Reduction in T_{LCO} and survival in a clinical population

Martin Raymond Miller[1] and Brendan G. Cooper[2]

[1]Institute of Applied Health Sciences, Birmingham University, Birmingham, UK. [2]Lung Function and Sleep, Queen Elizabeth Hospital, University Hospital Birmingham NHS Foundation Trust, Birmingham, UK.

Corresponding author: Martin Raymond Miller (m.r.miller@bham.ac.uk)

Shareable abstract (@ERSpublications)
A four-tier grading of lung function defined by z-scores of −1.645, −3 and −5 is simpler and better relates to subsequent survival than previous grading systems. Using ethnicity-specific GLI T_{LCO} prediction equations is appropriate. https://bit.ly/2PRxc0u

Cite this article as: Miller MR, Cooper BG. Reduction in T_{LCO} and survival in a clinical population. *Eur Respir J* 2021; 58: 2002046 [DOI: 10.1183/13993003.02046-2020].

Copyright ©The authors 2021. For reproduction rights and permissions contact permissions@ersnet.org

This article has supplementary material available from erj.ersjournals.com

This article has an editorial commentary: https://doi.org/10.1183/13993003.01326-2021

Received: 28 May 2020
Accepted: 29 March 2021

Abstract

Background How best to express the level of transfer factor of the lung for carbon monoxide (T_{LCO}) has not been properly explored.

Methods We used the most recent clinical data from 13 829 patients (54% male; 10% non-European ancestry; median age 60.5 years, range 20–97 years; median survival 3.5 years, range 0–20 years) to determine how best to express T_{LCO} function in terms of its relationship to survival.

Results The proportion of subjects of non-European ancestry with Global Lung Function Initiative (GLI) T_{LCO} z-scores above predicted was reduced, but was significantly increased between −1.5 and −3.5, suggesting the need for ethnicity-appropriate equations. Applying GLI forced vital capacity (FVC) ethnicity methodology to GLI T_{LCO} z-scores removed this ethnic bias and was used for all subsequent analysis. T_{LCO} z-scores using the GLI equations were compared with Miller's USA equations with median T_{LCO} z-scores being −1.43 and −1.50 for GLI and Miller equations, respectively (interquartile range −2.8 to −0.3 and −2.4 to −0.7, respectively). GLI T_{LCO} z-scores gave the best Cox regression model for predicting survival. A previously proposed six-tier grading system for level of lung function did not show much separation in survival risk in the less-severe grades. A new four-tier grading based on z-scores of −1.645, −3 and −5 showed better separation of risk with hazard ratio for all-cause mortality of 2.0, 3.4 and 6.6 with increasing severity.

Conclusion Applying GLI FVC ethnicity methodology to GLI T_{LCO} predictions to remove ethnic bias together with a new four-tier z-score grading best relates T_{LCO} function to survival.

Introduction

Single-breath transfer factor of the lungs for carbon monoxide (T_{LCO}) has been shown to be the best predictor of survival in the general population [1]. With the new Global Lung Function Initiative (GLI) prediction equations for T_{LCO} [2] now available it is important to look at verifying their validity by how well they fit survival in large datasets [3]. How lung function results are presented has been under scrutiny in recent years to find the best method. Using percentage of predicted (PP) was for many years an accepted means, but this method retains age, sex and size effects that may lead to false conclusions, such as the misconception that females were more susceptible to COPD [4]. Standardised residuals (that is, z-scores) are the preferred method for defining how far a subject's result is from their population norm [5, 6], but older subjects can never have z-scores as low as young people [7], which may lead to problems in grading the severity of an abnormality. T-scores have been proposed as an improved method for using spirometry data [8], which might overcome this latter issue because T-scores relate the subject's observed value to the subject's sex- and height-specific maximum predicted value seen in early adulthood and not to their age-related predicted value.

For grading the severity of airflow obstruction the previous American Thoracic Society (ATS)/European Respiratory Society (ERS) recommendations used cut levels of PP values [6] to give five levels of abnormality. This approach was subsequently adapted to be used with z-scores [9]. Because this approach

was just for airflow obstruction, it is uncertain whether it can be applied to other lung function indices and whether five distinct levels are justified. The ATS/ERS previously recommended three grades of severity for T_{LCO} [6]. Grading the degree of lung function reduction is arbitrary, in that the degree of reduction is a continuum and the previously proposed five levels of grading for airflow obstruction may be implying differences that are not substantive.

We looked at a large dataset of patients (>13 800) with their survival and compared two different prediction equations for T_{LCO}; we used different methods for presenting the results to look for the best method for grading T_{LCO} and spirometry result abnormalities.

Methods

We extracted anonymised patient data from a clinical lung function database set up in January 1996 and closed in October 2016 which stored all routine lung function tests performed at the Queen Elizabeth Hospital (University Hospitals Birmingham National Health Service Foundation Trust, Birmingham, UK). The data in this study complied with patient confidentiality and comprise the results of the most recent attendance tests for spirometry and diffusing capacity measured on the same day from 13 829 patients (54% male) referred for any reason for lung function tests. All patients had their survival registered up to 19 September 2016 in UK National Health Service data records. All tests were performed according to ATS/ERS criteria on equipment validated to conform to ATS/ERS specifications [10].

