accurate picture of what happened. However, in another investigation self-reports of regular, former or never smoking are found to be usually accurate and the validity of self-reporting smoking seems to be similar among persons from different ages and socioeconomic groups [39].

### Phase 3 - Accuracy of the lesion-specific reference models

An important aspect of prediction is to consider if a regression model can be used reliably in persons with comparable characteristics. The ICCs for the lung volume

parameters were poor and the Bland-Altman plots for FVC and FEV<sub>1</sub> showed relatively wide differences and limits of agreement between the measured and the predicted values (Table 4). According to the ATS/ERS documents the acceptable difference between the measured and the predicted value should be below 0.15 L for FVC, should not exceed 20% for FEV<sub>1</sub> and should be below 0.67 L/s for PEF [33]. In our study the difference between the measured and the predicted values for FVC (−0.45 L) is higher than the acceptable difference of 0.15 L and for FEV<sub>1</sub> (−0.31 L) and PEF (−0.53 L/s) they are within the acceptable range but the differences are higher than in the validation of the old models (Table 4). A positive trend appeared evident in all five Bland-Altman plots, as illustrated with the regression line of the difference and confidence interval limits [25] (Fig. 3). We conclude that for lung volume models in long-term SCI, other possible candidate predictors need to be evaluated in future research. The list of further predictors is long, ranging from clinical (e.g. co-morbidities) and laboratory (e.g. Chest X-ray) predictors to social/ psychological (e.g. ethnicity, motivation) predictors [40]. With the currently available candidate predictors, lung function cannot be well-predicted. As such, the “old” models appeared to be more accurate for predicting FVC and FEV<sub>1</sub>.

The use of different prediction equations can lead to different interpretation with results differing on models and geographical site of assesment [41]. The Global Lung function Initiative (GLI) provide standardized spirometry reference equations for able-bodied aged from 3 to 95 years to improve comparability and accuracy [42]. After comparing the GLI and another two commonly used spirometry prediction equations in a group of healthy Kenyan volunteers one can see that no equation consistently provides accurate estimates of normal lung function [43]. Even if the exposure to environmental pollutants in the Kenyan population is not comparable with the reference population the study brings into question the validity of these major published spirometry prediction equations [43]. The best accuracy is given with detailed information about age and height. A few months age difference can affect the predicted values by up to 8.5% [44]. Also an 1 cm error in height can lead to an error in the predicted value of 6% [42].

### Clinical relevance

Prediction models for respiratory function are useful to individually assess the respiratory function in persons with different levels of injury. By yearly measuring the respiratory function and comparing them to the predicted values, we can potentially identify lung dysfunction and deterioration with aging, posture, obesity or ascension of neurological level.

## Strengths and limitations

The strengths of this study are the appropriate set of respiratory function values and the dataset that best represents the target patient population (Table 1) [45]. Data of several laboratories are combined in which the techniques are performed in the same way as in daily routine so that sources of variability are minimized [45].

The development of these new models involved some compromise as data of respiratory muscle strength and PEF were only available in two of three countries. Ideally, reference values are calculated with models derived from measurements observed in a representative sample of a clearly defined and, as much as possible, corresponding population [46, 47]. With our study size of 276 and 393 participants for respiratory muscle strength and PEF (Table 3), respectively, these criteria can still be fulfilled.

The lung function measurements in the Netherlands and in Australia were especially performed for research purposes which bears the risk of a selection bias, but had the advantage of standardized measurement protocols.

The period of inclusion of participants from 1996 to 2018 is long and changes in treatment regime happened, e.g. new respiratory muscle training devices are used in clinical practice and more is known about the effect of in- and expiratory muscle strength training. However, since this is not a training- or longitudinal study (one measurement per participant only) this may only have a marginal influence on the results.

Differences in TPIs may have contributed to the heterogeneity of the results. In the Australian and Swiss centers the TPI started at two years in contrast to the Dutch centers where the TPI was defined with 10 years. Since TPI had been included as continuous parameter into the analysis, this fact should not be a limitation.

Our focus was on the most obvious and easy to assess candidate predictors for feasibility reasons in the daily clinical practice. The research team selected the candidate variables based on the available literature in the same population [15, 20] and with able-bodied persons [47]. In future, the Delphi method as a structured communication technique could help to update and reach consensus on candidate predictors [40]. Despite the fact that our models do still not contain all potential candidate predictors of relevance [40], our large sample size in general reflects a wide range of various levels of, e.g. activity and sports among the participants.

For the graphical illustration of the validation, Bland–Altman plots (Figs. 2 and 3) and residual plots (online supplement) were used. There is a discussion about the difference between these two statistical methods with the argument that in Bland–Altman plots a systematic

proportional bias can occur [48]. When looking at Figs. 2 and 3, in fact this type of bias is obvious. At lower respiratory function values the predicted values are higher than the measured values while at higher respiratory function values the predicted values are lower than the measured values.

