Dynamic Prediction of Pulmonary Hypertension in Systemic Sclerosis Using Landmark Analysis

Svetlana I. Nihtyanova,1 Benjamin E. Schreiber,2 Voon H. Ong,3 Athol U. Wells,4 John G. Coghlan,2 and Christopher P. Denton3

Objective. Pulmonary hypertension (PH) is a serious complication of systemic sclerosis (SSc). In this study, we explored the prediction of short-term risk for PH using serial pulmonary function tests (PFTs) and other disease features.

Methods. SSc patients in whom disease onset occurred ≥10 years prior to data retrieval and for whom autoantibody specificity and PFT data were available were included in this study. Mixed-effects modeling was used to describe changes in PFTs over time. Landmarking was utilized to include serial assessments and stratified Cox proportional hazards regression analysis with landmarks as strata was used to develop the PH prediction models.

Results. We analyzed data from 1,247 SSc patients, 16.3% of whom were male and 35.8% of whom had diffuse cutaneous SSc. Anticentromere, antitopoisomerase, and anti–RNA polymerase antibodies were observed in 29.8%, 22.0%, and 11.4% of patients, respectively, and PH developed in 13.6% of patients. Over time, diffusing capacity for carbon monoxide (DLCO) and carbon monoxide transfer coefficient (KCO) declined in all SSc patients (up to 1.5% per year) but demonstrated much greater annual decline (up to 4.5% and 4.8%, respectively) in the 5–7 years preceding PH diagnosis. Comparisons between multivariable models including either DLCO, KCO, or forced vital capacity (FVC)/DLCO ratio, demonstrated that both absolute values and change over the preceding year in those measurements were strongly associated with the risk of PH (hazard ratio [HR] 0.93 and 0.76 for KCO and its change; HR 0.90 and 0.96 for DLCO and its change; and HR 1.08 and 2.01 for FVC/DLCO ratio and its change; P < 0.001 for all). The KCO-based model had the greatest discriminating ability (Harrell’s C-statistic 0.903).

Conclusion. Our findings strongly support the importance of PFT trends over time in identifying patients at risk of developing PH.

INTRODUCTION

Pulmonary hypertension (PH) is an important complication of systemic sclerosis (SSc), ultimately affecting 10–15% of all patients with SSc. Unlike manifestations of SSc in other organs, which commonly develop early in the disease course, PH develops late in the disease course. The risk of developing PH is very low in the first 3 years of disease and thereafter is 1–2% per year (1–4). Although the prognosis of SSc-related PH (SSc-PH) is poor, with a median survival of ~3 years (5,6), emerging evidence suggests that earlier diagnosis and early treatment initiation despite mild symptoms is associated with a significant improvement in survival (6–10).

Potential screening tools for PH, separately or in combination, include clinical signs and symptoms, blood biomarkers, echocardiography, pulmonary function testing (PFT), cardiac magnetic resonance imaging, and cardiopulmonary exercise testing (11). Although several PH prediction models and algorithms exist, none utilize serial patient assessments and longitudinal data from PFTs over extended time periods (12–18).

PFTs are used routinely to screen for and monitor lung involvement in SSc. Annual PFT assessments are recommended for all SSc patients. Both spirometry and gas transfer should be assessed to evaluate for the possible presence of restrictive or obstructive pulmonary disease or pulmonary vasculopathy (19). The interpretation of PFT results in isolation can be challenging in...
the context of SSc. Carbon monoxide transfer coefficient (KCO), the uptake of carbon monoxide per unit of alveolar volume (VA), is relatively preserved in ILD, while serial decline in KCO is a specific measure of pulmonary vasculopathy (20). In contrast, both interstitial lung disease (ILD) and PH are associated with declines in diffusing capacity for carbon monoxide (DLCO, the calculated product of measured KCO and measured VA). The ratio of forced vital capacity (FVC) to DLCO (FVC/DLCO) is also used in PH screening, with higher ratios predicting greater PH risk (21,22). Although FVC/DLCO provides similar information as KCO, it is substantially more imprecise due to the variability of FVC, KCO, and VA from which it is calculated (20). In patients who develop both ILD and PH, the PFT changes are often mixed. In addition, smoking history and emphysema can further confound the interpretation of PFT measurements (23). The substantial intra- and interpatient variability in PFT measurements, with regard to DLCO in particular, also means that a single PFT assessment may not be very informative, and serial PFT results are needed for context.

The development of models for PH prediction in SSc is complicated by the interdependence of disease characteristics. For example, anticentromere antibody (ACA) is associated with low risk of ILD (1), higher FVC, and lower DLCO compared to other antibodies, even in subpopulations where both PH and ILD have been excluded (24). Anti-topoisomerase I antibody (ATA)–positive patients are at a much higher risk of ILD and lower risk of PH than any other antibody group (1), but patients with severe ILD can develop group 3 PH. The different PFT measurements are also strongly intercorrelated, which may lead to multicollinearity when they are included in the same model. Time-varying disease characteristics used in prediction models, with potentially time-varying effects on outcome, make analysis and interpretation of longitudinal data challenging (25).

Landmark analysis was first described by Anderson et al in 1983 (26). It was introduced as an unbiased approach to the comparison between survival time in responders and nonresponders to chemotherapy among cancer patients. The methodology avoids the bias resulting from grouping patients based on a diagnosis to investigate changes in PFT results in the years preceding development of PH. As the FVC/DLCO ratio had a skewed distribution, this was log-transformed and modeled as log(FVC/DLCO). Estimates were then exponentiated and reported as geometric means.

