Exercise SpO2 Accurately Reflects SaO2 and Predicts Mortality in Systemic Sclerosis

Jeffrey J. Swigris¹
Xianmei Zhou¹
Fred S. Wamboldt²
Roland du Bois¹
Rebecca Keith¹
Aryeh Fischer¹
Gregory P. Cosgrove¹
Stephen K. Frankel¹
Doug Curran-Everett³
Kevin K. Brown¹

National Jewish Health
Interstitial Lung Disease Program and Autoimmune Lung Center¹
Division of Psychosocial Medicine²
Division of Biostatistics and Bioinformatics³

Running Title: SpO2 in ILD
Word Count: 2274

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in Thorax and any other BMJPLG products to exploit all subsidiary rights, as set out in our licence (http://thorax.bmj.com/ifora/licence.pdf).

Author Correspondence:
Jeffrey J. Swigris, DO, MS
Assistant Professor of Medicine
Interstitial Lung Disease Program and Autoimmune Lung Center
National Jewish Medical and Research Center
1400 Jackson Street
Denver, Colorado 80206
Phone: (303) 398-1621
Fax: 303-398-1040
email: swigrisj@njc.org
ABBREVIATIONS
DLCO% = percent predicted diffusing capacity for carbon monoxide
ESHL = end stage honeycomb lung
FVC% = percent predicted forced vital capacity
NSIP = nonspecific interstitial pneumonia
SSc = systemic sclerosis
TLC% = percent predicted total lung capacity
UIP = usual interstitial pneumonia
KEY WORDS: lung diseases, interstitial; systemic sclerosis; pulmonary fibrosis; exercise test; exercise tolerance; oxygenation
ABSTRACT

Background: Measures of oxygenation have not been assessed for prognostic significance in systemic sclerosis-related interstitial lung disease (SSc-ILD).

Methods: We identified 83 subjects with SSc-ILD who performed a maximal cardiopulmonary exercise test (CPET) with arterial line at our center. We examined agreement between peripheral (SpO2) and arterial oxygen saturation (SaO2). Next, we analyzed survival differences between subgroups of subjects stratified on SpO2. Finally, we used Cox proportional hazards analyses to examine the prognostic capabilities of SpO2.

Results: At maximal exercise, the SpO2-SaO2 difference was 2.98 ± 2.98 and only 15 subjects had a SpO2-SaO2 difference > four points. The survival of SSc-ILD subjects whose maximum exercise SpO2 (SpO2_{MAX}) fell below 89% or whose SpO2_{MAX} fell > 4 points from baseline was worse than subjects in comparator groups (log-rank p = 0.01 and 0.01 respectively). The hazard of death during the median 7.1 years of follow-up was 2.4 times greater for subjects whose SpO2_{MAX} fell below 89% (hazard ratio [HR] = 2.4, 95% confidence interval [CI] 1.1-4.9, p=0.02) or whose SpO2_{MAX} fell > 4 points from baseline (HR = 2.4, CI 1.1-5.0, p=0.02).

Conclusion: Among patients with SSc-ILD, SpO2 is an adequate reflection of SaO2, and radial arterial lines need not be inserted during CPET in these patients. Given the ease of measurement and its prognostic value, SpO2 should be considered as a meaningful clinical and research outcome in patients with SSc-ILD.
INTRODUCTION

Measurement of static pulmonary physiology,\textsuperscript{1-4} and more recently, assessments of high-resolution computed tomography scans (HRCT),\textsuperscript{5} have been shown to provide important prognostic information about patients with fibrosing interstitial lung disease related to systemic sclerosis (SSc-ILD). For patients with idiopathic fibrosing ILD, assessments of blood oxygenation, particularly those done while patients are exerting, have also been shown to predict outcome.\textsuperscript{6-8}

