



# Two is better than one: the double diffusion technique in classifying heart failure

Gerald S. Zavorsky <sup>1</sup> and Piergiuseppe Agostoni <sup>2,3</sup>

<sup>1</sup>Department of Physiology and Membrane Biology, University of California Davis, Sacramento, CA, USA. <sup>2</sup>Department of Critical Cardiology, Centro Cardiologico Monzino IRCCS, Milan, Italy. <sup>3</sup>Department of Clinical Sciences and Community Health, Cardiovascular Section, University of Milan, Milan, Italy.

Corresponding author: Gerald S. Zavorsky ([gzavorsky@ucdavis.edu](mailto:gzavorsky@ucdavis.edu))



Shareable abstract (@ERSpublications)

*D*<sub>LNO</sub>+*D*<sub>LCO</sub> is better at classifying heart failure than either one alone <https://bit.ly/3upnR46>

Cite this article as: Zavorsky GS, Agostoni P. Two is better than one: the double diffusion technique in classifying heart failure. *ERJ Open Res* 2024; 10: 00644-2023 [DOI: 10.1183/23120541.00644-2023].

Copyright ©The authors 2024

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact [permissions@ersnet.org](mailto:permissions@ersnet.org)

Received: 1 Sept 2023  
Accepted: 15 Nov 2023

## Abstract

**Background** Heart failure (HF) is a chronic condition in which the heart does not pump enough blood to meet the body's demands. Diffusing capacity of the lung for nitric oxide (*D*<sub>LNO</sub>) and carbon monoxide (*D*<sub>LCO</sub>) may be used to classify patients with HF, as *D*<sub>LNO</sub> and *D*<sub>LCO</sub> are lung function measurements that reflect pulmonary gas exchange. Our objectives were to determine 1) if *D*<sub>LNO</sub> added to *D*<sub>LCO</sub> testing predicts HF better than *D*<sub>LCO</sub> alone and 2) whether the binary classification of HF is better when *D*<sub>LNO</sub> z-scores are combined with *D*<sub>LCO</sub> z-scores than using *D*<sub>LCO</sub> z-scores alone.

**Methods** This was a retrospective secondary data analysis in 140 New York Heart Association Class II HF patients (ejection fraction <40%) and 50 patients without HF. z-scores for *D*<sub>LNO</sub>, *D*<sub>LCO</sub> and *D*<sub>LNO</sub>+*D*<sub>LCO</sub> were created from reference equations from three articles. The model with the lowest Bayesian Information Criterion was the best predictive model. Binary HF classification was evaluated with the Matthews Correlation Coefficient (MCC).

**Results** The top two of 12 models were combined z-score models. The highest MCC (0.51) was from combined z-score models. At most, only 32% of the variance in the odds of having HF was explained by combined z-scores.

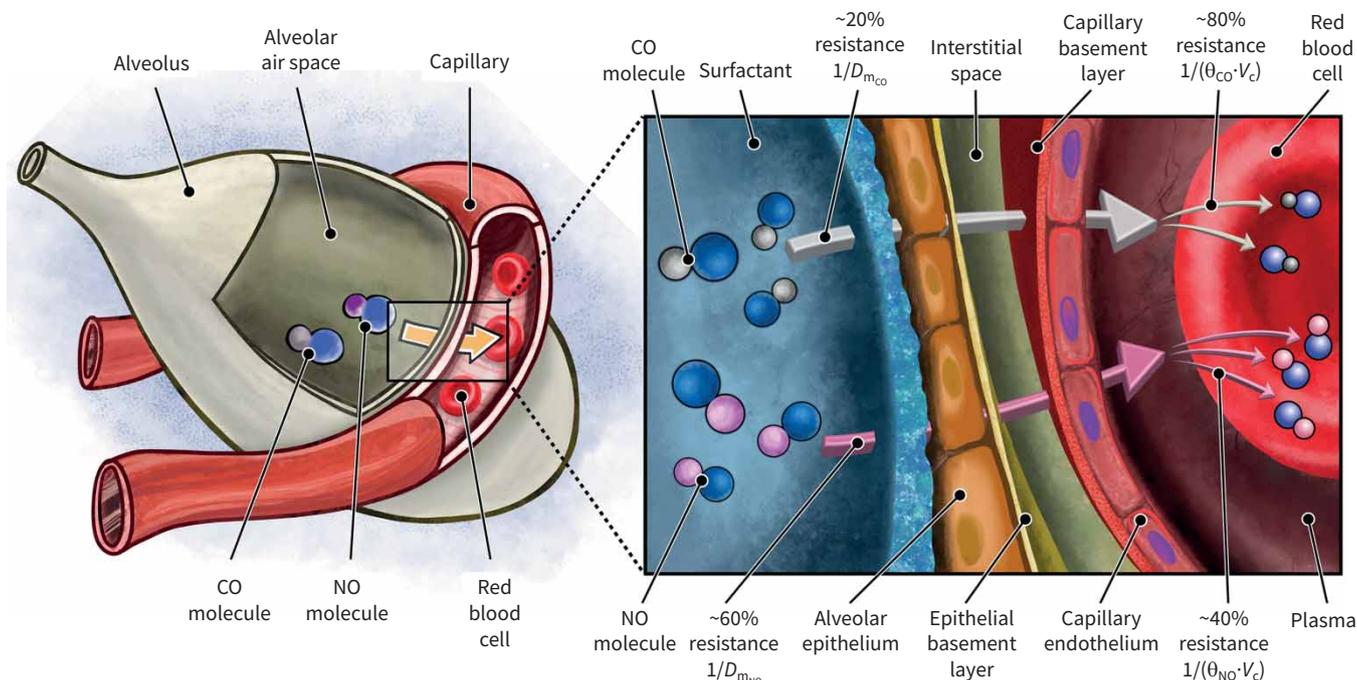
**Conclusions** Combined z-scores explained 32% of the variation in the likelihood of an individual having HF, which was higher than models using *D*<sub>LNO</sub> or *D*<sub>LCO</sub> z-scores alone. Combined z-score models had a moderate ability to classify patients with HF. We recommend using the NO–CO double diffusion technique to assess gas exchange impairment in those suspected of HF.