Mixed-effects modeling and PFT trajectory prediction. Separate mixed-effects models were constructed using the whole data set, in which time was included as a predictor and FVC, DLCO, KCO, or log(FVC/DLCO) were included as outcomes. For each measure we calculated patient-specific, model-derived best linear unbiased predictions of the random effects for the model parameters. For each patient we then calculated model-predicted values for FVC, DLCO, KCO, and log(FVC/DLCO) at 12-month intervals over the follow-up, as a sum of the fixed

PATIENTS AND METHODS

Patients and disease characteristics. All participants had a confirmed diagnosis of SSc and fulfilled the 2013 American College of Rheumatology/EULAR classification criteria for SSc (27). Patients were included in the study if SSc onset (defined as time of first non-Raynaud’s phenomenon symptom) occurred ≥10 years prior to data retrieval, they had been tested for autoantibodies, and they had undergone at least 1 PFT. Definitions for cutaneous subsets and organ complications are included in the Supplementary Materials and Methods, available on the Arthritis & Rheumatology website at http://onlinelibrary.wiley.com/doi/10.1002/art.42349.

As all patients were followed up and diagnosed prior to the release of the proposed update to the hemodynamic definition of PH in 2019 (28), we used the Venice classification for PH, which specified mean pulmonary artery pressure (mPAP) of ≥25 mm Hg at rest and pulmonary artery wedge pressure (PAWP) of ≤15 mm Hg on right-sided heart catheterization (RHC) as criteria for PH. As it is often difficult to distinguish group 1 (connective tissue disease–associated pulmonary arterial hypertension) and group 3 (ILD–associated PH) in patients with moderate-to-severe ILD, we included both groups in the analysis, and throughout this article the term PH refers to precapillary PH.

Most PFTs were performed at the Royal Free Hospital and the Royal Brompton Hospital, London, UK. However, some PFTs were conducted elsewhere, as many patients attending our specialist center were not local to the hospital and were under the care of several specialists.

Statistical analysis. Descriptive statistics were used to summarize the characteristics of the cohort. To study serial FVC, DLCO, KCO (% predicted), and FVC/DLCO ratio, we used random effects modeling. Possible nonlinear associations between time and PFT results were explored using polynomials. For patients who had not been diagnosed as having PH at the time of data extraction, the time variable was anchored at disease onset. In the PH patient group, time was centered at PH diagnosis to investigate changes in PFT results in the years preceding development of PH. As the FVC/DLCO ratio had a skewed distribution, this was log-transformed and modeled as log(FVC/DLCO). Estimates were then exponentiated and reported as geometric means.
portion of the linear prediction and the predicted patient-specific random effects. The calculated values for log(FVC/DLco) were then back-transformed and included in the PH prediction models as the FVC/DLco ratio.

**Landmarking.** The proposed method by Anderson et al involves a choice of a time point during the follow-up (deemed a landmark) and determining patient characteristics at the landmark time point (26). Patients who have died or been lost to follow-up before that are excluded from the analysis and change in the characteristics after the landmark is ignored. Following this first publication, landmark analysis has become widely used (29). The methodology was developed further by van Houwelingen and Putter, who proposed the use of short-term prediction within a narrow “sliding window” (30). In this way, multiple short-term predictions could be made, using the values of time-varying patient characteristics at the landmark and predicting the cumulative incidence of an outcome within the prespecified time window, starting from the landmark time point. The separate Cox models could further be combined into a stratified Cox “super model”, using landmarks as strata.

Using landmark methodology, we can predict the probability of PH development within a 12-month window from the landmark time point, accounting for the subject status at the landmark (30). To achieve this, we selected landmarks at yearly intervals, starting from 3 years from onset (Figure 1A). For each landmark we created a separate dataset, which included all patients who were alive and had not developed PH at that time point. Time-varying characteristics, including the presence of organ complications and model-predicted PFT measures (absolute values and change over the preceding year) were recorded as of the landmark time point. Time from the landmark to the PH diagnosis was recorded, if this occurred within 12 months after the landmark, otherwise this was censored at 12 months (Figure 1B). For the time-to-event analysis assessing predictors of PH development, all landmark data sets were stacked and analyzed together.

**Time-to-event analysis.** The Kaplan-Meier (KM) estimator was used to calculate survival. Cumulative incidence of PH over the whole follow-up period was calculated using the 1-KM estimate and the cumulative incidence function (CIF) estimate, adjusting for death as a competing risk (31). Comparisons were made between 1-KM and CIF estimates to inform the need for competing risk consideration when developing the Cox regression super models in the landmark data set. Small differences

![Figure 1. Landmarking of the data set. A, Landmarks (LMs) were selected at yearly intervals, between years 3 and 22 from systemic sclerosis (SSc) onset, and a separate data set was created for each landmark. B, As an example, data set 7 included all patients who were alive and had not developed pulmonary hypertension (PH) at 7 years from onset. The values of time-varying characteristics, including presence of organ complications and model-predicted pulmonary function test (PFT) measurements (absolute values and change over the preceding year) were recorded as the values at 7 years from onset. PH events were recorded if they developed between years 7 and 8 from SSc onset. Time from year 7 to PH diagnosis was calculated and recorded, if this was ≤12 months; otherwise time was censored at 12 months. For the time-to-event analysis assessing predictors of PH development, all landmark data sets were stacked and analyzed together.](image-url)

