Original Research

Emphysema Detection in Smokers: Diffusing Capacity for Nitric Oxide Beats Diffusing Capacity of Carbon Monoxide-Based Models

Gerald S. Zavorsky, PhD, RRT[1] Roberto W. Dal-Negro, MD[2] Ivo van der Lee, MD, PhD[3] Alexandra M. Preisser, MD[4]

[1]Department of Physiology and Membrane Biology, University of California, Davis, Davis, California, United States

[2]CESFAR – National Center for Respiratory Pharmacoeconomics and Pharmacoepidemiology, Verona, Italy

[3]Department of Pulmonology, Spaarne gasthuis, Haarlem, The Netherlands

[4]Institute for Occupational and Maritime Medicine, University Medical Center Hamburg–Eppendorf, Hamburg, Germany

Address correspondence to:
Gerald S. Zavorsky, PhD RRT, RPFT, FACSM
Department of Physiology and Membrane Biology
University of California, Davis
Davis, CA
Email: gszavorsky@ucdavis.edu

Running Head: **Emphysema Detection: Diffusing Capacity for Nitric Oxide**

Keywords: pulmonary diffusing capacity; z-scores; logistic modeling; predictive accuracy, emphysema.

Abbreviations:

Funding Support: None

Date of Acceptance: October 29, 2025 | ***Published Online***: November 10, 2025

Citation: Zavorsky GS, Dal-Negro RW, van der Lee I, Preisser AM. Emphysema detection in smokers: diffusing capacity for nitric oxide beats diffusing capacity of carbon monoxide-based models. *Chronic Obstr Pulm Dis*. 2025; Published online November 10, 2025. https://doi.org/10.15326/jcopdf.2025.0645

This article has an online data supplement

Abstract

Background: Pulmonary diffusing capacity for nitric oxide (DLNO) remains underutilized despite potential advantages over carbon monoxide diffusing capacity (DLCO). We evaluated whether DLNO better detects emphysema than DLCO, spirometry, or lung volumes in smokers.

Methods: We performed an individual participant data meta-analysis of adult smokers (14–43 pack-years) with and without CT-defined emphysema using a standardized 10 ± 2 s double diffusion protocol. Variables were converted to z-scores. Prespecified models contrasted DLCO- versus DLNO-based approaches. Model selection used the Bayesian Information Criterion (BIC) and Leave-One-Out Information Criterion (LOOIC); discrimination used area under the ROC curve (AUROC) and Matthews Correlation Coefficient (MCC) with repeated cross-validation.

Results: After harmonization and quality control, 408 participants (85 emphysema, 323 controls) were analyzed. The lowest BIC (164.6) occurred for the three-predictor model with TLC, FEV_1, and DLNO z-scores, with an 88% probability of being superior to the next-lowest BIC model (168.5). Discrimination (AUROC 0.97, 95% CI 0.95–0.98) and classification (MCC 0.80, 95% CI 0.69–0.89) were high. Hierarchical partitioning showed unique contributions from FEV_1 z-scores (R^2=0.35) > DLNO z-scores (R^2=0.21) > TLC z-scores (R^2=0.11), totaling McFadden's R^2=0.663. Adding DLCO z-scores increased the total R^2 trivially (by 0.003) and contributed largely shared information with DLNO (variance inflation factors ≤ 4.5). Category-free reclassification and Youden-threshold analyses showed small but favorable gains; the case–control risk gap improved by up to ~5% when adding DLNO to a DLCO-based model.

Interpretation: A parsimonious z-score model comprising TLC, FEV_1, and DLNO z-scores provides excellent performance and stable rank superiority.

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

Introduction

Emphysema remains a major health concern in the United States, affecting an estimated 3.8 million individuals in 2018[1], with an age-adjusted mortality rate of approximately 9.5 per 100,000 adults in 2020 [2]. This disease destroys alveoli, compromises lung elasticity, increases air trapping, and leads to dyspnoea. Chronic inflammation and progressive alveolar-capillary membrane damage culminate in irreversible airflow limitation. Cigarette smoking is the primary risk factor driving lung injury[3-6]. Despite improvements in diagnostics, early-stage emphysema often goes unrecognized, delaying treatment and negatively influencing clinical and quality-of-life outcomes.

Standard pulmonary function tests—such as spirometry (measuring FEV_1, FVC, and their ratio) and pulmonary diffusing capacity for carbon monoxide (DLCO)—are useful for aiding diagnosis and monitoring. However, spirometry primarily assesses airway obstruction, while DLCO is affected by pulmonary capillary blood volume and haemoglobin concentration[7] and may lack sensitivity to early alveolar damage. Pulmonary diffusing capacity for nitric oxide (DLNO), first introduced in 1983-84 as abstracts[8,9], offers a more direct assessment of alveolar-capillary membrane function[10].

According to the Roughton-Forster model, gas transfer resistance comprises both membrane and red blood cell (RBC) components[11]; RBC interior resistance primarily limits CO uptake (DLCO)[10], whereas DLNO uptake occurs largely in the plasma boundary layer and via reactions at the RBC surface membrane[12]—not within the cell itself (as once thought)[13]. The time-based model of NO and CO absorption [14], suggests DLNO reflects "surface absorption" in the plasma space while DLCO represents "volume absorption" influenced by haematocrit [15], emphasizing diffusion and reaction kinetics [16].

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

Despite its potential, the NO-CO double diffusion technique remains uncommon in clinical practice, even 42 years[8,9] after its introduction. While commercially available devices exist for this measurement[17], their clinical adoption is limited by a lack of clinician awareness and the absence of United States Food and Drug Administration approval for any NO-CO device. As a result, use of this method has largely been confined to research settings by a small group of specialized investigators. Given the distinct technical advantages of measuring DLNO[18], it is important for manufacturers to pursue the necessary regulatory approvals so that DLNO can become as routine as DLCO in standard pulmonary function testing.

Studies show that the alveolar uptake for NO (KNO) has better sensitivity in detecting emphysema than DLCO[19], and that summed DLNO+DLCO z-scores outperform DLCO z-scores alone in model performance, predictive accuracy, and classification scores[7,20]. DLNO and KNO also correlate more closely with computed tomography (CT) markers of emphysema better than DLCO or KCO [21,22]. By enabling earlier and more accurate detection of emphysema, DLNO could facilitate timely, targeted interventions. Early identification of disease is clinically important, as patients with undiagnosed COPD face poorer outcomes and reduced quality of life[23] while early diagnosis and management can reduce healthcare utilization and improve quality of life[24]. Detecting emphysema before significant functional decline allows implementation of evidence-based interventions—including risk stratification, smoking cessation support, pulmonary rehabilitation, and individualized care—which slow disease progression, reduce emphysema progression in quitters[25] and improve exercise tolerance, reduce dyspnoea, and improve quality of life through pulmonary rehabilitation[23,26,27].

This study examines DLNO's diagnostic performance, accuracy, and classification ability in emphysema patients compared to DLCO, spirometry, and lung volumes. Using a large cohort—

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

predominantly smokers—from three hospital centres, we apply z-scores derived from established reference equations[28-33]. We hypothesized that DLNO and DLCO z-scores would outperform conventional metrics in diagnosing emphysema. If DLNO proves more accurate, it could be adopted routinely alongside DLCO. This adoption could facilitate earlier diagnosis, improve patient-centred outcomes, and stimulate the development and regulatory approval of accessible DLNO measurement equipment—overcoming current technological and logistical barriers[17].

Study Design and Methods

Study Design and Population

We conducted an individual participant data (IPD) meta-analysis pooling raw, participant-level data to harmonize variables, standardize analyses, and increase precision[34]. The pooled dataset included 496 White participants (mostly smokers; interquartile range 6–43 pack-years): 126 with computed-tomography (CT)–confirmed emphysema and 370 without, from four European hospital centres[19,35-37]. After harmonization, three centres were retained because they consistently used the simultaneous 10-s NO–CO protocol NO–CO testing[19,35,36]; all four source datasets remain available in a public repository[38]. All original studies had ethics approval; this de-identified secondary analysis did not require additional review.