## Introduction

In patients with heart failure (HF), there is significant pulmonary gas exchange impairment even at rest, as the alveolar–arterial oxygen tension difference (*P*<sub>A–aO<sub>2</sub></sub>) is excessive (~28 mmHg) and the arterial oxygen tension (*P*<sub>aO<sub>2</sub></sub>) is low (mild hypoxaemia, ~72 mmHg) [1]. Diffusing capacity of the lung for carbon monoxide (*D*<sub>LCO</sub>) and nitric oxide (*D*<sub>LNO</sub>) can identify gas exchange issues in HF patients. *D*<sub>LNO</sub> z-scores share 48–56% variance with *P*<sub>A–aO<sub>2</sub></sub>, while *D*<sub>LCO</sub> z-scores share 38–49% with *P*<sub>aO<sub>2</sub></sub> [2, 3], aiding gas exchange assessment.

Combined *D*<sub>LNO</sub> and *D*<sub>LCO</sub> measurements together provide a detailed understanding of lung gas exchange issues. While *D*<sub>LCO</sub> reflects mostly pulmonary vascular/blood volume problems, accounting for ~70–80% of the red blood cell barrier, *D*<sub>LNO</sub> predominantly captures the diffusion between the alveolar–capillary membrane and red blood cell membrane (~60%) (figure 1) [4]. Together, they provide a comprehensive view of different diffusion pathways: what *D*<sub>LCO</sub> misses, *D*<sub>LNO</sub> captures. Despite its potential, the NO–CO double diffusion technique remains underutilised. There are companies that manufacture the device [5], but clinicians do not know about it, and if they do not know about it, they cannot request it. Furthermore, no devices that measure *D*<sub>LNO</sub> are yet US Food and Drug Administration approved, and despite several conversations with manufacturers, they have been reluctant to spend the time and effort to make them so.





**FIGURE 1** An illustration of how inhaled nitric oxide (NO) and carbon monoxide (CO) travel from the lungs' air sacs (alveoli) to combine with haemoglobin within red blood cells. The process can be described using the Roughton–Forster equation, which breaks down the total resistance to NO or CO uptake into two components: 1) membrane resistance ( $1/D_m$ , where  $D_m$  is the membrane diffusing capacity), which represents the resistance encountered as the gases move from the alveolar–capillary membrane to the red blood cell membrane, and 2) red blood cell resistance ( $1/(\theta \cdot V_c)$ , where  $V_c$  is the capillary blood volume), which encompasses both the diffusion resistance and the resistance due to the chemical combination within the red blood cell. The key takeaways are: 1) for CO uptake, most resistance (~70–80%) is inside the red blood cell, with the remaining ~25% occurring in the alveolar–capillary membrane, 2) in contrast, for NO uptake, most of the resistance (~60%) lies between the alveolar–capillary membrane and the red blood cell membrane, with the red blood cell resistance accounting for the remaining 40% of the resistance to NO diffusion [39], and 3) notably, the red blood cell's interior is crucial to determining the membrane resistance to NO.

Thus, the NO–CO double diffusion technique has only been used for research purposes by a select few scientists. Recently, the European Respiratory Society (ERS) 2017 Technical Standards for  $D_{LNO}$  [4] highlighted its use, increasing exposure to clinicians of  $D_{LNO}$ 's technical advantages over  $D_{LCO}$ , and favour its routine use in pulmonary function testing [6]. Thus, hopefully, companies will move towards obtaining the required governmental approvals so that  $D_{LNO}$  becomes as ubiquitous as  $D_{LCO}$  in routine pulmonary function testing.

HF leads to a decline in the alveolar membrane diffusing capacity ( $D_m$ ), especially as HF worsens [7]. Even though some HF research employs the single-breath NO–CO method [8–12], combining  $D_{LCO}$  and  $D_{LNO}$  in HF evaluation remains under-researched. Considering ~35% of HF patients also have COPD [13], which affects both pulmonary capillary blood volume and  $D_m$  [3], the NO–CO double diffusion method, capturing both these issues, may offer a more accurate HF assessment than just  $D_{LCO}$ , which primarily detects  $V_c$  and/or haemoglobin concerns.

The ERS/American Thoracic Society 2022 Technical Standard on lung function interpretation encourages the classification of the severity of lung function impairment based on z-scores [14]. A z-score is a standardised score that indicates how many standard deviations a value is from the mean of a reference population. The z-score, compared with the percentage predicted, is a better way to classify pulmonary diffusion impairment since the percentage predicted at the lower limit of normal (LLN) changes with age [15].

As such, this study aimed to 1) assess  $D_{LNO}$  z-scores' impact on  $D_{LCO}$  z-scores and their HF correlation and 2) classify HF using the Matthews Correlation Coefficient (MCC) [16]. We hypothesised that  $D_{LNO}$  z-scores added to  $D_{LCO}$  z-scores improves HF association and classification compared with  $D_{LCO}$  z-scores alone.

## Methods

This was a retrospective secondary data analysis that utilised four prior studies involving White HF patients and the single-breath NO–CO double diffusion technique [9–11, 17]. At Centro Cardiologico Monzino (Milan, Italy), all patients provided consent, allowing their data to be used for retrospective research in a completely anonymous manner. As the data were previously published by the same research group [9–11, 17], the transferred data were fully anonymised so that a data transfer agreement was not necessary. These four studies were a combination of cross-sectional or case–control research designs.

This article’s analyses focus on classifying those with New York Heart Association (NYHA) Class II HF from controls using  $D_{LNO}$ ,  $D_{LCO}$  or their combined z-scores, distinguishing it from the goals of previous studies [9–11, 17]. Throughout these studies, we maintained consistency regarding breath-hold time (BHT) (~5.5 s), testing apparatus and the research team involved. Since all four studies’ data were gathered at the same research centre, employing fixed statistical models was deemed suitable. We opted for a BHT of ~5.5 s because the NO electrochemical sensor used for  $D_{LNO}$  assessments functions within the ppm spectrum. A 10 s breath-hold could reduce the exhaled NO concentration to below the ppb spectrum, which surpasses the sensor’s detection limits. As the established reference equations are grounded on a BHT of  $6.2 \pm 1.3$  s [4, 15, 18], the potential influence of a slightly shorter BHT on gas diffusion in the lung is offset by comparing the HF patients to reference equations that employ similar BHTs.