GLI prediction equations [11] were used for spirometry with patients of non-European ancestry classified as ethnicity "other". For transfer factor the GLI equations [2] were compared with those of Miller *et al.* [12]. Initial analysis showed that GLI T_{LCO} predictions gave many subjects extremely low z-scores (values as low as −23.0), which was not seen for the Miller equations, and this was particularly seen in females. After contacting GLI about this, an error was discovered in their data, with the wrong sex attributed to some subjects in their dataset [13]. Revised GLI-2020 T_{LCO} prediction equations have been issued, which were used for all our analyses [14]. T-scores were derived as the number of standard deviations the index result was from the subject's sex- and height-specific highest GLI predicted value for that index, which was at an age in their early 20s and which differed between the different indices. We graded for lung function abnormality with the ATS/ERS six-tier system for grading airflow obstruction [6] using PP and Quanjer *et al.*'s [9] system using z-scores, and also with the ATS/ERS four-tier system suggested for T_{LCO} [6]. We looked to see if one grading system might work for T_{LCO}, forced vital capacity (FVC) and forced expiratory volume in 1 s (FEV_1).

All Cox regression models were created using Stata/SE v16.1 (StataCorp, College Station, TX, USA) using lung function indices as predictors for survival with stratification for age quintiles, sex and smoking status (see supplementary material for detail). Time to event was date of death or censored at 19 September 2016 (last entry into database) if no event occurred. Proportional hazard assumptions were tested by looking for parallel separation of log–log survival curves and with Schoenfeld residuals tests [15]. As a measure of the best fit we took the lowest Akaike information criterion (AIC), which is a measure of the relative amount of information lost by a given model [16] and calculated Harrel's C concordance index, where a value of 0.5 means the predicted outcomes are no better than guessing and 1.0 means the predicted outcomes are all completely accurate.

Results

Out of the 13 829, patients 54% were male, 37% had never smoked, 47% were ex-smokers and 16% were current smokers. There were 1435 (10.4%) patients of non-European ancestry. The lung function request forms gave putative diagnoses, relevant symptoms and the subject's therapy. 12.0% indicated COPD, 9.6% asthma, 9.0% interstitial lung disease, 4.1% bronchiectasis, 2.6% emphysema and 14.6% had breathlessness. There were 7508 (54%) patients who were not on any treatment.

Because patients of non-European ancestry appeared to have lower T_{LCO} z-scores than expected, suggesting that an adjustment for ethnicity might be appropriate, we then took the GLI ethnicity-specific FVC coefficients [11] and applied them to T_{LCO}, since any adjustment would probably need to correct for size differences (see supplementary material for detail). Figure 1 shows the percentage of subjects who were of non-European ancestry by bins of T_{LCO} z-score values with (pale grey) and without (dark grey) using the GLI ethnicity-specific FVC coefficients. It shows that there were significantly fewer subjects of non-European ancestry with GLI T_{LCO} z-scores better than predicted (z-score >0.0) and significantly more than expected in the range −1.5 to −3.5 (Chi-squared 79.5, p<0.0001). The GLI z-scores using ethnic specific FVC coefficients did not show this bias (Chi-squared 16.8, p=0.16). All of the T_{LCO} analyses presented used ethnicity-specific FVC coefficients. Alveolar volume (V_A) showed even more skewness in

