

A new maximal bicycle test using a prediction algorithm developed from four large COPD studies

Göran Eriksson^a, Finn Radner^a, Stefan Peterson^b, Georgia Papapostolou^a, Linnea Jarenbäck^a, Saga Jönsson^a, Jaro Ankerst^a, Alf Tunsäter^a, Ellen Tufvesson^a and Leif Bjermer^a

^aDepartment of Clinical Sciences, Respiratory Medicine and Allergology, Lund University, Skane University Hospital, Lund, Sweden;

^bRegional Cancer Centre South, Region Skane, Lund, Sweden

ABSTRACT

Background: Maximum exercise workload (W_{MAX}) is today assessed as the first part of Cardiopulmonary Exercise testing. The W_{MAX} test exposes patients with COPD, often having cardiovascular comorbidity, to risks. Our research project was initiated with the final aim to eliminate the W_{MAX} test and replace this test with a predicted value of W_{MAX} , based on a prediction algorithm of W_{MAX} derived from multicentre studies.

Methods: Baseline data (W_{MAX} , demography, lung function parameters) from 850 COPD patients from four multicentre studies were collected and standardized. A prediction algorithm was prepared using Random Forest modelling. Predicted values of W_{MAX} were used in a new W_{MAX} test, which used a linear increase in order to reach the predicted W_{MAX} within 8 min. The new W_{MAX} test was compared with the standard stepwise W_{MAX} test in a pilot study including 15 patients with mild/moderate COPD.

Results: The best prediction algorithm of W_{MAX} included age, sex, height, weight, and six lung function parameters. FEV_1 and DLCO were the most important predictors. The new W_{MAX} test had a better correlation ($R^2 = 0.84$) between predicted and measured W_{MAX} than the standard W_{MAX} test ($R^2 = 0.66$), with slopes of 0.50 and 0.46, respectively. The results from the new W_{MAX} test and the standard W_{MAX} test correlated well.

Conclusion: A prediction algorithm based on data from four large multicentre studies was used in a new W_{MAX} test. The prediction algorithm provided reliable values of predicted W_{MAX} . In comparison with the standard W_{MAX} test, the new W_{MAX} test provided similar overall results.

ARTICLE HISTORY

Received 12 August 2019

Accepted 8 November 2019

KEYWORDS

COPD; cardiopulmonary exercise testing; W_{MAX} ; Random Forest; prediction



Introduction

Spirometry is mandatory for establishing a diagnosis of chronic obstructive pulmonary disease (COPD) but is not sufficient for a general clinical assessment and evaluation of potential treatment effects. Functional performance assessed at walk or bicycle exercise tests together with measurements of symptoms and health-related Quality of Life (QoL) provide valuable additional information about present disease status, as well as prediction of future risk of exacerbations and disease prognosis [1–7]. The standard bicycle exercise protocol for cardiopulmonary exercise testing (CPET) measures the endurance time of cycling during a standard endurance test, at a constant rate of 75% of the maximum exercise workload (W_{MAX}) [7–11]. W_{MAX} is obtained in a preceding incremental W_{MAX} test, in which the patient is subjected to a stepwise increase in workload until the point of physical

exhaustion or another symptom-limited reason for stopping is reached.

CPET is technically demanding and requires expensive equipment [11]. A considerable number of participants are needed to detect potential therapeutic effect differences due to high inter- and intra-test variability of the standard endurance test [8,10,12,13]. In addition, CPET is associated with a certain increase in cardiac risks, particularly at the initial W_{MAX} test [7,8].

A way to address the above downsides would be to correctly predict W_{MAX} , which optimally could avoid the W_{MAX} test by using 75% of the predicted value for the endurance test. Several papers have identified predictors for W_{MAX} or maximal oxygen consumption (VO_{2MAX}) in COPD patients using the standard W_{MAX} or VO_{2MAX} test. Forced expiratory volume in 1 sec (FEV_1) has been identified as an important predictor [14–19]. Other predictors identified are static lung volumes (e.g. inspiratory capacity; IC) [14,15,18],

CONTACT Göran Eriksson,  goran.eriksson@med.lu.se  Department of Clinical Sciences, Respiratory Medicine and Allergology, Lund University, Skane University Hospital, Lund 22185, Sweden

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

small airway dysfunction, e.g. reduced mid-expiratory flow (FEF_{25-75}) [17,20,21], impaired gas exchange measured by e.g. diffusing capacity for carbon monoxide (DLCO) [14,15,18,20,22], maximal voluntary ventilation (MVV) [18,19,21], Medical Research Council (MRC) scale [16,17], sub-maximal exercise [23,24] and demographic data such as sex, age, body-mass index (BMI) or fat-free mass [14–17,20,21,25]. However, another article concluded that predicted VO_{2MAX} from baseline characteristics cannot be used for patients with stable COPD [26].

Previous papers claiming predictability of W_{MAX} (VO_{2MAX}) had different reasons for trying to replace the incremental test by a predictive (W_{MAX} or VO_{2MAX}) value, e.g. to determine an individual patient's future risks of, e.g., exacerbations, to avoid complication in a cardiopulmonary risk population, to reduce costs caused by expensive equipment and to achieve a reasonably accurate endurance value for a rehabilitation program. Only one study [27] has specifically addressed the possibility to predict the individual incremental step increase to keep the test to 8–10 min (see also [16,28]). Most previous prediction papers have used stepwise and not linear increase. A number of older studies concluded that a linear ramp test was no better than a stepwise incremental test (for an overview, see [8]), but these papers could be challenged.