### HF prediction

We derived the  $D_{LCO}$  and  $D_{LNO}$  z-scores from three primary sources: ERS Technical Standards [4], MUNKHOLM *et al.* [18] and ZAVORSKY and CAO [15]. Subsequently, we combined the individual z-scores ( $D_{LCO}$  z-score +  $D_{LNO}$  z-score) for each method. As a result, both the ERS Technical Standards [4] and MUNKHOLM *et al.* [18] yielded three outcomes per patient: one  $D_{LCO}$  z-score, one  $D_{LNO}$  z-score and one combined z-score. On the other hand, ZAVORSKY and CAO [15] presented prediction formulas using segmented linear regression and generalised additive models for location, scale and shape (GAMLSS). Consequently, this approach produced two sets of scores: two  $D_{LCO}$  z-scores, two  $D_{LNO}$  z-scores and two combined z-scores for each patient. It is worth noting that the ERS Technical Standards’ reference equations [4] drew from aggregated adult data [19–21], while ZAVORSKY and CAO’s [15] were based on a combined dataset of children and adults [18–22].

These 12 normalised z-scores, considering age, sex, height, altitude and lung function device, acted as independent variables in binary logistic regression for HF prediction. (For both linear and segmented linear regression, the z-score is determined by the formula: (measured value–predicted value)/residual standard error. For the z-score computation *via* the GAMLSS model from ZAVORSKY and CAO [15], please consult the footnote in table 3 of their publication.) The model with the lowest Bayesian Information Criterion (BIC), a typical selection method for lung function reference equations, was chosen for the best overall fit [23, 24].

Between-model differences in BIC were interpreted as follows: BIC difference 0–2: weak evidence (50–75% probability that the lower BIC model is better); BIC difference 2–6: positive evidence (75–95% probability that the lower BIC model is better); BIC difference 6–10: strong evidence (95–99% probability that the lower BIC model is better); and BIC difference >10: very strong evidence (>99% probability that the lower BIC model is better) [25].

To predict HF through pulmonary diffusion anomalies, we utilised the z-score threshold derived from the optimal combination of sensitivity and specificity from the receiver operating characteristic (ROC) curve analysis for all combined models. Additionally, we employed a z-score threshold of  $-1.645$  (corresponding to the 5th percentile, or LLN) for the standalone  $D_{LCO}$  and  $D_{LNO}$  z-score models. We applied two-sided independent t-tests to contrast absolute  $D_{LCO}$  with  $D_{LCO}$  z-scores, absolute  $D_{LNO}$  with  $D_{LNO}$  z-scores and the  $D_{LNO}/D_{LCO}$  ratio between different groups. We used the Benjamini–Hochberg procedure to control the false discovery rate among MCCs and set it at 0.05 [26].

### Classification of HF versus controls

Model efficacy was further gauged using the area under the ROC curve (AUC) and the MCC. Notably, the MCC is an especially reliable metric when evaluating binary categories in datasets where the number of disease cases does not match non-disease cases, as in our pooled dataset [16, 27, 28]. The MCC considers true positives, true negatives, false positives and false negatives to provide a comprehensive score for model classification. High MCC scores are only achieved when predictions accurately classify a significant proportion of diseased and non-diseased patients, regardless of any class imbalance. We also derived the 95% confidence interval for the MCC from 1000 bootstrap samples and evaluated each MCC accordingly [29, 30].

The prevalence of HF in Italians aged between 60 and 85 years is ~4% [31]. This starkly contrasts with the prevalence in our four-study cohort (~74%). Due to this disparity, we have chosen not to report the positive and negative predictive values, along with the false positive and false negative rates. Nevertheless, we did calculate parameters such as the true positive rate (sensitivity, probability of detection), true negative rate (specificity), false omission rate, false detection rate, and both positive and negative likelihood ratios since they are unaffected by disease prevalence.

SPSS Statistics version 29 (IBM, Armonk, NY, USA) and R version 4.2.2 (www.r-project.org) were used for statistical analyses (specifically, R packages AICcmodeavg version 2.3.1 and glmnet version 4.1.6). A p-value <0.05 signified statistical significance.

## Results

### General findings

140 patients with NYHA Class II HF (113 males, 27 females) and 50 control subjects (26 males, 24 females) were included in this analysis (table 1). The same researchers tested all subjects in the same hospital centre [9–11, 17]. Only baseline values were used when patients were tested multiple times or where they were in different experimental conditions. These were a cohort of ambulatory low ejection fraction (<40%) HF patients that were regularly followed. The subjects without HF, on average, were 4 years younger and had a mean body mass index  $2.5 \text{ kg}\cdot\text{m}^{-2}$  less than those with HF ( $p<0.05$ ). Approximately 50% of those with HF had an obstructive or restrictive spirometric pattern compared with 20% without HF ( $p<0.001$ ). The mean BHT for the diffusing capacity testing was  $5.5\pm 0.5 \text{ s}$  (range 4.2–7.4 s).

There were significantly different diffusing capacities between groups (z-scores in non-HF:  $D_{\text{LCO}} = -1.03 \pm 0.84$ ,  $D_{\text{LNO}} = -1.30 \pm 0.91$ , combined z-scores =  $-2.33 \pm 1.45$ ; z-scores in HF:  $D_{\text{LCO}} = -2.18 \pm 1.32$ ,  $D_{\text{LNO}} = -2.35 \pm 1.19$ , combined z-scores =  $-4.53 \pm 2.34$ ) ( $p<0.001$  between groups for all, z-scores created from segmented regression [15]).