Data Collection, Conversion, and Quality Control

Pulmonary function tests (PFTs) followed ATS/ERS guidelines[39-41], measuring DLNO, DLCO, alveolar volume (VA), KCO, and KNO with a 10 ± 2-second breath-hold time (denoted $DLNO_{10s}$, $DLCO_{10s}$, VA_{10s}, KCO_{10s}, and KNO_{10s}). Lung function variables were converted to z-scores using Global Lung Function Initiative (GLI) reference equations for spirometry[29], lung volumes[28], and $DLCO_{10s}$[30,31], adjusting for age, sex, and height. For the NO-CO double diffusion technique, z-scores were derived from reference equations developed with 10-s breath-hold manoeuvres[32] and

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

from equations that account for between device variability[33,42]. Because available DLNO reference equations were derived in White cohorts[32,33] and genetic ancestry influences DLNO[43-45], analyses were restricted to White participants.

Study-level quality was graded across nine items: inclusion of COPD and non-COPD participants; pack-years; radiologist-adjudicated CT emphysema (% volume); smoking history; mMRC dyspnoea; sex; height; weight (coded "not provided" if imputed); and technical quality control. Technical failure criteria were breath-hold outside 8–12 s; VA/TLC > 1.0; FEV1/FVC ≥ 1.0; RV/TLC < 0.20; or inspired-volume/FVC (IV/FVC) < 0.85. Studies with <5% failures met technical standards. The RV/TLC < 0.20 rule excluded physiologically implausible values. A summary score (0–9) tallied the eight availability items plus the quality control flag (**Table S1**). R packages are listed in **Tables S2–S3**.

Model discovery and comparator definition (post-selection)

We assembled 34 candidate logistic models from clinically plausible and data-driven combinations of z-scores (FEV_1, FVC, FEV_1/FVC, TLC, RV/TLC, VA, $DLCO_{10s}$, KCO_{10s}, $DLNO_{10s}$, KNO_{10s}). To encourage parsimony, we screened with east absolute shrinkage and selection operator (LASSO) and ranked all candidates using Bayesian Information Criterion (BIC) and Pareto-smoothed importance-sampling Leave-One-Out Information Criterion (PSIS-LOOIC) on the analysis dataset (lower values indicate superior expected out-of-sample fit and parsimony). Guided by this screening, we defined three focal comparators for all downstream evaluation:

- **Model A:** TLC z-scores + FEV_1 z-scores + $DLCO_{10s}$ z-scores
- **Model B:** Model A + $DLNO_{10s}$ z-scores
- **Model C:** TLC z-scores + FEV_1 z-scores + $DLNO_{10s}$ z-scores

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

TLC, FEV_1 and $DLCO_{10s}$ z-score were fitted using GLI equations[28-31] and DLNO z-scores were fitted using GAMLSS reference equations of Zavorsky & Cao (2022)[33].

This workflow is data-adaptive/post-selection: information-criterion screening of individual/summed predictors informed LASSO model building, and all candidates were ultimately compared on the same information criterion scale before selecting Models A–C.

Objectives, Endpoints, and Hypotheses

Based on the results above, a primary objective was to test whether the parsimonious three-predictor $DLNO_{10s}$ model (Model C) is non-inferior to the analogous $DLCO_{10s}$ model (Model A) for detecting CT-defined emphysema in adult smokers, compared to smokers without emphysema, while achieving better parsimony/generalizability. A key secondary objective was to assess whether adding $DLNO_{10s}$ to Model A (which is Model B) or expanding to higher-dimension variants yields clinically meaningful gains over Model C after accounting for complexity.

The primary endpoint was Out-of-fold (OOF) Matthews correlation coefficient (MCC) on held-out folds, using thresholds learned via Youden's J in training and applied to the paired test fold (MCC family under Benjamini–Hochberg control for prespecified contrasts). Another endpoint was the BIC and PSIS-LOOIC computed on the analysis dataset to quantify parsimony/generalization.

Secondary endpoints were test-fold area under the ROC curve (AUROC, threshold-free discrimination), decision curve analysis net benefit across thresholds 0–0.25, threshold based metrics (accuracy, balanced accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, Kappa statistic, false positive rate (FPR), False Negative Rate (FNR), False Discovery Rate (FDR), positive and negative likelihood (LR±), diagnostic odds ratio (DOR), computed out-of-fold at the Youden threshold. For the exploratory

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

endpoints**,** Category-free NRI, IDI, calibration intercept/slope with bootstrap CIs, PCA loadings/variance explained, hierarchical partitioning of McFadden's R^2.

The primary hypothesis was that Model C is non-inferior to Model A on out-of-fold (OOF) MCC and AUROC and shows lower BIC and/or PSIS-LOOIC (i.e., superior parsimony/generalization). The secondary hypotheses were that Model B and higher-numbered variable models do not provide clinically meaningful gains in OOF performance or decision-curve net benefit over Model C once complexity penalties are considered.

Decision rules, inference, and multiplicity

For each contrast we estimated ΔMCC and ΔAUROC using paired, fold-level bootstrap (10,000 resamples); two-sided bootstrap p-values were computed, with BH control applied only within the MCC family across contrasts. Parsimony/generalization superiority was judged by BIC/PSIS-LOOIC (with $|\Delta| \gtrsim 2$ typically indicating a small but non-trivial improvement). Secondary and exploratory endpoints were interpreted descriptively without alpha allocation.

Model fitting, selection, and internal validation

Logistic models were initially fit as generalized linear mixed models (GLMMs) with a study-level random intercept; model fit, and diagnostics favored generalized linear models (GLMs) without the random intercept for 8/34 models, which were used for primary analyses (**Table S10**). Primary selection used BIC; expected out-of-sample performance used PSIS-LOOIC. Discrimination used AUROC with 95% CIs (BH-controlled where applicable). We applied stratified 10-fold cross-validation × 1,000 repeats with within-fold standardization; thresholds were learned in training and applied to held-out folds. OOF probabilities were averaged across repeats; each model's global threshold was the mean of fold-level Youden-J thresholds. Uncertainty used paired, fold-level bootstraps; decision-curve analysis assessed standardized net benefit.

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

Ancillary and rank-based analyses

Principal component analysis (PCA) assessed latent structure/collinearity. Hierarchical partitioning decomposed McFadden's R^2 into unique and joint components for FEV_1, TLC, $DLCO_{10s}$, and $DLNO_{10s}$ z-scores Reclassification (category-free and threshold-based Net Reclassification Improvement, NRI; Integrated Discrimination Index, IDI) compared models B−A and models C−A. Calibration used logistic recalibration (intercept α, slope β) with bootstrap CIs and calibration plots. Sensitivity analyses included alternative operating points and leave-one-centre-out checks.

To synthesize signals across metrics, we performed rank-based comparisons of MCC, AUROC, BIC, and PSIS-LOOIC across the top 10 models.

To assess whether model comparisons were robust to the weighting of evaluation metrics, we prespecified six ranking schemes (Equal Weight, Weighted Average, Generalization-Emphasis, Discrimination-Emphasis, BIC-Omitted, AUROC-Emphasis). For each comparator we computed the rank difference (comparator – the best ranked model) within each scheme and then the mean Δ rank across schemes. To reflect sensitivity to weighting choice, we obtained 95% bootstrap intervals by resampling schemes ($n = 6$) with replacement ($B = 10{,}000$) and recomputing the across-scheme mean. We did not resample the derived "Average Rank" column; instead, intervals were based on the six scheme-specific differences. Intervals entirely > 0 indicate the comparator is consistently ranked worse than MODEL C across the prespecified schemes

Statistical software

Analyses were performed using RStudio (2025.09.0), Build 387, with R (version 4.4.2). Two-sided $p < 0.05$ was considered statistically significant. Additional information on the statistical analyses can be found in the online supplementary material in the "Supplementary Methods" section, and

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

Tables S1 to S9.