We, therefore, embarked on a research project aiming to improve CPET by

- (1) elimination of the incremental W_{MAX} test in order to avoid the cardiovascular risks associated with the test and, in addition, save clinical study resources. Instead, a carefully predicted value of W_{MAX} is to be used to identify the starting point for the 75% endurance test.
- (2) decreasing the usually high intra- and inter-variability of the 75% endurance test and thereby offer comparative COPD study designs needing less patients.

As an initial step, we designed a new model for the incremental W_{MAX} test involving three combined ideas to be tested: 1) a prediction algorithm, developed from four large pharmaceutical industry-sponsored clinical studies, in order to identify a predicted value of W_{MAX} to be reached within a certain exercise time (3 + 8 min; see Figure 1); 2) a linear instead of stepwise increase of workload to introduce continuous W_{MAX} values instead of categorical values of 10 W per step. 3) measurements of future predictors (e.g. low-intensity O_2/CO_2 kinetic evaluation, impulse oscillometry and activity scores).

The proposed new W_{MAX} test was compared with the standard incremental W_{MAX} test [8] in a single-centre pilot study. This report includes two major parts – the development of the prediction algorithm and the results from the pilot study.

The primary objective of the pilot study was to verify the prediction algorithm using the new W_{MAX} test. The secondary objective was to show that the new W_{MAX} test provided better results compared to the standard W_{MAX} test, for use in future studies within this research project. This new W_{MAX} test will be utilized until the prediction model provides values of W_{MAX} that are reliable enough to be used as the starting point for the 75% endurance test.

Materials and methods

Studies used for developing a prediction algorithm of W_{MAX}

Patients

Baseline data (demography, lung function parameters and W_{MAX} from incremental exercise tests) were provided from two pharmaceutical companies (AstraZeneca and Boehringer Ingelheim) via standard data-sharing agreements. Patients with COPD participated in four multicentre, randomised clinical studies examining the effects of tiotropium (Study A,B,D) [12,13,29] or budesonide/

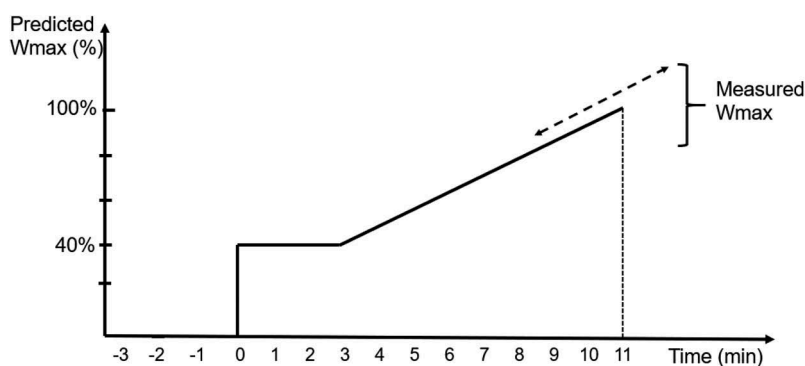


Figure 1. Schematic design of the new maximum test.

formoterol (Study C) [30] on exercise performance (Table 1). The inclusion criteria were similar in Study A-C, while Study D included patients with milder COPD. Patients were clinically stable patients with COPD, aged >40 years, with a smoking history of >10 pack-years, a pre-bronchodilator FEV₁ ≤ 65% (Study A and B) of predicted normal, ≤50% (Study C) or 50–80% (Study D). Studies were approved by medical ethical committees and all patients gave written informed consent before undertaking any study procedures.

Study design

In Study, A-D eligibility criteria were assessed during an initial screening visit. At this or a following pre-randomization visit, patients performed pulmonary function tests followed by a standard incremental W_{MAX} test. Study A-C were conducted using bicycle exercise testing while Study D used treadmill exercise testing.

The data were combined into a single dataset following transformations in order to account for differences in units, data formats and naming of the variables. Note that all variables were not available from all studies.

Outcome measures

For the post hoc analysis of Study A-D, the outcome was a prediction algorithm for the best possible prediction of W_{MAX}.

Statistical analyses

Prediction algorithms were formulated using Random Forest modelling [31] for each of Study A-D separately. The Random Forest model was used to automatically handle potential non-linear associations among the variables (predictors). A list of the ranking order among the variables was obtained, sorted in descending order with the best predicting variable first. Random Forest modelling was performed using R (v 3.5.1 for Mac) software. Univariate regression was performed in Microsoft Excel for Office 365. The Random Forest algorithms are available for external use upon request.

Pilot study to compare new and standard W_{MAX} test

Patients

In the pilot study, patients with COPD who had no exacerbations of COPD within the last 6 weeks, a post-bronchodilator FEV₁ of ≥40 to ≤80% of predicted normal and no cardiovascular co-morbidity preventing exercise testing were included. The Regional Ethical Review Board approved the study. Written informed consent was obtained from all patients prior to any study procedures. Patients were out-patients who had previously visited our clinic and registered for volunteering in other research studies.

Table 1. Patient baseline characteristics per study for study A-D and the pilot study.