About 24% of the subjects without HF showed mild  $D_{\text{LCO}}$  impairment (per the diffusion impairment classification scheme of ZAVORSKY and CAO [15]). Conversely, 65% of the HF patients displayed mild, moderate or severe  $D_{\text{LCO}}$  impairment. About 36% of the non-HF subjects showed mild  $D_{\text{LNO}}$  impairment. Conversely, 69% of the HF patients displayed mild, moderate or severe  $D_{\text{LNO}}$  impairment.

The absolute values for  $D_{\text{LCO}}$  and  $D_{\text{LNO}}$  for those with HF were  $17.0\pm 5.7$  and  $69\pm 23 \text{ mL}\cdot\text{min}^{-1}\cdot\text{mmHg}^{-1}$ , respectively. The absolute values for  $D_{\text{LCO}}$  and  $D_{\text{LNO}}$  for those without HF were  $21.8\pm 4.9$  and  $88\pm 21 \text{ mL}\cdot\text{min}^{-1}\cdot\text{mmHg}^{-1}$ , respectively. The 95% CI of the difference was 3.3–6.4 and 13–26  $\text{mL}\cdot\text{min}^{-1}\cdot\text{mmHg}^{-1}$  for  $D_{\text{LCO}}$  and  $D_{\text{LNO}}$ , respectively (two-sided p-value <0.001). The  $D_{\text{LNO}}/D_{\text{LCO}}$

TABLE 1 Subject characteristics

	Patients without HF (n=50)	Patients with HF (n=140)
<b>Sex</b>		
Female	24	27
Male	26	113
<b>Age (years)</b>	62±8 (36 to 75)	66±11 (29 to 86)
<b>Height (cm)</b>	169±8 (150 to 187)	170±8 (150 to 197)
<b>Weight (kg)</b>	69±12 (43 to 99)	77±15 (48 to 131)
<b>BMI (<math>\text{kg}\cdot\text{m}^{-2}</math>)</b>	24.1±3.1 (16.8 to 32.3)	26.6±4.2 (18.4 to 40.4)
<b>FEV<sub>1</sub> (L)</b>	2.75±0.54 (1.52 to 4.28)	2.23±0.67 (0.87 to 3.87)
<b>FVC (L)</b>	3.62±0.76 (1.91 to 5.49)	3.05±0.85 (1.52 to 5.62)
<b>FEV<sub>1</sub>/FVC</b>	0.77±0.07 (0.53 to 0.88)	0.73±0.09 (0.42 to 0.92)
<b>FEV<sub>1</sub> z-scores</b>	-0.29±1.25 (-2.41 to 2.71)	-1.51±1.18 (-4.50 to 1.74)
<b>FVC z-scores</b>	-0.12±1.11 (-2.43 to 2.26)	-1.38±1.23 (-4.07 to 1.92)
<b>FEV<sub>1</sub>/FVC z-scores</b>	-0.35±0.97 (-3.44 to 1.43)	-0.44±1.16 (-4.11 to 2.10)
<b>Individuals with obstruction (FEV<sub>1</sub>/FVC &lt;LLN)</b>	4 (8)	20 (14)
<b>Individuals with a restrictive spirometry pattern (FEV<sub>1</sub>/FVC &gt;LLN and FVC &lt;LLN)</b>	6 (12)	49 (35)
<b>Subjects with either obstruction or restriction</b>	10 (20)	69 (49)
Data are presented as n, mean±SD (range) or n (%). HF: heart failure; BMI: body mass index; FEV <sub>1</sub> : forced expiratory volume in 1 s; FVC: forced vital capacity; LLN: lower limit of normal.		

ratio was  $4.09 \pm 0.67$  for those with no disease and  $4.20 \pm 0.92$  for those with heart disease (95% CI of the difference  $-0.36$ – $0.12$ ) (two-sided p-value 0.36), indicating that the ratio cannot be used to distinguish those with and those without HF. The correlation between  $D_{LCO}$  and  $D_{LNO}$  in  $\text{mL} \cdot \text{min}^{-1} \cdot \text{mmHg}^{-1}$  was 0.79 (95% CI 0.73–0.85) and the correlation between  $D_{LCO}$  and  $D_{LNO}$  z-scores was 0.73 (95% CI 0.64–0.79) when the z-scores were created from segmented regression [15].

### Prediction results

Table 2 presents the summary results of the 12 models. The “ $\Delta$ BIC” column indicates the change in BIC compared with the first model. In this case, there is positive evidence that Model 1 is a better fit than Models 2 and 3, strong evidence that Model 1 is better than Models 4–7, and very strong evidence that Model 1 is better than Models 8–12. Thus, models that used ERS reference equations [4] or GAMLSS reference equations [15] were the best fit among the 12 models. Approximately 32% of the variability in the odds of having HF can be explained by combined z-scores when the z-scores were generated from the ERS equations. In other words, the combined z-scores from Model 1 provide information that helps explain 32% of the variance in the likelihood of an individual having HF. Four of the six highest-ranked models were those involving models with combined z-scores.

The “BIC weight” column in table 2 is used to see how much more likely one model is the correct model compared with the other models that have been fitted. In this case, Model 1 is 4.1 times more likely to be a better fit than Model 2 ( $0.69/0.17=4.1$ ). So, for example, Model 1 has a BIC weight of 0.69, which suggests that there is a 69% chance that Model 1 is the best model compared with the 11 other models tested. In the second row, Model 2 has a BIC weight of 0.17, which suggests that there is a 17% chance that Model 2 is the best model compared with the other 11 models tested. The “Cumulative weight” column in table 2 shows the percentage chance that a particular group of models is the best in the table. For example, Model 2 has a cumulative weight of 0.86, suggesting that there is an 86% chance that either Model 1 or Model 2 is the best model of all the models tested in table 2.