Measures of peripheral oxygen saturation (e.g., SpO2) are used to estimate arterial blood oxygenation (e.g., SaO2); however, the validity of the relationship between SpO2 and SaO2 is dependent on a number of factors, including the adequacy of peripheral perfusion.\textsuperscript{9} Given recent data questioning the reproducibility of the level of desaturation (as measured by SpO2) over short time intervals in patients with idiopathic pulmonary fibrosis (IPF)\textsuperscript{10} but continued enthusiasm to use this easily obtainable measure in patients with fibrosing ILD, we sought to investigate SpO2 in patients with SSc-ILD—a disease in which peripheral circulation is often impaired, an impairment considered to preclude reliable interpretation of peripheral oxygen saturation assessments. We performed this study to test two hypotheses: 1) SpO2 would inaccurately reflect SaO2 at rest, and the inaccuracy would be even greater during maximal exercise; and 2) despite the hypothesized inaccuracy, peripheral measures of blood oxygenation during exertion should provide significant prognostic information in patients with SSc-ILD. Thus besides assessing the utility of SpO2 as a prognostic marker, a goal of the study was to determine whether an arterial line is needed to accurately assess exertional oxygenation in patients with SSc-ILD.

METHODS

Subjects

We identified 83 patients with SSc-ILD and the absence of signs of pulmonary hypertension on physical examination who were evaluated in the ILD Program at National Jewish Health with pulmonary function tests and a maximal cardiopulmonary test between 1983 and 2005.

All subjects were evaluated clinically by using a standard protocol focused on the identification of the cause of ILD. Patients with SSc met diagnostic criteria adopted by the American College of Rheumatology\textsuperscript{11}; those with SSc sine scleroderma (ssSSc) met criteria suggested by Poormoghim and colleagues.\textsuperscript{12} Patients with overlap syndromes were excluded. In patients with SSc, the diagnosis of ILD was made by surgical biopsy (n=17), chest radiograph (n=60), or computed tomography (n=4).

Maximal Cardiopulmonary Exercise Test

All subjects underwent maximal cardiopulmonary exercise testing at our institution according to a standardized protocol. A radial arterial line and a peripheral pulse oximeter were placed prior to commencing exercise. The peripheral pulse oximeter was placed on the index finger of the hand opposite the arterial line. If an adequate pulse oximeter signal was not obtained (inability to obtain correct pulse), an earlobe probe was used. In greater than 95\% of subjects, finger probes were used. Baseline measurements were collected after the patient mounted the cycle ergometer (Vmax 29; SensorMedics Corp.; Yorba Linda, CA) and just prior to beginning pedaling. Subjects pedaled for three minutes at 60 revolutions per minute, and then work was incrementally added every minute; the goal was to reach a subject’s maximal exercise capacity within six to twelve minutes. Blood was drawn from the arterial line at baseline and after every minute of exercise, and it was analyzed with a co-oximeter (to measure the percentage of arterial oxygen saturation—SaO2) and blood gas analyzer.

Statistical Analysis

Categorical data are presented as counts or percentages. Continuous data are presented as means with standard deviations or medians with interquartile ranges. We used graphical methods, including Bland
Altman plots, to display agreement between SpO2 and SaO2. Here, Bland Altman plots graph, for each subject, the average of the SpO2 and SaO2 at maximal exercise against the difference between the SpO2 and SaO2 at maximal exercise. In certain analyses, we used a SpO2 - SaO2 difference greater than four percentage points as the variable of interest. Four percentage points was chosen as the difference value of significance, because this difference falls outside of the intrinsic variability of the pulse oximeters used and is felt to be clinically relevant. We used the product-limit method to estimate survival probabilities and the Kaplan-Meier method to generate survival curves that were compared by using the log rank test. Cox proportional hazard models were used to examine the prognostic capabilities of SpO2. We confirmed the proportionality assumption was met for the dichotomized SpO2 variables by examining log(-log) plots. All statistical analyses were performed by using SAS, version 9.1 (SAS Institute; Cary, N.C.). We considered a p value < 0.05 to be statistically significant.