	Study A	Study B	Study C	Study D	
Study code	205.223	205.334	D5892C00015	205.440	
Sponsor	BI	BI	AZ	BI	
Reference	[13]	[12]	[30]	[29]	
ClinicalTrials.gov identifier	NCT00274508	NCT00530842	NCT00489853	NCT01072396	Pilot study
Number of patients	261	344	119	126	15
Age, years	62 ± 7	61 ± 7	64 ± 8	61 ± 9	71 ± 5
Male, %	72	72	75	52	73
Height, cm	171 ± 9	170 ± 8	172 ± 8	169 ± 9	172 ± 9
Weight, kg	77 ± 17	77 ± 17	76 ± 16	79 ± 14	77 ± 12
BMI, kg/m ²	26 ± 5	26 ± 5	26 ± 4	28 ± 4	26 ± 3
Caucasian, % (race)	100	99	100	91	100
Afro-American, % (race)	0	0,6	0	9	0
Current smoker, %	40	44	40	47	7
Pack years	53 ± 28	46 ± 24	36 ± 15	47 ± 22	34 ± 17
COPD duration, years	9 ± 7	8 ± 6	10 ± 8	5 ± 5	*
LABA user, %	33	52	86	18	53
ICS user, %	54	44	39	18	60
LAMA user, %	0	29	44	22	93
FEV ₁ , L **	1.21 ± 0.42	1.36 ± 0.40	1.12 ± 0.32	1.95 ± 0.67	1,80 ± 0,44
FEV ₁ , % predicted ***	39.0 ± 15.6	43.8 ± 14.2	36.3 ± 18.8	66.3 ± 16.7	60,3 ± 8,6
FVC, L **	2.78 ± 0.83	2.85 ± 0.76	2.25 ± 0.68	3.43 ± 1.05	3,62 ± 1,07
FEV ₁ /FVC % **	43.9 ± 10.8	48.8 ± 11.6	51.4 ± 12.6	57.1 ± 8.5	0,51 ± 0,09

Values are mean ± standard deviation unless otherwise specified.

AZ, AstraZeneca; BMI, Body-mass index; BI, Boehringer Ingelheim; FEV₁, forced expiratory volume in 1 sec; FVC, Forced vital capacity; ICS, Inhaled corticosteroid; LABA, Long-acting β₂ adrenoceptor agonist; LAMA, Long-acting muscarinic receptor antagonist.

* 80% had COPD duration >5 years; 20% had COPD duration 2–5 years

** Pre-bronchodilator (Study A-D); post-bronchodilator (Pilot study).

*** % of predicted normal, calculated according to NHANES III.

Study design

The study was non-blinded and single-centre. At visit 1, demography data, modified medical research council (MMRC) dyspnea scale scores and clinical COPD questionnaire (CCQ) scores were collected. The COPD patients performed lung function tests (spirometry, body plethysmography, CO-diffusion test and impulse oscillometry) after inhalation of 400 µg salbutamol. The predicted value of W_{MAX} for each patient was calculated on a pre-programmed computer holding the Random Forest algorithm.

Patients then performed the standard or the new W_{MAX} test in a random cross-over fashion at the following visits. During the standard test, after a few minutes of sitting on the bicycle in order to stabilize the oxygen kinetics measurement equipment, the patient had an approximately one-minute warm-up period of loadless pedalling, followed by an incremental increase in workload with 10 W per minute until the patient reached his/her W_{MAX} at the point of exhaustion. During the new W_{MAX} test, the patients started cycling at a load of 40% of the predicted W_{MAX} for 3 min, followed by a linear increase in load, calculated to reach the predicted W_{MAX} after an additional 8 min (Figure 1). In both W_{MAX} tests, the same safety procedure were applied, as recommended for the standard W_{MAX} test [7,8]. Borg scale results were collected at every 2 min during both tests [32].

Outcome measures

The primary outcome measure was the ability of the prediction algorithm to successfully predict W_{MAX} as shown by the coefficient of determination (R^2) between

predicted and measured W_{MAX} at the new W_{MAX} test. Secondary outcome variables were descriptive comparisons between the new and the standard W_{MAX} test for averages of measured W_{MAX} and work performed, reasons for stopping exercise and Borg scale results.

Statistical analyses

We designed the study to include 15–20 patients in order to obtain enough data to test the prediction model and to compare the new test with the standard test, but no formal power calculation was performed. The comparisons between predicted and measured W_{MAX} , and the new and the standard W_{MAX} test were performed using univariate regression, descriptive statistics and Bland–Altman graphs [33]. Microsoft Excel for Office 365 was used for univariate regression analyses and Bland–Altman graphs. Demographics and patient data were expressed as mean \pm standard deviation.

Results

Studies used for developing a prediction algorithm of W_{MAX}

In total, 850 patients with COPD were included in the dataset for Study A–D, and their baseline data are summarized in Table 1. Using Random Forest modelling, all lung function and demographic variables for each of Study A–D were assessed for their ability to predict W_{MAX} . Results are presented in Table 2. The best percentage of variance explained was obtained for Study A, followed by Study B, C and D. The six best

Table 2. Rank order for prediction ability for different variables from Random Forest modelling in study A–D.

Study A		Study B		Study C		Study D	
Rank	Variable	Rank	Variable	Rank	Variable	Rank	Variable
1	FEV ₁	1	FEV ₁	1	FEV ₁	1	FEV ₁
2	DLCO	2	Weight	2	FVC	2	IC
3	FEF ₅₀	3	FVC	3	BODYB_VC	3	FVC
4	VA	4	SVC	4	IC	4	DLCO
5	FVC	5	Height	5	RV	5	SVC
6	FEF ₂₅₋₇₅	6	Age	6	FRC	6	Height
7	TLC	7	TGV	7	Weight	7	RV
8	FEF			8	SPIR_VC	8	Weight
9	SVC			9	TLC	9	TGV
10	RV			10	Height	10	FRC
11	Height			11	Age	11	RAW
12	Weight					11	Age
13	Age					11	TLC
14	TGV						
15	RAW						
16	SGAW						
Nd	FRC						
% variance explained	49		40		39		32

DLCO, diffusing capacity for carbon monoxide; FEF, forced expiratory flow; FEF₂₅₋₇₅, mid-expiratory flow; FEF₅₀, forced expiratory flow – 50%; FEV₁, forced expiratory volume in 1 sec; FRC, functional residual capacity; FVC, forced vital capacity; IC, inspiratory capacity; Nd, not determined; RAW, airway resistance; RV, residual volume; SGAW, specific airway conductance; SVC, slow vital capacity; TGV, thoracic gas volume; TLC, total lung capacity; VA, alveolar volume; VC, vital capacity.

