Intersession Variability in Single-Breath Diffusing Capacity in Diabetics without Overt Lung Disease

Michael B. Drummond[1], Pamela F. Schwartz[2], William T. Duggan[2], John G. Teeter[2], Richard J. Riese[2], Richard C. Ahrens[3], Robert O. Crapo[4], Richard D. England[2], Neil R. MacIntyre[5], Robert L. Jensen[4], and Robert A. Wise[1]

[1]Division of Pulmonary and Critical Care Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland; [2]Pfizer Global Research and Development, New London, Connecticut; [3]University of Iowa College of Medicine, Iowa City, Iowa; [4]University of Utah, and Pulmonary Division, LDS Hospital, Salt Lake City, Utah; and [5]Duke University Medical School, Durham, North Carolina

Rationale: American Thoracic Society guidelines state that a 10% or greater intersession change in diffusing capacity of the lung ($D_{L_{CO}}$) should be considered clinically significant. However, little is known about the short-term intersession variability in $D_{L_{CO}}$ in untrained subjects or how variability is affected by rigorous external quality control.

Objectives: To characterize the intersession variability of $D_{L_{CO}}$ and the effect of different quality control methods in untrained individuals without significant lung disease.

Methods: Data were pooled from the comparator arms of 14 preregistration trials of inhaled insulin that included nonsmoking diabetic patients without significant lung disease. A total of 699 participants performed repeated $D_{L_{CO}}$ measurements using a highly standardized technique. A total of 948 participants performed repeated measurements using routine clinical testing.

Measurements and Main Results: The mean intersession absolute change in $D_{L_{CO}}$ using the highly standardized method was 1.45 ml/minute/mm Hg (5.64%) compared with 2.49 ml/minute/mm Hg (9.52%) in the routine testing group ($P < 0.0001$ for both absolute and percent difference). The variability in absolute intersession change in $D_{L_{CO}}$ increased with increasing baseline $D_{L_{CO}}$ values, whereas the absolute percentage of intersession change was stable across baseline values. Depending on the method, 15.5 to 35.5% of participants had an intersession change of 10% or greater. A 20% or greater threshold would reduce this percentage of patients to 1 to 10%.

Conclusions: Intersession variability in $D_{L_{CO}}$ measurement is dependent on the method of testing used and baseline $D_{L_{CO}}$. Using a more liberal threshold to define meaningful intersession change may reduce the misclassification of normal variation as abnormal change.

Keywords: respiratory function testing; diffusing capacity; intersession variability; inhaled human insulin; methodology

Measurement of single-breath diffusing capacity of the lung ($D_{L_{CO}}$) is an integral component of the assessment of lung function in many diseases (1–5). As inhaled medications for nonpulmonary diseases are developed, $D_{L_{CO}}$ will play a more important role in monitoring for early lung toxicity in individuals. This assessment requires adequate characterization of the expected intersession variability in $D_{L_{CO}}$ in individuals without lung disease.

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

Intersession variability in single-breath diffusing capacity is poorly characterized in previous studies.

What This Study Adds to the Field

Our study shows that the current criterion of 10% for determining clinical change in diffusing capacity may be too sensitive, and we suggest a 20–25% threshold as more appropriate.

Intersession variability in $D_{L_{CO}}$ measurements exists as a consequence of biological and measurement variability (6, 7). Current American Thoracic Society/European Respiratory Society (ATS/ERS) guidelines state that a change of more than 10% in $D_{L_{CO}}$ over 1 year should be considered clinically significant (8). This value is derived from the work of Hathaway and colleagues, who reported on the intraindividual variability in five $D_{L_{CO}}$ measurements in eight healthy trained pulmonary function technicians over 1 year (9). Additional variability can be introduced by using different testing centers to serially monitor an individual. Wanger and Irvin have shown that healthy trained subjects had a substantial variation in measured $D_{L_{CO}}$ when tested at 13 pulmonary function laboratories in a major metropolitan center (10). To minimize this measurement variability, published guidelines provide standard methods for $D_{L_{CO}}$ measurement and quality control (8, 11). Although standard testing methods improve reproducibility within a single session at a single center (12), there are few data about the intersession reproducibility in $D_{L_{CO}}$ measurements in untrained individuals free of substantial lung disease. Moreover, it is unclear how much intersession reproducibility can be improved by instituting a centralized structured program of laboratory quality control.

Recently, a series of multicenter trials of inhaled human insulin (Exubera [insulin human (rDNA origin)] Inhalation Powder; Pfizer, Inc., New York, NY; and Nektar Therapeutics, San Carlos, CA) was completed in individuals with type 1 and 2 diabetes (13–22). Serial $D_{L_{CO}}$ measurements were obtained in participants receiving inhaled insulin and in comparator participants receiving no inhaled therapy. In the initial drug safety trials, lung function was measured in community pulmonary function laboratories (hereafter referred to as "routine testing"). In later trials, a comprehensive program of centralized quality control monitoring was instituted (referred to as "highly standardized testing") (23–27). Individuals enrolled in the comparator groups did not receive inhaled therapies during the study protocol.

The present analyses used $D_{L_{CO}}$ measurements from the comparator groups of these preregistration trials of inhaled insulin to address the following questions: What is the mean intersession difference in $D_{L_{CO}}$ measurements obtained using routine as compared with highly standardized testing in individuals free of overt lung disease? Is the magnitude of intersession difference best described as absolute change or percentage of change from baseline $D_{L_{CO}}$? Are the intersession differences observed in absolute and percent changes affected by the baseline $D_{L_{CO}}$ level? Is the current ATS recommendation of a 10% threshold to define clinically significant change appropriate for an untrained population free of overt lung disease?