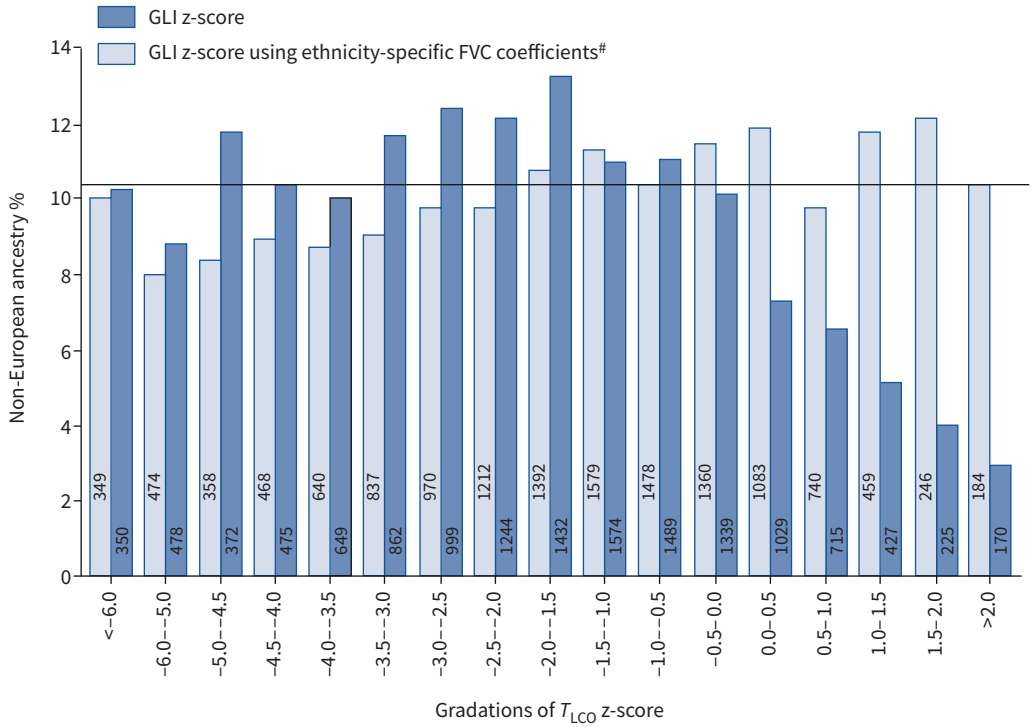

FIGURE 1 Columns representing the percentage of subjects who were of non-European ancestry in bins of gradation of Global Lung Function Initiative (GLI) transfer factor of the lung for carbon monoxide (T_{LCO}) z-score values, with the total number of subjects in each bin shown within each column. The horizontal black line indicates the percentage of non-European ancestry found in the dataset of 13 829 patients as a whole. FVC: forced vital capacity. #: derived using the coefficients for ethnicity used by the GLI methodology for FVC [11].

relation to ethnicity than did T_{LCO}, and the GLI FVC correction methodology did not satisfactorily correct these V_A differences (supplementary figures S1 and S2).

The median, interquartile range and range of values found for age, survival and lung function indices are shown in table 1. There were 4371 patients (32% of the total) who had normal FEV_1, FVC, FEV_1/FVC and T_{LCO} (all z-scores $\geqslant -1.645$) and of these 70% were not on any treatment. The distribution for GLI

TABLE 1 Results for the 13 829 patients expressed as z-scores from Global Lung Function Initiative (GLI) for spirometry [11] and transfer factor of the lung for carbon monoxide (T_{LCO}) [2]

	Median	Interquartile range	Range (minimum–maximum)
Age years	60.48	48.9 to 70.2	20.0 to 96.7
Survival years	3.52	1.4 to 7.7	0.0 to 19.9
FVC z-score	−0.872	−1.8 to 0.0	−6.7 to 9.8
FEV_1 z-score	−1.503	−2.6 to −0.6	−6.6 to 10.4
FEV_1/FVC z-score	−1.010	−2.3 to −0.1	−6.1 to 3.9
FEV_1/FVC	0.718	0.6 to 0.8	0.1 to 1.0
T_{LCO} z-score	−1.430	−2.8 to −0.3	−13.0 to 5.2
MT_{LCO} z-score	−1.499	−2.4 to −0.7	−6.3 to 5.3
K_{CO} z-score	−0.584	−1.8 to 0.5	−9.2 to 16.6
V_A z-score	−1.120	−2.2 to −0.2	−11.3 to 3.3
T_{LCO}PP	78.9	60.8 to 95.6	7.0 to 193.0

FVC: forced vital capacity; FEV_1: forced expiratory volume in 1 s; M: derived using Miller *et al.* [12] equations; K_{CO}: transfer coefficient of the lung for carbon monoxide; V_A: alveolar volume; PP: % predicted from GLI equations.

TABLE 2 Hazard ratios (HR) with 95% confidence limits (95% CL) for all-cause mortality from Cox regression using Global Lung Function Initiative and Miller (M) prediction equations with the American Thoracic Society/European Respiratory Society recommended cut-points for transfer factor of the lung for carbon monoxide (T_{LCO}) [6] using percentage predicted (PP) and for z-scores chosen to approximate the numbers found with the PP cut-points

	HR (95% CL)	Population distribution %	Male %	Age years
T_{LCO} z-score				
Normal ⩾ –1.645	1.0	54.5	49.7	56.4
Mild ⩾ –3.0	2.0 (1.8–2.1)	22.8	58.4	61.2
Moderate ⩾ –5.0	3.4 (3.2–3.7)	16.7	59.8	62.4
Severe < –5.0	6.6 (6.0–7.3)	6.0	57.2	61.6
T_{LCO}PP				
Mild ⩾60%	1.9 (1.8–2.1)	21.3	54.4	58.6
Moderate ⩾40%	3.3 (3.0–3.6)	17.0	61.1	63.9
Severe <40%	6.2 (5.6–6.9)	7.0	66.3	65.8
MT_{LCO}PP				
Mild ⩾60%	1.8 (1.6–1.9)	15.6	56.2	53.4
Moderate ⩾40%	2.8 (2.6–3.0)	21.1	58.2	62.2
Severe <40%	5.5 (5.0–6.0)	8.7	63.2	64.5

The comparator group (HR=1) included those with T_{LCO} ⩾ –1.645 (the lower limit of normal). Data are presented as mean, unless otherwise stated.

T_{LCO} z-scores was left-skewed as expected for patient data, whereas that for the Miller equations was not (supplementary figure S3). The distribution of T_{LCO} against age did not lend itself to the use of a T_{LCO} quotient index (that is, relating T_{LCO} to a sex-specific lower 1st centile value) in the way that a FEV_1 quotient has been used (supplementary figures S4 and S5 and supplementary table S1) [7].