lung function variables in Study A were FEV₁, DLCO, forced expiratory flow at 50% (FEF₅₀), forced vital capacity (FVC), alveolar volume (VA) and FEF₂₅₋₇₅. Random Forest modelling for Study B-D (Table 2) showed that FEV₁ was the highest ranked variable in all three studies. In patients with the most severe COPD disease (Study C), predictors after FEV₁ were FVC, vital capacity (VC), IC and residual volume (RV). In Study D (the study with subjects with milder COPD disease), FEV₁ was followed by IC, FVC and DLCO. In Study B, FEV₁, FVC and VC dominated as predictors. Neither Study B nor Study C included measurements of DLCO. Absolute values showed better results than % of predicted values. Univariate regression analyses of measured W_{MAX} versus individual values for each variable are presented in Table 3.

Table 3 shows that DLCO, together with FEV₁, was the best predictor in Study A. It was decided to base the prediction algorithm on the results from this study only, since the other study with measurements of DLCO (Study D) was a study using treadmill walking in patients with mild COPD. The lung function variables listed above from Study A were selected for the prediction algorithm, together with age, sex, height and weight. The results from the prediction algorithm are presented in the Figure 2, where predicted W_{MAX} has been plotted versus measured W_{MAX} for Study A. An R²-value of 0.95 and a slope of 0.77 was observed.

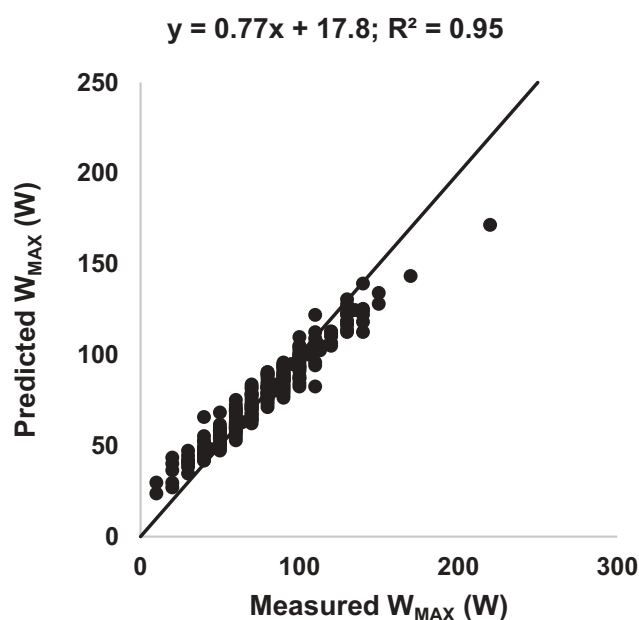


Figure 2. Predicted W_{MAX} versus measured W_{MAX} for Study A from the Random Forest prediction algorithm. Line is the line of identity.

Pilot study to compare new and standard W_{MAX} test

Patient population

A total of 19 patients were enrolled between July and October 2015. Fifteen of these were included in the study. Three were excluded due to FEV₁ ≥ 80% pre-

Table 3. Results from univariate regression analyses of individual values from each parameter versus measured W_{max} in study A-D and the pilot study.

Variable*	Study A	Study B	Study C	Study D	Pilot study, standard test**		Pilot study, new test***	
	R ²	R ²	R ²	R ²	R ²		R ²	
FEV ₁	0.41	0.32	0.30	0.28	0.44		0.67	
DLCO	0.37	nd	nd	0.21	0.78		0.75	
VA	0.26	nd	nd	nd	0.10		0.23	
FEF ₅₀	0.25	nd	nd	nd	0.43		0.49	
FVC	0.20	0.22	0.24	0.26	0.06		0.20	
SVC	0.16	0.20	0.27	0.26	0.04		0.17	
FEF ₂₅₋₇₅	0.15	nd	nd	nd	0.53		0.63	
Weight	0.14	0.16	0.23	0.11	0.25		0.32	
Height	0.13	0.21	0.12	0.21	0.29		0.29	
Age	0.07	0.07	0.02	0.05	0.16		0.15	
TLC	0.03	nd	0.0005	0.13	0.02		0.05	
RAW	0.03	nd	nd	0.05	nd		nd	
SGAW	0.03	nd	nd	nd	nd		nd	
TGV	0.004	0.002	nd	0.006	0.03		0.05	
RV	0.002	nd	0.10	0.008	0.004		0.02	
IC	nd	nd	0.23	0.31	0.29		0.36	
FRC	nd	nd	0.06	0.002	nd		nd	
VC	nd	nd	0.23	nd	0.04		0.17	

*Sorted by rank order no in Study A

** Value vs measured W_{MAX} in standard maximum test

*** Value vs measured W_{MAX} in new maximum test

DLCO, diffusing capacity for carbon monoxide; FEF₂₅₋₇₅, mid-expiratory flow; FEF₅₀, forced expiratory flow – 50%; FEV₁, forced expiratory volume in 1 sec; FRC, functional residual capacity; FVC, forced vital capacity; IC, inspiratory capacity; Nd, not determined; RAW, airway resistance; RV, residual volume; SGAW, specific airway conductance; SVC, slow vital capacity; TGV, thoracic gas volume; TLC, total lung capacity; VA, alveolar volume; VC, vital capacity.

dicted and one due to significant cardiovascular comorbidity. Fifteen patients with COPD were included in the pilot study. They had a mean age of 71 years and mean FEV₁ was 60% predicted normal. Further patient baseline characteristics are presented in Table 1.