---

**Table 1.** Demographic and clinical characteristics of the cohort of systemic sclerosis (SSc) patients used to develop prediction models for pulmonary hypertension (PH)*

| Characteristic                                      | Value            |
|----------------------------------------------------|------------------|
| Total number of patients                           | 1,247 (100.0)    |
| Male sex                                           | 203 (16.3)       |
| Age at onset, mean ± SD years                      | 46.6 ± 13.4      |
| Emphysema/COPD history†                             | 92 (7.4)         |
| Smoking history‡                                    | 529 (45.0)       |
| Diffuse cutaneous SSc                              | 446 (35.8)       |
| Overlapping syndromes                              | 250 (20.0)       |
| Autoantibodies                                     |                  |
| Anticentromere                                     | 371 (29.8)       |
| Anti-topoisomerase I                               | 274 (22.0)       |
| Anti-RNA polymerase                                | 142 (11.4)       |
| Anti-U3 RNP                                        | 53 (4.3)         |
| Anti-Pm/Sc                                         | 55 (4.4)         |
| Other antibodies§                                   | 201 (16.1)       |
| ANA positive and ENA negative                      | 178 (14.3)       |
| ANA negative                                       | 52 (4.2)         |
| Organ complications                                |                  |
| Interstitial lung disease, any                     | 564 (45.2)       |
| Clinically significant interstitial lung disease    | 521 (41.8)       |
| PH (group 1 and group 3)                           | 170 (13.6)       |
| PAH (group 1)                                      | 132 (10.6)       |
| Cardiac scleroderma                                | 58 (4.7)         |
| Scleroderma renal crisis                           | 87 (7.0)         |
| Death                                              | 404 (32.4)       |

* Except where indicated otherwise, values are the number (%) of patients. COPD = chronic obstructive pulmonary disease; ANA = antinuclear antibodies; ENA = extractable nuclear antigen; PAH = pulmonary arterial hypertension.
† Emphysema/COPD history was obtained from 1,245 patients.
‡ Smoking history was obtained from 1,175 patients.
§ Including anti–ribosomal RNP, anti–Th/To, anti–PL-4, anti–PL-7, anti–PL-12, and/or anti–Sm.
in the estimates over short follow-up periods would indicate that death as a competing risk would have minimal effect on the estimation of PH risk. For the stacked data set, we used stratified Cox proportional hazards regression analysis to assess predictors of PH, with each landmark data set being analyzed as a separate stratum (30). We tested the significance of interaction terms of the covariates and the landmark variable (treated as continuous) to assess for time-varying effects. Given the inherent associations between different PFT measurements, to avoid issues with multicollinearity, we developed multivariable models with either DLCO, KCO, or FVC/DLCO ratio included. Model fit was tested using Akaike’s information criterion and Bayesian information criterion. Predictive discrimination was assessed using Harrell’s C-statistic. All analyses were performed using Stata version 14.

RESULTS

Cohort description. A total of 1,247 participants with confirmed SSC diagnoses were included in the study. Demographic and clinical characteristics are summarized in Table 1. All patients had a disease onset between 10 and 22 years prior to data extraction, with a mean ± SD follow-up of 12.6 ± 5.4 years. Of the 1,247 patients included in the cohort, 133 (10.7%) were lost to follow-up (patients had not died and had undergone their last assessment more than 2 years prior to the year of data extraction). All patients had at least 1 PFT over the follow-up period, with 147 patients (11.8%) having 2 assessments and 945 patients (75.8%) having 3 or more assessments. In total, 8,165 PFT results were available and 6,769 (82.9%) of these were results for FVC, DLco, and KCO, while data were missing on 1 or 2 of the measurements for the remaining PFT results. FVC was available in 8,030 (98.4%), DLco in 7,899 (96.7%), and KCO in 6,838 (83.8%) of the available PFT results. FVC/DLco ratios could be calculated from 7,767 of the available PFT results (95.1%). The mean time between PFTs was 15.3 months.

Over the follow-up, PH developed in 170 (13.6%) of the patients. KM estimates of PH incidence at 5, 10, 15, and 20 years from disease onset were 3.9%, 9.4%, 16.5%, and 23.0%, respectively; CIF estimates of PH incidence at 5, 10, 15, and 20 years from disease onset were 3.8%, 8.8%, 14.8%, and 19.7%, respectively. Approximately one-third of

Figure 2. Measurements of serial forced vital capacity (FVC) (A and B) and diffusing capacity for carbon monoxide (DLco) (C and D) over the entire follow-up in SSC patients who had not been diagnosed as having PH at the time of data extraction (A and C), and in the years preceding PH diagnosis among patients who had been diagnosed as having PH at the time of data extraction (B and D). Each thin line represents multiple measurements over time in a single patient; blue dots represent measurements of patients in whom FVC or DLco was assessed only once; the thick green line is the model-predicted mean FVC or DLco measurement. See Figure 1 for definitions. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/art.42349/abstract.
the patients had died at the time of data extraction, with KM survival estimates at 5, 10, 15, and 20 years of 93.1%, 83.5%, 69.2%, and 55.2%, respectively.

**PFT results.** *FVC.* Among the 1,077 non-PH patients, 1,064 had available FVC measurements, while among the 170 patients with PH, 141 had undergone FVC assessments prior to being diagnosed as having PH. In the non-PH group, FVC demonstrated little change over time. The mean FVC at baseline was 89%, and over time there was a small but significant increase of <1% per year, with an estimated mean FVC of 93.6% at 10 years from onset and 93.0% at 20 years from onset. In the years preceding the development of PH, a significant but clinically small nonlinear decline in the FVC of ≤1% per year was seen, with an estimated mean FVC at PH diagnosis of 77.1%, while 5 and 10 years prior to PH diagnosis, mean FVCs were 81.9% and 84.0%, respectively (Figures 2A and B and Supplementary Tables 1 and 2, http://onlinelibrary.wiley.com/doi/10.1002/art.42349).