Cox models were prepared using severity grading by PP as previously proposed by the ATS/ERS for T_{LCO} [6] along with cut-points for z-scores (−1.645, −3 and −5) that were selected to roughly give similar numbers in each grade. Table 2 shows the results for these Cox regression models using the GLI and Miller equations. The models for T_{LCO} T-scores and for Miller z-scores did not comply with assumptions for proportional hazard and are not shown. The degree of fit for the GLI z-score model was similar to that for PP (AIC 51 392 and 51 402, respectively, and both had Harrel's C=0.66). The PP Miller model was slightly less good, with AIC 51 472. Both PP models showed a slightly older and more male distribution in the most severe grade. A sensitivity analysis using the 12 916 subjects aged <80 years (3903 deaths) gave hazard ratios (HR) (95% confidence limits) for T_{LCO} z-scores of 2.0 (1.8–2.2), 3.5 (3.2–3.8) and 6.4 (5.7–7.1) for the three grades *versus* 2.0 (1.8–2.1), 3.4 (3.2–3.7) and 6.6 (6.0–7.3) for the full dataset.

Figure 2 is a plot of T_{LCO} z-score against PP for 11 117 patients with values below predicted with the intersections of z-scores −3 and −5 and 40% predicted and 60% predicted, which shows that for a given z-score males had a lower PP value than females. The subjects in quadrants A and C are those where PP puts them in a more severe grade than z-score and quadrants B and D are those subjects where PP puts them in a less severe grade than z-score. Table 3 shows the percentages of subjects in these quadrants who were male, in the oldest tertile for age, in the tallest tertile for height and with ethnicity as "other". For all these indicators there was a significant difference between the groupings shown in table 3 ($p<0.005$, Chi-squared) with quadrants A and C (where PP grades more severely than z-score) comprising more males and more subjects of older age, taller height and European ancestry. While there was a higher percentage of deaths in groups A and C compared to B and D, this was accounted for by older age and greater proportion of male subjects, so in Cox regression with the comparator group being all those with T_{LCO} greater than lower limit of normal (LLN) and with age and sex accounted for by stratification, the HRs in A and C due to their lung function was found to be lower than that for groups B and D. Supplementary figures S6 and S7 show that the residual standard deviation (RSD) for T_{LCO} in the GLI equations varies with the subject's sex, age, height and ethnicity. A smaller RSD in comparison to the magnitude of predicted (being female, shorter, of non-European ancestry and younger) will lead to PP grading being better than z-score grading, and *vice versa*.

Cox models with T_{LCO} were prepared using the six tier gradations of z-scores and PP previously recommended for FEV_1 in the context of airflow limitation [9]. The results for T_{LCO} z-scores and PP

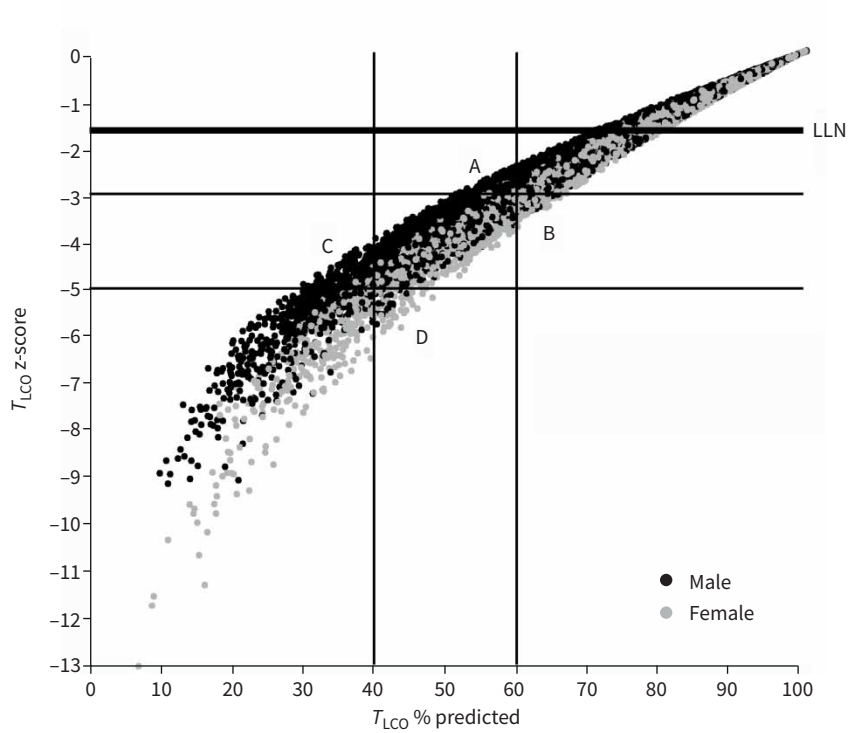

FIGURE 2 Plot of transfer factor of the lung for carbon monoxide (T_{LCO}) z-scores against percentage predicted (PP)T_{LCO} values for all the 11 117 subjects with results below predicted. Cut lines of z-score −3 and −5 and 40% predicted and 60% predicted are shown. Quadrants A and C indicate those subjects where PP puts them into a more severe grade than z-score. Quadrants B and D indicate those subjects where PP put them in a less severe grade than z-scores. LLN: lower limit of normal.