Results

Participant characteristics

Of 496 eligible individuals, 408 (85 emphysema, 323 controls) met harmonization and quality control criteria (See **Table S1, Figure S1**, from online supplement). Subject characteristics are presented in **Table 1**. The median breath-hold time for diffusing capacity was 10 seconds (IQR, 9.2–10). The median $DLNO_{10s}/DLCO_{10s}$ ratio was 4.43 (IQR 4.09–4.88) for emphysema and 4.50 (IQR 4.22–4.84) for non-emphysema ($p = 0.226$). Nearly all variables in **Table 1** were statistically different between the two groups except for pack-years of smoking and the proportion of subjects with pulmonary restriction. The pooled data demonstrate that 71% of the variance in $DLNO_{10s}$ z-scores is shared with $DLCO_{10}$s z-scores (**Figure S2**).

Information criteria and generalization

The top 17 models are ranked by their difference in BIC and LOOIC relative to the top-performing model (**Figure 1**). The bottom 18 models are presented in **Figure S3.** Absolute values for BIC and LOOIC for all 34 models are presented in **Table S10**. A compact cluster—including the three-predictor DLNO model (Model C)—lies near Δ0 for both BIC and LOOIC. Single-index models (e.g., VA alone; DLCO alone) are markedly inferior to the multiple variable predictor models. As such, the model with the lowest BIC is presented in **Table 2** with the AUROC and its precision-recall curves are presented in **Figure S4**.

Discrimination (AUROC) and MCC at the Youden Cut-Point

Figure 2 displays MCC (bars) and AUROC (points with 95% CIs) ranked by MCC at the Youden threshold for the top 17 models. The bottom 17 models are presented in **Figure S5**. Three- to six-predictor DLNO models achieve AUROC ≈0.96–0.97 with MCC ≈0.8. Model A (DLCO-based) is

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

modestly worse, and Model B (A + DLNO) trades sensitivity and specificity without clear net gain (for fold-averaged metrics and CIs).

Principal Component Analysis (PCA) and hierarchical partitioning.

PCA revealed Principal Component 1 (PC1) (gas transfer: DLNO/DLCO), PC2 (hyperinflation: TLC/VA), and PC3 (obstruction/air-trapping: FEV_1/FVC, RV/TLC) (**Tables S11-S13**). PC4 was not found to add any benefit (**Table S14, Figure S6**). Replacing z-score predictors with PC1–PC3 did not improve discrimination or generalization (ΔAUROC ≈ 0–0.01; ΔLOOIC/ΔBIC < 2); hence we favour Model C (TLC + FEV_1 + DLNO z-scores) specification for interpretability. Hierarchical partitioning ranked unique contributions as the FEV_1 z-scores > $DLNO_{10s}$ z-scores > TLC z-scores > $DLCO_{10s}$ z-scores (**Tables S15–S16**).

Classification, reclassification and decision analysis

Model B z-scores (FEV_1, TLC, $DLNO_{10s}$, $DLCO_{10s}$ z-scores) or Model C (z-scores FEV_1, TLC, $DLNO_{10s}$, z-scores) compared to Model A (FEV_1, TLC, $DLCO_{10s}$ z-scores) demonstrated no real difference in 17 metrics different when considering the 95% CI (**Tables S17-S18**). At Youden-optimized thresholds, the overall net improvement in reclassification when $DLNO_{10s}$ is added to the model A was not significant. However, the average predicted risk-gap between predicting smokers with and without emphysema improved by as much as 5% when $DLNO_{10s}$ is added to Model A (**Table S19**).

At category-free reclassification, there was a 34% overall net improvement in reclassification (95% CI = –12 to 96%) when $DLNO_{10s}$ was added to Model A. Simply, this means that, compared with the old model, the new model moved people in the right direction (up for true cases, down for true non-cases) 34 percentage points more often than it moved them in the wrong direction. Moreover, the average predicted risk-gap between predicting smokers with and without

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

emphysema improved by as much as 5% when $DLNO_{10s}$ was added to Model A (**Table S20**).

Decision-curve analysis using out-of-fold predictions (**Figure S7**) showed all three models (Models A, B, C) delivered positive net benefit across threshold probabilities 0–0.25, exceeding Treat None and—apart from the very lowest thresholds—exceeding Treat All. The Treat-All curve crossed zero at ~0.21 (cohort prevalence), while all model curves remained positive. The three curves overlapped closely; the 4-predictor model (Model B) offered no discernible advantage over either Model A or C. Absolute net benefit was ~0.16–0.20, i.e., ~16–20 more correctly flagged smokers with emphysema per 100 smokers than doing nothing, at thresholds 0–0.25

Model rankings

Across schemes, eight out of nine comparators ranked significantly worse than the best ranked model (MODEL C, Δ rank > 0; 95% CI is > 0). The five-predictor DLNO model was the only top 10 model had a ranking that was not different to MODEL C. (Δ=+0.22 [−0.09, 0.57]). As such, Model C (TLC z-scores, FEV_1 z-scores, & $DLNO_{10s}$ z-scores) offered near-top performance yet delivered comparable discrimination/generalization with fewer predictors with greater parsimony than any other model. Thus, Model C is the best choice (**Figure 3**).

Discussion

Our multi-centre individual participant data meta-analysis demonstrates that compact models incorporating DLNO offer robust, high-quality classification of emphysema compared to smokers without emphysema. In the three-predictor DLNO z-score model (Model C: TLC z-scores, FEV_1 z-scores, & $DLNO_{10s}$ z-scores), the TLC z-scores and FEV_1 z-scores were obtained using GLI equation for white subjects[28,29], while the best fitting DLNO z-scores were obtained by using the DLNO GAMLSS equation from Zavorsky & Cao (2022)[33]. The three-predictor DLNO z-score model (MODEL C) occupied a consistently superior or co-superior position across BIC/LOOIC

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

(**Figure 1**) and AUROC/MCC (**Figure 2**) and maintaining top ranks across weighting scenarios (**Figure 3**).

The incremental advantage in Model C (TLC z-scores, FEV_1 z-scores, & $DLNO_{10s}$ z-scores) versus a DLCO-based analogue (Model A) is modest, but consistency across metrics and resampling supports a genuine performance edge for Model C. Reclassification indices were small (**Tables S19–S20**), suggesting DLNO's benefits manifest more as improved overall performance and calibration (**Figure S7**) than as wholesale shifts in categorical assignment at a single threshold.

The pooled data demonstrate that 71% of the variance in $DLNO_{10s}$ z-scores is shared with $DLCO_{10}$s z-scores (**Figure S2)** displaying substantial collinearity; yet the residual ~29% 'unique' variance is not necessarily predictive for emphysema. Nevertheless, adding DLCO z-scores to model C, model barely changes performance (McFadden's $R^2 = 0.663 \rightarrow 0.666$; and BIC worsens), and AUROC gains were negligible. That's strong evidence that the "unique" DLCO portion doesn't add meaningful predictive signal for emphysema beyond the z-score model of TLC + FEV_1 + $DLNO_{10s}$ z-scores.

Physiologic plausibility of the results is strong. One view is that NO uptake primarily reflects membrane resistance, whereas CO uptake reflects the resistance that occurs within the red cell membrane (See **Figure 1** of Zavorsky *et al* (2025))[46]. The finding that adding DLCO z-scores to a DLNO z-score-based model contributes little, coheres with early membrane-dominant injury in emphysema. PCA structure (**Tables S11–S14**) and hierarchical partitioning (**Table S15–S16**) further support construct validity by aligning dominant components with expected physiologic domains.