METHODS

Study Participants

Data for this analysis were pooled from a database of participants enrolled as control subjects in 14 preregistration trials of inhaled insulin (Table 1). All participants were individuals with type 1 or type 2 diabetes using either subcutaneous insulin or oral hypoglycemic medications. Individuals enrolled in type 1 diabetes studies ranged from 18 to 65 years of age, whereas individuals enrolled in type 2 diabetes studies ranged from 35 to 75 years of age. Individuals with active lung disease, self-reported smoking within 6 months, previous abnormal lung function, inability to perform pulmonary function tests, or major organ system disease were excluded. Individuals were screened for lung disease with physical examination, chest radiograph, and pulmonary function testing (PFT). Individuals with stable mild asthma (determined by self-report) or mild chronic obstructive pulmonary disease (COPD) (ratio of FEV_1 to FVC <70%, FEV_1 ≥70% predicted, and a history of smoking) were permitted to participate in these studies. Participants underwent baseline and serial PFT.

Monitoring and Measurements

In the preregistration trials, individual participants performed $D_{L_{CO}}$ measurements at baseline and serially for 3 months to 3 or more years, depending on the specific trial protocol. For this analysis, the baseline and the last measurement at or before the Month 6 visit were used. Individuals underwent PFT measurements using either a highly standardized or routine approach. Methods for the highly standardized approach have been previously described (23). In brief, all sites used the same type of lung function analyzer (Collins CPL; Ferraris Respiratory, Louisville, CO), and all analyzers were calibrated and leak-tested daily. Technicians, many of whom had no previous experience in performing PFTs, underwent a 2-day training course followed by a written test and performance evaluation. For all PFT measurements in the highly standardized method, the prediction equations of Miller and

colleagues (28), Crapo and colleagues (29), and Hankinson and colleagues (30) were used for $D_{L_{CO}}$, total lung capacity (TLC), and FEV_1, respectively. A 12% adjustment in TLC and $D_{L_{CO}}$ predicted values was applied for African-American participants. All study data were transmitted to a reading center (Ferraris Respiratory), where they were reviewed within 24 hours for test quality, and feedback was given to technicians.

The pulmonary function laboratories using the routine testing approach were all community and hospital pulmonary function laboratories. By protocol, these testing laboratories were required to follow published guidelines for performance and reproducibility as recommended by the ATS (11). However, formal technician training and central review of results were not performed at these sites. For calculation of predicted values, the prediction equations in place at the particular PFT lab were used. In the routine testing group, performing serial testing at different pulmonary function laboratories or on different pieces of equipment was possible. Although not specifically tracked, testing at a different site was a very rare event insofar as patients were monitored at the same clinical site throughout. In both the highly standardized and routine testing approach, values of $D_{L_{CO}}$ were corrected for hemoglobin concentration per ATS guidelines (11).

Statistical Analysis

Baseline clinical and demographic characteristics were compared between groups using t tests for continuous variables and Fisher's exact test for categorical variables. Box-and-whisker plots of intersession differences between the first and 3- or 6-month session measurements were generated. For each subject, the intersession change was calculated as the second session $D_{L_{CO}}$ value minus the first (baseline) $D_{L_{CO}}$ value. The percentage of intersession change was calculated as the second session $D_{L_{CO}}$ value minus the baseline $D_{L_{CO}}$ value divided by the baseline session $D_{L_{CO}}$ value, multiplied by 100%. For both calculations, the absolute value was represented (i.e., always positive, irrespective of sign). Analysis of variance (ANOVA) was performed to compare mean intersession changes between different categories of baseline $D_{L_{CO}}$. Cumulative frequency distributions were generated to reflect the percentage of participants with various magnitudes of $D_{L_{CO}}$ difference between baseline and final $D_{L_{CO}}$. Kolmogorov-Smirnov analysis was performed to evaluate differences between cumulative frequency distributions. Stata version 10.0 (Stata Corp, College Station, TX), Splus version 7.0.6 for Microsoft Windows (Microsoft Corp., Redmond, WA), and SAS release 8.2 (SAS Institute, Cary, NC) were used for statistical calculations.

RESULTS

Baseline Demographic and Clinical Characteristics

A total of 739 participants were enrolled in four studies using the highly standardized procedure. Of these, 699 participants had

TABLE 1. STUDIES CONTRIBUTING TO $D_{L_{CO}}$ ANALYSIS

Study No.	ClinicalTrials.gov Registration	Diabetes Type	No. of Participants	Comparator	Study Date	Reference No.
			Highly Standardized			
1022	NCT00137046	1	290	SC Insulin	5/02–Present	24
1026	N/A	1	22	SC Insulin	4/02–4/03	25
1027	N/A	1	116	SC Insulin	4/02–10/03	27
1029	NCT00136916	2	311	SC Insulin	6/02–Present	26
			Routine Testing			
102	N/A	1	37	SC Insulin	12/96–9/97	17
103	N/A	2	28	SC Insulin	11/96–9/97	18
104	N/A	2	36	Oral Agents	10/97–7/98	19
106	NCT00424437	1	135	SC Insulin	9/99–9/00	14
107	NCT00424333	1	105	SC Insulin	5/99–10/00	15
108	NCT00424411	2	149	SC Insulin	9/99–12/00	13
109	NCT00370565	2	99	Oral Agents	6/99–9/00	16
110	N/A	2	68	Oral Agents	10/99–3/01	20
1001	N/A	2	210	Oral Agents	2/00–2/04	21
1002	N/A	2	231	Oral Agents	2/00–2/04	22

Definition of abbreviations: N/A = not applicable; SC = subcutaneous.

both baseline and follow-up $D_{L_{CO}}$ measurements within 6 months (Table 2). For 82% of these individuals, the first and last $D_{L_{CO}}$ measurements were 6 months apart, whereas 18% had a 3-month interval between the two tests. A total of 1,098 participants were included in 10 studies using the routine testing procedures. Of these, 948 had both baseline and follow-up $D_{L_{CO}}$ measurements within 6 months. For 76% of the participants in this group, 6 months elapsed between the first and last test, whereas 24% of participants had a 3-month interval between tests. Sex and smoking status were not significantly different between the highly standardized and routine testing groups. Subjects in the highly standardized group were slightly younger (45 vs. 52 yr), had a lower body mass index (27.2 vs. 29.1 kg/m^2), and more frequently had type 1 diabetes (58 vs. 25%). The mean age for all type 1 subjects was 38.0 ± 10.9 years. For the subjects with type 2 diabetes included in this analysis, the mean age was 56.16 ± 10.1 years. Of the 699 individuals in the highly standardized group, 36 (5.4%) had mild stable asthma ($n = 9$) or COPD ($n = 27$). In the routine testing group, 38 (4.0%) of the 948 individuals with this information available had mild asthma ($n = 19$) or COPD ($n = 19$). Absolute $D_{L_{CO}}$ measurements at baseline were similar for the two groups (25.90 ± 6.26 ml/min/mm Hg for the highly standardized group, 26.01 ± 6.52 ml/min/mm Hg for the routine

testing group). At baseline, percent predicted $D_{L_{CO}}$ was lower in the highly standardized group compared with routine testing (92.44 ± 12.6 vs. $98.31 \pm 17.16\%$ predicted, $P < 0.0001$).