models are shown in table 4. The numbers of subjects in the most severe grading varied considerably, from 325 with Miller z-scores to 1649 for GLI z-scores, with 647 in the worst PP grade. There was little difference in HR for all-cause mortality between the mild, moderate and moderately severe grading, suggesting that fewer grades were more appropriate. The results for the six-tier gradation for FVC in the whole group and for FEV_1 in those with FEV_1/FVC <LLN are shown in table 5. Only the model for FVC z-score met the assumptions for proportional hazard. In all the models shown there was little difference in HR for all-cause mortality for the lower grades, suggesting that fewer tiers would be more appropriate.

TABLE 3 Numbers of subjects in quadrants A, B, C and D (figure 2) and the percentages who were male, in the oldest tertile, in the tallest tertile and of non-European ancestry, coded as "other"

	Group	Subjects n	Died	Male	Oldest tertile	Tallest tertile	Other	HR (95% CL)
PP worse than z-score	A and C	570	54.4	84.6	78.4	35.9	4.2	
PP better than z-score	B and D	221	29.9	33.5	0.0	22.6	22.6	
Chi-squared				199.0	626.9	26.1	63.7	
PP worse than z-score	A	346	49.4	81.8	81.5	34.1	4.1	2.4 (2.0–2.9)
PP better than z-score	B	137	27.7	33.3	0.0	22.5	23.1	3.0 (2.1–4.2)
Chi-squared				110.2	405.2	11.2	42.7	
PP worse than z-score	C	224	62.1	88.8	73.7	38.8	4.5	4.3 (3.6–5.2)
PP better than z-score	D	74	37.8	33.8	0.0	23.0	21.6	5.0 (3.4–7.4)
Chi-squared				90.3	221.6	16.4	20.6	

Data are presented as %, unless otherwise stated. The hazard ratio (HR) for mortality shown was derived with the comparator group being all those with transfer factor of the lung for carbon monoxide (T_{LCO}) z-score greater than lower limit of normal. 95% CL: 95% confidence limits; PP: percentage predicted.

TABLE 4 Hazard ratios (HR) with 95% confidence limits (95% CL) for all-cause mortality from Cox regression using transfer factor of the lungs for carbon monoxide (T_{LCO}) with Global Lung Function Initiative (GLI) and Miller predictions according to the Quanjer six-level grading system from z-score cut-points [9] and the American Thoracic Society/European Respiratory Society six-level grading from percentage predicted (PP) [6] for airflow obstruction

	GLI T_{LCO} z-score				Miller T_{LCO} z-score				GLI T_{LCO}PP			
	HR (95% CL)	Population distribution %	Male %	Age years	HR (95% CL)	Population distribution %	Male %	Age years	HR (95% CL)	Population distribution %	Male %	Age years
Normal	1	54.6	49.8	56.4	1	54.6	49.9	58.1	1	53.4	49.8	56.5
Mild	1.6 (1.5–1.9)	7.0	55.0	61.0	1.7 (1.6–1.9)	11.0	51.8	60.4	1.7 (1.5–1.9)	8.8	45.7	54.3
Moderate	1.9 (1.7–2.1)	8.8	59.8	60.8	2.3 (2.1–2.5)	12.8	56.4	61.1	2.0 (1.9–2.2)	13.2	58.9	61.0
Moderately severe	2.5 (2.2–2.7)	7.0	60.1	62.0	3.1 (2.8–3.4)	9.6	59.1	60.0	2.8 (2.6–3.1)	10.3	58.6	63.1
Severe	3.2 (2.9–3.5)	10.7	59.7	62.7	4.8 (4.3–5.2)	9.7	63.6	58.2	4.0 (3.7–4.4)	9.6	64.0	65.0
Very severe	5.2 (4.7–5.6)	11.9	58.6	61.8	7.5 (6.4–8.7)	2.4	77.2	51.8	7.5 (6.7–8.3)	4.7	68.9	65.7

Data are presented as mean, unless otherwise stated.

Table 6 shows the new proposed four-tier grading applied to FVC in the whole dataset and for FEV_1 in those with airflow obstruction. The model for FVC was not as good as that for T_{LCO} with AIC 52 366 and Harrel's C=0.58 compared to 51 392 and 0.66, respectively, for the T_{LCO} model (table 2). The models for FVCPP and for FEV_1PP did not meet the assumptions of proportional hazard. Changing the thresholds to −3 and −4.5 gave slightly better fits for T_{LCO} and FVC, but the model for FEV_1 did not meet the assumptions of proportional hazard (supplementary material). Only with cut levels of −3 and −5 could satisfactory models be derived for T_{LCO}, FVC and FEV_1.