## Conclusion

The respiratory function prediction models published in 2012 showed to be not accurate enough for persons with long-term SCI. Thus, new statistical models have been developed to predict the respiratory function in persons with injuries more than two years ago. In summary, we did not find better models for lung function in long-term SCI but those for respiratory muscle strength showed better accuracy.

## Data archiving

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

**Acknowledgements** The contribution of the 10 participating rehabilitation centers, in particular that of the research assistants and/or medical practice assistants collecting all the data, is highly appreciated.

**Funding** The content of this publication was developed under grant from Wings for Life, grant number WFL-CH-017/14.

**Author contributions** AMR was responsible for designing the protocol, analysing data, interpreting results writing of this manuscript. SdG conducted the regression-analyses and supported with interpreting results and provided feedback on the manuscript. DJB supported with interpreting results and provided feedback on the manuscript. MWMP supported with interpreting results and provided feedback on the manuscript. JA helped with the data collection for the Dutch part of the study and provided feedback on the manuscript. MH provided feedback on the manuscript. GM was responsible for designing the protocol, interpreting results and provided feedback on the manuscript.

## Compliance with ethical standards

**Statement of ethics** We certify that all applicable institutional and governmental regulations concerning the ethical use of data of human volunteers were followed during the course of this research. All patients in the centers had been informed in written form about the potential retrospective use of their data for research purposes. All patients who had rejected the use of their medical or personal data were excluded.