Ability of the prediction algorithm to predict W_{MAX}

Figure 3 shows the correlation of predicted W_{MAX} versus measured W_{MAX} for the two tests (Figure 3(a,b)) plus the corresponding Bland–Altman plots (Figure 3(c,d)). Predicted values of W_{MAX} correlated well with measured W_{MAX} for the new test ($R^2 = 0.84$). For the standard test, a smaller determination coefficient was observed ($R^2 = 0.66$), even though the prediction algorithm was based on the standard W_{MAX} test performed in Study A. The slope for predicted W_{MAX} versus measured W_{MAX} was 0.50 for the new test and 0.46 for the standard test.

Comparison of the standard and new W_{MAX} test

The prediction algorithm from Study A was used for the new W_{MAX} test. Similar average results were obtained for W_{MAX} , work capacity, time of exercise,

Borg scale results and reasons for stopping exercise at the two tests (Table 4). For W_{MAX} reached (Figure 4, part A) and work performed in kW (Figure 4, part B), R^2 -values of 0.90 (W_{MAX}) and 0.95 (work capacity) were observed when the standard test and the new test were compared. A higher amount of work with a lesser percentage deviation was achieved during the new W_{MAX} test ($47.5 \text{ kW} \pm 36\% \text{ SD}$) than during the standard W_{MAX} test ($35.5 \text{ kW} \pm 45\% \text{ SD}$) (Table 4). The average duration was $9.3 \pm 2.6 \text{ min}$ (excluding the initial 3 min at 40% of predicted W_{MAX}) for the new test compared to $10.6 \pm 2.6 \text{ min}$ for the standard test.

Significance of DLCO

DLCO was the variable with the highest R^2 -value from univariate regression (0.78; Table 3) in the pilot study, which demonstrates its high ability for prediction of W_{MAX} in a small single-centre study. In Study A (a multi-centre study), univariate regression gives a mean R^2 -value of 0.37 for DLCO, but with a large inter-centre variability (range from 0.01 to 0.83 for the individual centres; data not shown).

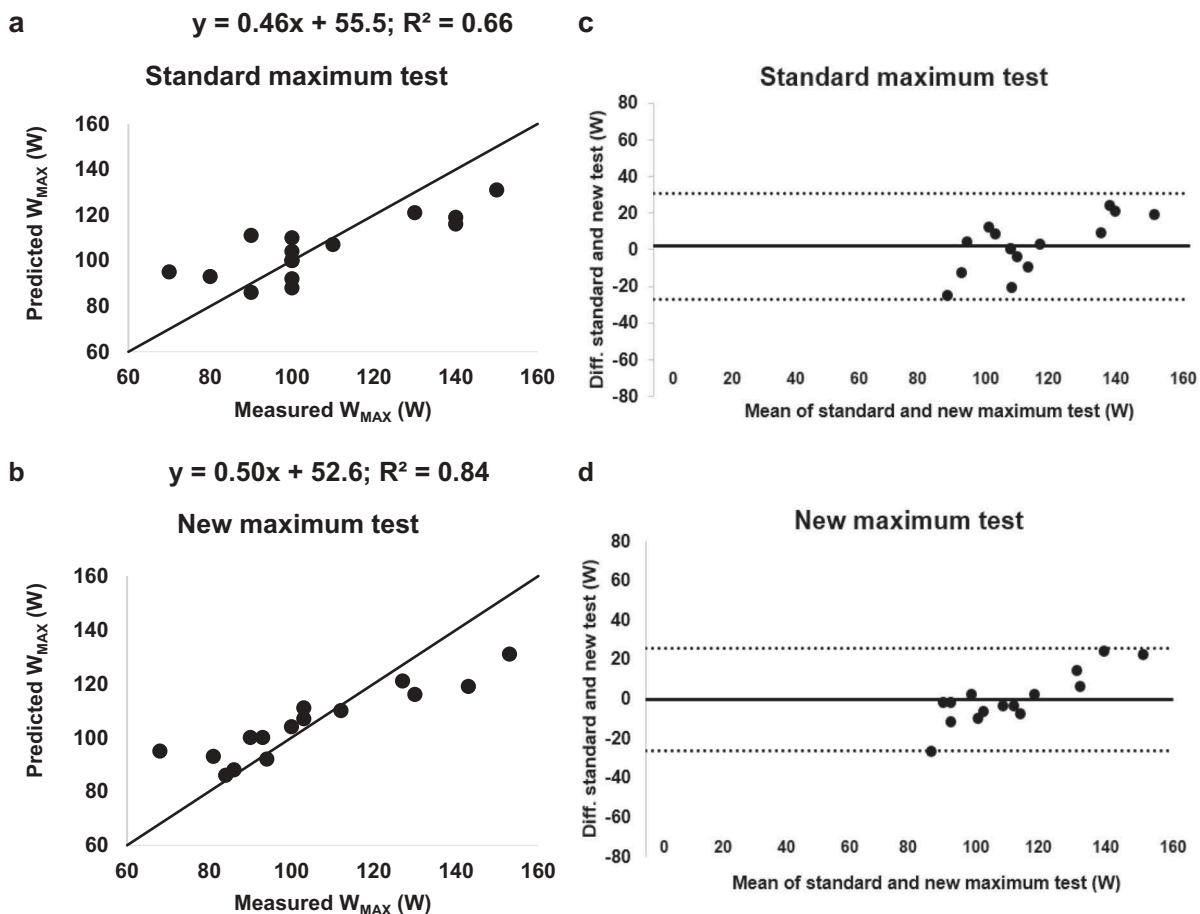


Figure 3. Predicted W_{MAX} versus measured W_{MAX} (a and b) and Bland–Altman plots (c and d) for the standard (a and c) and new (b and d) maximum test from the pilot study using the Random Forest algorithm. Line in A and B is the line of identity.