Discussion

We have shown that the GLI prediction equations for T_{LCO} gave the best survival prediction model for our population of patients, which was better than the survival predictions from spirometry and is in keeping with what was found for predicting survival in the general population [1]. We initially found that the original GLI equations gave many more extremely low z-scores than did the Miller equations, but this was due to an error in the original GLI data. Using revised GLI equations we found that the 10.4% of our

TABLE 5 Hazard ratios (HR) with 95% confidence limits (95% CL) for all-cause mortality from Cox regression using Global Lung Function Initiative predictions according to the Quanjer six-level grading system from z-score cut-points [9] and the American Thoracic Society/European Respiratory Society six-level grading from percentage predicted (PP) for airflow obstruction [6]

	FVC z-score	FEV_1 z-score	HR (95% CL)	Population distribution %	Male %	Age years	FVC PP	FEV_1 PP	HR (95% CL)	Population distribution %	Male %	Age years
All patients						n=13 829						
Normal	⩾LLN		1.0	70.8	52.7	59.7	⩾LLN		1.0	70.8	52.7	59.7
Mild	⩾−2.0		1.4 (1.2–1.5)	7.8	54.1	58.9	⩾70%		1.3 (1.1–1.4)	9.2	54.4	50.8
Moderate	⩾−2.5		1.6 (1.4–1.7)	7.8	55.1	57.9	⩾60%		1.6 (1.5–1.7)	10.2	55.7	59.9
Moderately severe	⩾−3.0		1.8 (1.6–2.0)	5.8	60.5	57.7	⩾50%		2.3 (2.1–2.5)	5.9	57.6	59.7
Severe	⩾−4.0		2.8 (2.5–3.1)	5.5	57.1	55.0	⩾35%		3.2 (2.9–3.6)	3.5	61.9	57.8
Very severe	<−4.0		4.0 (3.4–4.7)	2.3	59.2	45.7	<35%		4.6 (3.4–6.2)	0.4	52.1	52.1
FEV_1/FVC <LLN (airflow obstruction)						n=4862						
Normal		⩾LLN	1.0	21.4	57.1	61.7		⩾LLN	1.0	21.4	57.1	61.7
Mild		⩾−2.0	1.0 (0.8–1.3)	8.1	61.8	63.0		⩾70%	1.0 (0.8–1.3)	6.9	58.5	51.4
Moderate		⩾−2.5	1.1 (1.0–1.4)	13.8	59.8	63.3		⩾60%	1.1 (0.9–1.3)	15.1	59.5	60.8
Moderately severe		⩾−3.0	1.4 (1.2–1.7)	14.2	58.4	64.3		⩾50%	1.3 (1.1–1.6)	16.5	57.1	62.7
Severe		⩾−4.0	1.7 (1.4–2.0)	24.5	59.3	62.3		⩾35%	1.6 (1.4–1.9)	22.3	56.4	63.0
Very severe		<−4.0	2.7 (2.3–3.2)	18.1	56.0	52.5		<35%	2.6 (2.2–3.1)	17.7	59.2	59.2

Data are presented as mean, unless otherwise stated. FVC: forced vital capacity; FEV_1: forced expiratory volume in 1 s; LLN: lower limit of normal.

TABLE 6 Hazard ratios (HR) with 95% confidence limits (95% CL) for all-cause mortality from Cox regression using Global Lung Function Initiative predictions according to the proposed four-tier grading system from z-score cut-points

	FVC z-score	FEV_1 z-score	HR (95% CL)	Population distribution %	Male %	Age years	FVC PP	FEV_1 PP	HR (95% CL)	Population distribution %	Male %	Age years
All patients						n=13 829						
Normal	⩾LLN		1	70.8	52.7	59.7	⩾LLN		1.0	70.8	52.7	59.7
Mild	⩾−3.0		1.5 (1.4–1.7)	21.4	56.2	58.2	⩾60		1.4 (1.3–1.6)	19.4	55.1	55.6
Moderate	⩾−5.0		3.0 (2.7–3.3)	7.3	58.3	53.2	⩾40		2.6 (2.4–2.8)	8.7	58.8	59.1
Severe	<−5.0		4.9 (3.5–6.9)	0.5	48.6	39.0	<40		3.3 (2.7–4.0)	1.1	66.0	55.0
FEV_1/FVC <LLN						n=4862						
Normal		⩾LLN	1	21.4	57.3	61.7		⩾LLN	1.0	21.4	57.3	61.7
Mild		⩾−3.0	1.2 (1.1–1.4)	36.0	59.7	63.6		⩾60	1.1 (0.9–1.3)	22.0	59.2	57.8
Moderate		⩾−5.0	1.8 (1.6–2.1)	37.5	58.6	60.0		⩾40	1.5 (1.3–1.7)	31.7	56.0	62.8
Severe		<−5.0	4.3 (3.4–5.6)	5.1	52.6	44.3		<40	2.3 (1.9–2.6)	24.9	61.8	60.3

Data are presented as mean, unless otherwise stated. FVC: forced vital capacity; FEV_1: forced expiratory volume in 1 s; LLN: lower limit of normal.