**Conflict of interest** The authors declare that they have no conflict of interest.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Brown R, DiMarco AF, Hoit JD, Garshick E. Respiratory dysfunction and management in spinal cord injury. *Respir Care*. 2006;51:853–68.
- Galeiras Vazquez R, Rascado Sedes P, Mourelo Farina M, Montoto Marques A, Ferreira Velasco ME. Respiratory management in the patient with spinal cord injury. *BioMed Res Int*. 2013;2013:168757.
- Chen CF, Lien IN, Wu MC. Respiratory function in patients with spinal cord injuries: effects of posture. *Paraplegia*. 1990;28:81–6.
- Berlowitz DJ, Tamplin J. Respiratory muscle training for cervical spinal cord injury. *Cochrane Database Syst Rev*. 2013;7:CD008507.
- Van Der Molen T, Willemse BW, Schokker S, Ten Hacken NH, Postma DS, Juniper EF. Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health Qual life outcomes*. 2003;1:13.
- Mueller G, de Groot S, van der Woude L, Hopman MT. Time-courses of lung function and respiratory muscle pressure generating capacity after spinal cord injury: a prospective cohort study. *J Rehabil Med*. 2008;40:269–76.
- Haisma JA, Bussmann JB, Stam HJ, Sluis TA, Bergen MP, Dallmeijer AJ, et al. Changes in physical capacity during and after inpatient rehabilitation in subjects with a spinal cord injury. *Arch Phys Med Rehabil*. 2006;87:741–8.
- Hoh DJ, Mercier LM, Hussey SP, Lane MA. Respiration following spinal cord injury: evidence for human neuroplasticity. *Respir Physiol Neurobiol*. 2013;189:450–64.
- Grant RA, Quon JL, Abbed KM. Management of acute traumatic spinal cord injury. *Curr Treat options Neurol*. 2015;17:334.
- Postma K, Haisma JA, de Groot S, Hopman MT, Bergen MP, Stam HJ, et al. Changes in pulmonary function during the early years after inpatient rehabilitation in persons with spinal cord injury: a prospective cohort study. *Arch Phys Med Rehabil*. 2013;94:1540–6.
- Harms CA. Does gender affect pulmonary function and exercise capacity? *Resp Physiol Neurobiol*. 2006;151:124–31.
- Vaz Fragoso CA, Gill TM. Respiratory impairment and the aging lung: a novel paradigm for assessing pulmonary function. *J Gerontol A Biol Sci Med Sci*. 2012;67:264–75.
- Stanojevic S, Wade A, Stocks J, Hankinson J, Coates AL, Pan H, et al. Reference ranges for spirometry across all ages: a new approach. *Am J Respir Crit Care Med*. 2008;177:253–60.
- Berlowitz DJ, Wadsworth B, Ross J. Respiratory problems and management in people with spinal cord injury. *Breathe*. 2016;12:328–40.
- Mueller G, de Groot S, van der Woude LH, Perret C, Michel F, Hopman MT. Prediction models and development of an easy to use open-access tool for measuring lung function of individuals with motor complete spinal cord injury. *J Rehabil Med*. 2012;44:642–7.
- Wolfe R, Carlin J. Statistical models for respiratory disease diagnosis and prognosis. *Respirology*. 2015;20:541–7.
- van der Woude LH, de Groot S, Postema K, Bussmann JB, Janssen TW, Post MW. Active Lifestyle Rehabilitation interventions in aging spinal cord injury (ALLRISC): a multicentre research program. *Disabil Rehabil*. 2013;35:1097–103.
- Heus P, Damen J, Pajouheshnia R, Scholten R, Reitsma JB, Collins GS, et al. Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. *BMC Med*. 2018;16:120.
- Spong J, Graco M, Brown DJ, Schembri R, Berlowitz DJ. Subjective sleep disturbances and quality of life in chronic tetraplegia. *Spinal Cord*. 2015;53:636–40.
- van Silfhout L, Peters AE, Berlowitz DJ, Schembri R, Thijssen D, Graco M. Long-term change in respiratory function following spinal cord injury. *Spinal Cord*. 2016;54:714–9.
- Standardization of Spirometry. 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med*. 1995;152:1107–36.
- Karvonen J, Saarelainen S, Nieminen MM. Measurement of respiratory muscle forces based on maximal inspiratory and expiratory pressures. *Respiration*. 1994;61:28–31.
- Twisk JW. *Applied Longitudinal Data Analysis for Epidemiology, A Practical Guide*. Amsterdam: Cambridge University Press; 2003.
- Rasbash J, Browne W, Goldstein H, Yang M, Plewis I, Healy M, et al. *A User's Guide to MLwin*. Centre for Multilevel Modelling, Institute of Education, University of London. London, United Kingdom 2000.
- Giavarina D. Understanding Bland Altman analysis. *Biochem Med*. 2015;25:141–51.
- Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15:155–63.
- Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil*. 1998;12:187–99.
- Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19:453–73.
- van de Port IG, Kwakkel G, Schepers VP, Lindeman E. Predicting mobility outcome one year after stroke: a prospective cohort study. *J Rehabil Med*. 2006;38:218–23.
- van Stralen KJ, Jager KJ, Zoccali C, Dekker FW. Agreement between methods. *Kidney Int*. 2008;74:1116–20.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–10.
- Raab AM, Krebs J, Perret C, Michel F, Hopman MT, Mueller G. Maximum inspiratory pressure is a discriminator of pneumonia in individuals with spinal-cord injury. *Respir Care*. 2016;61:1636–43.
- Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26:319–38.
- ATS/ERS Statement on respiratory muscle testing. *Am J Respir Crit Care Med*. 2002;166:518–624.
- Zimmer MB, Nantwi K, Goshgarian HG. Effect of spinal cord injury on the respiratory system: basic research and current clinical treatment options. *J Spinal Cord Med*. 2007;30:319–30.
- Sherrill DL, Holberg CJ, Enright PL, Lebowitz MD, Burrows B. Longitudinal analysis of the effects of smoking onset and cessation on pulmonary function. *Am J Respir Crit Care Med*. 1994;149(3 Pt 1):591–7.
- Tow AM, Graves DE, Carter RE. Vital capacity in tetraplegics twenty years and beyond. *Spinal Cord*. 2001;39:139–44.
- Vold ML, Aasebo U, Melbye H. Low FEV1, smoking history, and obesity are factors associated with oxygen saturation decrease in an adult population cohort. *Int J Chronic Obstr Pulm Dis*. 2014;9:1225–33.
- Vartiainen E, Seppala T, Lillsunde P, Puska P. Validation of self reported smoking by serum cotinine measurement in a community-based study. *J Epidemiol Community Health*. 2002;56:167–70.
- Wiens MO, Kissoon N, Kumbakumba E, Singer J, Moschovis PP, Ansermino JM, et al. Selecting candidate predictor variables for the modelling of post-discharge mortality from sepsis: a protocol development project. *Afr health Sci*. 2016;16:162–9.

41. GOSH NHS Foundation Trust. Lung Function Update Exciting News! New recommendations for the interpretation of spirometric lung function measurements. London; 2013. p. 1.
42. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, 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.
43. Paul DW, Lagat DK, MacIntyre N, Egger JR, Murdoch DM, Que LG, et al. Validation of spirometry prediction equations in western Kenya. *Int J Tube Lung Dis*. 2018;22:112–8.
44. Quanjer PH, Hall GL, Stanojevic S, Cole TJ, Stocks J. Age- and height-based prediction bias in spirometry reference equations. *Eur Respir J*. 2012;40:190–7.
45. Weinberg J, Sinderby C, Sullivan L, Grassino A, Lindstrom L. Evaluation of diaphragm electromyogram contamination during progressive inspiratory maneuvers in humans. *Electro Clin Neurophysiol*. 1997;37:143–53.
46. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. *Eur Respir J*. 2005;26:948–68.
47. Rochat MK, Laubender RP, Kuster D, Braendli O, Moeller A, Mansmann U, et al. Spirometry reference equations for central European populations from school age to old age. *PLoS ONE*. 2013;8:e52619.
48. Hopkins WG. Bias in Bland-Altman but not regression validity analyses. *Sportscience*. 2004;8:42–6.