In our IPD meta-analysis, Model C had similar discrimination and classification of the

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

three-predictor model using TLC, FEV_1, and $DLCO_{10s}$ z-scores, while achieving lower BIC and LOOIC—indicating a more parsimonious specification with better expected out-of-sample fit. Practically, this suggests $DLNO_{10s}$ can substitute for $DLCO_{10s}$ as the gas-transfer input in emphysema when simultaneous NO–CO testing is available, and quality control is assured. We are not advocating $DLNO_{10s}$ in isolation from spirometry and lung volumes across all pathologies – summed DLNO+DLCO z-score models can outperform either measure alone in other conditions[20,46]. Rather, among transfer measures, $DLNO_{10s}$ alone (without $DLCO_{10s}$) appears sufficient in this emphysema-screening context. $DLCO_{10s}$ still has broader clinical roles (e.g., interstitial lung[47,48] and pulmonary vascular disease[49]), so laboratories without NO–CO capability can continue to rely on $DLCO_{10s}$, whereas centres with NO–CO may reasonably prioritize $DLNO_{10s}$ – or $DLNO_{5s}$ where validated – in parsimonious models. Decision-curve analyses indicate potential utility at low thresholds common to screening/case-finding (**Figure S5**).

Despite promising diagnostic performance, DLNO remains underutilized since its introduction in 1983-84 [8,9], largely due to limited awareness, regulatory hurdles, and the high cost of sensitive NO analysers. DLNO testing is available as an add-on to DLCO on at least one commercial platform (e.g., MGC Diagnostics), requiring only an NO sensor, cylinder, and minimal training. These systems typically use low-cost electrochemical sensors (e.g., 7NT CiTiceL®), but their limited range (0–100 ppm) and slow response (~15 s) make them unsuitable for longer breath-holds, where exhaled NO concentrations drop below detectable thresholds. Chemiluminescence remains the gold standard for NO detection due to its rapid response (<1 s) and wide dynamic range (1 ppb–100 ppm), but devices like the CLD 855 Yh (Eco Physics) cost ~$35,000 USD. Mass production could lower this to $10,000–$21,000 per unit if adopted across the estimated 2,000–5,000 DLCO-equipped sites in the U.S.[50].

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

Moreover, the similar median $DLNO_{10s}/DLCO_{10s}$ ratios between emphysema and non-emphysema groups (4.43 vs. 4.50) suggest that raw ratios lack discriminatory power—reinforcing the need for z-score standardization in diagnostic models. Future work should prioritize affordable, regulatory-approved $DLNO_{10s}$ systems and longitudinal studies to assess their clinical utility.

Limitations

Our pooled analytic sample comprised primarily White adults who were current or former smokers from European centres using a 10 ± 2 s simultaneous NO–CO protocol; only ~14% were never-smokers. Accordingly, generalizability to never-smokers, other ancestral groups, paediatric or very elderly populations, and to laboratories using different devices or protocols may be limited. CT-confirmed emphysema improved case specificity but may miss early/subclinical or airway-predominant disease. Although we harmonized variables across sites, applied uniform quality control, and used repeated cross-validation with leave-one-centre-out checks and calibration assessment, we lacked an independent external cohort; thus, performance estimates and optimal thresholds may shift in other settings. Furthermore, it is known that genetic ancestry affects DLNO[43-45], so standardization relied on DLNO reference equations and GLI equations for TLC and DLCO derived largely from White cohorts, can constrain calibration and promote bias. Reference equations developed for specific genetic ancestries (versus only white) are needed to maintain precision. We also lacked uniform data on comorbidities, medications, and socioeconomic context, limiting adjustment for potential confounding and spectrum effects. Finally, the dataset did not include longitudinal outcomes (e.g., exacerbations, CT progression, mortality), so we evaluated discrimination and reclassification rather than long-term clinical impact; strict and complete-case analysis may also introduce selection bias, and residual site/device effects may persist despite adjustment.

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

Broader Implications for DLNO Adoption in Clinical Practice

The demonstrated superiority of $DLNO_{10s}$ z-scores in predicting and classifying smokers with emphysema compared to smokers without emphysema offers a compelling case for DLNO inclusion in routine pulmonary function testing. Despite the current barriers to widespread adoption—such as the high cost of NO analysers and regulatory hurdles—this study provides strong evidence for the clinical value of DLNO. Its ability to detect subtle changes in alveolar-capillary membrane functionality positions it as a critical tool for early COPD diagnosis, particularly in high-risk populations such as smokers.

Future efforts should focus on developing cost-effective, FDA approved devices for DLNO measurement to facilitate its integration into pulmonary function laboratories. Additionally, prospective studies validating these findings in diverse populations are necessary further to establish DLNO's role as a diagnostic benchmark.

Conclusion

The three-predictor model incorporating $DLNO_{10s}$ z-scores with TLC z-scores and FEV_1 z-scores offers superior model performance, predictive accuracy, and classification for emphysema detection compared to $DLCO_{10s}$-based models. These findings advocate integrating $DLNO_{10s}$ into routine clinical practice, potentially improving early diagnosis and patient outcomes in emphysema management.

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

Author contributions

G.S.Z: Conceptual design of the work; data curation, formal analysis; methodology, software, validation. visualization, writing – original draft preparation and final draft completion, writing – review and editing.

R.W.D-N: Resources (provision of study subjects), investigation (data collection), writing – review and editing.

I.v.d.L: Resources (provision of study subjects), investigation (data collection), writing – review and editing.

A.M.P. Resources (provision of study subjects), investigation (data collection), writing – review and editing.

Funding Details

No funding was provided for this study

Conflicts of interest statement / Competing interests' statement

Gerald S. Zavorsky is a GLI Network member. The GLI Network has published reference equations for spirometry, DLCO, and static lung volumes using GAMLSS models. Gerald S. Zavorsky is the current co-chair of the ERS Task Force on the interpretation of DLNO. The remaining authors declared that no conflict of interests exists.

Data availability statement

The data supporting this study's findings are accessible on Mendeley Data [38], a cloud-based repository for research data.

Ethics approval statement

Ethical approval was obtained for all original studies, but this secondary analysis of de-identified data did not require IRB approval.

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

References

1. Trends in Emphysema by Race/Ethnicity, Sex, and Age (https://www.lung.org/research/trends-in-lung-disease/copd-trends-brief/data-tables/trends-in-emphysema-by-race-sex-age). The American Lung Association. Accessed February 25, 2025.
2. Saeed H, Arshad MK, Shahnoor S, et al. Temporal trends, gender, and ethnoracial disparities in mortality from pulmonary emphysema: A retrospective nationwide analysis. *Medicine (Baltimore)*. Dec 27 2024;103(52):e41032. doi:10.1097/MD.0000000000041032
3. Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. *J Am Med Inform Assoc*. May–Jun 2009;16(3):371–9. doi:10.1197/jamia.M2846
4. Adatia A, Wahab M, Shahid I, Moinuddin A, Killian KJ, Satia I. Effects of cigarette smoke exposure on pulmonary physiology, muscle strength and exercise capacity in a retrospective cohort with 30,000 subjects. *PLoS One*. 2021;16(6):e0250957. doi:10.1371/journal.pone.0250957
5. Hoffmann RF, Zarrintan S, Brandenburg SM, et al. Prolonged cigarette smoke exposure alters mitochondrial structure and function in airway epithelial cells. *Respir Res*. Oct 2 2013;14(1):97. doi:10.1186/1465-9921-14-97
6. Calverley PM, Walker P. Chronic obstructive pulmonary disease. *Lancet*. Sep 27 2003;362(9389):1053–61. doi:10.1016/s0140-6736(03)14416-9
7. Zavorsky GS, Barisione G, Gille T, et al. Enhanced detection of patients with previous COVID-19: Superiority of the double diffusion technique. *BMJ Open Respir Res*. 2025;http://dx.doi.org/10.1136/bmjresp-2024-00256110.1136/bmjresp-2024-002561
8. Borland C, Chamberlain A, Higenbottam T. The fate of inhaled nitric oxide [abstract]. *Clin Sci (Lond)*. 1983;65(3):37P. doi:10.1042/cs065037P
9. Borland C, Cracknell N, Higenbottam T. Is the measurement of "DLNO" a true measure of membrane diffusing capacity? [abstract]. *Clin Sci (Lond)*. 1984;67 (S9):41P. doi:10.1042/cs067041Pa