*DLco.* DLco results were available for 1,061 of 1,077 non-PH patients and 138 of 170 PH patients prior to PH diagnosis. In the non-PH group, DLco values demonstrated a small but consistent annual decline of 0.8–1.5% per year, with estimated mean DLco values at 1, 5, 10, 15, and 20 years of follow-up of 70.4%, 66.1%, 62.1%, 57.8%, and 51.4%, respectively. Modeling serial DLco values over the years preceding PH diagnosis revealed rates of decline similar to those among non-PH patients up to ~7 years prior to PH diagnosis. After this period, annual decline became greater than 1.5%, steadily increasing with every year closer to PH diagnosis (mean 1.6%, 1.9%, 2.3%, 2.8%, 3.3%, 3.9%, and 4.5% between year 7 and the time of PH confirmation on RHC) (Figures 2C and D and Supplementary Tables 3 and 4, http://onlinelibrary.wiley.com/doi/10.1002/art.42349). Model-estimated mean DLco values at 15 years, 10 years, 5 years, and 1 year before PH diagnosis were 64.6%, 59.7%, 52.9%, and 40.7%, respectively.

*Kco.* Kco results were available for 983 of 1,077 SSc patients without PH, while Kco results prior to PH diagnosis were available for 115 of 170 patients who were later diagnosed as
having PH. Analysis of serial Kco values revealed results similar to those of the analysis of DLco values. Among the non-PH patients, Kco declined over time at a rate of 0.7–1.4% per year. Estimated mean Kco values at 1, 5, 10, 15, and 20 years of follow-up were 87.0%, 83.0%, 79.6%, 75.5%, and 68.4%, respectively. Among the patients who developed PH, yearly decline increased steadily over the 6 to 7 years prior to PH diagnosis, with annual mean declines of 1.5%, 1.8%, 2.3%, 2.8%, 3.4%, 4.0%, and 4.8% per year between year 7 and the time of RHC diagnosis. Model-estimated average Kco measurements at 20, 15, 10, 5, and 1 year prior to PH diagnosis were 99.7%, 84.9%, 79.3%, 72.9%, and 60.5%, respectively (Figures 3A and B and Supplementary Tables 5 and 6, http://onlinelibrary.wiley.com/doi/10.1002/art.42349).

FVC/DLco. FVC/DLco ratios could be calculated in 1,048 of 1,077 non-PH patients and were available prior to PH development in 137 of 170 PH patients. Among the non-PH patients, there was a gradual mean increase in the FVC/DLco ratio over time by 0.02–0.04 per year. At 1 year from SSc onset, the estimated mean FVC/DLco ratio was 1.29, which increased to 1.42 at year 5, 1.54 at year 10, 1.67 at year 15, and 1.90 at year 20. In the PH group, there was a similar gradual increase in the FVC/DLco ratio over the years preceding PH diagnosis, which varied between a mean of 0.02 and 0.04 per year until ~7 years prior to PH diagnosis, when the estimated mean FVC/DLco ratio was 1.5. Following that, we observed much greater annual increases in the FVC/DLco ratio, which became greater than 0.1 per year in the final 3 years, with the mean estimated FVC/DLco ratio being 1.6 at 5 years, 2.0 at 1 year before PH, and 2.2 at PH diagnosis (Figures 3C and D and Supplementary Tables 7 and 8, http://onlinelibrary.wiley.com/doi/10.1002/art.42349).

**Landmark analysis.** *Univariable analysis.* Patient characteristics at each landmark are summarized in Supplementary Table 9 (http://onlinelibrary.wiley.com/doi/10.1002/art.42349).

| Table 2. Univariable analysis of the associations between SSc patient characteristics and the risk of PH development* |
|---------------------------------|-----------|--------|
|                                | HR        | 95% CI |
| Age at onset                    | 1.02      | 1.01–1.04 |
| Male                            | 1.80      | 1.23–2.64 |
| Smoking history                 | 1.23      | 0.88–1.71 |
| Emphysema/COPD                  | 1.92      | 1.22–3.01 |
| Overlap                         | 0.53      | 0.34–0.83 |
| Diffuse cutaneous subset        | 1.04      | 0.73–1.47 |
| Autoantibodies†                 |           |        |
| Anti-topoisomerase 1            | 0.67      | 0.39–1.14 |
| Anti-RNA polymerase             | 1.34      | 0.78–2.30 |
| Anti-U3 RNP                     | 2.60      | 1.41–4.80 |
| Anti-Pm/Scl                     | 0.61      | 0.22–1.71 |
| ANA positive and ENA negative   | 1.15      | 0.67–1.96 |
| Other antibodies                | 1.54      | 0.97–2.44 |
| Characteristics at LM          |           |        |
| Age                             | 1.02      | 1.01–1.04 |
| Clinically significant ILD      | 1.96      | 1.42–2.71 |
| Cardiac scleroderma             | 0.92      | 0.29–2.88 |
| Scleroderma renal crisis        | 2.01      | 1.11–3.63 |
| PFTs at LM                      |           |        |
| FVC % predicted                 | 0.97      | 0.96–0.98 |
| DLco % predicted                | 0.91      | 0.90–0.92 |
| Kco % predicted                 | 0.92      | 0.91–0.93 |
| FVC/DLco ratio*10‡              | 1.12      | 1.11–1.13 |
| Change in PFTs over 1 year preceding LM | | |
| ΔFVC % predicted                | 0.95      | 0.92–0.97 |
| ΔDLco % predicted*LM§           | 0.67      | 0.57–0.79 |
| ΔKco % predicted*LM§            | 0.98      | 0.96–1.00 |
| ΔFVC/DLco ratio*10‡             | 2.13      | 1.63–2.80 |
| ΔFVC/DLco ratio*10*LM§          | 1.07      | 1.02–1.11 |