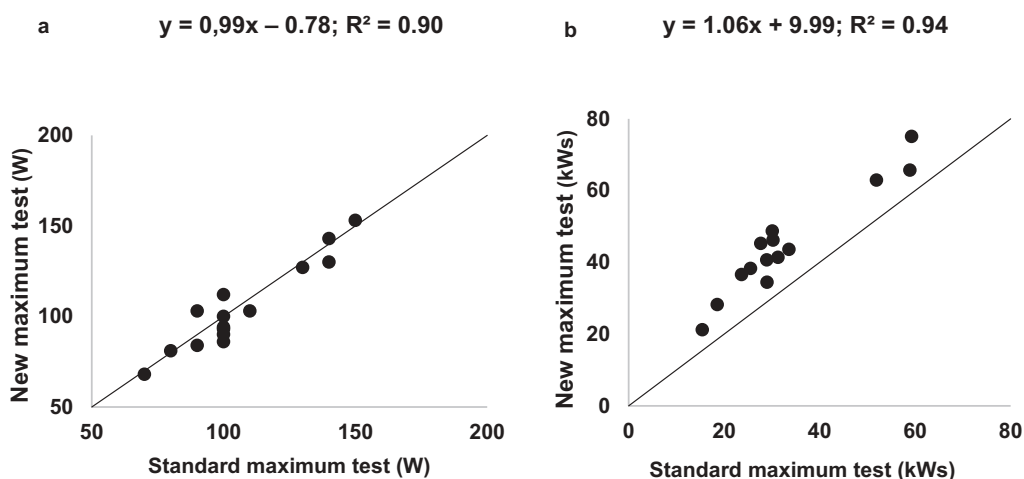


Figure 4. Measured W_{MAX} (A) from and measured work performed (B) during the new maximum test versus the standard maximum test. Line is the line of identity.

Table 4. Results of the standard maximum test and new maximum test performed in the pilot study.

	Standard maximum test	New maximum test
W_{MAX} (W)	107 ± 22	105 ± 23
Time of exercise (min)	10.6 ± 2.4	9.3 ± 2.4*
Work performed (kWs)	35.5 ± 16	47.5 ± 17**
Borg dyspnea score, peak	8.5 ± 2.0	8.7 ± 1.4
Borg leg discomfort score, peak	17.5 ± 1.8	17.6 ± 1.9
Reason for stopping exercise, n (%)***		
Dyspnea	7 (47)	8 (53)
Dyspnea plus leg discomfort	5 (33)	6 (40)
Leg discomfort	3 (20)	1 (7)

Values are mean ± standard deviation unless otherwise specified.

* Excluding the initial 3 min bicycling at 40% of predicted W_{MAX}

** Including the initial 3 min bicycling at 40% of predicted W_{MAX}

*** Reported as dyspnea, leg discomfort, both of these or other reasons

Discussion

In our search for a prediction algorithm with the ability to accurately predict W_{MAX} to a level where standard incremental W_{MAX} test can be eliminated, we investigated baseline lung function and demographic data together with results from a standard incremental W_{MAX} test from four large pharmaceutical industry-sponsored clinical studies. Since DLCO was found to have outstanding abilities for prediction of W_{MAX} , a prediction algorithm was built using Random Forest modelling of data from only one of these four studies, a multicentre study with 261 patients. FEV₁ was another an excellent predictor, as was also observed in the three remaining studies, but these studies either lacked measurements of DLCO or used treadmill rather than bicycle exercise testing. The obtained prediction algorithm was utilized to set the low-intensity start and the linear increase in the new linear W_{MAX} test in a pilot study. This study compared the new

W_{MAX} test with the standard incremental W_{MAX} test. The prediction algorithm was successful in providing reliable values of predicted W_{MAX} , and the measured value of W_{MAX} from the new test correlated better with the predicted value than did the corresponding value from the standard test. The R^2 -value (our primary variable) was 0,84 for predicted vs measured W_{MAX} for the new test. It should, however, be noted that the new W_{MAX} test will possess the same cardiovascular risks as the standard W_{MAX} test.

In comparison with the standard W_{MAX} test, the new W_{MAX} test provided similar overall results for work capacity, time of exercise, Borg scale results and reasons for stopping exercise. The slope for predicted W_{MAX} versus measured W_{MAX} was 0.50 (new test) and 0.46 (standard test), respectively, indicating that, for both tests, W_{MAX} in the high range obtained from high performers are underestimated by the prediction algorithm and that W_{MAX} in the low range (low performers) are overestimated by the prediction algorithm. The conformity was best for patients with mid-range W_{MAX} values and hence the prediction algorithm needs further improvement. An improved prediction algorithm would improve the precision of the new W_{MAX} test, so that the endurance test will be performed at an optimal level. If the patient starts the endurance test at too low or too high values as derived from the W_{MAX} test, too long or too short endurance times may be observed leading to an increased intra-test variability.