population who were of non-European ancestry had T_{LCO} z-scores that were distributed lower than anticipated. There was no *a priori* reason to expect that our patients of non-European ancestry would have a higher prevalence of moderate disease than those of European ancestry, so this suggested that accounting for ethnicity would be appropriate [17]. We chose to adopt the methodology used by GLI for ethnicity with respect to FVC and apply this to T_{LCO}, which assumes that the ethnic difference in T_{LCO} relates to size issues rather than any intrinsic difference in the ability to transfer carbon monoxide to the blood. Using these ethnicity-specific coefficients removed the uneven distribution of ethnicity across the T_{LCO} z-score values in our patients. GLI were originally unable to propose ethnicity-specific predictions for T_{LCO} due to a lack of available data. While our proposal is not as satisfactory as deriving ethnicity-appropriate reference values from suitable data from all ethnicities, it does offer a sensible approach until such data are available. This will mean that all patients can be assessed against the most appropriate standard available for them as an individual and so avoid unnecessary misclassification of lung function results. We have only been able to examine this using the GLI category "other" for ethnicity, and future work will need to see if the other GLI ethnicity FVC coefficients can improve T_{LCO} predictions for those ethnicities.

For spirometry, we previously found that in the extremely old, the GLI predictions were high [18], suggesting that rigorous entry criteria for data to be used in prediction equations might lead to using values from subjects not properly representative of usual people for that age. However, for GLI T_{LCO} equations the upper age limit was suggested to be 85 years [2], but coefficients are available up to the age of 90 years. Thus, our 42 patients who were older than this were being assessed against their 90-year-old predicted values. This means that the z-scores of these subjects may be inappropriately low. Despite this, there was no age effect on the severity grading categories for GLI z-scores (table 4), whereas there were possible age effects within the grading system using the Miller equations and with PP values.

We have proposed a new simplified four-tier system for grading lung function from z-scores based on survival analysis. The six-tier system of Quanjer *et al.* [9] proposed for assessing airways obstruction with FEV_1 did not show much separation in survival for the lower grades of severity (table 4). All gradings of severity are arbitrarily set on a continuous scale, and the six levels suggest a degree of separation between grades that does not appear to be justified. We think that using broader, but fewer, grades that reflect survival is a better and simpler approach. We have found marked differences in the severity groups identified by grading using z-scores *versus* PP for T_{LCO}. The analysis around the discordant subjects in quadrants A, B, C and D in figure 2 suggest that the z-score thresholds more correctly grade the severity with respect to all-cause mortality, since hazard for mortality better related to the z-score cut-offs than the PP cut-offs. These discordant groups showed significant skew in age, sex and height.

This difference in the way that z-scores and PP grade lung function deficit is related to the fact that PP takes no account of the scatter of the range of values in fit and healthy people and how this scatter relates to the magnitude of the index and the sex, ethnicity, age and height of the subject. PP assumes that the degree of lung function deficit is always simply proportional to the predicted. There is no evidence or law indicating that, for example, a 50% reduction in lung function for a 25-year-old is in some way equivalent

to the same reduction in a 75-year-old. Consider T_{LCO}, an index that has a numerically relatively large predicted value, which, for a 75-year-old male of average height (175 cm) has a predicted value of 8.14 mmol·min^{-1}·kPa^{-1} with a 5th centile value of 5.94 mmol·min^{-1}·kPa^{-1}, which is at 73.0% of the predicted value, whereas for a 25-year-old male of this height, the LLN is at 79.3% predicted. Now consider forced expiratory flow at 25–75% of FVC, an index that has a numerically smaller predicted value with a relatively wide range of values found in normal healthy people. For the same 75-year-old man, the predicted is 2.17 L·s^{-1}, 5th centile is 0.90 L·s^{-1}, which is at 41.4% predicted, whereas for a 25-year-old man of this height the LLN is at 63.6% predicted. So comparing degrees of lung function reduction by PP is going to be problematic, since 73% predicted and 79% predicted in one is equivalent in terms of grading lung function deficit to 41% and 64% predicted in the other. In most previous prediction equations, for example the European Community for Steel and Coal equations [5] and from the third National Health and Nutrition Examination Survey [19], the RSD from the regression equations was the same for all subjects. GLI have observed that the scatter of results in healthy subjects was not uniform for all subjects and this was accounted for in their predictions, so people of different age, height, sex and ethnicity may have different RSD values, as shown in supplementary figures S8 and S9. These differences in approach mean that z-scores will grade differently from PP and appear to relate better to outcome.