Mean Intersession Change in $D_{L_{CO}}$ Measurements

The mean $\pm$ SD intersession change in $D_{L_{CO}}$ measurements for participants in the highly standardized group was -0.31 ± 1.86 ml/minute/mm Hg. For participants in the routine testing group, the mean intersession change was -0.54 ± 3.50 ml/minute/mm Hg. The mean absolute intersession change in $D_{L_{CO}}$ measurements (change expressed as positive, regardless of direction) for individuals using the highly standardized approach was smaller than in the routine testing group (1.45 vs. 2.49 ml/min/mm Hg, $P < 0.0001$) (Figure 1, *top*). The highly standardized group also had a smaller interquartile range (75th–25th percentile) than did the routine testing group (1.51 vs. 2.61 ml/min/mm Hg). The 90th percentile for mean absolute intersession change was lower in the highly standardized group than in the routine testing group (3.06 vs. 5.40 ml/min/mm Hg).

The mean intersession change was also expressed in terms of the percentage of change from baseline (final value minus baseline value, divided by baseline value, times 100%). The mean absolute percentage of intersession change for the highly stan-

TABLE 2. BASELINE DEMOGRAPHIC AND CLINICAL CHARACTERISTICS

	Highly Standardized (n = 739)	Routine Testing (n = 1,098)	P Value
Age, yr	45.14 ± 13.85	51.9 ± 12.82	<0.0001
Race, n (%)			0.005
White	608 (82)	951 (87)	
Black	36 (5)	44 (4)	
Other	95 (13)	103 (9)	
Sex, n (%)			0.66
Female	310 (41.9)	474 (43.2)	
Male	429 (58.1)	624 (56.8)	
Body mass index, kg/m^2	27.21 ± 4.25	29.14 ± 4.98	<0.0001
Diabetes type, n (%)			<0.0001
Type 1	428 (57.9)	277 (25.2)	
Type 2	311 (42.1)	821 (74.8)	
Duration of diabetes, yr			
Type 1	18.21 ± 10.59	18.28 ± 10.80	0.89
Type 2	13.67 ± 7.55	8.96 ± 6.58	< 0.0001
Hemoglobin A1C, % Hb	7.59 ± 1.08	8.97 ± 1.33	<0.0001
Smoking status, n (%)			0.97
Never	450 (60.9)	667 (60.8)	
Former	289 (39.1)	429 (39)	
Current	0 (0)	2 (0.2)	
Pack-years*	14.2 ± 16.5	N/A[†]	
Interval between PFTs, %			0.002
3 mo	18.0	24.0	
6 mo	82.0	76.0	
Spirometry			
FEV$_1$, L	3.23 ± 0.78	3.07 ± 0.79	<0.0001
FEV$_1$, % predicted	92.19 ± 11.49	97.96 ± 15.06	<0.0001
FVC, L	4.08 ± 0.98	3.81 ± 1.01	<0.0001
FVC, % predicted	92.63 ± 11.30	97.66 ± 14.47	<0.0001
FEF$_{25–75}$, L	3.12 ± 1.10	3.19 ± 1.13	0.19
FEF$_{25–75}$, % predicted	98.62 ± 10.42[‡]	94.39 ± 30.9	0.52
Lung volumes			
TLC, L	5.75 ± 1.29	5.80 ± 1.28	0.43
TLC, % predicted	94.13 ± 11.13	97.61 ± 12.90	<0.0001
RV, L	1.77 ± 0.57	1.97 ± 0.65	<0.0001
RV, % predicted	98.39 ± 13.21	99.84 ± 28.48	0.18
D$_{L_{CO}}$, ml/min/mm Hg	25.90 ± 6.26	26.01 ± 6.52	0.73
D$_{L_{CO}}$, % predicted	92.44 ± 12.6	98.31 ± 17.16	<0.0001