The two tests correlated well with measured W_{MAX} , but the new linear test gave continuous data (not 10W stepwise data) along the regression line and achieved the predicted value of W_{MAX} in 9.3 ± 2.4 min (programmed value 8 min), which meant that the increase was adapted to the respective individual. This resulted in a higher amount of performed work with less deviation. The

standard test and the new test showed similar mean W_{MAX} , but the amount of work performed in kW was constantly higher for the new method for both low and high performers. This is partly explained by the initial 3 min of bicycling at 40% of W_{MAX} , but this accounts only for some 30–40% of the difference. The rest of the difference may be explained by the individualized increase in work overtime. In the standard test the load started at 10 W (but can be 30–40 W based on the experience of the investigator). The incremental steps were 10 W (but steps from 5 to 20 W have been utilized), and higher increases than 10 W have been pointed out to be beneficial [8]. In the new test, the initial low load and the following linear increase to W_{MAX} are personalized. The bicycle is programmed to reach the predicted W_{MAX} in 8 min; i.e. a predicted high performer faces a much steeper increase than a predicted low performer. Also, the precision of W_{MAX} is lower in the standard test, since work performed is calculated based on 10W-increments. This is exemplified in Figure 3 where five patients have a measured W_{MAX} of 100 W in the standard test.

A large number of prediction formulas for W_{MAX} or VO_{2MAX} in COPD patients have been presented [14–25,27,28]. Most of these have been derived from single-centre studies with low numbers of patients, and the generalizability of these formulas to broader COPD populations and to multicentre studies can be questioned. We took another approach by using data from large multicentre studies sponsored by pharmaceutical companies. The resulting prediction algorithm might be more accurate and generalizable as compared to most previous prediction equations of W_{MAX} due to the bigger and broader COPD populations included in the analyses. However, the multicentre data includes many centres with different qualities in performing a correct W_{MAX} test. For DLCO, problems with standardisation across centres and large coefficients of variation have been observed [34]. A good and generalizable prediction algorithm could be helpful to achieve the best possible measured W_{MAX} in order to have the 75% level in the endurance test as accurate as possible.

One of our major learnings is that the single and multicentre approach identified very similar predictors. Our prediction algorithm identified FEV_1 (flow), DLCO (gas-exchange), FVC and VA (volume), FEF_{50} (small airways) and demographic data, which agrees well with previous publications [14–25,27,28]. One study [26] criticised the use of baseline lung function to predict VO_{2MAX} . However, this study selected six prediction formulas from three papers [14,15,22] of which four included transplant candidates and one very severe COPD patients while the validation was

done in 60 COPD-patients equally divided into GOLD 2, 3 and 4. Other papers have not challenged the possibility of predicting W_{MAX} or VO_{2MAX} . However, Fregonezi [26] highlight a very important issue of all prediction models, i.e. generalisability, both related to disease severity and quality of assessment between centres.

The current study has a number of limitations. The prediction algorithm has been based on multicentre studies with high variability. Only one study was finally used as some of the key variables for prediction have not been collected in the other studies. A limitation with Random Forest is the non-intuitive relationship between the different variables making a judgement based on clinical knowledge important. In addition, Random Forest is usually applied to larger sample sizes. In the studies, the measurements of W_{MAX} and the demographic parameters were well documented but other good predictors had a lower quality or were missing (e.g. DLCO). The prediction algorithm does not account for differences in the level of activity among the patients, and whether the patients have little or much experience from bicycling. Also, the low number of patients in the pilot study makes the results from this study sensitive to the performance by individual patients.

Areas for further improvements of the prediction algorithm include QoL-variables, variables measuring daily activities, oxygen kinetic data and values from examinations of small airway functions such as impulse oscillometry measures. The possible role of hyperinflation as predictors in more severe patients should also be explored [35].

In summary, a prediction algorithm for W_{MAX} has been developed from incremental exercise bicycle tests in previously published multicentre clinical studies. The best predictors were FEV_1 and DLCO. A new linear W_{MAX} test including the use of the prediction algorithm correlated well with the standard incremental W_{MAX} test. The new W_{MAX} test had a better correlation to the prediction algorithm, even though the algorithm was developed from the standard W_{MAX} test. In addition, an improved prediction algorithm with a slope closer to line-of-identity will benefit the new W_{MAX} test. Therefore, the new test will be utilized in future studies within our research project until the prediction algorithm has reached a level of precision that enables the W_{MAX} test to be replaced with a predicted value of W_{MAX} .

Acknowledgments

We thank Boehringer Ingelheim and AstraZeneca for providing patient-level data for this post-hoc analysis, Dr Ulf Nihlén, Department of Respiratory Medicine and

Allergology, University Hospital, Lund, Sweden for valuable comments and the study nurses at the Department of Respiratory Medicine and Allergology, University Hospital, Lund, Sweden for their assistance during the pilot study.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Swedish Heart and Lung foundation.