* HR = hazard ratio; 95% CI = 95% confidence interval; LM = landmark; ILD = interstitial lung disease; PFT = pulmonary function test; FVC = forced vital capacity; DLco = diffusing capacity for carbon monoxide; Kco = carbon monoxide transfer coefficient (see Table 1 for other definitions).
† Anticentromere was used for reference.
‡ The FVC/DLco ratio has been transformed to (FVC/DLco)*10 for the model to enable more practical interpretation for change. Because 1 unit change in the ratio is too large, with the transformation, the beta coefficients indicate the effect on outcome of 0.1 units change.
§ Interactions between lung function measurements and the landmarks are indicated by *LM.
Associations between different variables and hazard of PH development from the univariable analyses are detailed in Table 2. There were significant associations between the development of PH within 12 months and all PFT measurements, including % predicted FVC, DLco, and Kco, and FVC/DLco ratios, as well as changes in the results of those PFTs over the preceding 12 months. Lower FVC, DLco, and Kco, and declines in those measurements increased the hazard of developing PH. Conversely, higher FVC/DLco ratios and increases in the ratio over the preceding 12 months were associated with higher risk of PH. There was evidence for interactions between change in PFT values and landmarks, suggesting that the increase in PH hazard for a unit drop in FVC, DLco, or Kco, or an increase in FVC/DLco ratio became greater with longer disease duration (Supplementary Table 10, http://onlinelibrary.wiley.com/doi/10.1002/art.42349).

**Multivariable analysis.** Multivariable models were developed with either DLco (with or without FVC), Kco (with or without FVC), or FVC/DLco ratio, and compared in terms of fit and discriminatory performance (Supplementary Table 11, http://onlinelibrary.wiley.com/doi/10.1002/art.42349). The comparisons suggested that FVC/DLco ratio is a poorer predictor of PH development compared to DLco or Kco measurements. The models including Kco performed better when adjusting for restrictive lung disease, using either FVC or the presence of ILD, while this was not the case when DLco was used. The 3 best models predictive of PH development based on each PFT measurement are summarized in Table 3.

### DISCUSSION

In this report we describe the trends in PFT measurements over a long period in a large, unselected cohort of SSc patients. We utilized landmark analysis to assess the associations between FVC, DLco, Kco, and FVC/DLco ratio, as well as their annual rates of change, and the short-term risk of developing PH in patients followed up for up to 22 years, using available serial assessments from that entire period. Our study expands the findings from previous work, demonstrating the importance of PFTs, and in particular gas transfer factor, in combination with autoantibodies, as predictors of PH development (18).

We confirmed that over time, DLco and Kco steadily declined in all SSc patients irrespective of PH diagnosis, although much greater annual decline was observed in the 5–7 years preceding PH diagnosis. While the average annual rates of change we report fall within the measurement variability of PFTs in individual patients, DLco and Kco trends over a 2-year period or longer will show definite decline, exceeding measurement variability, which is a robust PFT signal that can easily be applied in clinical practice and could be used in PH risk stratification. It is important to acknowledge that the average PFT measurements for the cohort over time in non-PH patients may have been influenced by the

| Table 3. | Multivariable analysis models for predicting the development of PH based on DLco, Kco, or FVC/DLco ratio* |
|----------|------------------------------------------------------------------------------------------|
|          | **DLco** | **Kco** | **FVC/DLco ratio** |
|          | HR | 95% CI | P | HR | 95% CI | P | HR | 95% CI | P |
| **Autoantibodies**† | | | | | | | | | |
| Antitopoisomerase 1 | 0.24 | 0.12–0.47 | <0.001 | 0.18 | 0.10–0.32 | <0.001 | 0.43 | 0.24–0.79 | 0.007 |
| Anti-RNA polymerase | 0.75 | 0.37–1.51 | 0.419 | 0.68 | 0.36–1.26 | 0.222 | 1.08 | 0.60–1.93 | 0.793 |
| Anti-U3 RNP | 2.78 | 1.44–5.36 | 0.002 | 2.58 | 1.39–4.82 | 0.003 | 2.64 | 1.38–5.07 | 0.004 |
| Anti-Pm/Scl | 0.37 | 0.13–1.09 | 0.072 | 0.36 | 0.13–1.01 | 0.053 | 0.58 | 0.20–1.65 | 0.307 |
| ANA positive ENA and negative | 0.44 | 0.23–0.85 | 0.014 | 0.48 | 0.27–0.84 | 0.01 | 0.83 | 0.46–1.49 | 0.532 |
| Other antibodies | 0.61 | 0.35–1.06 | 0.082 | 0.53 | 0.33–0.86 | 0.011 | 0.74 | 0.44–1.25 | 0.258 |
| **Age** | | | | 1.03 | 1.01–1.04 | 0.002 | | | |
| Scleroderma renal crisis | 2.57 | 1.21–5.48 | 0.014 | 2.42 | 1.21–4.82 | 0.012 | | | |
| Clinically significant ILD | | | | | | | | | |
| Emphysema/COPD | | | | | | | | | |
| **PFTs at LM** | | | | | | | | | |
| Kco % predicted | 0.93 | 0.92–0.94 | <0.001 | 0.83 | 0.82–0.84 | <0.001 | 1.09 | 1.08–1.10 | <0.001 |
| ΔKco % predicted | 0.76 | 0.67–0.88 | <0.001 | 0.76 | 0.68–0.86 | <0.001 | 0.94 | 0.93–0.96 | <0.001 |
| FVC % predicted | 0.97 | 0.96–0.98 | <0.001 | 0.97 | 0.96–0.98 | <0.001 | 0.97 | 0.96–0.98 | <0.001 |
| ΔFVC% predicted*LM‡ | 0.97 | 0.95–0.98 | <0.001 | 0.97 | 0.95–0.98 | <0.001 | 0.97 | 0.95–0.98 | <0.001 |
| DLco % predicted | | | | | | | | | |
| ΔDLco% predicted*LM‡ | | | | | | | | | |
| FVC/DLco ratio*10§ | | | | | | | | | |
| ΔFVC/DLco ratio*10§ | | | | | | | | | |

* PH = pulmonary hypertension; ANA = antinuclear antibodies; ENA = extractable nuclear antigen; COPD = chronic obstructive pulmonary disease (see Table 2 for other definitions).
† Anticentromere was used for reference.
‡ Interactions between lung function measurements and the landmarks are indicated by *LM.
§ The FVC/DLco ratio has been transformed to (FVC/DLco)*10 for the model to enable more practical interpretation for change. Because 1 unit change in the ratio is too large, with the transformation, the beta coefficients indicate the effect on outcome of 0.1 units change.

http://onlinelibrary.wiley.com/doi/10.1002/art.42349)
fact that patients dropped out of the study due to severe ILD or death, which may explain the small increase in average FVC observed over time.