Definition of abbreviations: Hb = hemoglobin; PFTs = pulmonary function tests.
Values are expressed as mean $\pm$ SD or mean (% of total).
* Pack-years = (cigarette/d) * (no. of yr subject smoked)/20.
[†] Pack-years data not collected in this group.
[‡] FEF$_{25–75}$ % predicted only collected in protocol 1026 ($n = 22$).
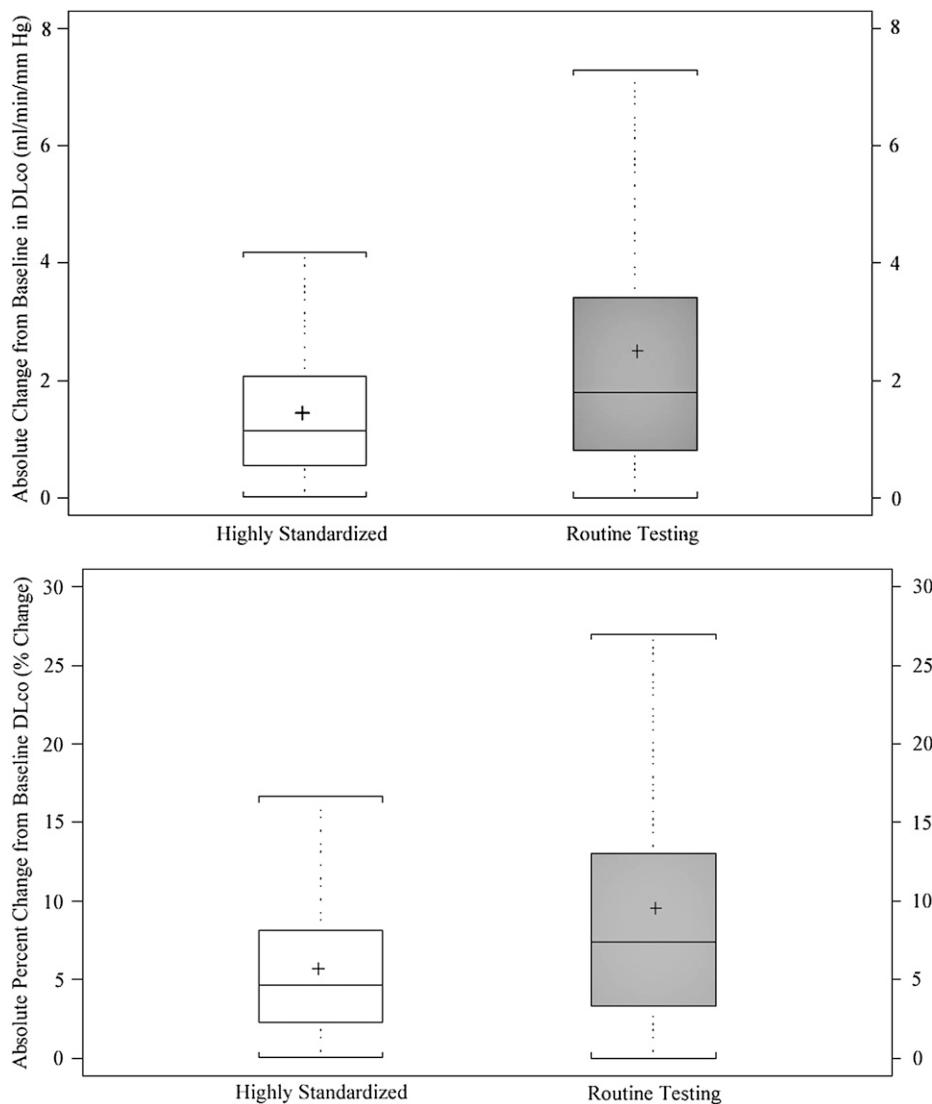
Figure 1. Box-and-whisker plots for the mean absolute intersession change in $D_{L_{CO}}$ measurement from first to final test in highly standardized (*white box*) and routine testing (*gray box*) groups. *Top panel* represents absolute change from baseline; the *bottom panel* shows percentage of change from baseline. *Boxes* represent interquartile range (IQR) bounded by the 25th and 75th percentiles with the *bar* as the median. The "+" denotes the mean value. The *whiskers* represent the distribution of values between the upper fence (75th percentile + 1.5 * IQR) and the lower fence (25th percentile − 1.5 * IQR). Values beyond the whiskers are not displayed in the graph. Change is represented as positive in direction, regardless of direction between the two tests.

dardized approach was statistically significantly smaller than that for the routine testing group (5.64 vs. 9.52%, $P < 0.0001$) (Figure 1, *bottom*). Similarly, the highly standardized group had a smaller interquartile range than did the routine testing group (5.82 vs. 9.73%). The 90th percentile for mean absolute percentage of intersession change from baseline was lower in the highly standardized group than in the routine testing group (11.3 vs. 20.0%). When the diffusing capacity was separated into components Kco ($D_{L_{CO}}$/Va) and Va, both Kco and Va components had smaller variability in the highly standardized method when compared with routine method. Both components contributed approximately equally to the observed variability.

Because the age of participants in the routine testing method was slightly older than those in the highly standardized method, we assessed the possible interaction of age on intersession variability in these groups. When analyzing mean absolute intersession change across age strata within a testing method, no difference was seen between age groups in the highly standardized ($P = 0.39$) or routine testing group ($P = 0.10$). Similarly, no difference in mean absolute percentage of intersession change was seen across age strata in the highly standardized ($P = 0.59$) or routine testing group ($P = 0.71$).

To assess the potential effect of individuals with mild asthma or COPD on the observed variability, a sensitivity analysis was performed excluding those individuals. After excluding the 5.4% of individuals in the highly standardized approach, we observed no difference in the mean absolute intersession change (1.43 ml/min/mm Hg) or interquartile range for mean absolute intersession change (1.53 ml/min/mm Hg). Similarly, there was no difference in mean absolute percentage of intersession change (5.57%) or interquartile range for mean absolute percentage of intersession change (5.89%). After excluding the 4% of participants in the routine testing group with mild asthma or COPD, we observed no difference in the mean absolute intersession change (2.50 ml/min/mm Hg) or interquartile range for mean absolute intersession change (2.65 ml/min/mm Hg). Similarly, there was no difference in mean absolute percentage of intersession change (9.52%) or interquartile range for mean absolute percentage of intersession change (10.12%).