References

- [1] Chawla H, Bulathsinghala C, Tejada JP, et al. Physical activity as a predictor of thirty-day hospital readmission after a discharge for a clinical exacerbation of chronic obstructive pulmonary disease. *Ann Am Thorac Soc.* 2014;11(8):1203–1209.
- [2] Iyer AS, Wells JM, Bhatt SP, et al. Life-space mobility and clinical outcomes in COPD. *Int J Chron Obstruct Pulmon Dis.* 2018;13:2731–2738.
- [3] Marin JM, Alfageme I, Almagro P, et al. Multicomponent indices to predict survival in COPD: the COCOMICS study. *Eur Respir J.* 2013;42(2):323–332.
- [4] Marin JM, Carrizo SJ, Casanova C, et al. Prediction of risk of COPD exacerbations by the BODE index. *Respir Med.* 2009;103(3):373–378.
- [5] Moy ML, Teylan M, Weston NA, et al. Daily step count predicts acute exacerbations in a US cohort with COPD. *PLoS One.* 2013;8(4):e60400.
- [6] Rodriguez DA, Kortianou EA, Alison JA, et al. Heart rate recovery after 6-min walking test predicts acute exacerbation in COPD. *Lung.* 2017;195(4):463–467.
- [7] Palange P, Ward SA, Carlsen KH, et al. ERS task force recommendations on the use of exercise testing in clinical practice. *Eur Respir J.* 2007;29(1):185–209.
- [8] American Thoracic Society. American College of Chest P. ATS/ACCP Statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med.* 2003;167(2):211–277.
- [9] O'Donnell DE, Lam M, Webb KA. Measurement of symptoms, lung hyperinflation, and endurance during exercise in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 1998;158(5 Pt 1):1557–1565.
- [10] Puente-Maestu L, Palange P, Casaburi R, et al. Use of exercise testing in the evaluation of interventional efficacy: an official ERS statement. *Eur Respir J.* 2016;47(2):429–460.
- [11] Singh S, Harvey-Dunstan TC. Walking for the assessment of patients with COPD. In: Palange P, Laveneziana P, Neder JA, et al., editors. *Clinical exercise testing (ERS monograph)*. Sheffield: European Respiratory Society; 2018. p. 175–195.
- [12] Magnussen H, Paggiaro P, Schmidt H, et al. Effect of combination treatment on lung volumes and exercise endurance time in COPD. *Respir Med.* 2012;106(10):1413–1420.
- [13] Maltais F, Hamilton A, Marciniuk D, et al. Improvements in symptom-limited exercise performance over 8 h with once-daily tiotropium in patients with COPD. *Chest.* 2005;128(3):1168–1178.
- [14] Cahalin L, Pappagianopoulos P, Prevost S, et al. The relationship of the 6-min walk test to maximal oxygen consumption in transplant candidates with end-stage lung disease. *Chest.* 1995;108(2):452–459.
- [15] Carter R, Holiday DB, Stocks J, et al. Predicting oxygen uptake for men and women with moderate to severe chronic obstructive pulmonary disease. *Arch Phys Med Rehabil.* 2003;84(8):1158–1164.
- [16] Hartman JE, Slebos DJ, Boezen HM, et al. Selecting the increment size for a maximal incremental cycle test in patients with COPD. *Respirology.* 2015;20(2):352–355.
- [17] Inal-Ince D, Savci S, Coplu L, et al. Functional capacity in severe chronic obstructive pulmonary disease. *Saudi Med J.* 2005;26(1):84–89.
- [18] Carter R, Peavler M, Zinkgraf S, et al. Predicting maximal exercise ventilation in patients with chronic obstructive pulmonary disease. *Chest.* 1987;92(2):253–259.
- [19] LoRusso TJ, Belman MJ, Elashoff JD, et al. Prediction of maximal exercise capacity in obstructive and restrictive pulmonary disease. *Chest.* 1993;104(6):1748–1754.
- [20] Behnia M, Wheatley C, Avolio A, et al. Influence of resting lung diffusion on exercise capacity in patients with COPD. *BMC Pulm Med.* 2017;17(1):117.
- [21] Efremitis G, Tsiamita M, Manolis A, et al. Accuracy of pulmonary function tests in predicted exercise capacity in COPD patients. *Respir Med.* 2005;99(5):609–614.
- [22] Chuang ML, Lin IF, Vintch JR. Comparison of estimated and measured maximal oxygen uptake during exercise testing in patients with chronic obstructive pulmonary disease. *Intern Med J.* 2004;34(8):469–474.
- [23] Coquart JB, Eston RG, Lemaitre F, et al. Prediction of peak oxygen uptake from ratings of perceived exertion during a sub-maximal cardiopulmonary exercise test in patients with chronic obstructive pulmonary disease. *Eur J Appl Physiol.* 2015;115(2):365–372.
- [24] Ekblom-Bak E, Bjorkman F, Hellenius ML, et al. A new submaximal cycle ergometer test for prediction of VO₂max. *Scand J Med Sci Sports.* 2014;24(2):319–326.
- [25] Cavalheri V, Hernandez NA, Camillo CA, et al. Estimation of maximal work rate based on the 6-minute walk test and fat-free mass in chronic obstructive pulmonary disease. *Arch Phys Med Rehabil.* 2010;91(10):1626–1628.
- [26] Fregonezi G, Resqueti V, Vigil L, et al. Maximal oxygen uptake cannot be estimated from resting lung function and submaximal exercise in patients with chronic obstructive pulmonary disease. *J Cardiopulm Rehabil Prev.* 2012;32(4):219–225.
- [27] Buchfuhrer MJ, Hansen JE, Robinson TE, et al. Optimizing the exercise protocol for cardiopulmonary assessment. *J Appl Physiol Respir Environ Exerc Physiol.* 1983;55(5):1558–1564.
- [28] Debigare R, Maltais F, Mallet M, et al. Influence of work rate incremental rate on the exercise responses in patients with COPD. *Med Sci Sports Exerc.* 2000;32(8):1365–1368.
- [29] O'Donnell DE, Maltais F, Porszasz J, et al. The continuum of physiological impairment during treadmill

- walking in patients with mild-to-moderate COPD: patient characterization phase of a randomized clinical trial. *PLoS One*. 2014;9(5):e96574.
- [30] Worth H, Forster K, Eriksson G, et al. Budesonide added to formoterol contributes to improved exercise tolerance in patients with COPD. *Respir Med*. 2010;104(10):1450–1459.
- [31] Couronne R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics*. 2018;19(1):270.
- [32] Borg GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc*. 1982;14(5):377–381.
- [33] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.
- [34] Hegewald MJ, Markewitz BA, Wilson EL, et al. Single-breath diffusing capacity for carbon monoxide instrument accuracy across 3 health systems. *Respir Care*. 2015;60(3):430–436.
- [35] Eriksson G, Jarenback L, Peterson S, et al. A new approach to assess COPD by identifying lung function break-points. *Int J Chron Obstruct Pulmon Dis*. 2015;10:2193–2202.