Most existing models and algorithms for the prediction of PH were derived in subjects who were considered at risk of PH development (12–14). This restricts the applicability of such models, and their extrapolation to the general SSc population could lead to incorrect predictions (32). In many models, predictions are based on cross-sectional data (14,15,17) or longitudinal data incorporated as a single summary value (for example, worst ever assessment or observation ever present over the entire follow-up) (4). A more recent study from the Canadian Scleroderma Research Group registry utilized a more sophisticated methodologic approach by including data from repeat visits prior to PH diagnosis, allowing for inclusion of both absolute PFT measurements and their change over the preceding year (16). In our center, previous research exploring predictors of pulmonary complications in SSc used information available at baseline (within a window from disease onset) to identify patients at risk of developing PH at any time during their follow-up (18). Of the multiple screening tools for PH that have been published, the most robust to date is the DETECT score (17), which was derived in an enriched patient population with disease durations greater than 3 years and DLco <60%.

Although landmark analysis has been used extensively in oncology research, to our knowledge, so far only 2 studies have used this approach for prognostication in PH patients (33,34). Mazurek et al utilized a single landmark—1 year after baseline—to test if tricuspid annular plane systolic excursion assessed after 12 months of PAH therapy was predictive of survival (33). McLaughlin et al (34) used 3 landmarks—at 3, 6, and 12 months—to investigate the association between PAH-related morbidity events and the risk of death in the pooled data from the SERAPHIN and GRIPHON clinical trials (35,36). For each of the landmarks, survival until the end of the study was compared between patients who had and patients who had not experienced a PAH-related morbidity event before the landmark.

Using dynamic prediction with multiple landmarks, we developed models applicable to the entire length of the studied period and focused on prediction within a narrow time window, which reflects real-life management of SSc-PH patients. We demonstrated how landmark analysis may be applied to a well-characterized cohort of unselected SSc patients to define changes in PFT trajectory that could predict the development of PH. This has relevance to current practice where routine interval assessment with PFT is incorporated into standard of care, and there is added benefit in considering each measurement in the context of preceding values. It can also improve the selection of cases for further assessment and can define the broader profile of patients at risk of developing PH to refine real-world detection. In addition, the longitudinal dimension of this analysis and the unselected nature of the cases included make it more relevant to current practice than some previous analyses.

As it is often difficult to clinically distinguish PH groups 1 and 3, we took a pragmatic approach and included both in this analysis. This has become even more relevant with the recently approved indication extension of inhaled treprostinil to include ILD-PH (37). The patients included in this study were seen and assessed prior to the publication of the proposed updated hemodynamic definition of precapillary PH (mPAP >20 mm Hg, PAWP ≤15 mm Hg, and pulmonary vascular resistance [PVR] ≥3 Wood units) and PH was defined by mPAP ≥25 mm Hg. Nevertheless, PVR ≥3 Wood units remains a stringent threshold, and several studies have presented evidence suggesting that the application of the new definition has had a limited effect on the diagnosis of PH overall, identifying only a small number of additional cases (38–40). Consequently, it is unlikely our results would have been substantially different if the new definition had been used.

Similar to previous analyses of our SSc cohort, we found that, of the SSc-specific autoantibodies, anti-U3 RNP was the strongest predictor of the development of PH. While this is a comparatively rare specificity among SSc patients, it is worth highlighting that it has a higher frequency among patients who are not Caucasian (41–43). ATA had the strongest negative association with PH development. In the multivariable analyses, ACA conveyed the second highest risk of PH development, which nevertheless was less than half the risk associated with anti-U3 RNP.

Important confounders of the association between DLco or Kco and PH are smoking history and a diagnosis of chronic obstructive pulmonary disease (COPD)/emphysema. In this analysis, smoking was not associated with PH, while a history of COPD or emphysema was associated with an increased risk of PH in the univariable analysis. The association did not hold consistently in the multivariable models based on DLco or Kco, possibly because the effect was conveyed through the PFT measurements, but it did remain a significant independent covariate in the model based on FVC/DLco ratio.

The best-performing models were based on Kco and adjusted for background restrictive lung disease (including either FVC or ILD as covariates). It is noteworthy that Kco is an alternative to the FVC/DLco ratio and consideration should be given to the variability of the different PFT measures. Compared to FVC and Kco, DLco has greater variability, as it is computed from measured Kco and VA. The FVC/DLco ratio then has the combined measurement variability of FVC, Kco, and VA. As a result, in serial assessments of PFTs, a 10% decline in Kco predicts mortality among SSc-ILD patients, but a 20% increase in FVC/DLco ratio is required for the same effect on mortality (44). Consequently, the good fit of the models using Kco, especially when adjusting for FVC, would likely be at least partly due to Kco and FVC being less variable than DLco and FVC/DLco ratio. At the same time, even though DLco-based models show poorer fit than Kco-based models, their discriminating ability is slightly better than that of Kco models that do not also adjust for FVC. One possible reason for this is that DLco captures decline in both Kco and VA (reflecting
worsening ILD), which can also lead to PH. It is also possible that the poorer performance of FVC/DL\textsubscript{CO} ratio is related to the inclusion of both group 1 and group 3 PH patients in the analysis.