Intersession Change in $D_{L_{CO}}$ as a Function of Baseline $D_{L_{CO}}$

The absolute intersession change varied with baseline $D_{L_{CO}}$ in both PFT methods (Figure 2, *top*). Across categories of baseline $D_{L_{CO}}$, the mean absolute change in the highly standardized group ranged from 1.13 ml/minute/mm Hg in the ⩽15 ml/minute/mm Hg category to 2.15 ml/minute/mm Hg in the >35 mL/minute/mm Hg category ($P < 0.0001$ for difference across means using ANOVA). In the routine testing group, the mean absolute change
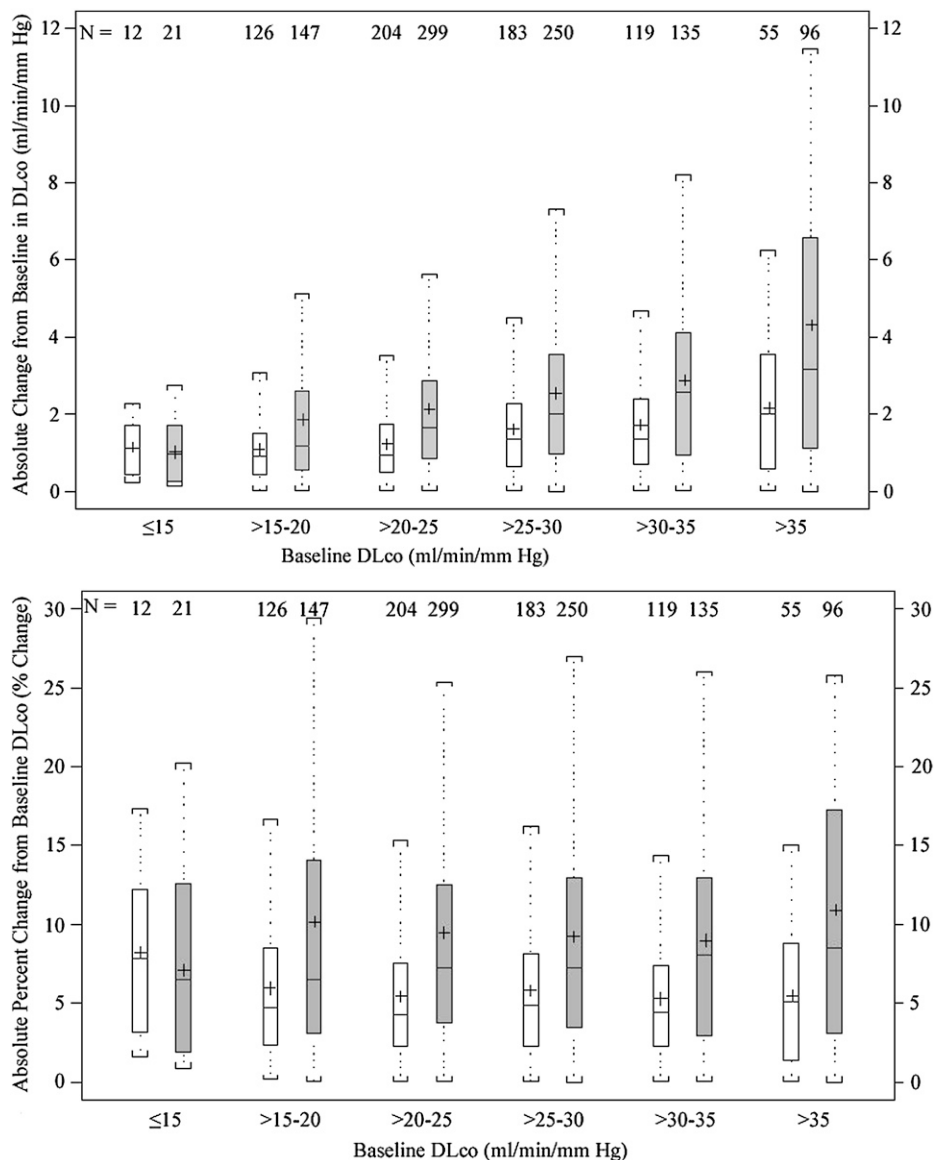
Figure 2. Box-and-whisker plots for the mean intersession change in DL_{CO} measurement from first to final test in highly standardized (*white boxes*) and routine testing (*gray boxes*) groups, binned by baseline DL_{CO} value. The *top panel* represents absolute change from baseline; the *bottom panel* represents percentage of change from baseline. *Boxes* represent interquartile range (IQR) bounded by the 25th and 75th percentiles with the *bar* as the median. The "+" denotes the mean value. The *whiskers* represent the distribution of values between the upper fence (75th percentile + 1.5 * IQR) and the lower fence (25th percentile – 1.5 * IQR). Values beyond the whiskers are not displayed in the graph. Change is represented as positive in direction, regardless of direction between the two tests.

was 1.01 ml/minute/mm Hg in the ≤15 ml/minute/mm Hg category, increasing to 4.29 ml/minute/mm Hg in the >35 ml/minute/mm Hg category ($P < 0.0001$ for difference across means).

In contrast, the mean absolute percentage of intersession change across categories of baseline DL_{CO} did not differ significantly, regardless of testing method used ($P = 0.28$ for the highly standardized group, $P = 0.39$ for routine testing) (Figure 2, *bottom*). The intersession percentage of change in the highly standardized group ranged from 5.28% in the >30–35 ml/minute/mm Hg category to 8.25% in the ≤15 ml/minute/mm Hg category. In the routine testing group, the percentage of change ranged from 7.06% in the ≤15 ml/minute/mm Hg category to 10.91% in the >35 ml/minute/mm Hg category.

Cumulative Frequency Distribution Curves

Cumulative frequency distribution curves were plotted to assist in developing criteria for intersession reproducibility. Across all ranges of absolute intersession change and absolute percentage of intersession change from baseline DL_{CO}, the highly standardized method had a tendency to exhibit less intersession variability than did the routine method ($P < 0.0001$) (Figure 3). In this study of individuals free of overt pulmonary disease, an intersession

absolute percentage of change from baseline of 10% or more was observed in 15.5% of those in the highly standardized group and 35.5% of participants in the routine testing group. Using a more conservative 15% threshold, only 3.9% of individuals in the highly standardized group and 19.7% of participants undergoing routine testing would have been classified as having a clinically meaningful change. A higher threshold of 20% reduced the percentage misclassified to 1% in the highly standardized group and 10% in the routine testing group. The table in the bottom panel of Figure 3 displays the percentages of individuals exceeding different thresholds of percentage of change from baseline DL_{CO}.