RESULTS
Baseline characteristic of subjects are displayed in Table 1. SpO2 overestimated SaO2 both at rest and

Table 1. Demographic and Clinical Characteristics of Subjects

<table>
<thead>
<tr>
<th>SS-c-ILD (N=83)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Age in yrs</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Smoking status</td>
</tr>
<tr>
<td>Current</td>
</tr>
<tr>
<td>Former</td>
</tr>
<tr>
<td>Never</td>
</tr>
<tr>
<td>Pulmonary physiology</td>
</tr>
<tr>
<td>Rest FVC%</td>
</tr>
<tr>
<td>Rest DLCO%</td>
</tr>
<tr>
<td>Surgical biopsy</td>
</tr>
<tr>
<td>Pattern</td>
</tr>
<tr>
<td>UIP</td>
</tr>
<tr>
<td>NSIP</td>
</tr>
<tr>
<td>ESHL</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Exercise variables</td>
</tr>
<tr>
<td>Oxygenation</td>
</tr>
<tr>
<td>Baseline SpO2</td>
</tr>
<tr>
<td>Baseline SaO2</td>
</tr>
<tr>
<td>Baseline PaO2</td>
</tr>
<tr>
<td>Max Ex. SpO2</td>
</tr>
<tr>
<td>Max Ex. SaO2</td>
</tr>
<tr>
<td>Max. Ex. PaO2</td>
</tr>
<tr>
<td>Exercise capacity</td>
</tr>
<tr>
<td>VO2_MAX in L/min</td>
</tr>
<tr>
<td>VO2_MAX%</td>
</tr>
<tr>
<td>Work in watts</td>
</tr>
<tr>
<td>Work%</td>
</tr>
</tbody>
</table>
maximal exercise. The average SpO2 - SaO2 difference (keeping overestimates as positive values and underestimates as negative values) was similar at maximal exercise and rest (1.5 ± 4 vs. 1.5 ± 3, p=1) (Figure 1). Results were similar when the absolute values of the SpO2 - SaO2 difference were used (2.24 ± 1.91 vs. 2.98± 2.98, p=0.06). SpO2 misclassified four subjects at maximum exercise: SpO2 was observed to be > 88% but SaO2 was ≤ 88%. For these subjects, median values for baseline SpO2, SaO2, and PaO2, and maximum exercise SpO2, SaO2, and PaO2 were 91.5 (89-96), 88.5 (88-90), 61 (59.5-67), and 89.5 (89-93), 82 (79-88), and 47.5 (43.5-62.5).

Over the study period (median follow-up 10.3 years, interquartile range [IQR] 4-17 years), there were 40 deaths. Truncating follow-up at 20 years, we observed 39 deaths over the study period. Median survival was 9.5 years, IQR 4-16 years. Subjects whose SpO2 at maximum exercise (SpO2_MAX) fell to < 89% had shorter survival than subjects whose SpO2_MAX remained ≥ 89% (log-rank p=0.01) (Figure 2). Results were similar when stratifying subjects on whether SpO2_MAX fell > four points from baseline (log-rank p=0.01). In Cox proportional hazards models, SpO2 was a significant predictor of mortality: over a median 7.1 years of follow-up, the risk of death was 2.4 times greater for subjects whose SpO2_MAX fell below 89% (HR = 2.4, CI 1.2-4.9 p=0.02) than for subjects whose SpO2_MAX remained 89% or greater. Similarly, the risk of death was 2.4 times greater for subjects whose SpO2_MAX fell > four points from baseline (HR = 2.4, CI 1.1-5.0, p=0.02) than for subjects whose SpO2_MAX remained within four points of baseline values. When analyzed as a continuous variable, the difference between baseline SpO2 and SpO2_MAX remained a significant predictor (HR = 1.07, CI 1.01-1.14, p=0.02). Figure 3 shows the relationship between DLCO% and SpO2_MAX.