While table 2 summarises the comparative results of different reference equations from three distinct studies [4, 15, 18], a detailed examination of each reference equation reveals a consistent trend. The BIC

**TABLE 2** Comparing the Bayesian Information Criterion (BIC) using diffusing capacity of the lung for carbon monoxide ( $D_{LCO}$ ), diffusing capacity of the lung for nitric oxide ( $D_{LNO}$ ) or the composite ( $D_{LCO}+D_{LNO}$ ) z-scores as the independent predictors of heart failure

Model number	BIC	$\Delta$ BIC	BIC weight	Cumulative weight	Log-likelihood	Nagelkerke $R^2$
Model 1. Combined z-scores obtained from ERS Task Force reference equations [4]	183.2	0.0	0.69	0.69	–86.4	0.32
Model 2. Combined z-scores obtained from GAMLSS reference equations of ZAVORSKY and CAO [15]	186.0	2.78	0.17	0.86	–87.8	0.30
Model 3. $D_{LCO}$ z-scores obtained from ERS Task Force reference equations [4]	187.6	4.42	0.08	0.94	–88.6	0.29
Model 4. $D_{LNO}$ z-scores obtained from GAMLSS reference equations of ZAVORSKY and CAO [15]	190.3	7.11	0.02	0.96	–89.9	0.27
Model 5. Combined z-scores obtained from segmented regression reference equations of ZAVORSKY and CAO [15]	190.6	7.41	0.02	0.98	–90.1	0.27
Model 6. Combined z-scores obtained from reference equations of MUNKHOLM <i>et al.</i> [18]	191.5	8.24	0.01	0.99	–90.5	0.27
Model 7. $D_{LNO}$ z-scores obtained from ERS Task Force reference equations [4]	191.8	8.54	0.01	1.00	–90.6	0.26
Model 8. $D_{LCO}$ z-scores obtained from reference equations of MUNKHOLM <i>et al.</i> [18]	196.1	12.93	0.00	1.00	–92.8	0.24
Model 9. $D_{LCO}$ z-scores obtained from segmented regression reference equations of ZAVORSKY and CAO [15]	197.2	13.94	0.00	1.00	–93.3	0.23
Model 10. $D_{LNO}$ z-scores obtained from segmented regression reference equations of ZAVORSKY and CAO [15]	198.1	14.91	0.00	1.00	–93.8	0.22
Model 11. $D_{LCO}$ z-scores obtained from GAMLSS reference equations of ZAVORSKY and CAO [15]	200.0	16.82	0.00	1.00	–94.8	0.21
Model 12. $D_{LNO}$ z-scores obtained from reference equations of MUNKHOLM <i>et al.</i> [18]	200.2	17.0	0.00	1.00	–94.9	0.21

Pooled data from Italy [9–11, 17]. ERS: European Respiratory Society; GAMLSS: generalised additive models for location, scale and shape. K (estimated parameters for each model)=2.

for combined z-scores was always superior to the BIC for standalone  $D_{LNO}$  or  $D_{LCO}$  z-scores within the same reference equation set, with a minimum BIC difference of 4 units. This pattern emerged across various models. Specifically, using the combined z-scores from the ERS reference equations [4] resulted in a lower BIC than using individual  $D_{LCO}$  or  $D_{LNO}$  z-scores from that set. A similar trend was observed in the reference equations from MUNKHOLM *et al.* [18], where combined z-scores led to a lower BIC than standalone  $D_{LCO}$  or  $D_{LNO}$  scores. In the segmented linear regression or GAMLSS reference equation models by ZAVORSKY and CAO [15], the combined z-scores again demonstrated a lower BIC than the individual  $D_{LCO}$  or  $D_{LNO}$  z-scores. Therefore, while Models 1 and 2 stood out as the best overall models among the other 10 presented, combined z-scores consistently offered better BIC results when partitioning the three studies separately [4, 15, 18].

### Odds ratios from the prediction models

Table 3 displays the odds ratio of heart disease for every 1 unit increase in z-scores. For example, when using the calculated  $D_{LCO}$  z-scores from the ERS Task Force equations [4], it can be shown that for every 1 unit increase in  $D_{LCO}$  z-scores, the odds of having HF are reduced by 45% to 74%.

The odds that  $D_{LCO}$ ,  $D_{LNO}$  or combined ( $D_{LCO}+D_{LNO}$ ) models are below the LLN when a patient has HF are presented in table 4. For example, someone with HF has a ~6–34 times increase in odds that the combined z-scores are below the LLN when a patient has HF (when using the ERS prediction equations [4]).

### Classification results

Table 5 shows that Model 1 (row 1) had a higher MCC than models presented in rows 6–12. The positive and negative likelihood ratios describe the probability of disease shifts when the finding is present and absent, respectively [32]. While the AUC for Model 1 is 0.79, demonstrating good discriminatory ability between positive and negative cases, this can be misleading as the MCC was only 0.51 (table 5). So, the power of Model 1 in determining true positives and true negatives while minimising false positives and false negatives was only “moderate”. (An MCC of +1 represents a perfect classification, a MCC of 0 indicates random classification and an MCC of –1 indicates total disagreement between observation and classification. Since the MCC of 0.51 is above the random classification level and indicates a correlation that is more than half-way towards perfect classification, in that sense one may describe an MCC of 0.51 as moderate, as it is past the half-way point but not near perfect.)

### Discussion

This study evaluated the combined effectiveness of  $D_{LNO}$  and  $D_{LCO}$  z-scores in predicting and classifying NYHA Class II HF patients from controls. Class II HF patients, given their mild exercise limitations and relatively few symptoms, are often less inclined to adopt the recommended comprehensive therapy regimen for HF. Our analysis revealed that the combined z-scores significantly outperformed individual z-scores, with an 86% probability that one of the first two combined z-score models was the best predictor of HF among the 12 we assessed. Interestingly, the first combined model alone had a 69% likelihood of being the top performer.