Although others have found that T-scores for FEV_1 were effective in determining abnormality in lung function [8], we have found that T-scores did not convincingly relate to survival in our patients. For FVC, T-scores were equivalently good at predicting survival to z-scores, but in all the other indices T-scores violated Cox regression assumptions and gave poorer predictions. The potential advantage of T-scores over z-scores is that in the older population z-scores cannot go so low as in younger subjects because the median value falls with age with the RSD not changing in proportion. Since T-scores are relating a subject's result to their predicted value at a much younger age, their values can go just as low in old age as in younger subjects. In our data this does not seem to be advantageous in predicting mortality since healthy older subjects have a much lower T-score than z-score due to age alone and this adversely distorts the survival analysis. Also the origins of T-scores with bone density [20] is predicated on that fact that values of bone density as a young adult can pertain into later life to define a normal range. This is not the case for lung function. The spread of lower values of T_{LCO} against age did not favour the easy adoption of a T_{LCO} quotient in the way that has proven effective for FEV_1 [7, 21–23]. This may just reflect the spread of our data and our referral pattern for younger people for lung transplantation, and so this should be further explored in other datasets.

The limitations to our study include that we could not separate patients into disease categories, since the referral information was not verifiable. Additionally, we did not know the cause of death, so we could not look for respiratory causes of mortality to enhance the analysis. Our results are only applicable for adults, and in paediatrics such survival analysis may not be appropriate. However, the z-score grading system should still be applicable and avoid sex and size effects. Our patients were classified as either "white" or "other" in terms of GLI with respect to ethnicity and it will be necessary to check from data in African American and Asian populations whether the application of the FVC methodology to T_{LCO} for these ethnic groups is appropriate.

We conclude that the revised GLI-2020 equations with the addition of ethnicity-specific coefficients give a better spread of z-score results compared to the Miller equations and this leads to better survival prediction. We have found that a four-tier grading system for grading T_{LCO} function (z-scores of −1.645, −3 and −5) was appropriately related to survival.

Acknowledgements: We thank the staff of the Lung Function department at the Queen Elizabeth Hospital (Birmingham, UK) for their dedication and professionalism in working with our patients to obtain the best possible test results.

Conflict of interest: M.R. Miller has nothing to disclose. B.G. Cooper has nothing to disclose.

References

1 Neas LM, Schwartz J. Pulmonary function levels as predictors of mortality in a national sample of US adults. *Am J Epidemiol* 1998; 147: 1011–1018.

2 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.

3 Ward H, Cooper B, Miller MR. Validation of lung function prediction equations from patient survival data. *Eur Respir J* 2012; 39: 1181–1187.

4 Jordan RE, Miller MR, Lam K-BH, *et al.* Sex, susceptibility to smoking and chronic obstructive pulmonary disease: the effect of different diagnostic criteria. Analysis of the Health Survey for England. *Thorax* 2012; 67: 600–605.

5 Quanjer P, Tammeling G, Cotes JE, *et al.* Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J* 1993; 16: 5–40.

6 Pellegrino R, Viegi G, Brusasco V, *et al.* Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948–968.

7 Miller MR, Pedersen OF. New concepts for expressing forced expiratory volume in 1 s arising from survival analysis. *Eur Respir J* 2010; 35: 873–882.

8 Lee SW, Kim HK, Baek S, *et al.* Development of a spirometry T-score in the general population. *Int J Chron Obstruct Pulmon Dis* 2016; 11: 369–379.

9 Quanjer PH, Pretto JJ, Brazzale DJ, *et al.* Grading the severity of airways obstruction: new wine in new bottles. *Eur Respir J* 2014; 43: 505–512.

10 Miller MR, Hankinson J, Brusasco V, *et al.* Standardisation of spirometry. *Eur Respir J* 2005; 26: 319–338.

11 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: 1324–1343.

12 Miller A, Thornton JC, Warshaw R, *et al.* Single breath diffusing capacity in a representative sample of the population of Michigan, a large industrial state. Predicted values, lower limits of normal, and frequencies of abnormality by smoking history. *Am Rev Respir Dis* 1983; 127: 270–277.

13 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* 2020; 56: 1750010.

14 GLI Consortium. Gas Transfer Project. http://gli-calculator.ersnet.org/ Date last accessed: 26 November 2020.

15 Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; 81: 515–526.

16 Bozdogan H. Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions. *Psychometrika* 1987; 52: 345–370.

17 Pesola GR, Sunmonu Y, Huggins G, *et al.* Measured diffusion capacity *versus* prediction equation estimates in blacks without lung disease. *Respiration* 2004; 71: 484–492.

18 Miller MR, Thinggaard M, Christensen K, *et al.* Best lung function equations for the very elderly selected by survival analysis. *Eur Respir J* 2014; 43: 1338–1346.

19 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: 179–187.

20 Faulkner KG. The tale of the T-score: review and perspective. *Osteoporos Int* 2005; 16: 347–352.

21 Pedone C, Scarlata S, Scichilone N, *et al.* Alternative ways of expressing FEV_1 and mortality in elderly people with and without COPD. *Eur Respir J* 2013; 41: 800–805.

22 Hegendörfer E, Vaes B, Andreeva E, *et al.* Predictive value of different expressions of forced expiratory volume in 1 second (FEV_1) for adverse outcomes in a cohort of adults aged 80 and older. *J Am Med Dir Assoc* 2017; 18: 123–130.

23 Huang T-H, Hsiue T-R, Lin S-H, *et al.* Comparison of different staging methods for COPD in predicting outcomes. *Eur Respir J* 2018; 51: 1700577.