10. 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*. Feb 2017;49(2):1600962. doi:10.1183/13993003.00962-2016
11. Roughton FJ, Forster RE. Relative importance of diffusion and chemical reaction rates in determining rate of exchange of gases in the human lung, with special reference to true diffusing capacity of pulmonary membrane and volume of blood in the lung capillaries. *J Appl Physiol*. Sep 1957;11(2):290–302.
12. Borland C, Patel R. Comparing in vitro nitric oxide blood uptake to its pulmonary diffusing capacity. *Nitric Oxide*. 2024/02/01/ 2024;143:29–43. doi:https://doi.org/10.1016/j.niox.2023.11.006
13. Borland C, Patel R. Comparing in vitro nitric oxide blood uptake to its pulmonary diffusing capacity. *Nitric Oxide*. Dec 21 2023;143:29–43. doi:10.1016/j.niox.2023.11.006
14. Kang MY, Sapoval B. Time-based understanding of DLCO and DLNO. *Respir Physiol Neurobiol*. May 2016;225:48–59. doi:10.1016/j.resp.2016.01.008
15. Kang MY, Grebenkov D, Guenard H, Katz I, Sapoval B. The Roughton-Forster equation for DLCO and DLNO re-examined. *Respir Physiol Neurobiol*. Jul 2017;241:62–71. doi:10.1016/j.resp.2016.12.014
16. Borland CDR, Hughes JMB. Lung Diffusing Capacities (D(L)) for Nitric Oxide (NO) and Carbon Monoxide (CO): The Evolving Story. [Corrigendum in Compr Physiol. 2021 Jan 24;11(1):1371: https://doi.org/10.1002/cphy.cv11i01corr]. *Compr Physiol*. Dec 18 2020;10(1):73–97. doi:10.1002/cphy.c190001
17. Giraud X, Le-Dong NN, Hogben K, Martinot JB. The measurement of DLNO and DLCO: A manufacturer's perspective. *Respir Physiol Neurobiol*. Jul 2017;241:36–44. doi:10.1016/j.resp.2017.02.006
18. 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*. Jul 2017;241:9–16. doi:10.1016/j.resp.2016.11.008
19. van der Lee I, Gietema HA, Zanen P, et al. Nitric oxide diffusing capacity versus spirometry in the early diagnosis of emphysema in smokers. *Respir Med*. Dec 2009;103(12):1892–7. doi:10.1016/j.rmed.2009.06.005

20. Zavorsky GS, Agostoni P. Two is better than one: the double diffusion technique in classifying heart failure. *ERJ Open Res*. Jan 2024;10(1)doi:10.1183/23120541.00644-2023
21. Dressel H, Filser L, Fischer R, et al. Lung diffusing capacity for nitric oxide and carbon monoxide in relation to morphological changes as assessed by computed tomography in patients with cystic fibrosis. *BMC Pulm Med*. 2009;9(30)doi:10.1186/1471-2466-9-30
22. Barisione G, Brusasco C, Garlaschi A, Baroffio M, Brusasco V. Lung diffusing capacity for nitric oxide as a marker of fibrotic changes in idiopathic interstitial pneumonias. *J Appl Physiol*. May 1 2016;120(9):1029–38. doi:10.1152/japplphysiol.00964.2015
23. Lin CH, Cheng SL, Chen CZ, Chen CH, Lin SH, Wang HC. Current Progress of COPD Early Detection: Key Points and Novel Strategies. *Int J Chron Obstruct Pulmon Dis*. 2023;18:1511–1524. doi:10.2147/COPD.S413969
24. Aaron SD, Montes de Oca M, Celli B, et al. Early Diagnosis and Treatment of Chronic Obstructive Pulmonary Disease: The Costs and Benefits of Case Finding. *Am J Respir Crit Care Med*. Apr 15 2024;209(8):928–937. doi:10.1164/rccm.202311-2120PP
25. Postolache P, Nemes RM, Petrescu O, Merisanu IO. Smoking cessation, pulmonary rehabilitation and quality of life at smokers with COPD. *Rev Med Chir Soc Med Nat Iasi*. Jan–Mar 2015;119(1):77–80.
26. Rochester CL, Alison JA, Carlin B, et al. Pulmonary Rehabilitation for Adults with Chronic Respiratory Disease: An Official American Thoracic Society Clinical Practice Guideline. *Am J Respir Crit Care Med*. Aug 15 2023;208(4):e7–e26. doi:10.1164/rccm.202306-1066ST
27. Mendes Xavier D, Lanza Galvao E, Aliane Fonseca A, de Souza GM, Pereira Lima V. Effects of Home-Based Pulmonary Rehabilitation on Dyspnea, Exercise Capacity, Quality of Life and Impact of the Disease in COPD Patients: A Systematic Review. *COPD*. Dec 2022;19(1):18–46. doi:10.1080/15412555.2021.2020234
28. Hall GL, Filipow N, Ruppel G, et al. Official ERS technical standard: Global Lung Function Initiative reference values for static lung volumes in individuals of European ancestry. *Eur Respir J*. Mar 2021;57(3)doi:10.1183/13993003.00289-2020

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

29. 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*. Dec 2012;40(6):1324–43. doi:10.1183/09031936.00080312

30. 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*. Sep 2017;50(3)doi:10.1183/13993003.00010-2017

31. Stanojevic S, Graham BL, Cooper BG, et al. [Corrigendum] Official ERS technical standards: Global Lung Function Initiative reference values for the carbon monoxide transfer factor for Caucasians. *Eur Respir J*. 2020;56:1700010.

32. van der Lee I, Zanen P, Stigter N, van den Bosch JM, Lammers JW. Diffusing capacity for nitric oxide: reference values and dependence on alveolar volume. *Respir Med*. Jul 2007;101(7):1579–84. doi:10.1016/j.rmed.2006.12.001

33. Zavorsky GS, Cao J. Reference equations for pulmonary diffusing capacity using segmented regression show similar predictive accuracy as GAMLSS models. *BMJ Open Respir Res*. Feb 2022;9(1)doi:10.1136/bmjresp-2021-001087

34. Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. *Stat Med*. Oct 15 1995;14(19):2057–79. doi:10.1002/sim.4780141902

35. 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*. Mar 1990;3(3):318–22.