The most informative metric we used was the MCC, which captures the true balance between positive and negative classifications, minimising errors. Four out of the top six MCC values came from models utilising combined z-scores. While the AUC is often used as a performance metric, it can be misleading, especially

**TABLE 3** A comparison of the odds ratios (95% CI) from different reference equations using the diffusing capacity of the lung for carbon monoxide ( $D_{LCO}$ )-only, diffusing capacity of the lung for nitric oxide ( $D_{LNO}$ )-only or combined ( $D_{LCO}+D_{LNO}$ ) z-score value as the independent predictor of heart failure (HF)

Reference equation used	$D_{LCO}$ -only z-score model	$D_{LNO}$ -only z-score model	Combined z-score model ( $D_{LCO}+D_{LNO}$ )
ERS [4]	0.37 (0.26 to 0.54)	0.38 (0.26 to 0.55)	0.56 (0.45 to 0.69)
MUNKHOLM <i>et al.</i> [18]	0.47 (0.35 to 0.64)	0.46 (0.33 to 0.63)	0.62 (0.52 to 0.74)
GAMLSS [15]	0.49 (0.36 to 0.66)	0.55 (0.44 to 0.69)	0.65 (0.56 to 0.76)
Segmented regression [15]	0.42 (0.30 to 0.60)	0.40 (0.28 to 0.58)	0.57 (0.46 to 0.71)

ERS: European Respiratory Society; GAMLSS: generalised additive models for location, scale and shape. For example, for every 1 unit increase in the  $D_{LNO}$  z-score obtained from the ERS Task Force reference equations, the odds of having HF decrease by 45% to 74%. The independent variables were the actual z-scores from various prediction equations. All  $p < 0.001$ .

**TABLE 4** The odd ratios (95% CI) that diffusing capacity of the lung for carbon monoxide ( $D_{LCO}$ ), diffusing capacity of the lung for nitric oxide ( $D_{LNO}$ ) or combined ( $D_{LCO}+D_{LNO}$ ) are below the lower limit of normal (LLN) when a patient has heart failure

Reference equation used	$D_{LCO}$ -only z-score model (z-score < -1.645)	$D_{LNO}$ -only z-score model (z-score < -1.645)	Combined z-score model ( $D_{LCO}+D_{LNO}$ )
ERS [4]	11.5 (5.0 to 26.4)	3.6 (1.7 to 7.4)	14.6 (6.2 to 34.0)
MUNKHOLM <i>et al.</i> [18]	5.3 (2.6 to 10.7)	4.0 (2.0 to 7.8)	26.9 (6.3 to 115.0)
GAMLSS [15]	5.1 (2.5 to 10.3)	5.6 (2.6 to 12.1)	14.1 (4.8 to 41.2)
Segmented regression [15]	5.9 (2.8 to 12.3)	4.0 (2.0 to 7.9)	17.6 (5.2 to 59.1)

ERS: European Respiratory Society; GAMLSS: generalised additive models for location, scale and shape. For the ERS combined z-score model, the LLN was defined as a combined z-score of < -3.76. For the MUNKHOLM *et al.* combined z-score model, the LLN was defined as a combined z-score of < -4.87. For the segmented combined z-score model, the LLN was defined as a combined z-score of < -4.38. For the GAMLSS combined z-score model, the LLN was defined as a combined z-score of < -5.92. All p<0.001.

in datasets with high disease prevalence [33]. For instance, an 80% accuracy in a population with 74% prevalence is not a notable achievement [27]. The MCC, which aggregates true positives, true negatives, false positives and true negatives, offers a more truthful representation, especially when both true positives and true negatives are equally significant.

**TABLE 5** Binary classification comparisons of heart failure (HF) versus no HF ranked from the highest Matthews Correlation Coefficient (MCC) to the lowest MCC

Row	Model number	z-score cut-off	MCC	AUC	TPR (%)	TNR (%)	FOR (%)	FDR (%)	+LR	-LR
1	Model 1. Combined z-scores obtained from ERS Task Force reference equations [4]	-3.76	0.51 (0.40 to 0.62)	0.79 (0.73 to 0.85)	74	84	47	7	4.6	0.31
2	Model 3. $D_{LCO}$ z-scores obtained from ERS Task Force reference equations [4]	-1.65	0.47 (0.34 to 0.58)	0.76 (0.69 to 0.84)	69	84	51	8	4.3	0.37
3	Model 6. Combined z-scores obtained from reference equations of MUNKHOLM <i>et al.</i> [18]	-4.86	0.44 (0.35 to 0.52)	0.74 (0.67 to 0.81)	53	96	58	3	13.2	0.49
4	Model 5. Combined z-scores obtained from segmented regression reference equations of ZAVORSKY and CAO [15]	-4.38	0.42 (0.31 to 0.51)	0.73 (0.66 to 0.81)	53	94	58	4	8.8	0.50
5	Model 2. Combined z-scores obtained from GAMLSS reference equations of ZAVORSKY and CAO [15]	-5.92	0.42 (0.31 to 0.52)	0.74 <sup>#</sup> (0.66 to 0.81)	55	92	58	5	6.9	0.49
6	Model 9. $D_{LCO}$ z-scores obtained from segmented regression reference equations of ZAVORSKY and CAO [15]	-1.65	0.36 <sup>#</sup> (0.23 to 0.48)	0.71 <sup>#</sup> (0.62 to 0.79)	67	72	56	13	2.7	0.46
7	Model 8. $D_{LCO}$ z-scores obtained from reference equations of MUNKHOLM <i>et al.</i> [18]	-1.65	0.35 <sup>#</sup> (0.21 to 0.47)	0.70 <sup>#</sup> (0.61 to 0.78)	67	72	56	13	2.4	0.46
8	Model 11. $D_{LCO}$ z-scores obtained from GAMLSS reference equations of ZAVORSKY and CAO [15]	-1.65	0.35 <sup>#</sup> (0.21 to 0.48)	0.69 <sup>#</sup> (0.61 to 0.78)	69	70	56	14	2.3	0.45
9	Model 4. $D_{LNO}$ z-scores obtained from GAMLSS reference equations of ZAVORSKY and CAO [15]	-1.65	0.34 <sup>#</sup> (0.18 to 0.50)	0.65 <sup>#</sup> (0.56 to 0.75)	89	42	43	19	1.5	0.27
10	Model 10. $D_{LNO}$ z-scores obtained from segmented regression reference equations of ZAVORSKY and CAO [15]	-1.65	0.30 <sup>#</sup> (0.16 to 0.43)	0.67 <sup>#</sup> (0.58 to 0.76)	69	64	57	16	1.9	0.48
11	Model 12. $D_{LNO}$ z-scores obtained from reference equations of MUNKHOLM <i>et al.</i> [18]	-1.65	0.30 <sup>#</sup> (0.15 to 0.44)	0.66 <sup>#</sup> (0.56 to 0.75)	77	54	54	18	1.7	0.42
12	Model 7. $D_{LNO}$ z-scores obtained from ERS Task Force reference equations [4]	-1.65	0.26 <sup>#</sup> (0.11 to 0.41)	0.62 <sup>#</sup> (0.53 to 0.72)	84	40	52	20	1.4	0.39

Brackets represent the 95% CI (bootstrapped and bias-corrected). AUC: area under the receiver operating characteristic curve; TPR: true positive rate (sensitivity); TNR: true negative rate (specificity); FOR: false omission rate; FDR: false discovery rate; +LR: positive likelihood ratio; -LR: negative likelihood ratio. #: the MCC in that row significantly differs from Model 1 (first row), controlling for the FDR.