DISCUSSION
With two goals in mind—1) to assess the utility of SpO2 as a prognostic marker in patients with SSc-ILD, and 2) to determine whether an arterial line is needed to accurately assess oxygenation—we conducted a study to first examine agreement between SpO2 and SaO2 at rest and maximal exercise and then analyze the ability of SpO2 to predict mortality in SSc-ILD. We hypothesized that in subjects with SSc-ILD, SpO2 would inaccurately reflect SaO2 at rest and the disparity would be even greater at maximal exercise. In contrast, we found that SpO2 was an accurate reflection of SaO2 both at rest and maximal exertion in these subjects. Moreover, we observed that SpO2, a simple, noninvasive, inexpensive measure to collect, was a predictor of mortality in patients with SSc-ILD.

Recently, there has been a groundswell of attention on and use of noninvasive markers of exertional blood oxygenation (e.g., nadir SpO2 during a 6-minute walk test (6MWT) or statistical manipulations of SpO2 over the course of a timed walked test) as outcome measures in therapeutic trials and clinical studies enrolling subjects with ILD.7,14,15 This increased attention raises three important distinct but related questions regarding the use of SpO2 at maximal exercise as an outcome metric: 1) is it valid—does it in fact measure what it is purported to measure (e.g., true blood oxygenation, or SaO2)?; 2) is it reliable—if it is measured at two separate time points in a subject whose clinical status has not changed, will it produce similar results?; and 3) is it responsive to underlying change—if a subject’s blood oxygenation at maximal
exercise changes from baseline, will SpO2 reflect those changes? The current study shows that SpO2 at maximal exercise is a valid measure of blood oxygenation at maximal exercise in patients with SSc-ILD, and SpO2 does in fact accurately track changes in SaO2.

The Bland Altman plots reinforce this finding. These plots give a graphical presentation of the agreement between two methods of measurement; they depict an estimate of the bias (or systematic error, which is simply the over- or underestimation of one measure compared with the other) as the mean difference between the two measures. The precision of that estimate is reflected in its standard deviation. Whereas correlation coefficients express the relationship between two variables, Bland Altman plots depict agreement between them. When one is trying to determine the accuracy with which one measure (e.g., SpO2) reflects another (e.g., SaO2), or whether one measure might be used in place of another measure, correlation may not tell the true story—there can be extremely high correlation between two measures but, at the same time, poor agreement. This study shows that, for patients with SSc-ILD, SpO2 is an accurate reflection of SaO2 at rest or maximal exercise.

Several studies have examined the agreement between SpO2 and SaO2, but, to our knowledge, this is the first in a cohort with SSc-ILD. The importance and clinical relevance of this study centers on the peripheral circulation issues in SSc that make most clinicians reluctant to place an arterial line (digits have been lost as a consequence) and wary of SpO2 accuracy in these patients. Thus, there is a need to validate SpO2 in SSc-ILD. In general, SpO2 may either over- or underestimate SaO2; in a meta-analysis, Jensen and colleagues reported that among 23 studies for which bias and precision estimates were available, the absolute mean bias was 1.99 ± 0.23 (i.e., on average, SpO2 overestimated SaO2 by 1.99 points). In those studies, the mean SpO2 - SaO2 difference ranged from -13.2 ± 8.0 to 12.0 ± 13.3. The authors commented that severe or rapid desaturation; hypotension, hypothermia, or other unstable hemodynamic or low perfusion states; dyshemoglobinemia or use of vital dyes; and motion may all confound agreement between SpO2 and SaO2. The mean SpO2 - SaO2 differences in the current study fall well within the range mentioned in that analysis.