36. Diener L, Herold R, Harth V, Preisser A. DLNO/DLCO in different patient collectives; influencing factors have different effects on DLNO and DLCO. From the 61st Congress of the German Society for Pneumology and Respiratory Medicine. [Abstract]. *Pneumologie*. Apr 2021;75:S3–S3. doi:10.1055/s-0041-1723272

37. Dal Negro RW, Turco P, Povero M. Single-Breath Simultaneous Measurement of DL(NO) and DL(CO) as Predictor of the Emphysema Component in COPD - A Retrospective Observational Study. *Int J Chron Obstruct Pulmon Dis*. 2024;19:2123–2133. doi:10.2147/COPD.S467138

38. Zavorsky GS, Dal-Negro R, van der Lee I, Preisser A. Data from: “Nitric oxide uptake in the lungs in smokers with emphysema ”, Mendeley Data, V2, doi: 10.17632/g4jp8wd6f6.2. 2025. Deposited Febrary 26, 2025.
39. Graham BL, Steenbruggen I, Miller MR, et al. Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. *Am J Respir Crit Care Med*. Oct 15 2019;200(8):e70–e88. doi:10.1164/rccm.201908-1590ST
40. Wanger J, Clausen JL, Coates A, et al. Standardisation of the measurement of lung volumes. *Eur Respir J*. Sep 2005;26(3):511–22. doi:10.1183/09031936.05.00035005
41. Macintyre N, Crapo RO, Viegi G, et al. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *Eur Respir J*. Oct 2005;26(4):720–35. doi:10.1183/09031936.05.00034905
42. Radtke T, de Groot Q, Haile SR, Maggi M, Hsia CCW, Dressel H. Lung diffusing capacity for nitric oxide measured by two commercial devices: a randomised crossover comparison in healthy adults. *ERJ Open Res*. Jul 2021;7(3)doi:10.1183/23120541.00193-2021
43. Zavorsky GS, Almamary AS, Alqahtani MK, Shan SHS, Gardenhire DS. The need for race-specific reference equations for pulmonary diffusing capacity for nitric oxide. *BMC Pulm Med*. Jul 13 2021;21(1):232. doi:10.1186/s12890-021-01591-7
44. Gochicoa-Rangel L, De-Los-Santos-Martinez A, Reyes-Garcia A, et al. Reference equations for DLNO and DLCO in Mexican Hispanics: influence of altitude and race. *BMJ Open Respir Res*. Oct 14 2024;11(1)doi:10.1136/bmjresp-2024-002341
45. Simaga B, Forton K, Motoji Y, Naeije R, Faoro V. Lung diffusing capacity in sub-Saharan Africans versus European Caucasians. *Respir Physiol Neurobiol*. Jul 2017;241:23–27. doi:10.1016/j.resp.2017.01.003
46. Zavorsky GS, Barisione G, Gille T, et al. Enhanced detection of patients with previous COVID-19: superiority of the double diffusion technique. *BMJ Open Respir Res*. Aug 25 2025;12(1)doi:10.1136/bmjresp-2024-002561
47. Alhamad EH, Cal JG. Predictors of mortality in interstitial lung disease patients without pulmonary hypertension. *Ann Thorac Med*. Oct–Dec 2020;15(4):238–243. doi:10.4103/atm.ATM_438_20

48. Wong AW, Ryerson CJ, Guler SA. Progression of fibrosing interstitial lung disease. *Respir Res*. Jan 29 2020;21(1):32. doi:10.1186/s12931-020-1296-3
49. Stadler S, Mergenthaler N, Lange TJ. The prognostic value of DLCO and pulmonary blood flow in patients with pulmonary hypertension. *Pulm Circ*. Oct–Dec 2019;9(4):2045894019894531. doi:10.1177/2045894019894531
50. Medical M. Understanding medical device manufacturing costs: key factors influencing medical device manufacturing costs. https://www.meridian-medical.com/key-factors-influencing-medical-device-manufacturing-costs/. September 21st, 2025, 2025. Accessed September 21, 2025, 2024.

Table 1. Subject characteristics.

Variable	Smokers with no-emphysema (n=323)	Smokers with emphysema (n=85)
Number of Males and Females	306M, 17 F (95% M)	78M, 7 F (92% M)
Pack-years of smoking (Interquartile Range)	8 to 44	22 to 42
Age (years)	58 (11)	61 (10)*
Height (cm)	177 (8)	174 (9)*
Weight (kg)	80.5 (9.2)	81.9 (14.3)
BMI (kg/m^2)	25.6 (2.5)	27.0 (3.9) *
Hb (g/dL)	14.6 (0.7)	15.1 (1.3) *
FEV_1 (L)	3.54 (0.54)	2.15 (0.73)*
FVC (L)	4.29 (0.66)	3.38 (0.73) *
TLC (L)	7.44 (0.95)	7.66 (1.38)
FEV_1/FVC	0.83 (0.05)	0.63 (0.13)*
FEV_1 z-scores (GLI equations)	–0.15 (0.69)	–2.26 (1.29)*
FVC z-scores (GLI equations)	–0.53 (0.64)	–1.46 (1.11)*
FEV_1/FVC ratio z-scores (GLI equations)	0.69 (0.79)	–1.79 (1.45)*
TLC z-scores (GLI equations)	0.35 (0.68)	0.97 (1.17)*
RV/TLC ratio z-scores (GLI equations)	1.16 (0.59)	2.09 (0.89)*
$DLCO_{10s}$ z-scores (GLI equations)	–0.24 (1.13)	–2.23 (1.54)*
VA_{10s} z-scores (GLI equations)	– 0.13 (1.09)	–0.66 (1.68)*
$DLNO_{10s}$ z-scores (SLR, Zavorsky & Cao (2022))	–0.75 (1.05)	–2.34 (1.27)*
Combined $DLNO_{10s}$+ $DLCO_{10s}$ z-scores (SLR, Zavorsky& Cao (2022))	–1.52 (1.92)	– 4.71 (2.38)*
Number and percentage of individuals with obstruction (FEV_1/FVC < LLN_5 and FVC > LLN_5)	1 (0.3%)	15 (18%)*
Number and percentage of individuals with mixed disorders (FEV_1/FVC < LLN_5 and FVC < LLN_5)	1 (0.3%)	20 (2%)*
Number and percentage of individuals with confirmed restrictions as assessed by total lung capacity (TLC < LLN_5)	0 (0%)	1(1%)
Number and percentage of individuals with hyperinflation (RV/TLC z-score > ULN)	66 (20%)	56 (66%)*
Number and percentage of individuals with a $DLCO_{10s}$ abnormality	46 (14%)	54 (64%)*
Number and percentage of individuals with a $DLNO_{10s}$ < LLN_5	61 (19%)	59 (69%)*
Number and percentage of subjects with either obstruction, restriction, hyperinflation, DLCO, or DLNO < LLN_5	105 (33%)	77 (91%)*

Mean (SD) *For all variables marked with an asterisk, $p < 0.05$. The *p*-values were determined after controlling for a false discovery rate of 1% (38). The combined Z-scores for diffusing capacity were

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

calculated using segmented (piecewise) linear regression (SLR) of Zavorsky & Cao (2022) [33]. LLN_5 = lower limit of normal defined as the 5th percentile.

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

Table 2. Binary Logistic regression results of the model with the best performance as determined by the lowest BIC. The z-scores of TLC, FEV_1, and $DLNO_{10s}$ (GAMLSS) best predict emphysema.

Predictors Fixed effects	Estimate [95% CI]	SE	z-value	VIF	McFadden's Pseudo R^2 [Overall = 0.66]	Nagelkerke R^2 [Overall = 0.77]	Odds ratios [95%CI]
Intercept	–5.27 [–6.33, –4.33]	0.51	–10.30	–	–	–	0.01 [0.00, 0.02]
FEV_1 z-scores	–1.43 [–1.85, –1.03]	0.21	–6.87	1.15	0.52	0.64	0.24 [0.16, 0.35]
TLC z- scores	1.55 [1.05, 2.08]	0.26	5.92	1.40	0.08	0.13	4.73 [2.87, 8.07]
$DLNO_{10s}$ z-scores (GAMLSS)	–0.73 [–1.02, –0.46]	0.14	–5.10	1.45	0.32	0.43	0.49 [0.36, 0.64]

We compared a 3-predictor logistic model with and without a study-level random intercept. The model without the study-level random intercept had slightly higher expected out-of-sample accuracy [The change in expected log predictive density (ΔELPD) = 0.22 ± 0.27 is within 1 SE, stacking weight = 1.00]. Therefore fixed-effects model without a study random intercept was used.