In this pooled dataset, the  $D_{LCO}$  z-scores and  $D_{LNO}$  z-scores shared 53% (95% CI 41–64%) of their variance. This implies that for HF patients, approximately half of the variability in one z-score can be predicted from the other. Since there is not complete overlap between the  $D_{LNO}$  and  $D_{LCO}$  z-scores (*i.e.* they do not share 100% of their variance), it is advisable to measure both  $D_{LNO}$  and  $D_{LCO}$  to capture the entirety of the variance present. This accentuates the advantages of employing the NO–CO double diffusion technique in pulmonary function tests over traditional  $D_{LCO}$  tests alone. The combined use of  $D_{LNO}$  and  $D_{LCO}$  z-scores provides a broader perspective, capturing variances that either score might miss individually. There are certain standalone conditions, like anaemia [34], polycythaemia, CO poisoning [35] or pulmonary capillary haemorrhage, where  $D_{LNO}$  is much less sensitive than  $D_{LCO}$  in detecting microvascular changes in the haemoglobin sink for oxygen transfer, but generally the combination proves more robust in classifying HF.

Although  $D_{LNO}$  is technically superior to  $D_{LCO}$  [6], it does not render  $D_{LCO}$  obsolete. The combined approach offers the most reliable identification of HF. While  $D_{LNO}$  can replace  $D_{LCO}$  in some cardiopulmonary diseases [6], the synergy of using both for impairment of pulmonary diffusing capacity is crucial for HF.

Our analysis also highlighted that only about a third of the variance in combined z-scores directly correlates with HF identification. The factors causing the remaining two-thirds of the variance remain an enigma. Despite the strength of combined z-scores with HF, established screening tools (e.g. brain natriuretic peptide (BNP), N-terminal pro-BNP and echocardiography) are still better [36, 37]. Our intent was not to substitute these tools but to showcase how  $D_{LNO}$  z-scores could complement  $D_{LCO}$  z-scores under specific circumstances. Indeed, combining  $D_{LNO}$  and  $D_{LCO}$  offers a more comprehensive insight into the specific location of resistance in the oxygen diffusion pathway, from the alveoli to the interior of the red blood cell.

### Limitations

Several challenges were present in this study. First, the limited application of the NO–CO double diffusion method constrained the amount of available data. Second, the dataset mainly represents an adult White population, excluding children and diverse ethnic groups. The equations in Model 1 do not account for child-related data or specific device adjustments. Third, even though Model 1 performed the best among the 12 models examined, it might not be the most suitable in specific conditions like anaemia [34], polycythaemia or elevated carboxyhaemoglobin levels [35], such as in cases of CO poisoning. Under these conditions, only  $D_{LCO}$  is only influenced. Fourth, we chose not to adjust for haemoglobin in  $D_{LCO}$  due to feasibility and model fit concerns. Notably, ~20% of NYHA Class II HF patients display signs of anaemia ( $11.7 \pm 1.1$  g·dL<sup>-1</sup>) [38]. At this concentration and prevalence,  $D_{LCO}$  results likely remain stable. Lastly, the participants labelled as “non-HF” were not perfect controls. Approximately 20% of these control participants displayed restrictive or obstructive patterns despite normal heart function. The MCC would likely be higher if these controls had no lung obstruction or restriction.

### Conclusions

In conclusion, combined  $D_{LNO}+D_{LCO}$  z-score models outperform single z-score models for HF fit and the binary HF classification benefits from combined z-scores over individual  $D_{LCO}$  or  $D_{LNO}$  LLN cut-offs. However, only ~32% of HF variance predicting HF is explained by combined z-scores, discouraging HF screening use. Instead, the NO–CO technique is recommended for gas exchange assessment and improved binary classification in HF patients.

Provenance: Submitted article, peer reviewed.

Acknowledgements: The authors wish to thank the colleagues of P. Agostoni, specifically Elisabetta Salvioni and Anna Apostolo (Centro Cardiologico Monzino IRCCS, Milan, Italy), for obtaining the de-identified raw data from four of their studies for use here.

Data availability: The data that were used in this study will be available from G.S. Zavorsky or P. Agostini, but only if the researchers requesting the data will correctly cite this work, and will allow G.S. Zavorsky or P. Agostini to be co-authors in any subsequent publication that uses this data.

Conflict of interest: None declared.

### References

- 1 Van Iterson EH, Smith JR, Olson TP. Alveolar air and O<sub>2</sub> uptake during exercise in patients with heart failure. *J Card Fail* 2018; 24: 695–705.