DISCUSSION

The primary finding of this study was that there is substantial intersession variability in DL_{CO} measurement in individuals free of overt lung disease and the degree of variability is dependent on the method of quality control. We found that in a large population of diabetic patients with no substantial lung disease, the 90th percentile for mean absolute intersession change could be as high as 3.06 ml/minute/mm Hg using a highly standardized approach
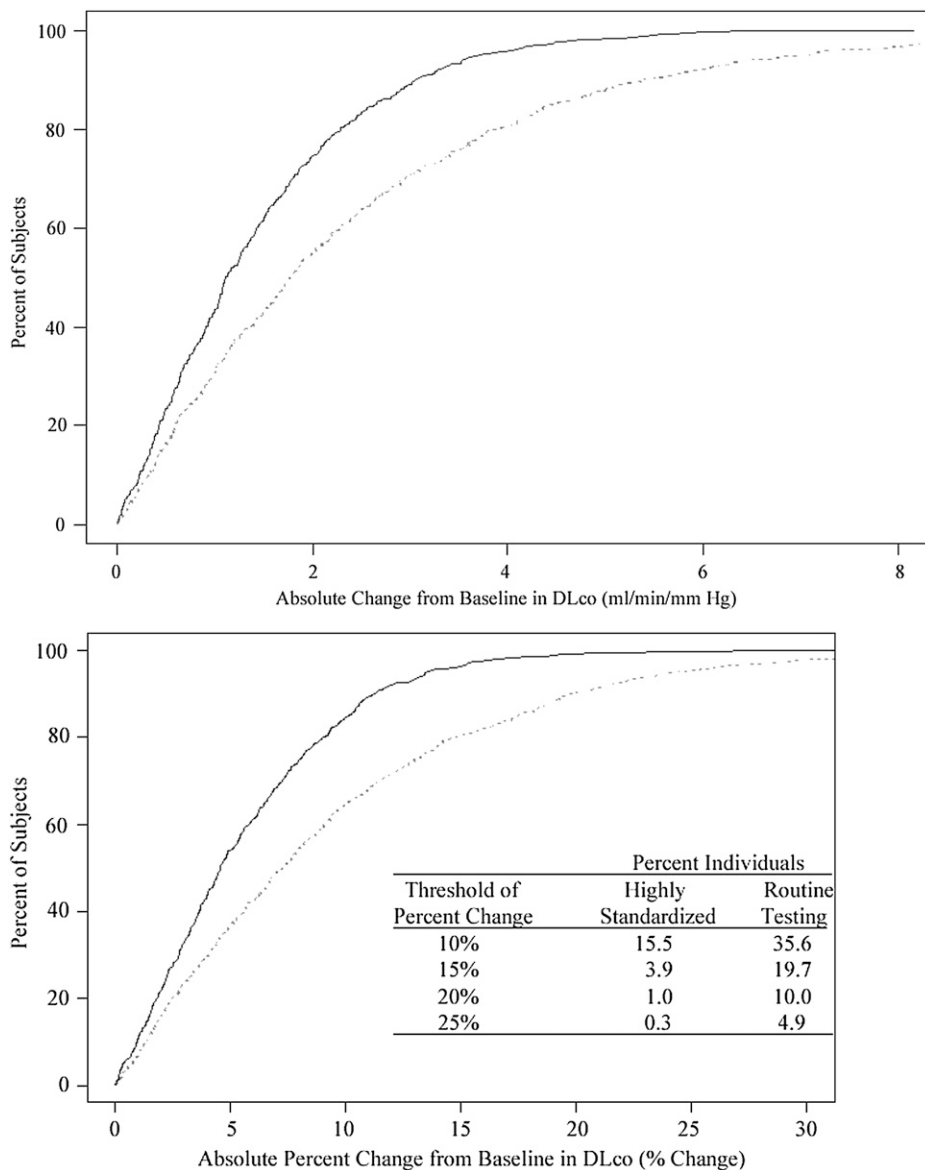
Figure 3. Cumulative frequency curve of absolute change (*top*) and absolute percent change (*bottom*) from baseline in $D_{L_{CO}}$. The highly standardized method is shown by the *solid line,* and the routine method by the *broken line.* The table (*bottom panel*) delineates the percentage of individuals exceeding different thresholds of percentage of change from baseline $D_{L_{CO}}$.

and 5.40 ml/minute/mm Hg in routine testing. The observed absolute intersession change increased with increasing baseline $D_{L_{CO}}$ value. However, when we expressed the intersession change as a function of percentage of baseline, the intersession variability did not change as a function of baseline $D_{L_{CO}}$. Because of this observation, the intersession percentage of change, rather than absolute intersession change from baseline $D_{L_{CO}}$, proved to be a more stable metric for the measure of variation in diffusion capacity.

Our analysis involved more than 1,600 participants with diabetes, the majority of whom were free of any lung disease. Before this analysis, the available data regarding the intersession variability in healthy individuals were limited to measurement of small numbers of trained individuals at one or a few centers (9, 10). The current guidelines for interpretation of clinically important changes in $D_{L_{CO}}$ measurement over time are limited by these small numbers of measurements. Well-characterized normative ranges of variability will become essential as new inhaled pharmacotherapies requiring serial $D_{L_{CO}}$ monitoring in untrained patients without underlying lung disease are developed. The data in this study will assist physicians in determining whether an observed change in $D_{L_{CO}}$ measurement is likely to

be within observed variability in a population of healthy individuals not exposed to potential toxic agents.

Although it may be desirable for pulmonary function laboratories to develop more rigorous centralized quality control procedures, such as those used in the highly standardized group, this may not be realistic in clinical practice. More importantly, the currently recommended 10% value for determining clinically significant change may be too sensitive. Approximately 15% of individuals who underwent highly standardized testing and 35% of individuals who underwent routine testing would have been characterized as having a clinically significant change according to the current ATS/ERS guidelines, with no other indication of a change in lung function. Therefore, it is reasonable to define a less restrictive threshold (e.g., 20% change from baseline $D_{L_{CO}}$) to determine a clinically meaningful change. Using a 20% threshold would reduce the percentage of individuals misclassified as having a clinically significant change to 1% in the highly standardized testing group and 10% among those undergoing routine testing, providing a better estimate of clinically important changes between two test sessions. However, expanded criteria for normal variation will reduce sensitivity of the test to detect clinically meaningful changes. Recognizing that practitioners

may desire different thresholds in different clinical situations, our study provides the percentage of individuals misclassified as having a clinically significant change across a spectrum of thresholds ranging from 10 to 25%.