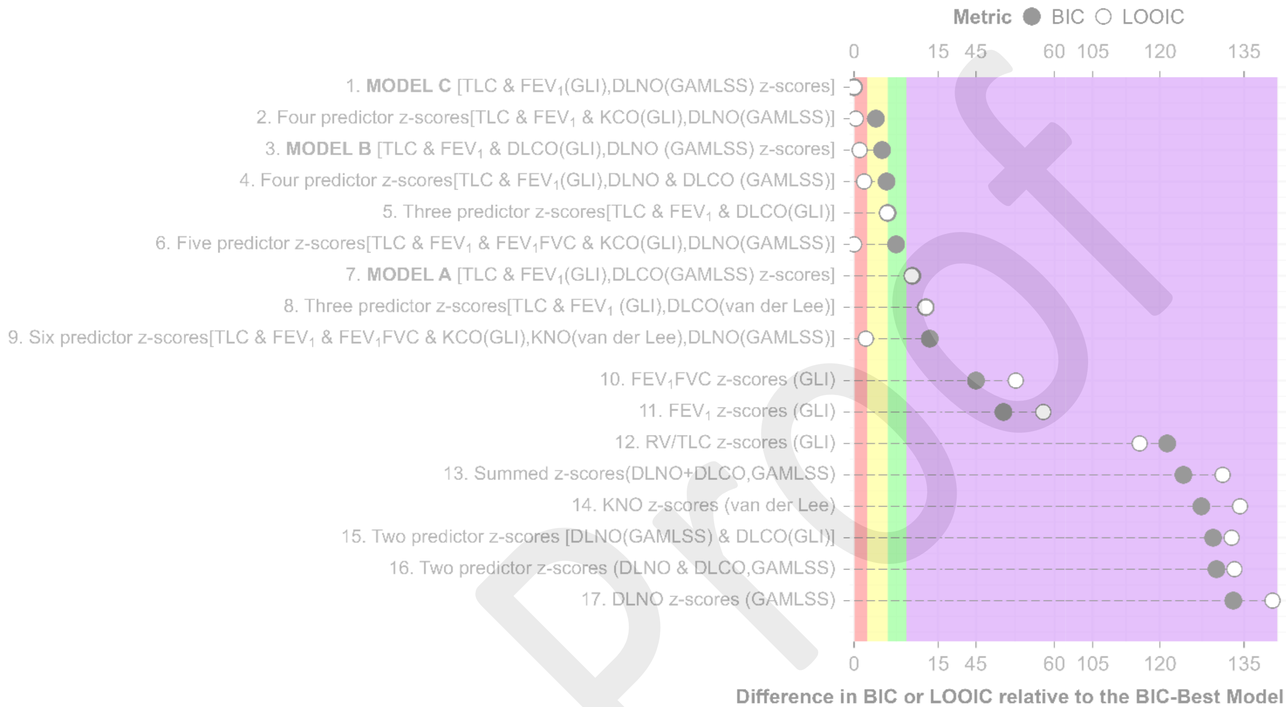
Null deviance: 417.6 on 407 degrees of freedom. Residual deviance: 140.6 on 404 degrees of freedom. AIC (null) = 419.6; BIC (null) = 423.6. AIC for the three-predictor model = 148.6; BIC for the three-predictor model = 164.6. McFadden's pseudo-R^2 is a measure that assesses how well the model fits the data relative to a baseline (null) model of no predictors. It is not the same as the R^2 in linear regression. McFadden's pseudo-R^2 is a scaled measure of improvement in log-likelihood by comparing the model's log-likelihood with the three predictors to the log-likelihood of a null model with no predictors, just the intercept. Overall, McFadden's pseudo R^2 = 0.66 suggests that TLC, FEV_1, and $DLNO_{10s}$ z-scores explain a significant portion of the log-likelihood compared to the null model, suggesting a strong fit. However, McFadden's R^2 does not directly quantify the percentage of variance in emphysema explained by those specific predictors. However, the Craig-Uhler R^2 (NagelKerke R^2) was 0.77, adjusted for the binary outcome, indicating that the three-predictor model explained 77% of the explainable variance.

Note: The overlapping McFadden pseudo-R^2 occurs because the predictors are correlated. For example, TLC z-scores and FEV_1 z-scores have related aspects of lung function sharing variance explaining disease. The TLC z-scores and FEV_1 z-scores are obtained from GLI equations [28,29], while the $DLNO_{10s}$ z-scores were obtained using GAMLSS equations of Zavorsky & Cao (2022) [33]. These specific reference equations presented here were a better fit than DLNO reference equations, be they segmented linear regression reference equations [33] or otherwise [32]. Furthermore, $DLCO_{10s}$ z-scores obtained from any reference equation, including GLI equations [30-32], displayed inferior fitting compared to $DLNO_{10s}$, and removing it resulted in a negligible loss of 0.2-0.3% in pseudo-R^2 and a slight increase in AIC (from 148.6 to 149.5) and a BIC (from 164.6 to

169.6). This implies that $DLCO_{10s}$ add little unique predictive value for emphysema in this dataset. Thus, the final three predictor models are included without $DLCO_{10s}$. In this analysis, there were 85 smokers with emphysema and 323 smokers without emphysema.

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645

Figure 1. Bayesian Information Criterion (BIC) and Leave-One-Out Information Criterion (LOOIC) of models for emphysema prediction and generalizability.

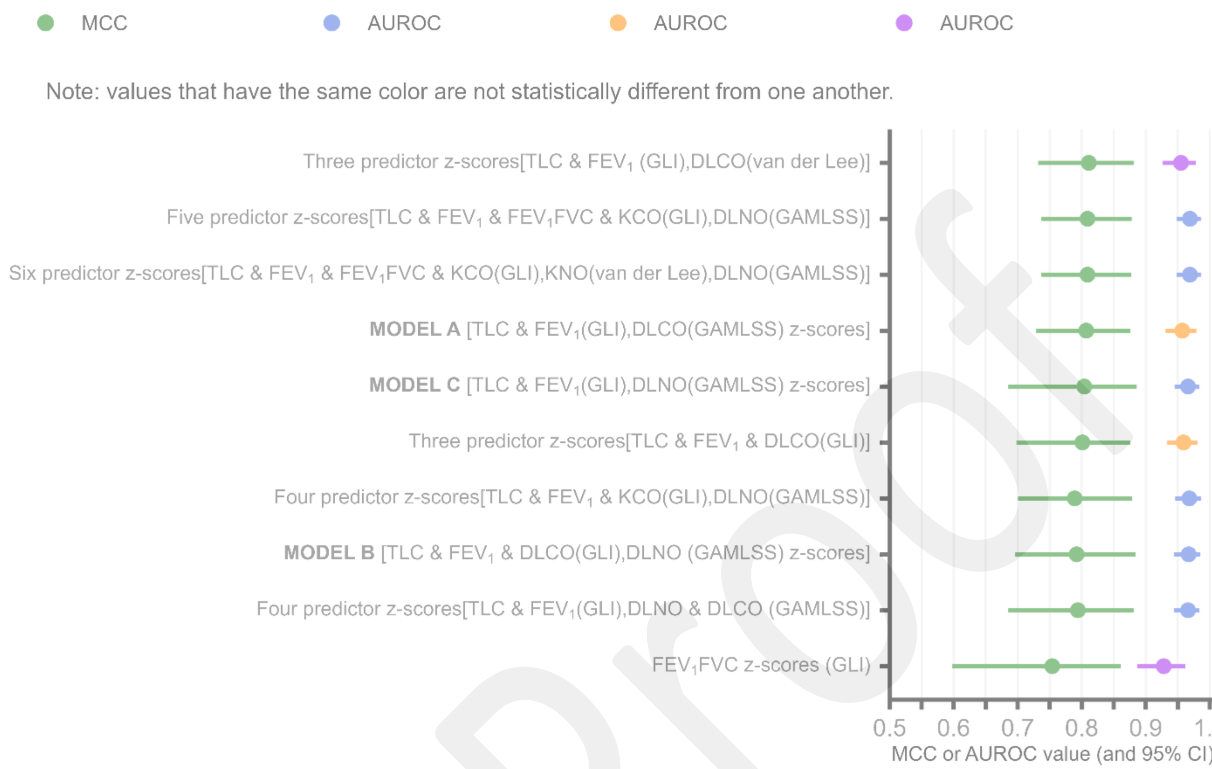


Models are ranked by their difference in BIC (black circles) and LOOIC (white circles) relative to the top-performing model (top = best). BIC penalizes model complexity; LOOIC evaluates predictive performance via cross-validation. Coloured zones indicate performance tiers:

- **Red zone (BIC or LOOIC difference ≤ 2.2):** Models nearly as good as the best model.
- **Yellow zone (BIC or LOOIC difference 2.3–5.9):** Models with substantial but acceptable performance differences compared to the best model.
- **Green zone (BIC or LOOIC difference 6.0–9.2):** Models with considerably weaker performance compared to the best model.
- **Purple zone (BIC or LOOIC difference ≥ 9.3):** Models with significantly poorer fit compared to the best model.

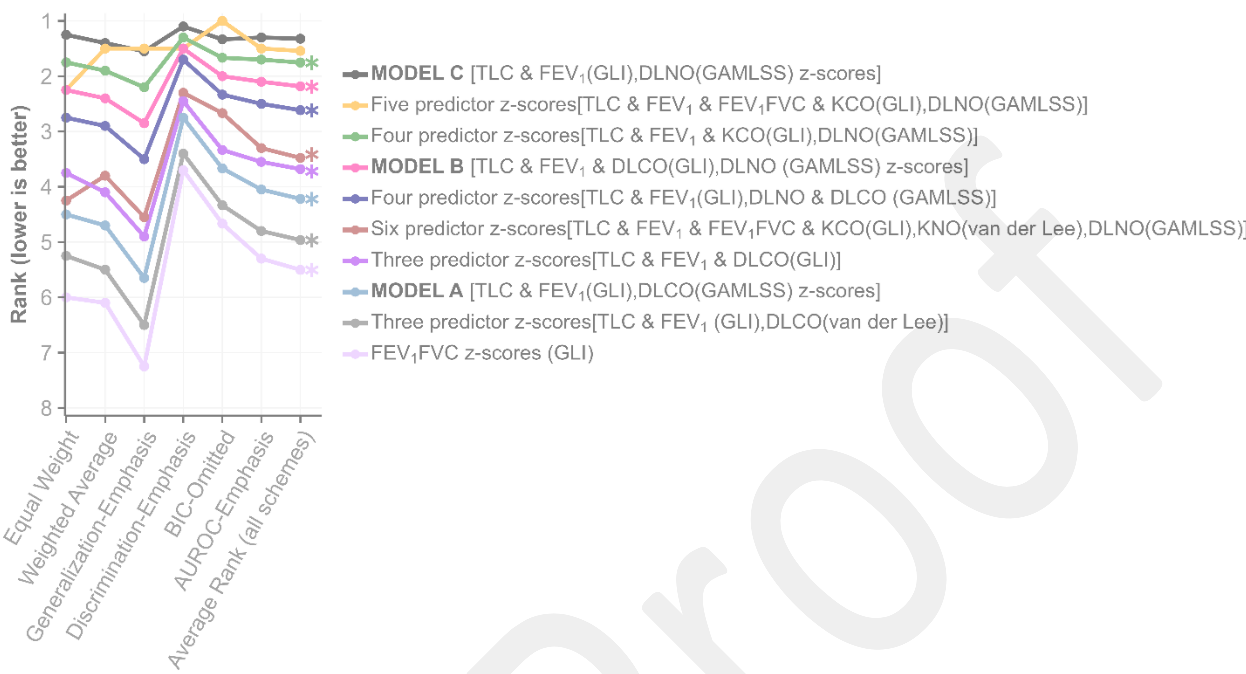
The x-axis shows the difference from the best model—smaller is better. The best-performing model is the three-predictor z-score model of TLC + FEV_1 + $DLNO_{10s}$ (GAMLSS) derived from the GLI equations [28,29] and DLNO z-scores from the GAMLSS equations [33] (n=323 smokers with without emphysema; n= 85 smokers with emphysema).

Figure 2. The discriminatory classification performance of the top 10 predictive models.



Models are ranked by Matthews Correlation Coefficient (MCC; blue bars and 95% CI) with corresponding AUROC values shown (red circles and error bars). CIs are based on 100,000 bootstrap samples. The top-performing model is a 3-predictor GAMLSS model using TLC, FEV_1, and DLNO z-scores, which achieved the highest MCC and AUROC. Lower-ranked models showed weaker predictive performance. Asterisks (*) indicate statistically significant differences from the top model ($p < 0.05$, Benjamini-Hochberg FDR 1%). Mean (SD). For all variables marked with an asterisk, $p < 0.05$ after FDR correction (1%). Combined diffusing capacity z-scores were calculated using piecewise linear regression Zavorsky & Cao (2022) [33]. LLN5 = lower limit of normal, 5th percentile.

Figure 3. Weight sensitivity: model rankings across weighted schemes.



For each of the six ranking schemes (Equal Weight, Weighted Average, Generalization-Emphasis, Discrimination-Emphasis, BIC-Omitted, AUROC-Emphasis), we computed a per-scheme rank difference (Δ rank) between each comparator and the baseline MODEL C [TLC & FEV_1(GLI), DLNO(GAMLSS) z-scores] (Δ rank = rank_comparator − rank_MODEL C; positive = worse). We then took the mean Δ rank across the six schemes for each comparator and obtained 95% CIs via nonparametric bootstrap by resampling schemes with replacement (B = 10,000). We computed per-scheme rank differences vs MODEL C for the six schemes, then took the mean of those six differences for each comparator, and then bootstrapped CIs by resampling schemes (n=6) with replacement Across-scheme mean Δ rank vs MODEL C (95% CI), Eight comparators have significantly higher ranking numbers (higher = worse) than MODEL C (CIs entirely > 0): FEV_1/FVC z-scores (GLI): +4.18 [3.36, 4.98]; Three predictor z-scores [TLC & FEV_1 (GLI), DLCO (van der Lee)]: +3.64 [2.98, 4.33]; MODEL A [TLC & FEV_1 (GLI), DLCO (GAMLSS)]: +2.90 [2.28, 3.53]; Three predictor z-scores [TLC & FEV_1 & DLCO (GLI)]: **+2.36** [1.88, 2.85]; Six predictor z-scores [TLC & FEV_1 & FEV_1/FVC & KCO (GLI), KNO (van der Lee), DLNO (GAMLSS)]: +2.16 [1.57, 2.72]; Four predictor z-scores [TLC & FEV_1 (GLI), DLNO & DLCO (GAMLSS)]: +1.29 [0.95, 1.64]; MODEL B [TLC & FEV_1 & DLCO (GLI), DLNO (GAMLSS)]: +0.86 [0.63, 1.08]; Four predictor z-scores [TLC & FEV_1 & KCO (GLI), DLNO (GAMLSS)]: +0.43 [0.32, 0.55]. The five-predictor DLNO model is not significantly different from MODEL C (CI includes 0); However, due to the desire of having a parsimonious model, MODEL C wins.

Published online November 10, 2025 https://doi.org/10.15326/jcopdf.2025.0645