Strengths of this study include the use of a robust data set, based on a large number of patients followed up for over 2 decades with serial clinical assessments done consistently in a single center with low threshold for referral to RHC. Nevertheless, it is important to acknowledge limitations, including that it is not possible to determine the exact timing of PH onset, and we used the date of RHC-based diagnosis for the time-to-event analysis, which is likely to introduce noise in the data. Moreover, hemodynamic data were not complete for all patients, as some of the diagnostic RHCs were performed prior to the use of electronic records or in other hospitals; therefore, it was not possible to explore in detail the associations between changes in PFT scores and PH severity at diagnosis. Additionally, it was not possible to evaluate the role of other known predictors of PH, such as echocardiogram-derived measurements and N-terminal pro-brain natriuretic peptide levels, as those have not been measured routinely in all SSc patients in the past.

A source of potential bias in the data is that patients with ILD who also showed modest evidence for PH may have been less likely to be catheterized, while breathlessness and drop in DL\textsubscript{CO} or increase in FVC/DL\textsubscript{CO} will often prompt catheterization in the absence of other evidence of PH if no ILD is present. In addition, PFTs would likely be done more frequently in symptomatic patients with suspected PH development or ILD development or progression. The number and frequency of PFT assessments varied between patients, which necessitated the use of model-derived rather than observed values for the PFT measurements and their change and assumed data were missing at random. As a result of the PFTs being performed in several hospitals, part of the variability in results may be related to different approaches for calculation of % predicted values of PFT measurements. While PFT standardization is important, it is not possible in the context of prolonged observational studies. The data we analyzed reflect real world clinical practice and the findings are therefore more broadly applicable.

In conclusion, we demonstrated that an increase in FVC/DL\textsubscript{CO} ratio and an incremental decline in DL\textsubscript{CO} and K\textsubscript{CO} precede PH development by more than 5 years, and therefore could be used as early indicators of PH development. Prospective validation of our findings is needed for them to be applicable in practice, but they nevertheless strongly support the importance of PFT trends over time in identifying patients at risk of developing PH.

**ACKNOWLEDGMENTS**

We are grateful to Professor Bianca De Stavola for her invaluable advice on the statistical aspects of this project and, in particular, the landmark analysis.

**AUTHOR CONTRIBUTIONS**

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Denton had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study conception and design.** Nihtyanova, Denton.

**Acquisition of data.** Nihtyanova, Schreiber, Ong, Wells, Coghlan, Denton.

**Analysis and interpretation of data.** Nihtyanova, Wells, Coghlan, Denton.

**ADDITIONAL DISCLOSURES**

Author Nihtyanova is an employee of GSK.

**REFERENCES**

1. Nihtyanova SI, Sari A, Harvey JC, et al. Using autoantibodies and cutaneous subset to develop outcome-based disease classification in systemic sclerosis. Arthritis Rheumatol 2020;72:465–76.
2. Hachulla E, de Groote P, Gressin V, et al. The three-year incidence of pulmonary arterial hypertension associated with systemic sclerosis in a multicenter nationwide longitudinal study in France. Arthritis Rheum 2009;60:1831–9.
3. Vandecasteele E, Driege B, Meleens K, et al. Screening for pulmonary arterial hypertension in an unselected prospective systematic sclerosis cohort. Eur Respir J 2017;49.
4. Morrisroe K, Huq M, Stevens W, et al. Risk factors for development of pulmonary arterial hypertension in Australian systemic sclerosis patients: results from a large multicenter cohort study. BMC Pulm Med 2016;16:134.
5. Ranjug S, Hussain N, Hurdman J, et al. Idiopathic and systemic sclerosis-associated pulmonary arterial hypertension: a comparison of demographic, hemodynamic, and MRI characteristics and outcomes. Chest 2017;152:92–102.
6. Lewére G, Dauchet L, Hachulla E, et al. Survival and prognostic factors in systemic sclerosis-associated pulmonary hypertension: a systematic review and meta-analysis. Arthritis Rheum 2013;65:2412–23.
7. Humbert M, Yaici A, de Groote P, et al. Screening for pulmonary arterial hypertension in patients with systemic sclerosis: clinical characteristics at diagnosis and long-term survival. Arthritis Rheum 2011;63:3522–30.
8. White RJ, Yonk-Noordegraaf A, Rosenkranz S, et al. Clinical outcomes stratified by baseline functional class after initial combination therapy for pulmonary arterial hypertension. Respir Res 2019;20:208.
9. Brown Z, Proudman S, Morrisroe K, et al. Screening for the early detection of pulmonary arterial hypertension in patients with systemic sclerosis: a systematic review and meta-analysis. Semin Arthritis Rheum 2021;51:495–512.
10. Gaîle N, Rubin L, Hooper M, et al. Treatment of patients with mildly symptomatic pulmonary arterial hypertension with bosentan (EARLY study): a double-blind, randomised controlled trial. Lancet 2008;371:2093–100.
11. Bruni C, De Luca G, Lazzaroni MG, et al. Screening for pulmonary arterial hypertension in systemic sclerosis: a systematic literature review. Eur J Intern Med 2020;78:17–25.
12. Hsu VM, Chung L, Hummers LK, et al. Development of pulmonary hypertension in a high-risk population with systemic sclerosis in the Pulmonary Hypertension Assessment and Recognition of Outcomes in Scleroderma (PHAROS) cohort study. Semin Arthritis Rheum 2014;44:55–62.
13. Schreiber BE, Valerio CJ, Keir GJ, et al. Improving the detection of pulmonary hypertension in systemic sclerosis using pulmonary function tests. Arthritis Rheum 2011;63:3531–9.