It should be noted that, although the participants included in this study were free of overt lung disease, all had underlying diabetes mellitus. Some have speculated that diabetics may be prone to microvascular injury of the lung, similar to that seen in the kidney (31, 32). Such a process could potentially affect the intersession variability in $D_{L_{CO}}$ measurements. Although it is possible that this situation may impair the ability to generalize our findings to a population free of any systemic illnesses, the mean baseline $D_{L_{CO}}$ in our study population was within the normal range, and each participant acted as his or her own control over the study period. In addition, the mean change over the short period of observation did not differ from zero, suggesting that the populations did not experience a significant decline or improvement in their lung function. These factors should minimize any potential impact that diabetic microangiopathy of the lung might have on our interpretation. One might also speculate that shifts in pulmonary blood volume due to changes in glycemic control might increase the variability of $D_{L_{CO}}$ in patients with diabetes, but such an effect has yet to be established. Although our study included a follow-up observation period of only 3 to 6 months, analysis of 2-year data from two protocols included in our study (studies 1022 and 1029) demonstrated an annualized change in $D_{L_{CO}}$ of -0.222 ± 0.073 ml/minute/mm Hg for patients with type 1 diabetes (study 1022) and -0.292 ± 0.071 ml/minute/mm Hg in patients with type 2 diabetes (study 1029). These observed annual changes are consistent with what has been reported in nondiabetic patients (33), and counter the possibility of the presence of a diabetic lung syndrome accounting for the variability observed in the present study. Further studies are necessary to determine whether these measures of intersession variability would be similar in individuals with lung disease. When reanalyzing our population after excluding the 5% of individuals with mild stable asthma or COPD, we found no change in the observed intersession variability in either testing method. We suspect that in the more severely diseased lung, intersession variability would be even higher, thereby necessitating more liberal demarcation of a significant percentage of change in $D_{L_{CO}}$.

Regardless of whether our results apply to other populations of patients with chronic nonpulmonary diseases, these results can clearly be applied to the interpretation of monitoring of inhaled insulin in diabetic populations if this therapy becomes available. Current U.S. prescribing information for Exubera recommends measuring $D_{L_{CO}}$ but does not provide guidance as to the magnitude of change that would require further action. On the basis of our study, when using routine PFT methods, we suggest a 20 to 25% change in intersession $D_{L_{CO}}$ as a more useful threshold for determining a clinically meaningful change in intersession $D_{L_{CO}}$, because this cutoff would minimize the percentage misclassified to 5 to 10%.

In summary, we have shown that the expected variability in $D_{L_{CO}}$ measurements over time in stable individuals is substantial and depends on the method of testing used. Using the percentage of change from baseline, rather than absolute change, minimizes the effect of baseline $D_{L_{CO}}$ on the assessment of intersession variability. Finally, by studying a large population of individuals with diabetes, we have characterized the expected variability in intersession change in $D_{L_{CO}}$ measurements using both highly standardized and routine testing conditions. The current recommendation of a clinically significant change, defined as 10% from baseline, appears too stringent and may misclassify as many as one out of three normal subjects as abnormal. On the basis of these findings, we propose that a 20 to 25% change from baseline

$D_{L_{CO}}$ is a more reasonable threshold when using routine PFT methods, thereby allowing the practicing physician to better estimate the clinical significance of an observed change in intersession single-breath diffusing capacity.

References

1. Martinez FJ, Flaherty K. Pulmonary function testing in idiopathic interstitial pneumonias. *Proc Am Thorac Soc* 2006;3:315–321.
2. Nakano M, Hasegawa H, Takada T, Ito S, Muramatsu Y, Satoh M, Suzuki E, Geiyo F. Pulmonary diffusion capacity in patients with systemic lupus erythematosus. *Respirology* 2002;7:45–49.
3. Takeda S, Yasunobu F, Kadota Y, Koma M, Maeda H, Kawamura S, Matsubara Y. Fall in diffusion capacity associated with induction therapy for lung cancer: a predictor of postoperative complications? *Ann Thorac Surg* 2006;82:232–236.
4. Stam H, Splinter T, Versprille A. Evaluation of diffusing capacity in patients with restrictive lung disease. *Chest* 2000;117:752–757.
5. Proudman S, Stevens W, Sahhar J, Celermajer D. Pulmonary arterial hypertension in systemic sclerosis: the need for early detection and treatment. *Int Med J* 2007;37:485–494.
6. Jenson RL, Teeter JG, England RD, White HJ, Pickering EH, Crapo RO. Instrument accuracy and reproducibility in measurements of pulmonary function. *Chest* 2007;132:388–395.
7. Jenson RL, Teeter JG, England RD, Howell HM, White HJ, Pickering EH, Crapo RO. Sources of long-term variability in measurement of lung function. *Chest* 2007;132:389–402.
8. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, Coates A, van der Griten CPM, Gustafsson P, Hankinson J, *et al.*; American Thoracic Society/European Respiratory Society Task Force. Standardisation of lung function testing: interpretative strategies for lung function testing. *Eur Respir J* 2005;26:948–968.
9. Hathaway E, Tashkin D, Simmons M. Intraindividual variability in serial measurements of DLCO and alveolar volume over one year in eight healthy subjects using three independent measuring systems. *Am Rev Respir Dis* 1989;140:1818–1822.
10. Wanger J, Irvin C. Comparability of pulmonary function results from 13 laboratories in a metropolitan area. *Respir Care* 1991;36:1375–1382.
11. American Thoracic Society. Single-breath carbon monoxide diffusing capacity (transfer factor): recommendations for a standard technique—1995 update. *Am J Respir Crit Care Med* 1995;152:2185–2198.
12. Punjabi NM, Shade D, Patel AM, Wise RA. Measurement variability in single-breath diffusing capacity of the lung. *Chest* 2003;123:1082–1089.
13. Hollander PA, Blonde L, Rowe R, Mehta A, Milburn J, Hershon K, Chiasson J, Levin S. Efficacy and safety of inhaled insulin (Exubera) compared with subcutaneous insulin therapy in patients with type 2 diabetes: results of a 6-month, randomized, comparative trial. *Diabetes Care* 2004;27:2356–2362.
14. Quattrin T, Bélanger A, Bohannon NJV, Schwartz SL; for the Exubera Phase III Study Group. Efficacy and safety of inhaled insulin (Exubera) compared with subcutaneous insulin therapy in patients with type I diabetes: results of a 6-month, randomized, comparative trial. *Diabetes Care* 2004;27:2622–2627.
15. Skyler JS, Weinstock RS, Raskin P, Yale J, Barrett E, Gerich J, Gerstein H; for the Inhaled Insulin Phase III Type I Diabetes Study Group. Use of inhaled insulin in a basal/bolus insulin regimen in subjects with type 1 diabetes: a 6-month, randomized, comparative trial. *Diabetes Care* 2005;28:1630–1635.