The results of the current study not only suggest that SpO2 is a valid surrogate for SaO2 in patients with SSc-ILD, they suggest that, desaturation as measured by SpO2, is a significant predictor of mortality in this patient group. Our results are in line with the work by Lama and her colleagues that suggested desaturation (as measured by SpO2) during a 6MWT is an important prognostic indicator in patients with idiopathic interstitial pneumonia (IIP). In so far as the 6MWT accurately reflects functional exercise capacity in patients with SSc-ILD—it does so in patients with fibrotic IIP—we hypothesize our results would hold for measures of SpO2 collected during 6MWT in this patient population. Not surprisingly, we found DLCO% to be a strong driver of SpO2MAX (data not shown)—in fact, among several candidate variables including age, gender, FVC%, and baseline SpO2, DLCO% was the only significant predictor. Like other investigators, we also found DLCO% to be a potent predictor of survival in our cohort (data not shown). Because of the strong relationship between DLCO% and SpO2MAX, and because our goal was merely to begin to examine SpO2MAX as a prognostic marker, we performed our survival analysis adjusting for FVC% and not DLCO%.

Although results are novel and clinically relevant, this study has limitations, including that this is a retrospective analysis of data collected prospectively over a period of three decades. Different pulse and co-oximeters were used during different time periods; however, each instrument is purported to be accurate within two percentage points for SaO2 values from 70-100%, thus, we can be confident in the readings. Data for this study were collected at a center situated 5280 feet above sea level. In Denver, patients probably “live” closer to the steep portion of the oxygen dissociation curve—likely on or very close to the shoulder—than patients at lower altitudes. How this affects results merits consideration and examination in
future studies. Given the lack of systematic examinations for pulmonary hypertension and the changes in available technology to assess for PH over the study period, we can not be certain how many subjects truly had PH. Even more complex is the issue of exercise-induced PH: how many subjects had it is unknown, but as with other studies of subjects with ILD, the possibility of its presence and its effects on exercise SpO2 must be considered. Given these limitations, we believe the results should be viewed as hypothesis-generating and will hopefully spark continued investigation in this area. Moving forward, these results will need prospective confirmation (at other altitudes). Future studies should examine whether SpO2 values collected during 6MWT are as meaningful as those collected during CPET, and efforts should be made to further delineate the relationship between resting or exercise-induced PH and SpO2.

CONCLUSION
In SSc-ILD, at both baseline and maximal exercise, SpO2 is an accurate reflection of SaO2. In patients with SSc-ILD, SpO2 carries prognostic value; because of the ease with which it is assessed, consideration should be given to measuring exercise SpO2 as a marker of clinical status or as an outcome in clinical trials enrolling subjects with SSc-ILD. Future research could clarify a number of outstanding and important questions related to SpO2 in this patient population.
REFERENCES
Figure 1. Plots of SpO2 vs. SaO2 and Bland Altman Plots for Subjects with SSc-ILD at Baseline and Maximum Exercise.

PANEL A. Baseline.

In the upper graph, the diagonal line marks the line of unity. In the lower graph, the middle line marks the mean difference between SpO2 and SaO2, and the two outer lines mark twice the standard deviation in either direction.
PANEL B. Maximum Exercise.

In the upper graph, the diagonal line marks the line of unity. In the lower graph, the middle line marks the mean difference between SpO2 and SaO2, and the two outer lines mark twice the standard deviation in either direction.

In the upper graph, the diagonal line marks the line of unity. In the lower graph, the middle line marks the mean difference between SpO2 and SaO2, and the two outer lines mark twice the standard deviation in either direction.
Figure 2. Kaplan-Meier Survival Curves for SSc-ILD Subjects Stratified on SpO2_{MAX} < 89% or ≥ 89%

Solid line represents subjects whose SpO2_{MAX} remained ≥ 89%. Dashed line for subjects whose SpO2_{MAX} fell to < 89%. Tic = censored observation. Median survival for the entire cohort = 9.5 years.
Figure 3. Relationship Between SpO2 at Maximum Exercise and Resting DLCO%.

Maximum Exercise SpO2 %

DLCO%