14. Gladue H, Steen V, Alliandre Y, et al. Combination of echocardiographic and pulmonary function test measures improves sensitivity for diagnosis of systemic sclerosis-associated pulmonary arterial hypertension: analysis of 2 cohorts. J Rheumatol 2013;40:1706–11.

15. Meune C, Avouac J, Airò P, et al. Prediction of pulmonary hypertension related to systemic sclerosis by an index based on simple clinical observations. Arthritis Rheum 2011;63:2790–6.

16. Semalulu T, Rudski L, Huynh T, et al. An evidence-based strategy to screen for pulmonary arterial hypertension in systemic sclerosis. Semin Arthritis Rheum 2020;50:1421–7.

17. Coghlan JG, Denton CP, Grünig E, et al. Evidence-based detection of pulmonary hypertension in systemic sclerosis: the DETECT study. Ann Rheum Dis 2014;73:1340–9.

18. Nihtyanova SI, Schreiber BE, Ong VH, et al. Prediction of pulmonary complications and long-term survival in systemic sclerosis. Arthritis Rheumatol 2014;66:1625–35.

19. UK NPHCot, Ireland. Consensus statement on the management of systemic sclerosis without established or suspected pulmonary disease. Clin Exp Rheumatol 2018;36:115–26

20. Hughes JM, Pride NB. Examination of the carbon monoxide diffusing capacity (DL(CO)) in relation to its KCO and VA components. Am J Respir Crit Care Med 2012;186:132–9.

21. Steen V, Medsger TA Jr. Predictors of isolated pulmonary hypertension in patients with systemic sclerosis and limited cutaneous involvement. Arthritis Rheum 2003;48:516–22.

22. Steen VD, Graham G, Conte C, et al. Isolated diffusing capacity reduction in systemic sclerosis. Arthritis Rheum 1992;35:765–70.

23. Antoniou KM, Margaritopoulos GA, Goh NS, et al. Combined pulmonary fibrosis and emphysema in scleroderma-related lung disease has a major confounding effect on lung physiology and screening for pulmonary hypertension. Arthritis Rheumatol 2016;68:1004–12.

24. Gunn J, Pauling JD, McHugh NJ. Impact of anti-centromere antibodies on pulmonary function test results in patients with systemic sclerosis without established or suspected pulmonary disease. Clin Rheumatol 2014;33:869–71.

25. Daniel RM, Cousens SN, De Stavola BL, et al. Methods for dealing with time-dependent confounding. Stat Med 2013;32:1584–618.

26. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. J Clin Oncol 1983;1:710–9.

27. Van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2013;65:2737–47.

28. Simonneau G, Montani D, Celermajer DS, et al. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. Eur Respir J 2019;53:1801913.

29. Daññ U. Landmark analysis at the 25-year landmark point. Circ Cardiovasc Qual Outcomes 2011;4:363–71.

30. Van Houwelingen H, Putter H. Dynamic prediction in clinical survival analysis. CRC Press 2012.

31. Pepe MS, Mori M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? Stat Med 1993;12:737–51.

32. Lau EM, Humbert M, Celermajer DS. Early detection of pulmonary arterial hypertension [review]. Nat Rev Cardiol 2015;12:143–55.

33. Mazurek JA, Vaidya A, Mathai SC, et al. Follow-up tricuspid annular plane systolic excursion predicts survival in pulmonary arterial hypertension. Pulm Circ 2017;7:361–71.

34. McLaughlin WV, Hoeper MM, Channick RN, et al. Pulmonary arterial hypertension-related morbidity is prognostic for mortality. J Am Coll Cardiol 2018;71:752–63.

35. Sitbon O, Channick R, Chin KM, et al. Sildenafil for the treatment of pulmonary arterial hypertension. N Engl J Med 2015;373:2522–33.

36. Pulido T, Adzerikho I, Channick RN, et al. Macitentan and morbidity and mortality in pulmonary arterial hypertension. N Engl J Med 2013;369:809–18.

37. Wiauxman A, Restrepo-Jaramillo R, Thenappan T, et al. Inhaled treprostinil in pulmonary hypertension due to interstitial lung disease. N Engl J Med 2021;384:325–34.

38. Yamamoto K, Tanabe N, Takahashi Y, et al. Characteristics of patients meeting the new definition of pre-capillary pulmonary hypertension (Nice 2018) in a single Japanese pulmonary hypertension centre. BMC Pulm Med 2021;21:260.

39. Xanthouli P, Jordan S, Milde N, et al. Haemodynamic phenotypes and survival in patients with systemic sclerosis: the impact of the new definition of pulmonary arterial hypertension. Ann Rheum Dis 2020;79:370–8.

40. Jaafar S, Visovatti S, Young A, et al. Impact of the revised haemodynamic definition on the diagnosis of pulmonary hypertension in patients with systemic sclerosis. Eur Respir J 2019;54:1900586.

41. Steen V, Domisc RC, Lucas M, et al. A clinical and serological comparison of African American and Caucasian patients with systemic sclerosis. Arthritis Rheum 2012;64:2986–94.

42. Otero CM, Assaifi S, Hudson M, et al. Antibfibrilfan antibodies are associated with Native North American Ethnicity and poorer survival in systemic sclerosis. J Rheumatol 2017;44:798–805.

43. Mendes C, Viana VS, Pasoto SG, et al. Clinical and laboratory features of African-Brazilian patients with systemic sclerosis. Clin Rheumatol 2020;39:9–17.

44. Goh NS, Hoyles RK, Denton CP, et al. Short-term pulmonary function trends are predictive of mortality in interstitial lung disease associated with systemic sclerosis. Arthritis Rheumatol 2017;69:1670–8.