16. Rosenstock J, Zinman B, Murphy LJ, Clement SC, Moore P, Bowering CK, Hendler R, Lan S-P, Cefalu WT. Inhaled insulin improves glycemic control when substituted for or added to oral combination therapy in type 2 diabetes: a randomized, controlled trial. *Ann Intern Med* 2005;142:549–558.

17. Skyler JS, Cefalu WT, Kourides IA, Landschulz W, Balagtas C, Cheng S, Gelfand R; for the Inhaled Insulin Phase 2 Study Group. Efficacy of inhaled human insulin in type 1 diabetes mellitus: a randomized proof-of-concept study. *Lancet* 2001;357:331–335.

18. Cefalu WT, Skyler JS, Kourides IA, Landschulz W, Balagtas C, Cheng S, Gelfand R; for the Inhaled Insulin Study Group. Inhaled human insulin treatment in patients with type 2 diabetes mellitus. *Ann Intern Med* 2001;134:203–207.

19. Weiss SR, Cheng S-L, Kourides IA, Gelfand R, Landschulz W; for the Inhaled Insulin Phase II Study Group. Inhaled insulin provides improved glycemic control in patients with type 2 diabetes mellitus inadequately controlled with oral agents: a randomized controlled trial. *Ann Intern Med* 2003;163:2277–2282.

20. DeFronzo RA, Bergenstal RM, Cefalu WT, Pullman J, Lerman S, Bode BW, Phillips LS; for the Exubera Phase III Study Group. Efficacy of inhaled insulin in patients with type 2 diabetes not controlled with diet and exercise. *Diabetes Care* 2005;28:1922–1928.

21. Barnett AH, Dreyer M, Lange P, Serdarevic-Pehar M. An open, randomized, parallel-group study to compare the efficacy and safety profile of inhaled human insulin (Exubera) with metformin as adjunctive therapy in patients with type 2 diabetes poorly controlled on a sulfonylurea. *Diabetes Care* 2006;29:1282–1287.

22. Barnett AH, Dreyer M, Lange P, Serdarevic-Pehar M. An open, randomized, parallel-group study to compare the efficacy and safety profile of inhaled human insulin (Exubera) with glibenclamide as adjunctive therapy in patients with type 2 diabetes poorly controlled on metformin. *Diabetes Care* 2006;29:1818–1825.

23. Wise RA, Teeter JG, Jensen RL, England R, Schwartz P, Giles D, Ahrens R, MacIntyre N, Riese R, Crapo R. Standardization of the single-breath diffusing capacity (DLCO) in a multicenter clinical trial. *Chest* 2007;132:1191–1197.

24. Skyler JS, Jovanovic L, Klioze S, Reis J, Duggan W; for the Inhaled Human Insulin Type 1 Diabetes Study Group. Two-year safety and efficacy of inhaled human insulin (Exubera) in adult patients with type 1 diabetes. *Diabetes Care* 2007;30:579–585.

25. Heise T, Bott S, Tusek C, Stephan J-A, Kawabata T, Finco-Kent D, Liu C, Krasner A. The effect of insulin antibodies on the metabolic action of inhaled and subcutaneous insulin: a prospective randomized pharmacodynamics study. *Diabetes Care* 2005;28:2161–2169.

26. Lange P, Klioze S, Foyt H, Ogawa M, St. Aubin L, Duggan W, Riese R, Teeter J; for the Exubera 1029 Study Group. Pulmonary function of inhaled human insulin (Exubera®) in patients with type 2 diabetes: results from a 2-year study using standardized methodology [abstract E4257]. *Eur Respir J* 2006;28:742s.

27. Norwood P, Dumas R, Cefalu W, Yale JF, England R, Riese R, Teeter J. Randomized study to characterize glycemic control and short-term pulmonary function in patients with type 1 diabetes receiving inhaled human insulin (Exubera). *J Clin Endocrinol Metab* 2007;92:2211–2214.

28. Miller A, Thornton JC, Warshaw R, Anderson H, Teirstein A, Selikoff I. Single-breath diffusing capacity in a representative sample of the population of Michigan, a large industrial state: predicted values, lower limits of normal and frequencies of abnormalities by smoking history. *Am Rev Respir Dis* 1983;127:270–277.

29. Crapo RO, Morris AH, Clayton PD, Nixon C. Lung volumes in healthy nonsmoking adults. *Eur Physiopathol Respir* 1982;18:419–425.

30. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;159:179–187.

31. Ljubic S, Metelko Z, Car N, Roglic G, Drazic Z. Reduction of diffusion capacity for carbon monoxide in diabetic patients. *Chest* 1998;114:1033–1035.

32. Guvener N, Tutuncu NB, Akcay S, Eyuboglu F, Gokcel A. Alveolar gas exchange in patients with type 2 diabetes mellitus. *Endocr J* 2003;50:663–667.

33. Sherrill DL, Enright PL, Kaltenborn WT, Lebowitz MD. Predictors of longitudinal change in diffusing capacity over 8 years. *Am J Respir Crit Care Med* 1999;160:1883–1887.