- 2 Zavorsky GS, Christou NV, Kim DJ, *et al.* Preoperative gender differences in pulmonary gas exchange in morbidly obese subjects. *Obes Surg* 2008; 18: 1587–1598.
- 3 Moinard J, Guénard H. Determination of lung capillary blood volume and membrane diffusing capacity in patients with COLD using the NO–CO method. *Eur Respir J* 1990; 3: 318–322.
- 4 Zavorsky GS, Hsia CC, Hughes JM, *et al.* Standardisation and application of the single-breath determination of nitric oxide uptake in the lung. *Eur Respir J* 2017; 49: 1600962.
- 5 Giraud X, Le-Dong NN, Hogben K, *et al.* The measurement of DLNO and DLCO: a manufacturer’s perspective. *Respir Physiol Neurobiol* 2017; 241: 36–44.
- 6 Zavorsky GS, van der Lee I. Can the measurement of pulmonary diffusing capacity for nitric oxide replace the measurement of pulmonary diffusing capacity for carbon monoxide? *Respir Physiol Neurobiol* 2017; 241: 9–16.
- 7 Puri S, Baker BL, Dutka DP, *et al.* Reduced alveolar-capillary membrane diffusing capacity in chronic heart failure. Its pathophysiological relevance and relationship to exercise performance. *Circulation* 1995; 91: 2769–2774.
- 8 Magini A, Zavorsky GS, Apostolo A, *et al.* Week to week variability of pulmonary capillary blood volume and alveolar membrane diffusing capacity in patients with heart failure. *Respir Physiol Neurobiol* 2021; 290: 103679.
- 9 Magini A, Contini M, Spadafora E, *et al.* Variability in pulmonary diffusing capacity in heart failure. *Respir Physiol Neurobiol* 2020; 280: 103473.
- 10 Apostolo A, Paolillo S, Contini M, *et al.* Comprehensive effects of left ventricular assist device speed changes on alveolar gas exchange, sleep ventilatory pattern, and exercise performance. *J Heart Lung Transplant* 2018; 37: 1361–1371.
- 11 Magini A, Apostolo A, Salvioni E, *et al.* Alveolar–capillary membrane diffusion measurement by nitric oxide inhalation in heart failure. *Eur J Prev Cardiol* 2015; 22: 206–212.
- 12 Contini M, Spadafora E, Barbieri S, *et al.* Effects of beta2-receptor stimulation by indacaterol in chronic heart failure treated with selective or non-selective beta-blockers: a randomized trial. *Sci Rep* 2020; 10: 7101.
- 13 Iversen KK, Kjaergaard J, Akkan D, *et al.* Chronic obstructive pulmonary disease in patients admitted with heart failure. *J Intern Med* 2008; 264: 361–369.
- 14 Stanojevic S, Kaminsky DA, Miller MR, *et al.* ERS/ATS technical standard on interpretive strategies for routine lung function tests. *Eur Respir J* 2022; 60: 2101499.
- 15 Zavorsky GS, Cao J. Reference equations for pulmonary diffusing capacity using segmented regression show similar predictive accuracy as GAMLSS models. *BMJ Open Respir Res* 2022; 9: e001087.
- 16 Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 2020; 21: 6.
- 17 Apostolo A, Vignati C, Della Rocca M, *et al.* Why Levosimendan improves the clinical condition of patients with advanced heart failure: a holistic approach. *J Card Fail* 2022; 28: 509–514.
- 18 Munkholm M, Marott JL, Bjerre-Kristensen L, *et al.* Reference equations for pulmonary diffusing capacity of carbon monoxide and nitric oxide in adult Caucasians. *Eur Respir J* 2018; 52: 1500677.
- 19 Aguilaniu B, Maitre J, Glenet S, *et al.* European reference equations for CO and NO lung transfer. *Eur Respir J* 2008; 31: 1091–1097.
- 20 Zavorsky GS, Cao J, Murias JM. Reference values of pulmonary diffusing capacity for nitric oxide in an adult population. *Nitric Oxide* 2008; 18: 70–79.
- 21 van der Lee I, Zanen P, Stigter N, *et al.* Diffusing capacity for nitric oxide: reference values and dependence on alveolar volume. *Respir Med* 2007; 101: 1579–1584.
- 22 Thomas A, Hanel B, Marott JL, *et al.* The single-breath diffusing capacity of CO and NO in healthy children of European descent. *PLoS One* 2014; 9: e113177.
- 23 Bowerman C, Bhakta NR, Brazzale D, *et al.* A race-neutral approach to the interpretation of lung function measurements. *Am J Respir Crit Care Med* 2023; 207: 768–774.
- 24 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.
- 25 Raftery AE. Bayesian model selection in social research. *Sociol Methodol* 1995; 25: 111–163.
- 26 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc B* 1995; 57: 289–300.
- 27 Chicco D, Totsch N, Jurman G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min* 2021; 14: 13.
- 28 Boughorbel S, Jarray F, El-Anbari M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS One* 2017; 12: e0177678.
- 29 Diedenhofen B, Musch J. cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS One* 2015; 10: e0121945.
- 30 Zou GY. Toward using confidence intervals to compare correlations. *Psychol Methods* 2007; 12: 399–413.
- 31 Piccinni C, Antonazzo IC, Simonetti M, *et al.* The burden of chronic heart failure in primary care in Italy. *High Blood Press Cardiovasc Prev* 2017; 24: 171–178.

- 32 McGee S. Simplifying likelihood ratios. *J Gen Intern Med* 2002; 17: 646–649.
- 33 Wald NJ, Bestwick JP. The area under the ROC curve: is it a valid measure of screening performance? *J Med Screen* 2014; 21: 220.
- 34 van der Lee I, Zanen P, Biesma DH, *et al.* The effect of red cell transfusion on nitric oxide diffusing capacity. *Respiration* 2005; 72: 512–516.
- 35 Zavorsky GS. The rise in carboxyhemoglobin from repeated pulmonary diffusing capacity tests. *Respir Physiol Neurobiol* 2013; 186: 103–108.
- 36 Castiglione V, Aimò A, Vergaro G, *et al.* Biomarkers for the diagnosis and management of heart failure. *Heart Fail Rev* 2022; 27: 625–643.
- 37 Horodinschi RN, Bratu OG, Dediu GN, *et al.* Heart failure and chronic obstructive pulmonary disease: a review. *Acta Cardiol* 2020; 75: 97–104.
- 38 von Haehling S, van Veldhuisen DJ, Roughton M, *et al.* Anaemia among patients with heart failure and preserved or reduced ejection fraction: results from the SENIORS study. *Eur J Heart Fail* 2011; 13: 656–663.
- 39 Borland CD, Dunningham H, Bottrill F, *et al.* Significant blood resistance to nitric oxide transfer in the lung. *J Appl Physiol* 2010; 108: 1052–1060.